Institutional Data Management in Higher Education

Ronald Yanosky, ECAR

Volume 8, 2009
This research study is available online at the ECAR website (www.educause.edu/ecar). The content of this study is restricted to authorized ECAR subscribers and to those who have separately purchased this study. The username and password below are required to gain access to the online version and are to be used only by those who may legally access the content.

Username: ERS0908
Password: DataMgmt0912
Institutional Data Management in Higher Education
EDUCAUSE is a nonprofit association whose mission is to advance higher education by promoting the intelligent use of information technology.

The mission of the EDUCAUSE Center for Applied Research is to foster better decision making by conducting and disseminating research and analysis about the role and implications of information technology in higher education. ECAR will systematically address many of the challenges brought more sharply into focus by information technologies.

Copyright 2009 EDUCAUSE. All rights reserved. This ECAR research study is proprietary and intended for use only by subscribers and those who have purchased this study. Reproduction, or distribution of ECAR research studies to those not formally affiliated with the subscribing organization, is strictly prohibited unless prior written permission is granted by EDUCAUSE. Requests for permission to reprint or distribute should be sent to ecar@educause.edu.
# Contents

**Foreword** ............................................................................................................................... 5
The Challenges of Digital Data ◆ Much Ground to Cover and Much Variation in Coverage ◆ What Can We Conclude from This Report? ◆ Too Many People to Thank

**Chapter 1** Executive Summary ............................................................................................... 11
Defining Institutional Data Management ◆ Methodology ◆ Key Findings ◆ Conclusion

**Chapter 2** Introduction and Methodology.................................................................................. 23
Data Management Concepts and Definitions ◆ Study Scope and Objectives ◆ Research Approach ◆ Overview of Respondents ◆ Study Organization

**Chapter 3** The Data Management Environment......................................................................... 33
Key Findings ◆ Data Growth: Volume and Variety ◆ Central IT Disk Storage Capacity and Growth ◆ Identity Management ◆ Data Management Drivers and Barriers ◆ Organizational Context ◆ Surfing, Swimming, or Drowning?

**Chapter 4** Data Quality and Analytics ....................................................................................... 49
Key Findings ◆ Assessing Enterprise Data Quality ◆ Data Quality Improvement Initiatives ◆ Data Analytics ◆ Views from the Qualitative Research: Toward a Data-Driven Culture ◆ Data Quality and Analytics: How Can We Do Better?

**Chapter 5** Data Stewardship, Security, and Policies ................................................................. 75
Key Findings ◆ Business/Academic Unit and Central IT Data Responsibilities ◆ Data Stewardship Policies ◆ Data Use, Security, and Access Policies ◆ Classifying Data by Risk or Sensitivity ◆ Data Breach Notifications ◆ Data Stewardship, Formal and Cultural

**Chapter 6** Institutional Content Management ............................................................................ 93
Key Findings ◆ Institutional Content Management: An Overview ◆ The Institutional Website ◆ Electronic Records Management ◆ Facing the Content Challenge

**Chapter 7** Research Data Management .................................................................................... 119
<table>
<thead>
<tr>
<th>Chapter 8</th>
<th>Data Management Outcomes........................................................................................................ 141</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Key Findings • Assessing Data Management Outcomes: An Overview • Outcomes Factors: Broad Impact • Outcomes Factors: Targeted Impact • Toward Better Data Management</td>
</tr>
<tr>
<td>Chapter 9</td>
<td>Institutional Data Management in an Age of Superabundant Information .............................. 155</td>
</tr>
<tr>
<td></td>
<td>Data Quality • Institutional Content Management • Research Data • Governing Institutional Data Management</td>
</tr>
<tr>
<td>Appendix A</td>
<td>Institutional Respondents to the Online CIO Survey............................................................... 167</td>
</tr>
<tr>
<td>Appendix B</td>
<td>Interviewees in Qualitative Research ......................................................................................... 171</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Supplementary Tables .................................................................................................................. 173</td>
</tr>
<tr>
<td>Appendix D</td>
<td>Bibliography ............................................................................................................................ 177</td>
</tr>
</tbody>
</table>
Foreword

The EDUCAUSE Center for Applied Research (ECAR) was launched on January 1, 2002, to create a body of research and analysis on important issues at the intersection of higher education and information technology (IT). ECAR is fulfilling its mission through a program of symposia and through the publication of biweekly research bulletins, detailed research studies, occasional papers, executive roadmaps, and case studies. These publications are designed to highlight effective practices, lessons learned, and other insights from the practical experience of campus leaders. Since ECAR’s inception, 14 symposia have been held, and more than 400 research publications have been issued.

The Challenges of Digital Data

The area of data management is particularly dear to me. Few people may know that unlike many, I began my career in IT management through the information resources management career pathway. Although most baby boomers in our community began their careers in programming, computer operations, production control, and other aspects of data processing, and although increasing numbers of today’s IT leaders are coming from the professoriate or other core constituencies of higher education, some of us come from the library, from records management, from archives administration, and from other cognate fields that deal more with the management and use of information than of technologies per se. In my case, I worked first as an archivist and then a records manager and was professionally preoccupied with the life cycle of information, regulatory compliance, information flows, information asset management, and the preservation of a meaningful and accurate historical record.

The introduction of computers and, later, networks has made information management at once harder and easier. Paper-based information artifacts take up a lot of space, but they can be relatively easily “marked” and “parked” for future retrieval. Further, our inability to control the management and use of individual words on individual pages made it possible for us to socialize administrative expectations that were in truth blunt. You were to either keep or discard all correspondence written between one date and another, pursuant to an institutional policy, or a local, state, or federal regulation. Records managers in the late 20th century could not find proverbial needles, so we contented ourselves with the management of haystacks instead.

As information became digital, the very basis of information management became
Institutional Data Management

unclear. Documents have a clear structure, and their management can be and is defined and prescribed by that structure. Dimensions of document structure include date of origin, author, recipient, document type, length, and so forth. Information in digital form can be structured or unstructured. And even when it is structured, the dimensions that define this structure can be problematic. What, for example, is the unit of data management for the purposes of retention, destruction, storage, labeling, and so forth? Is it the data element? If we alter data elements, do we inadvertently obscure or even obliterate the meaning of other data elements that together constitute a greater information whole?

The digitization of information is dramatically changing the landscape of its management. Information that is digital is easy and relatively cheap to replicate. What is the record copy of information? The easy replication, transport, and storage of data have conspired to make information superabundant. What does superabundance of information mean in the context of laws, regulations, and practices that seek to create an accurate and coherent record of human activity? How many times have we seen the proverbial digital smoking gun disappear from the record only to be resurrected from an audit log, a backup file, or from files completely outside a coherent record system?

As we have undertaken to increase the human capacity to create, move, and store information, so have the media of storage available to us become more and more transitory. Cuneiform impressions on clay, carvings on cave walls, hieroglyphs on papyri, or inks on animal skins have endured for centuries or even millennia, whereas we struggle to read the digital record of the NASA Viking missions of 1976.

Digital techniques have lowered the cost of retaining information, giving rise to new economics of information. In the paper world, today’s news was tomorrow’s fish wrap. All but the most important historical documents—including most books—marched through a life cycle that imbued them with great value at their birth and then eroded that value quickly over time as owners and publishers moved to liberate precious real estate from the demands of “inventory.” In the digital context, multiple copies of data are regularly created for any number of reasons (for example, caching to accelerate search). As a result, whole businesses are arising to understand and exploit the residual value of information in the so-called long tail.

Digital information is not only relatively ephemeral, it can also be altered by a variety of people. Ink on paper can be altered as well, but not easily and rarely without a trace. The ephemeral nature of digital information contributes to making its creation, management, and use a “team sport.” Archivists, in particular, and historians have long struggled with the concept of provenance—the lineage of a document. In the digital context, data can be mashed up and easily altered and can appear in a multiplicity of contexts. Understanding provenance, and therefore the accuracy, reliability, authenticity, and credibility, of information sources is made more and more complex and problematic. Not surprisingly, this problem is especially vexing in academic contexts. Do our historical concepts of sole authorship, originality, or plagiarism hold in this digital context?

The challenges of digital data go well beyond these few areas and examples. The complexity associated with managing data is increasing faster than our capacity to keep up.

This ECAR study chronicles the current state of the race between the capabilities that we now enjoy to create, manipulate, store, disseminate, and find information and our ability to organize, contextualize, interpret, or preserve that information.
**Much Ground to Cover and Much Variation in Coverage**

ECAR principal investigator Ron Yanosky has looked closely at five aspects of data management:

- data integrity and quality,
- analytics,
- data stewardship and security,
- content and records management, and
- management of research data.

Not surprisingly, Yanosky found tremendous variation in survey responses across these aspects of the problem. As the reader would expect, survey respondents from research-intensive universities expressed far greater concerns about an impending data deluge than did counterparts from associate’s or bachelor’s institutions. Notwithstanding many of the findings that corroborate our expectations, several stand out:

- The good news is that respondents are generally in agreement that their institutions have today and will continue to have in three years the IT infrastructure they need to manage data in administrative systems, learning management systems, web content, e-mail, research systems, video and audio files, and digital library systems. This confidence answers the question, Can you create it, move it through networks, compute it, find it, and store it?

- The less good news is that despite the progress that has been made in the past 20 years to bring the campus CIO to the cabinet level, managing the “I” in CIO continues to be a challenge. Respondents to this survey were not confident that the data in their systems are accurate, well defined, or synchronized properly across systems, nor were they confident that the processes to ensure things such as data quality are in place. Among our respondents the formal assignment of responsibility for managing data resources to data stewards was “the exception rather than the rule.”

- In administrative contexts we are still not a culture of evidence. On balance, our research participants described environments that are data rich but information poor. Poor data quality, a paucity of analytical tools, and a workforce that is largely unschooled in quantitative analytical methods conspire to yield a picture of relatively low adoption of analytics in higher education. Further, in cases where responses could be compared with those made to identical questions asked in 2005, we can only conclude that little if any progress has been made in this area of activity. These findings are in sharp relief to the vocal call among state and federal lawmakers and regulators to use data in more sophisticated ways to support decision making, student success, and other purposes.

- Not surprisingly, IT security—an area of rather significant investment over the past five years—is one of the areas that survey respondents were most satisfied with.

- Research university respondents were nearly unanimous in their expectations that research data volumes will grow significantly in the next three years. Keeping up with a possible deluge appears to be considered an economic issue more than anything.

- Survey respondents do not believe that their institutions are getting maximum academic and business value from the information they have. In fact, nearly half of the survey respondents disagreed or strongly disagreed with the statement “We get maximum academic (business) value from institutional data.” Although the standard
set in this question is a high one, it is clear from other responses that the situation regarding “value capture” is unsatisfactory.

**What Can We Conclude from This Report?**

We may certainly conclude that many of our IT leaders are charged as—or otherwise behave as—chief technology officers rather than chief information officers. The state of data management in this sample of higher education institutions is mixed. Most respondents appeared to be reasonably strong where it comes to managing an infrastructure for data and to securing that data. By the same token, many respondents were clear that data quality is lackluster, that little effort is made to mine institutional data to promote better institutional outcomes, and that although there is a lot of data to manage, the institutions are not getting full value currently from the data they have.

These findings leave me and likely many readers confused. Is this a scandal? A long-standing expectation? A critical weakness? Is higher education substantially different in this arena of endeavor than corporate organizations, governments, or others? Should these findings be the basis of a clarion call to action, or are they a reflection of a broader and ultimately sensible set of priorities?

My instinct on reading this fine report is that effective data management depends on ascertaining the dependencies among the aspects studied here. At bedrock, if regulators, educational leaders, politicians, and others want to use information to promote student success, to lower rates of student attrition, to improve institutional decisions, and to promote transparency, then attention must be paid first and foremost to data quality. Are the data in college and university systems untrustworthy, and if so, why? Building a superstructure of complex data mining, analytics, simulation, and benchmarking capabilities is likely wasteful at best if the underlying data cannot be trusted. Creating meaningful data quality is an enterprise activity that is reasonably well understood. It is also an activity that is rife with politics. EDUCAUSE clearly has an opportunity here to transmit best practices in the areas of data governance, knowledge bases, data architecture, quality assurance and management, and so forth. Campus IT leaders will need to build social networks that will make it possible to rationalize institutional dictionaries and to re-architect accidental management environments that contribute to low data quality. Yanosky provides helpful and practical summaries—with practical advice in every chapter.

**Too Many People to Thank**

The ECAR study of data management was an ambitious undertaking, and of course there are too many people to thank. Yanosky is a master at making sense of very complex data and at keeping his eyes fixed on big and actionable questions, even while wrestling with large volumes of cacophonous data. ECAR Fellows Susan Foster, Judith A. Pirani Pirani, Gail Salaway, and Mark Sheehan conducted qualitative research to gain deeper insights into the research data and to capture additional ideas and viewpoints for the study.

To enrich this core study, we are able to add two case studies. The first, *Revitalizing Data Stewardship through Risk Reduction: Managing Sensitive Data at the University of Virginia*, written by Bob Albrecht and Judith A. Pirani, presents the institution’s multilayered strategy that remediates past data usage practices, creates an updated data policy framework, and better protects university data by strengthening the IT infrastructure security. Pirani and Donald Z. Spicer delve into technical and archival issues surrounding the long-term digital preservation of research data in the second case study, about the
Chronopolis Digital Preservation Initiative, a multi-institutional grid-based project involving the San Diego Supercomputer Center, the University of California San Diego Libraries, the National Center for Atmospheric Research, and the University of Maryland’s Institute for Advanced Computer Studies. Many people graciously took time out of their busy schedules to assist us with these case studies, most notably James Hilton, Vice President and CIO, and Shirley Payne, Assistant Vice President for Information Security, Policy, and Records, at the University of Virginia; Brian Schottlaender, University Librarian, and Ardys Kozbial, Technology Outreach Librarian, University of California San Diego Libraries; and David Minor, Head of Curation Services, San Diego Supercomputer Center.

We are especially grateful to our colleagues at a number of institutions: Shelton Waggener, Associate Vice President and CIO at the University of California at Berkeley, and Bruce Maas, CIO at the University of Wisconsin–Milwaukee, provided much-appreciated advice about institutional data management strategy and practices. James Phelps, Senior Information Technology Architect at the University of Wisconsin–Madison, helped us understand the role of the data architect and current practices in that area, and Shirley Payne generously contributed to our understanding of data stewardship and security policy. Our ECAR colleagues Donald Spicer and Susan Foster, experienced CIOs both, reviewed our survey questions and made suggestions that greatly improved the study. Finally, Julie Curtis and her colleagues at SunGard Higher Education provided valuable insights about the technology and cultural issues affecting academic analytics and data management.

Of course, after all of this work on content development is complete, the work of the production team begins. The care of our investigators and fellows in constructing and designing surveys and in analyzing responses and checking analyses is matched by a team of editors under the guidance of Gregory Dobbin and Nancy Hays. They are thorough people and work with a team of editors, proofreaders, digital compositors, and printers. In studies where a quarter-inch shift in a column can obliterate a careful analysis, one cannot understate the effort these people make or the successes they claim. And last, but of course not least, Toby Sitko resides at the interface of the research team and the production team and orchestrates the overall project with the skill of a symphony conductor. ECAR depends on her every day.

Richard N. Katz
Boulder, Colorado
Executive Summary

Not for nothing is the modern era frequently called the Information Age. Computing and storage hardware, following Moore’s law for more than half a century, has brought us the 32-gigabyte pocket device and the multiterabyte personal computer. We don’t seem to have any difficulty at all finding new ways to fill these ever-fattening data piggy banks. The IT research firm IDC, taking a shot at the admittedly elusive question of how big the “digital universe” is, suggests a suitably astronomical size: 281 exabytes (billions of gigabytes) of data created, captured, and replicated in 2007 alone, growing at 60% annually.¹

Higher education has played a big role in making this data “big bang” possible, and the purpose of this study is to examine its effects on colleges and universities and the ways their IT organizations deal with it. The challenges facing institutions may be understood by considering three broad areas of data impact.

First, like all big organizations running a complicated business, higher education institutions consume and generate a lot of operational information. Digital use of this kind of information goes back to the days when the 80-column punch card symbolized “data processing,” and many of that era’s classic problems remain with us today: making sure data is accurately recorded, kept up to date and in sync with business processes, and available for use beyond the narrow operational need it was originally created to serve. For the most part, this kind of data is structured, i.e., modeled according to a rigorous scheme that defines its size, its type, and its legitimate values. Though modern relational databases and enterprise applications suites have made operational use of business information much speedier and more flexible, they have not been as successful at opening it up for broader analysis and management use. It remains much easier today to place the right student in the right course than to use aggregated information about students and courses to do enrollment and curriculum planning. Yet funders and accreditors increasingly ask for metrics that permit commensurate comparisons of unit and institutional performance, and higher education reformers are envisioning a world of outcomes-based assessment and easily exchanged student records. To the extent that these trends continue, the quality of operational data and the ability to use data to analyze and manage will become ever more significant.

A second domain of data challenge is the enormous body of data known as content. Fed by the intersection of new networked modes of delivering information (above all, the web) and by the shift of all kinds of media

¹Executive Summary
toward digital formats, content is a necessarily vague category yet an essential one to manage effectively. At a modern university, content includes not only the material available on the institution’s website(s) but also documents such as memoranda and spreadsheets, its publications, the institution’s online teaching materials, a growing proportion of its library and museum collections, and much else besides. Most content is unstructured, meaning that it contains more or less free-form information rather than the sort of modeled data consumed in a typical business application. Because content is often formatted for use in website templates or for recall from a database, however, it increasingly takes the form of semistructured data. Much content—a policy statement, a website describing academic programs, a course catalog entry—requires some kind of official approval; much of it also needs to be formally described and cataloged in order to be trackable and searchable and available when needed. And much of it is ephemeral junk. The great challenge of institutional content management is to provide the tools that allow the right people to create, publish, find, and preserve or winnow the right content according to the needs of the institution. Yet the technology for handling unstructured or semistructured data remains far less mature than that for handling structured data.

A final domain of the higher education data challenge is research data: Whether from experiment or computation, modern science is producing enormous quantities of it. Digital data is more portable and searchable than traditional forms of data capture—so much so that some science visionaries, backed by funding agencies, believe that the time has come for a new era of shared data cyberinfrastructure and open data access. Yet besides sharing many of the problems of operational business data and content, research data raises unique problems relating to ownership, preservation, and interpretation. Solving these problems could enable a dramatic increase in research productivity; failing to solve them could mean that much of the data now being collected with such creativity and effort could be lost through deterioration or simply because nobody knows it exists.

Our capacity to produce data and expose it over networks has far outstripped our ability to reform or replace the business, legal, and cultural practices that defined our relationship to information in the era before data superabundance. As a result, requirements relating to new forms of data are piled onto ongoing challenges from more familiar forms, while government and institutional leadership subject data to new kinds of regulation, often in a reactive way not very well informed by an appreciation of data management principles.

Our study looks at how higher education institutions are responding to these challenges, with particular focus on institutional content, research data, and operational business/academic data and its analysis. Our aim has been to provide institutional CIOs and other concerned parties with information about the state of higher education data management and to identify practices associated with good data management outcomes.

**Defining Institutional Data Management**

In the context of this study, *institutional data management* refers to the policies and practices by which higher education institutions effectively collect, protect, and use digital information assets to meet academic and business needs. It’s worth noting that though our overall guiding concept is data management, our definition refers broadly to information about the state of higher education data management and to identify practices associated with good data management outcomes.
practical concerns of institutional CIOs will lose more than it gains by insisting on too sharp a distinction between the two. We believe that the term institutional data management is elastic enough to reflect our interest in enterprise information needs while speaking the language of working IT professionals.

**Methodology**

Our study of institutional data management in higher education took a multipart approach that consisted of:

- a literature review to examine data management frameworks and definitions, to define issues and establish research questions;
- consultation with higher education IT administrators and data management experts to identify and validate survey questions;
- a quantitative web-based survey of EDUCAUSE member institutions that received 309 responses, 78% of which were from the institutional CIO or equivalent;
- qualitative interviews with 23 higher education IT leaders and staff; and
- two case studies, one examining a sensitive-data risk reduction initiative at the University of Virginia, and one concerning Chronopolis, a multi-institutional grid-based project for the long-term preservation of research data.

Certain questions in our quantitative survey were identical to questions posed in several earlier ECAR surveys. Where applicable, we carried out a longitudinal analysis of responses to these questions on the basis of institutions that had responded to both the earlier and the current survey.

**Key Findings**

Our study was chiefly concerned with how institutional IT units are coping with an explosion in the volume and variety of data that they have to manage, and how well they think they’re doing at such data management goals as keeping data secure and getting maximum academic and business value from it. We drew eclectically (and selectively) from a number of data management frameworks and standards to identify topics of research interest. Areas of survey coverage, which map roughly to the organization of this study report, included the following:

- **Institutional context**—size and scale of the institutional data environment, data-related staffing, institutional drivers, and barriers to data management.
- **Data integrity and quality**—planning and control activities that measure and ensure the fitness of data for use, respondent assessment of institutional data quality, and management of “master” data.
- **Analytics**—processes to provide decision support (aka business intelligence) for institutional leaders and other campus management and staff; and supporting infrastructure, including data warehouses, data marts, and related tools.
- **Data stewardship and security**—assignment of responsibilities for the security and quality of data; exercise of authority and decision making over the management of data assets and creation of data policy; and related data policies.
- **Content and records management**—infrastructure and practices to store, access, and (as applicable) archive data within content systems (e.g., web, e-learning, documents) and institutional records systems; and assessment of the institution’s compliance with records retention requirements.
- **Research data management**—distribution of research-related data responsibilities among institutional units, research data ownership and other
research-related policies, and institutional ability to support the long-term preservation of research data.

- **Data management outcomes**—respondent assessment of institutional performance on summary measures of data management effectiveness.

In the following sections, we summarize and synthesize our main findings.

**Drowning in Data?**

Clichés about today’s data environment often invoke catastrophic images involving water. But are higher education institutions “drowning” in data, as is often said? Our respondents certainly reported high levels of data growth. When we asked them to characterize changes in the volume of institutional data in the past 12 months using a scale from 1 (greatly decreased) to 5 (greatly increased), the mean response for seven out of eight categories of data was at or slightly above 4 (increased). The fastest-growing types of data by volume were learning management systems data and content (4.22) and e-mail (4.18), and the slowest-growing was research data (3.74). The apparently low figure for research, however, was chiefly an effect of combining research-oriented institutions (4.27) and teaching-oriented ones (3.52).

Respondents were fairly positive in assessing whether their institutions have the infrastructure needed to effectively manage various kinds of data (see Figure 1-1). Except for research data and video/audio content available online—two types of data that are particularly hungry consumers of data storage and bandwidth—respondents tended toward agreement with this proposition, and they were still more positive about their infrastructure in three years. About half of the central IT units among our respondent institutions managed a total disk storage capacity of 20 terabytes or less, and the median estimated growth in that

![Figure 1-1. Institution Has Infrastructure Needed to Manage Institutional Data, Now and in Three Years](image)

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree*
capacity over the next 12 months was 20%. In short, though our respondents report data volume growth and plenty of it, they don’t seem to indicate that they’re “drowning,” at least in terms of capacity and infrastructure.

So where are the pain points? Lack of funds was by far the barrier to investment in institutional data management that our respondents most often named, followed by lack of staff expertise and a decentralized or informal institutional culture. As for investment drivers, institutional “sticks” (improving data security and regulatory compliance) stood somewhat ahead of “carrots” (improving business/academic unit operational efficiency and enhancing data-driven decision making).

Identity management is a crucial enabler of good data management because so much about data security and appropriate use hinges on the ability to authenticate and authorize users. Our respondents, however, averaged only an uninspiring “neutral” response when we asked whether they agreed that their institutions’ identity management provided sufficiently granular data access authorization.

**Enterprise Data Quality**

What makes data useful? Many things, no doubt, but common to most business uses of data would be its accuracy, along with confidence that it is well defined and stored in the right place, that its value is in sync everywhere it’s used, and that it resides in an environment attuned to good data quality.

One of the major themes of our study is that these questions of data quality seem to be foundational to many aspects of data management. And as Table 1-1 shows, when we asked respondents a series of data-quality-related questions pertaining to major data elements within their institutions’ major administrative systems and data stores (such as HR, finance, and student administration

---

**Table 1-1. Administrative Enterprise System Data Quality Measures**

<table>
<thead>
<tr>
<th>Characteristics of Major Data Elements</th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A system of record is identified for each major data element.</td>
<td>3.54</td>
<td>301</td>
<td>0.981</td>
</tr>
<tr>
<td>Each major data element has a single definition that is recognized across the institution.</td>
<td>3.29</td>
<td>308</td>
<td>1.117</td>
</tr>
<tr>
<td>Each major data element is coded consistently across systems and data stores.</td>
<td>3.12</td>
<td>301</td>
<td>1.075</td>
</tr>
<tr>
<td>When the value of a major data element changes, the change propagates across all enterprise systems and data stores that use it.</td>
<td>3.11</td>
<td>308</td>
<td>1.140</td>
</tr>
<tr>
<td>When the value of a major data element changes, the change propagates across all business/academic unit (“shadow”) systems and data stores that use it.</td>
<td>2.52</td>
<td>299</td>
<td>1.082</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Quality Processes</th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processes are in place to assure data quality at the point of capture/origin (e.g., data entry).</td>
<td>3.24</td>
<td>306</td>
<td>0.975</td>
</tr>
<tr>
<td>Automated systems are in place to validate data across enterprise systems and data stores.</td>
<td>2.89</td>
<td>302</td>
<td>1.035</td>
</tr>
<tr>
<td>Processes are in place for documenting and reviewing all identified data quality issues.</td>
<td>2.76</td>
<td>304</td>
<td>0.995</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
systems), they were far from enthusiastic. Mean responses on a scale from 1 (strongly disagree) to 5 (strongly agree) were for the most part only slightly above neutral, and in some cases were below neutral.

We averaged each institution’s responses to these questions to create a single enterprise data quality score. This score, which for all respondents averaged a middling 3 on our 5-point scale, proved to be positively associated with many other measures in the study, suggesting (though not proving) that better enterprise data quality is a factor in better data management outcomes. We also found some evidence that perceived enterprise data quality is susceptible to educational efforts: Where respondents agreed more strongly that executives, managers, and staff receive training in data quality issues, enterprise data quality scores tended to be higher.

Data Analytics

Modern administrative systems promise not just to handle real-time transactions such as paying a bill or enrolling a student in a course, but also to make the aggregate information collected by these systems available for institutional management and planning. How institutions approach reporting, modeling, analysis, and decision support—or more simply, analytics—was one of the key concerns of our study.

Effective use of analytics is often said to be as much a cultural matter as a technological one, because institutional executives and managers simply may not be accustomed to using empirical data to make decisions, or they may wish to restrict access to relevant data. Our respondents, however, were fairly positive about such cultural factors. Majorities of them agreed or strongly agreed with such propositions as that their institution’s leadership was committed to evidence-based decision making and that their institution was committed to making management information widely available.

At the same time, we found a fairly modest infrastructure for analytics. Of four different kinds of analytics-related data stores we asked about, including data warehouses, data marts, and operational data stores in each of two different usage modes, 41% said they had none of them in use, and another 18% reported only one. Roughly one-third reported data warehouses in use institution-wide, with a few more reporting them in use at the school, college, or department level. The incidence of most of the analytics infrastructure items we asked about was unchanged since ECAR’s 2005 survey on academic analytics.

As in that study, we found that the most common primary use of analytics tools by far was for extraction and reporting of transaction-level data. Advanced uses of analytics such as identifying at-risk students or auto-triggering business processes were widespread in terms of the percentage of institutions using them at least occasionally, but relatively few said they made use of them “usually” or “almost always.” Higher-frequency use of advanced analytics, however, proved to be positively associated with a number of good data management outcomes, as we report later in this chapter.

Data Stewardship and Security Policies

Authorities on data management stress that data should be “owned” by the people who know it best and have the best incentives to care for it and keep it secure. Our results relating to the division of data responsibilities suggest that this advice is being taken seriously at most of our respondent institutions. When we asked whether primary responsibility for various data-related items lay with business/academic units, central IT, or neither, we found a generally reasonable division of labor. Business and academic units most often had primary responsibility for data accuracy, fitness for purpose, and data access decisions, and central IT most often
Institutional Content Management

It's increasingly difficult to think of some aspect of institutional work that doesn’t create a digital document, spawn an entry on a website, get communicated by electronic means, or somehow contribute to the body of digital objects known loosely as “content.” Some portion of this digital content meets the business, regulatory, or historical significance criteria that make content into records. Our study looked at a number of facets of institutional content management, including the overall environment for handling content, management of and plans for the institutional website, and the institution’s ability to meet the requirements of records management.

IT analysts often promote the notion of enterprise content management, that is, integrated systems and processes for dealing with a wide variety of digital content relevant to business operations. Only about 12% of our respondents, however, said that their institution has an integrated enterprise content management solution managing the widest possible range of content (see Figure 1-2). Most often, they said that they have a mix of enterprise and local best-of-breed solutions varying by content type (36%), and about 40% reported no enterprise solutions for content, only local/departmental solutions or a generally ad hoc approach to digital content management. Looking ahead three years, most institutions expected to move in a more enterprise-oriented direction, either with an integrated enterprise content management system (32%), an enterprise best-of-breed–predominant environment (28%), or a mixed enterprise/local best-of-breed environment (31%).

Web content management was a particular focus of our survey. Respondents averaged roughly neutral responses to a statement that the institutional website routinely exceeds the expectations of constituents; they were most
positive with respect to public users (mean 3.23 on a scale of 1 to 5) and least positive about faculty users (mean 2.80). Perceptions of exceeding constituent expectations were associated with a number of web content management best practices that we asked about. For example, where respondents said that characteristics such as a database-driven web environment or consistent look and feel of web pages were more prevalent, they also tended to agree more strongly that their institutional website exceeded constituent expectations.

About half of institutions reported either an archivist or a records officer responsible for electronic records, a group responsible for overseeing electronic records management, or both. Respondents were not particularly positive about their institution’s ability to comply with records retention and disposition requirements throughout the institution; those who characterized that ability as poor or fair (53%) slightly outnumbered those who rated it good, very good, or excellent. Institutions that reported e-records retention schedules in more areas of content, and those with higher enterprise data quality scores, tended to rate this ability higher, but somewhat surprisingly we found no difference based on the existence of an archivist/records officer or a group overseeing electronic records management.

Research Data Management

The capacity of digital technologies to capture research information in a form that can be readily analyzed and shared inspired the National Science Foundation to observe in 2005 that “digital data collections are at the heart” of “fundamentally new approaches to research and education.” They are also, however, at the heart of an emerging debate about how to store, share, and properly manage this huge, rapidly accumulating, and highly diverse body of data. Our study examined the context of digital research data management at respondent institutions, with a particular focus on the long-term preservation of data.
We began by asking about the direction of change in the basic IT infrastructure of research, including the use of high-performance computing and networking and the amount of research data storage. Overall, about 39% of institutions reported that use of high-performance computing had increased or greatly increased in the previous three years, and slight majorities reported such increases in high-performance networking and research data storage. Increases were much more common among research-oriented institutions than among teaching-oriented ones, and expectations of increased use in the next three years were more dramatic still: 98% of research-oriented institutions expected to increase or greatly increase amounts of research data storage over that time period.

When we asked respondents how often various campus entities were involved in research data support activities, we found widespread and rather diffuse participation in different activities. The activities central IT most often supported to a large or very large extent were data storage and data backup/recovery—classic central IT tasks—though about one in five institutions said that central IT provided assistance with creating research data management plans and support in the selection and use of research tools. The role of individual investigators, labs, or teams in providing their own research data management support was considerably higher at research-oriented institutions than at teaching-oriented ones.

Respondent institutions did not provide much evidence of aggressively undertaking long-term research data preservation activities, though our results were somewhat diluted by an unusually large proportion of “don’t know” answers. Only about 31% of respondents said their institution assumes responsibility for archiving research data after investigator projects conclude, and most of these said the institution does so only in a few or some cases. A 41% “don’t know” response, however, suggests that many respondents were not acquainted with their institution’s activities in this area. When we asked whether respondents agreed that their institutions have the necessary commitment and funding mechanisms to support the long-term preservation of research data, their average answers fell well short of even a neutral (3) on our 5-point agreement scale. Two-thirds disagreed or strongly disagreed with the funding mechanisms item. Those institutions where central IT was more involved in providing research data management support tended to have somewhat more positive views.

Given these results, it’s not surprising that respondents were rather lukewarm when asked if they agreed that their institutions met the data-related needs of institutional researchers. The average response overall was almost exactly neutral, but doctoral institutions averaged well below that; their mean agreement was 2.54, about midway between a disagree and a neutral response. As with the long-term preservation items mentioned above, agreement about meeting the data-related needs of researchers tended to be higher where central IT was more involved in research data support activities.

Data Management Outcomes
Data management is an especially complex area of IT practice. To tie its many facets together and determine how well respondents think their institutions are achieving goals that good data management ought to serve, we asked them to state their level of agreement with a series of statements about data management outcomes (see Table 1-2).

Given the amount of angst that has been expressed about IT security in the past decade, we were surprised to find that institutions agreed most strongly with the statement that restricted or sensitive data was secure at their institutions. This was the only outcome that
came close to a mean agree (4) response. Respondents were also relatively positive regarding their institutions’ ability to support anticipated growth in the volume of data in the next three years—another item, given the “deluge” theme often associated with contemporary data management, that we might have expected to take a lower relative place in the outcomes.

By contrast, the two outcomes with the lowest average responses were the ones about getting maximum academic and business value from institutional data. Granted, these statements, by referring to “maximum” value, set a high standard, but it was striking that both fell well below even a neutral mean response. Nearly half of respondents disagreed or strongly disagreed with each of these statements.

Despite the diversity of concerns that our outcomes statements represented, we found a number of factors that were positively associated with most or all of them. Most interesting, and one of the critical findings of the study, was a strong relationship to enterprise data quality score: Institutions with higher scores tended to agree more strongly with all five of the outcomes items. These differences could be dramatic; for example, although agreement among institutions at the low end of our enterprise data quality score range averaged 2.01 on the maximum business value statement, those at the high end averaged 3.21.

Among the other factors we found positively associated with better outcomes measures were

- stronger agreement that the institution provides effective training to users of analytics tools (all outcomes except the data security item),
- an early adopter approach to technology adoption (as opposed to mainstream and late adopters), and
- stronger agreement that the institution has the necessary commitment to support long-term preservation of research data.

In addition, where institutions made frequent use of advanced analytics practices, and where they agreed more strongly that institutional leadership was committed to evidence-based decision making, they tended to report higher agreement about getting maximum academic and business value from institutional data.

### Conclusion

As with a lot of higher education IT concerns, data management has the aspect of a workaday task being handled competently by busy IT departments with a lot of other tasks to perform. We didn’t conclude that institutions are losing a desperate battle with a data deluge, though we did note that they seem to be more worried about handling the variety of data that the modern environment involves, as opposed
to its volume. And that makes sense: So far, at least, Moore’s law has held steady for the density of storage technologies, whereas our mastery of data complexity—figuring out, for example, how to search a video file or cost-effectively tag thousands of documents with useful metadata—is evolving a lot more slowly.

In fact, it was not the poster concerns of data volume and security that produced the most notable results of our study, but the more mundane and operational ones. Institutions gave themselves relatively weak marks for data quality, and their analytics infrastructure and use seem not to have changed much since we investigated those topics in 2005. Yet our summary enterprise data quality score turned up again and again as a positive factor in desirable data management outcomes: Institutions with higher data quality scores gave better marks to their identity management capabilities, their institution’s analytics training, their ability to manage all the varieties of data and digital content their institution needs, their websites’ ability to exceed the expectations of users, and all of the general outcomes shown in Table 1-2. Likewise, effectiveness of analytics training proved to be one of the most powerful factors in data management outcomes, perhaps because it speaks of a commitment to do something with data, and use of advanced analytics techniques was strongly associated with two of the most important—and lowest-rated—outcomes, getting maximum academic and business value from institutional data.

That is, after all, what it’s there for. In the Y2K era, higher education interest and investment in enterprise administrative systems surged, and at about the same time IT security and data privacy emerged as the great hot-button issues of IT administration. Each of these initiatives can claim its successes, even if they may be hard to recognize in the busyness and confusion of day-to-day IT operations. But it may be time to declare that success more explicitly and consider how institutions can work toward their goals by giving as much attention to data’s long-term value as to its real-time transactional value, and by investing in the ability to analyze data as well as the ability to lock it down.

To extend attention from security and transactional systems to include data quality and analytics would, we believe, help institutions get more business value from their structured institutional data. As for the much more amorphous yet vital realm of content, it seems only fair to note that if institutional content management capabilities seem pretty immature, so too do content management technologies themselves. Our study, however, suggests some practical steps that institutions can take toward taming the content beast: approaching content from an enterprise view, being diligent about identifying and documenting the subset of documents that needs special treatment as records, and (to return to our favorite theme) maintaining a high-quality data environment.

As for research data, institutional and indeed scholarly needs are still inchoate, and though we noted some outlines of an institutional response to these needs, it’s not at all clear that higher education institutions are committed to shouldering the burden of research data preservation, or even that they should. Just the same, we see some basis for IT attention to institutional research data needs, partly because our respondents on the whole seem to think their institutions are doing a mediocre job supporting their researchers, and partly because IT has skills—including data center administration, managing high-reliability applications, and large-scale data administration in an on-demand environment—that few other units on campus can boast of. At the very least, central IT can be an advocate for ensuring that the necessary expertise exists somewhere on campus, and IT can be a partner in contributing to its success.
Endnotes


In the second decade of the Internet era, the world is awash in a stunning quantity and variety of information. Exactly how much digital information we’re producing and living alongside is a tough question to answer, but there’s no doubt that the answer involves the kinds of numbers we used to associate mainly with extragalactic distances. A study by the University of California, Berkeley, estimated that in 2002 a total of 5 exabytes (5 billion gigabytes) of new data was recorded on media (including paper)—37,000 times the size of the Library of Congress—and another 18 exabytes flowed through electronic channels.¹ The IT research firm IDC, estimating the size of what the company calls the digital universe, produced an even more eye-popping figure: 281 exabytes of digital information created, captured, and replicated in 2007 alone, a quantity IDC expects to grow by 60% annually through 2011.²

These numbers reflect a particularly fast-growing component of image and video capture, but there seems to be hardly any endeavor that isn’t confronted with an exploding body of digital material. Scientific data captured in experiments or generated in computational research has reached heroic proportions, with no end in sight; at the upper end of data density, the Large Hadron Collider at CERN is expected to produce 15 million gigabytes of raw experimental data per year, a DVD every five seconds.³ The more bookish and text-bound side of the academy is undergoing a digital revolution of its own, as Google and other entities produce a searchable online corpus of books that could only have been found at the finest research libraries a decade ago—if there. Perhaps the most remarkable efflorescence of digital content has come from a combination of widening Internet access and web interactivity, a phenomenon almost impossible to quantify but suggested in outline by one estimate that almost 1.7 billion people worldwide now have access to the Internet, and another estimate that the number of available websites exceeds 225 million.⁴ And that doesn’t count more than 400,000 terabytes of new information generated annually by e-mail!⁵

Nor are more traditional sorts of structured business data lying still. In a 2007 study by the IT advisory firm Ventana Research, more than half of organizations surveyed reported that their structured business data was growing by 20% or more annually, and their unstructured content was growing faster still.⁶ For that matter, such neat categories as structured and unstructured data are getting intertwined, as back-end databases dynamically feed websites and XML-related technologies contribute to

---

A wealth of information creates a poverty of attention.
—Herbert Simon

©2009 EDUCAUSE. Reproduction by permission only.
what one analyst calls “a blurring distinction between data and documents.”

