Chapter 7. Information Institutions and Infrastructures

Large institutions often have complex information needs. Moreover, some institutions are devoted to supporting social information needs. Increasingly, such large institutions are closely integrated with information systems. Information systems facilitate the management (e.g., selection, organization, access, and preservation) of collections of information. These institutions include libraries and archives which serve as information resources to the community, ensuring trustworthiness and accessibility. In addition, we consider system development procedures and metrics.

Architecture implies coordination of physical resources and services. We can characterize the types of services provided by information systems. A set of coordinated information services compose an information system. We have already examined databases (3.9.0) and here we consider other sets of information services such as libraries and archives. The details of how the services work together compose the system's architecture. Services are the building blocks that compose the information system. They support the basic functions that are required to run the system. The services themselves may be defined by user needs (desired functionality, for example) or by the system needs (operating programs), and may range from content management to end-user services.

7.1. Information Institutions

7.1.1. Types of Information Institutions

Here, we consider some categories of information institutions in detail: libraries, archives, museums, as well as the related scenario of information management in business and other dynamic organizational settings. They can be examined as sets of services and they can be implemented with service-oriented architectures. Organizational and institutional connections. Common functionality of information systems. Many of the traditional distinctions are becoming blurred. s Many digital collections are established with the policies common for traditional institutions. Different traditions based on the purpose of the institution. Many pieces have to work together. Content, tasks, people, tools. In the next three sections we consider libraries, enterprise architectures, and archives. They are so well established that they are associated with established models for knowledge institutions. But, they can also be broken down into sets of services. they are still socio-technical systems and, hence, they help to define an ecology of content, use practices, and technologies. Nature of institutions (5.8.1). Institutions need to follow procedures and policies. Institutions have a need for continuity and stability. Types of institutions: Memory, Educational, Cultural, Research, and Knowledge. Moreover, we routinely distinguish among sub-categories of each of these, though these distinctions can be fuzzy. In addition, there are many ad hoc institutions such as for the management of data sets.

Management of Information Institutions. Management (8.11.0). Library and archive management. Because of disruptive technologies, many information institutions are in flux. But there are also many new opportunities.

7.1.2. Information Services and Institutional Practice

Institutions generally implement specific formalized services. These are traditional sets of services, but increasingly other sets of services are being defined. Increasingly, there are media intensive applications such as films from Major League Baseball. Science data sets and data libraries (9.6.3). PIM (4.11.0). Digital Asset Management systems which are generally extensions of database management systems. Assets of an organization may be kept in a digital asset management system. Supporting branding. Media libraries. Personal information management (4.11.0). Information commons. Personal archives. Possibly developed with use cases (3.10.2). Project management (8.11.3) for the infrastructure. Services could be implemented on a SOA. Creativity and collaboration. Web publishing platforms. Moreover, institutions need to include evaluation of their services.

Selection	Deciding what content to include.
Organization	The content should be described clearly. and structured to make materials easy to find.
Access	It should be available for users and facilitate collaboration.
Persistence	It should be available into the future.

7.1.3. The Nature of Collections

Collections allow rapid access to sets of related materials. For instance, they allow botanical research to identify new species of plants by comparing specimens to examples of other species. Coverage of a given topic.

Figure 7.1: Curiosity box. (check permissions)

Many motives for collecting. Difficulty of defining digital collections. Personal collections (4.11.1). Developing a collection for education and for reference. Formal collections require thought and effort and large collections need to be supported by an institutional framework.

Libraries are traditionally based around collections of published materials but with the advent of the web the boundaries of collections are much more porous. Collections are still important, they are less distinct when a lot of material can be presented remotely and at relatively low cost. One possibility is to redefine the concept of collection to consider it broadly as the information context rather than as a specific set of information resources [?]. Still, there are notable exceptions such as collections which are distinctive or exhaustive. Collecting for evidence of transactions.

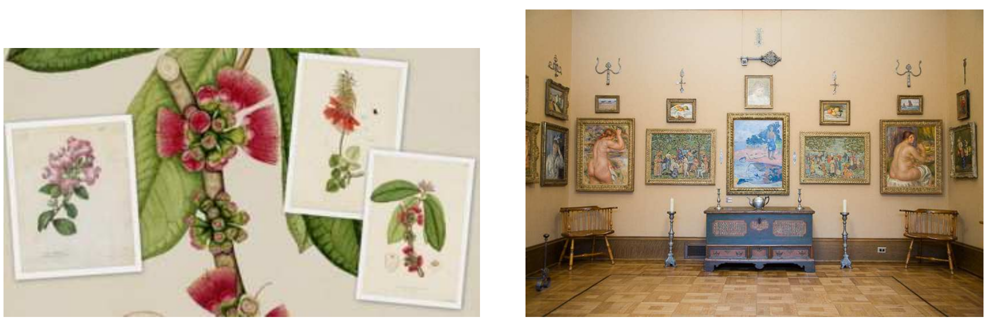

Figure 7.2: Collections are created for many purposes. Left: A herbarium data sheet from a botanical reference collection. Right: The Barnes Collection is a famous art collection in Philadelphia. Its presentation is based on the unique theories of Mr. Barnes for combining color and shape. (check permissions)

7.2. Libraries and Library Services

7.2.1. Libraries

Formally, a library is an institution that manages one or more collections of information resources.

Many types of libraries (**??**). Special collections.

There are several ways to view libraries: a community resource, a set of services, and as a collection of collections. Clearly, libraries are rich and complex knowledge institution (8.13.2), Traditionally, libraries have provided access to formally published materials. This can be an educational mission. Technology is greatly changing the distribution of books.

The very nature of collections is changing [?] since it is now possible to obtain digital copies of books. eCopies. Relationship of libraries and publishers. Libraries and publishers traditionally have had a symbiotic relationship.

One view is if libraries are community information centers and that the role of librarians is to facilitate management and sharing of information about a community. Librarians as experts in supporting the knowledge creation process. Libraries relationship to publishers (8.13.4).

Figure 7.3: Types of libraries: A) Public library, B) Hathi Trust, C) Rare books. D) Story hour. (check permissions)

Some other common types of libraries include law libraries, business libraries, map libraries, and music libraries. These are often known as special libraries. Academic libraries serve universities.

Libraries collect and manage content, Primarily, libraries manage collections of information resources. This is accomplished by a set of services. We will examine library services in more detail later in the context of digital repositories. The table gives a simple set services for content management. Physical libraries are changing; increasingly, physical space is used for research interaction and also for reflection. Libraries can be seen as places which provide reliable information for a group of constitutions regardless of how that information is organized. Access genres (3.2.2) and catalogs (2.4.3). Libraries as places where people facilitate specific type of interaction with information. Academic libraries in universities (8.13.2) and (9.1.1).

Beyond collection development by libraries have an active role in providing services which support the curriculum and educate the public. Libraries for scholarship (9.0.0) and learning.

Integrated library system (ILS). Platform.

Library Business Models

Several business models for libraries. Public libraries. Cost-center or profit center. Subscription library especially for eBooks. Digital lending rights. Interdependence of libraries and publishers. Value of library services and budgeting[57],[59]. Information economics (8.13.3). Increasingly, collections are being maintained by publishers. Though, in some cases, universities are trying to maintain a collection of their own intellectual products.

Library Customer Service and Service Populations

Libraries and customer service. It should also be built and maintained for a specific service population, whether that be young adult readers (as in a middle-school library), senior citizens (as in a retirement community library), or a cross-section of the population as in a general community library.

Information needs and approaches for youth. Youth services.

Libraries are developed for many purposes and there are many models, digital libraries being just one. Libraries may be grouped into categories such as community libraries, research libraries, special libraries, and university libraries. These categories are based on a library's user groups, on the type of collection it houses or on the specific types of services it provides.

The main function of libraries is to serve patron information needs. Thus, we should measure the library's effectiveness at doing that [44]. But, we also know libraries as cultural organizations (5.9.3). and, indeed, many physical libraries also serve as community education centers [65]. Most traditional libraries facilitate collaboration. Focusing on collecting information versus what's popular.

7.2.2. Library Services and Collection Management

We considered reference service earlier (3.3.0)and here we consider collection services. Libraries are sometimes described as a collection of collections. One of the changes for digital libraries is that the nature of collections is becoming less distinct as digital copies of texts are available. We have already discussed bibliographic metadata (2.4.0).

Workflows for collection managment.

For many tasks it is difficult to tell what type of information will be needed. Therefore, library collections emphasize a wide range of perspectives on a topic. Information resources and scholarship are continually evolving. The collection needs to evolve as new content becomes available and the needs of users change (Fig. 7.4). They facilitate information and data sharing by providing one or more collections (7.1.3) of information resources and associated services. Collection management can include establishing collection policies which define what types of materials should be acquired for the collection and what materials should be removed. The basic collection management services can be summarized with the acronym S.O.A.P. which stands for Selection, Organization, Access, Persistence. Focusing on the services which form he basis of a library can be helpful when developing services for a digital repository (7.8.1).

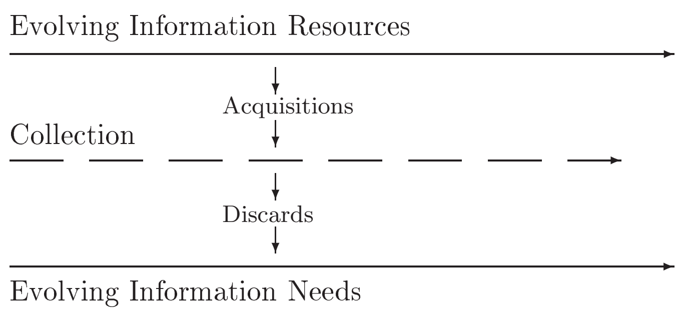

Figure 7.4: A collection is a set of selected information resources. Both the of information resources which the collection represents and the information needs of users evolve through time (left-to-right); thus, the collection also has to evolve through time. The collection policies define what types of materials should be acquired for the collection and what materials should be discarded. (redraw)

A library collection could provide a broad perspective on a given topic. Building a good collection [43] may include a variety of perspectives on a given issue. Special collections, which are used is special libraries (7.2.1)need to design their policies to match the requirements of the institution is which they are based.

Collection development. Determining what's missing in a collection and attempt to complete it. Collection development focuses on building and enhancing a collection. A collection development policy can be created by analyzing the stakeholders and the needs of likely users. Log files can be useful in filling gaps in a digital collection by showing what a person is looking for. Basically, this can be thought of as another way to predict what is likely to be helpful to users of the library. For digital collections this means determining the supplement to the information environment available to patrons.

Libraries must select what information to store in their collections. Many books, journals, reports, and other publications are produced each year, and librarians must determine which acquisitions will best serve the information needs of the library's users. An adequate collection provides the possible coverage across a given topic, gives people the broadest possible perspective on information. The library's collection level and commitment goal are part of its stated policy. A "research level" goal of the collection is described in Fig. 7.5. This means that the collection is intended to have resources available for serious research. Other policies might highlight other collection goals, such as increasing circulation statistics, securing privacy, limiting information access by children (8.1.0), or promoting "community standards". Although a given collection may change slowly, all collections become stale. Periodically, it is desirable to eliminate, or "cull," items that are no longer useful. Here again, a collection policy might state the basis for this culling such as that the materials are outdated or that they simply are being used. This is a re-appraisal or re-selection.

> The Berkeley Digital Library SunSITE collects... at the research level in the areas of digital library development and information technology. In other subject areas, the resources and projects we select are reviewed by the appropriate selector for that discipline and gauged against our existing collecting levels for print publications in that area.

Figure 7.5: A brief collection policy.[5]. (check permission)

We can ask how well an actual collection matches to ideal standards for a collection. Measuring the quality of a collection. Such as whether it covers the topics specified in the collection policy. Scope of the collections[89]. We have discussed repository metadata (7.8.4). Increasingly, we think of objects as multiple levels. An entire collection can have metadata; this would include the collection name and its scope. Indeed, the metadata should reflect the collection policy. Conspectus.

Digital libraries can be service-oriented architectures with repositories. Virtual repositories. Library workflow. Complex digital object (7.8.4). Hathi Trust (10.1.6). Providing services such as a pipeline model for decomposing a complex problem into manageable pieces (3.7.1).

7.3. Organizational Information Systems and Knowledge Management

Organizations have several types of information needs. Earlier, we considered organizational structures. Here, we consider the use of information.

7.3.1. Knowledge Management and Organizational Information Systems

How to fit information systems to match the complexity of organizations. We have, for instance, considered DSS (3.4.2). Special libraries. EDMS.

Knowledge managing knowledge assets and databases. Knowledge management for models of formal inferences via the Semantic Web. Knowledge management and internal corporate communication [**?**]. Knowledge management in medical settings (9.9.3). This involves systems, users, and content. Avoids redundancy in an organization.

Unlike libraries enterprise content is an ongoing part of business activities. Mining organizational data. Active information which is more directly task oriented. Management information systems (7.3.2). Records management. Business classification systems.

Describing how documents in a work process are related to each other. Documents show work processes. Modeling Roles and Activities. Indexing and retrieval of these documents.

Creating knowledge. Sharing knowledge across silos. Organizing knowledge. Disseminating knowledge. Tacit and explicit knowledge.

The context of archival resources help us interpret them. There are many difficulties such as where in the organization the knowledge is found. Unlike libraries which collect available resources, information in an organization needs to be more systematically assembled and managed. These active information repositories can be a dynamic record of a community's activities and thoughts. They may also have special services such as information brokering and organizational memories. The effective use of information and knowledge can help an organization increase its efficiency and accomplish its goals.

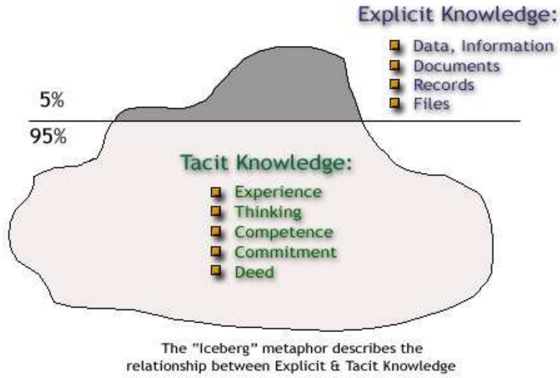

Figure 7.6: Tacit knowledge. (new figure showing person) (redraw).

Knowledge in organizations is kept in many forms not just to documents. Information processes can improve the efficiency of business and organizations (8.8.1)[92] but also introduces difficulties.

