We live in an era of data superabundance. All kinds of things—communication, media, work activities—have become data by the simple fact of being captured or expressed using a digital tool. Computation and digitization have revolutionized scientific and scholarly research, generating stupendously large data flows and unprecedented archival challenges. At the same time, computing has become a mass social activity that swamps formerly adequate tools
for categorizing and storing information. This Internet-fed data deluge is one reason why so much discussion of contemporary academic and popular culture turns on concepts like open access, the wisdom of crowds, and new modes of “social production.”

Higher education institutions aren’t the only organizations grappling with these challenges, but their historic mission of producing and aggregating information, and the complexities of institutional administration in an era of accountability and information regulation, make colleges and universities especially sensitive to them. It seems that every aspect of higher education IT administration, from the most traditional, such as managing business applications, to the most emergent, such as online video and support for computational research, brings its associated bag of tricky data-related problems.

As a result, data management has become a topic with powerful implications for the way institutions carry out instruction, research, and administration. ECAR’s recent study of institutional data management practices, summarized in this Roadmap, took a broad look at institutional data infrastructure and policies and at respondent assessments of data-management-related performance.

Deluge and Diversity

We found lots of signs that keeping up with today’s data growth is a struggle. On average, institutions reported increases in the volume of every kind of data we asked about. Learning management systems data, e-mail, and, at research institutions, research data were the fastest-growing types. Looking forward, institutions anticipated a median 20% increase in central IT disk storage needs in the next year.

Still, respondents were generally positive about having the infrastructure needed to handle institutional data, now and in the near future. They seemed to be less concerned about the sheer volume of data they need to accommodate than about the variety of data they’re called upon to manage. On an agreement scale of 1 (strongly disagree) to 5 (strongly agree), institutions averaged agreement of 3.32 (above neutral) that they can support growth in the volume of data over the next three years, but only a below-neutral 2.81 agreement that they can effectively manage all the varieties of data and digital content that the institution needs.

We found some echoes of these results when we asked institutions how they managed the large and loosely defined universe of data known as “content”—website materials, video and audio files, documents, and other digital assets that lack the rigid structure that characterizes traditional business application data. Institutional content management environments were generally distributed and diffuse, most often characterized by a mix of enterprise and local best-of-breed products focused on individual content types. Many institutions characterized their content environments as purely local or ad hoc.

However, respondents seemed to look forward to a different environment in the future, many of them anticipating a move toward more enterprise control. While only 12% said they now had integrated enterprise content management systems, 32% said they expected to have them within three years. We also found some evidence that these investments may pay off: Institutions with more enterprise-oriented content management environments tended to report higher agreement that they could effectively manage all the varieties of data and digital content that they needed.

Enterprise Data Quality

To deliver value, business data has to be accurate, well defined, kept up to date, and generally of high quality. Our respondents, however, were lukewarm at best in their assessments of the quality of data in major administrative enterprise systems and data stores. While they tended to agree that a system of record was identified for each major data element, they gave lower marks to the institution’s ability to ensure that when a data element changed, the change was propagated to all relevant enterprise and local systems. In addition, mean agreement that systems and processes were in place to validate enterprise data and to document and
review data quality issues was below neutral. When all of the data quality items we asked about were averaged into an enterprise data quality score, the mean score (3.06) was dead in the middle of our 5-point scale.

Data quality has been a business and an IT issue for many years, and it may be tempting to consider such mediocrity as a perennial, unavoidable, and ultimately tolerable problem. Yet our study also produced clear evidence that good data quality is foundational to many aspects of effective institutional data management. As we note below, better results for most of the major data management outcomes we asked about were strongly and positively associated with better perceived enterprise data quality. What’s more, where respondents agreed more strongly that executives, managers, and staff received training in data quality issues, enterprise data quality scores were higher.

**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 strongest incentives to care for it. Our study suggests that this advice is being taken to heart at our respondent institutions: Business and academic units most often had the primary responsibility for such matters as data accuracy, fitness for purpose, and data access, while IT units tended to have the primary responsibility for specifying data formats and maintaining data definitions. This division of labor notwithstanding, at two-thirds of institutions it was the central IT unit that had responsibility for notifying parties affected by data breaches. Such breaches were not uncommon: 30% of institutions reported at least one incident in the previous 24 months requiring notification.