The sheer volume of information is only the beginning. As data flows with greater speed and ease, it overwhelms the physical and social restrictions that once kept it private or profitable, and those traditional constraints have increasingly been replaced by waves of privacy, security, and property rights regulation. What were once adequate tools for categorizing and storing information have been swamped, creating the need for new ways of describing and discovering information and of integrating its bewildering variety of expressions. As data gets more plentiful, ways of combining it grows exponentially, and people and institutions are asked to do more with it: to glean more from it; to put all of it at every (legitimate) user’s fingertips; and to link it, sync it, and use it to manage.

Higher education institutions aren’t the only organizations facing these challenges, but their historic mission of producing and aggregating information, along with the complexities of institutional management in an era of information overload, makes them especially sensitive to them. It seems that every aspect of IT administration, from the most traditional (major administrative applications) to the most emergent (online video and support for computational research), brings its associated bag of tricky data-related problems.

What’s more, with their often distributed organizational structures and highly expressive, data-loving constituents, higher education institutions can rarely impose centralized order on institutional data chaos even where that might seem like a good idea. The daily press of urgent business also seems to militate against attending to data issues that require a good deal of coordination and investigation even to describe, let alone solve. A survey conducted by the higher education IT architects’ organization ITANA in 2008 found that when asked to characterize the maturity of various aspects of their data management practice on a scale from 1 to 10, most institutions answered with a 5 or less in such areas as data quality and master data management, data warehousing and business intelligence, document/content/records management, and metadata management.

Data Management Concepts and Definitions

The body of IT expertise and practice surrounding these issues is variously known as data management, information management, enterprise information management, or some close variant. Some authors stress the conceptually useful distinction between data (symbolic elementary units of information) and information (data that is interpreted, processed, or put to a purpose). Attempts to carry the hierarchy further, to what systems theorist Russell Ackoff denoted as knowledge and wisdom, have been influential in management theory, though working “knowledge management” systems are not common in most IT environments, and wisdom management still appears to be a job for elders and philosophers, not IT professionals.

In practice, however, these distinctions may be more refined than necessary. It is clear that in an increasingly digital world, IT expertise surrounding data is the foundation for an environment from which individuals can either directly access information or create it through analysis or transformation. It’s interesting to note that DAMA International, a certification organization for data professionals, describes data resource management as “organizing people, energy, practices and procedures into activities that properly manage the information needs of an enterprise.” Our approach in this study has been to presume the fluid relationship between data and information, primarily employing a data-oriented vocabulary while still being mindful of the information needs of
the institution and the reasons why IT leaders are called chief information officers.

As the guiding concept for this study, ECAR uses the term institutional data management, which we define as the policies and practices by which higher education institutions effectively collect, protect, and use digital information assets to meet academic and business needs. We believe that this term stretches well enough to reflect our interest in enterprise information needs while speaking the language of working IT professionals.

**Study Scope and Objectives**

In a world where all manner of processes and types of content are being virtualized, a study of data management runs the risk of being a study of everything. We therefore looked for frameworks that could suggest some structure and boundaries to the project.

Data management is not governed by a single controlled vocabulary or by well-established comprehensive frameworks or standards. The most thorough framework is DAMA International’s Data Management Body of Knowledge (DAMA DMBOK), which divides data management into 10 different functions. The DMBOK framework, however, pays little explicit attention to strategic or regulatory issues that are likely to be of concern to IT leaders, and it is still in draft form. Many implications for data management are distributed throughout the IT Governance Institute’s COBIT 4.1 framework, but its “manage data” control process is narrowly concerned with backup, storage, and disposal of data. The ISO/IEC 27002 information security standard (formerly ISO 17799) covers security issues relevant to data management, but no comprehensive ISO standard dedicated specifically to data management exists.

The absence of a single dominant model of data management has led us to adopt an eclectic but selective approach to the topic that draws both on the assorted models mentioned above (especially DMBOK) and on ECAR’s understanding of the particular concerns of higher education, informed in part by prior ECAR research. Our study focuses on the following main topics:

- **Institutional context**—size and scale of the institutional data environment, data-related staffing, institutional drivers, and barriers to data management.
- **Data integrity and quality**—planning and control of activities that measure and ensure the fitness of data; respondent assessment of institutional data quality; and management of “master” data.
- **Analytics**—processes to provide decision support (aka business intelligence) for institutional leaders and other campus management and staff; and supporting infrastructure, including data warehouses, data marts, and related tools.
- **Data stewardship and security**—assignment of responsibilities for the security and quality of data; exercise of authority and decision making over the management of data assets and creation of data policy; and related data policies.
- **Content and records management**—infrastructure and practices to store, access, and (as applicable) archive data within content systems (for example, web, e-learning, documents) and institutional records systems; and assessment of the institution’s compliance with records retention requirements.
- **Research data management**—distribution of research-related data responsibilities among institutional units; research data ownership and other research-related policies; processes for determining investigator compliance with research data management...
Institutional Data Management

regulations; and institutional ability to support the long-term preservation of research data.

- **Data management outcomes**—respondent assessment of institutional performance on summary measures of data management effectiveness.

We have chosen not to include a number of topics that might be considered aspects of data management, either because they are things institutions have largely mastered or because ECAR has recently conducted studies related to them. These include database operations, backup/recovery, business continuity, and IT governance.

**Research Approach**

Our research consisted of four different components: a literature review, a quantitative web-based survey of IT leaders at EDUCAUSE member institutions, qualitative interviews with CIOs and others at selected institutions chosen on the basis of their responses to the survey, and case studies.

The literature review helped us determine workable definitions, and a reasonable scope, for this notoriously expansive concept, and it suggested hypotheses we could test while providing secondary studies for context and support. We examined articles and monographs from journalistic, academic, IT practitioner, and IT advisory sources. We also investigated relevant frameworks and standards, including DAMA International’s DAMA-DMBOK Functional Framework and the IT Governance Institute’s COBIT 4.1 framework.

With input from a number of higher education CIOs and data management experts, we designed a web-based survey aimed at the senior-most IT leader at respondent institutions. A copy of the survey resides at http://www.educause.edu/ECAR/InstitutionalDataManagementSur/163709. The survey was released in early February 2009 and remained in the field for approximately four weeks. Appendix A lists institutions that responded to the survey.

ECAR used qualitative interviews to gain insight into the quantitative results and to capture additional thoughts and ideas we would not have been able to explore through the survey alone. A total of 23 IT leaders and staff representing 19 institutions took part in our interviews; they are identified in Appendix B.

Finally, this study is accompanied by case studies that look in depth at data stewardship at the University of Virginia and at the Chronopolis project, a multi-institutional project sponsored by the Library of Congress that aims to create a grid-based environment for the long-term preservation of research data.

**Classification Schemes**

For comparison, we grouped institutions using categories derived from the 2000 edition of the Carnegie Classification of Institutions of Higher Education, developed by the Carnegie Foundation for the Advancement of Teaching. To obtain adequate numbers for statistical and descriptive purposes, we collapsed the Carnegie 2000 classifications as follows:

- **Doctoral institutions (DR).** The study grouped the doctoral-extensive and doctoral-intensive universities together.
- **Master’s institutions (MA).** The study grouped master’s colleges and universities I and II together.
- **Baccalaureate institutions (BA).** The study combined the three Carnegie 2000 baccalaureate groups together.
- **Associate’s institutions (AA).** These are the same as the Carnegie 2000 associate’s category.

To characterize respondent demographics, we also report an “other Carnegie” category that includes specialized institutions and U.S. higher education offices. Owing to the diversity and small size of this category, it does not
In our detailed data analysis by Carnegie class. We also tracked Canadian institutions in a separate, single category.

In November 2005, the Carnegie Foundation for the Advancement of Teaching introduced a new classification scheme employing additional institutional characteristics. We have not provided a crosswalk to the new scheme, in large part because we suspect that our readers will be more familiar with the older 2000 taxonomy.

We also follow the lead of several previous ECAR studies by analyzing results according to respondent institutions’ self-classification regarding the relationship of research and teaching in their institutional missions. As Table 2-1 shows, we asked respondents which of four statements best described their institution, using the terms research essential, balanced, teaching favored, and teaching essential as short names for each mission category. To group institutions at a still higher level, we refer to the two categories in which research is identified as primary to the mission (research essential and balanced) as “research oriented,” and those where teaching is the primary mission (teaching favored and teaching essential) as “teaching oriented.”

These categories avoid the blending of missions that takes place in the Carnegie 2000 classification, allowing us, for example, to combine responses from research-focused master’s and bachelor’s institutions with the bulk of the doctoral institutions. Where the mission categories prove to have greater statistical significance or explanatory power than the Carnegie classifications, we report them accordingly.

### Analysis and Reporting Conventions

We observed the following conventions in analyzing the data and reporting the results:

- Some tables and figures presented in this study report results for fewer than the total number of survey respondents (309). They were adjusted for missing information or to reflect some subset of responses.
- Sums of percentages in some charts and tables may not add up to 100.0% due to rounding.
- The data for each question in the online survey was analyzed for differences in patterns of response among Carnegie classes, Canadian and U.S. institutions, private and public institutions, U.S. region, research/teaching mission, and institutions of varying size. Institution size is determined by the number of full-time-equivalent (FTE) enrollments. We also looked for associations between other combinations of variables as appropriate. Differences that were both meaningful and statistically significant were noted in the text and/or the supporting figures and tables. Note that

<table>
<thead>
<tr>
<th>High-Level Category</th>
<th>Category</th>
<th>Mission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research oriented</td>
<td>Research essential</td>
<td>Research and teaching are the primary missions, but research is what really drives faculty and institutional success.</td>
</tr>
<tr>
<td></td>
<td>Balanced</td>
<td>Research and teaching are both primary missions, and they are equally important for faculty and institutional success.</td>
</tr>
<tr>
<td>Teaching oriented</td>
<td>Teaching favored</td>
<td>Teaching is the primary mission, but faculty research is rewarded.</td>
</tr>
<tr>
<td></td>
<td>Teaching essential</td>
<td>Teaching is the primary mission, and faculty research does not factor heavily in faculty and institutional success.</td>
</tr>
</tbody>
</table>
a statistically significant relationship between variables does not necessarily indicate a causal relationship.

- The Likert scales used in the online surveys are footnoted in the tables and figures showing results for these survey questions.

- Following what we believe to be the predominant practice among IT professionals, we treat the word “data” as singular rather than plural. Aside from noting that this seems to be the trend in ordinary speech, we believe that IT administrators tend to look at data as a resource or entity to be managed rather than as a collection of datum units. We expect, for example, that not many IT professionals would be inclined to say that “few data are stored on floppy disks anymore, but more and more are stored on key disks,” however much they might agree with the proposition.

**Overview of Respondents**

Invitations to respond to our survey were sent to the ECAR representative or (where the institution was not an ECAR subscriber) to the EDUCAUSE institutional representative at each EDUCAUSE member institution. In most cases, the invitee was the CIO. From 1,733 invited institutions we received 309 responses (a 17.8% response rate). Of the 309 respondents, 290 were from the United States or its territories and 19 were from Canada.

Figure 2-1 shows how respondents’ institutions were distributed among Carnegie categories, alongside the distribution of EDUCAUSE members and the overall population within each category at the time we closed our survey in March 2009. Proportionately, we had the highest participation from doctoral institutions. We also had more respondents from public institutions (58.6%) than from private ones (41.4%). Figure 2-2 shows the distribution...
and Carnegie class composition of the research/teaching mission categories.

The median FTE enrollment of respondent institutions was 4,219, and the mean was 7,642. As Figure 2-3 shows, smaller institutions contributed disproportionately to the results; 70.8% of respondent institutions had FTE enrollments of 8,000 or fewer students.

The survey was completed predominantly (78.4%) by the senior-most IT leader (CIO or equivalent; see Figure 2-4). Another 7.9% were directors of academic or administrative computing; combined with other IT management titles, a total of 96.8% reported IT-related titles. About 7 out of 10 respondents (68.5%) agreed or strongly agreed that they were personally very involved in data management policy and planning at their institution, yielding a median 4.00 and mean 3.68 level of agreement on our 5-point scale (where 1 = strongly disagree and 5 = strongly agree).

**Study Organization**

The rest of this study presents the results of our quantitative survey and the qualitative research that accompanied it.

To set the stage for later analysis, Chapter 3 looks at the institutional context for data management. We assess what respondents told us about trends in the amount of storage they manage and how well they believe their data infrastructure can handle the different varieties of data and content that their institutions need to accommodate. We look as well at drivers and barriers to institutional investment in data management and at the organizational context for data management expertise and improvement.

Chapters 4 through 7 look at different aspects of data management and the infrastructure, policies, and practices brought to bear on them. Chapter 4 examines in detail respondent assessments of enterprise data quality and initiatives to improve it, then looks at institutional analytics capabilities. In
Figure 2-3. Survey Respondent Institutions, by FTE Enrollment (N = 301)

1-2,000, 23.3%
2,001-4,000, 24.9%
4,001-8,000, 22.6%
8,001-15,000, 13.3%
15,001-25,000, 11.0%
More than 25,000, 5.0%

Figure 2-4. Survey Respondent Titles (N = 305)

Senior-most IT leader (e.g., CIO or equivalent), 78.4%
Director of academic computing, 2.3%
Director of administrative computing, 5.6%
Other IT management, 10.5%
Other academic management, 0.7%
Other administrative management, 0.3%
Vice president/provost/vice provost or equivalent (non-CIO), 2.3%
Chapter 5, we consider the extent and impact of data stewardship policies and practices and look at the policy environment related to data security, use, and access. Chapter 6 reports on institutional content management environments, looking at the extent of content management solutions and their impact on managing data variety, as well as examining practices related to the institutional website and records management. Chapter 7 looks at the increasingly complex problem of research data management, with a focus on how research data management support is distributed throughout the institution and on how institutions approach long-term research-data preservation issues.

In Chapter 8, we look synthetically at five outcomes measures chosen to represent different aspects of institutional data management, considering how desirable outcomes are related to respondent practices and policies. Finally, Chapter 9 looks at the future of data management in higher education, with an eye on the implications of our study results.

Endnotes
2. John F. Gantz et al., The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth through 2011 (Framingham, MA: IDC, 2008), http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf. This IDC white paper was sponsored by the storage products and services company EMC.
3

The Data Management Environment

The more the data banks record about each one of us, the less we exist.
—Marshall McLuhan

Key Findings

- Institutions reported increases over the past 12 months (mean near 4 on a scale of 5) in the volume of every kind of data we asked about. Learning management systems data and e-mail were the data types with the highest mean increase.

- Respondents generally agreed that they have the infrastructure they need to manage various kinds of institutional data; one exception is video/audio content available online, for which responses were neutral, on average. Views of data infrastructure capability three years in the future were more positive than those for current capability.

- Total disk storage capacity managed by central IT is expected to grow fast; the median anticipated growth rate for the 12 months following the survey was 20%.

- Respondents were only neutral, on average (near 3 on a scale of 5), when asked whether their institution’s identity management capabilities permit sufficiently granular user access to data. We found modest growth in the use of role-based authorization between 2005 and 2009. Federated identity was operational or being implemented by only about one institution in five, and it was disproportionately confined to research institutions.

- Leading drivers for data management investment were improved data security and regulatory compliance, followed by improved operational efficiency and enhanced data-driven decision making. The overwhelmingly highest-ranked barrier was lack of funds, followed by lack of staff expertise and decentralized institutional culture.

Something about data seems to remind people of water—lots of it. Here and there what’s happening with data might inspire an organic metaphor (mushrooming) or one related to chemical force (explosion), but mostly it’s all about water: flood, deluge, tsunami, ocean, fire hose. These not very tranquil associations summon up a few notions of how institutions might be responding to the emerging data environment. We wondered if our respondents were surfing the data wave (moving easily and skillfully over it), swimming in it (working hard to make progress), or drowning in it (here the meaning is self-evident).
To distinguish between those circumstances and to sound the depths of institutional data management (and here the metaphors will stop for a while), we collected a variety of subjective and objective measures about such basic factors as how fast various kinds of data are growing, how much storage institutions are managing, what drivers and barriers influence investment in data management, and how effectively respondents think their data infrastructures can manage assorted kinds of data.

**Data Growth: Volume and Variety**

Some kinds of data have been with us for a long time, especially structured business data such as student records, personnel files, and accounts payable and receivable. Other kinds, particularly unstructured data such as web content and digital video and audio, are relatively new. We wondered whether our respondents were experiencing data growth at different rates for different kinds of data, so we asked them to characterize how the volume of various sorts of data had changed at their institutions in the 12 months prior to our survey. Though we specified that we were referring to data as measured in bytes, we used a subjective scale ranging from greatly decreased (1) to greatly increased (5).

Big differences between types of data? Actually, similarity rather than disparity characterized our results. As Table 3-1 shows, respondents on average reported a level between 3.95 and 4.22, pretty close to “increased,” for every kind of data we named except research data (3.74). This exception was mainly due to differences between research and teaching institutions. Research-oriented institutions averaged 4.27, whereas teaching-oriented institutions averaged 3.52—considerably lower, though still indicating some research data growth.

Other than this, we found little relationship between reported growth in the different kinds of data on the basis of demographic factors such as Carnegie class, institution size, and public/private control.

Perhaps driven by the growing use of rich media in online learning, learning management systems data and content was on average the highest-growth data type among institutions overall, though e-mail growth was very close, and the difference between the two types is probably not meaningful. These two types of data had the highest percentages of “greatly increased” responses, at 28.1% and 30.7% respectively. Majorities of respondents reported increased or greatly increased volume for every data type, and in no case did even 2% of respondents report a decrease in volume in any data type.

Virtually all of our qualitative interviewees noted rapid data growth at their institutions, though they varied in their estimations of where it was growing fastest. It seems that scenarios for growth are unfolding in just

<table>
<thead>
<tr>
<th>Administrative Systems Data</th>
<th>Learning Management Systems Data and Content</th>
<th>Institutional Web Content</th>
<th>E-Mail</th>
<th>Research Data</th>
<th>Video/Audio Content Available Online</th>
<th>Library Digital Content Available Online</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean*</td>
<td>4.04</td>
<td>4.22</td>
<td>4.06</td>
<td>4.18</td>
<td>3.74</td>
<td>4.09</td>
</tr>
<tr>
<td>N</td>
<td>305</td>
<td>302</td>
<td>305</td>
<td>306</td>
<td>264</td>
<td>301</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>0.564</td>
<td>0.550</td>
<td>0.606</td>
<td>0.657</td>
<td>0.716</td>
<td>0.658</td>
</tr>
</tbody>
</table>

*Scale: 1 = greatly decreased, 2 = decreased, 3 = stayed the same, 4 = increased, 5 = greatly increased
about every type of digital information. The online-learning-related growth prominent in our survey results is evident at Bates College. Gene Wiemers, vice president for Information and Library Services, noted that Bates has “a mature environment for administrative computing,” whereas “the growth really has been in our course management system.” One major source of learning-related growth is faculty and student use of video, which has led Bates to set up protocols for storage and retrieval of video files in student labs. Clinton Smullen, director of Academic and Research Computing Services at the University of Tennessee at Chattanooga, has observed a similar connection. “Faculty members are adding media elements to their coursework,” Smullen said. “It’s the most rapidly growing part just because the files are so large.”

At Arizona State University, on the other hand, “traditional business data is still the fastest-growing area,” said John Rome, associate vice president for University Technology. Growth in the student population is one reason, he noted, but he also has seen a trend toward “data coming into central IT that was previously kept and managed locally.” And research data is “exploding” at Brown University, said Terri-Lynn Thayer, assistant vice president and deputy CIO. “Instruments are producing data in unprecedented volumes, so that it’s not just a question of storage but of how to ever find what you’re looking for again.” Brown has recently partnered with IBM and a number of research entities to create a high-performance computing cluster with half a petabyte of storage in order to provide a centrally managed location for research data; in addition, Brown’s library is spearheading a digital repository initiative. At other institutions, interviewees mentioned e-mail archiving, document imaging, and other kinds of content digitization as dominant areas of data growth.

Data volume, in short, is increasing across the board. Does this mean, then, that our respondents are struggling to provide the infrastructure that today’s data environment demands? Our qualitative interviewees frequently said that they were and reported various initiatives to add more storage. Jack Rutt, director of Information Systems at Eastern Mennonite University, noted that a recent ERP implementation and growing use of digital media have dramatically increased pressure on the university’s storage infrastructure. “We got a storage network three years ago and at that time projected a three-year life,” Rutt said. “After a year and a half, we already had to add more storage.” At the University of North Texas, acting vice president for information technology and CIO Maurice Leatherbury complained, “We’re spending an extra several hundred thousand dollars annually to buy storage. I’m not sure how long this rate of spending can be sustained.”

Our quantitative results, however, suggest that most respondents felt that their institutions are meeting the challenge, at least when it comes to infrastructure. When we asked whether they agreed that their institutions currently have the infrastructure needed to effectively manage each kind of data and then asked whether they agreed their institutions would have it in three years, respondents were generally positive about the present and slightly more so about the future (see Figure 3-1).

Mean agreement regarding current infrastructure was above the midway point between neutral (3) and agree (4) for every type of data we asked about except video/audio content available online and research data. The online video/audio item seems to be creating more infrastructure challenges than other kinds of data: 40.5% of respondents disagreed or strongly disagreed that their institutions had the needed infrastructure. The next highest item in combined disagreement was research data (22.5%). As with our findings regarding research data volume growth, the research data infrastructure item was associated with
institutional mission: Teaching-essential institutions agreed more strongly (mean 3.41) that they currently have the needed infrastructure than did research-essential institutions (2.63). Balanced research/teaching institutions and those where teaching is favored but research is also rewarded had similar means in between (3.24 and 3.25, respectively).

The results for institutional infrastructure in three years correlate to those for the present but are higher, in every case exceeding the midway point between neutral and agree. The outlook for online video/audio was especially optimistic: In contrast with the 40.5% of respondents who disagreed or strongly disagreed about having the needed infrastructure in the present, only 14.5% felt the same looking ahead three years. Similarly, the 22.5% combined disagreement about research data in the present fell to less than half that figure (10.7%) for the three-year question. Interestingly, research and teaching institutions did not differ significantly in their agreement on the three-year infrastructure questions.

Our respondents, then, reported an increase in all varieties of data, and extra pressure in some, but they don’t seem to feel they are losing control of the situation, at least to the extent that infrastructure influences effective data management. Infrastructure, however, may not be where the most serious challenges lie. After all, increases in storage density have reliably kept pace with Moore’s law over the years, while our understanding of how to integrate, document, search, secure, and interpret data has been far more uneven, and what we do know is often hard to apply in an organizational context. “My problem is not the fastest-growing data type, but the interconnection between the data,” said Marc Hoit, vice chancellor for IT and CIO at North
Carolina State University. “So much of this data is siloed and not integrated, and where it’s duplicated, it’s not in sync.”

**Central IT Disk Storage Capacity and Growth**

So far, we’ve looked at subjective evaluations about changes in the volume of different kinds of data and institutional data management infrastructure. We also asked for three more concrete measures: the total capacity of disk data storage (in terabytes) managed by central IT (not including PCs), the percentage of that total managed by central IT on behalf of institutional researchers, and how much the total was expected to grow in the next 12 months.

Most institutions reported totals that, although stupendous by historical standards, were not enormous by the measure of modern device capacity. About half of institutions (47.0%) reported a total central-IT-managed capacity of 20 terabytes or less, and half of these reported 10 terabytes or less (see Figure 3-2). Another 12.8% reported 51 to 100 terabytes, while 14.7% gave responses of more than 100 terabytes. Perhaps reflecting the complexity of central IT data operations, our capacity question yielded an unusually high 7.0% “don’t know” response.

As Figure 3-3 shows, we found variations by Carnegie class that help explain the somewhat bimodal distribution of Figure 3-2. Doctoral institutions reported much higher capacity than did the other Carnegie classes: more than half of them (52.9%) reported 101 terabytes or higher. By contrast, majorities of all the other Carnegie types reported a total capacity of 20 terabytes or less. (These Carnegie results do not include “don’t know” responses, which came from all Carnegie types.)

In a follow-up question, we asked about the percentage of total disk capacity managed by central IT on behalf of institutional researchers. A high proportion of “don’t know” responses (28.6%, mainly from teaching-oriented institutions) suggests that

![Figure 3-2. Total Capacity of Disk Storage Managed by Central IT (N = 298)](image-url)
the question was either hard to answer or not applicable for many respondents. Among the remainder, however, it was clear that central IT is managing substantial research data assets on behalf of researchers. Among research-oriented institutions, respondents reported an average 15.2% and median 10.0% of total capacity under such management. Teaching-oriented institutions were managing quite a bit less, with a 6.5% mean and a 2.0% median.

These figures suggest that many researchers find sufficient advantages to central IT management of resources, such as environmental and system administration economies, to give up their traditional do-it-yourself autonomy. The results are consistent with findings in several ECAR research studies reporting an increasing role by central IT in research storage, especially at research-essential institutions. One driving factor may be the growing role of digitization in disciplines relatively less experienced in large-scale data system management. “It used to be that those most in need of computing solutions were in physics and engineering,” said Brown’s Thayer. “Now there is a ton of stuff coming from the biomedical and life science areas and fields like geology—disciplines whose science and instrumentation didn’t previously need this amount of data storage, and that don’t already have the infrastructure to take care of it themselves in the way those other disciplines do.”

Respondents reported substantial expectations of growth in their central IT disk storage needs in the next year (see Figure 3-4). The most common single growth figure reported was 10% (66 respondents), and putting aside the small number of respondents anticipating either a decrease or growth greater than 100%, responses averaged 22.7%, and the median
anticipated growth was 20.0%. Growth is a widespread and undifferentiated phenomenon: We didn’t find any significant patterns by Carnegie class, FTE enrollment, or mission, nor did we find any association between the reported current total capacity of central-IT-managed disk storage and expected growth. Among the small outlier group of respondents reporting greater than 100% expected growth, five were research-oriented and seven were teaching-oriented institutions.

With many institutions anticipating central IT disk storage capacity growth in the neighborhood of 20% in a single year, and some quite a bit more than that, it seems clear that however confident respondents may be about their institutions’ data infrastructures, they face a challenge in building out their raw capacity.

### Identity Management

The sheer growth of data and its related storage infrastructure, however, isn’t all that institutions must confront. Flexible and secure identity management lies at the heart of a secure and effective data environment. Identity management “has all sorts of relations to data management,” said Steve Gallagher, CIO at the University of San Francisco. “It’s key to ensuring that you’re exposing appropriate data to appropriate people. Even more important is ensuring that you cut access to people at the time of employee separation or job change.”

Our survey asked about respondent adoption of several technologies related to identity management:

- **enterprise directories**, directory services capable of including all persons affiliated with the institution and usable by multiple applications;
- **automated role- and privilege-based authorization** (or more simply role-based authorization, abbreviated here as RBA), systems that give access to electronic resources using privileges or permissions derived automatically from affiliations and groups; and
- **federated identity solutions**, automated management of identity information.
between a campus and other institutions or organizations to facilitate collaborative or business initiatives.

We present our findings about the adoption and use of these technologies below. Thanks to a previous ECAR survey on identity management that was fielded in 2005, we were able to compare some of our 2009 results with earlier ones. Where we do so, we restrict our analysis to the 116 institutions that responded to both surveys.

Adoption of Identity Management Technologies

We found a robust and active identity management environment on our respondent campuses, but one still largely focused on internal capabilities rather than extra-institutional outreach.

The most mature of the three technologies we asked about is enterprise directories, and not surprisingly it was the one most frequently reported (see Figure 3-5). Altogether, two-thirds (66.0%) of respondents said their institutions were either in the process of implementing an enterprise directory or had one partially or fully operational. Enterprise directories were somewhat more prevalent among research-oriented institutions, but small majorities of teaching-oriented institutions reported one of these statuses as well.

Role-based authorization was somewhat less common, with 54.6% either implementing or operational; about half of these said they were partly operational, which was the most common RBA status reported. We found no significant associations between RBA adoption and Carnegie class, FTE enrollment size, or mission.

By contrast, federated identity solutions remain relatively marginal and are strongly associated with institutional size, research/teaching mission, and Carnegie class. Overall, only one in five institutions (19.8%) reported in-progress implementation or partial or full operation of federated identity solutions,
and federated identity had by far the highest proportion of respondents (35.3%) who were not considering the technology. The bigger and more research-oriented the institution, the more likely it was to adopt the technology. Among doctoral institutions, 44.1% reported either implementing federated identity or having it partially or fully operational, versus 17.6% for MA institutions and less than half that for BA (6.2%) and AA (8.7%) institutions. As Figure 3-6 shows, Shibboleth was by far the most popular federated identity solution reported by institutions that were either implementing or had an operational solution, though substantial proportions reported SAML 2.0 or other solutions.


Comparing 2009 with 2005 identity management survey results suggests a degree of growth in the less mature areas of identity management during the past four years. Among the 116 institutions responding to both surveys, enterprise directory adoption remained flat, with no significant difference between the 73.0% reporting in-progress implementation or operation in 2009 and the 69.5% in 2005. Given the relative maturity of enterprise directories, this is not surprising; institutions have had a long time to settle on an enterprise directory strategy.

There was, however, significant growth in the prevalence of role-based authorization among the 2005/2009 survey respondent institutions. As Figure 3-7 shows, the total proportion of institutions reporting in-progress implementations or partial or fully operational RBA rose from 37.4% in 2005 to 55.7% in 2009. All of the difference in these statuses came from higher reported rates of implementation in progress or fully operational RBA, rather than from partially operational RBA.

It’s harder to state definitively that we saw an increase in federated identity solutions, because this technology was so new in 2005 that we didn’t ask about it in the same detail as in 2009. However, a rough comparison for which we cannot claim statistical significance

![Figure 3-6. Federated Identity Technologies in Use (Institutions Implementing or Operating Federated Identity Technologies, N = 77)](image-url)
suggests a degree of growth has occurred. Only 14.0% of 2005/2009 responders told us in 2005 that their institutions were implementing Shibboleth, Liberty Alliance, or another federating technology. In 2009, among the same institutions, 16.3% reported partially or fully operational federated identity solutions (overwhelmingly Shibboleth), and another 8.6% said they had implementations in progress. Although this suggests some progress has been made, it’s still striking that the key technology enabling extra-institutional authentication and authorization remains so narrowly implemented.

**Granularity of User Data Access**

One of the primary purposes of identity management tools is to assemble the information and rules necessary to ensure that users and data are properly matched—to permit employment records, for example, to be available to HR staff but not to accounts payable staff who have no legitimate need to see them. Creating a system sufficiently refined to authorize accurately and reliably can be a challenge.

On the whole, our respondents were less sanguine about their institutions’ identity management capabilities for providing sufficiently granular user data access authorization than they were about their institutions’ having the infrastructure needed to effectively manage different kinds of data (see Figure 3-8). Their agreement about having a sufficient capability averaged an almost exactly neutral response (3.03, SD = 1.008), and combined disagreement (35.5%) was not much less than combined agreement (40.4%).

These patterns of agreement and disagreement were broadly generalized. We found no significant differences based on Carnegie class, FTE enrollment size, public/private control, or institutional mission. Though
Institutions at more advanced stages of adoption of enterprise directories and RBA tended toward higher agreement that their institutions had sufficient granularity of access, the difference was very modest for enterprise directories, and for RBA the difference was most evident among the relatively small percentage (12.7%) reporting fully operational RBA systems.

Interviews suggest that much of the challenge of refined data access lies in the variety of data elements that bear on access rules and the frequency with which the rules and data values can change. Georgia State University, for example, has a “reasonably good” identity management system, said J. L. Albert, associate provost and CIO, thanks to an LDAP-based identity vault that manages pertinent identity information and feeds Active Directory in near-real time. “The big issue is having all of the right attributes of data in it for building workflow,” said Albert. “There are still some issues about having the right pieces of data in the right locations and having somebody responsible for keeping them current. The issue is made more complex by individual identities with multiple roles.”

The University of San Francisco’s Gallagher also noted an issue with the applications that consume services from the identity infrastructure. At USF, he said, “the granularity is there at the core level, but the various spokes and the different integration points, especially third-party services, handle different levels of granularity.”

### Data Management Drivers and Barriers

Reasons why institutions might invest in data management vary: to keep data secure, to get more value out of it, to deliver new services, or to make IT systems more integrated and efficient. Overall, “safety first” seems to be the principle most often driving data management investment at our respondent institutions, followed by efforts to promote better business results in the major units of the institution.
We asked respondents to choose up to three primary drivers (from a list of 11 items) for data management investment at their institutions. By a considerable margin, improved data security was the item most often chosen, named by 6 in 10 respondents (see Figure 3-9). Another security-related item, regulatory compliance, was chosen by 4 in 10, making it the second most chosen driver.

If the top two items both seem to be related to data (and professional!) safety, the next two both spoke to making better use of data for institutional purposes. In each case, about a third of respondents chose improving business/academic unit operational efficiency and enhancing data-driven decision making, placing these two related items close together behind the security and regulatory items. The contrast between the relatively high rate of selection for these items and the bottom-ranking two items—teaching/learning support and research support—suggests that institutions are being driven more by a traditional “business intelligence” conception of data management than by more innovative uses relating to instruction and research. At the same time, business purposes seem to trump purely IT-related items, which fell in the lower range of responses.

Finding any pattern in the barriers to investing in data management that our respondents reported (choosing up to 3 primary barriers from our list of 11 items) is more difficult. The standout barrier was lack of funds (see Figure 3-10), named by 69.9% of respondents. The fact that our survey was in the field in early 2009, during the spreading financial crisis and recession, may help explain why this figure was so high, though lack of funds is a perennially top-ranking item in ECAR results relating to IT initiative barriers. The next-highest (but much less frequently chosen) items were basically an even split between lack of staff expertise (34.0%) and a decentralized/informal institutional culture (33.0%), though a number of other items in the middle ranks, such as the difficulty of developing campus policies and procedures (27.2%) and resistance from data owners (16.2%), suggest that basically...
cultural issues such as organizational stove-piping, decentralization, and institutional inertia may be bigger barriers overall than lack of expertise.

Barriers relating to expertise and culture are not mutually exclusive, but we found some evidence that different kinds of institutions struggle more or less with each. The very things that make data management expertise harder to develop and staff at simpler institutions—flatter organizational structures, leaner IT units, and relatively more focus on basics—may help minimize cultural obstacles, whereas at more complex institutions, expertise may be more abundant but bureaucracy and inertia proportionately greater. As Figure 3-11 shows, associate's institutions were much more likely than doctoral institutions to report lack of expertise as a barrier, whereas doctoral institutions were much more likely than associate's to report decentralized or informal institutional culture. Master's and baccalaureate institutions tended to fall between the other Carnegie types on both measures.

**Organizational Context**

We noted above that after security and regulatory issues, improved operational efficiency and supporting data-driven decision making were the top drivers for investing in data management. We found supporting evidence along the same lines by asking our respondents whether they agreed that leaders and users were demanding improved institutional data management capabilities. Slight majorities agreed or strongly agreed about each group, and mean agreement was in each case just about halfway between neutral and agree (see Table 3-2). Respondents were slightly more strongly in agreement as well that improved institutional data management was a strategic priority.

Exactly two-thirds of our respondents (66.7%) told us that their institutions had either a stand-alone IT strategic plan or one integrated into an institutional strategic plan. We found no differences in agreement about the demand for or priority of improved institutional data management between...
Table 3-2. Demand for/Priority of Improved Data Management

<table>
<thead>
<tr>
<th></th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leadership is demanding</td>
<td>3.46</td>
<td>309</td>
<td>1.103</td>
</tr>
<tr>
<td>improved institutional data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>management capabilities.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Users are demanding</td>
<td>3.48</td>
<td>309</td>
<td>1.021</td>
</tr>
<tr>
<td>improved institutional data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>management capabilities.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improved institutional data</td>
<td>3.68</td>
<td>309</td>
<td>0.983</td>
</tr>
<tr>
<td>management is a strategic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>priority.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree

Institutions that did and those that didn’t have such a plan. However, among those reporting a plan, those saying that the plan explicitly addressed data management issues reported significantly higher agreement on all three of the demand/priority items.

We did not find respondent institutions heavily endowed with offices or governance bodies related to data management (see Figure 3-12). A slight majority (55.4%) said that their institution has a central IT group for data (including, for example, data architects), but such a group was twice as prevalent as the other two entities we asked about, an institutional data policy office or an IT governance committee dedicated to data issues. Though larger institutions were more likely to report a central IT group for data than smaller ones, the association was a modest one, and we found no differences based on Carnegie class or FTE enrollment for the other data management entities.

None of this necessarily means that institutions lacking dedicated data management entities entirely lack the expertise or championing of issues that such groups might
provide. Some of our interviewees told us that they have sufficient data management expertise at hand, but not in dedicated positions or units “There are no people with the title ‘data architect’ on our staff,” said Andrew Lawlor, associate vice president for Technology and Communications at Edinboro University of Pennsylvania. “There are data experts, who may apply their knowledge as part of their duties.”

Likewise, respondents from some institutions seemed to feel that general IT governance structures can get the job done without a special data committee. At the Community College of Rhode Island, Stephen Vieira, CIO and executive director of IT, uses an Institutional Technology Advisory Committee, which is a subcommittee of the president’s council and has representatives from all major functional groups. “We can feed anything to do with data right into that committee,” said Vieira. “It’s nice that they are all in one room, so the integration of data comes into play and everyone gets a feel for it.”

**Surfing, Swimming, or Drowning?**

Though it remains a popular metaphor, our respondents don’t really say they are drowning in data. The sheer volume of information that institutions manage is undoubtedly impressive, but when asked about how it has changed in a year across various data types, respondents were more likely to say it increased than that it greatly increased, and they were generally positive about the capability of their institutions’ data infrastructure to manage all those types of data, now and in the future. They do give the impression of a certain struggle to keep up: The mean 22% growth in central IT disk storage that they anticipate over a year’s time is not going to permit a relaxed approach to data infrastructure, nor will it address the relative weakness they seem to recognize in their ability to handle audio and video content. But our respondents showed no signs of a feeling of impending breakdown.

But neither, it must be added, do respondents seem to be surfing happily on the great
wave of data. Keeping up with research data seems to be a real challenge at research-essential institutions, which reported a below-neutral agreement about data infrastructure sufficiency for this item as well as relatively high levels of central IT storage dedicated to institutional researchers. Funding for data management was a widely reported barrier, and drivers for investing in data management seem focused on defensive security matters and business-oriented tasks rather than potentially transformative uses in instruction and research. Qualitative interviewees expressed concerns about integrating and documenting data, and about the conceptual challenges of managing unstructured and nontraditional data. And although we saw some progress over the past four years in the creation of an identity infrastructure that can authorize access by role, the basic tool for integrating identity (and thus data regimes) across institutional lines—federated identity—remains the preserve of a small fraction of predominantly research-oriented institutions.

So our overview of the data management environment pictures most institutions safely, if a bit doggedly, swimming across a challenging sea of data. As with a lot of higher education IT concerns, data management has the aspect of a workaday task being handled competently by busy IT departments with many other tasks to do. Yet in an age when users can increasingly get instant data and content gratification through a multitude of commodity sources, higher education will have to do better than supply a local infrastructure for doing so. Data delivery is frequently the end for which IT is the means, and as the most commodified forms of data become almost effortlessly available, higher education will increasingly satisfy its mission and its users by providing higher value-added, more authoritative, better-vetted, and less generally accessible data and content. To do that, it must imaginatively rethink how data abundance reshapes its business and create the infrastructure and policies enabling a new era of data discovery and access. The swimmer needs to learn how to surf.

