The Information Architecture Imperative

Karin Steinbrenner, University of North Carolina at Charlotte
Overview

Colleges and universities are information-intensive. Faculty rely on data, information, and knowledge for their scholarship and instruction. Students consume information of all types—administrative (course schedules, tuition rates), course-related (readings, assignments), and personal (music, e-mail). Staff collect data and manage information for a variety of operational and decision-making purposes. College and university constituents create and use data and information daily for decisions about admission, registration, room scheduling, how many meals to prepare, which courses to take, what library material to access, who the most likely donors are, or which faculty receive tenure.

There are two challenges for information-intensive organizations. One is that information has historically been kept in silos. Administrative data is handled separately from research data, teaching material, or library resources. This separation inhibits users from finding and linking related pieces of information.

A second challenge is that most information is handled in a laissez-faire manner. Ninety percent of all information is unstructured, maintained in documents or other formats. While a great deal of attention has been paid to administrative data (managed by an enterprise resource planning [ERP] system), little thinking has gone into systematically assessing the various types of information an institution uses and how to manage that information to provide an integrated, central point of access to both structured and unstructured information.

Consider the following scenario, which illustrates the implications of a laissez-faire information environment:

A faculty member intends to prepare from his home an online IT security seminar for a small group of honor students. He would like to customize the seminar to meet the needs of the honor students. From the course management system (CMS), he would like to link to relevant library material. He is also interested in learning if others on campus specialize in this area; perhaps they can assist him with the seminar by contributing papers or serving as experts. Although the faculty member has access to the student information system (SIS), he cannot integrate information in the SIS with the CMS. He is also prevented from viewing information about students other than his advisees. The SIS might not be very helpful anyhow, since it only maintains course and grade history, not information about student expertise or interests. However, the institution uses e-portfolios, so the information is available elsewhere. Unfortunately, the faculty member does not know how to access them. The faculty member attempts to search the institution’s Web site for papers related to IT security, hoping to find another on-campus expert. The search yields no results. As a result of his experiences, the frustrated faculty member gives up on customizing the seminar.

Now consider a similar scenario, only this time in an institution that has developed an integrated information architecture:
The faculty member would enter the institution’s portal using a single sign-on. From his personal home page, he could check the library system and transfer relevant online library material into the Web pages of the CMS. Because the SIS and CMS are integrated, his seminar is automatically set up in the CMS, and the class list is populated from the SIS. From the class list, he can see a photograph of each student and access their e-portfolios, from which he learns about their academic backgrounds and areas of interest. He discovers that most participants have an extensive IT background. As a result, he begins the seminar at a much higher level than he originally intended. Searching the Web, he finds several papers written by an IT staff member for a professional organization. After a quick review, he clicks on the author’s e-mail address (which is provided with the content) and asks if he may use the papers as part of the course. He also learns that the staff member is willing to participate in class discussions. Because information has been categorized, related, and integrated, the faculty member can offer a much richer, better tailored seminar that suits the level of his students. It also saves him significant time and frustration.

At colleges and universities, everyone is a creator—as well as a consumer—of information. What students, faculty, and staff create and consume ranges from administrative data to images, documents, and Web pages. Unless information across institutional silos can be categorized, organized, and directed to the appropriate audience, how does a person find the needed information? How does a user know if the content is relevant and correct?

Many tools are available for information management: database management, records management, imaging, workflow, and, lately, enterprise content management systems (ECMS). However, implementing these individual technologies will not be successful unless the institution develops an overarching information architecture, which identifies what information to capture, how it is used, where it is stored, who needs it, when it is relevant, and why it should be maintained.

This Research Bulletin describes the need for an information architecture and proposes a framework for institutions to follow. Content management technologies are described, as are information delivery technologies. Finally, a description of “the information organization” is offered to help institutions integrate the many information components that must be managed in today’s information-intense colleges and universities.

**Highlights of an Information Architecture**

In order to develop an information architecture, colleges and universities must identify and categorize a range of information types (see Table 1). The two main categories are structured and unstructured data. Structured data is information that can be managed and manipulated by a database management system (DBMS). Structured data is used to generate reports and is analyzed using statistical analysis or decision support systems. Unstructured data is more common; the majority of an institution’s information is contained in documents and other non-structured formats, such as images and articles.
Table 1. Information Types of Structured and Unstructured Data