Organizational Knowledge Management

Like individuals, organizations need information to make decisions. We considered organizations themselves earlier (5.7.0); here we focus on information within organizations. Knowledge management tries to make sure the right information is available when making a decision. There are many types of information in organizations. Knowledge management focuses on effectively collecting and organizing the information that is used in organizations. Information in an organization can be very fluid, and organizations use different types of information systems than those utilized by libraries. Thus, knowledge management often includes capturing and distilling organizational knowledge. An organization may not maintain clearly defined collections as do libraries, but simply sets of documents that are frequently updated. Useful in the design of information systems (7.9.1). Collaboration. Teams.

7.3.2. Organizational Information Needs and Use

We have described organizations earlier (5.7.0); here we explore them in more detail. Internal and external information needs. Organizations possess large stores of both formal and informal information. However, this knowledge must be codified for capture in a knowledge management system. Typically, this capturing process must occur both at the front and back end of an organization's system. New knowledge must be incorporated into the overall system at the same time that old information is reformatted for consistency. There are many areas of application for information in an organization and, indeed, there may have to be several different information systems.

Efforts to make sense of what is really happening in an organization's environment [86]. Like more general sensemaking (3.1.1).

Managers have particular needs for information and information management. Indeed, Management information systems (MIS) are designed to provide managers information environments. Decision support systems (3.4.2). Enterprise Resource Planning system and a lot of data available. These generally keep active business document and records which are then transferred to records management systems (7.4.1) as those documents and records become less actively used for the purpose for which they were created.

Organizations often have individual or departmental information systems based on specified roles. But, the larger question is whether or not the roles or departments themselves are designed to handle the flow of information as it actually happens, and not as it is supposed to happen. It is an enormous challenge to design, or redesign, an organization to handle the flow of information effectively for routine activities (8.11.2).

Task specification in an organization may be described as workflow (5.7.2). The structure of a group and its tools has an effect on the way that information is used and transmitted. Roles such as "manager" are assigned certain tasks within the hierarchy of an organization and they are often given procedures for completing those tasks. These structures and procedures either facilitate or inhibit information flow. Genres of organizational communication. Regularity of organizational processes. Organizational grammars. Organizational communication genres.

In many organizations, there are groups whose main purpose is managing information relevant to the organization. The Chief Information Officer (CIO) is typically responsible for this at the organizational level, and will manage several other departmental information officers.

An individual functions in an organization and has his/her own perspective on it. Individuals may mental models of how thing work in the organization. These may include not only formal specification but also of the human interactions. People try to determine what is really going on in the organization. Related to structuration. Ultimately, the information system needs to meet the organization's information needs.

7.3.3. Supporting Organizational Culture, Planning, and Change

Organizations need management. While it may seem obvious that information processing plays an important part in the success of any organization, effective information management is necessary. Moreover, there are often communication channels that supplement these the formal channels. Formal and informal models of information flow in organizations. Organizations often do not follow their own formal policies.

The organizational culture and the manner in which information is handled are often counterproductive to efficient information systems. Also, because organizations necessarily have a mix of goals, information often gets distorted when being transferred from one location within an organization to another[13].

Figure 7.7: Many organization do not survive change. Here are some logos of organizations which no longer exist.

There are many types of organizations and many strategies for meeting organizational goals. Organizations have a distinctive style, ranging from regimented, button-down conservative organizations to the more open dot-com era organizations of the past twenty years. One of the most indicative elements of organizational culture is often how information is shared among different levels and departments.

This sharing can be the result of many factors. The expectations of the upper-level managers and officials regarding an organization's culture, financial goals, productivity, and public image is one means by which organizational culture is created. Another is the type of people who populate an organization, which is the result of the industry in which the organization is located, its geographical location, and the hiring practices (7.3.6). Problem of resistance to information sharing.

Organizational culture has many ramifications. It can both cause and reduce inter-organizational power struggles, it can stimulate and stifle the sharing of ideas, and it often determines what people actually know about the organization of which they are a part. The extent to which the management shows a willingness to hear contrary opinions. The willingness to learn from bad news allows the organization to respond to that bad news[3]. Information systems need to be compatible with the organizational culture. Open discussion forums may be appropriate for entrepreneurial and research organizations while more secure communications are appropriate for financial organizations. Change is inevitable and it has been particularly common in this era of disruptive technologies.

Strategic planning seeks to utilize all of the internal strategic knowledge within a company to come up with a plan for company competitive advantage. An organization may explore how to collect information. A SWOT (Strength, Weakness, Opportunity, Threat) analysis maps possible activities on the four dimensions of a graph. Like affinity diagrams, SWOT diagrams can be a GDSS tool (3.4.3). The horizontal axis runs from weaknesses to strengths, and the vertical axis runs from threats to opportunities (Fig. 7.8).

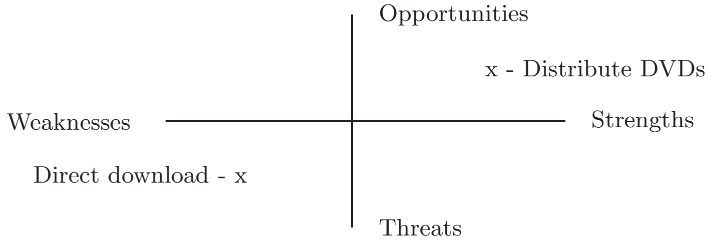

Figure 7.8: Example of a SWOT analysis for the video-store example.

7.3.4. Tacit Organizational Knowledge, Organizational Memories, and Lessons Learned

Tacit Organizational Knowledge

Tacit knowledge is difficult to capture. It includes know-how and the knowledge employees have about "the way things actually work". It is the context-laden, implicit knowledge that people have about their work and its procedures, which people follow often without being able to articulate their reasoning for it. If the tacit information possessed within organizations could be captured systematically, it could be put to many practical uses, such as training new employees, as well as simply being a vast source of useful information. Capturing this information creates an "organizational memory," which could be invaluable if the staff changes, for example. Tacit knowledge can sometimes be captured with improved procedures. Knowledge creation from tacit knowledge[68].

In other cases, it is inherently ill-defined. Artifacts such as jotted notes and emails can be physical examples of this tacit knowledge, however they are generally not satisfactory for creating organizational memories because they lack sufficient context [25]. A variety of techniques have been proposed for capturing tacit knowledge organization members and other "experts"; the techniques are generally known as "elicitation" or "extraction". Another strategy explores informal workplace interaction. Dialog and task description (6.4.0). Formalize tacit knowledge. For instance, an internal Wiki allows sharing and discussion of ideas.

Bureaucratic Organizations versus Learning Organizations

Organizations must find a balance between being rigidity and flexibility. Some organizations tend to become bureaucracies — they adopt routinized practices[84]. Having a strict routine can be very efficient when the environment is highly predictable. However, it is very difficult for such organizations to adapt to new conditions. Disruptive changes [42]. When the environment in which an organization operates changes, and all environments eventually do change, the organization needs to adapt. Knowledge capture and reflection: after-action review, mentoring, communities of practice.

Structured organizations often develop an orthodoxy which makes them slow to change. Making traditional organizations more adaptive. Encouraging organizational innovation. Learning in organizations has been characterized as involving a double loop (Fig. 7.9), that is there can be adaptation within a given structure or the structure itself can change. Some organizations are so flexible that they can make these loop changes seamlessly, but those organizations also often have high overhead.

Records and data warehouse (7.4.4). Sales. CRM (8.12.5).

Focusing on quality rather than immediate costs may result in a long-term reduction of costs. Mature businesses may focus on Six Sigma[21] efficiency and TQM[47]. Quality improvement can drive customer

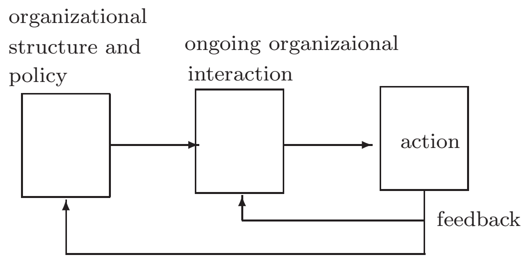

Figure 7.9: In most cases, an organization adapts to its environment but sometimes the organization itself needs to be changed. This is an example of double-loop learning[28].

focus and process improvement.

Organizational Memories and Lessons Learned

A learning organization collect information about itself which can be used in reflection [76]. There is often resistance to knowledge sharing in an organization. Adaptive organizations have a feedback mechanism in which the experiences and knowledge of organization members is displayed throughout the organization as a whole. These are learning organizations[28], which work by changing the representation and processing of information from an individual perspective to a community perspective. A learning organization typically learns to process information more efficiently.

Mechanisms for learning facilitating learning by organizations. Representations of organizational knowledge. Attempts to create learning organizations have sometimes focused on lessons-learned, which are attempts to summarize what was learned by different projects. It is necessary to describe these lessons by metadata, however, and there is a problem of context in the search field. What exactly were the "lessons-learned," for a given project, and how are they best described. Furthermore, because tacit knowledge of the lessons-learned variety is typically internalized by individuals, it is a social artifact and may reflect the social perspectives and demands of the individual who observed or created them and do not apply easily to other individuals. Lessons learned can be considered a type of failure analysis. Probably the most effective way to capture these lessons is in policies of organizational procedures. Science (9.2.0). Math (9.7.0).

7.3.5. Information about the World Outside the Organization: Environmental Scanning, Business Intelligence, and Competitive Intelligence (CI)

Organizations are affected by the environment beyond their boundaries. Therefore, it is essential to understand the environmental trends that may affect an organization or an industry. This knowledge can to determine business model and strategy. Environmental trends do not simply mean the natural environment — an organization's environment consists of all the variables that affect its existence in the world, including regulatory and political considerations, market forces, technological innovations, worker relations, competitors, and innumerable other factors. Developing a sound strategy necessarily involves taking stock of these trends as well as the position that competitors are taking in response to them. The more complete a picture an organization has of its environment, particularly the place its competitors occupy in that environment, the less uncertainty it will face and the better strategies it can devise. One method of reducing uncertainty is to increase environmental scanning. This might include observing how other businesses are dealing with the uncertain environment. Normally an environmental scan would use public resources but these analysis techniques such as data mining can also be applied. Asking the right question is a challenge. Intelligence analysis. (sec:intelligenceanalysis). Potentially, this can be captured and stored in model such as those employed by DSSs and MIS environments.

Business intelligence report identifying trend in computing and evaluating their potential for disruption.

An important aspect of any business is a thorough understanding of competitors' strategies and capabilities; this is known as "competitive intelligence" (CI). Two companies in the same field may have

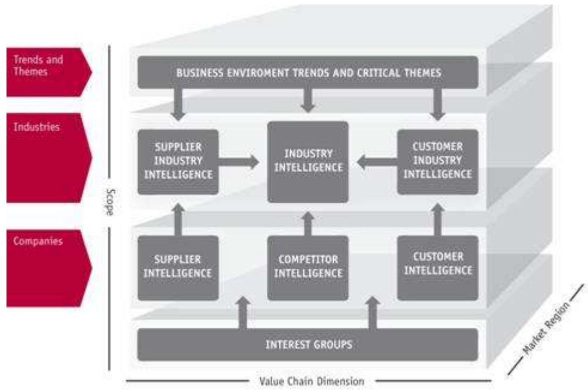

Figure 7.10: Competitive intelligence. (check permissions) (redraw)

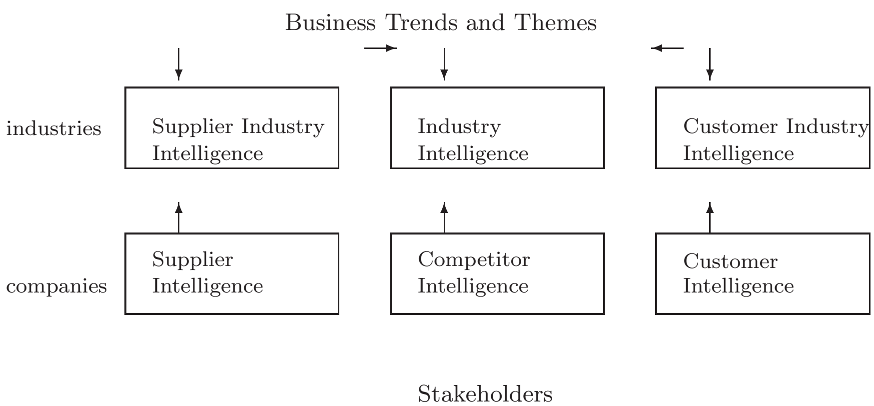

Figure 7.11: Competitive intelligence requires intelligence at several levels including broad trends, developments within industries, and specific companies. (check permission)

very different strategies and different business information needs. Collecting information on each other may show that one has a better strategy, which could force the other to mimic it or be outperformed. Understanding the context in which organizations operate is essential to understanding gathered information. Information is valuable only if it is known how to use it. Competitive intelligence ranges in scope from researching a competitor's publicly available company profile to developing a computer model of the competitor using insider tips. More than simply seeking any information about a competitor, CI seeks to answer the question "what types of information would be most useful," and attempts to acquire it. understanding the competitive advantage can help formulate a business strategy.

Forming a mental model of the competition. An analysis of competitors could include the way they are reacting to the environment. A further level analysis would examine the competitors' competitive strategy. That could employ game theory (3.4.1). Developing models, such as financial models (8.11.4) of how competitors are responding to competitive threats.

Ethical principles should always be applied in information collection. There are permitted resources as well as illegal ones. CI is not business espionage — it is simply a means of collecting, analyzing, and using publicly available information to form organizational strategy. Others do not always play by the rules, and information security is needed to prevent business espionage (7.10.3).

Due diligence about a company. Voluntarily released information for instance before entering into a contract. Checking personal information but also confidential.

7.3.6. Organizational Information Management Systems

An Organization's Strategic Knowledge

The most important information in organizations can be called "strategic knowledge". What does the organization know and what does it need to know in order to facilitate decision making. Having a top-to-bottom understanding of a company, with solid information, can help in reaching effective decisions. What is an organization all about? Understanding an organization's niche is central to sound decision making. Does it make sense for an organization to both make medical instruments and run a fast food restaurants? An organization's core competencies (8.11.2) are its foundation.

Information management can provide a competitive advantage for an organization. Information about a company could be contained in formal documents, but could also be the situated knowledge of several key employees. Dispersing that knowledge throughout the company creates the conditions for informed strategic planning. It is important to identify the bottlenecks that prevent the flow of information need to identified and removed.

Business processes (8.11.2).

Strategic management seeks to combine and organize all of the necessary requirements for maneuvering a company to a better fiscal position. This involves much in the way of decision making, which in turn, requires excellent information. Specific systems for processing information in an organization can be instituted to facilitate strategic management processes, such as decision support systems (3.4.2). These seek to utilize all of the internal strategic knowledge that is available within a company and make it centrally available for use. Indeed, this can lead to an electronic records effort.