Endnotes
1. SD = 0.574 and 0.652, respectively.
2. Harvey Blustain, with Sandra Braman, Richard N. Katz, and Gail Salaway, *IT Engagement in Research: A Baseline Study (Research Study, Vol. 5)* (Boulder, CO: EDUCAUSE Center for Applied Research, 2006), available from http://www.educause.edu/ecar. The authors reported that data storage was growing faster than such other research-related IT resources as high-performance computing and high-performance networks (p. 52), and most respondents expected central IT funding for research-related data storage to increase. For similar findings in medical school settings, see Mark R. Nelson, *IT Engagement in Research: A View of Medical School Practice (Research Study, Vol. 1)* (Boulder, CO: EDUCAUSE Center for Applied Research, 2008), 17, 26, available from http://www.educause.edu/ecar.
Data Quality and Analytics

Key Findings

- Overall, respondents were lukewarm in their self-assessments of data quality in major administrative enterprise systems. An enterprise data quality score summarizing eight different data-quality-related measures averaged 3 (on a scale of 1 to 5) across the whole respondent population.

- Higher data quality scores were associated with earlier adoption of new technologies, stronger agreement that the institution provides data quality training to employees, and stronger agreement that central IT works actively with business/academic units to ensure data quality.

- Respondents tended to agree that their institution’s leadership was committed to evidence-based decision making, and that their institution was committed to making management information widely available.

- Infrastructure for analytics was modest among our respondent institutions, with 4 out of 10 reporting that they didn’t have any of the four data store types we asked about, and another 2 in 10 reporting only one. About one-third of institutions said they have an institution-wide data warehouse.

- The primary use for analytics tools is either extraction and reporting of transactional data or analysis of operational performance. Only about one in eight institutions reported what-if decision support, predictive modeling, or auto-triggering processes as their primary use. However, these advanced uses are commonly employed for specific situations on at least an occasional basis.

- More than half of institutions said they plan to upgrade their analytics capability in the next two years. At the same time, we found no significant differences in analytics infrastructure or use between 2005 and 2009 among the 112 institutions that responded to ECAR academic analytics surveys in both years.

Data that isn’t accurate can easily do more harm than good, and structured data of the sort used by applications needs to be up to date, coded consistently, and defined precisely. Maintaining a high level of data quality is the sine qua non of getting data to do...

©2009 EDUCAUSE. Reproduction by permission only.
work—providing information that helps the institution fulfill its mission.

But even perfectly complete and accurate data doesn’t just jump up and tell administrators how their institution is doing. A lot of infrastructure is needed to make data readily available for analysis. Just as important is an institutional culture that’s committed to sharing data as needed and bringing it to bear on decision making. ECAR studies commonly find that the use of metrics and data-driven management processes is related to successful IT outcomes.

In this chapter, we consider the “hands-on” aspects of data management, with an emphasis on structured business data. We begin by looking at how institutions assess the documentation, consistency, and reliability (in short, the quality) of data in major administrative systems and at what they’re doing to maintain and improve data quality. Next, we look at how institutions make data available for analysis and the kinds of analytical processes they use to manage the institution.

Assessing Enterprise Data Quality

Often, the most powerful and valuable data in an institution passes through many hands, because it’s used for different purposes by different departments. Because of the challenges of managing this kind of enterprise data and its potential for helping administrators improve institutional performance, our survey looked in detail at how institutions approach data quality in their major administrative enterprise systems and data stores.1 We began by probing respondents’ views of their institutions’ performance in assorted enterprise-data-related measures, which we examine below first in detail and then in a more summary fashion by calculating a data quality score based on the detailed responses. Because promoting data quality relies on the awareness and participation of many parties, we also asked respondents whether their institutions provided data quality training and to assess how well constituents understood data quality issues. Finally, we looked at institutional initiatives to improve data quality.

Enterprise Data Characteristics and Quality in Detail

We asked a series of questions about data-quality-related characteristics of major data elements (defined as data items essential to conducting basic business and academic operations, such as student name or employee status) and about processes to validate data and ensure its quality. Most of our questions related to major administrative enterprise systems and data stores, though one item asked about business/academic unit (“shadow”) systems.

We didn’t find respondents terribly enthusiastic about their institutions’ performance on these enterprise data quality measures. Means were for the most part only mildly positive and in some cases fell short of an average neutral (3) response (see Table 4-1). For only one item, that a system of record was identified for each major data element, did a large majority of respondents (64.2%) agree or strongly agree. A more modest majority (53.3%) agreed or strongly agreed that each major data element had a single definition recognized across the institution; combined agreement was below 50% for all the other items. For the statement that changes in the value of a major data element propagate across all business/academic unit (“shadow”) systems and data stores that use it, combined disagreement was 58.8%.

Regarding data quality processes, although respondents were modestly positive in agreeing that processes were in place to assure data quality at the point of capture or origin, their mean agreement fell below neutral on items relating to the existence of automated systems to validate data across enterprise systems and data stores (mean 2.89) and to
having processes in place for documenting and reviewing all identified data quality issues (mean 2.76).

Taken as a whole, we interpret our enterprise system data management measures results as evidence that institutions are struggling to treat data in a truly enterprise manner. Although major data elements generally seem to have a home in some system of record, presumably enterprise resource planning (ERP) systems, respondents were neutral on average—and substantial minorities disagreed—that data is consistently coded and that changes propagate appropriately even in enterprise systems. When major data elements make their way into business and academic unit systems, the results are less positive still. No one who has attempted to collect and interpret accurate, synchronized data from across different business processes and time frames will fail to see the potential for error, inefficiency, and unrealized data value in such results.

Even IT administrators selected for qualitative interviews because of their institutions’ relatively high self-ratings for data quality freely acknowledged the difficulties of addressing data quality in a systemic, enterprise-oriented way. “Data quality is one of my real concerns,” reported Marc Hoit of North Carolina State University. “A few data elements, such as grade and salary, are self-correcting. And accountants have been doing reconciliations for what?—at least 300 years. But this only verifies some things.” Hoit believes that data quality, often seen as a precondition for an analytics capability, could in turn best be improved through analytics. “I think we can use trend analysis, detecting outliers, getting information on data flow and data changes to identify problems. Are things positively correlated that should be?”

Table 4-1. Administrative Enterprise System Data Quality Measures

<table>
<thead>
<tr>
<th>Characteristics of Major Data Elements</th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A system of record is identified for each major data element.</td>
<td>3.54</td>
<td>301</td>
<td>0.981</td>
</tr>
<tr>
<td>Each major data element has a single definition that is recognized across the institution.</td>
<td>3.29</td>
<td>308</td>
<td>1.117</td>
</tr>
<tr>
<td>Each major data element is coded consistently across systems and data stores.</td>
<td>3.12</td>
<td>301</td>
<td>1.075</td>
</tr>
<tr>
<td>When the value of a major data element changes, the change propagates across all enterprise systems and data stores that use it.</td>
<td>3.11</td>
<td>308</td>
<td>1.140</td>
</tr>
<tr>
<td>When the value of a major data element changes, the change propagates across all business/academic unit (“shadow”) systems and data stores that use it.</td>
<td>2.52</td>
<td>299</td>
<td>1.082</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Quality Processes</th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processes are in place to assure data quality at the point of capture/origin (e.g., data entry).</td>
<td>3.24</td>
<td>306</td>
<td>0.975</td>
</tr>
<tr>
<td>Automated systems are in place to validate data across enterprise systems and data stores.</td>
<td>2.89</td>
<td>302</td>
<td>1.035</td>
</tr>
<tr>
<td>Processes are in place for documenting and reviewing all identified data quality issues.</td>
<td>2.76</td>
<td>304</td>
<td>0.995</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
At the Community College of Rhode Island, CIO Stephen Vieira stated that a lot of effort has been put into training data entry personnel and establishing data quality standards and guidelines—‘‘data quality is an issue we are constantly working on,’’ he said—but he also noted the potential for advanced, downstream data management capabilities to feed back into the earlier stages of the data life cycle. Now in the process of implementing a new operational data store and data warehouse, Vieira observes that ‘‘the data warehouse part takes in ERP and shadow database elements so that users can incorporate both into the reports they write. The shadow data has to be consistent with the ERP data, and if not, the authoritative database takes precedence.’’

Enterprise Data Quality Score

To get an overall measure of enterprise data quality, we created a score for each institution by calculating the mean of the eight enterprise system data characteristics and data quality process items shown in Table 4-1. For all respondents, the data quality score average fell right in the middle of the scale at 3.06, though there was considerable variation: 22.7% scored below 2.50 and 27.9% scored above 3.50. To distinguish this score from the agreement scales it is derived from, we characterize its value range simply as low (1) to high (5).

This enterprise data quality score proved to be powerfully associated with many other measures, and we report these associations throughout this study. In particular, as we discuss in Chapter 8, we found strong associations between enterprise data quality score and data management outcomes: Every outcome measure we asked about was significantly higher among institutions with higher data quality scores. Another interesting set of positive associations, reported in the section below (see Table 4-3), relates to data quality training and constituent understanding of data quality issues.

For present purposes we note two other intriguing associations: Early technology adopters tended to have higher scores than late adopters, and institutions reporting greater agreement that their identity management capabilities provided sufficiently granular user data access authorization also tended to have higher scores (see Table 4-2).

It may be that earlier adopters of new technologies are better equipped with the tools to manage data quality (for example, they agree more strongly than do late adopters that their institutions have automated systems in place to validate data), or they may tend to deploy systems with more modern architectures and stronger integration tools that minimize the amount of replicated and unsynchronized data.

Identifying even speculatively a link between identity management authorization granularity and data quality is more difficult, though in general, robust authorization implies scrutiny of identity and role information, good integration, and a more formal data access regime. In ECAR’s 2006 study of identity management, we found many interviewees voicing some variation on the lament that ‘‘we have 25 definitions of what a student is.” Cleaning up and rationalizing such environments is often a prerequisite to improving identity management. However, as with all the associations we note here, the relationship could work either way; we cannot say whether better data quality leads to more granular identity management, or the other way around.

Training, User Understanding, and Central IT/Business Unit Interaction

Our respondents tended to represent training in data quality largely as something staff receives, rather than management and leadership. As Figure 4-1 shows, large majorities of respondents disagreed or strongly disagreed that either executive leaders or business and
academic unit leaders receive training in data quality, though combined disagreement was lower for the unit leaders than for the executives. Even for staff, nearly 3 in 10 respondents disagreed that staff receives training, though a slight majority agreed that they did. Respondents agreed more strongly that each of these groups understood data quality.
In our results suggests that respondents see users often picking up knowledge of data quality issues through on-the-job experience rather than through formal training. Just the same, where users are getting training, there is a greater perception of better understanding. Although the picture overall is far from bleak, it does suggest that CIOs will find it more challenging to communicate data quality issues the farther they go up the management chain, and that a reservoir of misunderstanding of data quality issues exists at some institutions.

Perhaps one factor leading respondents to be more positive about data quality understanding than training is that they seem to work often with users on data quality issues. Three-quarters of respondents (76.4%) agreed or strongly agreed that central IT works actively with business and academic units to ensure data quality (see Figure 4-2), and overall mean agreement was 3.88, close to an average agree (4) response. Agreement
about training and user understanding of data quality issues for all constituent groups was higher at those institutions that reported higher agreement about working actively with business and academic units. This makes sense, given the complex way that functional and technical issues can intertwine with data items. Stephen Vieira of the Community College of Rhode Island noted that at his institution, data quality training is largely concentrated in the user departments, with assistance from the institutional research group, but “IT is there to assist. We listen to learn what the stumbling blocks are, so we can look into how to change the system to make things easier and more accurate. For their part, the users get a good picture of why errors are occurring.”

The good news cutting across all these results is that judging from the associations we found, data quality is associated with active effort and constituent understanding. As Table 4-3 shows, for every item reported in this section—training and understanding among all three constituent groups, and central IT working actively with business and academic units—mean enterprise data quality scores were positively and significantly associated with higher respondent agreement.

Table 4-3. Enterprise Data Quality Score, by Respondent Agreement on Data Quality Training, User Understanding, and Central IT Interaction

<table>
<thead>
<tr>
<th>At my institution...</th>
<th>Agreement Level</th>
<th>Enterprise Data Quality Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean* N Std. Deviation</td>
</tr>
<tr>
<td>Executive leaders receive training in data quality issues.</td>
<td>Strongly disagree/disagree</td>
<td>2.94 225 0.763</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>3.49 45 0.744</td>
</tr>
<tr>
<td></td>
<td>Agree/strongly agree</td>
<td>3.28 31 0.797</td>
</tr>
<tr>
<td>Business/academic unit leaders receive training in data quality issues.</td>
<td>Strongly disagree/disagree</td>
<td>2.85 180 0.745</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>3.31 45 0.692</td>
</tr>
<tr>
<td></td>
<td>Agree/strongly agree</td>
<td>3.42 79 0.777</td>
</tr>
<tr>
<td>Staff receive training in data quality issues.</td>
<td>Strongly disagree/disagree</td>
<td>2.65 85 0.720</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>2.83 56 0.669</td>
</tr>
<tr>
<td></td>
<td>Agree/strongly agree</td>
<td>3.36 163 0.739</td>
</tr>
<tr>
<td>Executive leaders understand data quality issues.</td>
<td>Strongly disagree/disagree</td>
<td>2.75 101 0.776</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>3.11 80 0.758</td>
</tr>
<tr>
<td></td>
<td>Agree/strongly agree</td>
<td>3.30 121 0.735</td>
</tr>
<tr>
<td>Business/academic unit leaders understand data quality issues.</td>
<td>Strongly disagree/disagree</td>
<td>2.70 71 0.781</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>3.02 76 0.714</td>
</tr>
<tr>
<td></td>
<td>Agree/strongly agree</td>
<td>3.28 156 0.750</td>
</tr>
<tr>
<td>Staff understand data quality issues.</td>
<td>Strongly disagree/disagree</td>
<td>2.63 52 0.782</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>2.82 59 0.741</td>
</tr>
<tr>
<td></td>
<td>Agree/strongly agree</td>
<td>3.26 195 0.738</td>
</tr>
<tr>
<td>Central IT works actively with business/academic units to ensure data quality.</td>
<td>Strongly disagree/disagree</td>
<td>2.50 24 0.790</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>2.62 48 0.759</td>
</tr>
<tr>
<td></td>
<td>Agree/strongly agree</td>
<td>3.21 236 0.732</td>
</tr>
</tbody>
</table>

*Scale: 1 = low, 5 = high
On the other hand, we found no differences in enterprise data quality scores on the basis of Carnegie class, research versus teaching mission, FTE enrollment size, or public/private control. Higher enterprise data quality, it appears, doesn’t just happen, and it isn’t just the preserve of a certain kind of institution. It flourishes in environments where training, understanding, and central IT involvement with other units take place.

**Data Quality Improvement Initiatives**

We found a lot of activity relating to data quality improvement. Three in 10 institutions had a data quality improvement initiative under way at the time of our survey, and an additional 4 in 10 said they had completed at least one such initiative in the past 24 months (see Figure 4-3). Another 13.9% were planning one. This is probably a floor for activity, since we suspect that respondents probably would not regard routine data-quality-related work as rising to the level of an “initiative.” The institutions that reported completed, planned, or in-progress initiatives seemed to perceive greater pressures from outside central IT than did institutions with no initiatives. They averaged higher agreement that users and executive leaders were demanding improved institutional data management capabilities and were more likely to agree that improved data management was a strategic priority.

As for more specific incentives, as Figure 4-4 shows, the most common factor causing institutions to undertake data quality initiatives was improving reporting and analysis. Initiatives related to specific application or data warehouse/data mart implementations or upgrades were less common, though roughly half of respondents reported them in each case. Leadership mandates for data quality improvement were the least often selected among our five specific factors, though even so, slightly over a third of respondents chose them.

IT and administrative officers dominated the results from a question asking those respondents at institutions with completed, planned, or in-progress data quality initiatives to identify who sponsored the initia-
tives (choosing all that applied from a list of nine options). As Figure 4-5 shows, the CIO or equivalent was by far the most common sponsor, and administrative executives followed. The top five sponsors were all either from IT or administration. Presidents were named as sponsors by a surprisingly high 25% of respondents with initiatives, and they were a little more likely to be sponsors at institutions whose respondents agreed or strongly agreed that leadership was demanding improved institutional data management. Though academic titles fall toward the bottom of our chart of sponsors, overall 40.2% of respondents in the active initiative categories named either an academic executive or other academic management, or both, as sponsors (this percentage does not include presidents).

Where initiatives have been completed, respondents do not seem to feel they have simply patched problematic data. Asked if their data quality initiatives resulted in process changes that improved data quality throughout the data life cycle rather than just fixing bad data, 68.4% agreed or strongly agreed, and only 11.1% disagreed or strongly disagreed (see Figure 4-6). Mean agreement was 3.62 on our 1-to-5 agreement scale, and the median response was 4.0.6

**Data Analytics**

Much institutional data originates in the need to carry out a transaction—to register a student in a course or pay a bill. Administrative systems in the last generation have worked a revolution in getting these mundane jobs done, not just by automating and accelerating processes, but also by empowering constituents to directly address their needs in real time, without the once-feared trip to the administration building.

But modern administrative systems have promised something more: the ability to aggregate all that transactional data into an accurate picture of a dynamic, living institution. In principle, the data flowing through today’s relational databases and sophisticated system architectures should allow administrators not just to put the right student in the right course, but also to assess the student’s, or the course’s, history and compare it with the history for all other students or courses. With so much of the institution’s activity mediated through shared systems, it should be possible to mix and match data elements across time and organizational boundaries and to analyze them not just to find past patterns and current performance but also to create informed models and scenarios looking to the future. Likewise, it should be
Figure 4-5. Sponsors of Data Quality Initiatives (N = 256)

Figure 4-6. Data Quality Initiatives Resulted in Process Changes (Institutions Completing Initiative in Past 24 Months, N = 117)
possible to distribute this information in role-appropriate ways and facilitate its use within the institution’s decision-making culture. But are institutions really leveraging institutional data and systems in these ways?

We pursued this question by asking respondents about how their institutions approach reporting, modeling, analysis, and decision support—that is, the technologies that provide access to data and the tools that support operational reporting, institutional decision making, and regulatory compliance. For the sake of convenience, we refer to this set of technologies as analytics tools, and the broader capability to collect, analyze, and act on institutional information as analytics.

An ECAR study of academic analytics published in 2005 found institutions relying most often on transactional systems for reporting and analysis, and very few whose primary use of analytics was for such advanced capabilities as predictive modeling or scenario building. In the current study, we returned to some of the topics covered in that earlier work to see where the state of analytics practice stands today and how it has or hasn’t changed in the last four years.

Management Information Climate

One commonly mentioned barrier to the wider adoption of analytics within institutions is that leaders are accustomed to an intuitive management style and therefore aren’t really committed to evidence-based decision making. Another is that institutions lack a commitment to making management information widely available, deferring instead to the turf-protecting restrictions of data owners. We wanted to investigate whether our respondents shared these perceptions and to what extent their agreement about them was associated with other aspects of data management.

On the whole, respondents tended to agree that institutional leadership was committed to evidence-based decision making and that their institutions were committed to making management information widely available. As Table 4-4 shows, their mean agreement on both items was nearly identical and was more than halfway between a neutral (3) and agree (4) response. Responses were closely correlated; level of agreement on one item tended to be similar to that on the other. Majorities agreed or strongly agreed on both items, and fewer than 14% disagreed or strongly disagreed on either.

Our qualitative interviewees recognized that decisions may be driven by many forces other than empirical data, but on the whole they described constituents receptive to, and even hungry for, management information. “You get a certain number of people who seem to say, ‘Sure, show me the data—as long as it doesn’t challenge or mitigate what I already believe,’” acknowledged Samuel Levy, vice president for Information Resources and Technologies and CIO at the University of St. Thomas. “But that said, I’ve seen significant changes in the past few years.”

Table 4-4. Commitment to Evidence-Based Decision Making and to Making Management Information Widely Available

<table>
<thead>
<tr>
<th></th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>My institution’s leadership is committed to evidence-based decision making.</td>
<td>3.58</td>
<td>281</td>
<td>0.957</td>
</tr>
<tr>
<td>My institution is committed to making management information widely available.</td>
<td>3.58</td>
<td>282</td>
<td>0.967</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
decisions based on objective analysis of data in the key business units over the years. I know this because if there’s any disruption in the decision support systems, the users let me know immediately.”

Likewise, interviewees emphasized the important role of executive leadership in setting the tone for a culture of evidence. At Eastern Mennonite University, an upcoming accreditation review led the institution’s provost to press for better ways to empirically demonstrate continuing improvement, an accreditation criterion. The result was an annual institutional effectiveness report process that requires departments to identify data-driven outcomes and measure progress toward them. The university’s Information Systems organization provided technical support to build the reporting application and derive data from multiple internal and external systems, and it has a seat on EMU’s Institutional Effectiveness Committee. “As the departments see the data that’s available and become more data oriented, they move toward more standardized rubrics and test measures where applicable,” said Jack Rutt, director of Information Systems. “It’s becoming an institutional culture, practice, and habit, and there are ways IS can facilitate it, like moving next to dashboard reporting formats.”

Although agreement about leadership and institutional commitment was widespread, we did find some differentiating factors:

- Respondents who more strongly agreed that improved data management was a strategic priority at their institution also tended to agree more strongly with these two items.

- Late technology adopters averaged lower agreement on both items than did mainstream and early adopters, suggesting that more aggressive technology adoption may feed these leadership and institutional commitments or be in part an expression of them.

- Though we did not find an association between data quality score and either leadership commitment to evidence-based decision making or institutional commitment to making management information widely available, where respondents more strongly agreed that executives understood data quality issues, they tended also to more strongly agree on these two items.

We did not find a lot of difference on the basis of demographic factors, but there was some variation by Carnegie class on institutional commitment to making management information widely available. Baccalaureate institutions averaged the lowest agreement (mean 3.33), and associate’s institutions averaged the highest (mean 4.00); doctoral and master’s institutions were in between.

There are, of course, many nuances to institutional culture, and our questions may not have captured some that are relevant to the role of analytics in institutional administration. Our questions, for example, asked about commitment rather than realization, and they did not specifically name analytics as a vehicle for achieving either evidence-based decision making or wide availability of management information. However, most of our respondents seemed to perceive that leaders are far from indifferent or hostile to an evidentiary approach to decision making and that their institutions are committed to making management information available.

### Analytics Tools: Data Stores

ERP systems optimized to carry out transactions fast in real time aren’t well adapted to aggregating that information and making it available for complex, multidimensional reporting and analysis. Institutions that want to go beyond “canned” ERP reports without getting bogged down in programming-intensive custom reporting typically find it necessary to move transactional data into more analytically oriented data stores. These can
be highly focused or more general in nature, and they may stand independently or feed one another, but as a whole they make up a foundation for a systematic approach to institutional analytics.

Our survey asked about the use of three kinds of data stores, one of them in two different usage contexts:

- **Operational data stores (ODSs)**—databases containing information collected from transaction systems, typically from multiple feeds but concentrating on shorter-term data and with relatively limited scope. Because ODSs may be used directly for transactional reporting or for staging of data to another data store, our survey asked about both uses.

- **Data marts**—specialized data stores that usually are built with a particular purpose or subject in mind.

- **Data warehouses**—typically fairly comprehensive data stores with longer-term data tracking and more general-purpose analysis in mind. Besides receiving feeds from ODSs and data marts, data warehouses may integrate historical data from across multiple generations of transaction systems.

We asked respondents to describe the status of each data store type (and both uses of ODSs) as part of their institution’s reporting, modeling, analysis, and decision support platform. The statuses respondents chose from were as follows: in use at the institution-wide level; in use at the school, college, or department level (also referred to here as the distributed level); currently in implementation; implementing within 12 months; considering implementing within 24 months; or not considering it.

Use of these kinds of data stores was widespread but far from universal (see Figure 4-7).
Roughly one-third of institutions reported data warehouses, ODSs for transactional reporting, and ODSs for staging in use at an institution-wide level, and in each of these cases another 7% to 10% reported such use at a distributed (school, college, or departmental) level. Data marts were not as common; about one in five institutions reported them at the institution-wide level, and another 10.3% at a distributed level. These figures may slightly understate usage, since we received relatively high “don’t know” responses, ranging from 6.6% to 8.6%, for all items except data warehouses (2.0%).

Usage of each type of data store at either the institution-wide or distributed level varied a good deal by Carnegie class. As Figure 4-8 shows, usage levels were highest for all four data store items at doctoral institutions, whereas usage at other Carnegie types was lower and similar across Carnegie types. Incidence was also associated with FTE enrollment size; smaller institutions reported lower incidence of each type of data store than larger institutions.

Differences of each data store type’s incidence by Carnegie class and institution size don’t tell the whole story, however, because there was considerable variance by Carnegie class in the number of items respondents reported on average. Doctoral institutions were the most likely to report all four items (35.6%), and BA institutions were the least likely (7.7%). Overall, only 16.8% of respondents said they had all four items in use, 18.1% reported only one, and 40.5% didn’t report using any of the four.9

Data store infrastructure, then, was fairly modest among our respondent population, with 6 out of 10 institutions reporting one or none of the items we asked about, and multiple-item users were concentrated disproportionately among doctoral and larger non-doctoral institutions. Presumably, those institutions not using these technologies rely on their transaction systems for reporting, or on shadow applications and databases that didn’t fall into our categories. The absence of these data store infrastructure items may make deep-diving institutional analysis more difficult in many cases, but on the other hand, we did not find any associations between having these items in use and better results in the data management outcomes reported in Chapter 8.

Figure 4-8. Incidence of Institution-Wide or Distributed Data Stores, by Carnegie Class*
Active implementation projects and expressed plans to implement suggest potential for usage growth in all these data store types. At the same time, however, a comparison with the results of our 2005 academic analytics survey reveals no significant differences in usage (up or down) for any of the data store items, despite the fact that many of the 2005 respondents, too, reported implementations planned and in progress. (This comparison was limited to the 112 institutions responding to both the 2005 and 2009 surveys.) For some reason, our respondent institutions don’t seem to be reporting much progress on these elements of data analytics infrastructure.

**Analytics Tools: Data Manipulation, Documentation, and Reporting**

To round out our view of institutional reporting, modeling, analysis, and decision support platforms, we looked at the status of tools used to get data into and out of data stores, maintain data quality, and document data. In addition, we wanted to know about reporting solutions that might be separate from an institutional analytics platform. The items our survey asked about were:

- metadata servers or data dictionaries for documenting characteristics of data elements;
- data cleaning tools that help maintain accuracy and consistency of data;
- extract, transform, and load (ETL) tools, which permit the harvesting of data from transactional systems (and other databases) and its formatting and transfer between data stores; and
- vendor-supplied reporting solutions delivered as part of another system—for example, integrated ERP reporting modules.

Among these, vendor-supplied reporting solutions were by far the most commonly reported in use, with 36.3% of respondents reporting institution-wide use and another 15.8% reporting use at a distributed level. This made it the only analytics tool we asked about that was reported in use at some level by a majority of respondents. Although native ERP “canned report” tools falling under this heading can be used in the absence of a separate analytics platform, it appears that our respondents also reported vendor-supplied reporting solutions that were used in conjunction with analytics platform elements. For every analytics tool we asked about (i.e., those in Figures 4-8 and 4-9), use was higher among respondents reporting a vendor-supplied reporting solution in use than among those who did not report such use.

Among the other items in Figure 4-9, ETL tools were the most prevalent and data cleaning tools the least. For two of the items, data cleaning tools and metadata servers, data quality scores were slightly higher among respondents reporting the item in institution-wide or distributed use than among those who did not. Some evidence also suggests that metadata servers contribute to enterprise data coherence: Respondents reporting them in institution-wide or distributed use averaged much higher agreement that each major data element had a single definition recognized across the institution (mean 3.63) than did those not reporting such use (mean 3.08).  

Metadata servers, data cleaning tools, and ETL tools were reported in use more often at institutions with higher student FTE enrollments than at smaller institutions, but use of the more prevalent vendor-supplied reporting solutions was not significantly associated with FTE enrollment size or with Carnegie class or public/private control.

As with data stores, the status of metadata servers, data cleaning tools, and ETL tools did not differ significantly between 2005 and 2009 among the 112 institutions that answered both academic analytics surveys. Once again, however, vendor-supplied reporting solutions were the exception: They were the
only analytics tool item we asked about that showed a significant status change over the four-year period. Institution-wide use increased from 30.5% in 2005 to 44.2% in 2009, and distributed use grew even more, from 5.3% to 20.2%, while the proportion of those saying they were not considering the item declined from 24.2% to 15.4%. Although our survey questions may have missed some elements of investment, our respondents apparently are choosing to deploy or upgrade commercial reporting solutions more often than the other analytics platform elements we asked about.

Analytics Use and Users

Our survey questions about how institutions use analytics were based on the notion that such use might follow a hierarchy from more basic to more sophisticated uses. To learn how far our respondent institutions may have progressed along these lines, we asked them to describe the primary use of reporting, modeling, analysis, and decision support tools at their institutions, using one of five stages:

- Stage 1: extraction and analysis of transaction-level data,
- Stage 2: analysis and monitoring of operational performance (e.g., via dashboards),
- Stage 3: what-if decision support (e.g., scenario building),
- Stage 4: predictive modeling and situations, and
- Stage 5: use of information to automatically trigger a business process
(e.g., automatically scheduling an adviser appointment with an at-risk student).

In addition, respondents could indicate that their institution was not an active user of reporting, modeling, analysis, and decision support.

We found primary use of analytics reported mainly in the two “earliest” of these stages, extraction and analysis of transaction-level data, and analysis and monitoring of operational performance (see Figure 4-10). Almost 6 in 10 reported the former, and about 2 in 10 reported the latter; if we add to them the 9.5% who said their institution was not an active user, this leaves only 12.5% across all three of the most advanced stages.

We found a similar concentration in the first two stages in our 2005 academic analytics study, and indeed there was no statistically significant difference between the 2005 and 2009 results among the 112 institutions that answered both surveys. Finding the relatively simpler stages of use to be the dominant ones is not surprising; after all, these are the foundation for more sophisticated stages, and our question’s focus on primary use may, as we note below, be obscuring more sophisticated use of analytics for particular purposes or in certain departments. At North Carolina State, for example, Marc Hoit noted that the analytics needs and capabilities of departments vary widely. “Some departments, like education and engineering, have much stronger accreditation requirements than others and have done much more with data,” he reported. “The financial people are generating all kinds of scenarios on budget cuts. About 80% of users just want their standard reports; another 10% to 15% do a little more refined analysis but aren’t complex statistically; and then there are the 5% to 10% that really dig in.”

Though this may well be a common scenario, we find it noteworthy that no trend toward an institutional use of analytics primarily devoted to higher-order and more strategic management activities—to pursuing scenarios, modeling or predicting outcomes, or integrating information dynamically with

![Figure 4-10. Primary Use of Analytics (N = 305)](chart.png)
business processes—is evident in our results. In its primary uses, analytics seems to remain grounded in transactional and operational performance monitoring.

A closer look at certain specific advanced uses of analytics shows, however, that some degree of use of advanced analytics is widespread. Majorities reported at least occasionally employing all of the six advanced analytics use cases that we asked about (see Figure 4-11).

Still, much of the use seems infrequent, and most of the uses mentioned were “almost never” used by substantial minorities. Monitoring performance by, for example, comparing budget to actual data (a stage 2 use) was by far most often reported in use: Only 6.4% of respondents said their institutions almost never make use of it. The two predictive modeling (stage 4) uses produced mixed results: Identifying students at risk was the second most common use, with 15.2% reporting almost never doing so, but modeling the potential impact of strategic decisions was much less common, with 38.3% almost never doing it. The auto-trigger (stage 5) uses we asked about were also less common, with between 34.0% and 39.9% reporting almost never using them.

Means of responses on our 5-point scale (which ranged from 1 = almost never to 5 = almost always) reflected these differences:

---

**Figure 4-11. Selected Uses of Advanced Analytics**

- **Monitor operational performance such as budget to actual data (N = 297):**
  - Almost always: 26.6%
  - Usually: 11.7%
  - Sometimes: 4.5%
  - Occasionally: 6.8%
  - Almost never: 10.7%

- **Identify students who may be at risk academically (N = 290):**
  - Almost always: 33.8%
  - Usually: 20.0%
  - Sometimes: 21.4%
  - Occasionally: 17.0%
  - Almost never: 12.6%

- **Model the potential impact of strategic decisions (N = 290):**
  - Almost always: 39.8%
  - Usually: 17.0%
  - Sometimes: 22.4%
  - Occasionally: 16.4%
  - Almost never: 19.9%

- **Automatically alert appropriate officials when a financial indicator falls outside a desired range (N = 294):**
  - Almost always: 39.9%
  - Usually: 16.4%
  - Sometimes: 22.5%
  - Occasionally: 12.6%
  - Almost never: 19.9%

- **Automatically alert appropriate officials when an enrollment metric falls outside a desired range (N = 293):**
  - Almost always: 34.0%
  - Usually: 19.3%
  - Sometimes: 19.9%
  - Occasionally: 18.6%
  - Almost never: 18.6%

- **Automatically alert an appropriate official when an academic intervention with a student is warranted (N = 291):**
  - Almost always: 34.0%
  - Usually: 15.2%
  - Sometimes: 38.3%
  - Occasionally: 39.8%
  - Almost never: 39.9%
Views from the Qualitative Research: Toward a Data-Driven Culture

Qualitative interviewees chosen for the high marks they gave their institutions’ ability to get maximum business value from data tended to describe an evolutionary pattern that began with an institutional commitment to management metrics rooted in reliable and accessible data, which cleared the way for enhanced analytics capabilities and cultural change.

University of St. Thomas

An ERP implementation with a shared centralized database set the stage. “Free the captive data’ was the catchphrase during the implementation,” said CIO Samuel Levy. “We divested the institution of the cultural artifact of a small number of people controlling access to data and analysis.” The university instituted a “shared data owners” group to work out data ownership and standardization decisions, and when it added an operational data store, a data warehouse, and analytics tools, it also provided IT staff to assist users with analytics. “As you can imagine, at a university you have some folks who are very good at that kind of thing and others, perhaps in the smaller units, who need help,” Levy added. “The larger departments tend to push the data further, asking more questions, and thereby developing more expertise.” As the institution’s analytics capability improved, it began to be integrated more deeply into administration. “The data have all been tied to institutional budgeting and planning processes, and the goals of the president, vice presidents, and on down are increasingly stated in empirical terms,” Levy noted. “Here, you don’t go question-begging into meetings—you walk in with the numbers. It’s no longer valid to make excuses about the data not being available or accurate.”

University of North Texas Health Science Center

The university’s health science center once had an “old school academic atmosphere,” reported Renee Drabier, vice president for Information Resources and Technology and CIO at UNTHSC. New presidential leadership and the senior executive team instituted a requirement that departments develop outcome measures, align with institutional goals, and report on their success. To “close the assessment loop,” Drabier said, the institution adopted an assessment management system oriented toward performance improvement and accreditation standards. “You put in your department goals aligned with institutional goals and tactical initiatives, document how you’re going to measure them, when you’re going to achieve them, and write in outcomes and lessons learned and how you’ll proceed from there with updated processes and measures.” This in turn has fed UNTHSC’s investment in executive dashboards. The campus developed web-based dashboards driven by extracted ERP data, which has led to consistent definitions for key data elements across UNT campuses. “The items we monitor right now extend down to the VP, dean, and college level,” Drabier said. “The use of analytics continues to be expanded to reach into all departments so that each has its own dashboards that align with university goals and reporting needs.”

University of Central Missouri

Describing the data-based decision-making culture at the university as “strong,” CIO Russell Helm noted that the university’s original decision to implement an integrated ERP suite and subsequent analytics enhancements arose from leadership demands for better data. “We relied on them to take the lead,” Helm said. The chief champions of the project were the provost and chief financial officer, who “sold the project to the cabinet, the board of governors, and the campus. They orchestrated the requirement that when the Banner system went in, Banner data would be the institution’s official data, period. They wanted sound data.” The university has subsequently added a data warehouse and a number of data marts. The institutional research division, which reports to the provost, has had the main responsibility for managing time-series information available from the data warehouse. But thanks to new high-level key performance indicators being developed by the president and board, Helm said, “I wouldn’t be surprised if there were a push for analytical tools at the staff and managerial level to make the data easier to use at all levels of the institution.”
The monitoring operational performance responses averaged 3.62, more than halfway between “sometimes” and “usually,” and identifying students at risk averaged 3.08 (“sometimes”). The strategic decision impact modeling and auto-trigger item means were all below 2.50.

We did not find any significant differences in advanced analytics use by demographic factors such as Carnegie class, FTE enrollment size, or public/private control, though we did find that the auto-trigger items were on average more frequently used by institutions reporting higher data quality scores. This makes sense, since presumably institutions would be reluctant to implement automatic alerts to officials unless they felt confident about the data that fed the process. In addition, where respondents agreed more strongly that institutional leadership was committed to evidence-based decision making, they tended to report more frequent use of operational performance monitoring, automatic alerts regarding financial indicators, and modeling the impact of strategic decisions.

Perhaps our most important finding relating to these advanced analytics use items was that for every one of them, more frequent mean use was associated with higher respondent agreement about getting maximum business and academic value from institutional data. We report the results relating to these key data management performance outcomes in detail in Chapter 8. Given this association, we were surprised to find no significant increase in the frequency of use of any of these items among the institutions that responded to both the 2005 and 2009 surveys.

There was a similar continuity in respondents’ reports about who uses reporting, modeling, analysis, and decision support tools. Respondents choosing up to three “most active” users from a list of 12 user entities most often selected institutional research, central admissions/enrollment management staff, and central business/finance administrative staff (see Figure 4-12). Despite some apparent differences in rank order caused by percentage differences too small to be

Figure 4-12. Most Active Users of Analytics (N = 309)
meaningful, these results were very similar to those reported in the 2005 analytics study, and among institutions responding in both 2005 and 2009, we found no statistically significant changes in the activity of the various user entities.

Just the same, we did find some variation, particularly relating to the role of fundraising/advancement staff and institutional research. Private institutions were far more likely to identify central fundraising/advancement as a most active user (27.0%) than were public institutions (3.5%), presumably reflecting their relatively greater dependence on endowments for operating revenues. As for institutional research (IR), it was most frequently reported as a most active user at associate’s institutions (80.4%) and least often at BA institutions (49.2%), possibly reflecting less prevalence of IR units, or simpler IR analytical needs, among liberal arts colleges.

**Analytics Training**

Respondents were far from enthusiastic about the analytics training provided by their institutions. Asked their level of agreement that their institution provides effective training to users of reporting, modeling, analysis, and decision support tools, they averaged a 2.77 response, below neutral (3) on our 5-point agreement scale. Only about one in four agreed or strongly agreed (see Figure 4-13). This was consistent with previous findings; there was no significant difference in agreement about training effectiveness between our 2005 and 2009 surveys among the 112 institutions that answered both.

This is an unfortunate situation, since as we report in detail in Chapter 8, among all study variables, effective training in analytics had some of the strongest associations we found with good data management outcomes, and we found particularly strong relationships between training and agreement that the institution gets maximum business and academic value from institutional data. For the most part, training had a stronger relationship with positive outcomes than did deployment of analytics platform items or use of advanced analytics techniques.