Institutions with formal policies assigning data stewardship responsibilities (about a third of our respondent institutions) tended to report slightly better enterprise data quality than those without them. But somewhat surprisingly, these institutions didn’t differ significantly from others in the perceived security of data from unauthorized access. On the other hand, institutions that had documented policies on the acceptable use of institutional data (which were much more common) did do better on perceived security of data than those lacking them.

**Research Data Management**

A profoundly important discussion about the need for a new paradigm in managing research data has been unfolding for the last decade. When we looked at the context of digital research data management at our respondent institutions, the results suggested that the institutional role remains as inchoate and elusive as the larger paradigm.

The research data problem has many dimensions, including direct support of research data infrastructure, complex metadata and access issues, and the long-term preservation of data in an archival setting. We found that research data management support was widely distributed among various entities, with central IT’s largest role in traditional activities such as providing storage and backup/recovery services. More emergent activities, especially metadata creation, seemed somewhat orphaned, drawing high percentages of non-involvement or “don’t know” responses. Despite a good deal of forward-looking thought about the role librarians might play in research data management, from our respondents’ standpoint libraries were the least involved of the entities we asked about. Overall, respondents gave their institutions poor marks when evaluating their capabilities for the long-term preservation of digital research data and only modest marks for meeting the data-related needs of institutional researchers. For the latter measure, doctorals had the lowest mean result among all the Carnegie types.

**Getting Value from Institutional Data**

Data management is a complex, multifaceted discipline, but in the end perhaps the most important function it can serve is ensuring that the institution gets the maximum academic and business value from its data. Yet as the chart that begins this Roadmap shows, of all the major data management outcomes we asked about, our items relating to this goal inspired the weakest respondent agreement. Granted,
Based on the findings in *Institutional Data Management in Higher Education*, ECAR recommends that institutional leaders:

1. **Be proactive, not fatalistic, about improving institutional data quality.**

   No practice we looked at more frequently traveled in company with desirable circumstances and good outcomes than higher perceived enterprise data quality. Efficient operations, informed tactical and strategic planning, and getting maximum value from information assets all presume accurate, timely, well-defined data. Staff throughout the institution should be trained in data quality issues, and processes for improving data quality should be systematic and ongoing, not dedicated to spot fixes.

2. **Invest in analytics.**

   Given the enormous investments institutions have made in creating powerful integrated administrative systems, it’s remarkable how little progress they’ve made toward building the infrastructure and culture necessary to put their business data to management use. Respondent institutions that went beyond basic reporting to make use of advanced analytics techniques were more positive about their ability to get maximum value from institutional data.

3. **Think about data value, not just data security.**

   For all the angst that surrounds data security and privacy issues, security of data from unauthorized use was an area of relative confidence among our respondents. On the other hand, getting maximum value from institutional data was at the bottom of the outcomes we asked about. Due diligence calls for continued attention to data security, and probably always will, but good data management has to be measured by how effectively data is put to work, as well as how tightly it’s secured.

4. **Confront the digital research data quandary.**

   Digital research data is exploding in volume and complexity, and it is a major intellectual and economic resource. Information technology has the potential to make it far more accessible and productive, yet digital data is remarkably fragile, and its very abundance works against its effective use. Higher education executives need to work actively to ensure that high-quality, sustainable archival solutions become available, and they must define the institutional role in that process.

In asking whether institutions were getting “maximum” value, we set a high standard—but isn’t maximizing value the appropriate standard for today’s data-driven institutions? Despite generally low self-evaluations, however, we found a wide distribution in respondent agreement about getting value from data and significant associations between stronger agreement and certain institutional practices. As with every outcome we asked about, better perceived enterprise data quality went along with stronger agreement about getting maximum academic and business value from data. Institutions that gave themselves higher marks for training staff in data quality issues and analytics did better as well. Though on the whole we found only a modest infrastructure for data analytics among our respondent institutions and a tendency to use only rather basic transactional reporting rather than advanced analytics, those institutions that more frequently employed advanced analytics techniques also tended to report higher agreement about getting maximum value from data. Finally, in a relationship that suggests the importance of executive support and a culture of evidence in institutional administration, where respondents more strongly agreed that institutional leadership was committed to evidence-based decision making, and where they more strongly agreed that the institution was committed to making management information widely available, they also did better on our data value outcomes items.