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 averse 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—but 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...
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
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
Institutional Data Management

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.”

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 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. (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.reg hardware.co.uk/2009/06/05/hdd_data_density_drive/.