![Figure 4-13. Institution Provides Effective Training to Users of Analytics Tools (N = 279)](image-url)
So what distinguishes institutions that report stronger agreement about this evidently influential variable? Our results can’t demonstrate causality or the direction of any connection that exists, but several factors we’ve already mentioned in this chapter do tend to be seen alongside greater agreement about training effectiveness. As Table 4-5 shows, institutions perceiving better data quality (as reflected in data quality scores), those seen as committed to evidence-based decision making, and those seen as committed to making management information widely available also tend to report higher analytics training effectiveness. We suspect that such factors are all components of a general cultural disposition in favor of making effective use of analytics, which in turn drives funding, availability, interest in, and expectations of pursuing analytics training.

Analytics Upgrade Plans

Hope, they say, springs eternal, and if we found little evidence that institutions have significantly enhanced their analytics infrastructure or use since 2005, our respondents nonetheless tended to report intentions to improve their capabilities. One item in evidence is the substantial percentages of respondents saying they will, or may, deploy the assorted analytics platform items shown in Figures 4-7 and 4-9. Another, more direct, indicator was a solid majority of respondents who agreed or strongly agreed that their institutions planned to significantly upgrade their reporting, modeling, analysis, and decision support capability in the next two years (see Figure 4-14).

Averaging a 3.52 response (midway between neutral and agree), the responses to this statement were remarkably diffuse. We found no associations between agreement about upgrading analytics and any demographic variables such as Carnegie class, FTE enrollment size, research mission, or public/private control, nor, more surprisingly, any relationship between perceived demands for improved data management capabilities from leadership or users. Neither did we find institutions with lower data

Table 4-5. Training Effectiveness, by Data Quality Score, Commitment to Evidence-Based Decision Making, and Commitment to Availability of Management Information

<table>
<thead>
<tr>
<th>Item</th>
<th>Item Value</th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data quality score</td>
<td>Low (less than 2.5)</td>
<td>2.33</td>
<td>60</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td>Medium (2.5–3.5)</td>
<td>2.79</td>
<td>140</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td>High (greater than 3.5)</td>
<td>3.10</td>
<td>78</td>
<td>1.014</td>
</tr>
<tr>
<td>My institution's leadership is</td>
<td>Strongly disagree/disagree</td>
<td>2.12</td>
<td>34</td>
<td>0.977</td>
</tr>
<tr>
<td>committed to evidence-based decision making</td>
<td>Neutral</td>
<td>2.71</td>
<td>79</td>
<td>0.908</td>
</tr>
<tr>
<td></td>
<td>Agree/strongly agree</td>
<td>2.97</td>
<td>163</td>
<td>0.990</td>
</tr>
<tr>
<td>My institution is committed to</td>
<td>Strongly disagree/disagree</td>
<td>2.26</td>
<td>39</td>
<td>1.019</td>
</tr>
<tr>
<td>making management information widely</td>
<td>Neutral</td>
<td>2.35</td>
<td>72</td>
<td>0.937</td>
</tr>
<tr>
<td>available.</td>
<td>Agree/strongly agree</td>
<td>3.10</td>
<td>166</td>
<td>0.909</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
quality scores, or those reporting lower agreement about getting maximum business or academic value from institutional data, significantly higher or lower in agreement about upgrading.

Accounting for the absence of a relationship is difficult. But given that expectations of analytics platform improvement among our 2005 respondents were not (except for vendor-supplied reporting solutions) clearly realized among the same respondents in 2009, and also given our inability to find a link between upgrade agreement and logical indicators of a need to improve, we are led to speculate that these expectations of analytics improvement derive more from optimism and a general hope for progress than from concrete plans.

### Data Quality and Analytics: How Can We Do Better?

Assembling a timely, complete picture of how the institution is functioning and where it’s headed, and using that information to manage and lead, has been something of a Holy Grail for higher education over the years. Yet this vision continues to be surprisingly difficult to realize. Many of our respondent institutions seem to be relying on their ERP systems for their business reporting needs despite the fact that online transactional systems aren’t well architected to support deep dives into data. Our respondents also seem distinctly unenthusiastic about most facets of enterprise data quality, averaging only a 3 response on a scale of 1 to 5 across eight different data-quality-related items. We found rather modest infrastructure beyond ERP for data analytics and little change in the analytics infrastructure landscape between our 2005 and 2009 surveys.

Why such unprepossessing results so many years into the Information Age? Though we didn’t ask respondents directly about obstacles to better data quality and more robust analytics, we found some suggestive clues. As we reported in Chapter 3,
the most commonly reported barrier to investing in data management was lack of funds, whereas the top drivers for investment were improving data security and regulatory compliance. Improving operational efficiency and enhancing data-driven decision making—items that rely heavily on data quality and analytics—were distinctly secondary drivers. It may be that an understandable concern for data security is taking resource precedence over other kinds of improvements to the data environment.

A second possible explanation is that institutions see their transactional ERP systems as “good enough” platforms for maintaining data quality and serving basic reporting needs. Among the eight detailed data-quality-related items we asked about, those that elicited the strongest mean agreement—that each major data element has an identified system of record and a single definition recognized across the institution—were the ones most likely to be improved by the existence of an integrated ERP system with a shared database. The data quality items that fared worst, such as value changes propagating across non-enterprise systems and processes for documenting and reviewing data quality issues, were those that depended most on cross-system integration and cross-department coordination and feedback.

Likewise, extraction and reporting of transaction-level data was by far the most common primary use cited for analytics tools, and 4 in 10 institutions reported no use of the data store types we asked about, presumably relying instead on their ERP transaction systems. Although ERP systems demand a certain amount of interdepartmental communication and cooperation, especially in their deployment phase, we speculate that this kind of interaction is focused mainly on transaction workflows. Institutions may find it difficult to adapt these to the feedback loops, big-picture perspectives, and sophisticated techniques involved in both systemic data quality improvement and model- or scenario-oriented analytics.

It’s possible as well that mediocre data quality and unambitious analytics are mutually reinforcing circumstances. In principle, it seems reasonable that the more data that is fed into advanced analysis for high-stakes decision making, the more likely the data will be scrutinized and improved; contrarily, unreliable data makes aggressive analysis difficult or pointless. Our empirical evidence for this is spotty, however. Institutions with higher data quality scores tended to report more frequent mean use of advanced auto-trigger analytics techniques, but the same relationship did not hold up for other techniques. Data quality scores were also associated only with usage of a few analytics tools.

One might also speculate that cultural issues, so often described as more important than tools or resources, are somehow undermining data quality and analytics development. We did find that institutions reporting stronger agreement about leadership commitment to evidence-based decision making were more likely to be early technology adopters and averaged more frequent use of a few advanced analytics techniques. But most respondents did agree that their institution’s leadership had such a commitment, and as we reported in Chapter 3, even more agreed that improved data management was a strategic priority. These findings hardly scratch the surface of possible cultural dimensions, of course, and the signs of inadequate interdepartmental coordination that we noted above could be considered a cultural issue deeper down in the organization.

What is clear, however, is that both data quality and use of advanced analytics are strongly related to our study’s most prominent data management outcomes, getting maximum business and academic value from institutional data. These factors, in turn, are positively associated with things that institutional administrators can do something about:
providing data quality and analytics training, pursuing an active partnership between IT and business units, and undertaking data quality improvement initiatives. For institutions with lots of problems to solve and not a lot of resources to bring to bear, these would be good places to start looking for improvements in institutional data management.

Endnotes
1. The issues we cover in this section include reference and master data management as well as data quality management. For convenience, we refer to the whole range of issues under the term data quality. See Mark Mosley, “DAMA-DMBOK Functional Framework Version 3.02,” DAMA International (2008), 11, http://www.dama.org/i4a/pages/Index.cfm?pageID=3548.
2. Cronbach’s alpha for the eight items included in the score was 0.881.
4. N = 85, SD = 1.090, and N = 164, SD = 0.621, respectively, for these groups.
5. N = 309, SD = 0.863.
6. N = 117, SD = 0.955. Because of the skewed distribution of these responses, the median is the more reliable summary measure.
8. SD = 0.998 and 0.913, respectively.
9. The results in this paragraph and in Figure 4-8 exclude respondents who answered “don’t know.”
10. SD = 1.049 and 1.140, respectively.
11. SD = 1.188 and 1.268, respectively.
12. SD = 1.009.
5

Data Stewardship, Security, and Policies

Discretion is the perfection of reason, and a guide to us in all the duties of life.
—Sir Walter Scott

Key Findings

- Data responsibilities tended to follow an organizational division of labor within institutions. Business and academic units most often had primary responsibility for data accuracy, access decisions, and privacy, whereas central IT units tended to have primary responsibility for data formats, metadata management, and notification of affected parties following data breaches.
- Only one-third of institutions reported formal data stewardship policies. Institutions with such policies were more likely to have formal processes for classifying data according to risk, and they had slightly higher mean enterprise data quality scores, but they did not differ from other institutions in respondent agreement about the security of data from unauthorized access.
- The data-related policy portfolios at respondent institutions were modest. Although majorities reported policies on acceptable use of data and employee responsibilities regarding data security, and half had policies regarding notification of parties following data breaches, fewer than half had policies regarding sharing, storage, or transmission of institutional data; data encryption; or cloud-related data practices.
- About 30% of institutions reported at least one data breach in the 24 months prior to the survey that required notification of affected parties. Incidents were more common at larger institutions and more common at doctoral and master’s institutions than at other Carnegie types.

One of the ruder shocks of the Internet era has been the discovery that reducing data friction—the difficulty of finding, copying, and transporting information—often creates a corresponding need to enhance and formalize data security. Not long ago, the greatest danger to institutional information security (barring the odd spy with a pocket micro-camera) was the departmental Xerox machine, and a few locks and keys, a logbook, and an electromechanical copy counter constituted reasonable protection from data breaches. Today, data practically begs to be copied, and a thumbnail-sized key disk can collect more information in a few seconds than a fleet of photocopiers could reproduce in a year.

©2009 EDUCAUSE. Reproduction by permission only.
Small wonder that data security has been among the top three items named as critical for strategic success in EDUCAUSE’s annual Current Issues Survey every year since 2003.1 But taking good care of institutional data means much more than just keeping it out of the wrong hands. It also means ensuring that data is accurate and timely, that it’s formatted in ways that make sense, and that when it does leak, the institution responds appropriately. This range of responsibilities, sometimes conceptualized under the name stewardship, raises many questions about who should shoulder which tasks and how they should be carried out. In this chapter, we investigate how institutions approach these challenges.

Business/Academic Unit and Central IT Data Responsibilities

“Data,” writes Burton Group analyst Noreen Kendle, “is a representation of a business.”2 This simple fact sums up one of the key dilemmas of data management: When should responsibilities for managing data fall on those whose competencies lie in representing and processing information—typically the IT unit—and when should they fall on those who best know the processes that generate, consume, and are managed by that information—the business and academic units? IT administrators rarely wish to second-guess business managers’ judgments about their own data needs, yet IT professionals can often improve the accuracy, usability, and security of data if they have a hand in managing it.

To find out how institutions divide up these tasks, we asked respondents whether business/academic units or central IT had primary responsibility for each of 10 data management activities. We knew ahead of time that our question was somewhat artificial, because it seemed likely—and results reported in Chapter 3 confirmed—that central IT and business/academic units often work together on data issues. Still, we thought it would be revealing to see how respondents, given the choice, would divide IT and line unit activities.

We found a fairly clear division of labor, with aspects regarding the content and use of data generally assigned to the business/academic units and those concerned with data formatting or classification to central IT. This was especially evident in activities related to data quality and data specification. As Figure 5-1 shows, the great majority of respondents told us that business and academic units had primary responsibility for data accuracy, timeliness, and fitness for purpose, whereas only one in five or fewer said central IT was primarily responsible. (A small number of respondents in each case said that neither was involved, or that the item was not applicable at their institution.) Two items involving the need for technical skills—specifying data formats and metadata management—followed an almost exactly opposite pattern: The great majority of institutions said central IT was primarily responsible (though the “neither/not applicable” responses were more numerous for these activities).

Business and academic units also predominately held primary responsibility for data access; at about three-quarters of institutions, they were responsible for deciding both who has routine access to data and who could have access on an as-needed or exception basis (see Figure 5-2). This makes sense, since the business and academic units are the ones most likely to know both the nature of the data and the parties asking to use it. Still, given the identity management implications of access, it seems likely that close coordination with central IT is common. At Georgia State University, for example, “access is automated based on being in a role,” said J. L. Albert, associate provost and CIO. “Each of the major function areas has a data steward who applies granularity of access. They set the rules,
Institutional Data Management

Figure 5-1. Primary Responsibility for Data Quality and Specification Activities

Figure 5-2. Primary Responsibility for Data-Privacy-Related Activities
and IT implements them.” A similar process applies at the University of North Texas, according to Maurice Leatherbury, acting vice president for information technology and CIO. “Authorization is primarily a business unit concern,” said Leatherbury, “though IT sometimes brokers approvals where people on campus don’t understand the distributed nature of data stewardship.” Another indication that access decisions can be a multiparty process lies in the fact that 46.4% of institutions told us they have a governance structure for setting data access policy.

Classification of data for risk or sensitivity was most often reported as a central IT responsibility. However, with 38.4% of institutions saying business/academic units were primarily responsible and a relatively high “neither/not applicable” response, this was the most evenly divided of the activities we asked about.

Somewhat surprisingly, about two-thirds of institutions said central IT was primarily responsible for notifying affected parties following a data breach, whereas only about a quarter selected business/academic units. This was something of a contrast with our results’ overall picture of business/academic units as the main data owners. However, the virtues of a centralized notification process are easy to see: It can help ensure that regulatory protocols are followed to the letter, and it can enhance the enterprise visibility of data breaches that might otherwise be known only to local units. Also, where breaches result from technical vulnerabilities, central IT is in the best position to carry out security forensics and apply solutions.

Although a pattern of divided labor was fairly evident in these results, few other patterns were. We didn’t find significant differences in the way institutions assigned responsibilities on the basis of Carnegie class, FTE enrollment size, or public/private control, nor any relationships with enterprise data quality scores, the existence of data stewardship policies, or better or worse performance in data management outcomes.

**Data Stewardship Policies**

If, as we noted above, data represents the business of an organization, it is very much a traveling representative. Optimizing business processes and coordinating them at multiple levels of business often requires that data move through a lot of hands. With the advent of real-time integrated administrative suites and sprawling institutional websites, movement of both structured and unstructured data has expanded astronomically. But these enormous increases in the availability of data have not been met by any corresponding decrease in the need for security, privacy, and accuracy—quite the contrary.

To combat this danger, modern data management has introduced a concept of data stewardship—the formal definition and assignment of responsibility for the management of data resources. Acknowledging that data may flow through few or many channels, data stewardship explicitly describes who is primarily responsible for seeing to it that data represent what it’s supposed to represent, what rules and regulations apply to it, and who may see or manipulate it.

For data management expert Robert Seiner, formality rather than assignment of responsibilities is the real key to successful data stewardship. Seiner notes that de facto stewardship often arises out of the logic of the business and its organizational structure. “Data stewards are already there,” Seiner writes. “They may not know they are stewards and the organization may not recognize them as such, but there are already people in your organization that have accountability for the management of data. The challenge is getting them to participate and operate more efficiently and effectively in these data management roles. The data will not govern itself.”
Our survey examined data stewardship by asking whether institutions had documented policies for it, and where they did, by looking at what kinds of data were covered and to whom data steward responsibilities were assigned. In addition, in our qualitative interviews we talked to interviewees about their experience with data stewardship and their approaches to it. We examine the results in this section.

Data Stewardship Policies
Institutions with documented policies defining data steward responsibilities were the exception rather than the rule among our respondents. Overall, only about one in three reported them. As Figure 5-3 shows, data stewardship policies were most common at institutions with FTE enrollments greater than 15,000, though even among these they constituted only a modest 56.3% majority. At the smallest institutions, those with enrollments of 2,000 or fewer, 15.7% reported data stewardship policies, and roughly one-third of institutions in the size categories in between reported them.

Does having a documented data stewardship policy make a difference in desirable outcomes? We found some limited evidence that it does. Institutions reporting such policies averaged slightly higher enterprise data quality scores than those without them, and they were almost three times as likely to say that they had a formal process for classifying institutional data by risk or sensitivity level (an item we discuss in more detail below). Furthermore, data stewardship policies do seem to help match data with responsible parties: Although 70.7% of respondents at institutions with data stewardship policies agreed or strongly agreed that all institutional data had a designated data steward, fewer than half as many (31.2%) did so where there was no policy. These findings all lend some weight to the idea that good data management practices are mutually reinforcing, an idea that Neil McElroy, dean of Libraries and Information Technology Services at Lafayette College, expressed in one of our interviews. “I suppose we might have understood this better 10 years ago,”

Figure 5-3. Institution Has Documented Institutional Policy Defining Data Steward Responsibilities
he said of changes he witnessed after a “reimplementation” of the institutional ERP system led to systemic improvements to the institution’s data environment. “But it’s all coming together now. Data stewardship depends on shared definitions and data integrity, and it all contributes to more productive work for the college.”

At the same time, we did not find a significant association between having a data stewardship policy and any of the key data management outcomes we asked about, such as getting maximum business or academic value from data, securing sensitive data from unauthorized access, or employee understanding of her or his responsibilities in the use of data. Formalizing data stewardship policies may have practical benefits such as documenting an institutional commitment to responsible data use or permitting better communication about data rules, especially in large institutions. Furthermore, as our finding on data quality scores suggests, data stewardship policies may foster other improvements that in turn ultimately contribute to better data management outcomes. But as far as our evidence permits, we cannot say that documented data stewardship policies are related to either better or worse major outcomes.

### Data Stewardship Scope and Assignments

Most of the advisory literature on data stewardship focuses on structured business data, though in principle the idea might apply to many other kinds as well. When we asked about the kinds of data covered in documented data stewardship policies, we found that although administrative systems data seemed to be the chief concern—every single institution with a data stewardship policy said that this kind of data was covered by it—coverage was often broadly inclusive (see Figure 5-4).

Although administrative systems data ruled the roost, learning management systems data and content, e-mail, and institutional web content were all reported as covered by more than 6 in 10 institutions. Research data was covered less frequently, perhaps due to its distributed nature or because protocols separate from those in the institutional data stewardship policy govern it; even so, a slight majority (52.8%) covered it. Surprisingly, we found no association between research data coverage and Carnegie class or research/teaching mission emphasis. E-mail and learning management systems data and content were most likely to be covered at doctoral institutions and least likely at BA institutions, but on
the whole we found few differences based on institutional demographics.

As for who the data stewards are, we found assignments concentrated among administrative officers and staff and an overall pattern suggesting the mutual involvement of business/academic units and central IT (see Figure 5-5). Almost 9 out of 10 institutions reported business/academic unit managers as being explicitly assigned data steward responsibilities in institutional data stewardship policies. Central IT leadership was the next most often reported (78.1%), and business/academic executives and staff and other central IT participants all were assigned responsibilities by between 68.4% and 72.0% of institutions. There was a clear break between these parties and other institutional officers such as researchers and teaching staff, auditors, and institutional counsel; none of these parties were assigned responsibilities by even half of institutions, and although higher “don’t know” responses for these groups may indicate that our results somewhat understate the actual figures, they also suggest that these groups are less visibly associated with the data stewardship process.

Gartner analyst Ted Friedman (as well as other data management advisers) recommends that “data stewards should reside in the business, not the IT organization.” Do our assignment results, in which central IT officers figure so prominently, suggest that our respondent institutions are violating this best practice?
practice? We did not ask fine-grained questions about the nature of the assigned responsibilities, but on the whole we are inclined to interpret the results as suggesting cooperation and division of duties rather than undue domination by IT over data stewardship.

The data responsibilities results reported in Figures 5-1 and 5-2 suggest a reasonable division of labor in which business and academic units are more often responsible for data content and use, whereas IT is more often responsible for specification and classification. Data stewardship policies have room for both, since they commonly define a variety of data responsibility profiles and a data sensitivity or risk classification scheme that might involve both business unit knowledge of regulatory and use issues and IT knowledge of data formatting and system integration issues. In addition, IT may play a broader role educating users about data security and privacy issues beyond those applying to a particular functional set of data elements.

Our qualitative interviews suggest that institutions implementing data steward policies “get it” when it comes to giving business/academic unit officers the primary role in owning and stewarding data while also giving IT an important supporting role. At Bates College, said Gene Wiemers, vice president for Information and Library Services, “data stewardship is decentralized and the responsibility of the individual departments. But in central IT we see ourselves as educators for creating awareness. We try to teach everyone from data entry clerks to vice presidents about their data-related responsibilities.”

Andrew Lawlor, associate vice president for Technology and Communications at Edinboro University of Pennsylvania, recounts IT’s part in getting data stewardship moving and also its emphasis on business unit ownership. “Our information security policy for administrative information covers data stewardship and defines data custodian roles and other requirements. IT developed it based on practices already in place, and the business managers reviewed it. The data custodians call the shots—for example, if somebody wants access, we require them to get approval from the data custodian, then we enable the appropriate process.” And at the University of San Francisco, CIO Steve Gallagher reported a similar sensibility. “We have designated data stewards who have been identified by the appropriate division lead in each of the business units, and they control access to the various administrative functions and the associated data,” Gallagher said. “IT is the facilitator, not the decision maker, and we’re pleased with that arrangement.”

More worrisome than the promiscuous mixing of business/academic and IT officers in our assignment results is the relatively lower level of involvement by faculty, in their roles as researchers and teachers, and by other officers whose understanding of institutional risk could be vital to successful data stewardship. Not only are academic officers now involved in the digital management of large amounts of regulated data, but also their options for collecting, storing, and sharing those data are growing as consumer-oriented cloud tools increasingly compete with institutionally delivered tools. It may be that the data responsibilities of these positions are being addressed in other ways, of course, but incorporation into a broad-based notion of data stewardship could help send the message that everyone who handles institutional data—not just administrators—has a role in minimizing institutional exposure.

**Data Stewards’ Councils**

Besides their responsibilities in the operational oversight of data, data stewards are sometimes conceived of as subject matter or policy advisers with a role in data governance (or IT governance generally). A council of data stewards is one mechanism for encouraging communication between data stewards and uncovering common practices or problems
Institutional Data Management

The institutional data stewards’ council is, however, a relatively rare body among our respondent institutions. Only 30 of 100 institutions with documented data stewardship policies said that they have such a council. Among those that did, however, most seemed to give it an influential role. Although the most common of the five roles we asked about was to provide advice, 22 institutions also said their council set policy and 21 said it set priorities, and councils at 18 institutions adjudicated conflicts (see Figure 5-6). The power of the purse was largely absent, however: only one institution said its council could authorize funding.

Data Use, Security, and Access Policies

Although only about a third of our respondent institutions reported data stewardship policies, we found more widespread incidence of other policies relating to data use, security, and access. As Figure 5-7 shows, nearly three-quarters of respondent institutions (73.4%) said they have a documented policy on the acceptable use of institutional data, and a solid majority (59.8%) have one defining individual employee responsibilities for data security practices. It is possible that some institutions see these as alternatives to a more formal data stewardship policy, though institutions reporting a data stewardship policy were also more likely to have these policies.

The policy groundwork on data issues related to storage and use outside the institution, as in outsourcing, was modest. Slightly fewer than half of respondents reported documented policies for sharing, storing, and transmitting institutional data (for example, with or to ISPs, external networks, or contractors’ systems), despite the common practice of trading data with outside entities. What’s more, given how easy it’s becoming for users to directly use consumer-oriented cloud services (often for free) without any IT assistance, it’s noteworthy that well under one in five institutions (17.6%) reported a documented policy for storage of institutional data on extra-institutional web-delivered (cloud) systems, and fewer still (10.9%) had one for use of extra-institutional web-delivered services to do institutional business. Perhaps some institutions feel that these issues are implicitly covered by their acceptable-use or sharing/storing/transmitting policies. We speculate, however, that most institutions simply have not yet caught up with the marketplace for cloud services.

Given the recent proliferation of laws relating to compromised personal information, we were surprised to find that only half

Figure 5-6. Role of the Data Stewards’ Council (N = 30)
of institutions reported a documented policy for notification of affected parties following data breaches. In addition, though encryption of institutional data on user devices is one way of mitigating the exposure caused by data breaches and may be grounds for exemption from notification requirements, it has not found a lot of traction among our respondents: Only about a quarter said they have documented policies for it.

The most commonly reported policies were sufficiently widespread that we found few distinguishing demographic patterns in their adoption. However, policies for data breach notification; sharing, storing, and transmitting data; and encryption were all more common among larger institutions (see Figure 5–8). The difference was especially notable for data breach notification policies, which were twice as prevalent (74.5%) among institutions with enrollment greater than 15,000 than among those with enrollment of 4,000 or fewer (36.9%).

One can infer from our results that many institutions feel the lack of documented policies relating to data use, security, and access and aspire to have them. In many cases the percentage of institutions reporting a policy under development would, if the policy were completed, dramatically increase the proportion of those with policies. The most striking examples were encryption of institutional data on user devices and the two cloud-related items, for which the percentages reporting a policy under development equaled or exceeded those of institutions that have a policy already.

Although these initiatives may well represent aspiration to a more formal data policy regime, policy development can be informal, low priority, unfunded, and fraught with political obstacles—and can fail. Simply adding the “under development” results to the “have policy” results in Figures 5–7 and 5–8 would probably overstate the likely future incidence of these policies. The percentages
of institutions reporting policies for employee information security responsibilities and for sharing, storing, and transmitting institutional data were actually lower in these results than those saying they had “implemented” similar policies in ECAR’s 2006 IT security study. Differences in the way the questions were posed in the two surveys prevent comparing them conclusively, but our findings certainly do not suggest a substantial increase in data policy formulation.

**Classifying Data by Risk or Sensitivity**

Regulatory requirements, access authorization rules, and better organization of data for analytics use are only some of the reasons why institutions might wish to categorize data, and in particular to systematically distinguish data whose inappropriate use creates institutional exposure. Although we recognize that many institutions might handle these risks informally or in a distributed fashion, we wanted to know how many have created a formal process for classifying institutional data by risk or sensitivity level.

Not many, as it turned out: About one in four institutions told us they have such a process. As Figure 5-9 shows, data classification processes were most common among larger institutions, though even among those with more than 15,000 FTE enrollments, fewer than half (45.8%) reported one. Among Carnegie classes, doctoral institutions were the most likely to report a classification process (49.2%) and baccalaureate institutions the least likely (10.8%), with master’s and associate’s institutions in between.

Data classification is often a component of data stewardship programs, so we were not surprised to find that institutions reporting a documented policy defining data steward responsibilities were almost three times as likely to report a formal data classification process.
process (45.0%) than were those without a data stewardship policy (16.3%).

Clinton Smullen, director of Academic and Research Computing Services at the University of Tennessee at Chattanooga, described how the data stewardship and classification systems work at UTC. “The data stewardship program began in tandem with the data classification program,” he said, “and when departments classified their data, they nominated a steward too.” The classification system is based on a state framework that attempts to identify key areas of exposure without getting bogged down in unproductive details. “The state of Tennessee devised a three-tier classification scheme,” Smullen explained. “There’s public data and proprietary data, which we don’t have to report about should it get exposed, and then there’s sensitive data—information about individuals, which requires reporting. The classifications are gross, but if you get any finer, you get into real trouble trying to classify or control it.” Departments take a first cut at classifications, then negotiate them with IT.

As these remarks suggest, data classification is a big and possibly never-ending process, and very few of our respondents (4.0%) were so bold as to tell us that they had classified all their data (see Figure 5-10). On the other hand, no institutions at all said they had classified none of their data, and more than three-quarters (77.3%) said some or most was classified.

Do classification processes for restricted/sensitive data, and for that matter progress in classifying data, make a difference? We found some evidence that they do. Institutions with data classification processes averaged stronger agreement than those without when assessing the statement that restricted or sensitive data at their institution was identified and appropriately classified (see Table 5-1). Looking at the data another way, 80.8% of institutions with data classification processes agreed or strongly agreed with the statement, whereas only 58.4% of those without such processes agreed or strongly agreed. Among institutions with classification processes, those reporting
more progress in classifying data also agreed more strongly with the same statement than did those reporting less progress.

Although these results lend some weight to the hypothesis that formal classification processes make a difference in the important matter of appropriately classifying sensitive data, they also give some hints about why relatively few institutions choose to create them. Nearly 6 in 10 institutions without such processes agreed or strongly agreed that they are accomplishing the task of classifying data appropriately, and most of the rest were neutral rather than disagreeing, for a mean response of 3.47 out of 5.00. What’s more, we found no significant agreement difference between those with and without formal classification processes regarding the key practical end of data classification—“restricted/sensitive data is secure

Table 5-1. Restricted/Sensitive Data Is Identified and Classified, by Data Classification Characteristics

<table>
<thead>
<tr>
<th>Data Classification Characteristic</th>
<th>Restricted/sensitive data is identified and appropriately classified.</th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does institution have a formal process for classifying institutional data by risk or sensitivity level?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3.47</td>
<td>226</td>
<td>0.925</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3.96</td>
<td>78</td>
<td>0.780</td>
<td></td>
</tr>
<tr>
<td>How much data classified?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A little/some</td>
<td>3.63</td>
<td>38</td>
<td>0.714</td>
<td></td>
</tr>
<tr>
<td>Most/all</td>
<td>4.27</td>
<td>37</td>
<td>0.732</td>
<td></td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
from unauthorized access”—or in their agreement that the institution gets maximum business value from data. Even if they think they could do a better job by formalizing the classification process, many institutions may feel that existing informal or decentralized approaches are good enough and that they can get a better return by investing scarce resources elsewhere.

**Data Breach Notifications**

In recent years, probably nothing has caused more lost CIO sleep than the possibility of an embarrassing data breach exposing sensitive information. Since the passage of California's landmark Senate Bill 1386 in 2003, more than 40 U.S. states have adopted laws requiring formal notification of individuals when private information about them has been compromised. Similar legislation has become commonplace internationally as well. Despite the highly distributed nature of institutional data and the near impossibility of establishing central IT control over it all, institutions tend to look to the IT organization when breaches occur. As we reported in Figure 5-2, 67.0% of institutions told us that central IT was primarily responsible for notification of affected parties following a data breach.

If actual occurrences of this particular source of CIO insomnia aren’t exactly widespread, they aren’t so rare as to be negligible, either. Asked how many data breaches requiring notification of affected parties had occurred at their institutions in the past 24 months, 30.3% reported one or more. Of these, about half (14.8% of all institutions) reported a single incident, and the remaining institutions reported between 2 and 10, though only a handful (1.8%) said they had had more than 4 (see Figure 5-11).

Data breach notification incidents were reported by institutions in all major Carnegie classes and FTE enrollment size ranges, but they were more common at master’s and doctoral institutions than at associate’s and baccalaureate institutions, and they were more often reported by larger than smaller institutions. As Figure 5-12 shows, exactly half of responding doctoral institutions reported...
one or more data breach notification incidents in the past 24 months, close to five times the rate among baccalaureates. Fewer than one in five institutions with FTE enrollments of 4,000 or below reported one or more incidents, whereas more than half of those with enrollments above 15,000 did so.

Aside from these demographic associations—which may simply reflect the fact that having more data, devices, and people creates more chances for exposure—we found few variables related significantly to data breach notification incidents. One intriguing exception was the existence of a documented policy for notifying affected parties following a data breach; those with such a policy were far more likely to report one or more incidents (46.6%) than those without a policy (8.8%) or those with a policy under development (20.0%). This doesn’t necessarily mean that institutions with policies experienced more data breaches per se, but it may be that institutions with documented policies feel more obliged to make notifications, suggesting that policies may take a more aggressive view of the need to notify than the law strictly requires. Or the existence of a documented policy may simply increase awareness of the need to notify.

Institutions reporting data stewardship policies were not significantly more or less likely than those without them to have experienced incidents, nor did we find that having an institutional data policy office or an IT governance committee dedicated to data made a difference. Institutions that had experienced incidents did not do better or worse in enterprise data quality score or in major data management outcomes, either. Although suggesting that such incidents are merely random would be going much too far, it does appear that they strike all kinds of institutions pursuing many kinds of data management regimes.

**Data Stewardship, Formal and Cultural**

We confess to being surprised that formal data stewardship policies didn’t draw a
bright line down the middle of the findings presented here. We did, of course, find some evidence that these policies, so prominent in data management best practices advice, make a difference in desirable outcomes. Institutions with documented data stewardship policies do a little better than others on enterprise data quality, and they are more likely to have a process for classifying data by sensitivity, and slightly more likely to agree that data is, in fact, appropriately classified. But data stewardship policies did not prove to be a magic bullet: The one-third of institutions reporting them didn’t do better (or worse) on major measures such as getting business value from data, agreeing that restricted data is secure, or being less likely to have experienced a data breach notification event.

Looking at our data as a whole, however, we noted other results suggesting that some of the spirit of data stewardship may be found in broader practices related to data responsibilities. Concerned that institutions might regard central IT as the default data owner, we found in fact that the great majority assign business and academic units the primary responsibility for such critical matters as data accuracy, access decisions, and regulatory compliance, leaving central IT mainly responsible for appropriate format and metadata management duties. Though we found a rather modest portfolio of data-related policies in place, 60% or more of institutions had policies defining employee data security responsibilities and acceptable use of data (and others reported such policies in development). Many institutions may simply regard data stewardship as part of professional practice and culture and see no extra benefit in formalizing it, regardless of the data management advisory literature.

It’s striking, in fact, that after a decade or more of intense concern over data security and privacy issues, our study found institutions more strongly in agreement that data is secure and appropriately classified for risk than they are about other key measures, such as getting business value from data and effectively managing all the varieties of data and digital content that the institution needs. (We examine such data management outcomes in more detail in Chapter 8.) Although security and privacy concerns remain essential to responsible IT administration, they have also been lavished with resources and attention in recent years. Declaring victory in this arena would be premature (and always will be), but it might be reasonable to consider that progress has been made, that awareness is broad, and that what was once a set of issues primarily understood only by IT professionals has become part of a larger set of professional practices.

Despite some lingering concerns, then, that relatively few institutions have chosen to follow the commonly recommended practice of establishing formal data stewardship policies, we note that this doesn’t necessarily distinguish successful from unsuccessful data management. Although formal data stewardship may be one way for institutions to inculcate security awareness and a culture of responsibility toward data, other initiatives—in data quality, analytics, training, and content management—deserve consideration as well when addressing data-related needs.

Endnotes

7. A single outlier institution reporting more than 30 data breach notification incidents was not included in the results reported here.
6
Institutional Content Management

Yea, from the table of my memory I’ll wipe away all trivial fond records.

—William Shakespeare

Key Findings

♦ The most common environment respondents reported for enterprise management of digital content was a mix of best-of-breed enterprise and local/departmental solutions (35.9% of institutions). Four in 10 institutions reported only local or ad hoc solutions (that is, no enterprise solutions at all), whereas slightly more than 1 in 10 said they have an integrated enterprise content management solution. Institutions generally anticipated having more enterprise-oriented content management environments three years in the future.

♦ Institutions with more enterprise-oriented content management environments tended to report greater agreement that they could effectively manage all the varieties of data and digital content that their institutions need.

♦ Public relations and marketing organizations were responsible for the look and feel of the institutional website at about two-thirds of institutions.

♦ Video streaming/download and content formatted for handheld devices were the top web environment enhancements institutions selected as most important to meeting constituent expectations in the next three years.

♦ Agreement that the institutional web environment routinely exceeds user expectations was higher among institutions that reported more extensive use of such best practices as a database-driven web environment, website owner ability to perform routine maintenance without IT intervention, consistent web page look and feel, and granular control over user access to website content.

♦ About half of institutions overall had some kind of entity responsible for electronic records management.

♦ Seven in 10 institutions reported having documented electronic records retention schedules for financial, student, and/or HR records; about half have such schedules for faculty and/or staff e-mail.

♦ Respondents gave lackluster ratings to their institutions’ ability to comply with records retention and disposition requirements throughout the institution, averaging 2.5 (between fair and good) on a 5-point scale. Average ratings were higher where institutions reported more documented electronic records retention schedules, where enterprise data quality scores were higher, and where an enterprise records management solution had been implemented.
Information technology units once dealt almost exclusively with structured data that flowed inside structured processes. Designing an information system meant mapping the way data elements changed as they entered processes and were transformed by them. Every data element had a discrete type, length, and name. As some of the findings of preceding chapters make clear, managing that kind of data has its challenges, but it also works according to some fairly clear and powerful rules.

Structure, however, no longer governs the world of data. All kinds of things have become data by the simple fact of being captured or expressed using a digital tool. Not just the budget and enrollment figures of the English department, but its announcements and course descriptions, photographs of its faculty members, social interactions between students, even texts taught, lectures delivered, and discussions led are all now potential elements of the institutional data universe. Much of this data is entirely unstructured except at the most basic formatting level, and some of it can be considered semistructured in the sense that it is meant to be inserted into structured templates or tagged according to some schema.

Making this huge and complex emerging digital content environment usable and manageable is one of the great challenges of modern IT administration. In this chapter, we examine a few facets of institutional content management, focusing on those aspects where we believed we would find the most pressing institutional challenges and the most mature (if emergent) practices. Beginning with an investigation of the extent to which institutions are taking an enterprise approach to content management, we turn next to practices and future plans for the institutional web environment, and then to management of electronic records.

**Institutional Content Management: An Overview**

Institutions have a profound interest in the many varieties of digital content they generate. The institutional website is the world’s window on the institution and provides many critical constituent services. Important information is locked up in e-mails, spreadsheets, and documents, some of which may meet legal or archival criteria that make them official records. At some level, the institution needs a way to discover, organize, and manage important digital assets and to realize the efficiencies that come with standardizing on common tools.

On the other hand, digital content presents classic enterprise dilemmas. Not everything needs to be (or can be) managed centrally, nor do different types of content necessarily lend themselves to a single management platform. Integrated enterprise-grade solutions may be functionally inferior to separate best-of-breed products that address specific content types. Users with special requirements will demand special tools to meet them, and local units have the budgets and (sometimes) the expertise to go their own way. Considering the slow and uneven progress institutions have made toward achieving enterprise control over even the most heavily structured administrative data, the challenges of managing the Wild West of digital content stand out in high relief. A group of IT administrators who implemented a content management system for learning content at Athabasca University sum up the potential, and the challenges, from a higher education perspective: “A content management system can and should manage the institution’s full gamut of formalized content...[but] until processes and applications mature further, full implementation of a CMS within an institution will be constrained, largely due to the difficulty of creating simplified algorithms to describe and deal with dissimilar content sets.”1
Nevertheless, the concept of enterprise content management (ECM)—that is, integrated systems and processes for dealing with a wide variety of digital content relevant to business operations—has been promoted for a decade or more, and Gartner Inc. forecasts that the integrated ECM suites that embody the concept are within a few years of reaching maturity (or, in the company’s language, the “plateau of productivity”). We wanted to know if institutions are embracing this concept, a more modest or partial enterprise-oriented approach, or some version of distributed or ad hoc approaches. We also looked in more detail at specific kinds of ECM solutions and at the way institutions are approaching the management of learning content.

**Content Management Environment**

For a big-picture look at institutional content management environments, we asked respondents to choose which of five different environments best characterized their institution. The options, shown in Table 6-1 (along with short descriptors we’ll use to refer to them), ranged from integrated enterprise-level solutions, through mixed enterprise and local, to purely local or ad hoc environments. We also asked respondents which environment they anticipated having in three years.

As Figure 6-1 shows, we found a wide distribution of current environments emphasizing content-specific tools, and not much adoption of integrated ECM solutions. The most common environment (35.9% of institutions) was a mix of best-of-breed enterprise and local or departmental solutions varying by content type. Four in 10 institutions reported only local or ad hoc solutions—that is, no enterprise solutions at all. Only 11.5% said they have an integrated ECM solution.