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Considerations</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Structured Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operational data</td>
<td>Operational data are managed and manipulated by the institution’s administrative systems. In most instances operational data are maintained in an integrated database. Operational data are seldom used for ad hoc reporting or decision-making because their structure has been optimized for operational transactions, such as with a purchase requisition.</td>
<td>Student records, personnel records, payroll, course catalog, financial data</td>
</tr>
<tr>
<td>Decision-making data</td>
<td>Data warehouses and/or data marts are derived from operational data. They represent a simplified, sometimes aggregate or time-stamped, version of the operational database. Data warehouses are designed to enable queries, comparisons, analysis, and reporting, all of which are important components of decision-making. Research data, such as census or GIS data, can be used in decision-making as well.</td>
<td>Student retention information linked to demographic data and grade reports; data used to support research</td>
</tr>
<tr>
<td><strong>Unstructured Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Books, journals, and articles</td>
<td>Typically an institution’s library manages books, journals, and professional articles. An increasing amount of this material is being converted to electronic formats, making it accessible on the Web. Library systems organize content by subject matter, allowing students and faculty not only to search by title or author but also to search the text itself. Beyond the library, the other major source of books, journals, and articles is course material, such as textbooks or course packs. Many textbook publishers release textbooks on CDs, often in a format compatible with the institution’s CMS.</td>
<td>Reference books, textbooks, encyclopedias, journals</td>
</tr>
<tr>
<td>E-mail</td>
<td>E-mail is an accepted mode of official communication and contains much valuable information. However, little has been done to manage e-mail as an institutional information asset in the same way that paper documents or operational data is managed.</td>
<td>E-mail records</td>
</tr>
<tr>
<td>Unstructured Data (cont.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td><strong>Web pages</strong></td>
<td>Higher education is creating massive quantities of Web pages. Faculty create Web pages for course material and research projects. Students create online portfolios. Many administrative offices publish Web pages as well. While most colleges and universities have a Webmaster who manages the Web infrastructure, it is rare that someone manages the content. Web content management systems are evolving to manage Web content.</td>
<td>Course Web pages, student service Web sites, benefits Web site</td>
</tr>
<tr>
<td><strong>Word processing documents</strong></td>
<td>Word processing documents have rarely been viewed as institutional information assets. Where required, official documents are archived and stored, but they typically are not linked with related types of information.</td>
<td>Public documents, drafts</td>
</tr>
<tr>
<td><strong>Images</strong></td>
<td>Images greatly enhance other data (photos in textbooks or graphs that help readers better understand the material) and can be information resources. For example, a scanned image of a newspaper article about an alumnus, when paired with the alum’s giving history, may help the development officer to better target a donation request.</td>
<td>Photographs, digital images, scanned documents</td>
</tr>
<tr>
<td><strong>Rich media</strong></td>
<td>Rich media is quickly becoming another form of data requiring management. A voice or video clip inserted in an online textbook can augment a point made in the text. As professors use audio or video, the institution is challenged to maintain the content and make it available, on-demand, with other course-related material.</td>
<td>Audio files, video</td>
</tr>
</tbody>
</table>
College and university constituents are information creators. Instructors develop academic information and research articles, as well as operational data (grades). In the process of registering, students create new student course records. By paying online, students and parents create entries in the accounts-payable system. Administrators, who operate the institution’s ERP system, generate data when issuing purchase orders or paying invoices. When a faculty member, student, or administrator makes a long-distance phone call, a record is created. Needless to say, everyone generates e-mail and creates documents, and many craft Web pages.

Most content was designed to be shared and used. Unfortunately, in the absence of an information taxonomy or categorization scheme, much of the information institutions create is lost or destroyed.

Institutions face a huge challenge in organizing the vast amounts of content created daily to give consumers ready access to all relevant information. Two major sets of tools are involved in an information architecture: those that create, capture, and categorize information, and those involved with information delivery.

Enterprise content management systems represent the latest iteration of many earlier information management systems, such as database management systems, record management tools, or integrated document management systems that often include imaging and workflow. ECMSs are tools that focus on content creation and management; they assist the institution in developing an information taxonomy and a methodology for information categorization. ECMSs employ schemes to identify information domains and classifications. Each information resource is placed into predefined categories; its attributes and content are specified through XML (extensible markup language). To understand a data element, the data needs to be described. For example, an address needs to be described not only as an address but also as to whether it is a person’s home, business, or vacation address; when it is valid; who has access to it; and who created it. XML can describe structured as well as unstructured information and associate multiple attributes with each data item (such as its creator, category, or time dependency). XML also allows for exchange of data between independent systems.

Enterprise information portals (EIPs) focus on information delivery and presentation. These tools have matured over the past few years and are designed to deliver the right information to the right person at the right time. Portals use directories that maintain information about individuals as well as a person’s role relative to the institution. Directories can deliver the correct information to the right person based either on that person’s role or on self-expressed interest. For example, a faculty member who does DNA research might automatically receive any new research papers on that topic. Other
tools, such as calendars, enable EIPs to deliver time-sensitive information when it is needed. However, portals can only deliver information that is properly organized and categorized.