Enterprise Architectures

Describe enterprise-wide data and information management. This might be implemented with a service-oriented-architecture as a service-oriented enterprise (SOE). Business reference models. Common information model. Office information systems. Roles for managing access and permissions. Emphasis on business processes but remember the perspective of practical action (3.5.1). Paper is still common in offices and photocopying is often critical to information management in offices.

Organizational Information Structure: Silos and Stovepipes Organizations need to balance the control of information with the advantages of the free exchange of information. Sharing Information Resources within a Knowledge Community. Many organizations keep information for different departments or divisions in separate locations. These different locations can easily become over compartmentalized and turn into "silos" or "stovepipes" Sometimes there are good reasons for carefully managing data, such as security and organizational culture. On the other hand, excessive protection can lead to poor communication and lack of information where it is needed. However, there has to be a balance between the need for security and knowledge. Data management issues (9.6.3). Organizations may stake out turf by trying to control how language is used. Beyond boundary spanning.

> It is well established that United States' intelligence gathering capabilities are severely hampered by the lack of information sharing between agencies and further weakened by the absence of an adequate system to analyze the intelligence they collect.

Figure 7.12: Testimony about the problem of "silos" in the US government at the time of the 9/11 attacks[30].

While one of the primary functions for organizations is the effective coordination of information, aspects of organization that sometimes lead to the distortion of information. There are many impediments to information flow within an organization. There are institutional procedures that promote information security that prevent communication between various departments; there are individuals who hoard information out of contempt for others, or because they do not understand either the need or the technology of the information system. Sometimes individuals or departments have conflicting goals and seek their own achievement at the expense of another individual or department, or fear the reverse. Indeed, organizational players can take advantage of these tendencies[92].

Document Management Systems

Document management systems are information systems that are designed to be closely related to the tasks of knowledge workers. These area also closely linked to Electronic Records Systems. Electronic document and records management systems (EDRMS). Combined with records management (7.4.1). Document management focuses on providing knowledge workers access to a variety of materials that are related to current work. Document management systems are like digital repositories, though they are often more interactive than ordinary repositories. They may include versioning capabilities, allowing documents to be time and date stamped to determine whether they are the final iteration of the ideas they represent. Document management to support the entire document lifecycle (1.4.1). For instance, after the active use of a document is completed, it may be "locked down" in the repository so it can be viewed but not revised or deleted. Office information systems. Tools and environments for sharing information resources. Chat services within an organization. Records (7.4.1).

Roles. Permissions. Workflows.

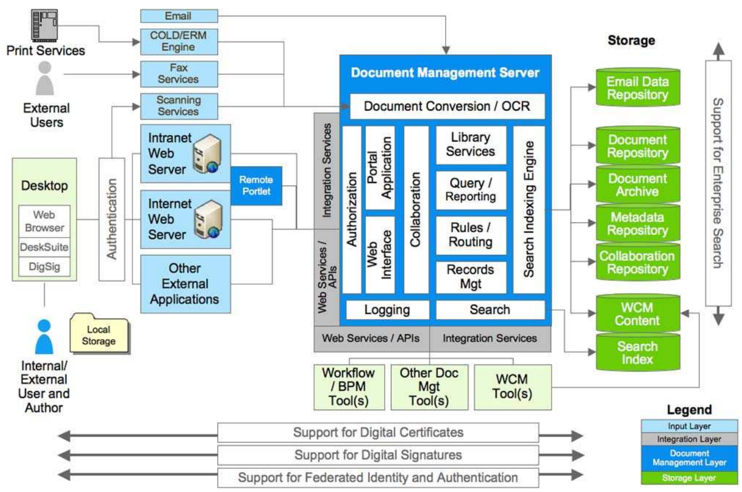

Figure 7.13: Organizational document management system. (check permission)

Enterprise Content Management

Keeping the information systems consistent in a very large, distributed, or dynamic organization can be difficult. Enterprise content management systems are total systems designed to manage all the information associated with any given enterprise. They control and track the information flow in organizations and make knowledge available where it is needed.

In addition to organizing and storing the information contained in an organization, enterprise content management systems are also designed to control the flow of information and put in the hands of those who need it. Taxonomies for organizing enterprise information. This many be use because almost every business has a different set of attributes.

topics

Figure 7.14: Organizational thesaurus.

Accurate descriptions of the resources make the content model useful. Capturing, coordinating, and updating organizational ontologies. Finding organizational information with enterprise search (10.11.1). Many indexing systems have been developed for business and legal applications. Semantic Web and ontologies for organizing information. Document management systems. Specifying an organization's information architecture. Enterprise classification, metadata, and taxonomies. From content management to eCommerce. Also need to manage access permissions. Thesauri are common for describing enterprise content. (2.2.2). Models based on organizational contexts. Enterprise applications. Composable applications.

In many cases, enterprise information may not be well enough defined that it can be easily managed (7.3.4). The management of information should aim to make information available when and here it is needed in an organization. That is, the information value chain [9], which we introduced back in the first chapter, needs to be managed. By streamlining organizational processes and facilitating communication, knowledge management systems can also produce better decisions, and thereby reduce risk. Risk is inherent in business, and acts as a constant. Institutions are often risk averse. Knowledge management systems that allow the sharing of information across a wide spectrum of organizational operating arenas produce an information environment in which better decisions can be made. By knowing the downstream consequences of actions, risk can be managed.

In some cases, especially for some businesses, information resources have value. Publishing and management of information as an asset (8.13.1). Information is an asset of an organization. Because information has value, it is worthwhile spending capital on it. Enterprise level risk analysis and assessment.

Procedures for information security in organizations. Part of organizational roles and norms. DSS (3.4.2). Intranets (Fig. 7.15).

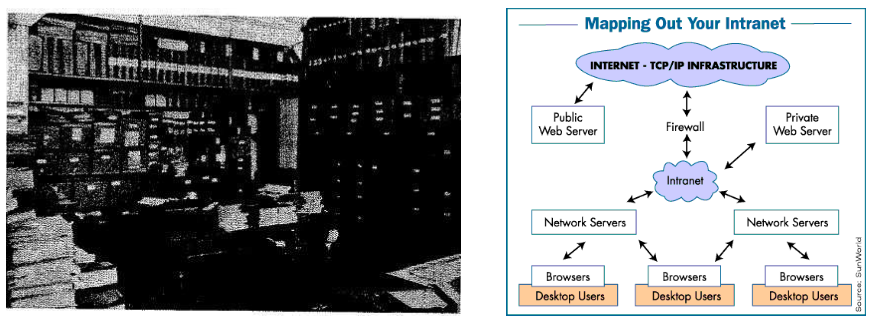

Figure 7.15: Internal information about an organization (right). These days, the information would be kept on the organization's intranet (right). (check permission) (redraw)

Policies are needed for managing corporate information such as employee email.

Impact of Information Systems in Organizations

In every model, however, information, about the company in question and its competitors, plays a central role. The idea that knowledge management systems offer a competitive advantage over other companies is itself not without debate, however. It has been argued that while a lack of any sort of information technology system is a disadvantage to a given company, knowledge management systems do not offer a specific advantage in today's world because the technology for them is almost ubiquitous among all companies [39]. Business process re-engineering (8.11.2)when business conditions change.

Information map. Knowing where to find information in an organization. Overview of an entire organization's information products.

Changing organizational culture is not an easy task. Often, the procedures and outlooks of an organization are firmly entrenched in the mindset of all members. However, because organizational culture is at least partially a result of business practices and operations, often the first step toward cultural changes is to change the way an organization is run. This is sometimes described as un-freeze, move, and re-freeze an organization's practices. Learning organizations (7.3.4). Metrics for knowledge management.

Cost-benefit analysis (7.10.3). Impact of computerization on organizations[75].

Organizational transitions. Mergers, acquisitions, change of management. Difficulty of changing organizational culture.

7.4. Records and Records Management

A record describes a transaction. There may be records for government, businesses, and even individuals. Because it describes a transaction, a record can be very useful as evidence. Thus, it is important to ask how should we manage the record so that it remains as persuasive evidence.

In some areas, records overlap significantly with archives which are considered below (7.5.1). In addition, Records also overlap with the management of data sets (9.6.0). Good records management is an essential aspect of accountability and transparency.

While there is a lot of variability in the way the terms are used, we may say that records management is concerned with the information while an archive is concerned with keeping the physical or perhaps digital manifestation of the transaction. In addition, business records management is closely related to knowledge management (2.2.2). Knowledge management often focuses on supporting the information needs of an organization while records management emphasizes the storage and re-use of records. Obviously, these are often highly interrelated.

There are generally many legal requirements for formal records. For instance, they should managed in a way that they can provide legal evidence (8.5.4). Reliable records in laboratory notebooks are needed for patent claims (9.6.3). Guaranteeing authenticity. Fixity is the requirement that a document retrieved from an information system is identical to the original. Properties of record-keeping systems. Information security (7.10.3). Distinguish between keeping the document and keeping the record. Data governance.

7.4.1. Types of Records

Records are evidence of events or transactions by organizations and governments. The functional goal is accountability. Typically, they describe administrative activities. In a business, records can include invoices, shipping tickets, payroll accounts, and database transactions. Many formal records management systems focus on business records and often don't consider other types of records.

Records found in other areas such as medicine and government often have specific policies and systems. In medicine, patient records are essential for effective treatment (9.9.3). Documents such as marriage licenses, birth certificates, immigration records, property deeds, and government records such as budgets and laws, as well as transactions such as real-estate sales, need to be open for public inspection. Family records. Records of natural phenomena such as weather. Scientific data sets can also be considered a type of record. The concept of cultural records is closely related to the topics addressed in archives.

Records may be seen permeate modern society. This are somewhat closer to archival records. Web archives as electronic records. Such records may include: QUOTE emails, faxes, spreadsheets, databases, maps and plans, samples and objects, information in business systems, letters, text messages, minutes, policy and briefing papers, photographs, research data, social media sites.
locate emails, documents or information when needed, reuse valuable work that you or someone has done in the past, determine the most recent version of a document, produce evidence as to why a particular decision was made, protect yourself, your clients, citizens and the Government, support cultural, social and historical values, helping future generations understand history, society, culture and people. ENDQUOTE

Genealogy. Family.

Special requirements for some types of records.

Authenticity, reliability, integrity, usability.

Figure 7.16: Birth certificate. (check permission)

7.4.2. Records as Evidence of Transactions Relevant to Organizational Current Needs

Presumably, records will provide accurate information. Indeed even more strongly, records many need to provide legal evidence. There are often definite legal requirements to keep records. Records and lifecycle management. Fixed form, stable content, stable media, archival bond. Social function of records to provide continuity across time. Evidence versus legal evidence (8.5.4).

The management of these records is essential for the effective functioning of society. Archived records of property deeds are essential to property rights (8.7.1). Government accountability is greatly facilitated by public access to government records. Government open data. As discussed with regard to knowledge management, the ability of individuals to view the records and rationales for government actions is an important element of creating an informed citizenry. Archives are also important as evidence and for legal records. Indeed, the roles of Archives and Electronic Records management are closely related. Archives tend to emphasize the memory function of the collection but that is not a boundary. Furthermore, Archives tend to maintain documents which electronic records systems keep database like records. Records are a foundation for bureaucracy-oriented organizations and institutions.

Formal records are saved intentionally and they should be distinguished from other types of organized information. Records in organizations may be required to demonstrate compliance with regulations such as accounting records and standards and they often have a legal mandate.

To be useful they must have certain properties. Digital provenance. Accurate records are integral to any organization, large or small. A record is authentic to the extent that it accurately reflects the past.

7.4.3. Records Management

Records systems are more than databases management systems, because they include organizational policies. The policies and technique needed to ensure the integrity of records – that is, in their role for providing evidence. Security concerns. Ensure that the features of records shown in Fig. 7.17 are implemented. Not only are records need to be managed but non-records – those information resources which are not to be saved also need to be properly managed. This highlights the problems of using simple backup tapes as storage – they keep a lot of material beyond the intended records. A records management protocol has to manage: Legal holds in records require that they cannot be deleted because of a court injunction. Records generally intentionally destroyed when the legal requirement expires. Data preservation issues.

We need to have confidence that the records have not been altered or corrupted. Indeed, the legal admissibility of electronic records as evidence in a court depends on such confidence. Reliability can be enhanced with transparent and routine processes for record development and handling. best practice guidelines could formalize the procedure. Records could always be: created (a set of procedures for what constitutes a recordable action/event), created at the same point in the process, stored in the same place, maintained in the same way, and accessed by the same people for specific purposes. The need for trust in the records is similar to archives which we consider in Section 7.5.1. Procedures include regular audits (7.10.2), record security (7.10.3) and preservation policies. More broadly, the organization also needs a supporting culture for information management [**?**], an information culture.

Records are essential for complex social and organizational structure. Records management policies for businesses should follow the business processes and business rules (8.11.1). They must also follow legal requirements. A proposed set of functionalities, which indeed, could be implemented as services, are shown in Fig. 7.17. The attributes of effective records include attributes of the organization, of the system, and of the content. Ultimately, this can become a type of metadata. Functional requirements (7.9.1). Functional descriptions may minimize the need for subjective archival descriptions. Scenarios as an alternative to functional requirements.

Conscientious Organization
  Compliant (1)
Accountable Recordkeeping System
  Responsible: Assigned (2), Documented (3)
  Implemented (4)
  Consistent (5)
Functional Records
  Comprehensive (6)
  Identifiable (7)
   Complete: Accurate (8), Understandable (9), Meaningful (10)
  Authorized (11)
   Preserved, Inviolate (12), Coherent (13), Auditable (14)
  Removable (15)
  Exportable (16)
   Accessible: Available (17), Renderable (18), Evidential (19), Redactable (20)

Figure 7.17: Functional requirements for authenticity. Records should link back to the business processes which generated them. Attributes of an effective business records management program[32].

Records are live documents which are part of the company's activities.

Automatic capture of metadata upon creation. In practice these policies are summarized by ARMA GARP. Record keeping and economic development. Records serve a social function. Court records. Metadata for permission to access.

Edit and audit rules. Records audit and procedural audit. Transactional database. Record-keeping metadata.

Electronic mail which is required for ongoing legal, fiscal, audit, administrative or operational purposes must be converted to the most suitable storage medium for retention (e.g., bond paper, microform or electronic format). These materials must be classified and scheduled according to ARCS or ORCS (Operational Records Classification System). They are not classified as transitory records.

Figure 7.18: British Columbia government retention policy on treating email as electronic records[4].

It is, at times, necessary to make certain documents and records, particularly those involving government agencies, sensitive or top-secret. However, decisions such as those made by government agencies may later be disputed, and records of how those decisions were reached must be kept as part of the public record and made available publicly when the necessity of their secrecy has passed. The "Freedom of Information Act," or FOIA in the U.S. allows individuals to make specific requests for information from any U.S. federal agency, and these agencies must honor the requests as long as they do not violate particular amendments of the law. However, there are generally some limits of Freedom of Information Act releases such as national security.