Three years out, however, things look quite different. Although about a third of institutions still anticipated a mixed enterprise/local best-of-breed environment—slightly fewer than now—the percentage of those expecting to have an integrated ECM environment was almost triple the current number. Correspondingly, the percentage of those anticipating an environment of local solutions by content type or an ad hoc environment dropped dramatically from the current levels.

In short, institutions seem to anticipate moving “up” the enterprise content management ladder at least to some degree. Table 6-2 shows how respondents expect current environments to change in three years. (Note that we combined the local solutions by content type and ad hoc categories into a single category to simplify this table.)

<p>| <strong>Table 6-1. Institutional Content Management Environment Descriptions and Short Names</strong> |</p>
<table>
<thead>
<tr>
<th><strong>Short Description</strong></th>
<th><strong>Environment Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated enterprise content management (ECM)</td>
<td>An integrated enterprise content management solution that manages the widest possible range of content</td>
</tr>
<tr>
<td>Enterprise best-of-breed predominant</td>
<td>Best-of-breed enterprise solutions for most or all major content types</td>
</tr>
<tr>
<td>Mixed enterprise/local best-of-breed</td>
<td>A mix of best-of-breed enterprise and local/departmental solutions varying by content type</td>
</tr>
<tr>
<td>Local solutions by content type</td>
<td>No enterprise solutions, only local/departmental solutions for particular content types</td>
</tr>
<tr>
<td>Ad hoc</td>
<td>Digital content management is generally ad hoc</td>
</tr>
</tbody>
</table>
an integrated ECM environment, only 6.1% expected changing to a different environment. By contrast, 80.4% of those in local/ad hoc environments expected to be in a different, more enterprise-oriented category. The largest group of these respondents (42.0%) expected to move to a mixed enterprise/local best-of-breed environment, but almost a quarter (23.2%) anticipated having an integrated ECM solution. Those respondents who currently have a mixed enterprise/local best-of-breed environment also often seem to aspire to something more centralized; 61.0% of them anticipated moving to either integrated ECM or an enterprise best-of-breed–predominant environment.

Assessing the likelihood of these plans being realized is difficult, though as we have noted elsewhere in this study, planned enhancements don’t always materialize. These
figures do, however, seem to be a vote of confidence in the value of more enterprise-oriented content management and (to a lesser degree) in the near-term maturity of integrated ECM solutions. At the University of North Texas, Maurice Leatherbury, acting vice president for information technology and CIO, is looking forward to the planned implementation of an ECM system. “Right now, we only have loosely enforced policies about who can manage and edit content,” Leatherbury said. “The ECM implementation will provide better templating, rollback, better look-and-feel control and change control, and it will simplify the editing process and make it easier to maintain web pages.”

Do content management environments vary significantly by institution type? Interestingly, we found that they did for current environments but not for those three years out. As Figure 6-2 shows, among the major Carnegie classes, baccalaureate institutions were the most likely to report integrated ECM solutions. This suggests that the best fit for current ECM products is with smaller, organizationally flatter institutions, though baccalaureates were actually slightly more likely to report mixed enterprise and local best-of-breed solutions. Doctoral institutions had the highest rate of mixed enterprise and local best-of-breed solutions, which probably reflects both the characteristically distributed nature of research institutions and a wider range of content management needs. Perhaps the main impression these results leave, however, is variety: Every one of the environment types was reported by at least some institutions of every Carnegie class.

Effective Management of Varieties of Data and Digital Content

One of the key promises of ECM is that it brings coherence and manageability to
content creation and distribution, allowing the institution to manage content consistently across different content types. But do institutions pursuing enterprise approaches to content management actually do better at achieving this goal?

Judging by one rough measure, we found at least some indication that they do (see Table 6-3). Asked to express their agreement with the statement that their institutions effectively manage all the varieties of data and digital content that the institutions need, respondents on average weren’t inclined to agree: Their mean response was 2.81, short of even a neutral (3) response. But those reporting local or ad hoc solutions only were even more critical, averaging a 2.56 response. By contrast, those with enterprise best-of-breed–predominant or mixed enterprise/local environments averaged close to a neutral answer, and those reporting integrated ECM had a slightly higher than neutral response. It gives a sense of the magnitude of the difference to note that among the local/ad hoc group, only 12.4% agreed or strongly agreed with the statement, whereas among the integrated ECM group, 45.7% did. The uninspiring overall response to this statement strongly suggests that institutions are struggling to manage all the varieties of data and digital content they need. It’s hard to say that even those with integrated ECM systems are achieving the goal of effective management when their average answer barely tips past neutral—and this group represents fewer than one in eight of our survey respondents.

Some limits to these results should be kept in mind. Since our statement referred to all varieties of data and digital content, it didn’t consider whether the benefits of managing particular types especially well might be more important to some institutions than managing the whole range effectively. Likewise, it may be that institutions willingly maximize other factors, such as local autonomy or preserving freedom to innovate, at some cost in effective management of the whole spectrum of content.

Still, combined with the patterns we found in anticipated shifts from local to enterprise environments over the next three years, and in the context of an explosion of new data and content types, these assessments about effectively managing variety seem to suggest that an enterprise orientation helps with this challenging task.

**ECM Components**

In addition to the overall institutional content management environment, we wanted to know about the status of key major components of environments that contain enterprise solutions. Respondents

<table>
<thead>
<tr>
<th>Current Environment</th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated enterprise content management</td>
<td>3.17</td>
<td>35</td>
<td>1.071</td>
</tr>
<tr>
<td>Enterprise best-of-breed predominant</td>
<td>2.95</td>
<td>37</td>
<td>0.780</td>
</tr>
<tr>
<td>Mixed enterprise/local best-of-breed</td>
<td>2.93</td>
<td>108</td>
<td>0.964</td>
</tr>
<tr>
<td>Local/ad hoc</td>
<td>2.56</td>
<td>121</td>
<td>0.795</td>
</tr>
<tr>
<td>Total</td>
<td>2.81</td>
<td>301</td>
<td>0.913</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
who reported an environment containing at least some enterprise aspects (that is, those reporting integrated ECM, enterprise best-of-breed–predominant, or mixed enterprise and local best-of-breed environments; see Figure 6-1) were asked a follow-up question about the status of each of three major ECM solutions:

- web content management (solutions permitting enterprise management of web content creation and publishing),
- records management (solutions managing/archiving official records for regulatory, financial, or historical needs), and
- document management (solutions for imaging, search, and other document-related functions focused on operational needs).

Respondents were asked to say whether each solution was implemented, being implemented, planned but not started, or not planned. Note that our question asked only about enterprise solutions of these types, whether best-of-breed or part of an integrated suite. Purely local solutions for these purposes did not fall within the scope of the question.

As Figure 6-3 shows, only web content management solutions were implemented at as many as half of institutions with enterprise-oriented content management environments. Records management solutions were implemented at 37.6% and document management solutions at 31.8% of the institutions with enterprise-oriented environments. But existing implementations don’t tell the whole story. When institutions currently implementing and planning to implement these solutions are included, our respondents seem overwhelmingly to feel that each of these items belongs in its enterprise IT environment. In each case, fewer than 1 in 10 reported not planning the solution.

Our qualitative interviewees gave us some insights into both the dynamics currently restraining content management approaches and those that might stimulate interest in them. For Linda Deneen, director of Information Technology Systems and Services (ITSS) at the University of Minnesota Duluth, web content management solutions don’t make the cost-benefit cut. “We investigated them,” she said, “but they seemed much too complex for our level of use, as well as too expensive.” UMD’s ITSS unit delivers Dreamweaver templates for use within the

---

**Figure 6-3. Status of Enterprise Content Management Solutions (Institutions with Enterprise-Oriented Content Management Environments)**

<table>
<thead>
<tr>
<th></th>
<th>Implemented</th>
<th>Implementation in progress</th>
<th>Planned, implementation not started</th>
<th>Not planning to do</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web content management (N = 180)</td>
<td>50.0%</td>
<td>30.6%</td>
<td>10.6%</td>
<td>8.9%</td>
</tr>
<tr>
<td>Records management (N = 170)</td>
<td>37.6%</td>
<td>24.7%</td>
<td>31.8%</td>
<td>5.9%</td>
</tr>
<tr>
<td>Document management (N = 176)</td>
<td>31.8%</td>
<td>30.7%</td>
<td>29.0%</td>
<td>8.5%</td>
</tr>
</tbody>
</table>
institutional web environment, “and that guides users to a common look and feel, depending on how they’re used.”

At Bates College, by contrast, Gene Wiemers, vice president for Information and Library Services, describes the rationale that led Bates toward a “very tightly managed” approach to web content management. “We recognized early that the web is an official publication of the college,” he said. “We felt that if the faculty member wants to be creative with his or her course description, the way to do it is through an official approval process.” So Bates employs a content management system combined with an official approval protocol and the ability to draw some content, such as course descriptions, directly from enterprise administrative systems. Wiemers noted, however, that the institution’s web content management system needs much more flexibility, especially the ability to incorporate internally and externally hosted resources within the same framework. “That will enhance our ability to use tools like Flickr and YouTube within our institutional look and feel,” Wiemers said, “something we can’t even dream of now.”

Our interviewees tended to see web content as the area with the most mature content management tools, and a number of them particularly wished to see better tools to address pressing issues with document management. Terri-Lynn Thayer, assistant vice president and deputy CIO for Computing and Information Services at Brown University, said that “document management systems have to support the various document flow processes, and I don’t think the technology is there yet.” At the same time, Thayer added, business units must change business processes to facilitate, and exploit the advantages of, enterprise document management. “We’d love to find a common infrastructure that can be leveraged across units,” she noted. At the University of St. Thomas, Samuel Levy, vice president for Information Resources and Technologies and CIO, identified the central problem as making the information in documents discoverable and usable by applications. “Key information is in e-mails and their attachments,” Levy said. “What’s the taxonomy for that? You have these various repositories sitting in various databases and you need to find some way to abstract from the application to the repository in a way that’s meaningful to the organization. That’s the biggest single challenge.”

Since only respondents reporting some sort of enterprise-oriented content management environment were asked about the solutions listed in Figure 6-3, the incidence of each in the overall respondent population may be considerably smaller than shown there. On the other hand, the great majority of those institutions reporting a purely local or ad hoc content management environment today expect to move to an enterprise orientation within three years. If they realize those expectations and are as inclined to adopt ECM solutions as those who already have enterprise-oriented environments, the demand for these solutions could be even larger than implied in Figure 6-3.

**Management of Learning Content**

In Chapter 3, we reported that respondents on average assessed learning management system (LMS) data and content as one of the fastest-growing categories of data we asked about (see Table 3-1). Thanks to the richness of learning content and its increasing interactivity and presence within Web 2.0–influenced contexts, it is also one of the most varied and complex forms of institutional content. Efforts to make learning content more discoverable, reusable, standardized, and pedagogically targeted add still further elements to the mix, as do intellectual property considerations and institutional workflows. As a result, LMSs that once stored content in fairly simple application databases have developed more sophisticated data management schemes and are also often
integrated with external content stores with advanced content management capabilities.

Our survey asked respondents about their institutions’ use of three different kinds of content stores used to deliver digital learning content via institutional LMSs:

- LMS database(s),
- dedicated learning content management systems (LCMSs), and
- other content management systems.

As Figure 6-4 shows, LMS databases were the most frequently reported content store, but dedicated LCMSs were surprisingly common as well. While 8 in 10 respondents reported using the LMS database, about 7 in 10 reported a dedicated LCMS. Though it seems clear that some institutions are using the LCMS in lieu of the LMS database, for the most part the dedicated LCMS is a supplement rather than an alternative. Of 200 institutions that reported using a dedicated LCMS, 163 (81.5%) also reported using an LMS database.

Evaluating the use of other kinds of content management systems definitively is more difficult, since this item attracted a large 14.0% “don’t know” response. It appears, however, that these content management systems are mainly supplemental to dedicated LCMSs. Of the 89 institutions reporting use of other content management systems, 68 (76.4%) also reported using a dedicated LCMS. Other content management systems were more commonly reported at doctoral institutions and at larger institutions, especially those with FTE enrollments of more than 15,000. Overall, it looks like these systems tend to be used in complex environments where there might be a wide variety of content types, LMSs, or both.

Among three kinds of content stores asked about, the largest proportion of respondents reported having two (44.6%), and a substantial additional group reported three (19.9%) (see Figure 6-5). Overall, only slightly more than a third of institutions used just one of the content stores we asked about. Most of these used the LMS database.

The Institutional Website

It’s hard to believe that colleges and universities somehow functioned without websites until about 15 years ago. Today, an institution without a website is as unthinkable as an institution without a campus—more unthinkable, in fact, in part because websites have made it so much easier to create an institution without a physical campus. Websites present the institution to the external world and facilitate a huge amount of internal communication;

![Figure 6-4. Content Stores Used to Manage Digital Learning Content Delivered via Institutional Learning Management Systems]
they are a sort of campus-branded universal platform.

And that’s what makes them so hard to manage. Probably no other technology has taught so many hard lessons about content management over the last decade or so. The web in its early years was a poorly understood yet immensely attractive environment that promiscuously mixed technically and creatively inclined users; everybody was an amateur. Campus web environments were a collage of inconsistently formatted, eccentrically edited, technically quirky sites, loosely overseen by nervous central IT units that lacked clear guidelines regarding authority or best practice. Although publication workflows that had some relevance to the problem were well known for years before the web was invented, they were largely found in special professional environments far from the IT culture, and it took a while to cross the barrier.

Since then, institutions have benefitted enormously from progress in the technical, policy, and procedural aspects of website management. At the same time, the web environment has become ever more complex—absorbing more types of content, interconnecting with more underlying systems, and facilitating more social practices. Thanks to a looming mobility revolution, the one semi-stable aspect of traditional web management—the PC as delivery platform—is fracturing, and the cloud computing paradigm is effacing the distinction between local applications and the web.

Given these challenges, we wanted to know where institutions situate responsibility for website design issues, to what extent they provide content management capabilities, what plans they have to enhance or add to their web environments, and whether they think they are exceeding the expectations of key website constituents.

**Responsibility for Website Look and Feel**

Assignment of responsibility tells us something about whether a technically enabled process is seen as a business or a technical matter. One of the pleasant surprises of this
study has been discovering that a reasonable division of responsibilities governs many data-management-related concerns. In Chapter 5, for example, we reported that key business-related issues such as data accuracy and regulatory compliance were most often the primary responsibility of business and academic units rather than central IT (see Figures 5-1 and 5-2).

When we asked a similar question regarding primary responsibility for determining the look and feel of the institution’s primary website (i.e., domain.edu homepage and subordinate pages), we found an even more pronounced assumption of business unit responsibility (see Figure 6-6). At about two-thirds of respondent institutions, this responsibility fell primarily to the public relations or marketing office, strongly suggesting that at this level the main website is conceived of primarily in terms of its business function of representing the institution. Central IT had the primary responsibility at only 5.6% of institutions.

This connection to the public relations or marketing function may actually be stronger than even these figures indicate, yet at the same time central IT’s role may not be quite as marginal as its infrequent assumption of primary responsibility suggests. At 23.4% of institutions, primary responsibility fell to a web policy/planning committee, and we asked these respondents which entities (from a list of 11 possibilities) were represented on the committee (see Figure 6-7). Out of 70 institutions replying to this question, 67 (95.7%) said that central IT was represented, and public relations (91.4%) and marketing (87.1%) rounded out the top of the list. They seem to have plenty of company as well: The median number of represented entities was seven, and 22.1% of institutions reported nine or more.

Thus, at the great majority of respondent institutions, public relations and marketing—campus offices typically charged with outreach, brand management, and publishing responsibilities—either have direct primary responsibility for website look and feel or are represented on the committee that has it. We speculate as well that even where some entity other than a policy/planning committee has primary responsibility, mechanisms exist to involve central IT and get input from a variety of campus organizations, whether through committees that are subordinate to the department in charge or through informal or ad hoc processes.

Certainly our qualitative interviewees described an active central IT role in web policy, planning, and administration, as well as confirming our survey results emphasizing public relations and marketing units as the entities with primary responsibility for overall look and feel. At Georgia State University,
for example, J. L. Albert, associate provost and CIO, explained that University Relations approves website templates, but IT programs the templates on the basis of frameworks in the university’s content management system; in doing so, Albert noted, “IT is controlling the ergonomics of navigation.” Steve Gallagher, CIO at the University of San Francisco, described an environment in which IT is both technical enabler and enterprise coordinator. “Academic enrollment and services manages content in admissions-related areas, and the advancement area, which includes public affairs and communications, handles other institutional content,” Gallagher said. Schools and departments must also keep their own content up to date, whereas “IT’s role is to provide the infrastructure and the information channeling capability.” The planned implementation of a new content management system, however, is leading USF to rethink its basically federal structure and consider greater central oversight. “The challenge,” Gallagher noted, “is the natural tension between providing a central framework that will ensure consistency while empowering the local levels.”

The predominance of outward-facing departments in determining website look and feel is no doubt partly explained by their professional competencies. But it also has something to say about the strategic nature of the website. Because it’s a virtual embodiment of the institution, it must convey a positive and consistent image; because it communicates so much information, it must be suited to frequent and diverse updates, and speak with authority. These requirements in turn imply the need for a web environment with robust content management characteristics.

**Web Environment Characteristics**

Like previous IT realms, web management has been forced by its own success to develop the markers of mature information practice: separating data and logic components that were once intertwined,
leverage technologies that take routine work out of the hands of programmers and put it into the functional staff hands where it belongs, and formalizing access rules and workflows even as participation in the technology broadens.

To learn more about how widespread practices like this are, we presented respondents with five statements relating to web content management and asked them the extent to which each characterized their institution, using a scale from 1 (in no cases) to 5 (in all cases). The statements and mean responses appear in Table 6-4.

Most of the characteristics we asked about were fairly prevalent. Consistent look and feel, ability of website owners to perform routine maintenance without IT intervention, and granular control over user access to website content all had mean responses above the midway point between “in some cases” and “in most cases,” and for each item, more than 60% of respondents said the characteristic applied in most or all cases. At the bottom of the list was one of the more advanced aspects of web content management, having adequate workflows in place for the editing and approval of website content. Even here, however, 41.4% said that such workflows were in place in most or all cases.

Broadly speaking, then, our respondents painted a relatively positive picture of the extent of these good (if basic) web content management practices. In addition, we found a strong positive association between having an enterprise web content management solution and higher mean results for all of the web environment characteristics we asked about. As Table 6-5 shows, institutions that reported an implemented enterprise web content management system reported on average a considerably higher prevalence of each web characteristic than those that did not plan, or had not started, such an implementation. Those in the process of implementing were in between. (Only institutions with an enterprise-oriented environment for digital content management—one of the three left-most categories shown in Figure 6-1—were included in this analysis.)

Although we can’t prove a causal link, and other factors surely are involved, it seems reasonable to interpret this relationship as evidence that enterprise web content management solutions do help achieve the ends they’re designed for. As Wayne Powel and Chris Gill of Gonzaga University argue in a 2003 article in the EDUCAUSE Quarterly, “Web content management systems present the best of both worlds: They give the people who know the

Table 6-4. Web Environment Characteristics

<table>
<thead>
<tr>
<th>At my institution...</th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web pages have a consistent look and feel.</td>
<td>3.80</td>
<td>309</td>
<td>0.816</td>
</tr>
<tr>
<td>Website owners can perform routine site maintenance without IT intervention.</td>
<td>3.73</td>
<td>308</td>
<td>0.976</td>
</tr>
<tr>
<td>We have granular control over user access to website content (e.g., create, edit, publish, read access).</td>
<td>3.66</td>
<td>305</td>
<td>1.185</td>
</tr>
<tr>
<td>The web environment is database-driven (i.e., content is stored separately from templates).</td>
<td>3.33</td>
<td>300</td>
<td>1.094</td>
</tr>
<tr>
<td>Adequate workflows are in place for the editing and approval of website content.</td>
<td>3.15</td>
<td>299</td>
<td>1.146</td>
</tr>
</tbody>
</table>

*Scale: 1 = in no cases, 2 = in a few cases, 3 = in some cases, 4 = in most cases, 5 = in all cases
content control over it, but retain the university’s ability to define a look and feel consistent throughout all or parts of an entire site.™

Future Enhancements to Institutional Web Environment

Websites evolved rapidly from the simple display of static information to more dynamic responses to user inputs, integration with back-office systems, incorporation of personalizable portals, and the addition of Web 2.0 tools. Recently, institutional websites have also come up against new competition, including extensible, platform-based social networking sites like Facebook that users can orient their online lives around and cloud-based office applications and collaboration tools users can access directly without the mediation of an institutional bureaucracy.

To keep up, institutional websites have to evolve constantly, incorporating new kinds of data and content that add their weight to an already overburdened institutional data management scene. We were interested to know how respondents expected demands on the institutional website to change in the near future, so we asked them to identify the three additions or enhancements to their institutional web environment (from a list of 10 items) they thought would be most important to meeting constituent expectations in the next three years (see Figure 6-8).

Table 6-5. Web Environment Characteristics, by Enterprise Web Content Management Solution Status*

<table>
<thead>
<tr>
<th>At my institution...</th>
<th>Enterprise Web Content Management Solution Status</th>
<th>Not Planning/Planned, Not Started</th>
<th>Implementation in Progress</th>
<th>Implemented</th>
</tr>
</thead>
<tbody>
<tr>
<td>The web environment is database-driven (i.e., content is stored separately from templates).</td>
<td>Mean**</td>
<td>2.59</td>
<td>3.46</td>
<td>3.70</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>32</td>
<td>54</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.911</td>
<td>0.966</td>
<td>0.982</td>
</tr>
<tr>
<td>Website owners can perform routine site maintenance without IT intervention.</td>
<td>Mean**</td>
<td>3.43</td>
<td>3.71</td>
<td>4.07</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>35</td>
<td>55</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>1.037</td>
<td>0.975</td>
<td>0.790</td>
</tr>
<tr>
<td>Adequate workflows are in place for the editing and approval of website content.</td>
<td>Mean**</td>
<td>2.35</td>
<td>3.23</td>
<td>3.61</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>34</td>
<td>53</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>1.203</td>
<td>1.050</td>
<td>0.961</td>
</tr>
<tr>
<td>Web pages have a consistent look and feel.</td>
<td>Mean**</td>
<td>3.57</td>
<td>3.64</td>
<td>4.08</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>35</td>
<td>55</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.739</td>
<td>0.847</td>
<td>0.674</td>
</tr>
<tr>
<td>We have granular control over user access to website content (e.g., create, edit, publish, read access).</td>
<td>Mean**</td>
<td>3.03</td>
<td>3.58</td>
<td>4.14</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>34</td>
<td>55</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>1.446</td>
<td>1.049</td>
<td>0.868</td>
</tr>
</tbody>
</table>

*Data excludes institutions reporting only local/departmental or ad hoc solutions for enterprise management of digital content.

**Scale: 1 = in no cases, 2 = in a few cases, 3 = in some cases, 4 = in most cases, 5 = in all cases
Two items stood out from the rest: video streaming and download, named among their top three by 69.9% of respondents, and web content formatted for handheld devices, selected by 55.3%. The video streaming result echoed a theme that has appeared several times in this study: the challenge of meeting growing demand for video resources. In Chapter 3, we noted that online video and audio content was the data type that institutions least agreed they had the infrastructure to manage (see Figure 3-1). At the same time, respondents were relatively positive about their ability to manage this kind of data in three years, and our results regarding enhancements to the web environment are consistent with those expectations.

One probable stimulus of online video demand is lecture capture systems, which permit institutions to create a library of streamable course lectures; likewise, media, arts, and education programs all increasingly involve video content creation. Institutions may also be responding to the growing prevalence of video in web content, stimulated by the popularity of YouTube, “vlogs” (video blogs), and the increasing amount of digitally delivered video entertainment. ECAR’s 2009 study of students and IT found that 84% of undergraduate respondents download music or video and 45% contribute content to video websites.5

Handheld devices may be on the cusp of an even bigger transformation of the web. The success of Apple’s iPhone redefined the smartphone as an applications-consuming information appliance with a “real” browser. A new generation of competing devices is now following Apple’s lead. Though device and data plan costs remain high, they are likely to drop, and even at their current levels they have already permitted the lively growth of mobile computing at some campuses. Stanford University, MIT, and the University of California at San Diego have been conspicuous among institutions developing campus mobile computing strategies,6 and Blackboard’s 2009 acquisition of Terriblyclever Design, an iPhone

---

Figure 6-8.
Additions/Enhancements to Web Environment Most Important to Meet Constituent Expectations in Next Three Years (N = 309)
Institutional Data Management

applications start-up founded by Stanford students, signals the incorporation of mobility into mainstream enterprise environments. ECAR’s 2009 student study found a large, though diverse, student user base for Internet-enabled handheld devices: Half of students said they owned such a device, and though a substantial 35% of these never accessed the Internet from them, 29% did so daily and another 21% did so weekly or several times per week. These figures suggest that institutions are wise to put web content formatting for handheld devices on their near-term web development agendas.

Qualitative interviews made it clear that many IT administrators are keenly following these developments. “We plan to aggressively push out new mobile services, to students in particular,” said USF’s Gallagher. “The overarching goal is to create more touch points to the institution.” At the University of Tennessee at Chattanooga, Clinton Smullen, director of Academic and Research Computing Services, reported that, besides improved content management, the most important enhancement the university will add to its web environment is support for mobile devices. “Blackboard has developed an iPhone app to connect to the course management system, and we think that’s a good thing to support,” said Smullen. “We’re also looking at ways to reformat our ERP content appropriately when the system detects a mobile browser.”

Though Web 2.0 capabilities weren’t especially conspicuous in our survey results regarding future web plans, interviewees often mentioned plans related to social networking and interactivity. Neil McElroy, dean of Libraries and Information Technology Services at Lafayette College, summed up the college’s plans this way: “We want to see the website ingest content from the community, not just from the central communications office. That’s the essential characteristic of Web 2.0.” But, McElroy adds in a comment echoed by many of our interviewees as they discussed their website plans, “You can’t do that without a good content management system.”

The Institutional Website and Constituent Expectations

Given this picture of what institutions think they need to do to meet constituent expectations, it’s only natural to wonder how well they think they’re meeting those expectations now. In our survey, we set a high bar of performance, asking respondents their level of agreement that the primary institutional website at their institutions routinely exceeds the expectations of each of four constituent groups: faculty, staff, students, and public users.

The results were a little ambiguous but leaned against the interpretation that institutions think they routinely exceed expectations. As Figure 6-9 shows, “neutral” was by far the most common response for each constituent group, and for all groups except public users, more respondents disagreed than agreed. Notwithstanding this slightly stronger agreement overall that public user expectations are routinely exceeded, the most striking thing about the responses was their similarity. The mean responses for all groups were close to an average neutral response. Among the on-campus constituent groups, means hovered slightly below or around an average neutral response for staff (2.99), faculty (2.80), and students (2.89), whereas mean agreement about exceeding public users’ expectations was 3.23.8

If these top-level results feel a bit blasé and homogeneous, though, we found more differentiation when we compared mean agreement about exceeding expectations with the extent to which the web environment characteristics we discussed above (Table 6-4) prevailed at the institution. As Table 6-6 shows, mean agreement about exceeding expectations rose for every constituent group where the extent of each of four web environment characteristics was greater:
Table 6-6. Primary Institutional Website Routinely Exceeds Expectations of Constituents, by Web Environment Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Faculty Mean*</th>
<th>Faculty Std. Deviation</th>
<th>Staff Mean*</th>
<th>Staff Std. Deviation</th>
<th>Students Mean*</th>
<th>Students Std. Deviation</th>
<th>Public Users Mean*</th>
<th>Public Users Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The web environment is database-driven (i.e., content is stored separately from templates).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In no or few cases (N = 65)</td>
<td>2.49</td>
<td>0.859</td>
<td>2.67</td>
<td>0.894</td>
<td>2.60</td>
<td>0.915</td>
<td>3.05</td>
<td>0.991</td>
</tr>
<tr>
<td>In some cases (N = 77)</td>
<td>2.82</td>
<td>0.914</td>
<td>2.95</td>
<td>0.986</td>
<td>2.86</td>
<td>0.990</td>
<td>3.05</td>
<td>0.902</td>
</tr>
<tr>
<td>In most or all cases (N = 139)</td>
<td>2.94</td>
<td>0.841</td>
<td>3.16</td>
<td>0.850</td>
<td>3.05</td>
<td>0.837</td>
<td>3.42</td>
<td>0.800</td>
</tr>
<tr>
<td>Adequate workflows are in place for editing/approval of website content.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In no or few cases (N = 79)</td>
<td>2.37</td>
<td>0.894</td>
<td>2.56</td>
<td>0.953</td>
<td>2.56</td>
<td>0.975</td>
<td>2.96</td>
<td>0.947</td>
</tr>
<tr>
<td>In some cases (N = 88)</td>
<td>2.86</td>
<td>0.847</td>
<td>3.00</td>
<td>0.892</td>
<td>2.90</td>
<td>0.905</td>
<td>3.25</td>
<td>0.925</td>
</tr>
<tr>
<td>In most or all cases (N = 116)</td>
<td>3.07</td>
<td>0.789</td>
<td>3.28</td>
<td>0.799</td>
<td>3.14</td>
<td>0.801</td>
<td>3.42</td>
<td>0.776</td>
</tr>
<tr>
<td>Web pages have a consistent look and feel.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In no or few cases (N = 23)</td>
<td>2.13</td>
<td>0.920</td>
<td>2.29</td>
<td>1.160</td>
<td>2.17</td>
<td>1.129</td>
<td>2.48</td>
<td>0.947</td>
</tr>
<tr>
<td>In some cases (N = 56)</td>
<td>2.61</td>
<td>0.779</td>
<td>2.76</td>
<td>0.844</td>
<td>2.72</td>
<td>0.840</td>
<td>3.09</td>
<td>0.950</td>
</tr>
<tr>
<td>In most or all cases (N = 211)</td>
<td>2.92</td>
<td>0.860</td>
<td>3.13</td>
<td>0.855</td>
<td>3.02</td>
<td>0.859</td>
<td>3.35</td>
<td>0.821</td>
</tr>
<tr>
<td>We have granular control over user access to website content (e.g., create, edit, publish, read access).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In no or few cases (N = 54)</td>
<td>2.44</td>
<td>0.883</td>
<td>2.53</td>
<td>0.920</td>
<td>2.52</td>
<td>0.991</td>
<td>2.91</td>
<td>0.986</td>
</tr>
<tr>
<td>In some cases (N = 52)</td>
<td>2.69</td>
<td>0.843</td>
<td>2.89</td>
<td>0.904</td>
<td>2.81</td>
<td>0.930</td>
<td>3.15</td>
<td>0.988</td>
</tr>
<tr>
<td>In most or all cases (N = 180)</td>
<td>2.95</td>
<td>0.861</td>
<td>3.15</td>
<td>0.876</td>
<td>3.04</td>
<td>0.855</td>
<td>3.36</td>
<td>0.802</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
the web environment is database-driven,
adequate workflows are in place for
the editing and approval of website
content,
web pages have a consistent look and feel,10 and
the institution has granular control over user access to content (e.g., to create, edit, publish).

Interestingly, despite the sometimes dramatic differences between institutions where these characteristics were less extensive and those where they were more extensive, we did not find any significant difference based on the status of enterprise web content management solutions. Neither did we find significant differences related to demographic factors such as Carnegie class, FTE enrollment size, or public/private control.

However, one other factor was strongly associated with higher agreement about exceeding constituent expectations about the institutional website: data quality. In Chapter 4, we introduced an enterprise data quality score based on the average response to eight different data-quality-related survey items (see Table 4-2 and related text). As Table 6-7 shows, for all constituent groups, mean agreement about exceeding expectations was roughly 0.7 higher (on a scale of 1 to 5) among institutions with high enterprise data quality scores than among those with low scores.11

Taken together, these findings suggest that investments in web content management best practices and in data quality overall do have an impact on how the institution’s web environment is perceived. The web environment characteristics reported here are intended to make websites easier to navigate and maintain, to keep up to date, and to manage for appropriate access, and it’s not hard to see how these characteristics might directly affect user experience. The connection with enterprise data quality is less obvious, especially since our data quality questions referred to administrative systems rather than to the web environment. But the institutional website is often fueled by underlying enterprise data, particularly through enterprise portals and self-service sites, and sound data may make those services not only more trustworthy but also easier to implement in the first place. Likewise, as we reported in Chapter 4, better data quality is associated with the granularity of user data access authorization, perhaps by making identity information more robust, and this may help make websites more powerful or convenient. Finally, given the frequency with which a data quality connection has arisen in

| Enterprise Data Quality Score | Faculty | | Staff | | Students | | Public Users |
|------------------------------|---------|---|---|---|---|---|
| | Mean* | Std. Deviation | Mean* | Std. Deviation | Mean* | Std. Deviation | Mean* | Std. Deviation |
| Low (less than 2.5), N = 64 | 2.39 | 0.920 | 2.63 | 0.960 | 2.40 | 0.871 | 2.83 | 0.985 |
| Medium (2.5–3.5), N = 144 | 2.83 | 0.817 | 2.99 | 0.868 | 2.99 | 0.901 | 3.24 | 0.819 |
| High (greater than 3.5), N = 77 | 3.09 | 0.835 | 3.29 | 0.860 | 3.14 | 0.823 | 3.52 | 0.821 |

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
In this study, it may be that good enterprise data quality is a marker for a general atmosphere of good data management, with pervasive effects that improve user satisfaction.

**Electronic Records Management**

Within the large and expanding universe of information used by an organization, some elements have qualities that give them special status as records. Exactly what constitutes a record is a matter of some debate, but ISO standard 15489:2001, which addresses the practice of records management, defines records as “information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business.” These attributes are fairly elastic, especially if, as is often the case, records are understood to have historical as well as legal and business importance. But this definition does make it clear that records stand out as particularly significant documents requiring special treatment in the context of complex regulatory and business needs. For these reasons, we dedicated a portion of our survey to find out more about how institutions manage records, especially those that exist in electronic form.

**Electronic Records Management in the Organization**

Records management professionals can contribute a range of skills that complement those of IT professionals. Although IT staff have a professional familiarity with data management grounded in the disciplines of computer science and the management of information systems, records professionals know the regulatory environment that applies to the whole spectrum of records, both paper and electronic, and draw on expertise from archival and library science. As law fitfully catches up with developments in IT and as the technology itself provides better ways to get value from data, professional practices in both areas have had to evolve. The goal of this convergence of skills is to better manage the fuzzy zone between data design and operational use on the one hand and regulatory compliance and full life-cycle information management on the other. As one author puts it, records managers and IT administrators “must educate each other and their organizations on the technology and the policies to ensure that their organization makes efficient, effective use of its information in a compliant environment.”

We were therefore interested to know to what extent institutions have designated archivists or records officers and have assigned them responsibility for electronic records. As Figure 6-10 shows, the two aspects of our inquiry divide neatly in halves: 50.5% of institutions reported having a designated archivist or records officer, and those institutions in turn were almost equally divided between those where the archivist/officer is or is not responsible for electronic records. Altogether, about a quarter of respondents (25.9%) reported an archivist/officer with electronic records responsibility. (The overall percentage for archivists/officers may be slightly low due to a relatively large 8.4% who didn’t know whether their institution had a designated archivist or records officer.)

Though we found institutions with designated archivists or records officers across all Carnegie classes and institution size ranges, the incidence of such officers and the likelihood that they had responsibility for electronic records increased with FTE enrollment size (see Figure 6-11; these percentages exclude respondents who answered “don’t know”). Research-oriented institutions were considerably more likely to report having an archivist or records officer (72.3%) than were teaching-oriented institutions (48.0%). Our figure for research-oriented institutions was a little
Figure 6-10. Institution Has Archivist or Records Officer, and Archivist/Officer Responsibility for Electronic Records (N = 309)

- No archivist/officer, 41.1%
- Archivist/officer, 50.9%
- Responsible for electronic records, 25.9%
- Not responsible for electronic records, 24.6%
- Don't know, 8.4%

Figure 6-11. Institution Has Designated Archivist/Records Officer, by FTE Enrollment and Responsibility for Electronic Records

- 1-4,000 (N = 133):
  - No archivist/officer: 30.1%
  - Archivist/officer not responsible for electronic records: 17.3%
  - Archivist/officer responsible for electronic records: 52.6%

- 4,001-15,000 (N = 97):
  - No archivist/officer: 25.8%
  - Archivist/officer not responsible for electronic records: 28.9%
  - Archivist/officer responsible for electronic records: 45.4%

- More than 15,000 (N = 47):
  - No archivist/officer: 57.4%
  - Archivist/officer not responsible for electronic records: 21.3%
  - Archivist/officer responsible for electronic records: 21.3%
higher than the 66% reporting a records management function in a 2008 survey of member institutions of the Association of Research Libraries.14

A designated archivist or records officer isn’t the only path to managing electronic records, so we also asked respondents if they had a group responsible for overseeing electronic records management. Four out of 10 (40.7%) said they did, and although some of these institutions were no doubt referring to their archivist or records officer, 31 institutions that said they had no archivist/officer reported having such a group. Altogether, 65.9% of institutions said they have an archivist/officer, an electronic records management group, or both. When all overlaps and archivist responsibilities are taken into account, 49.1% of institutions officially assigned responsibility for electronic records to either an archivist/officer or to an electronic records management group.15

There are probably other ways to skin the electronic records cat than by having a designated archivist records officer or an electronic records group with responsibility for them. Still, the fact that barely half of respondent institutions reported one or both of these organizational entities for managing the exploding and increasingly regulated realm of electronic records suggests that this aspect of records management has fallen through organizational cracks at many institutions.

**Electronic Records Retention Schedules**

One of the central concepts of records management is the information life cycle, and a key aspect of that life cycle is deciding whether, and in what way, records should be retained beyond the period of their operational use. Many requirements, both external (such as regulations and legal liability) and internal (such as preserving information with business or historical value), affect institutional records retention needs, and formal documentation of records retention schedules is a commonly noted best practice in records management.16 Electronic records (or e-records) present particular challenges, since they are extremely abundant, their physical preservation and readability depend on specific technologies and systems, and full-spectrum understanding of their nature may require expertise rarely found in a single individual or department.

Retention schedules for electronic records were quite common among our respondent institutions, at least for core areas of administration (see Figure 6-12). About three-quarters or more of institutions said that they have documented e-records retention schedules for financial (81.8%), student (78.4%), and human resource (73.3%) records. Outside these domains with long traditions of administrative practice and regulatory oversight, retention schedules were less common, though by no means rare. Schedules for course-related and LMS e-records, research grant e-records, and faculty and staff e-mail were in each case reported by roughly half of institutions. Somewhat complicating the interpretation of these figures were unusually high “don’t know” response rates, particularly in relation to HR, course/LMS, and research grant e-records.

These retention schedule rates hold up well in comparison with the results of a 2007 survey of records management professionals conducted by two industry groups (ARMA International and AIIM International) and by Cohasset Associates, a records management consulting firm. In that survey, 46% of organizations said they had electronic records retention schedules for data objects such as application data—well below the rates we found for financial, student, and HR data—and 45% said they had schedules for communications including e-mail; these were about the same as the rates we found for staff and faculty e-mail.17
Overall, among our respondents the median number of documented e-records retention schedules reported was five (of seven asked about), and the average number of schedules tended to rise with FTE enrollment size—that is, bigger institutions tended to report more. Schedules also tended to be more numerous where institutions had implemented an enterprise records management system, perhaps because such systems provide a framework for creating and maintaining schedules. However, we did not find any difference between institutions on the basis of having or not having an archivist/records manager or a group responsible for overseeing e-records management.