**Information Architecture Framework**

Neither an ECMS nor an EIP can substitute for an information architecture. In fact, for the successful deployment of either system, organizations must develop a comprehensive information architecture (see Figure 1) that provides a framework for determining what information should be maintained and how to categorize and manage it.

**Figure 1. Information Architecture Framework**

[Diagram showing the information architecture framework with nodes labeled Creator, XML, Directory, Content Creators, Enterprise Content Management, Enterprise Information Portal, Content Consumers, Repository, and Consumer. The diagram illustrates the relationships and questions associated with creating and consuming content.]
Zachman's framework for enterprise architecture\textsuperscript{1} can be adapted to develop an information architecture (see Table 2). Such a framework provides a methodology to determine the what, how, where, who, when, and why of an information architecture.

**Table 2. Categories within an Information Architecture**

<table>
<thead>
<tr>
<th>Category</th>
<th>Question</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>What</td>
<td>What information should be maintained?</td>
<td>The institution's mission, strategic plan, goals, and objectives can serve as a good starting point to address this issue. Business rules and legislative requirements also may determine what data must be maintained.</td>
</tr>
<tr>
<td>How</td>
<td>How is the information used?</td>
<td>The “how” identifies in which process—manual or automatic—the information is used. It also defines how the information relates to other information (parent-child relationship, same category, and so on).</td>
</tr>
<tr>
<td>Where</td>
<td>Where is the information created? Where is it stored?</td>
<td>This identifies where or on which system the information was created, its original data source, and where the information is stored. It should also specify other systems that can use the information.</td>
</tr>
<tr>
<td>Who</td>
<td>Who is the information creator and/or custodian? Who are its consumers?</td>
<td>This identifies which constituents should have access to what information, based on institutional roles. The data custodian is responsible for information accuracy.</td>
</tr>
<tr>
<td>When</td>
<td>How long is the information good?</td>
<td>Information is time-sensitive. Every data item should be associated with a time attribute that identifies the time period for which the data element is valid.</td>
</tr>
<tr>
<td>Why</td>
<td>Why should the information be maintained?</td>
<td>This is probably the most difficult question to answer. The institution's goals and objectives may serve as a general guideline. Business practices and rules are other reasons for keeping data.</td>
</tr>
</tbody>
</table>

The information architecture sets the rules for the institution's ECMS, determining what information to keep, where it should be maintained, where it will be stored, when it is relevant, and why it should be kept. The EIP, using directory services, determines who needs what information, and when. It either pushes the information to the user or makes it readily accessible.
When Web content is posted within an information architecture framework, an ECMS template would gather metadata, which can be used to place the new content in an institutional scheme. The metadata could include, but is not limited to, items such as the content creator, the subject area, themes in the document, the potential audience, the time frame for which this information is valid, and its relation to other categories within the established taxonomy.

Needless to say, building the information architecture requires involvement from many constituents. At the center of the framework is an information taxonomy that defines information hierarchies, classifications, and relationships. The taxonomy will evolve continually. Because there is no automated way to categorize or to decide what information to keep, the taxonomy will need to be refined manually.

**What It Means to Higher Education**

In the early days of data processing, information and applications were tightly coupled; data was often embedded in the actual program. ERP systems consolidate many critical applications and store data in a single database, separate from the application systems. Database technology allowed relationships to be established among records, which identified, for example, which students attended which courses, placing data in context and outside the application. Next, data warehouses and data marts were developed to represent a simplified, aggregate view of all operational data and to serve as the source for end-user reporting. In the same way that electronic information has grown from structured data to rich media, DBMSs are being superseded by ECMSs that can manage all electronic information formats. For higher education, a comprehensive information architecture provides the roadmap for the next step in the integration of institutional resources.

**Anticipated Impacts**

An information architecture will have a profound impact on all areas of the institution. The ability to personalize instruction, to provide the right information to a colleague, and to make information-based decisions are just a few of the potential changes. Equally important is that information will be less likely to get “lost.” As a result, information can be built upon, leading to new information or knowledge; this is the concept behind a learning organization. These benefits are possible throughout the institution—in instruction, student services, research, and administration.

**Challenges of Integration**

Since most higher education institutions are highly decentralized, they will face great challenges in integrating their information assets. Faculty see themselves as owning research data and course content. Administrative units have adopted a similar culture; for example, the registrar may be perceived as the owner of student records. Higher education’s culture and values focus on independence and autonomy even though individuals depend on the institution’s information assets.
However, interdepartmental and inter-institutional initiatives are thriving. Many faculty realize the synergies created from sharing information resources. Many academic and administrative constituents are frustrated by the difficulties (or inability) to find or to access the information they need.