Freedom of information and privacy protection (FIPP). Records and classification of content. There are several special applications of archives. Medical records. HIPAA for managing the privacy of medical records (9.9.3).

XBRL - extensible business reporting language [23]. This can provide consistency and greater transparency across companies.

7.4.4. Data Warehousing

Databases have been used in active business information system to long-term storage. The actual collection of records would probably be done with a database. Although many organizations do not employ full records management procedures, their databases are, nonetheless, a valuable asset of that organization. Indeed, complex organizations may employ many different databases. For long-term reference, these data need to be linked together and stored into "data warehouses", although, increasingly, businesses run integrated content management systems where they apply several of the techniques we have discussed such as organizational workflow (8.11.3). Enterprise Resource Planning system and a lot of data available. Big data (9.6.0).

Preservation of data becomes entangled with broader questions about the business rules reflected in those data and the metadata associated with the data. In short, the data needs to be treated as business records. Need by an organization for business records for auditing. The issue becomes collecting the right data across the enterprise. This often means reconciling the attribute definitions of the data, as the data models that contribute are not always consistent.

Data mining may be applied to a data warehouse. Increasingly, there are unified company-wide systems. However, one strategy involves organizing data in a cube — a so called "data cube" (e.g.,[88]) as a "star diagram" (Fig. 7.19). The Dimensional Data Model is less efficient than the Relational Data Model since it does not normalization of attributes but it allows greater portability. This is useful for obtaining a broad view of the interaction of components of the business (8.12.0). and the large amounts of data can be mined for potentially interesting patterns (9.6.5). Conceptual models for capturing organizational structure[79]. Data warehouses often use data mining. Increasingly, external information sources need to be incorporated into the data warehouse.

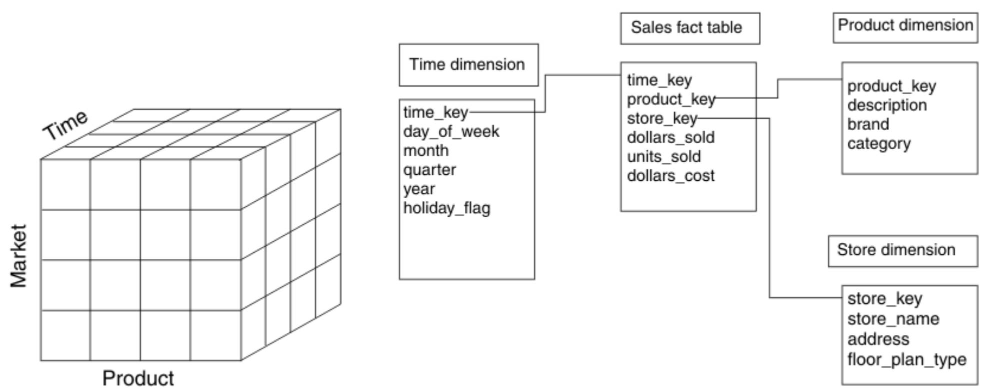

Figure 7.19: To consolidate and save the complexity of an organization's data files, an data cube (left) may be developed. This may implement dimensional data model (right), which is a simplified version of the relational model. (check permission) (redraw)

7.5. Archives

7.5.1. What is an Archive

People like to save things including information resources. Sometimes they save them as memories and sometime they save them as evidence. In some cases, the information resources may be saved systematically and specific procedures developed for them. Indeed, in some cases, that preservation may involve and entire institution.

While libraries typically hold published resources, archives hold documentary heritage. Cultural memory. Typically, this includes unpublished materials such as working notes, letters, and manuscripts. Compared to Records Management Systems, Archives hold information resources (even if they are digital) records systems focus on the information itself rather than the resource. Ultimately, we will need to go back to the fundamental concept of what makes a document (2.3.1) to understand the issues of how they should be saved. Indeed, the fact that archived documents are often carefully examined out of the original context of their creation means raises many deep questions about the nature of documents.

Archives covers a range of topics, from collecting and managing business records to preserving historical materials to preserving ancillary materials such as fan blogs.

There is a close relationship between Archives and Records. Indeed, some argue that they are identical. Although it is subtle and sometime contentious issue, we will associate the simple recording of transactions with electronic records and complex objects (whether physical or digital) with archives.

Whereas records often meet specific legal requirements, archives tend to emphasize cultural and societal materials. It may also include less that official organizational records such as email. Digital records do more than simply change the form of paper records. The cost on managing information is much most than the basic storage cost. Memory practices in science (9.2.0). The designated community is the community for which the archive is being developed.


Figure 7.20: Types of archives: A) U.S. National Archives, B) NASA data library. C) Film archive.

Transfer to an independent organization. Archival barrier.

Authenticity and integrity.

Archives keep materials for which there is just one copy. Uniqueness. Often, these were created without focusing on whether they would be saved. Providing authenticity and reliability. There are several types of archives, cultural, organizational, government, business, and personal. Each of these has distinct challenges but they also have issues in common. There are several viewpoints about the way for archives to record social and organizational activities. Here, we consider three of them:
1) Collection of documents from an organization.
2) Capturing organizational practice.
3) Documentation based on a specific initiative or activity.
Evidence and memory. Archives can provide valuable documentary evidence for history (5.13.0). More than organizational archives. Cultural archives.

Ideally, archives would dovetail with knowledge management. Similarly, the goals of archives are similar to electronic records. Together, archives and records provide evidence to provide accountability and information. Archives are views as providing support for journalists and historians while records are more often associated with business compliance. Archives and the content lifecycle (1.4.1). Archives and workflow. Archival information model.

Difficulties in the literal meaning for preservation since the meaning will be different for different viewers. This does not keep the context of the original records (Fig. 7.21).

Archives, like libraries and other information systems, need information management policies. In some cases, the goal is legal evidence (like electronic records) and the value of cultural memory. Game preservation (11.7.3). Archives of scientific data (9.6.0), technical, social, and medical data (9.9.3),

In many cases, archival material is effectively a record but we may make a distinction between the

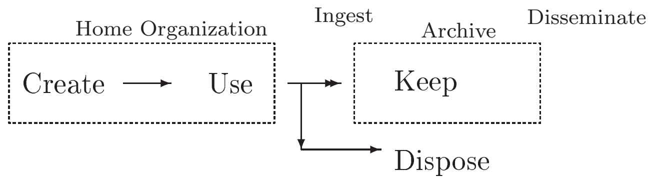

Figure 7.21: In the "Lifecycle model" of archives, material is transferred to an archive from its home organization. Among other issues, this creates a strong distinction between Electronic Records and Archives. A second model, the Continuum Model, emphasizes the continuity of records and the frameworks of record-keeping systems.

physical objects and the record it contains. Archives are repositories that store and protect the physical manifestations. Community memory can be a type of social capital.

Archives are often confused with libraries, but they serve a different function; archives emphasize preservation. While a library usually collects published works, most of the holdings in an archive are unpublished. Furthermore, the archive is typically focused on preservation while the library is focused on providing access to information. These materials are kept for historical analysis. However, maintaining archives for an entire government or organization is a difficult task. An archive should have standards for record selection, maintenance, preservation, and conservation. Preservation for long-term access.

Changing nature of organizations and continuum model (Fig. 7.21). Archives must be supported by an organizational commitment. Hand-off from the originating organization to the preserving organization.

Documenting versus traditional archives.

The United States National Archives in Washington, D.C. is the main repository for U.S. Federal Government records. The U.S. National Archives also houses important historical material, including the originals of the Declaration of Independence, the Constitution, and treaties such as the approval of the Louisiana Purchase. Treaties and agreements between the U.S. and other governments are also kept in the archives.

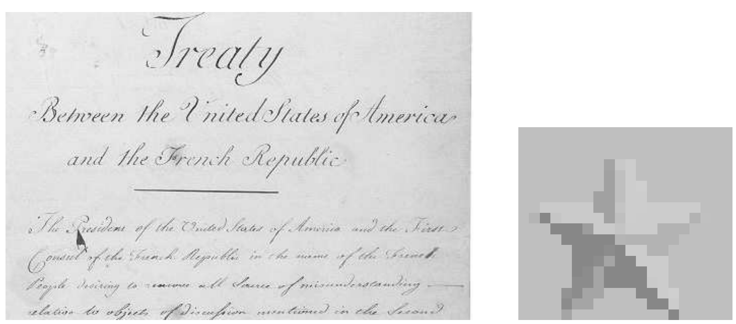

Figure 7.22: The original of the (left) Louisiana Purchase Treaty is kept in the U.S. National Archives. It helps to demonstrate that the Louisiana Territory belongs to the United States. Forgery (right) can be especially easy with digital documents.

Trust from several sources. From well managed repositories and from inspection.

Integrity and authenticity. Cannoicalization.

Documenting vs archival practice.

The Role of Archives

Film archives. Historical societies are also related to archives. Historically, one of the best ways to validate the authenticity of records was to verify the authenticity and age on the medium on which they are inscribed. Convergent documents provide evidence vs. Smoking gun. What to make on digital documents that have not be kept in a trusted environment.

Archives serve the need to keep information resources for the long-term. These are intended to be used as support for investigations by historians (5.13.0). This is traditionally focused on access. Archives are often associated with processing cultural and materials of historical significance while electronic records (7.4.1) are more often based on legal requirements.

Thus, the provide organizational or social memories." Institutional archives versus personal and informal archives. Many other types of preservation efforts: Meeting archive (5.6.4). Group and design archives. Design rationale (3.8.7). Data warehouses (7.4.4). The goals and ambitions of archival institutions are already changing. Popular memory archive. Archives versus memorials (4.6.2).

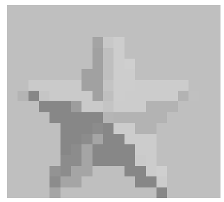

Figure 7.23: CEISMIC disaster archive.

7.5.2. Persistence, Preservation, and Conservation

Obviously, we want confidences that the material we save in an archive is not tampered with. An emphasis on persistence is the most distinctive aspect of archives. While we emphasize capturing practice, many archives emphasize capturing and management of documents. Some information resources may need to be kept for a very long time. Records and archives can give future historians the evidence on which to base their analyses. Because there is often social re-interpretation of history, documents can be valuable for "showing history as it really happened"[52] (5.13.0). Sustainability.

Government documents, such as treaties need to be protected (Fig7.22). Archives have more emphasis on persistence than libraries, and attempt to follow the S.O.A.P. doctrine: selection, organization, access, and persistence. While libraries periodically cull their collections of outdated and unused materials, the content of an archive is often kept on a specific schedule. Preservation plan.

These materials may need to be kept indefinitely because they are of historical interest, or because they have a very long life, such as treaties, contracts, or blueprints of buildings. There are certain terms associated with archiving that are important to understanding the field. Preservation only focuses on preserving the content of an information resource, and not the original resource medium. Preservation of the Declaration of Independence would be to copy the words. When a fire damages a valuable or famous painting, great effort is made to minimize the damage and restore it as nearly as possible to its original condition. In contrast to preservation, this process of protecting and restoring content in its original form is known as "conservation. Focusing on the carrier of the information or the content. In many cases, the medium and the content are not as inseparable as a painting on a canvas; in those cases we care more about the content than the medium. Keeping the original paper census records from 1920 (conservation) is not as important as keeping the data that was obtained in that census (preservation).

While procedures for archives are now well established for organizational records on paper, preservation of many other materials Preservation of multimedia or hypermedia content. Not only does each component need to be preserved, but the coordination among them also needs to be preserved. In the extreme, we may wonder how well would our culture be preserved if a new dark ages started in the next few years[11].

Just as human memory can be distorted by changes across time, cultural memories can also be distorted with time. Some of these issues can be reduced if we know that the material has not been directly corrupted and if some of the context (e.g., the workflow which generated the artifact) are maintained. Nonetheless, there remains a difficulty in interpreting archived materials.

Keeping track of the use (or re-use) of records and archival materials. Indeed, the reuse is often as informative and interesting as the original document.

Archives may also store documents from meetings, or scientific information such as astronomical and environmental datasets (9.6.3). These are typically collected via the agency for which the research was done. Personal papers (10.1.6) are also often archived, though on a scale of lesser magnitude. Archives need to provide at least limited access for use.

With archives, the emphasis is typically not on ease of access but on the integrity of the records. When records are stored, we need to know whether or not they have been tampered with. Trust in the authenticity of documents must be based on effective policies for the management of documents — without confidence that the archived materials are authentic, the entire reason for archiving them in the first place is lost. Increasingly, cultural records are stored in electronic form. Thus, the traditional methods of preserving them by saving papers and books will no longer work. Moreover, the Web is much more dynamic and contextual than printed archives. It may be necessary to save entire Web sites and not just Web pages, because a page should retain the context of its links. Culture (5.8.2). Lost books[58] and lost movies.

Policy-driven repositories. Capture early in the workflow process.

InterPARES rule.

Figure 7.24: Example of policies for developing a trusted repository.

7.5.3. Developing an Archive: Appraisal, Selection, Ingest, and Retention

Imagine that you have been asked to archive somebody's computer files. One of the greatest puzzles is what to save. Would you save everything, or just "important" documents? Only material that was selected to be saved is available later. How would you know what was important? Although it is tempting to try to save everything, it simply isn't possible. Is there a bias toward collecting only certain types of material – for instance, material representing only a given sector of society?

Ideally, archives select systematic and value-neutral selection of material. It is difficult to decide what cultural objects should be saved[81]. Often, those with power and money are preserved. Underserved communities are often not well represented in archives.

Appraisal determines the archival value of documents. Collecting the evidence in the context of the workflow. No interpretation of the documents.' Automated preservation assessment. Risk analysis. Earlier we considered the document lifecycle (1.4.1). Where in document lifecycle is the document brought into the archive? Continuum model[82] and duality of viewpoints about the archive. Documentation strategy. Documentation could also apply to scientific data sets. Attempting to determine the temporal dynamics of the value of data is difficult. Macro-appraisal emphasizes determining the context. This is often associated with finding organizational function of documents.

Protecting personal and private information.

Archives might be based in the organization where the documents originated. But it is often better to turn over control to an independent organization. Moreover, archiving is a specialized skill and organizations are not well suited to judge what should be archived due to changing policies and goals. Indeed, they may be tempted to selectively pick records which cast them in a favorable light and to destroy any unfavorable records. In other cases, the selection may be integrated into ongoing workflow. Functional appraisal.

Retention is built into organizational policies These policies may include what description of content to keep and for how long it wants to keep it. These are known as permanence ratings (7.5.4). Dynamic Web (10.10.1). In other cases, records need to be examined by an archivist as they are transferred into the archive. An archivist will choose documents for archiving based on a set of retention guidelines (Fig. 7.25).