**Records Retention Compliance**

Documenting records management policies is one thing; following them is another. Like a lot of institutional processes, records management can be difficult to carry out at the institution-wide level because of the distributed nature of records and the competing demands of higher-priority (or at least higher-urgency) operational tasks. Nor are these difficulties limited to higher education. In the 2007 ARMA/AIIM/Cohasset survey mentioned above, 45% of organizations answered with one of the two lowest-compliance categories—“not regularly” and “when time permits”—when asked whether they followed their records retention schedules. The study authors suggest that noncompliance is often the result of a lack of senior-level commitment to the schedules and the failure of records managers to take a strong role asserting consistent schedules throughout their organizations.

Such factors may help explain why, when we asked our respondents to rate their institution’s ability to comply with records retention

---

**Figure 6-12. Items for Which Institution Has Documented Electronic Records Retention Schedule**

- Financial records: 81.8% have retention schedule, 11.4% do not have, 6.8% don’t know.
- Student records: 79.4% have retention schedule, 15.7% do not have, 5.9% don’t know.
- Human resource records: 73.3% have retention schedule, 15.0% do not have, 11.7% don’t know.
- Course-related and learning management system records: 52.8% have retention schedule, 37.9% do not have, 9.4% don’t know.
- Research grants records: 51.3% have retention schedule, 27.3% do not have, 21.4% don’t know.
- Staff e-mail: 48.5% have retention schedule, 47.9% do not have, 3.6% don’t know.
- Faculty e-mail: 47.1% have retention schedule, 49.4% do not have, 3.6% don’t know.

Don’t know

Does not have retention schedule

Has retention schedule
and disposition requirements throughout the institution, most answered at the low end of our rating scale of 1 (poor) to 5 (excellent). As Figure 6-13 shows, 53.0% rated their institution as either poor or fair, and only 13.4% said it was either very good or excellent. The mean answer overall was 2.5, halfway between fair and good.20

The records management challenges implied by these results seem to be widespread. The response pattern didn’t vary significantly by Carnegie class, institutional FTE enrollment, or research/teaching orientation. Neither did we find that having a designated archives or records officer made a difference.

We did, however, find some other indications that good data management practices travel alongside better perceived ability to comply with records retention and disposition requirements throughout the institution (see Table 6-8). Mean perceived compliance tended to rise with the number of electronic records retention schedules the institution reported; among those with no schedules or only one, it averaged fair, whereas among those reporting six or seven, it averaged just slightly under good. As with a number of other important outcomes measures, we also found that institutions with higher enterprise data quality scores tended to report higher records retention and disposition compliance. Finally, respondents whose institutions had implemented an enterprise records management solution rated compliance substantially higher than did those that did not plan, or had not started, such a solution. (Only those institutions that told us they had an enterprise-oriented environment for digital content management—one of the three leftmost categories shown in Figure 6-1—were included in the enterprise records management solution result.)21

**Facing the Content Challenge**

Taken together, three measures reported in this chapter suggest a sense of unease among our respondents about meeting the challenges of the modern content

![Figure 6-13. Institution’s Ability to Comply with Records Retention and Disposition Requirements throughout the Institution (N = 306)](image-url)
environment. Asked about their institutions’ ability to effectively manage all the varieties of content their institutions need, to routinely exceed the expectations of major on-campus constituents regarding the institutional website, and to comply with records retention and disposition requirements, in each case respondents answered, on average, below or at the middle value of the response scales. Granted, one of these measures (exceeding expectations) set a high bar of performance. But then again, the institutional website is such a vital and heavily used resource that a high standard of satisfaction is worth pursuing.

We also found a rough outline of institutional response to the content challenge. Current content management environments tend to be distributed, and indeed for 4 out of 10 institutions they are entirely local or ad hoc; but looking to the near future, institutions seem to express the ambition for a more enterprise-oriented approach. That has some potential to help, considering that we found that institutions with more enterprise-oriented environments reported higher agreement about effectively managing content variety, and they were more likely to report web content management best practices that, in turn, we found positively associated with higher agreement about exceeding constituent expectations for the website. Although institutions aren’t on the whole very optimistic about complying with records retention needs, those that have documented more e-records retention schedules give themselves higher compliance ratings, as do those that have implemented enterprise

<table>
<thead>
<tr>
<th>Count of E-Records Retention Schedules</th>
<th>Rate your institution’s ability to comply with records retention and disposition requirements throughout the institution.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean*</td>
</tr>
<tr>
<td>0–1</td>
<td>2.00</td>
</tr>
<tr>
<td>2–3</td>
<td>2.14</td>
</tr>
<tr>
<td>4–5</td>
<td>2.45</td>
</tr>
<tr>
<td>6–7</td>
<td>2.93</td>
</tr>
<tr>
<td>Enterprise Data Quality Score</td>
<td></td>
</tr>
<tr>
<td>Low (less than 2.5)</td>
<td>2.22</td>
</tr>
<tr>
<td>Medium (2.5–3.5)</td>
<td>2.49</td>
</tr>
<tr>
<td>High (greater than 3.5)</td>
<td>2.79</td>
</tr>
<tr>
<td>Enterprise Records Management Solution Status**</td>
<td></td>
</tr>
<tr>
<td>Not planning/planned, not started</td>
<td>2.42</td>
</tr>
<tr>
<td>Implementation in progress</td>
<td>2.55</td>
</tr>
<tr>
<td>Implemented</td>
<td>3.03</td>
</tr>
</tbody>
</table>

*Scale: 1 = poor, 2 = fair, 3 = good, 4 = very good, 5 = excellent

**Data excludes institutions reporting only local/departmental or ad hoc solutions for enterprise management of digital content.
records management solutions. And, finally, in a recurring theme in this study, better data quality keeps company with higher agreement about both exceeding website constituent expectations and better compliance with records retention requirements.

Our outline suggests, then, that institutions can take some practical steps toward taming the content beast: approaching content from an enterprise view, being diligent about indentifying and documenting the subset of content and data that needs special treatment as records, and in general maintaining a high-quality data environment. Yet we can’t help but speculate that even institutions attempting to make these improvements at the margins will struggle to keep up with increasing content scope and complexity in the absence of some new, radically more potent paradigm for managing content according to its value and use, rather than its type and point of origin. Speaking of the difficulties of managing the modern data environment, Renee Drabier of the University of North Texas Health Science Center confessed, “I feel a little shortsighted about what we might be using in 10 years. So far it’s just been bigger, faster, more reliable. But I feel there has to be a sea change coming. It’s hard to prepare for the unknown. You just have to make the best possible decisions now and deal with what comes as best you can.”

Endnotes
3. The results in Table 6-5 apply only to the 180 respondents who reported an enterprise-oriented environment for digital content management, because only those respondents were asked about the status of an enterprise web content management solution (see Figure 6-3 and related discussion). On the hypothesis that most or all of those respondents who had non-enterprise-oriented environments could be assumed not to have enterprise web content management solutions, we tested this relationship for the whole population of 309 institutions. It was also statistically significant, though the association was not as strong.
6. For an analysis of MIT’s approach to mobility, see Bob Albrecht and Judith A. Pirani, “Massachusetts Institute of Technology: Transforming the Campus Experience with the MIT Mobile Web” (Case Study 3) (Boulder, CO: EDUCAUSE Center for Applied Research, 2009), available from http://www.educause.edu/ecar.
8. Ns/standard deviations were 292/0.878 (faculty), 296/0.913 (staff), 292/0.911 (students), and 287/0.889 (public users).
9. Ns shown on detail lines in Table 6-6 represent the smallest N of the separate means reported on each line; the actual Ns vary slightly.
10. A small number of “in few or no cases” responses to the consistent look-and-feel item make the means for this response slightly less reliable than the other means reported in this table.
11. Ns shown on detail lines in Table 6-7 represent the smallest N of the separate means reported on each line; the actual Ns vary slightly.
15. These figures exclude those who answered “don’t know” to the question about having a designated archivist or records officer.

18. Analysis of median and mean numbers of documented retention schedules excluded “don’t know” responses.

19. Williams and Ashley, *Call for Collaboration*, 22–23. The other two answer categories were “generally” (50%) and “always” (14%).

20. N = 306, SD = 0.906.

21. We only included the 180 respondents who reported an enterprise-oriented environment for digital content management because only these respondents were asked about the status of an enterprise records management solution (see Figure 6-3 and related discussion). Some respondents answered “don’t know” to the records management solution question or skipped it, and these are not included in the Ns in Table 6-8. On the hypothesis that most or all of those respondents who had non-enterprise-oriented environments could be assumed not to have enterprise records management solutions, we tested this relationship for the whole population of 309 institutions, and it was also statistically significant.
7 Research Data Management

A record, if it is to be useful to science, must be continuously extended, it must be stored, and above all it must be consulted.
—Vannevar Bush

Key Findings
- About 39% of institutions reported increase or great increase in the use of high-performance computing during the last three years, and small majorities reported growth in high-performance networking and research data storage. Research-oriented institutions were much more likely to report increases than teaching-oriented institutions.
- Substantial majorities of institutions expected increases in the use of high-performance computing and networking and in the amount of research data storage in the next three years.
- Central IT units provided the greatest extent of research data management support in the areas of research data storage and backup. Individual investigators, labs, or teams provided a much higher extent of support for research data management activities at research-oriented institutions than at teaching-oriented institutions.
- Respondents were not sanguine about their institutions’ ability to support the long-term preservation of research data. Of five items rated on a 5-point agreement scale, none reached as high as a mean 3.0. Infrastructure was the most highly rated aspect of long-term research data preservation (mean 2.96); funding mechanisms were the lowest (mean 2.15).
- About 4 in 10 respondent institutions said they have a documented policy defining ownership of research data; these policies were about twice as common among research-oriented institutions as among teaching-oriented ones.
- Overall agreement that the institution meets the data-related needs of institutional researchers was neutral (3 on a 5-point scale). Teaching-oriented institutions tended to agree more strongly than research-oriented ones; doctoral institutions had the lowest mean agreement among Carnegie classes, and baccalaureate and associate’s institutions had the highest.

Nature may still be mostly analog, but science is increasingly digital—in the computational methods that have become a new way of doing science, in the way that experimental results are captured, and in the modes scientists use to communicate with
each other, their students, and the public at large. In an influential report published in 2005, the National Science Foundation recognized that information technology had ushered in “fundamentally new approaches to research and education” and noted that “digital data collections are at the heart of this change.”1 Humanities scholarship, too, is being transformed by mass digitization projects and network resources that, as a recent report from the Council on Library and Information Resources puts it, “have created collections [of] a quantity, diversity, and scale hitherto unknown, as well as a community of scholars, librarians, and archivists with a common interest in long-term preservation of digital content.”2

Yet what sort of container this elixir of modern science and scholarship can or should be put in is not at all clear. Research data is enormously abundant—both in the staggering output of some cutting-edge instruments (the output of the Large Synoptic Survey Telescope, currently under construction, is expected to be 30 terabytes per night3) and in the collective size of the many smaller data sets and project communications that might have long-term documentary importance. In addition, investigators may be loath to archive research data whose interpretation requires intimate knowledge of the experiments that produced it, while tagging it with metadata robust enough to make it usable presents difficult technical and standards development issues. And unlike traditional paper records, digital data sets are only readable within specific technology environments. How many modern university IT operations still have the ability to read a deck of punched cards or even a truly “floppy” 5.25-inch disk?

Traditionally, researchers have kept and maintained their own data. But the size and complexity of modern data sets increasingly make this impractical, while obvious economies and research benefits would be realized from making data available to a wider community of investigators. As Clifford Lynch of the Coalition for Networked Information has argued, “The demands for data management and curation to facilitate data sharing and reuse form one of the fundamentally new aspects of e-research.”4 But where and how will we find a home for all the new research-related data we’re so busily creating?

A complete answer to this question is beyond the scope of this study, but we did investigate a number of areas related to research data management, focusing on institutional initiatives. Building on work from two previous ECAR studies, we look in this chapter at:

- how use of research computing infrastructure has changed in recent years and how respondents expect it to change in the near future;
- the extent of various kinds of research data management support offered by different institutional entities;
- respondents’ views on their institutions’ ability to support the long-term preservation of research data;
- policies regarding the ownership, archiving, and sharing of research data;
- the status and use of institutional research data repositories; and
- respondents’ overall assessments of whether their institutions meet the data-related needs of institutional researchers.

### Research Computing Infrastructure

To get a sense of the overall research computing environment on campuses, we asked respondents how three key activities—use of high-performance computing, use of high-performance networking, and the amount of research data storage—had changed in the past three years and how they expected them to change in the next three years.
Respondents made one thing clear: This was not a story about decreasing activity (see Table 7-1). Only 3.4% said use of high-performance computing had decreased or greatly decreased in the past three years—and that was the greatest percentage of decrease we measured for any item or time frame.

On the other hand, respondents were split about whether things had stayed the same, increased, or greatly increased in the last three years. Almost 6 in 10 (58.0%) said use of high-performance computing had stayed the same, and though majorities said use of high-performance networking and the amount of research data storage had increased or greatly increased, “stayed the same” percentages were sizable for these items. Looking forward three years produced a stronger emphasis on increase: 58.6% expected high-performance computing to increase or greatly increase, compared with 70.2% for high-speed networking and 68.7% for the amount of research data storage.

As might be expected, the results varied substantially by institutional mission, with research-oriented institutions more consistently reporting increases across the board. As Table 7-2 shows, research-essential and balanced institutions averaged well above 3.5 (midway between stay the same and increase on our 5-point scale) for all items regarding the past three years and near or above 4.0 for the future-oriented items. Among teaching-favored and teaching-essential institutions, average responses were considerably lower on all items, though still above an average “stay the same.”

In short, institutions overall reported a good deal of both past and anticipated increase in the use of these research infrastructure items, but increase was not universal, thanks to the research/teaching divide. For example, although 90.1% of research-oriented institutions reported increased or greatly increased amounts of data storage in the last three years and a whopping 97.5% expected such increases in the next three years, the corresponding figures for teaching-oriented institutions were 36.9% and 56.6%.

At least in the broad terms captured by our subjective decrease/increase scale, the patterns reported here seem to be part of a long-term process. Our results were close to those of a 2005 ECAR survey on IT support for research that asked similar questions. Like our results, that study’s results indicated substantial overall past and anticipated growth for these research infrastructure items, differentiated by research and teaching missions. A comparison of the results from the 112 institutions that answered both the 2005 and 2009 surveys found no significant difference in the responses relating to use of high-performance computing and high-

<table>
<thead>
<tr>
<th>How have activities changed/how expected to change...</th>
<th>Greatly Decreased</th>
<th>Decreased</th>
<th>Stayed the Same</th>
<th>Increased</th>
<th>Greatly Increased</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In the Past Three Years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of high-performance computing (N = 288)</td>
<td>1.7%</td>
<td>1.7%</td>
<td>58.0%</td>
<td>30.9%</td>
<td>7.6%</td>
</tr>
<tr>
<td>Use of high-performance networking (N = 287)</td>
<td>0.7%</td>
<td>0.3%</td>
<td>42.9%</td>
<td>45.6%</td>
<td>10.5%</td>
</tr>
<tr>
<td>Amount of research data storage (N = 272)</td>
<td>0.4%</td>
<td>0.7%</td>
<td>46.7%</td>
<td>43.8%</td>
<td>8.5%</td>
</tr>
<tr>
<td><strong>In the Next Three Years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of high-performance computing (N = 285)</td>
<td>0.4%</td>
<td>0.7%</td>
<td>40.4%</td>
<td>43.5%</td>
<td>15.1%</td>
</tr>
<tr>
<td>Use of high-performance networking (N = 289)</td>
<td>0.3%</td>
<td>0.3%</td>
<td>29.1%</td>
<td>51.9%</td>
<td>18.3%</td>
</tr>
<tr>
<td>Amount of research data storage (N = 275)</td>
<td>0.0%</td>
<td>0.4%</td>
<td>30.9%</td>
<td>48.0%</td>
<td>20.7%</td>
</tr>
</tbody>
</table>
performance networking, and inconclusive results for research data storage, due to slight changes in question wording.\(^5\)

**Research Data Management Support: Who Does What?**

We reported above that past and anticipated future increases in the amount of data storage were nearly universal among research-oriented respondent institutions and far from uncommon among teaching-oriented ones. But this pressure for more data storage is only the beginning of the story. All the services associated with a good data management environment, including backup and recovery, metadata management, and regulatory compliance, apply to research data as well.

What’s more, granting agencies trying to improve research efficiency are paying more attention to the concept of data management itself. For example, the National Science Foundation’s grant application guidelines specify that proposals include, as appropriate, “plans for preservation, documentation, and sharing of data,” and its *Grant General Conditions* require that investigators “share with other researchers...the data, samples, physical collections, and other supporting materials created or gathered in the course of the work.”\(^6\) Such ideas form one part of a broader vision of cyberinfrastructure or e-science, which promotes a set of shared resources, including advanced networks and data management tools, that can help support and tie together a whole universe of research initiatives.

To some extent, all this runs counter to entrenched practices in campus research. An entrepreneurial system of grants competition has encouraged something of a cottage

---

Table 7-2. Change in Research-Related Activities, Past Three Years and Next Three Years, by Institution Mission

<table>
<thead>
<tr>
<th>Institution Mission</th>
<th>Use of High-Performance Computing</th>
<th>Use of High-Performance Networking</th>
<th>Amount of Research Data Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Past Three Years</td>
<td>Next Three Years</td>
<td>Past Three Years</td>
</tr>
<tr>
<td>Research Essential</td>
<td>Mean*</td>
<td>3.89</td>
<td>4.33</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.916</td>
<td>0.679</td>
</tr>
<tr>
<td>Balanced</td>
<td>Mean*</td>
<td>3.75</td>
<td>4.12</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>59</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.822</td>
<td>0.677</td>
</tr>
<tr>
<td>Teaching Favored</td>
<td>Mean*</td>
<td>3.24</td>
<td>3.58</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>98</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.690</td>
<td>0.706</td>
</tr>
<tr>
<td>Teaching Essential</td>
<td>Mean*</td>
<td>3.24</td>
<td>3.47</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>101</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.513</td>
<td>0.625</td>
</tr>
</tbody>
</table>

*Scale: 1 = greatly decrease, 2 = decrease, 3 = stay the same, 4 = increase, 5 = greatly increase*
industry of autonomous investigators and laboratories with independent funding. Investigators often want a lot of control over their IT environments, both to better serve their highly specialized needs and to keep grant funds within the team. Despite their visionary celebration of shared infrastructure, the funding agencies have not altogether resolved the tension between the current funding model and the vision. “There’s some sense that the grants process favors local over centralized storage,” said Terri-Lynn Thayer, assistant vice president and deputy CIO at Brown University. “There is a high incentive for people to buy all their own stuff from these grants, and that’s not in the best interest of any of us in the long term.”

But researchers have also reached out to central IT and other institutional units in order to control costs—for example, by housing research clusters in the central data center—and to find skills not easily provided within a research team. Institutions, in turn, may be crucial in facilitating the development of research cyberinfrastructure that is beyond the means of individual investigators or labs. A 2008 ECAR study on higher education IT and cyberinfrastructure found that where central IT was a provider or funder of high-performance computing and research data storage and management resources, respondents were more likely to agree that the institution realized significant economies of scale in use of cyberinfrastructure resources.7

We wanted to know how these forces are playing out with respect to research data management among our respondent institutions. Are investigators largely taking care of their own data management needs, or are they looking to central IT or other entities? And to what extent do different entities provide different services?

We asked respondents to what extent (from not at all to a very large extent) each of four different entities provided the following with respect to data resulting from research conducted by institutional investigators:

- research data storage,
- research data backup and recovery,
- assistance with creating or fulfilling research data management plans,
- metadata creation, and
- support in selection and use of research tools (e.g., visualization, data mining, statistical analysis).

The four entities we asked about were central IT; individual investigators, labs, or teams; school, center, or department IT; and the library.

Before exploring our results, it’s worth noting a few limits that relate to them. Our respondents were primarily CIOs and were probably more knowledgeable about central IT roles than about those of other entities, and this may have led them to underestimate the extent to which those entities are involved in research data management. Previous ECAR research has found that CIOs rate their overall knowledge about the research use of data storage cyberinfrastructure technologies as good rather than excellent (close to 3 on a 5-point scale from poor to excellent).8 This may help explain the relatively high rate of “don’t know” responses to some of our questions, which in some cases are sufficient to materially affect the results, depending on how the “real” answers are distributed.

**Major Providers: An Overview**

Here, we provide an overview of research data management support activities, focusing on those institutional entities that respondents said provided support to a large or very large extent. We look in detail at the results for each of the activities in the next section.

Central IT’s major contribution to research data management support was in the areas of data storage and data backup/recovery, provided to a large or very large extent by 40.3% and 52.2% of institutions, respectively (see Figure 7-1). These items suggest
that researchers are taking advantage of traditional central IT capabilities and strengths: ownership of the data center with its cost-effective environmentals, enterprise systems administration skills, and perhaps disaster recovery planning. These are the sorts of arrangements John Rome, associate vice president for University Technology at Arizona State University, described at ASU. “We work closely with the College of Engineering to host some of their systems,” Rome said. “We can offer researchers physical space, and primary and backup storage. Managing these services gives the researchers and departments one less thing to worry about.”

Central IT was less conspicuous in other areas we asked about, though about one in five institutions (20.9%) said that central IT was involved to a large or very large extent in creating or fulfilling data management plans. This sort of support may be an outgrowth of the work associated with data storage and backup/recovery; extent of central IT support for creating or fulfilling data management plans was correlated positively with each of these other two items.

Besides central IT, the other entity most often mentioned as providing each activity to a large or very large extent was individual investigators, labs, or teams. Investigators were more often reported as large-extent providers for the data management plan, metadata creation, and research tools items than central IT, though except for the research tools item, the differences are so small that they may not be statistically meaningful.

The extent to which different entities supported research data management activities varied by institutional mission, though not as much as we expected. There wasn’t much difference, for example, between research-oriented and teaching-oriented institutions in the frequency with which central IT and the library provided large-extent support to the activities we asked about.
However, as Table 7-3 shows, dramatic mission-related differences were evident in the percentages of institutions reporting large-extent support by investigators, labs, or teams and by school, center, or departmental IT units. At research-oriented institutions, these entities were several times more likely to provide a large/very large extent of support than they were at teaching-oriented institutions. (Though the differences aren’t shown in Table 7-3, “not at all” responses were correspondingly much more common among teaching than among research institutions.) This greater degree of investigator and local IT unit provision of data support services isn’t terribly surprising, since research institutions are more likely to have active investigators and laboratories with their own grant resources, and also more likely to have separate school, center, or departmental IT units. Perhaps more surprising is the relatively small role the local IT units played in research data management support, compared with both investigators and central IT.

No doubt the question of how to support researcher needs and encourage a productive research environment will always involve messy boundary issues and a complicated interaction between researchers, their extra-institutional research colleagues, and the different entities within the institution. Our results surely reflect the messiness to some degree, but just the same we found evidence both for a high level of investigator self-help in matters of research data management at research-oriented institutions and for substantial contributions generally by other entities, particularly central IT.

**Research Data Management Support in Detail**

By focusing only on responses indicating a large or very large extent of research data management support, Figure 7-1 underscores some dramatic contrasts in the roles of different institutional entities. However, the overall picture of support is more complex, as the full details presented in Table 7-4 show.

The complete results suggest a distributed pattern of research data management support spread across multiple entities at the institution. For example, although rather few respondents reported that school, center, or

<table>
<thead>
<tr>
<th>Institutional Entity</th>
<th>Activity</th>
<th>Institutions Where Entity Provides Activity to Large/Very Large Extent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Research Oriented</td>
</tr>
<tr>
<td>Investigators or Labs</td>
<td>Research data storage (N = 296)</td>
<td>54.7%</td>
</tr>
<tr>
<td></td>
<td>Research data backup/recovery (N = 293)</td>
<td>32.6%</td>
</tr>
<tr>
<td></td>
<td>Assist with data management plans (N = 288)</td>
<td>46.4%</td>
</tr>
<tr>
<td></td>
<td>Support metadata creation (N = 287)</td>
<td>31.3%</td>
</tr>
<tr>
<td></td>
<td>Support selection/use of research tools (N = 295)</td>
<td>49.4%</td>
</tr>
<tr>
<td>School, Department IT</td>
<td>Research data storage (N = 290)</td>
<td>29.4%</td>
</tr>
<tr>
<td></td>
<td>Research data backup/recovery (N = 291)</td>
<td>12.9%</td>
</tr>
<tr>
<td></td>
<td>Assist with data management plans (N = 288)</td>
<td>10.7%</td>
</tr>
<tr>
<td></td>
<td>Support metadata creation (N = 289)</td>
<td>7.1%</td>
</tr>
<tr>
<td></td>
<td>Support selection/use of research tools (N = 296)</td>
<td>15.1%</td>
</tr>
</tbody>
</table>
departmental IT units provided a large/very large extent of support (from a low of 4.1% for supporting metadata creation to a high of 13.0% for research data storage), considerably more indicated that they provided a moderate extent (from 11.0% for metadata creation to 21.9% for data storage).

In fact, when all respondents are considered together, the only activity/entity combination with a median response (that is, the point at which half of responses are above and half below) reaching as high as “large extent” was central IT provision of research data backup/recovery. Most other combinations had median responses ranging from very small to moderate extent, and for six combinations the median was “not at all.”

Five of these six combinations related to the library, which on the whole was the least involved of the entities we asked about.
Although this may seem unremarkable, given the library’s traditional concerns, a good deal of emerging thought about the problems of research data management has noted that librarianship has many appropriate skills to contribute, such as experience with the design and application of metadata, archival and curatorial expertise, an ethic of open access, and experience with multi-institutional cooperation. The library is also in a position to fill the conspicuous vacuum left by the absence of existing alternative models for the more generalized, comprehensive preservation of research data. Institutional digital repositories, for example, though still in an experimental phase, are commonly administered by institutional libraries. Such cutting-edge developments, however, haven’t left much of a mark on our quantitative results, and if libraries are becoming more heavily involved in research data management support, the news doesn’t seem to have reached many of our CIO respondents.

It’s worth mentioning, however, that the whole world of research data management is in flux and that even a small role can potentially be a significant one. One of central IT’s emerging roles may be educating users about where skills and resources can be found, and coordinating or catalyzing services others can provide. John Rome at Arizona State, for example, saw facilitation rather than direct services provision as the future of central IT support for research. “We think we’ll become more of a facilitator for departments to find services. We see ourselves as a resource broker for researchers seeking access to systems and support.”

**Long-Term Preservation of Research Data**

One of the key emerging needs of the era of digital data-intensive research is some means of preserving data over the long term. Data that is kept in safe, undegraded, discoverable form is available for reproducing scientific results and for new investigations, and it documents the history of the research enterprise. Preserving data beyond the immediate needs of its creators, however, involves costs that research teams can’t bear and skills they are unlikely to possess.

Unfortunately, no really comprehensive model for preserving research data over the long run has emerged, despite widespread discussion of its potential benefits. Although some research organizations maintain their own archives, and assorted project-specific or disciplinary repositories exist, at best they constitute a patchwork with many gaps. Clearly, given the variety of types of data and consumers that research involves, some sort of global, networked solution would be optimal. But this is not to say that institutions have no role to play. Besides creating repositories for local research that could act as the nodes in a global system of research archives, institutions can provide curatorial and technical services to investigators much as they provide laboratory space, utilities, administration, and other services today.

In a recent article, Michael Witt of Purdue University noted some of the requirements of the data curation enterprise:

> The preservation and archiving of data so that they can be accessed and used in the distant future necessitates economically and technologically sustainable systems for ongoing curation activities such as data integrity checking and reversioning...appropriate points of access with both human and machine interfaces....[and] proper metadata [to] describe the data to support functions such as discovery, use, preservation, and administration. The provenance of data needs to be recorded.... Intellectual property rights must be determined, [and] policies are
needed to govern submissions, selection, usage, and levels of service, at a minimum. This list of challenges only begins to scratch the surface.12

With these requirements in mind, we decided to scratch the surface a bit ourselves. In the following sections, we examine whether respondent institutions assume archiving responsibilities, how well our respondents thought their institutions were prepared for such tasks, and to what extent institutions were making use of institution-wide or distributed research data repositories.

**Institutional Responsibility for Archiving Research Data**

Although we knew that research data preservation is largely uncharted ground at the institutional level, we were still surprised at how circumscribed and uncertain a picture our survey questions uncovered. Asked whether their institutions assume responsibility for archiving research data after investigator projects are concluded, 4 in 10 of our respondents said they didn’t know (see Figure 7-2). Almost half of the remainder (27.8% of all respondents) said that their institutions didn’t assume responsibility in any case. Where the institution did assume responsibility, it was usually only in a few or some cases; fewer than 1 in 10 respondents reported an institution’s assuming responsibility for archiving research data in most or all cases.

Many of our respondent institutions had a teaching rather than a research orientation, of course, and this may help account for the large proportion of “don’t know” responses. Almost half of teaching-oriented institutions (46.8%) gave “don’t know” responses, and we speculate that some of these were indicating something more like “not applicable.” But a substantial 28.1% of research-oriented institutions also gave “don’t know” replies, suggesting that these issues lack visibility on many campuses, either due to inactivity or because what activity there is takes place

![Figure 7-2. Institution Assumes Responsibility for Archiving Research Data after Investigator Projects Concluded (N = 309)](image)
outside the domains of our predominantly CIO respondents. Although research-oriented institutions were somewhat more likely to report that the institution assumes responsibility for archiving research data, the number of those reporting “in no cases” was not enormously lower among them (22.5%) than it was among the teaching-oriented institutions (29.8%).

We asked institutions that reported some level of institutional responsibility for archiving research data who was involved in the process (see Figure 7-3). Not surprisingly, among the parties we asked about, faculty were easily the most commonly involved (84.3%); it would be strange if institutional researchers were not involved in archiving their own data. But at 67.4%, central IT was virtually tied with research administration (69.7%) as the second most involved party and was much more commonly reported than was the library (36.4%). Keeping in mind the high “don’t know” response rate that hovers over all these numbers, we surmise that central IT’s prominence in these results may simply reflect respondents’ greater familiarity with their own roles than with those of other parties. Still, where some institutional role in archiving research data exists, central IT apparently is often involved. We did not find any substantial differences between research- and teaching-oriented institutions in the participation of the various parties.

Although these quantitative results suggest a contrast between a more active central IT organization and a less active library, qualitative interviews at institutions pursuing a research archiving initiative tended to reveal an important role for the library and close cooperation between the library and central IT. At Georgia State University, for example, J. L. Albert, associate provost and CIO, said, “The vice president for research sets archiving policy, IT is the storage provider, and the library supports archival data and the institutional repository. We hope to evolve to the point where the library provides the front-end data services and IT provides the back end.” At the University of North Texas, deployment of a new research cluster with massive storage has catalyzed a similar relationship. “Policy about data archiving will be drafted by data management experts in the library and the College of Information,” said Maurice Leatherbury, acting vice president for information technology and CIO. “The IT, library, and

![Figure 7-3. Officers/Organizations Taking Part in Institutional Responsibility for Archiving Research Data](image-url)
college partnership will evolve over time. We’ll provide the technology to the people who have expertise in managing data, and they in turn will assist the users.”

Assessing Institutional Capabilities

Besides asking about institutional responsibility for archiving research data, we asked respondents to express their level of agreement with a set of statements relating to their institutions’ ability to support the long-term preservation of research data. Their mean responses (excluding roughly 10% who answered “don’t know” to each item) appear in Table 7-5.

Not a single item we asked about rose as high as an average neutral response (3 on our 5-point agreement scale), though agreement that institutions have the necessary infrastructure to support long-term preservation of research data came very close (mean 2.96). Respondents also summoned something close to neutral agreement about having the necessary expertise. However, when it came to institutional commitment, policies and procedures, and especially funding mechanisms, respondent mean responses were considerably lower, ranging from the midway point between disagree and neutral to close to outright disagreement. Two-thirds of respondents (68.5%) either disagreed or strongly disagreed about having the necessary funding mechanisms.

Given the uncertainty many respondents expressed when we asked whether their institutions assumed responsibility for archiving research data (Figure 7-2) and given the rather modest levels of involvement in research data management activities that they reported (Table 7-4), it’s possible that these low figures reflect a lack of awareness or experience rather than low absolute institutional capacity to support long-term preservation of research data. Institutions reporting a greater extent of central IT support for some research data management activities—for example, assisting researchers with creating or fulfilling data management research plans—averaged higher agreement on these long-term preservation items. Although this may reflect a more preservation-capable environment, it may also be that these respondents simply have a fuller understanding of what’s going on at their institutions. Then again, as the fairly dismal level of agreement about funding mechanisms suggests, having services to offer and being able to sustain them economically are two different things. Georgia State’s J. L. Albert lamented that “researchers are looking to IT to provide data architecture and metadata expertise, as well as capacity, backup, restore, and archival services. It’s a daunting prospect in these economic times.”

Low agreement on these research preservation capability items was widespread across all kinds of institutions and in particular did not differ significantly between

Table 7-5. Support for Long-Term Preservation of Research Data

<table>
<thead>
<tr>
<th>To support the long-term preservation of research data, my institution has the necessary...</th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infrastructure</td>
<td>2.96</td>
<td>271</td>
<td>1.167</td>
</tr>
<tr>
<td>Expertise</td>
<td>2.85</td>
<td>267</td>
<td>1.098</td>
</tr>
<tr>
<td>Institutional commitment</td>
<td>2.54</td>
<td>264</td>
<td>0.950</td>
</tr>
<tr>
<td>Policies and procedures</td>
<td>2.27</td>
<td>268</td>
<td>0.854</td>
</tr>
<tr>
<td>Funding mechanisms</td>
<td>2.15</td>
<td>266</td>
<td>0.892</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
research-oriented and teaching-oriented institutions. It’s hard to avoid the conclusion that the lively and important discussion now going on about the long-term preservation of research data has not so far had a lot of practical impact on the capabilities of our respondent institutions.

Research Data Repositories

For the better part of the last decade, higher education administrators have been pondering the value of an electronic institutional repository that holds and makes available content produced by institutional researchers and scholars. What might be called the institutional repository movement was well under way by 2005, stimulated by open source repository software such as DSpace, Fedora, and EPrints, and by late 2009 a registry of open-access repositories counted more than 1,300 academic repositories worldwide.13 In U.S. colleges and universities, institutional repositories are typically administered by the library, with the expectation that faculty and other appropriate parties will voluntarily contribute content.14

Institutional repositories are controversial, in part owing to fundamental questions about the relationship of institution-specific collections to global needs and in part due to practical questions about their scope, technical direction, and connection to the fast-changing domain of scholarly communication. One repository manager has complained that faculty need dynamic data management tools with capabilities such as versioning and access control, whereas the typical repository is “like a roach motel. Data goes in, but it doesn’t come out.”15 Many institutional repositories have struggled to collect a critical mass of institutional research output, stymied by complex intellectual property rules, faculty apathy, and what one observer calls “an inherent [researcher] culture of self-sufficiency in the generation and organisation of data.”16

Such challenges have so far tended to keep repositories relatively small, experimental, and focused on scholarly (including student) output, collecting such materials as electronic theses and dissertations, e-prints, conference proceedings, and “gray literature” (unpublished or prepublication working papers). However, some repository advocates have argued for a broader role for repositories in research data management, noting that they can provide a platform for research data curation and archiving of digitized special collections and e-learning content.

This conception of the institutional repository as an infrastructure for research data management led us to ask our respondents about the status of institution-wide and local (school, center, or departmental) repositories for institutionally produced digital research data. As Figure 7-4 shows, institution-wide repositories were unusual; fewer than 1 in 10 respondents reported one. Interest in the idea, however, is reflected in the fact that almost as many institutions reported an institution-wide repository in development. Research-oriented institutions were about twice as likely to report institution-wide repositories as were teaching-oriented ones (13.5% versus 7.3%), and there was a slightly wider difference in repositories under development (14.6% versus 6.4%). Institution-wide repositories certainly don’t seem to be designed as data “roach motels”: 81.1% of those reporting them in operation or under development said that their repositories supported standards-based metadata harvesting such as that offered by the Open Archival Information System (OAIS) model—a key component for external discovery of repository materials.

Research repositories were more common at the school, center, or department level (see Figure 7-5), though a high rate of “don’t know” responses makes this a somewhat soft observation. A little more than one in five institutions said they have such local repositories, matched (actually, slightly
exceeded) by the proportion of respondents who said they didn’t know. We presume that many of these are modest repositories dedicated to special purposes. Among those institutions reporting local repositories, the disciplines most frequently named as represented in the repositories were physical and life sciences, followed by engineering and health sciences, social sciences, business, humanities, and arts.
Respondents at institutions with repositories (whether institution-wide or local) tended to rate their institution’s commitment to the long-term preservation of research data somewhat higher than those lacking them. They were also more likely to say that their institution assumed responsibility for archiving research data after investigator projects were concluded. Still, we emphasize that our findings relating to institutional repositories should be read with caution. The technology, policy context, and organizational absorption of these resources still seem to be highly immature, and the curation of research data, as opposed to research end products such as dissertations and e-prints, is still a leading-edge concept within the repository movement. We’re inclined to believe that research data storage is more often an incidental or potential activity at most of the repositories our respondents reported rather than the repositories’ primary purpose.

Research Data Policies

Privacy and other regulatory concerns, questions of research ethics and intellectual property, funding agency requirements, and accounting for research productivity are only some of the dimensions of research data management that call for a supporting environment of institutional policy. We reported above (Table 7-5) that policies and procedures were one of the lower-rated aspects of institutional support for the long-term preservation of data. We also wanted to know how many institutions maintained specific documented policies on data retention and other issues.

One issue fundamental to determining how to manage research data is who owns it. In their role as grantees, institutions are recognized by the NSF, NIH, and other funding agencies as the owners of data generated by grant-funded research. Some sponsors, however, may claim full or partial ownership, and rights to data may be differentiated by use (for example, research, educational, or commercial). The desire of many institutions to exploit intellectual property developed by their researchers, the growing practice of industry-sponsored research, the need to reconcile investigators’ physical custody of data with institutional responsibilities regarding retention and oversight, and the questions that arise when investigators change institutions are all reasons why the Council on Governmental Relations recommends that institutions “consider the establishment and communication of policy describing rights and obligations of all parties in the management and retention of data.”

Among our respondent institutions, however, that practice remains the exception rather than the rule. Overall, about 4 in 10 institutions said that they have a documented policy defining ownership of research data (see Figure 7-6). About two-thirds of research-oriented institutions either have a policy (62.9%) or have one under development (5.6%), considerably more than among teaching-oriented institutions (30.3% and 10.1%, respectively). Respondents who told us that investigators at their institutions conduct contract research for third parties (such as commercial entities) were much more likely to report a policy (60.6%) than were other respondents (27.2%), suggesting that the involvement of external parties may encourage ownership policy development. A high rate of “don’t know” responses to the contract research question, however, makes this a tentative association.

Respondents with documented ownership policies gave us a wide range of answers when we asked about the extent to which the policy asserted institutional ownership of research data, but on the whole responses favored substantial institutional claims (see Figure 7-7). A total of 36.9% said that the institution asserted ownership in most or all cases, and due to another high proportion of “don’t know” responses, this constituted just over half of all the respondents who provided
In no cases
In a few cases
In some cases
In most cases
In all cases
Don’t know

Figure 7-7. Extent to Which Policy Asserts Institutional Ownership of Research Data (Institutions with Policy, N = 149)

Figure 7-6. Status of Institutional Policy Defining Ownership of Research Data (N = 309)

Institutions in the extent of institutional assertion of ownership they reported, nor did we find any differences based on Carnegie class,

an answer other than “don’t know.” We did not find a significant difference between research-oriented and teaching-oriented
FTE enrollment, or public/private control. Only about one in four institutions with documented data ownership policies said that their policies permitted a third party to own research resulting from contract research, but here again the dominant response was “don’t know.”