Institutions will need to resolve a number of questions. For example, who will determine what information is worth keeping and how to categorize it? Should building the taxonomy be a top-down or a bottom-up approach? Or, must it be a combination of both? Who will determine who has access to specific pieces of information? Who determines when it is time to archive information or destroy it? Who owns data? How can individuals be encouraged to work across organizational lines? One of the first challenges for institutions will be to create an awareness of the importance of an information architecture and craft a plan for how one might be developed.

Organizational Implications

Colleges and universities have independent information resources organizations—silos—whose primary mission is the management and dissemination of information. Traditionally the largest are the library, institutional research (IR), and information technology services (ITS).

- The library manages hardcopy and electronic information in direct support of teaching and learning. As information is increasingly available in electronic form, the library relies on the Web and other commonly available electronic information management and distribution mechanisms.

- IR manages institutional data for decision support, creates new data through surveys, and furnishes standard reports for state and federal government. The primary data source for IR is excerpts from operational data managed by the institution’s ERP system.

- The central IT organization manages real-time, operational information through the institution’s business systems. Many IT departments have created data warehouses and, like IR, furnish information for decision support. In addition, the IT organization provides and manages technology tools that facilitate the capture, storage, manipulation, and transmission of information.

All three units are in the information management “business” and employ the same or similar technology tools. Often each unit develops the information management strategies most appropriate for its type of information. However, without losing their identities or unique cultures, the library, IR, and the central IT organization can work together to develop a campus-wide information architecture. The included informational resources might span relevant and balanced collections of print and electronic materials, information maintained and managed by the administrative systems, electronic images or video, and external electronic information. By adopting a common information architecture, all the institution’s constituents can intelligently create and manage all types of content—documents, Web pages, rich media—using a campus-wide content management platform, information access, and delivery system. The outcome will be a
system that captures and maintains information and provides timely access to relevant information.

Beyond creating a common information architecture, colleges and universities might choose to create a division for information resources and technologies (IRT). The IRT Division could provide information resources and manage technologies to meet the institution’s instructional, research, and planning needs, as well as facilitate effective administrative processes.

Along with a different type of organization, the leadership role may change. Although many institutions have created CIO positions, the majority of CIOs are viewed as chief technology officers (CTOs) and are charged with the management of the technical infrastructure. However, the CIO title implies a mandate to manage the institution’s information assets. It is the CIO’s role to ensure that diverse information assets are identified, captured, categorized, and effectively managed so that relevant information is made available to the right audience at the right time, anywhere, and on any device. As more institutions develop an information architecture, more true CIOs will emerge.

Evolution of Information Resources

According to a report from Gartner,

During the next five years, digitized data will continue to surge. This data will come from many internal and external sources and be better integrated. Enterprises must have the capability to use this data for improved decision making and will need new tools to make sense of structured data (for example, in tabular form) and unstructured data (text, images, audio, or video).

Information architecture systems will aggregate the data and may also store structured data in centralized databases and unstructured data in content or media repositories. In the structured environment, Gartner foresees a continued evolution of business intelligence and data mining. In the unstructured environment, we anticipate document and content management and enterprise portal solutions evolving and also starting to merge.

Information access involving customers and employees will also become enhanced. The electronic workplace, for example, will use IT in applications that generate or access information—e.g., groupware, collaboration tools, enterprise information portals, and expert registries.²

Until institutions embrace an enterprise-wide approach to an information architecture, no automated system, however sophisticated, can manage or deliver critical information to internal and external constituents. Without an information architecture, the ECMS is likely to manage irrelevant information, and the EIP cannot deliver important and significant information.

In the same way that organizations integrated their networks, operating systems, and applications, creating an information architecture can be viewed as the next step in the evolution of information resources. People and the information they create are the

---

² Source: Gartner, “Evolution of Information Resources”
essence of any institution and a source of its competitiveness. Those institutions that are best at capturing, managing, and leveraging their information resources will be in the best competitive position. As electronic information resources expand exponentially, institutions need to implement advanced information management tools and develop an information architecture to ensure they are in control of the information that is one of higher education’s most valuable assets.

**Key Questions to Ask**

- Are faculty, staff, or administrators frustrated by their inability to access the information they need?
- Does the institution have an integrated information resources framework that brings together structured as well as unstructured data?
- Are there silos of information resources across campus?
- Does the institution use an ECMS as well as an EIP?
- Does the institution have an information resources framework that provides a methodology for determining what information should be retained, how it will be used, where it is created and stored, who serves as its custodian, how long it is relevant, and why it should be maintained?
- Would better management of information resources enhance the competitiveness of the institution?
- Would better management of information resources enhance the effectiveness of faculty, students, staff, and administrators?

**Where to Learn More**

Endnotes


About the Author

Karin Steinbrenner (ksteinbr@email.uncc.edu) is Associate Provost and CIO at the University of North Carolina at Charlotte.