> The Preservation Department should seek to raise awareness of the importance of library preservation, and of the need to identify priorities for preservation in [Glasgow University Library] GUL as a whole. This will involve discussions with departments and subject specialists within GUL, so that specific recommendations can be put forward regarding the selection of materials of significant bibliographical, historical or aesthetic value which should be preserved for the benefit of research and scholarship in the longer term. Consideration should be given to the rarity or uniqueness of the materials in our collections, their national or international significance, and to our own local needs and special interests, bearing in mind the possibility of placing such materials under restricted conditions of use, and of storing them in a specially controlled environment. Where appropriate, material will be selected for re-formatting as microfilm and/or digitally scanned files to limit unnecessary handling of originals and to facilitate access.

Figure 7.25: Preservation policy for the Glasgow University Library[51]. (check permission)

Broadening the base of archival practice through social media. Moreover, the availability and relative low cost for digital materials make preservation and linking across a wide variety of materials more feasible.

The development of retention policies. should help determine the materials to save, as it is difficult to know what might be important in the future. Indeed, because documents help to define history, the selection of which documents to save can, in a sense, affect history itself. There are different strategies for archiving. Some say that documents that are used regularly should be saved. Others argue that random sampling is the strategy, because it is difficult to predict what may be useful in the future. Once the material is selected, it must be "accessioned" or "ingested". This means that ownership of the material is typically transferred to the archival organization and it is incorporated into the archive. Workflow processes for ingest. Less process more product. Supporting access to materials. Risk analysis in copyright.

7.5.4. Description and Access for Archival Items

Earlier, we considered bibliographic metadata (2.4.0); here, we consider archival metadata. There are several particular challenges for archival metadata. It should be able to be understood by people from a different cultural background. In a different sense, the context of one item in a collection of items is essential. There is a particularly close connection archival description and electronic records (7.4.1). Traditionally, individual items are not described but sets of materials with similar provenance are described. The ordering often provides valuable context. Increasingly, we see multiple modes of description. Quality in the collection and management of data. Rights and responsibilities of data.

Organization related to the preservation model of the archive. Fonds, Series vs. Function.

Description to Support Users of Archives

What sort of description do historians need?

Provenance and Provenance Metadata

Provenance is the history of an object's ownership or, at least, control of it. Thus, provenance can help to guarantee its authenticity. Across a sets of records, provenance may be seen as a process for ensuring authenticity (7.5.2). One of the strongest ways to establish provenance is tracing a chain of custody. Provenance has applications in the history and origin of information resources: what was created, where, when, and for what purpose. This helps to ensure that information is not tampered with after it has been deposited.

Provenance tells us about the origin of information and it provides an indication that the information has not been changed. For instance, attribute of records is their organizational history. It is helpful to maintain an account of where and when each piece of information originates; this is known as its "provenance". Managing the provenance for information resources is analogous to tracking the provenance of physical objects (7.6.1). Provenance through chain of custody in trusted repositories. Tracking provenance helps to demonstrate the authenticity of the materials because the origin of information can be authenticated either by eyewitness testimony or by its logical consistency when compared to other information from that time and place. Digital provenance for conversions between formats. Provenance data models and transaction logs. Provenance metadata.

The full import of some material is not evident from the material itself. It is only when it is seen in its proper context that its meaning becomes clear. For some archives, it can be important to maintain the context of the source material. This can be troublesome; an archive usually holds only a fragmentary record of material that is old and was often not created to be widely distributed, and recreating its context might not be possible. Sets of documents may belong together because of the activities which generated them [77]. Difficulty of capturing tacit knowledge. Context tells us a great deal about information resources. Keeping a rich context.

Formal models for provenance.

Preservation Metadata

We have already noted that preservation is a integral part of the information lifecycle. And, of course, preservation is the most distinctive function of archives. As with other types of repositories, metadata are essential for supporting access to archival records (2.4.0). PREMIS vs Functional Requirements for Records. PREMIS is especially useful for digital objects. For instance, it describes Events with change the object such as transformations. Close to some standard which are familiar to the records management community. Technical environment metadata.

Metadata for dynamic documents (10.10.1). However, there are several distinctive aspects of archival material that needs to be considered. Preservation and context. Thus, "preservation metadata" might include a description of the reason the material was selected for preservation (e.g., [14]). Events associated with maintenance. Some archival content is intended to be kept forever. Other content has a finite lifetime, which may be indicated by "permanence ratings". These describe the permanence of the informational content itself, and not its desired archival duration [16] (Fig. 7.26).

Label	**Description/Example**
Permanent: Unchanging Content	The original copy of a treaty.
Permanent: Stable Content	An edition of a book.
Permanent: Dynamic Content	Today's newspaper.
Permanence Not Guaranteed	An ordinary Web page.

Figure 7.26: Permanence levels as descriptions for records[16].

Another problem is guaranteeing that a digital document is the original and has not been tampered with or altered. This can be done with hashing and checksums (-A.20.2). Another possible solution is to compare the file with other archived copies of the document to see if they match [72].

Formats and Representation Information

Representation Information Media formats versus file formats. How is a file format structured? Format repositories. Global format registry and validation. Format validation [54]. Some formats are so complex, it is difficult to understand them out of context. Preservation for especially complex design tools such as CAD systems here are particular challenges in the preservation of complex material such as architectural design, descriptions of cultural heritage sites, games, and new media art. Desiccated formats reduce the complexity of data set to just a few salient features. Self-describing content. Metadata can be wrapped with the object. Self-archiving content.

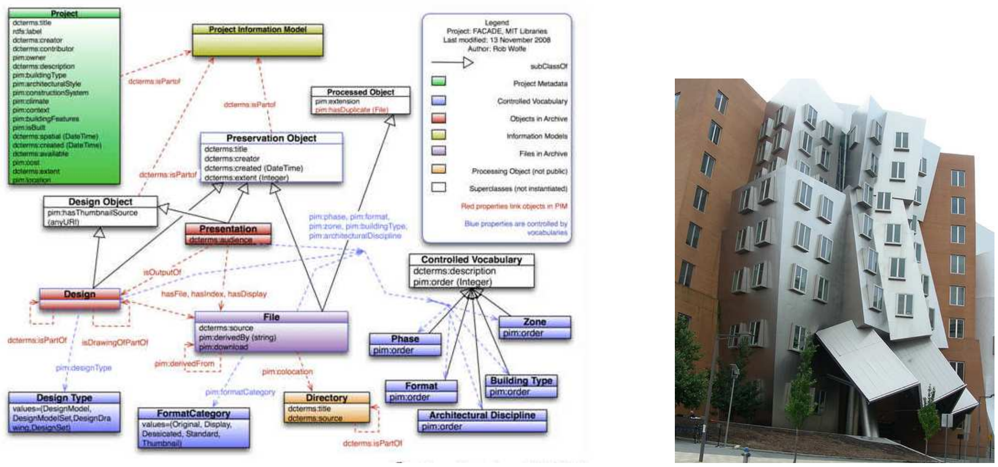

Figure 7.27: Describing the properties and relationships of a complex set of files related to a archivtecture. FACADE project ontology[64]. (check permission)

Archival Context

Describing objects in context is helpful because individual item will not tell the whole story. One organizing principle, the materials an attempt is often made to keep the context in which material was developed and used. This often mans keeping the materials in their original order; that is in the order in which they produced and used rather than the order in which they were stored or processed by the archive. The original order can be useful in providing the context in which they were crated.

Indeed, the entire collection of items may be considered as basic unit. This can be complex because because the actual collection may not be complete or it may be biased because it reflects selection decisions.

This is often different from the description of separate objects. That is, they are organized according to the way they were originally developed and used. Email messages might be kept in the order they were received, grouped by year, rather than by topic. For material stored in this way, one would access a record group, and then search through that group to find a specific record.

Figure 7.28: Archives typically try to capture the context of objects. Indeed, they consider several senses of context – how the object relates to other objects in the way it was used. (check permission)

Organization or Community Role Describing a complex combination of organizational functions. Archival series. Activity theory for archives.

Use after Ingest Interpretation of the artifact based on how the artifact in used and interpreted in that re-use. Continuum. Commons.

Finding Aids

Some archives collect assorted materials such as the files from a person's desk when they retire. In those cases, a "finding aid" could help a searcher to find out what sort of materials are included. The

Encoded Archival Description (EAD) is used to describe records in traditional archives. This finding aid is based on XML and it defines parts of an archived document. Increasingly, finding aids are being adapted to be sort of under interface to an archival collection.

Next generation finding aids and Web 2.0[91].

Overview of the Collection
Biographical History
Scope and Contents of the Collection
Arrangement of the Collection
Restrictions
Subject and Genre Headings
Related Material
Administrative Information
Inventory

Figure 7.29: Because most items in a traditional archival collection are not cataloged directly, sets of items are described in Finding Aids. Typical top-level sections of a finding aid. (check permission)

7.5.5. Preserving Complex Digital Materials

While conservation efforts generally focus on traditional materials (7.5.2), they are also necessary for digital media. Although the bits which comprise digital objects are ephemeral, they must be stored in some physical medium, which may be fragile. The iron oxide holding the bits may fall off a magnetic tape, making it unreadable. Archiving techniques must address the formats to be used, and devise methods for ensuring their future viability. Data curation (9.6.3).

Modern information systems have created new demands for archiving practices. There is an enormous volume of email, word processing files, and Web pages already extant, with more being produced every day. Should all of these documents be archived? Increasingly, there is a convergence between digital libraries and digital archives because digital copies are managed rather than physical copies.

Email preservation. Beyond preservation of traditional documents. However, preservation of letters and manuscripts has been a traditional role for some institutions.

Game preservation (11.7.3). Preservation of multimedia depends in part on the resolution of the original. New media preservation. Describing the experience. Fan blog. Because of the rapid increase in the availability of disk storage, the criteria for selecting material can be loosened. Indeed, it been proposed that we can save everything[61] though that is often broad and we want to save a lot but not literally everything. We cannot literally save everything but we can save a great deal more than we used to.

Born digital materials often have links to other resources and those links are fragile. Evolutionary systems — those whose contents are continuously changing such as databases — provide a particular challenge to a database may record many of the transactions of a business, but it is subject to frequent changes. Is it necessary to keep all transactions and versions of that database (9.6.3).

One factor to consider is whether it is necessary to simply display the archived materials (maintain what the documents say), or whether their behavioral capabilities also need to be maintained such as the buttons on Web pages. Because of word processors and Web publication, revisions are easy to make. This has resulted in a proliferation of versions.

Preservation of complex objects such as games (11.7.3). Preservation of new media. Data curation (9.6.3).

Figure 7.30: Frame from a reconstructed NASA video of the first step on the moon. The original was lost.

There are a great many attributes of digital objects which could be saved. Among these are attributes of the computing environment which generated and accessed the digital object. This might include the version of the operating system and file creations software. Current operating systems makes this difficult. Indeed, there are so many possible attributes to save that some selection policies are needed to prioritize them.

Perhaps surprisingly, digital media are not necessarily better for preservation than traditional media. Indeed, modern information systems are very complex, and materials created by them often depend on the system that created it.

The recreation of every digital object requires a program. The management of relative standard programs such as word processors is challenging but the management of ad hoc programs can be even more challenging.

As with traditional archives, there is a range of content types for digital materials with different expectations about how they should be handled. One policy is that there should be multiple copies of every digital object so that if one is lost, the others will survive. Replication of the entire information resource can also be useful in case the originals of the resources are corrupted or lost. If this is managed automatically, these are "self-healing archives". LOCKSS (-A.20.2).

Preservation of functionality.

Both formats and base technologies change and several strategies have been proposed for keeping media current: migration, emulation, encapsulation. Other strategies include desiccated formats and re-enactment. Migration from one compression format to another is transcoding (-A.2.1). There are limits to this approach, however. For one, the amount of data to be transferred would increase every time. For another, it is rare that we have "lossless" format conversion, which allows the original to be exactly copied. This would mean that some quality is lost every time a transfer is made. Backward compatibility of standards and programs.

Attacks on cultural materials. Suppression of certain viewpoints.

Another possibility is to keep at least one example of every old computer, its environment, and the programs that created the records running, but this does not seem practical. The underlying operating system and hardware would have to be emulated and the emulations would have to be cumulative such that newer emulations would run on previous emulations. In the end, it is not possible to keep all versions of all base technologies running. Evolution by virtual machines (7.7.6). Fig **??** rendering (7.5.5).

Probably, the most practical current strategy for complex digital preservation is encapsulation, for which metadata is essential. Encapsulation associates metadata with objects, and organizes the objects

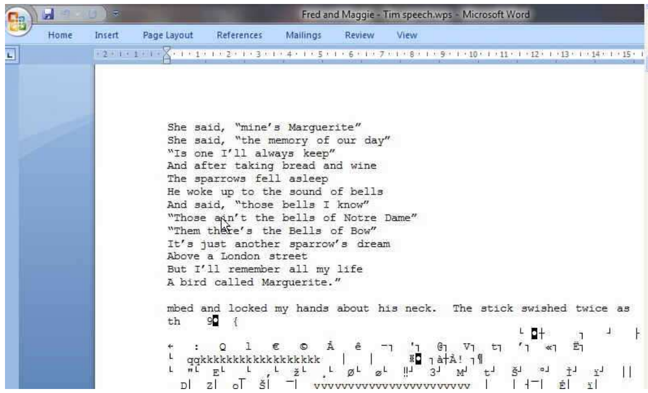

Figure 7.31: In many cases, an exact match for the software environment on which a program originally ran may not be available. In those cases, partial information may be retrieved. Here MS Word accurately recovered some information from a file but it also inserted some noise (from [?]). (check permission)

according to metadata similarities or logical connections. Encapsulation (7.5.5) is often applied when packaging digital objects for including in the file to facilitate later searching. Emulation appears increasingly to be possible. The application of emulation requires detailed knowledge of the software versions. Indeed, both migration and emulation have strengths and we may use both of the them.

Web Archives, Preserving the Cultural Record, and Mass Preservation

Many large digital collections are now being developed. Complete copy of the Web. Big-data (9.6.0). Mass digitization (10.1.6). Digital surrogates. How often should the web pages be sampled. Web curation tool for capturing and storing web content. Records need to be made for every page. WARC. Internet Archive and Wayback Machine. Managing copies of Web pages as legal evidence. Part of regulatory compliance for electronic records.

Figure 7.32: Democratic (left) and Republican (right) political party Web pages (cropped) from the 1996 presidential election. These are among the earliest sites preserved. This is an archive of political web pages[55]. Wayback museum.

Evidence and Forensics

Determining evidentiary value of an information resource. One set of issues for the evidence about physical media. Additional considerations for digital media. Several senses of forensics. In this case, we're considering whether a photograph has been tampered with Fig. 7.33. Trusted digital objects (7.5.6). Both legal implications and implications for scholarly accuracy. Determining legal evidence from examination of digital objects.