Our additional questions about research data policies yielded generally similar patterns (see Figure 7-8). We asked respondents whether their institutions had documented policies regarding

- investigator responsibility for compliance with privacy/confidentiality regulations relating to research data,
- investigator responsibility for sharing research data with other researchers,
- research data security, and
- retention of research data.

As Figure 7-8 shows, research-oriented institutions were in every case much more likely to report a documented policy than were teaching-oriented ones. “Don’t know” responses were sizable across the board, particularly at teaching-oriented institutions, but overall the item relating to investigator compliance with privacy/confidentiality regulations was the most frequently reported. Rates for the other policy items hovered around 50% at research-oriented institutions and around 20% at the teaching-oriented ones. Given the high “don’t know” rates, which in some cases exceeded the rates of reported policies, it’s hard to be conclusive, but apparently in their research data policy initiatives, our respondent institutions have been driven more by regulatory concerns than by other considerations.
Meeting Researchers’ Data-Related Needs

In this chapter, we’ve reported on numerous aspects of research data management support, from the participation of various parts of the institution in support activities to long-term data preservation and the policy environment. We’ve noted along the way that often research institutions are more involved or more likely to document a given policy than teaching institutions. But different kinds of institutions have different needs and pursue different kinds of research. We wanted to know how institutions looked upon research data management support in a more summary, if also more subjective, way: Do respondents agree that their institutions are meeting institutional researchers’ data-related needs?

On the whole, they were far from enthusiastic. The mean answer on our 5-point agreement scale was 3.02, almost exactly a neutral response. What’s more, research-oriented institutions averaged lower agreement (2.81) than teaching-oriented ones (3.12). Doctoral institutions in particular were less optimistic than institutions in the other Carnegie classes; as Table 7-6 shows, their mean agreement (2.54) was about midway between a disagree and a neutral answer. Half of doctoral institutions (51.8%) disagreed or strongly disagreed that they meet the data-related needs of institutional researchers; at MA institutions, with the next highest rate of disagreement, only 29.3% answered with one of these levels of agreement.

Given that research-oriented institutions tended to have the “bigger and better” rates of activity in the research data management items we asked about, we suspect that these differences in perceived institutional ability to meet researchers’ needs relate more to a disproportion in demand than supply; that is, research data management needs are simply much greater at doctoral institutions than even at master’s and baccalaureate institutions where research plays a larger than usual role. In addition, doctoral institutions tend to be more distributed, making it harder to mount a coordinated and coherent institutional response to researcher needs—or even to determine whether that’s the desired approach. Gene Wiemers, vice president for Information and Library Services at Bates College, said that a combination of centralized administration and relatively modest research computation needs permits his office to fine-tune its support for researchers without overstretching. “Bates is a small liberal arts college and we manage things centrally,” Wiemers said. “This isn’t like a big university, which may provide bandwidth and storage, and it’s somebody else’s problem how to use it.” Where researchers do have large-scale needs, Wiemers added, it’s important to work with them closely “because certain types of research activities can really have an impact on the institution as a whole if you don’t manage them properly.”

We found some evidence that the role of one key central player—central IT—makes a difference in respondents’ perceptions about their institutions’ ability to meet researcher

<table>
<thead>
<tr>
<th>Carnegie Class</th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR</td>
<td>2.54</td>
<td>56</td>
<td>0.894</td>
</tr>
<tr>
<td>MA</td>
<td>3.01</td>
<td>82</td>
<td>1.000</td>
</tr>
<tr>
<td>BA</td>
<td>3.29</td>
<td>58</td>
<td>0.817</td>
</tr>
<tr>
<td>AA</td>
<td>3.20</td>
<td>45</td>
<td>0.694</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
needs. For all of the research data management activities we discussed earlier in this chapter (refer to Figure 7-1 and Table 7-4), a greater mean extent of central IT involvement was positively associated with mean agreement about meeting researchers’ needs (Table 7-7). As with all such associations, we’re unable to say which direction the relationship runs, or whether some common external factors might be affecting the results. Still, these findings are generally consistent with previous ECAR studies noting an association between greater central IT engagement and perceived better research support outcomes.22

**Facing the Research Data Management Challenge**

A profoundly important discussion about the need for a new paradigm in research data management has been unfolding for the last decade, but the paradigm itself remains

---

**Table 7-7. Institution Meets Data-Related Needs of Researchers, by Extent Central IT Provides Research Data Management Support Activities**

<table>
<thead>
<tr>
<th>Extent to which central IT provides...</th>
<th>We meet the data-related needs of institutional researchers.</th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research Data Storage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all</td>
<td>2.13</td>
<td>24</td>
<td>0.900</td>
<td></td>
</tr>
<tr>
<td>Very small/small extent</td>
<td>2.84</td>
<td>77</td>
<td>0.947</td>
<td></td>
</tr>
<tr>
<td>Moderate extent</td>
<td>3.14</td>
<td>58</td>
<td>0.945</td>
<td></td>
</tr>
<tr>
<td>Large/very large extent</td>
<td>3.25</td>
<td>118</td>
<td>0.876</td>
<td></td>
</tr>
<tr>
<td><strong>Research Backup/Recovery</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all</td>
<td>2.31</td>
<td>29</td>
<td>0.930</td>
<td></td>
</tr>
<tr>
<td>Very small/small extent</td>
<td>2.75</td>
<td>56</td>
<td>0.939</td>
<td></td>
</tr>
<tr>
<td>Moderate extent</td>
<td>2.96</td>
<td>45</td>
<td>0.952</td>
<td></td>
</tr>
<tr>
<td>Large/very large extent</td>
<td>3.26</td>
<td>148</td>
<td>0.876</td>
<td></td>
</tr>
<tr>
<td><strong>Assistance with Data Management Plans</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all</td>
<td>2.46</td>
<td>52</td>
<td>0.979</td>
<td></td>
</tr>
<tr>
<td>Very small/small extent</td>
<td>2.94</td>
<td>87</td>
<td>0.992</td>
<td></td>
</tr>
<tr>
<td>Moderate extent</td>
<td>3.12</td>
<td>69</td>
<td>0.777</td>
<td></td>
</tr>
<tr>
<td>Large/very large extent</td>
<td>3.41</td>
<td>59</td>
<td>0.893</td>
<td></td>
</tr>
<tr>
<td><strong>Support Metadata Creation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all</td>
<td>2.83</td>
<td>106</td>
<td>1.055</td>
<td></td>
</tr>
<tr>
<td>Very small/small extent</td>
<td>2.91</td>
<td>81</td>
<td>0.840</td>
<td></td>
</tr>
<tr>
<td>Moderate extent</td>
<td>3.27</td>
<td>33</td>
<td>0.911</td>
<td></td>
</tr>
<tr>
<td>Large/very large extent</td>
<td>3.56</td>
<td>34</td>
<td>0.860</td>
<td></td>
</tr>
<tr>
<td><strong>Support Selection/Use of Research Tools</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all</td>
<td>2.66</td>
<td>41</td>
<td>0.938</td>
<td></td>
</tr>
<tr>
<td>Very small/small extent</td>
<td>2.85</td>
<td>114</td>
<td>0.998</td>
<td></td>
</tr>
<tr>
<td>Moderate extent</td>
<td>3.19</td>
<td>62</td>
<td>0.846</td>
<td></td>
</tr>
<tr>
<td>Large/very large extent</td>
<td>3.38</td>
<td>58</td>
<td>0.895</td>
<td></td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
elusive. It seems clear that higher education institutions, including their central IT organizations, will have to be part of the solution. Our look at respondent institutions’ research data management practices, however, suggests that that institutional role remains as inchoate as the larger paradigm.

To be sure, an outline of institutional response is apparent in our findings. The research data problem has many dimensions that require assistance from many quarters, and we did find widely distributed involvement in research data management support activities. Despite its organizational distance from the centers of research on campus, central IT does seem to play a large role, though mainly through traditional activities such as providing storage and backup/recovery services. This role seems to be proportionately larger at teaching-oriented institutions, where researchers are less likely to have generous resources of their own or the assistance of local IT units.

The emergent data management practices most germane to an era of superabundant digital data, such as devising and fulfilling data management plans and creating metadata, also seem to be attracting at least a small amount of central IT attention, and our qualitative interviews suggested that some IT units are acting as services brokers even if they don’t provide (or wish to provide) such services themselves. Just the same, more emergent data management activities, metadata creation in particular, seem somewhat orphaned at our respondent institutions, drawing high percentages of non-involvement or “don’t know” responses. And despite a good deal of forward-looking thought about the role librarians might play in fulfilling these functions, from our respondents’ standpoint, the libraries are the least involved of the entities we asked about.

The critical matter of long-term research data preservation also seems to lack definitive scope and commitment at the institutional level. Fewer than a third of our respondents overall, and only about half at research-oriented institutions, were prepared to say that their institution assumed responsibility for archiving research data, and most of those indicated that it was only in a few or some cases. Of course, other entities might undertake this sort of activity—research institutes and disciplinary organizations, for example—but in many cases these lack the clear long-term institutional viability and continuity that is vital to long-term preservation. At any rate, our respondents averaged less than neutral agreement about the proposition that they have the necessary expertise, policies, funding, and institutional commitment to support long-term data preservation. One possible infrastructure component for data preservation, the institutional digital repository, is still fairly rare and remains mired in questions about scope and functionality. Overall, institutions gave themselves modest marks for meeting the needs of institutional researchers, and doctoral institutions, which have the greatest need, gave themselves the lowest marks of all.

We confess to a good deal of uncertainty about our findings in this chapter. One conspicuous theme running through our results can be summed up in the prosaic words “don’t know”; some of the questions we asked elicited that response from a quarter or more of respondents. CIOs, of course, aren’t researchers, and even at the most eminent institutions, supporting the research enterprise is likely to be only one item in a long list of CIO responsibilities. It’s perfectly understandable that respondents from many central IT units would have little opportunity to get involved in such matters as research data preservation, ownership, and policy formation. Maybe all this is just somebody else’s job.

We’d be more inclined to leave it at that if we were more confident that both the job and the somebody exist. At any rate, a strong case can be made for the proposition that
to some extent, such matters do belong in the CIO’s admittedly crowded job description. Central IT units, our respondents tell us, frequently provide infrastructure and services for institutional researchers. It seems reasonable that the entity that enables research data storage should acquaint itself with the policies and services needed to give that data a stable long-term home, and indeed, where we find central IT more involved, we also find that institutions more strongly agree that the data-related needs of researchers are being met. Furthermore, no one on campus has more experience than central IT in data center administration, high-reliability applications, and large-scale data administration in an on-demand environment. At the very least, central IT can be an advocate for ensuring that this sort of expertise exists somewhere on campus, and as the research data management capability develops, central IT can and should be a partner in ensuring its success.

Endnotes

5. Though the 2009 results indicate a slightly lower rate of past-three-years increase than the 2005 results, the 2009 survey specified “amount of research data storage,” whereas the earlier survey specified simply “amount of data storage.” See Harvey Blustain et al., IT Engagement in Research: A Baseline Study (Research Study, Vol. 5) (Boulder, CO: EDUCAUSE Center for Applied Research, 2006), available from http://www.educause.edu/ecar.
8. Ibid., 50–51.
9. The medians in Table 7-4 refer to the original six levels of involvement extent we asked about, rather than the combined levels shown in the four columns in the left of the table.
17. These findings exclude those who answered “don’t know” to the local repositories question.
19. Ibid., 11.
20. Among respondents, 35.1% said investigators conducted contract research for third parties at their
institutions, 38.5% said no, and 26.4% said they didn’t know.

21. Of 145 respondents, 24.1% said their institution’s policy permitted third-party ownership, 15.9% said it didn’t, and 60.0% said they didn’t know.

22. Sheehan, Higher Education IT and Cyberinfrastructure, especially p. 87; more limited associations are reported in Blustain et al., IT Engagement in Research, 78–79.
Data Management Outcomes

However beautiful the strategy, you should occasionally look at the results.
—Winston Churchill

Key Findings

- When assessing data management outcomes, respondents were much more positive about the security of data from unauthorized access than they were about their institution’s ability to get maximum academic and business value from data.
- A slight majority of respondents agreed that their institutions can support growth in the volume of data over the next three years; the mean response was 3.32 on a scale of 1 (strongly disagree) to 5 (strongly agree).
- Among the factors positively associated with stronger agreement about all five of our outcomes measures were higher enterprise data quality scores, an early-adopter approach to technology, and stronger agreement that the institution is committed to long-term preservation of research data. Stronger agreement that the institution provides effective training in analytics tools was positively associated with better results in four of five outcomes measures.
- Agreement that the institution gets maximum academic and business value from institutional data was higher among institutions that reported frequent use of advanced analytics techniques and where respondents agreed more strongly that institutional leadership was committed to evidence-based decision making.

As we’ve worked on this project, it hasn’t been lost on ECAR that the study mirrors some aspects of data management itself: lots of data; a lot of separate and discrete yet interacting subtopics; the challenge of finding patterns in a topic of potentially boundless scope. We think it’s particularly important to consider how (and whether) the many parts of the data management domain work together and to ask how well institutional data practices serve major institutional needs. With an eye toward identifying practices associated with data management success and tying together material covered in earlier chapters, in this chapter we examine how well respondents think their institutions have achieved key data-management-related outcomes.

We begin with a top-level look at how respondents assess their institutions’ performance in five outcomes measures. Next, we
examines factors associated with good performance across most or all outcomes, that is, practices that seem to be common to broadly successful data management environments. Finally, we look at some additional success factors related to particular outcomes.

Assessing Data Management Outcomes: An Overview

We asked respondents to state their level of agreement (on a 5-point scale from strongly disagree to strongly agree) with the following five statements that all begin “At my institution...”:

- restricted/sensitive data is secure from unauthorized access,
- we can support anticipated growth in the volume of data over the next three years,
- employees understand their responsibilities in the use of data,
- we get maximum academic value from institutional data, and
- we get maximum business value from institutional data.

Average responses appear in Table 8-1, and Figure 8-1 shows the distribution of responses across agreement levels.

Perhaps the most common themes in discussions of data management in recent years have been the dangers of security breaches in an age of privacy regulation and an impending “deluge” of data growth. Yet security of restricted or sensitive data from unauthorized access was the outcomes item that our respondents were most positive about; three-quarters of them (74.7%) agreed or strongly agreed with our security statement, and the mean response of 3.87 was the only one of the five to approach an average agree (4). Respondents were more muted about their institutions’ ability to support anticipated data growth over the next three years, but still a majority (53.3%) agreed or strongly agreed with the statement, and the mean response of 3.32 was well above neutral (3). This relative unconcern about the volume of data makes an interesting comparison with another item, reported in Chapter 6, about data variety: Respondents averaged only a response of 2.81 (short of neutral) about their institutions’ ability to effectively manage all the varieties of data and digital content that the institutions needed (refer to Table 6-3).

Relatively high agreement about what might be called data management’s poster concerns stood in contrast to lower agreement about more mundane and “ordinary” issues. Respondents averaged only slightly above neutral (mean 3.14) in agreement that employees understand their responsibilities in the use of data, and they fell below neutral in agreement that their institutions got maximum

<table>
<thead>
<tr>
<th>At my institution...</th>
<th>Mean*</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted/sensitive data is secure from unauthorized access.</td>
<td>3.87</td>
<td>305</td>
<td>0.770</td>
</tr>
<tr>
<td>We can support anticipated growth in the volume of data over the next three years.</td>
<td>3.32</td>
<td>304</td>
<td>1.053</td>
</tr>
<tr>
<td>Employees understand their responsibilities in the use of data.</td>
<td>3.14</td>
<td>304</td>
<td>0.898</td>
</tr>
<tr>
<td>We get maximum academic value from institutional data.</td>
<td>2.72</td>
<td>303</td>
<td>0.936</td>
</tr>
<tr>
<td>We get maximum business value from institutional data.</td>
<td>2.67</td>
<td>304</td>
<td>0.963</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
academic value (2.72) or business value (2.67) from institutional data. Nearly half of respondents disagreed or strongly disagreed with each of the value statements.

Our academic and business value statements set a high standard by asking about “maximum” value from institutional data, and this may be part of the reason why these two vital items stood at the bottom of our outcomes measures. On the other hand, getting value from data is what data management is all about, and about one in four institutions did agree that they met the maximum-value standard. We’re inclined to believe that our respondents view the performance of their institutions in these areas, and to a lesser degree in the area of employee understanding of responsibilities in the use of data, as more in need of improvement than the more often acknowledged areas of security and data growth.

The attitudes expressed in these results seem to be widespread. We found no significant differences between respondents on the basis of Carnegie class, FTE enrollment size, public/private control, or research/teaching mission. But that’s not to say that we didn’t find factors differentiating institutional responses, and in the remainder of this chapter, we examine practices or circumstances that distinguished higher from lower agreement about performance on these outcomes. This look at associations between outcomes and other variables isn’t exhaustive,
Institutional Data Management

but it does focus on the strongest and/or broadest relationships we found.

Outcomes Factors: Broad Impact

Data management is a big topic, and we chose our outcomes measures in order to look for factors associated with success across its multiple dimensions. In this section, we analyze practices that were significantly associated with at least four of the five outcomes items mentioned above. Although the existence of statistically significant associations doesn’t prove a causal relationship (influence might flow in either direction or be related to other factors we didn’t ask about), we hope these findings will suggest places where IT administrators might begin when they consider how to improve institutional data management performance.

For the sake of clarity, we’ve expressed some of the findings that involve mean values in the form of bar charts, without the full statistical information we usually include with means. See Appendix C for the complete information related to these figures.

Enterprise Data Quality

One of the most striking themes in this study is a frequent and often strong association between perceptions of enterprise data quality and other desirable data management practices and outcomes. In Chapter 4, we reported results for eight survey questions relating to aspects of enterprise data quality, such as whether major data elements have a system of record and a definition recognized across the institution, whether changes to data elements propagate across all systems using them, and whether data quality processes are in place (refer to Table 4-1 and the related text). We also derived an enterprise data quality score for each respondent institution, calculated as the mean of the responses to all eight enterprise data quality questions.

Enterprise data quality scores proved to be positively associated with all five of the data management outcomes we asked about (see Figure 8-2). The spread between low and high scores was especially notable for the academic and business value outcomes, where high-scoring respondents averaged a point or more higher agreement with each outcome than did low-scoring respondents.

It’s not hard to see how a strong foundation in data quality could be pervasively related to good data practices across the institution, and ultimately to better success in achieving institutional goals. As we reported in Chapter 4, higher data quality scores were associated with more positive respondent assessments about training, leadership and staff understanding of data issues, and active cooperation between IT and business/academic units to ensure data quality (refer to Table 4-3). Trustworthy and consistently formatted data should be easier data to analyze and share for both academic and business purposes, and a widespread grasp of data handling issues should serve to improve both employees’ understanding of their responsibilities and their awareness of security concerns.

That’s the good news. The bad news is that institutions on the whole aren’t enthusiastic about their enterprise data quality practices; as we reported in Chapter 4, the mean score for all respondents was a mediocre 3.06 on a scale of 1 (low) to 5 (high). Though data quality is a highly traditional and far from sexy area of practice, it may be the best place to start as institutions look for ways to improve their data management game.

Aside from the enterprise data quality score itself, we also found that most of our outcomes items were strongly and positively related to items concerning data quality training and understanding. Agreement with outcomes items was stronger where respondents more strongly agreed that:
business/academic unit leaders receive training in data quality issues (all outcomes except the data security item);

- business unit leaders, executive leaders, and staff understand data quality issues (all outcomes except supporting anticipated growth in volume of data); and
- central IT works actively with business/academic units to ensure data quality (again, all items except growth in volume of data).

All of these bulleted data quality training and understanding items were, in turn, positively associated with higher enterprise data quality scores.

### Analytics Training

We noted in Chapter 4 that most institutions have a fairly modest infrastructure to support analytics tools (or, as we referred to them in our survey questions, tools for reporting, modeling, analysis, and decision support) and that advanced uses of analytics tools tend to be selective, whereas the primary use most often was extraction and reporting of transaction-level data. As we report later in this chapter, we found a strong association between two of our outcomes (getting maximum academic and business value from institutional data) and frequent use of advanced analytics.

The aspect of analytics most broadly and strongly associated with data management outcomes, however, was respondent agreement that their institution provided effective training to users of reporting, modeling, analysis, and decision support tools (see Figure 8-3). Our data security item was the only one of the outcomes we asked about that was not...
significantly higher at institutions agreeing more strongly about effective training in analytics tools.

Findings of strong relationships between training and good outcomes are common fare in ECAR studies, even where, as in this case, the subject of the training is associated with the outcomes in a narrower way (or not at all). We suspect that in many such cases, training is a marker of a stronger commitment, a more formal approach, or a better-funded initiative, in addition to making its own contribution by enhancing staff skills. We speculate that something like this may be reflected in the current findings. At any rate, institutions that do a better job training users in the use of analytics tools may be spreading good data use habits while they enable users to get more value from that data. And, looking at it from another direction, a willingness to invest in analytics training may flow from greater confidence in enterprise data overall; institutions that agreed more strongly about analytics training effectiveness also tended to have higher enterprise data quality scores.

Data and the Organization

We found several relationships between aspects of respondent institutions’ organizational climate and data management outcomes. As is often the case in ECAR studies, one of these related to the institution’s approach to adopting new technologies. As Figure 8-4 shows, early adopters reported higher mean agreement across all five major outcomes measures than did mainstream or late adopters. The difference was dramatic for most outcomes; the mean difference in getting maximum academic and business values exceeded a full point on our 5-point agreement scale.
Such differences could be related to strictly technological advantages. Early adopters, for example, may have more advanced IT architectures that facilitate good data management practices. But they may gain advantages as well from a leadership environment that is friendly to technological initiative. Early adopters also are more likely to agree that improved data management is a strategic priority at their institutions and that leadership is committed to evidence-based decision making.

Among other organizational factors, institutions with a central IT group for data (for example, data architects) averaged stronger agreement with all our outcomes measures except supporting anticipated growth in the volume of data, and institutions reporting a dynamic organizational climate (where change is continuous, orderly, planned, and navigable) agreed more strongly on all outcomes than did institutions reporting stable, volatile, or turbulent organizational climates.

Research Data Preservation and Records Retention/Disposition Compliance

In Chapter 7, we discussed the challenges of research data preservation as science and scholarship increasingly turn to digital data-intensive techniques. Though a compelling case can be made for the long-term preservation (and thus the sharing) of research data, our respondents were not especially positive in assessing their institutions’ support for it. Overall, their mean agreement that their institutions had the necessary commitment to support long-term research data preservation was midway between disagree and neutral (refer to Table 7-5).
Perhaps the very complexity of this challenge, with its mix of technological, cultural, and regulatory issues, helps explain why institutions that assess this institutional commitment more positively also tend to do better on all of our data management outcomes (see Figure 8-5). Like many aspects of data management, creating an institutional commitment to long-term data preservation is likely to demand an enhanced awareness of data issues, the cooperation of stovepiped units and mobilization of expertise distributed among them, and a supporting framework of policy and protocols. It’s also possible that activities set in motion to address research data preservation could have spillover effects on other data-related areas, and vice versa.

Something similar might be said about records management, and not surprisingly we found an association as well between how respondents rate their institutions’ ability to comply with records retention and disposition requirements throughout the institution and how strongly they agree with our outcomes items; across the board, higher records retention compliance was associated with better outcomes. To choose only one example, respondents who rated records retention compliance poor or fair averaged a 2.45 response to our item about getting maximum business value from institutional data, whereas those rating compliance very good or excellent averaged a 2.95 response.

**Outcomes Factors: Targeted Impact**

The broad-impact factors we considered above give some clues about how IT administrators might pursue initiatives that can have an impact on data management across multiple outcomes of interest. But each of our outcomes is important in its own right, and each has

![Figure 8-5. Data Management Outcomes, by Agreement That Institution Has Necessary Commitment to Support Long-Term Preservation of Research Data](image)

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree*
significant associations of its own. In this section we look at factors that were particularly strongly related to specific outcomes.

**Keeping Restricted Data Secure**

The strongest associations we found relating to higher agreement that restricted/sensitive data is secure from unauthorized access are, for the most part, already broadly accounted for in the section above. Factors related to data quality, such as consistent coding of major enterprise data elements, cooperation between central IT and business/academic units, and training executives and staff in data quality issues, were prominent among the variables positively associated with agreement about data security. So were a number of variables connected to the long-term preservation of data, including agreement about institutional commitment to, and funding mechanisms for, such preservation.

In addition, we found stronger mean agreement that restricted/sensitive data is secure from unauthorized access among institutions that reported certain security-related policies, including policies for defining individual responsibilities for data security practices and for the acceptable use of institutional data.

However, several other practices we thought might be associated with the data security item, such as having a documented institutional policy defining data steward responsibilities or a formal process for classifying institutional data by risk or sensitivity level, turned out not to be significantly associated.

**Employee Understanding of Responsibilities in Use of Data**

Besides the associations reported in the broad-impact factors section above, the policy environment at respondents’ institutions seems to be a factor in their perception of employees’ understanding of their responsibilities in the use of data. As Table 8-2 shows, agreement about this outcome was considerably stronger where institutions have a policy defining individual employee responsibilities for data security practices and a policy for acceptable use of institutional data, compared with institutions with no policy or one in development.

We also found some evidence that employee understanding of data responsibilities is related to the institutional content.

<table>
<thead>
<tr>
<th>Institution has a documented policy for...</th>
<th>Employees understand their responsibilities in the use of data.</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean*</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Defining Individual Employee Responsibilities for Data Security Practices</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No policy</td>
<td>2.75</td>
<td>61</td>
<td>0.925</td>
</tr>
<tr>
<td>Policy under development</td>
<td>2.89</td>
<td>61</td>
<td>0.839</td>
</tr>
<tr>
<td>Has policy</td>
<td>3.35</td>
<td>179</td>
<td>0.851</td>
</tr>
<tr>
<td><strong>Acceptable Use of Institutional Data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No policy</td>
<td>2.50</td>
<td>40</td>
<td>0.906</td>
</tr>
<tr>
<td>Policy under development</td>
<td>2.88</td>
<td>41</td>
<td>0.781</td>
</tr>
<tr>
<td>Has policy</td>
<td>3.31</td>
<td>222</td>
<td>0.854</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree*
management environment. Institutions reporting integrated enterprise content management solutions had the highest agreement regarding this outcome among the range of content management environments we asked about, while those with only local or ad hoc content management had the lowest. Those with best-of-breed enterprise solutions or a mix of enterprise best-of-breed and local/departmental solutions varying by content type were in between.

Supporting Anticipated Growth in the Volume of Data
With an overall mean agreement of 3.32, this was one of the outcomes items respondents seemed to be most confident about. Other than the factors reported in the broad-impact section above, we found that some of the strongest associations relating to agreement about the institution’s ability to support anticipated growth in data volume in the next three years had to do with respondents’ confidence about their institutions’ current data infrastructure. For example, where respondents more strongly agreed that their institution currently has the infrastructure it needs to effectively manage such challenging and fast-growing data types as online video/audio content, research data, and e-mail, they also agreed more strongly about supporting anticipated growth. Respondent agreement about this outcome was also strongly associated with agreement about multiple items relating to supporting the long-term preservation of research data, including having the necessary infrastructure, funding, policies and procedures, and expertise.

Getting Maximum Value from Institutional Data
Perhaps the most critical outcomes we asked about were our two items concerning getting maximum academic and business value from institutional data. Though they undoubtedly involve some subjective judgment about where value lies, these items get at the heart of what data management is ultimately all about: helping the institution use its data to accomplish its goals. The below-neutral mean answers these items elicited suggest that our respondents don’t think their institutions are getting “maximum” value or even something close to it.

These outcomes were strongly associated with the data quality and analytics training practices we mentioned in the broad-impact section above. As we drilled down in our analysis of the two items, we found additional factors consistent with this pattern. Of particular interest was a relationship with advanced analytics practices, shown in Table 8-3. In Chapter 4 (refer to Figure 4-11), we reported on use of six advanced analytics techniques and noted that more frequent use of some of them was associated with stronger agreement that institutional leadership was committed to evidence-based decision making.

In addition, more frequent use of all six of these advanced analytics items was associated with stronger agreement about getting maximum academic and business value from institutional data. Table 8-3 summarizes this relationship by showing that the more of the six named techniques institutions said they usually or almost always use, the stronger their agreement about getting maximum academic and business value tended to be. Institutions that didn’t report usually or almost always using any of the techniques averaged a dismal 2.27 response on the business value outcome—not much above a mean disagree—whereas those reporting five or six averaged an above-neutral 3.33. The pattern was much the same for the academic value item.

Our point here is not to encourage the use of some particular advanced analytics technique, but rather to note that more frequent across-the-board use, without regard to specific combinations of techniques, is broadly associated with stronger
agreement about getting value from institutional data. Since our results in Chapter 4 suggest an underdeveloped analytics infrastructure overall and relatively infrequent use of many of the advanced techniques referred to here, we interpret these results as an indication that institutions could profit from more aggressive use of (and presumably greater investments in) advanced analytics techniques.

It’s a commonplace of analytics best practice that culture and executive support are every bit as important to analytics success as infrastructure and tools. We found some support for this proposition as well; as Table 8-4 shows, where respondents more strongly agreed that their institutional leadership is committed to evidence-based decision making and that their institution is committed to making management information broadly available, they also agreed more strongly with the academic and business value outcomes items. Combined with the strong relationship between analytics training and data management outcomes that we reported in Figure 8-3, these results, we believe, add still further to the case for a broad institutional commitment to analytics as a way of getting better value from data.

### Toward Better Data Management

Two major themes stand out from our data management outcomes results. First, the complex of things that add up to enterprise data quality—including not just specific data quality practices but also training in data quality and executive, management, and staff understanding of the issues—seems to have a broad impact on the way our respondents perceive data management outcomes at their institutions. Even more than in the piecemeal results reported in earlier chapters, data quality emerges from our outcomes analysis as the foundation of data management success. Our results don’t tell us about the exact mechanisms involved—no doubt there are many—but it isn’t hard to think of ways

---

**Table 8-3. Institution Gets Maximum Academic and Business Value from Data, by Count of Advanced Analytics Practices Used**

<table>
<thead>
<tr>
<th>Count of Advanced Analytics Practices Usually/Almost Always Used</th>
<th>We get maximum academic value from institutional data.</th>
<th>We get maximum business value from institutional data.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean* 2.26</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>N 68</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation 0.745</td>
<td>0.827</td>
</tr>
<tr>
<td>1–2</td>
<td>Mean* 2.70</td>
<td>2.64</td>
</tr>
<tr>
<td></td>
<td>N 148</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation 0.929</td>
<td>0.938</td>
</tr>
<tr>
<td>3–4</td>
<td>Mean* 3.05</td>
<td>2.90</td>
</tr>
<tr>
<td></td>
<td>N 60</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation 0.946</td>
<td>0.961</td>
</tr>
<tr>
<td>5–6</td>
<td>Mean* 3.26</td>
<td>3.33</td>
</tr>
<tr>
<td></td>
<td>N 27</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation 0.859</td>
<td>0.961</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
in which better data quality practices can stimulate, and in turn benefit from, better data management throughout the institution. Administrators who aren’t constantly trying to fix or work around faulty data can put more effort into more advanced data use and can trust the data they have when they’re making decisions.

This leads to the second major theme: the relationship between analytics training and use, and data management outcomes. Respondents’ agreement that their institutions provide effective training to users of analytics tools proved to be one of the broadest and strongest factors in good outcomes results. In addition, frequent use of advanced analytics techniques was strongly associated with our two key outcomes, getting maximum academic and business value from institutional data. As with data quality, we’re inclined to see analytics as a broad marker of a pervasive institutional determination to overcome data stovepiping and localization and to liberate data for higher and better uses than its immediate transactional purposes.
To these major themes we might add that two different ways of answering the question “what do we do with all this data?” are reflected in our outcomes results. Institutions didn’t give themselves very high marks either in their ability to comply with records retention and disposition requirements or in their support for the long-term preservation of research data. Yet where they thought they’re doing better with these difficult challenges, they also more strongly agreed with our outcomes statements across the board. These factors, too, might be more related to cooperation and communication and an institutional willingness to address matters whose long-term importance exceeds their short-term urgency than they are to infrastructure and specific practices.

We confess that our outcomes assessments surprised us in a number of ways. Certain factors prominent in the data management literature, including data stewardship policies, enterprise content management solutions, and the existence of a designated archivist or a group responsible for electronic records management, played either no role in our outcomes results or a smaller role than we expected. We were also somewhat surprised to discover that, on the whole, respondents seem to be less concerned about the security of their data than about employee understanding of responsibilities in data use and about getting value from data. They also seem less concerned about the growing volume of data than about how to manage its myriad varieties. It may be time for higher education IT to broaden its concerns beyond the regulatory- and fear-driven issues of the past decade and to focus on foundational data quality and analytics issues that, like a rising tide, could lift all our data boats.

**Endnote**

1. N = 160, SD = 0.903, and N = 41, SD = 0.921, respectively.
Institutional Data Management in an Age of Superabundant Information

The most incomprehensible thing about the world is that it is comprehensible.
—Albert Einstein

Where are we going to put it all? one might wonder. If, as the IT advisory firm IDC asserts, we created or replicated 281 billion gigabytes of digital data in 2007, and it’s growing at 60% per year, today that’s... We leave the answer to the student as an exercise. (Hint: a lot.)

True, there is some reason to think that the smooth progression of memory densities we’ve become accustomed to will soon hit a bump in the road. Gordon Moore himself has repeatedly predicted the approaching end of his famous law, and some analysts have suggested that the physical and/or economic limits of semiconductor fabrication will be reached at around half the current production feature resolution of 32 nanometers.1

Yet similar fears have proved to be unfounded before, and we continue to be extremely clever about mapping bits onto ever smaller physical entities. Hard disk technology, currently capable of a density of about 400 gigabytes per square inch, is expected to reach 2.4 terabytes per square inch by 2014.2 Magnetic-based technologies promise denser, faster, less power-hungry alternatives to electronic RAM, and three-dimensional geometries are being developed for greater density in both silicon-based and optical storage technologies. Among more exotic technologies, carbon nanotubes with walls an atom thick are being called into storage service. Progress is being made on the long-term storage problem as well. Recently a team of Berkeley researchers demonstrated a nanotube-based electromechanical device that, besides achieving ultrahigh density, can store bits at room temperature without degradation for more than a billion years. “This system,” the team drily noted, “has the potential to store information stably for any practical archival time scale.”3

The more pressing problem isn’t where, but how, we’re going to store it all—and find it all, and use the parts we need. No theme has been more prominent in recent writing about cyberculture than the interaction of masses of Internet users with masses of data, interactions that themselves generate copious amounts of data and metadata. The prevailing theme in this literature is to stop worrying, because scale generates its own solutions. “Ontology is overrated,” writes the Web 2.0 analyst and NYU professor of communications Clay Shirky; linking and tagging give information users “much more organic ways of organizing information than our current categorization schemes allow.”4 Shirky argues that we should “publish, then filter” because the information signal can be separated from noise only within some context.5 Philosopher-marketeer David Weinberger similarly writes...
of a “third order of order” created when users begin to see all the connections made visible by de-physicalizing and virtualizing information. “Information doesn’t just want to be free,” Weinberger writes, “it wants to be miscellaneous.” Writers James Surowiecki and Cass Sunstein draw on social science research suggesting that collective wisdom often outperforms individual expertise in all manner of things, including structuring information, and point out that the social web has introduced new ways of harnessing this almost magical power.

Information superabundance is one of the reasons that concepts such as tipping points, black swans, and the wisdom of crowds have become not just topics for unlikely bestsellers but markers of the zeitgeist. These ideas—leveraging older ones, from the invisible hand to chaos theory—help us make sense of phenomena that were far less palatable when “mass communication” was a one-way process and the richest information source in the average middle-class household was a shelf of lightly used encyclopedias. Instant communication and the ready discoverability of content mean that today cultural memes can “go viral” with astonishing speed; masses of data feed the policy wonk and the conspiracy theorist alike; and everyone can fact-check everyone else—and does, in millions of blogs and comm boxes every day. Even as they celebrate the breakdown of traditional ontologies and ways of ordering information, the new cyberculture pundits promise a deeper, better, folksier order, emerging with a kind of probabilistic grandeur from apparent chaos. It’s a profoundly appealing idea to the culture-war-weary citizen and the hip, edgy blogger alike.

But as is often the case in grand social theory, these ideas have a hard time accommodating institutions. Homo cyberneticus may not be the rational utility maximizer that homo economicus is—there are no trolls in The Wealth of Nations—but he or she is usually conceptualized as a free agent pursuing individual interests even while seeking out like-minded people to bond with. Although the new cybercultural ideology includes a thick streak of utopian communitarianism, grounded in a belief in what Shirky calls “our strong talents and desires for group effort,” like so many previous utopias its vision is highly aversive to subordination and bureaucracy. This cyberworld is communal precisely because it’s anti-institutional; its advocates want to uproot creative talent from the sterile reign of management and plant it in the rich soil of self-directed “social production.” If institutions from newspapers to movie studios fear the result, it’s because they don’t (and can’t) get it. Instead, they fight costly rearguard actions in the courts and confront the “innovator’s dilemma” with paralytic denial.

One can almost hear the collective sigh of the officers who are charged with administering the information needs of higher education institutions. Colleges and universities are situated as deeply in cyberculture as any entities around—they invented a good bit of it, after all—yet they remain stubbornly institutional, and they are subject to the direction of that still-more-stubborn institution known as the government. Many legislators and regulators remain unmoved by the idea that information wants to be either free or miscellaneous, and IT administrators are caught between the two fires of open access and state regulation. Meanwhile, institutions, unlike infotopias, must formally account for the resources they expend, support non-virtualized activities and the infrastructures that make them possible, and organize themselves for long-term survival.

But it would be a mistake to wave away the new thinking about cyberculture and information superabundance just because its proponents aren’t very interested in institutional dynamics. The truth is that the special role that colleges and universities have historically played in providing lots of sophisticated
information to society at large is eroding, while new modes of personal discovery and interaction are being created. Higher education institutions have a long history of building satisfying environments for social interplay and discovery, and despite all the innovations of the Internet era, millions of students still compete fiercely to take advantage of those environments. By drawing on their heritage of parsing new information environments and mastering them, higher education institutions can move beyond the relative decline in their roles as repositories of information and take the lead in showing students and society at large how to navigate the ocean of data.

That will mean, however, a big and complex process of rethinking the institution’s information culture. No longer society’s prime solution to the problem of information scarcity, the university will have to evolve new solutions to the problems (and ways to exploit the opportunities) of information superabundance. We believe that the topics we’ve investigated in this study will play a major role in higher education’s reconciliation with an emerging information environment that is in some respects more consumerized and commodified, in other respects more professionalized and customized. We look here at three dimensions of the institutional information challenge that are rarely considered together: institutional data and its uses in administration and leadership, the institutional content environment, and the special problems that attend the production and preservation of research data.

**Data Quality**

One of the effects of information superabundance is that expectations about the availability and usability of data rise. The characteristic solution to many economic and social problems these days is to free up the flow of data in order to make markets more efficient and to force firms—and institutions—to compete more aggressively. This idea was central to the reforms of the Spellings Commission report of 2006, which complained of “a remarkable shortage of clear accessible information about crucial aspects of American colleges and universities” and recommended collecting student-level data at the national level that would permit better tracking of educational outcomes and make it easier for students to shop for courses and educational programs. Similarly, investment in electronic medical records and other information tools is prominent in the current discussion of U.S. health care reform. Wherever the notion of accountability is raised, some kind of information reform is likely to lie underneath it, and in this context institutional data is no longer merely an institutional concern.

Nor are these ideas being promoted only by policymakers. The large investments in ERP systems that institutions made in the Y2K era were often justified by claims that they would feed a “transformation” of higher education, allowing institutions to be data- rather than tradition-driven organizations. These hopes were rarely realized. In part this was because transactional systems, for all their advantages over earlier generations of mainframe systems, were far better at turning batch processes into real-time processes than they were at assembling an aggregate view of institutional performance. Turning transactional data into actionable intelligence requires an information infrastructure that, as Chapter 4 of this study shows, relatively few institutions have pursued aggressively. But it is hard to picture any sort of progress on the major issues confronting higher education today—recruiting and enrollment in a competitive atmosphere, student retention and success, research productivity, and above all gaining control over the cost of education—that doesn’t flow from a close and sophisticated scrutiny of trustworthy operational data suitable to the task.

It’s also likely that technological pressures will put new stresses on the institutional data environment. Chief among the most discussed...
Institutional Data Management

Technologies of the day are service-oriented architecture (SOA) and the distributed, services-based mode of computing known as the cloud. Today’s monolithic business applications often suffer from Rube Goldberg-ish data complexity, thanks to legacy survivals and the persistence of code lines across evolving technologies and architectures, but they can make up for these eccentricities to some degree through applications logic and suite-specific tools. The major data pain points for institutions arise when it’s necessary to stitch together applications or platforms.

But in the abstracted and modularized world of SOA, applications are decomposed into much smaller services that can be more flexibly combined; integration isn’t an exception condition but the basis on which systems are assembled, and data must pass across many more seams. Consistent and well-documented data models will be essential in this environment. Cloud computing will add to the pressure in two ways: by introducing new options for data storage, and by balkanizing user applications through software-as-a-service solutions. Whether adopted as enterprise solutions in the manner of salesforce.com’s CRM products or as user-chosen tools with no enterprise visibility, cloud solutions by their nature imply a proliferation of data models, greater movement of data, and the declining ability to tolerate eccentricities and errors.

All of this suggests that for CIOs managing institutional data, it’s not ontology but the death of ontology that’s overrated. Superabundance won’t mean the end of structured data; on the contrary, it will require far better solutions to age-old problems of error, inconsistency, integration, and poor metadata and semantics management. The results of our study are clear: Long-understood best practices in the handling of structured data, such as identifying a system of record, maintaining consistent definitions and coding schemes, and creating processes to document and resolve data problems, are associated with better outcomes in a wide range of data-management-related activities. Most institutions that carry out data quality improvement initiatives, furthermore, don’t seem to think that they’re just playing a fruitless game of whack-a-mole; almost 7 in 10 agree that their initiatives result in process changes that improve data quality throughout the data life cycle.

Yet despite this positive evaluation of data quality initiatives, it’s also important to remember that though nearly 80% of our respondent institutions said that they had an initiative under way, had completed one within 24 months, or planned one, respondents overall described a pretty lackluster enterprise data quality environment, averaging an almost exactly midscale 3.06 enterprise data quality score. We suspect that though institutions are sophisticated enough to know that data quality problems need systemic solutions rather than mere patches—hence the optimism about process improvements—they are conducting narrowly focused initiatives in a basically reactive or at best evolutionary mode rather than thinking about data quality at an enterprise-wide and architectural level. We also can’t help but suspect that despite fairly high agreement that institutional leadership is committed to evidence-based decision making, limited investments in analytics infrastructure and spotty use of advanced analytics techniques hinder systemic data quality improvement. Where data are exercised more strenuously, their imperfections are more likely to be noted and corrected, and the feedback loop is extended beyond operational needs to management and leadership areas, where data is scrutinized for its strategic value.

And it is, in fact, a strategic view of data that justifies investing in the familiar yet easily neglected area of data quality. It seems unlikely that books singing the praises of enterprise data quality will top the bestseller list any time soon, but nothing we found in our study seemed to have more potential for
stimulating pervasive improvements in institutional information environments, nor did we find anything that so clearly falls within the historic competencies and likely future responsibilities of IT professionals. Ten years after enthusiastic talk of transforming higher education through IT, an almost perfectly opposite view has taken shape in claims that IT is becoming a mere utility incapable of differentiation and empty of strategic value. This idea focuses heavily on the ostensibly declining need for on-site hardware administration and software deployment; it has far less to say about data, with its essentially non-commodifiable operational uses and its potential as performance metric, strategic guide, and product. As ECAR Fellow Philip Goldstein has argued, the CIO role in an era of utility computing could evolve more clearly into that of a data evangelist and services architect, providing institutions with expertise and advocacy in the management of data and acting as “the bridge between the technology, the data, and the decision makers within the institution.”

But if decades of handling institutional data have brought us to no better quality state than the one we occupy today, how will we improve our practice? Technology may give us a hand, though probably a pretty modest one, at least in the near term. Data services developed under the broader category of SOA may help us with the transformations necessary to deliver data from one storage mode or application to another seamlessly. More investment in metadata repositories and in data quality tools that help identify conflicts and inconsistencies, and that continuously filter and monitor data, could take some of the burden off institutional staff and create a better data foundation for analytics tools.

It seems likely, however, that improved process and collective action will do more to improve institutional data quality than technology. For one thing, though bustling with ideas, the data management technology tools market is fairly immature overall. Gartner Inc. regards data quality tools as a relatively mature technology, but it is one of only 4 out of 25 items tracked on the company’s 2009 data management “hype cycle” that have passed beyond the critical “trough of disillusionment.”

More important, as this study has shown, responsibilities for handling and using data are distributed throughout the institution, and awareness of data issues and commitment to quality must be similarly distributed. To some extent, departments everywhere need to be inculcated in the notion that theirs is an information-dependent business and that the uses they make of data locally are only part of a larger matrix of institutional value. Although executive leadership and business/academic unit management will be crucial to getting the data quality message out, no center on campus has a more comprehensive view of the issues involved, or a better grasp of how the technology interacts with institutional business, than central IT. We noted in Chapter 4 that where respondents agreed more strongly that central IT worked actively with business/academic units on data quality issues, perceived enterprise data quality was substantially higher. We doubt that any data management technology in the pipeline has more to contribute to getting better value from institutional data than such interaction.

Institutional Content Management

There’s probably no area of institutional data management where IT administrators feel the pressures of superabundance more than in the huge and loosely bounded domain known as content. It’s hard even to account for all the ways the institutionally owned content heap gets added to: web pages and blogs, learning objects, forms and records, e-mails and collaboration threads, photographs, videos, podcasts, and in principle every word processing file, form, and
spreadsheet created throughout the institution. Add to this the growing amount of content the institution pays to access and the universe of extra-institutional material that local IT units must in some way facilitate, from search to optimizing networks to setting up federated identity access, and it can be hard to believe that IT administrators do anything but manage content.

It’s this huge amorphous nebula of information that the new cyberculture thinking takes as its chief subject, and it’s here that the “free and miscellaneous” ideology is most relevant. The web is a big place, and colleges and universities will necessarily be spectators to much of what’s going on there, though it’s hard not to consider what might have been had higher education been more agile and visionary about extending its special concerns into cyberspace. It was not a given, for example, that course evaluation sites such as RateMyProfessors.com would have developed entirely outside institutional higher education or that the driving force in creating a global virtualized collection of books should have been a search engine owner rather than the owners of the world’s best physical book collections. Impressive as some academic content initiatives have been, their largely scattershot nature and their tendency to lack a sustaining economic basis once grant funds have run out have kept higher education at best a secondary player in the content revolution that’s transforming the world’s information landscape.

From the standpoint of institutional IT administrators, however, one capability that lies within reach is to draw into the institutional web environment some of the resources and styles of interaction that have helped add information manageability and sociability to the larger web environment. Looking at ways of enhancing their web environments over the next three years, our survey respondents were most concerned with video streaming/download and content for handheld devices—and rightly so, since in the post-YouTube era video is emerging as a medium that masses of people create as well as consume, and mobile devices have for all intents and purposes become pocket computers. But the further stimulation these drivers will add to the overall explosion of content is all the more reason why institutions might give greater consideration to resources that were less prominent in our web-enhancement results. In particular, relatively few institutions seemed to be giving high-priority consideration to Web 2.0–flavored interactivity enhancements such as wikis, social tagging, and Epinions-like ratings systems, despite the huge popularity of social networking among students. The “social commerce” interactivity that has revolutionized retail marketing on the web at the expense of formerly sacred marketing principles such as message control is rarely to be found in institutional websites, at least in their “official” precincts. To the extent this attempts to preserve institutional control and protect interested constituencies, it’s futile; it is simply another invitation to external entities to disintermediate the institution while sacrificing the valuable information and constituent attention such tools can generate.

This is not to say that institutions can or should adopt a “free and miscellaneous” attitude toward all content or all the interactions content engenders. One of the key responsibilities for institutional IT in the emerging information environment will be to manage the multiple boundaries between bodies of content in the wild, those that are available to the institution through formal contracts and those that belong to the institution. And although the metaphor of an institutional core surrounded by concentric layers of increasingly less institutionally controlled content remains useful, the ability to express it physically is in decline; the cloud computing paradigm decouples physical possession from ownership and control, exposing our reliance on the somewhat accidental modes of security and
oversight that come from flipping a switch or locking something in a closet.

Creating the appropriate degree of what Yochai Benkler calls institutional “permeability” will demand a sophisticated infrastructure in itself, including not only the tagging, commenting, and interaction tools we mentioned above but also a robust identity management capability. Indeed, a virtual, reliable means of identifying users is perhaps the single most important tool that can help institutions manage data superabundance and the spreading virtualization of computing infrastructure. There is some reason to expect that identity itself will be decoupled from institutions; “user-centric” identity belongs in the same box as wikis, folksonomies, social networks, and other resources that leverage the logic of the web to free personal interaction from institutional structures. But uptake of user-centric identity tools such as OpenID and Microsoft Windows CardSpace remains very small, and institutions may well have needs and concerns of their own that aren’t well served by these more generic tools. We expect that federated identity in particular, with its promise to permit interconnection within a regime defined by institutional needs, will be a critical enabler of interinstitutional collaboration and a means of navigating the complex intellectual property and security requirements of the ultra-digitized world.

It’s well to remember in addition that much of an institution’s most valuable content is purely internal and may be governed by stringent management or legal requirements. Though content is commonly thought of as “unstructured” data, institutions have many of the same concerns about it that they do with structured business data: They want to be sure that it’s of high quality, discoverable, and secure, and they want to get the maximum value from it that they can.

All of these considerations help explain why, notwithstanding all the excitement about free and miscellaneous information, institutions seem to be committed to extending enterprise management over content assets in much the same way they have long managed structured business information. (See Chapter 6 for our findings on enterprise content management environment trends.) Technology is increasingly capable of lending a hand: As Burton Group analyst Peter O’Kelly argues, the incorporation of documents and other forms of content into structured content frameworks like websites and document management systems is blurring the line between structured and unstructured data. “Many of today’s data and document management and integration challenges,” O’Kelly writes, “stem from what, in hindsight, seem like relatively arbitrary product categories that resulted from now fast-fading hardware, software, and economic constraints.” Technologies such as SQL/XML and XQuery increasingly permit data manipulation of documents in ways that bring to mind the manipulation of more atomic structured data in databases.

Though it’s hard to imagine a completely comprehensive version of enterprise content management at institutions as distributed and driven toward innovation as colleges and universities, it seems reasonable to believe that a large and growing portion of content would benefit from it. In particular, the institutional website, transactional documents, and records significant for their legal or business impact are likely to come under increasingly formal management within information systems, and learning objects and intellectual property assets are likely candidates as well. Our study findings show that institutions implementing web and records management systems report better outcomes relating to those types of content, and it’s easy to see, underneath the complexity of the content management challenge, principles of quality, stewardship, security, and decision support familiar from IT’s heritage of managing structured data increasingly applied to institutional content as well. In short, institutional IT units will
have to have a foot in each of two worlds, developing skills suitable for both the free and miscellaneous and the captive and categorized realms of content.

**Research Data**

Our study devotes a full chapter to the question of how institutions should manage the immense amount of data that modern research produces, and it’s unnecessary here to repeat in detail the complications this difficult question poses. Suffice it to say that science is generating unprecedented quantities of digitized data even as less abundant older data of unquestioned significance deteriorates for lack of good archival solutions; that the storage, sharing, and analysis of research data poses great potential as well as many problems; and that our results produced little evidence of a mature or even coherent institutional practice regarding these matters, at least from the standpoint of our CIO respondents.

Scientists and scholars often profess an open-access ethic that overlaps in many ways with the “free and miscellaneous” thinking of cyberculture pundits. Yet in some respects research data helps illuminate the weaknesses of information models that presume that data abundance generates its own solutions. Much research data is extremely recondite material that can only be handled properly by investigators with a deep and thorough understanding of the experiments and apparatuses that created it; it’s not clear that if crowds examine such data, they’ll produce wisdom. Attaching illuminating metadata is labor intensive and may be futile in the absence of standards that make the metadata broadly comprehensible. Above all, the availability and longevity of the data is questionable, given the lack of a widely shared, robust process for long-term archival storage. Many models for such preservation have been proposed or are functioning, but their viability over time is dubious or at least unknown; what is clear is that preservation will require some kind of ongoing commitment. Indeed, many research preservation advocates have argued that the researcher’s home institution will often be the most appropriate archive.

It may seem odd to propose that institutions should be repositories of the research data they produce, since even the most eminent institution contributes only a small fraction of all the work related to a given research problem, and deep understanding of research data sets requires the correspondingly deep domain expertise that aggregates around disciplines, not institutions. No institution is likely to be able to archive a definitive collection of data purely on the strength of its own research contributions, and the whole point of keeping data around for a long time is to make it available to the largest possible audience of researchers—including those at competing institutions and those with no institutional affiliation.

But then again, it’s also hard to imagine any single entity, including governments, institutes, or disciplinary bodies, that can hope to assemble a comprehensive body of research data relating to a given problem, let alone the whole range of science and scholarship. Preservation of research data therefore presupposes a distributed network of many archival nodes within a global system of search and access. When this issue is put aside, institutions do indeed have some characteristics that align well with the mission of data preservation. Unlike grants, projects, teams, experiments, and even many laboratories or institutes, higher education institutions are built on a model of perpetual continuity of operations. Historically, they have demonstrated a remarkable talent for longevity; universities are among the oldest institutions in the Western world. As the formal owners of data in many circumstances, they have a vested interest in the long-term economic value of research produced on campus, and they share in the glory of discovery even
Institutional Data Management

when it can’t be monetized. Thanks to their size and the range of disciplines they support, they have a broad information infrastructure that includes expertise no individual research group is likely to possess on its own.

Of course, institutions may sometimes better serve their longevity by declining new responsibilities than by taking them on, and particularly in these straitened times it’s hard to see how colleges and universities would be able to fund and support long-term research data preservation. Even if they could, it’s not clear that central IT units, for all their experience with data management and the frequent coordination with researchers, should play a major role. As we noted in Chapter 7, this may be somebody else’s job (the library is often mentioned), though in the absence of a clear responsible party, central IT surely belongs on the short list of candidates.

More definitely, however, we believe that an institutional dialogue on these issues is vital and that central IT should approach that dialogue equipped with its particular knowledge and concerns regarding data management. There’s a lot that IT can bring to the table regardless of the role the institution chooses to undertake. If institutions prefer to take on the responsibility of maintaining their own or others’ research data, they’ll need to be realistic about the long-term costs of maintaining a stable data environment, and they’ll need good advice not just about storage trends but also about the standards, identity management environment, and security and privacy policies needed to maintain a responsible archive. Where institutions choose not to be repositories of research data, they will surely still want to have the capacity to archive their data elsewhere, to keep track of it, and to access the repositories that do emerge. Central IT units are unlikely to be called on to deliver all these services, but they will certainly be involved in providing some of them, and by working with librarians, archivists, and researchers as well as institutional administrators, they can make an invaluable contribution to the institution’s research data legacy.

**Governing Institutional Data Management**

We don’t doubt that there’s a great deal of value in the new thinking about mass access to masses of data. The Internet and its many data tributaries make up far too large and chaotic a universe to be governed by information models that struggled to keep up even in the age of print and microfilm. The web’s new sociability is an enormously valuable tool, not just on its own merits but also as a means of interpreting, annotating, and assessing an otherwise impossibly complex ocean of information.

Nor do we doubt that the social web will in some ways displace the traditional functions and communities associated with higher education institutions. Contemporary business is filled with too many examples of industries in which some essential relationship between consumer, product, and/or distribution channel has been irreparably interrupted by “free and miscellaneous” information for higher education institutions to rest too comfortably on their long history as purveyors of information *par excellence*. The web has a powerful ability to foster what the University of Virginia’s James Hilton calls “unbundling”: the consumer’s ability to choose a single product out of what was once a mutually reinforcing set of products. Today it’s easy for a student to avoid a trip to the library by conducting a Google search, to fulfill a prerequisite with an online course from a university on another continent, or to discuss the book he’s been assigned with his high school friends on Facebook rather than in a midnight dorm room session. In small ways and large, such actions reduce the number of touch points that institutions so highly value. Higher education institutions need to understand the centripetal forces of
the modern information environment in order to better understand the centrifugal forces it can contribute to enrich community life; as Richard Katz writes, “The institution’s information system is increasingly the means by which the institution regulates the boundaries and conditions of its community, the rules of community engagement, and the boundaries, scope, and nature of the community’s access to scholarly resources.”14

If the institution has an advantage over the shifting communities of convenience and the anonymous mass behavioral phenomena that give the new cyberculture its power, it lies in the persistence and tangible reality of its community. Operating at a human scale and capable of intimacy as well as bureaucratic alienation, the institution provides an extremely sophisticated social and cultural experience that students continue to pay a premium to partake in, and it can dedicate itself consistently, across many areas of endeavor and over time, to coherent practical, ethical, and intellectual goals. It’s not at all clear that wise virtual crowds, even those consuming superabundant free and miscellaneous information, can do that.

The institution, in short, is capable of governing itself, and although these themes of community and culture may seem far removed from the topic of institutional data management, we believe that they are in fact closely bound to it. Today more than ever, data is politics, and ultimately the future of data management lies in the way we govern our institutions more than in the technologies or even the policies we adopt.

In a study of IT governance published in 2008, ECAR found that institutions reporting higher levels of constituent participation in IT governance also tended to report better governance effectiveness.15 (We also, alas, found that institutions rated the maturity of their IT governance on the low side.) In a modest way, this reflects the wise-crowd phenomenon within institutional politics; generally speaking, the benefits of participation (and particularly the participation of key figures, including executive leaders) seem to outweigh the disadvantages of inclusive governance. It seems especially desirable to have an inclusive IT governance structure when dealing with data issues, which so often have multiple business, academic, and technology components that call for different (and hopefully compatible, or at least reconcilable) viewpoints. Add this to the current study’s finding that institutions reporting better understanding of data quality issues by executives, managers, and staff also tend to have not just better enterprise data quality scores, but also better major data management outcomes across the board, and we believe that there is a sound basis for taking data management issues to governance bodies.

We believe that issues such as data stewardship and access, data quality, institutional capacity for analytics and decision support, and content management strategy are sufficiently significant to merit the attention of institutional IT governance bodies not just because they involve expenditure and regulatory exposure issues, but also because a discussion of the information environment has a lot to do with the nature and culture of the institution as a whole. Data issues such as these are emphatically not just IT issues, or the preserve of any department; they demand a truly institutional airing. Whether the institution formalizes data stewardship, for example, is a decision that will proceed in part from an estimation of whether an informal or formal approach is more appropriate within the institutional culture. How deep and how fast to proceed with an analytics initiative likewise calls for an evaluation of what kind of decision making (and decision obeying) prevails in the institution and whether a change is feasible or merited.

Above all, though, we would call on IT governance to address questions of data governance as a way to spread awareness and
understanding of the many facets of the data management challenge among those who set priorities, make major IT decisions, and have the ear (and often enough are the ear) of executive leadership. The fact is, though our study avoids the conclusion that institutions are “drowning” in data, we found plenty of evidence of muddling through on matters of vital strategic concern. Half of our respondent institutions disagreed that their institutions get maximum value from business data today. Where do you want your institution to be in three years?

Endnotes
2. Tony Smith, “Hard Disk Density to Hit 2.4 Tb/in² by 2014,” “Register Hardware (June 5, 2009), http://www.reghardware.co.uk/2009/06/05/hdd_data_density_drive/.
Appendix A

Institutional Respondents to the Online CIO Survey

A.T. Still University of Health Sciences
Antioch University System Administration
Aquinas College
Arizona State University
Athens State University
Auburn University
Austin Community College District
Austin Peay State University
Azusa Pacific University
Baker College System
Baker University
The Banff Centre
Barnard College
Barton County Community College
Bates College
Berry College
Bethel University
Big Bend Community College
Black Hills State University
Blue Ridge Community and Technical College
Bluefield State College
Board of Governors State University System of Florida
Bridgewater State College
Brock University
Broome Community College
Brown University
Bryant & Stratton College System Office
Bucknell University
California Institute of Technology
California Lutheran University
California Polytechnic State University, San Luis Obispo
California State University, Bakersfield
California State University, Channel Islands
California State University, Monterey Bay
California State University, Sacramento
California State University, San Bernardino
Calvin College
Camosun College
Canadian University College
Cardinal Stritch University
Carlos Albizu University
Case Western Reserve University
Cecil College
Central Connecticut State University
Central Michigan University
Central Piedmont Community College
Chattanooga State Technical Community College
Chicago State University
Christopher Newport University
The Citadel
Citrus College
Clarion University of Pennsylvania
Clark University
Clarkson College
Clemson University
Clovis Community College
College of Mount Saint Joseph
The College of New Jersey

©2009 EDUCAUSE. Reproduction by permission only.
College of the Holy Cross
Collin County Community College District
Colorado College
Colorado Mountain College
Colorado State University
Columbia College Chicago
Community College of Rhode Island
Concord University
Concordia College–Moorhead
Concordia Seminary
Concordia University Texas
Corban College & Graduate School
Cornell University
Dabney S. Lancaster Community College
Dakota County Technical College
Dakota Wesleyan University
Dalton State College
Dartmouth College
Davenport University
Delaware State University
DeVry University–Corporate Office
Dickinson College
East Stroudsburg University of Pennsylvania
Eastern Iowa Community College District
Eastern Mennonite University
Eastern Oregon University
Eastern Shore Community College
Edinboro University of Pennsylvania
Elmhurst College
Elmira College
Elon University
Emory University
Emporia State University
Fayetteville State University
Ferrum College
Florida International University
Franklin University
Franklin W. Olin College of Engineering
Fresno City College
Furman University
Gallaudet University
Galveston College
Genesee Community College
Georgia Southern University
Georgia State University
Glendale Community College
Grace College and Seminary
Guilford College
Harper College
Houston Community College
Hudson Valley Community College
Illinois State University
Illinois Valley Community College
Illinois Wesleyan University
Indiana University Southeast
Jamestown College
Kankakee Community College
Kenyon College
Keyano College
Lafayette College
Lander University
Lesley University
LeTourneau University
Lewis & Clark College
Lewis University
Lincoln Memorial University
Linn-Benton Community College
Loras College
Luther College
Lynchburg College
Lyon College
Manhattan College
Mansfield University of Pennsylvania
Maricopa Community College District
Marietta College
Marquette University
Marymount University
Marywood University
Massachusetts College of Liberal Arts
Memorial University of Newfoundland
Mercyhurst College
Metropolitan Community College
Metropolitan State College of Denver
Miami University
Michigan State University
Middle Tennessee State University
Middlebury College
Millersville University of Pennsylvania
Millikin University
Mills College
Missouri Western State University
Monmouth College
<table>
<thead>
<tr>
<th>Institution Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Montana State University–Great Falls, College of Technology</td>
</tr>
<tr>
<td>Montana State University Billings</td>
</tr>
<tr>
<td>Montgomery County Community College</td>
</tr>
<tr>
<td>Morgan State University</td>
</tr>
<tr>
<td>Morningside College</td>
</tr>
<tr>
<td>Moss Landing Marine Laboratories</td>
</tr>
<tr>
<td>Mott Community College</td>
</tr>
<tr>
<td>Mount Mary College</td>
</tr>
<tr>
<td>Mount Royal College</td>
</tr>
<tr>
<td>Mount Vernon Nazarene University</td>
</tr>
<tr>
<td>Naropa University</td>
</tr>
<tr>
<td>New College of Florida</td>
</tr>
<tr>
<td>New England College of Optometry</td>
</tr>
<tr>
<td>New Mexico Institute of Mining and Technology</td>
</tr>
<tr>
<td>Norfolk State University</td>
</tr>
<tr>
<td>North Carolina State University</td>
</tr>
<tr>
<td>North Shore Community College</td>
</tr>
<tr>
<td>Northern Arizona University</td>
</tr>
<tr>
<td>Northwood University</td>
</tr>
<tr>
<td>Nova Scotia Community College</td>
</tr>
<tr>
<td>Oakland University</td>
</tr>
<tr>
<td>Oberlin College</td>
</tr>
<tr>
<td>Occidental College</td>
</tr>
<tr>
<td>Ocean County College</td>
</tr>
<tr>
<td>Ohio Northern University</td>
</tr>
<tr>
<td>Okanagan College</td>
</tr>
<tr>
<td>Oregon Institute of Technology</td>
</tr>
<tr>
<td>Ouachita Technical College</td>
</tr>
<tr>
<td>Pace University</td>
</tr>
<tr>
<td>Pennsylvania College of Technology</td>
</tr>
<tr>
<td>Phoenix College</td>
</tr>
<tr>
<td>Plymouth State University</td>
</tr>
<tr>
<td>Pomona College</td>
</tr>
<tr>
<td>Portland State University</td>
</tr>
<tr>
<td>Prince George's Community College</td>
</tr>
<tr>
<td>Princeton University</td>
</tr>
<tr>
<td>Queen's University</td>
</tr>
<tr>
<td>Queens University of Charlotte</td>
</tr>
<tr>
<td>Ramapo College of New Jersey</td>
</tr>
<tr>
<td>Randolph-Macon College</td>
</tr>
<tr>
<td>Raritan Valley Community College</td>
</tr>
<tr>
<td>Regis University</td>
</tr>
<tr>
<td>Rhode Island School of Design</td>
</tr>
<tr>
<td>Rhodes State College</td>
</tr>
<tr>
<td>Rice University</td>
</tr>
<tr>
<td>Ripon College</td>
</tr>
<tr>
<td>Rosalind Franklin University of Medicine and Science</td>
</tr>
<tr>
<td>Rutgers, The State University of New Jersey</td>
</tr>
<tr>
<td>Rutgers, The State University of New Jersey/Camden</td>
</tr>
<tr>
<td>Saint Mary's College</td>
</tr>
<tr>
<td>Saint Mary's College of California</td>
</tr>
<tr>
<td>Saint Mary's University of Minnesota</td>
</tr>
<tr>
<td>Salisbury University</td>
</tr>
<tr>
<td>Samford University</td>
</tr>
<tr>
<td>San Diego State University</td>
</tr>
<tr>
<td>San Francisco State University</td>
</tr>
<tr>
<td>San Jose State University</td>
</tr>
<tr>
<td>Santa Barbara City College</td>
</tr>
<tr>
<td>Santa Clara University</td>
</tr>
<tr>
<td>Santa Fe College</td>
</tr>
<tr>
<td>Savannah College of Art and Design</td>
</tr>
<tr>
<td>Seattle Pacific University</td>
</tr>
<tr>
<td>Seneca College of Applied Arts and Technology</td>
</tr>
<tr>
<td>Sheridan College</td>
</tr>
<tr>
<td>Sheridan Institute of Technology &amp; Advanced Learning</td>
</tr>
<tr>
<td>Siena College</td>
</tr>
<tr>
<td>Solano Community College</td>
</tr>
<tr>
<td>South Dakota State University</td>
</tr>
<tr>
<td>Southern Oregon University</td>
</tr>
<tr>
<td>Southside Virginia Community College</td>
</tr>
<tr>
<td>Southwest Baptist University</td>
</tr>
<tr>
<td>Southwest Minnesota State University</td>
</tr>
<tr>
<td>Southwest Tennessee Community College</td>
</tr>
<tr>
<td>Southwestern Oregon Community College</td>
</tr>
<tr>
<td>Southwestern University</td>
</tr>
<tr>
<td>Springfield Technical Community College</td>
</tr>
<tr>
<td>St. Cloud State University</td>
</tr>
<tr>
<td>Stephen F. Austin State University</td>
</tr>
<tr>
<td>SUNY College at Fredonia</td>
</tr>
<tr>
<td>SUNY College at Oswego</td>
</tr>
<tr>
<td>Swarthmore College</td>
</tr>
<tr>
<td>Sweet Briar College</td>
</tr>
<tr>
<td>Tennessee Technological University</td>
</tr>
<tr>
<td>Texas A&amp;M University–Corpus Christi</td>
</tr>
<tr>
<td>Texas Wesleyan University</td>
</tr>
<tr>
<td>Troy University</td>
</tr>
</tbody>
</table>
United States Naval Academy
Universal Technical Institute Corporate Office
Universidad del Sagrado Corazón
University of Alabama at Birmingham
University of Baltimore
The University of British Columbia
University of California Office of the President
University of Central Missouri
University of Colorado at Boulder
University of Delaware
University of Denver
University of Florida
University of Kansas
University of La Verne
University of Louisville
University of Manitoba
University of Maryland Eastern Shore
University of Maryland, Baltimore
University of Maryland, Baltimore County
University of Massachusetts Boston
University of Medicine & Dentistry of New Jersey
The University of Memphis
University of Minnesota Duluth
University of Mississippi
The University of Montana
University of Nebraska at Kearney
University of Nevada, Las Vegas
University of North Carolina at Chapel Hill
University of North Carolina at Greensboro
University of North Carolina at Pembroke
University of North Texas
University of North Texas HSC at Fort Worth
University of Oregon
University of Ottawa
University of Puget Sound
University of Regina
University of Richmond
University of San Diego
University of San Francisco
University of South Carolina
University of Southern California
University of St. Thomas
University of Tennessee at Chattanooga
University of Texas at Dallas
University of Texas of the Permian Basin
University of the Incarnate Word
University of the Sciences in Philadelphia
University of the Southwest
University of Toronto
University of Vermont
University of Victoria
University of Washington, Tacoma
University of West Florida
University of Windsor
University of Wisconsin–Eau Claire
University of Wisconsin–Madison
University of Wisconsin–Milwaukee
University of Wisconsin–Platteville
University of Wisconsin Extension
Virginia Tech
Wagner College
Walsh University
Wentworth Institute of Technology
Wesleyan University
West Virginia School of Osteopathic Medicine
Western Carolina University
Western New Mexico University
Western State College of Colorado
Westfield State College
Whitman College
Wilberforce University
Williamette University
Wofford College
Appendix B

Interviewees in Qualitative Research

**Arizona State University**
John Rome, Associate Vice President, University Technology, UTO Enterprise Data Systems

**Azusa Pacific University**
Don Davis, Associate Vice President and Chief Information Officer

**Bates College**
Gene Wiemers, Vice President for Information and Library Services

**Brown University**
Terri-Lynn Thayer, Assistant Vice President/Deputy Chief Information Officer, Computing and Information Services

**Community College of Rhode Island**
Stephen Vieira, Chief Information Officer and Executive Director of IT

**Eastern Mennonite University**
Jack Rutt, Director, Information Systems
Ben Beachy, Application Development Manager

**Edinboro University of Pennsylvania**
Andrew C. Lawlor, Associate Vice President, Technology and Communications

**Georgia State University**
J. L. Albert, Associate Provost and Chief Information Officer

**Lafayette College**
Neil McElroy, Dean of Libraries and Information Technology Services

**North Carolina State University**
Marc Hoit, Vice Chancellor for IT and Chief Information Officer

©2009 EDUCAUSE. Reproduction by permission only.
North Shore Community College
Janice Forsstrom, Vice President, Administration and Finance
Nancy Sherwood, Project Manager
Patrick Thomas, Director of Administrative Technology

University of Central Missouri
F. Russell Helm, Chief Information Officer

University of Minnesota Duluth
Linda L. Deneen, Director of Information Technology Systems and Services

University of North Texas
Maurice Leatherbury, Acting Vice President for Information Technology and Chief Information Officer

University of North Texas Health Science Center
Renee Drabier, Vice President Information Resources and Technology and Chief Information Officer

University of San Francisco
Steve Gallagher, Chief Information Officer

University of St. Thomas
Samuel J. Levy, Vice President Information Resources and Technologies and Chief Information Officer
Chris Gregg, Director of Information Technology

University of Tennessee at Chattanooga
Clinton W. Smullen III, Director of Academic and Research Computing Services

University of Virginia
Shirley C. Payne, Assistant Vice President for Information Security, Policy, and Records
### Table C-1. Supplement to Figure 8-2: Data Management Outcomes, by Enterprise Data Quality Score

<table>
<thead>
<tr>
<th>Enterprise Data Quality Score</th>
<th>Outcome Measure</th>
<th>Restricted/ sensitive data is secure from unauthorized access.</th>
<th>We can support anticipated growth in the volume of data over the next three years.</th>
<th>Employees understand their responsibilities in the use of data.</th>
<th>We get maximum academic value from institutional data.</th>
<th>We get maximum business value from institutional data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low DQ score (less than 2.5)</td>
<td>Mean*</td>
<td>3.54</td>
<td>2.91</td>
<td>2.72</td>
<td>2.16</td>
<td>2.01</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>68</td>
<td>69</td>
<td>68</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.871</td>
<td>1.222</td>
<td>0.990</td>
<td>0.784</td>
<td>0.782</td>
</tr>
<tr>
<td>Medium DQ score (2.5–3.5)</td>
<td>Mean*</td>
<td>3.86</td>
<td>3.35</td>
<td>3.17</td>
<td>2.75</td>
<td>2.68</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>151</td>
<td>150</td>
<td>150</td>
<td>151</td>
<td>151</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.693</td>
<td>0.962</td>
<td>0.823</td>
<td>0.903</td>
<td>0.876</td>
</tr>
<tr>
<td>High DQ score (greater than 3.5)</td>
<td>Mean*</td>
<td>4.15</td>
<td>3.57</td>
<td>3.44</td>
<td>3.14</td>
<td>3.21</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>85</td>
<td>84</td>
<td>85</td>
<td>83</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.716</td>
<td>0.960</td>
<td>0.823</td>
<td>0.885</td>
<td>0.919</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
Table C-2. Supplement to Figure 8-3: Data Management Outcomes, by Agreement That Institution Provides Effective Training to Users of Analytics Tools

<table>
<thead>
<tr>
<th>Agreement That Institution Provides Effective Training to Users of Analytics Tools</th>
<th>Outcome Measure</th>
<th>We can support anticipated growth in the volume of data over the next three years.</th>
<th>Employees understand their responsibilities in the use of data.</th>
<th>We get maximum academic value from institutional data.</th>
<th>We get maximum business value from institutional data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree/disagree</td>
<td>Mean*</td>
<td>3.08</td>
<td>2.80</td>
<td>2.27</td>
<td>2.19</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>120</td>
<td>119</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>1.066</td>
<td>0.907</td>
<td>0.827</td>
<td>0.873</td>
</tr>
<tr>
<td>Neutral</td>
<td>Mean*</td>
<td>3.37</td>
<td>3.23</td>
<td>2.88</td>
<td>2.83</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>103</td>
<td>104</td>
<td>102</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.970</td>
<td>0.803</td>
<td>0.787</td>
<td>0.760</td>
</tr>
<tr>
<td>Agree/strongly agree</td>
<td>Mean*</td>
<td>3.61</td>
<td>3.54</td>
<td>3.22</td>
<td>3.22</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>76</td>
<td>76</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>1.047</td>
<td>0.824</td>
<td>0.947</td>
<td>0.974</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree

Table C-3. Supplement to Figure 8-4: Data Management Outcomes, by Institution Technology Adoption Approach

<table>
<thead>
<tr>
<th>What best characterizes your institution in terms of adopting new technologies?</th>
<th>Outcome Measure</th>
<th>Restricted/sensitive data is secure from unauthorized access.</th>
<th>We can support anticipated growth in the volume of data over the next three years.</th>
<th>Employees understand their responsibilities in the use of data.</th>
<th>We get maximum academic value from institutional data.</th>
<th>We get maximum business value from institutional data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early adopter</td>
<td>Mean*</td>
<td>4.06</td>
<td>3.81</td>
<td>3.53</td>
<td>3.47</td>
<td>3.38</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>32</td>
<td>31</td>
<td>32</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.716</td>
<td>0.749</td>
<td>0.842</td>
<td>0.915</td>
<td>0.871</td>
</tr>
<tr>
<td>Mainstream adopter</td>
<td>Mean*</td>
<td>3.92</td>
<td>3.37</td>
<td>3.25</td>
<td>2.80</td>
<td>2.76</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>199</td>
<td>199</td>
<td>198</td>
<td>197</td>
<td>198</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.768</td>
<td>1.006</td>
<td>0.841</td>
<td>0.845</td>
<td>0.885</td>
</tr>
<tr>
<td>Late adopter</td>
<td>Mean*</td>
<td>3.64</td>
<td>2.96</td>
<td>2.64</td>
<td>2.19</td>
<td>2.14</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.756</td>
<td>1.180</td>
<td>0.893</td>
<td>0.914</td>
<td>0.954</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
Table C-4. Supplement to Figure 8-5: Data Management Outcomes, by Agreement That Institution Has Necessary Commitment to Support Long-term Preservation of Research Data

<table>
<thead>
<tr>
<th>My institution has the necessary institutional commitment to support the long-term preservation of research data.</th>
<th>Outcome Measure</th>
<th>Restricted/ sensitive data is secure from unauthorized access.</th>
<th>We can support anticipated growth in the volume of data over the next three years.</th>
<th>Employees understand their responsibilities in the use of data.</th>
<th>We get maximum academic value from institutional data.</th>
<th>We get maximum business value from institutional data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree/disagree</td>
<td>Mean*</td>
<td>3.78</td>
<td>3.01</td>
<td>2.86</td>
<td>2.47</td>
<td>2.38</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>124</td>
<td>123</td>
<td>125</td>
<td>124</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.716</td>
<td>1.149</td>
<td>0.886</td>
<td>0.949</td>
<td>0.940</td>
</tr>
<tr>
<td>Neutral</td>
<td>Mean*</td>
<td>3.85</td>
<td>3.47</td>
<td>3.27</td>
<td>2.86</td>
<td>2.80</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>100</td>
<td>99</td>
<td>98</td>
<td>98</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.757</td>
<td>0.951</td>
<td>0.856</td>
<td>0.908</td>
<td>0.915</td>
</tr>
<tr>
<td>Agree/strongly agree</td>
<td>Mean*</td>
<td>4.28</td>
<td>3.74</td>
<td>3.64</td>
<td>3.05</td>
<td>3.16</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.605</td>
<td>0.850</td>
<td>0.811</td>
<td>0.887</td>
<td>0.945</td>
</tr>
</tbody>
</table>

*Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree
Appendix D

Bibliography


©2009 EDUCAUSE. Reproduction by permission only.
Institutional Data Management

ECAR Research Study 8, 2009