Figure 7.33: Digital Forensics: The ducks have been added to this photo. You can detect that by noting the shadows on the ducks but note that there's no shadow on the people. [?]. (check permission)

7.5.6. Archival Repositories

Trusted Archival Repositories and Trustworthy Digital Objects. Documents from archives come with a special expectation of authenticity. Trust from policies and procedures along with physical controls. The storage environment helps ensure trustworthy documents. For archives there is a special emphasis on ensuring the persistence and provenance of the stored materials. Thus, there is an emphasis on ensuring that repositories are trusted. This is similar to the issues of information security and information assurance and information system security (7.10.3). This can be accomplished with reliable storage media, stable organizational support, carefully considered formats.

Cloud repositories. Description of basic services. Open Archival Information Systems (OAIS) [40]. Reference model. Standard operating procedure. Policy-based service definitions, especially those that lead to a trusted repositories. Audits of the system (7.10.0). Certification of repositories as being trustworthy [53].

7.5.7. Archival Policies and Institutions

Information culture (5.9.5).

Access Policies for Archival Materials

Part of the the braoder question of policies (7.5.3). Ease of access has not been the primary objective for paper-oriented archives because of the difficulty involved in handling and indexing each page. Documents are often not indexed simply because of the sheer volume of material received; there are so many records that they cannot all be cataloged. Archives have not, typically, supported end-user access but that is rapidly changing with digital content. Because digital material is so easily copied it dramatically changed the possibility of access to archival materials. Access restrictions. Archival escrow.

Providing access can add a great deal to the cost of preservation. Dark archives emphasize storage of material but have a minimal concern with providing access [38]. This may be done to reduce costs or simply because of access restrictions.

Risk Assessment and Audit

Many data servers were lost in the 2011 Japanese earthquake. Potentially, many family records would have been lost but the records had be backed up at the national level.

Management of the preservation and migration of information resources. Project management taking care of these processes efficiently. Risk and threat analysis (7.10.3) for what could go wrong.

Preservation Planning Policies

Preservation planning is an integral part of a trusted repository. As with other information systems, there are many social dimensions to consider. Organizational aspects of information systems. Requirements for the archive. Risks for the integrity of the collection (7.10.3). Audits based on the OAIS model: Ingest, Storage, Dissemination. Organization structures for supporting the development and long-term support for preservation system. Cost-benefit analysis.

Sustainability of archives and organizational and institutional commitment.

Preservation of files versus preservation of data records such as in a data-warehouse (7.4.4).

7.6. Collections of Physical Objects

Physical objects such as artifacts and specimens can also be collected and those collections need collection management. Data schema (2.4.1) but also information about the relationship of the object to the institution. Metadata-like descriptions can be applied to physical objects. Emphasis on item-level description.

There are a great many types of physical objects and there are many dimensions on which they can be described. In other places we have discussed description of designs (3.8.4).

Here, we desribe the description of cultural objects. Primarily we consider museums which are cultural institutions but there are other descriptive approaches in fields such as archaeology.

7.6.1. Museums

Museums hold collections of physical objects. A particular sense of empathy from viewing objects. The primary role of museums is the presentation and interpretation of those physical objects for the public. While the main goal is education, preservation of rare or unique objects may also be important.

Types of museums shown in Fig. 7.34: Art, history, natural history, science museums. There are also many museums dedicated to unusual or distinctive materials. Research museums keep reference collections for use by researchers.

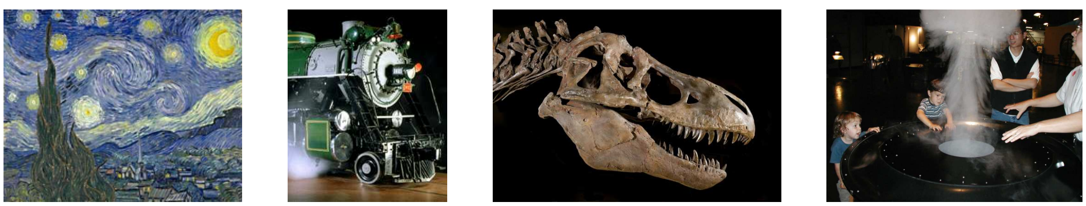

Figure 7.34: Types of museums: A) Art museum, B) History museum, C) Natural history museum, D) Science museum. (check permissions)

Interpretation and Recreating Experiences. Museums as keepers of cultural records serve a sort of storytelling function. Virtual and augmented reality reality for museums. Museums may try to provide an experience, related to the way artifact was originally used. Or, at least, they may try to create an understanding of the experience. What do the objects in a historical museum actually tell us?

Interactive, hands-on exhibits. Art museums may collection digital art works. Material culture.

Exhibits. Period room.

Museum Informatics

Museums are primarily based on the preservation of physical artifacts. They serve both research and education missions. However, this is tricky because memory can also be distorted. Indeed, science museums generally focus on demonstrations. Some of these descriptive systems are similar to those for data about entities in the world (9.6.0). Because there is an emphasis on public education in some museums, there may be an emphasis on interpretation of the collection. Reference collections. Science museums are equal parts educational institutions. There are controversies around the ownership of some cultural objects. Provenance of physical objects. Educational goals. However, they also attempt to make educational exhibits. Objects as evidence. Primacy of the physical object. This can be distinct from archives which preserve information resources. Thingness.

CIDOC-CRM for physical objects. Coordinating CIDOC with intellectual objects with Dublin Core and FRBR. This is also related to Schema.org (2.4.1).

Describing Cultural Objects

We also need to consider collections of physical objects. Some types of museums hold cultural objects. One difficulty is that physical objects are generally most meaningful in their original context [7]. Bibliographic description (2.4.0). Photos of physical objects as works. Compare physical objects to reproductions of those objects. Metadata extensions include Cataloging Cultural Objects (CCO) [7]. CIDOC-CRM and events.

Figure 7.35: Small statue of a Japanese god (add metadata, from the Spurlock Museum.)

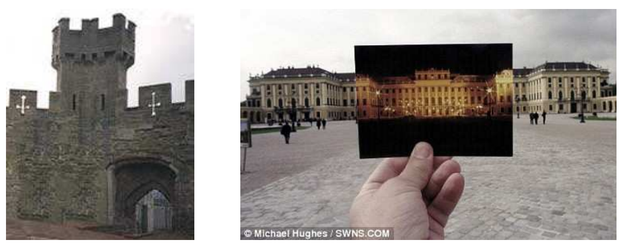

Figure 7.36: Describing parts of a building. (check permission)

7.6.2. Other Collections of Physical Objects

Objects in the field such as buildings. Some types of physical objects need to be indexed. Linking to explicit taxonomies and databases such as zoological and botanical records (9.8.1). These include zoos and botanical gardens. Data schema (2.4.1).

Interactive Museum Displays

Dioramas, exemplary specimen, comprehensive collection, interactive example.

Exploratorium.

7.7. Information System Architectures and Web Engineering

For the remainder of this chapter, we turn to considering technical issues which are associated with collections of information. Because information systems are complex, it is helpful to consider some principles. Through the evolution of information systems, there has been a decrease in hardware dependencies and increased alignment with services. Effectively using resources. Design (3.8.0). Reliability. Security. Scalability. Interoperability. Using resources effectively.

7.7.1. Networking and Distributed Systems

Networking (-A.16.0). OSI layers for managing complexity in a network (Fig. 7.37).

Network forensics.

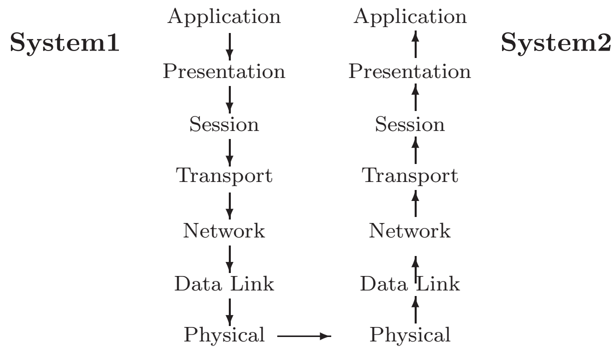

Figure 7.37: The combination of multiple network protocols makes for a very complex system. The OSI stack for networking manages this complexity with layering. In most cases, network protocols operate at only one of these layers.

Networking security (7.10.3). Playback attack. Some simple techniques are useful such as nonces and sequence numbers. Intrusion detection system.

Peer-to-Peer Systems

Peer-to-peer (P2P) systems are interacting sets of peer computers. Each of these are "peers" in that they are all equal within the network and there is no specialized servers or central coordination among the components. Thus, these are different from client-server architectures. Fig. 7.38 shows a peer-to-peer model in which each node has its own index and resources. Messages are passed from the originator to neighboring machines, with requests spreading outward from the center.

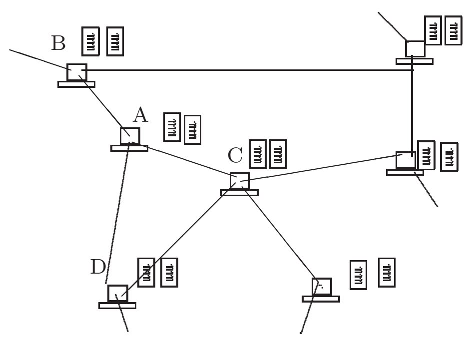

Figure 7.38: Peer-to-peer systems allow distributed searches without a centralized index. Rather, each peer has its own search index. A, B, C.

Searching in a P2P system spreads across the set of servers. With peer-to-peer systems, searches and document retrieval can be largely anonymous; the searcher can hide their identity. This anonymity is accomplished by stripping off identifying information as the requests pass each machine. Thus, peer-to-peer systems allow anonymous file sharing, but sometimes this may run afoul of intellectual property laws (8.2.2). Without a central index, the chance of finding a given item decreases as the number of items and nodes increases. In other versions of peer-to-peer systems, indexes can be progressively developed on each system. This is the basis of services such as Skype and Bit-torrent. Onion routing.

7.7.2. Global-Scale Information Service Platforms

Database foundations (-A.14.1). Networking (-A.15.0).

Grid and Cloud Computing

Support collaboration by distributed organizations. Overlay networks. Contracts for data management services. Service-level agreements. Many users are interested in protecting the value of the data

The Grid is designed to provide services on a very large scale. Later, we will also consider the "semantic grid" or data grid (9.6.0) but here we are focused on the network infrastructure. Grid coordinated database updates. To support the coordination of components there need to be standards such as consistent name spaces, metadata, and service descriptions. Cloud computing. Allows efficient use of computational resources. Security of cloud computing can be an issues because the data are not directly under the control of the owner. Federated searching has distinct types of searches in different clusters.

Europeana project has a federation of cultural objects. NIF has a federation of datasets.

Security issues for cloud computing.

An example of grid computing is the Search for Extra Terrestrial Intelligence (SETI; Fig. 7.39). SETI utilizes the spare computing power of thousands of ordinary home PCs to analyze radio signals coming toward earth from distant stars for any sign of extraterrestrial attempts at communication. This is a loosely connected type of Grid. Grid computing offers extraordinary computing power, but its development is still in its infancy. It will be necessary to develop user standards as well as security measures for grid computing. Other similar systems include. There are a lot of computationally intensive tasks which can be processed with this approach such as searching for the cure malaria.

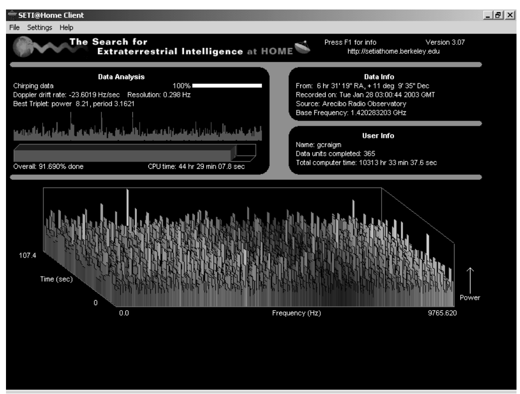

Figure 7.39: SETI@home distributed processing allows a networked computer to contribute un-used cycles to searching for extra-terrestrial intelligent life. This screen shows the results of an analysis of data from a radio telescope. However, no clear signs of extra-terrestrial life have yet been found! (check permission)

Tasks can be spread across computers on a Grid, workflow can be specified for completing tasks on the various computers. However, the management of data management on the multiple computers is particularly difficult because of synchronization.

Reliability of distributed network services. Stress test. Chaos Monkey.

File operations allowed. Public versus private clouds. Managing security for cloud computing.

7.7.3. Managing Networked Data

Different requirements for different services which provide the data.

Search engines have primarily push data. But there are back end challenges for web crawling. There are also challenges for indexing and real-time news feeds.

Data centers to support social media typically have a whole different set of challenges. Symmetry of friend relationships need to be updated. In addition, these have particular challenges for security.

Attributes of effective data specification (-A.14.1).

Cloud Computing

Virtual machines.

Data storage on the Grid is often provided by data centers (Fig. 7.40). Exactly what is being offered by the cloud. Service-level agreement. Quality of service (QoS) guarantees. Security issues and caution about placing sensitive records in the cloud. Difficulties of data centers for social network sites. Perhaps the greatest concern is whether the company or the country running the service is itself trustworthy.

Risk assessment for loss of a data center (7.5.7). Issues with legal access. For instance, what happens to data if the company which run the cloud goes bankrupt.

Figure 7.40: The Google data center in The Dalles, Oregon. Dams on the Columbia River (background) provide hydro power. (check permission)

Figure 7.41: This shipping container holds the entire Internet Archive. Potentially, portable data centers could be moved to places which had unexpected need for intensive data storage and processing. (check permission)

How much data storage is there? This is quite different from the question of how much information is stored but it is much easier to calculate.

Data Storage

Need increasing amounts for storage. Distributed file systems.

Challenges of cloud storage. While cloud storage is possible because communication costs are relatively small, they are not negligable. Is the data is replicated, then that replication must be synchronized.

Long term-costs of storage.

MapReduce. Google File System. Big Table[6]

7.7.4. Resiliance and Reliability of Information Infrastructure

As an engineering and policy question.

7.7.5. Information Infrastructure Governance

Information policies.

Coordinate with social policy. ICANN, IETF, Local laws versus cross-border law. Internet gambling.

7.7.6. From CPUs to Software

Virtual Machines

As an example of handling complex, we consider virtual machines (Fig. 7.42).

Figure 7.42: Virtual machines help to separate software from hardware dependencies. (hand drawing)

7.7.7. Software Architectures

Organizations do computing and we considered good organizational structure to make that efficient. As described above, developing modularity is helpful. What architecture do we need for general purpose computing. A simple functional breakdown of program components.

Networking allows independent computers to communicate. This can add complexity since data and computing resources may be in different locations. At the least, this can add time to computations because communication between computing resources will be slower than time for centralized computing resources. Distributed systems include redundant components (e.g., computers) which are located in different places. This can provide for a distributed system's reliability. Distributed systems are robust and not prone to failure because not only are the processing capabilities of the system dispersed. both of which prevent the entire system from collapsing even if one (or several) parts of it do — the remaining parts will compensate for the damaged ones. A large, well-known search engine company finds it cost effective to utilize thousands of low-cost PCs rather than a few large servers to run its system, even though individually the PCs have a high failure rate, because the engine is almost immune to system-wide crashes due to the built-in redundancy across the multiple processors.

There are advantages and disadvantages to centralized and distributed computer systems. As we have already noted, the primary advantages of a distributed system are its reliability, unlimited scalability and computing power, and relatively low cost. The use of distributed systems also makes sense when data repositories must remain separated. Tasks in a distributed system can sometimes pile up on one processor leaving the other processors lightly used. The processing can be spread across several machines or "load balance" so that no one processor is overly busy. There may be legal restrictions that require information to reside on different networks. The physical separation may also be inherent to the task, as with construction robots that work in different places.

7.7.8. Autonomous Agents

Autonomous agents are programs which interact with each other but do not have central control. They may have goals and an approach to meeting those goals as conditions change. Furthermore, they may interact with other agents. This is the logical extreme of distributed programs. If the framework is well constructed, such programs can adapt social and economic models similar to those which provide structure in human society. More agents and artificial life (-A.10.4).

Multi-agent systems. Multi-agent semantic web. Coordination (3.5.3). Because these agents are autonomous, coordination among them can be like coordination in economic systems (8.7.0). Useful in simulation (9.5.0). Learning SOAR (-A.7.3). Reactive agents[15]. Agents protocols (7.7.8). Alife (-A.10.4).

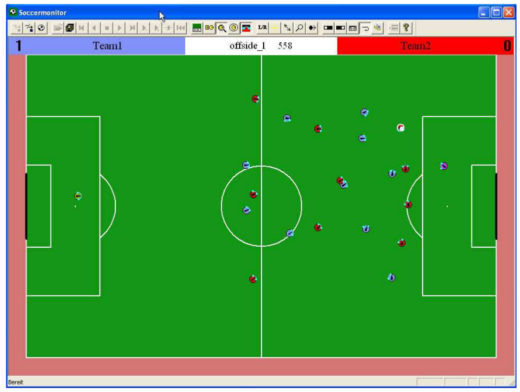

Figure 7.43: In RoboCup, teams of agents compete in a simulated soccer game.

7.8. Platforms for Services

We have already seen several examples of digital repositories. Web servers make documents available but they are not service oriented. A content-management system holds digital objects. The systems may also employ version control. A repository stores content, but what makes it unique is the services it provides. Incorporate annotations. Applications of repositories include digital asset management. Types or genres of information systems. Here, we focus on libraries, archives, and enterprise content management. Specification of services such as with UML (3.10.2).

Repository servers typically support collections. A collection of digital objects does not have to be on one server. It may be a virtual collection rather than a physical collection. As with all distributed systems, organization is particularly important (7.7.1). The Fedora digital library implements the following: Ingest, Dissemination, OAI provider, Search, Preservation monitoring,

7.8.1. Service Model and Service-Oriented Architectures (SOA)

We have emphasized services so it makes sense to develop computing platforms to support services. A first step is to be able to run processes on remote machines. the original Web was designed simply to deliver documents, but it can also provide a foundation for many other services. The key is to be able to run programs on remote systems that is to execute a “remote procedure call” (RPC) Specific service platforms include: Web Services Definition Language (WSDL) Service Discovery. UDDI.

7.8.2. Interlocking Services for Platforms

Search, ecommerce support, social networks. We have emphasized services. Repositories hold collections of information resources. Computing utility which saves local computing and storage resources. Micro-services. Service science. Ultimately, Internet features need to be based on the user needs and Usage patterns. In other words, the Internet itself is a socio-technical system. Composition of services. Social, mobile, cloud. This combination of services is termed a platform. Platforms include semantics such as from the media industry. Interlocking social interaction, content access and management, and ecommerce.

7.8.3. Distributed Repositories with Centralized Metadata

A good collection allows for many perspectives on a single topic Separate collections, one on Polish patriots and the other on piano players would both include material on Jan Paderewski. While a traditional library needs to manage its collection all in one place, a digital collection management node can provide services and manage a collection that is spread around the world. Fig. 7.44 shows connections to independent indexes and repositories from a meta-server. Relevant metadata from any digital document can be harvested from its source and built into a virtual collection. One example of collection indexing nodes is provided by the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH, Fig. 7.45). Access to selected documents can be controlled with link-resolution protocols.

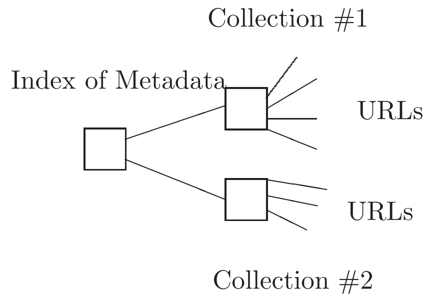

Figure 7.44: A virtual collection can be created by collecting and indexing metadata from remote repositories.

This leads some to an emphasis on information objects which are minimally structured. Or, at least, structure is not important. This view is quite different from the view of documents as works. As information objects, they can be shuffled, re-organized, and aggregated into new resources. This is view implemented by OAI-ORE (Object Reuse and Exchange) and in this approach, information resources are defined by resource maps. Learning objects and object reuse. Supporting other types of service such as discovery with distributed repositories.

```
<header>
  <identifier>oal::arXiv:9901001</identifier>
  <datestamp>1999-01-01</datestamp>
</header>
<metadata>
  <dc xmins“http://www.openarchives.org/OAI/dc.xs”>
    <title>Quantum slow motion</title>
    <creator>Hug, M.</creator>
    <creator>Milburn, G.J.</creator>
    <date>1999-01-01</date>
    <type>e-print</type>
    <identifier>“http://arXiv.org/abs/9901001”</identifier>
  </dc>
</metadata>
<about>
  <dc xmins“httpd://www.openarchives.org/OAI/dc.xsd”>
    <rights>Metadata may be used without restrictions</rights>
  </dc>
</about>
```

Figure 7.45: Example of Open Archive specification[60]. for a document in the “arXiv” repository (www.arXiv.org). (check permission)

Aggregation of Collections OAI.

URI.

Distributed Search Indexes

There are drawbacks to distributed collections. One of them is the difficulty of distributed searching. When the content of a collection is spread across many servers, a full-text search of the collection may be accomplished by contacting each of the servers (Fig. 7.46). Each remote server conducts the search and each has to respond. The problem with this arrangement is that the response time for the entire system is as slow as the response time of the slowest server. Indeed, if one of the remote servers is unavailable, the search may not be completed. Search engines (10.7.4).

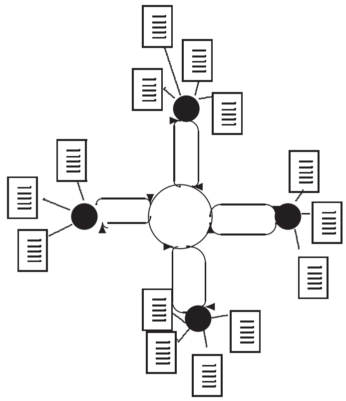

Figure 7.46: In distributed searching, the user contacts the central server and that request is in turn sent to remote servers (solid circles). Metadata for matching documents are then sent back to the central server. This approach can be slow to answer the query if any of the remote servers is slow to respond.

7.8.4. Metadata for Complex Information Resources

Earlier, we considered metadata frameworks (2.4.3). Metadata on the network.

Preservation of complex objects. Games (11.7.3). New media (7.5.5).

Metadata for Complex Digital Objects

Digital objects. Information Packages. Actual information resources are complex. They objects involve multiple metadata standards. For instance, a multimedia object. Packaging Services and Digital Objects in a Repository Micro-services. Use-cases. Information package includes the digital resource and associated metadata. Even more complex objects may include interactivity.

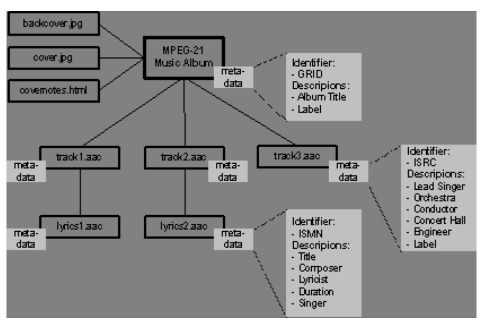

Figure 7.47: Describing a CD album with MPEG-21. MPEG-21 is a framework for metadata for each of the components. (check permission)

Front cover art

Back cover art

Figure 7.48: Describing a CD album with MPEG-21. MPEG-21 is a framework for metadata for each of the components. (check permission)

PREMIS is an XML standard is used with METS. METS from complex objects to composite packages.

Semantic annotation.

Content wrappers. Complete understanding of information resources requires many levels of description. A comprehensive encapsulation of content is a packaging standard. There are several of these such as XFDU, MPEG21. For instance, an album of music may be composed of several songs.

Here, we focus on the Multimedia Encoding and Transmission Standard (METS) (Fig. **??**) This is particularly important for archival description. METS example document. Resource maps. StructMaps are hierarchical. Limitations of hierarchies for document descriptions. (2.3.3). Beyond that, there are other standards such as MPEG-A.

This is particularly applicable to complex objects such as newspapers, movies.

From METS to genres.

MPEG-21 and MEPG-A. MAF and coordination of standards in application devices such as audio players.

Metadata Registries

They need to be indexed in order to be found. Information seeking may start with determining what strategy to employ. If a collection is adopted, the user must start with source selection. Collections in a networked environment. Repository severs and the services they provide are decentralized and it may be difficult to know what resources or services are available. There are several models of advertising services such as UDDI. is a generalized Web service registry. Moreover, because standard descriptions are so important for coordinating distributed services, it is helpful to have registries of metadata standards. Such as the Open Metadata Registry. Difficulty of interoperability. Cross-repository services. Storage resource broker. ISO 11179 is a standard which facilitates the reuse of metadata; it defines metadata on registries.

7.9. Software Engineering

Information systems can be complex and developing them a challenge requiring engineering. This process is itself highly information intensive. Software engineering attempts to develop quality software in an efficient manner.

Factors to consider are cost, complexity, disruption. These activities need to be coordinated with various stakeholders. a particularly complex set of design tradeoffs including content, people, and systems. There are both social and technical issues in the design and implementation of information systems. Bad design of information systems can be very costly (e.g., Fig. 7.50). Requirements specification and system design.

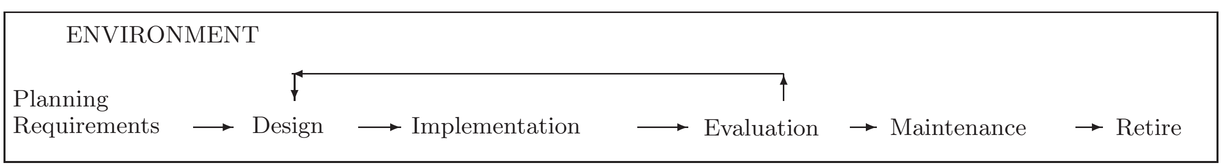

Figure 7.49: We briefly considered the system lifecycle earlier. We will go into much more depth and breadth. System design must consider the environment in which it occurs.

The software lifecycle depends on the system lifecycle – sometimes the software composes most of the system so they are close to identical. Once software is installed, it becomes enmeshed in an organization's other business processes (8.11.2); it is often extremely difficult to change. Thus, it is known as legacy software. Software as an information resource (8.13.8). Increasingly include off-the-shelf tools rather than custom-designed tools. Knowledge intensive firms. Intellectual capital: people, processes, relationships.

Challenge of legacy system upgrades.

> The FBI recognizes that its success will depend heavily on a greatly expanded role for IT. Given the complexity of its traditional and new activities, however, the FBI has found introduction of new IT technologies to be very challenging.
>
> Information technology's inherent capabilities must be fully exploited if the FBI is to use IT to help meet its operational needs. At this point, however, the FBI is not on a path to success in doing so.
>
> The most significant concern is that the FBI's technology efforts do not appear to be driven in a coherent manner by a strategic view of its mission and operational needs. The highest priority to address this problem is to formulate an enterprise architecture incorporating a detailed characterization of the Bureau's goals, tasks, and strategies. The FBI's efforts to do so, however, are late, limited, and far short of what is needed.

Figure 7.50: The FBI spent more than $170M on software development that was not useful. Here is an excerpt from a report about the management of that project[45].

7.9.1. System Analysis and System Requirements

We have encountered many types of systems. Collecting data and analysis of those systems is valuable for helping to understand them.

This approach can be applied to understanding information systems embedded in an organization. Requirements form constraints which **must** be satisfied by an implementation. This is often the first step in the design of information systems. Many aspects of a system need to be specified.

Functional Requirements

We might require that a car have seat belts or we might simply require that provide a certain level of safety of passengers in a crash.

Most often, we focus on functionality - those which describe what the system needs to do to satisfy its goal – but there are also requirements that go beyond the immediate functionality. These are so-called "non-functional requirements" which include content, usability, costs, and environmental requirements.

> The Shuttle will transport cargo into near Earth orbit 100 to 217 nautical miles (115 to 250 statute miles) above the Earth. This cargo – or payload – is carried in a bay 15 feet in diameter and 60 ft long.
> Major system requirements are that the orbiter and the two solid rocket boosters be reusable.
>
> Other features of the Shuttle:
> The orbiter has to carry a flight crew of up to eight persons. A total of 10 persons could be carried under emergency conditions. The basic mission is 7 days in space.

Figure 7.51: Part of NASA's requirements for the Space Shuttle[20].

Critical systems engineering.

Developing Requirements

A system developer might say "if you can just tell me what you want, I'll build it for you". This seems simple enough; however, some projects such, say, the design of the space shuttle are so complex and expensive that it is not practical to proceed by trial-and-error. There many tasks for which the goals are less clear than they were for the Space Shuttle. The goals may be highly dependent on cost or other parameters. Nonetheless, it is helpful to systematically organize and clarify the details into specific requirements.

There are many approaches for developing requirements. One strategy is to interview a range of stakeholders. Requirements elicitation procedures can be used. Techniques for extracting tacit knowledge from analyses of knowledge management (7.3.1). Information systems themselves can be helpful for keeping track of requirements.

A second strategy is collaborative development. Requirements elicitation is ultimately a human process

[41]. Joint application development[90]. A JAD is similar to a GDSS (3.4.3) but devoted to design and development. JAD sessions are: Focused and conducted in a dedicated environment JAD participants typically include: Facilitator, End Users, Developers, Tie breaker, Observers, Has social dynamics. This approach emphasizes the social aspects and compromise.

Stages in the development of requirements. Early Requirements are often general. The need for a system upgrade is often determined by strategic planning. This planning process requires understanding the environment and the tasks completed in that environment. We will explore an approach which incorporates goals into modeling and specification of business processes (8.11.2).

Validating and Evolving Requirements. Are they what the user wants?

The application profile is the range of tasks to which a system might be applied.

Requirements Management System (RMS). ISO 9000.

Requirements creep.

Structuring Requirements

Requirements engineering and knowledge representation. If the requirements are specified with a formal language, logical inference can generate implications. The more subtle aspects of use case design such as coordinating the use case with organizational and user interface design constraints and in the internal system constraints. The difficulty of capturing tacit knowledge (7.3.4). There are many variations to user analysis. Activity diagrams can be applied for developing use cases (3.10.2). While UML shows the expected interaction, a great deal more detail is needed. Processing flows, System architecture, Detailed requirements, System Interfaces/Impacts, Global requirements, Testing. A requirements specification document itemizes the requirements. Systematic presentation of requirements for clarity and completeness

Use Cases

Cast the requirements into a formal specification. One of the main applications for formal models such as Unified Modeling Language (UML) (3.10.2) which we have discussed are for the formal specification of requirements. We should have a systematic exploration of what needs to be examined with use cases. What tasks the user needs to accomplish. These are task models dovetails with interface design where the task contributes to the organizational activity. Need more complexity for use cases. One of the most common problem with user acceptance is that important functionality needs have not been considered. A further problem is it's difficult to extrapolate to an entirely new point in a product space (e.g., iPhone). Beyond use cases to Scenarios ((sec:scenarios)).

Interaction with users is the most immediate aspect of an information system. For most information systems, the interaction is essential. Ultimately, the design of information systems is about linking processes to people. This is sometimes accompanied by type of UML diagram known as an interaction diagram (3.10.2).

Use Cases: Specifying User Needs and Activities. User needs analysis. Difficulty of predicting user needs for changing environments. Use case diagrams (3.10.2). User driven service specification. For highly structured information tasks it is possible to associate user roles with the types of tasks those users may be expected to do. As mentioned earlier (1.5.2), these types of information should be matched with access points in the data set (Fig. 7.52). Use cases don't necessarily describe the full environment.

Ideally such process-based design would also consider social interaction action issues (7.9.6). Elaborating use cases is a step toward system implementation (Fig. 7.53). Fully dressed use cases. Designing tools and task models.

7.9.2. Implementing the Requirements

Business process engineering (8.11.2).

Type of User	Tasks
Customers	Select video
	Place order
	Change address
Managers	Check inventory
	Accounting
System Developers and Administrators	Monitor system performance

Figure 7.52: Some types of users for a Web-based online video store.

Tasks	Activities
Select video	search, read description, listen to preview,
Place order	submit order, pay with credit card

Figure 7.53: A few use cases for the end-user tasks described above.

Formal specification of a activities.

Design.

Formative and Agile Development Strategies

For many projects, it is difficult to develop complete formal requirements before beginning development. Or, at least, the requirements cannot be clearly specified. Formal requirements are often overly constrained — projects change and requirements evolve across time. In these cases, requirements development begins to resemble design (3.8.0). This is formative design and development strategy (3.8.3). Modularize with many benchmarks as a development strategy. Software prototyping[35]. Formative design with lots of testing. The idea of agile development has become a catch-all for a number of interrelated issues with traditional requirements development. Evaluating prototypes. User needs in the context of task completion. Software testing (7.9.5) is an essential aspect of agile development. Testing procedures need to be well specified. A variety of techniques have been proposed. Rapid application development (RAD). Extreme programming. Story-based design and scenario-based design. Interaction design (4.8.1). Formative design. Beta release.

7.9.3. Software Development

Once the requirements have been determined, the system must be built. That may involve customizing software or even developing new software modules. The software is a design artifact which is both highly malleable and complex. Service architectures such as described in ((sec:architecture)). System design is interwoven with evaluation metrics which we sill discuss later in the section (7.10.2).

Refactoring and code management for reuse.

The issues range from specific tools to managing very complex systems. Managing complexity (3.8.3) in code. Software is highly structured and that structure can facilitate development such as editing. Many programmers work by implementing models for their programs and integrating them. These can be explored with techniques such as "pseudo-code" (Fig. 7.54) and UML.

```
while not at end of list
  compare adjacent elements
    if second is greater than first
      switch them
    get next two elements
    if elements were switched
      repeat for entire list
```

Figure 7.54: Pseudo-code provides a conceptual overview of the program modules. Here, a procedure for sorting a list of numbers is described.

Cognitive aspects of programming. Programming languages as building representations[24]. Modularity and increasing degrees of abstraction. Ways of specifying data structures such as trees which we have found useful. Software development (7.9.3). Differences among programming languages.

CASE Tools and Software Development Tools

Programming implements a design (7.9.3). Many tools have been developed to support programming. A syntax-directed structure editor employs these syntax of the programming language to support the guide the user. Similarly, debugging tools. Syntax debugging. Semantic debugging. Software development can be supported with integrated development environments (IDEs) (Fig. 7.55). Modern development environments include the ability to search the web. Computer-Aided Software Engineering (CASE) tools include UML (3.10.2). A complete UML specification is often close to being able to be compiled into an object-oriented program. In addition, the the raw specification, several other support tools are useful such as version management tools (-A.5.8).

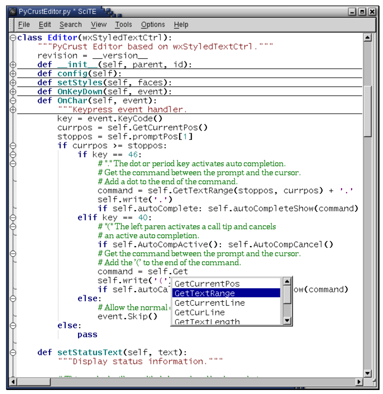

Figure 7.55: A screen from a software Integrated Development Environment (IDE)[8]. The same system could include compilers, version management, and even requirements. The editor for subroutines from a computer program "knows" the structure of the programming language. (check permission)

Software Libraries and Re-Use of Software Modules

Following the principle of object-oriented design, modules should maximize cohesion and minimize complexity (3.9.3, -A.3.4). Software itself can be treated as an information resource. Software can be collected into libraries in the hope that modules can be re-used. Information seeking for software re-use. Some software libraries, such as those which provide common mathematical functions, such as finding square roots, are routine. More specialized software could also be reused. One of the ideals of object-oriented design is the capability to share modules. Other collections are based around specific genres or media. Software and design libraries can facilitate reuse as part of the development process. Software collections must be carefully selected and organized; they should be easily accessible and persistent. Code metadata can describe the attributes of a program. Faceted classification systems (2.5.3) seem particularly effective.

Software, like other information resources, may be collected in libraries. Software management as information seeking. Software repositories[27]. Users are almost always interested in the functionality of the software more than the details of the code, so metadata and text descriptions are especially useful. Sourceforge — the collection of freeware source code — is a type of software library[22].

Software Development Organizations and Teams

Most software development is too much for one person to complete. Collaboration tools for supporting programming teams (5.6.2). Software development in an organizational context and is often a collaborative process. Because software development is an intellectual activity, it is difficult to understand the steps involved. Project management (8.11.3). These tools can also facilitate software project management. Coordinating large software projects. Even with detailed requirements, a great deal of development is needed.

Open source software (8.13.8) has a fundamentally different approach to software development[71]. Freelibre. The nature of the team may be affected by the development strategy and by business considerations. However, open source development is not effective for cutting-edge innovation [?].

Managing Software Development Teams Indeed, the development teams may be geographically distributed so this is an area in which collaborative support is helpful (5.6.7). Distributed management and consensus building. Software development communities. Management of software teams (e.g.,[36]). Implicit understanding of the task and the team members and project management. [74].

Adding people to a project team late in the software development cycle when that project is already late makes the project even later. This is has been demonstrated so many times that it is now known as Brooks' Law[37] after the software engineer who first proposed it.

The size of the development team needs to be coordinated with the software modules that are to be developed and there needs to be knowledge management for these teams. Information systems can improve communication and distribute necessary knowledge throughout the production process. Efficient processes coupled with improved communication allow organizations to confront unique situations.

A scrum is a type of programming team which employs overlapping phases of development[19]. A type of project management (8.11.3).

Coordination across software teams.

Tools for Supporting Software Teams Modularity is helpful for convenient development. Managing complexity of several interacting modules.

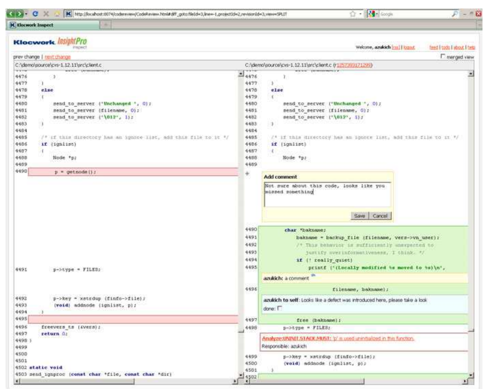

Figure 7.56: A collaborative code review interface allows comments from the review to be inserted and ultimately for approvals to be given. (check permission).

The process of systematically, recording bugs and attempts to fix them is known as bug tracking. Fig. 7.56 shows an example of collaborative development interface

7.9.4. Software Evaluation and Software Performance Standards

Difficulty of cost estimates for uncertain projects.

Software Quality Control

Attributes of quality software[34]. Consistency, impact. The principles we discussed earlier for quality control (1.4.3) can be applied to software quality. Design techniques and processes may improve quality. More formal methods and organizational processes. Software inspections. Code reviews. Quality and consistency of the design (e.g., ISO 9000). Total Quality Management (TQM) [47]. Process of improving quality. However, there are costs in over-emphasizing processes for quality. By allowing project managers to analyze the past production schedules of different products, prior methods can be used, discarded, or altered to meet the current demands. This allows to the organization to be adaptive, or to learn. Formal methods such as proofs of program correctness have been difficult to apply.

$$Plan > Do(Implement) > Check > Act$$

7.9.5. Software Testing and Software System Testing

Increasingly, we depend on software for critical functions but software sometimes fails (Fig. 7.57). Indeed, almost all complex software probably has bugs. When software is used in a critical system, failure can be dangerous. Controls may prevent such disasters when computers are involved in banking, airlines, and nuclear power plants. Software quality is essential for trustworthy systems. Confirmation testing.

> The failure of the Ariane 501 was caused by the complete loss of guidance and attitude information 37 seconds after start of the main engine ignition sequence (30 seconds after lift-off). This loss of information was due to specification and design errors in the software of the inertial reference system.

Figure 7.57: The Ariane 501 rocket was lost in 1996 because of a software failure[63].

Software needs to be tested throughout the software development process and as it is released for use, This testing can compare its performance to the requirements. Indeed, the testing performance can be part of the metrics (7.10.2) and it is integral to agile software development (7.9.2).

There are several testing strategies and methodologies: Testing can be intertwined with agile requirements development (7.9.2).

Development of performance standards for tests as the primary criteria for system features. This exactly specifies the criteria the system must meet. Test first[56]. The testing can be diagnostic – the test results can also be useful for suggesting bugs in the development processes.

Testing a program after changes have been made to it. Testing systems under load can often be helpful in revealing bugs. Testing can also look for unexpected conditions.

Normative software problem maintenance. Bug reports and bug logs. Open bug reports.

Many software systems are so large that they cannot be tested as a whole and must be tested as modules. Systems such as the telephone control switches may have more than one million lines of code. Can the system be tested as a whole, or is it adequate to test just the components. The testing of those components which have been changed is known as "regression testing"[17].

The use cases lead directly to testing. Joint Review and Testing by team (JRT).

Legal responsibility for software failures. Broader social context for software tools.

7.9.6. System Configuration, Documentation, and Roll-out

When an information system has been deployed its settings need to be configured. There are many setting in complex information systems. Keeping track of a settings is itself a task. Matching parameters of the environment in which it be used. Organizations and culture. Social media such as for support of trouble shooting.

Configuration Management

Tailoring setting of a complex system for a specific situation (Fig. 7.58). Configuration of setting on a personal computer. Software configuration. Configuration rules (-A.5.6) and constraints (3.7.2). Simple example. Dependency graph for showing constraints (8.11.2). Ultimately, we would like the configuration to match specification of the requirements and constraints of this particular application.

User Training and Documentation

We also need to make sure the users are on-board. Although we might like all interfaces to be intuitive, that is rarely the case. Users with different needs. Tradeoffs in training and documentation. Ideally, an information system would be intuitive, so the user can understand how to complete actions with a

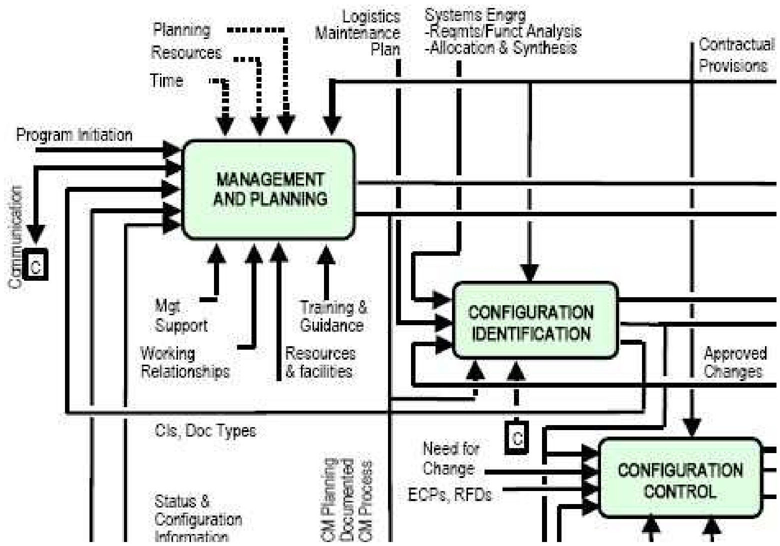

Figure 7.58: Complex information systems have many settings to adjust them to a given environment. The settings are often interdependent and changing one may also require changes to others. (redraw) (check permission)

minimum of confusion but this is rarely entirely feasible. Similarly, the system should require as little training as possible. When these ideals are not possible, supplemental training and documentation is required. Training and documentation can have a large impact on effective use and satisfaction.

Task-based documentation.

Training about information systems needs to be coordinated with other work processes. Effective training can be facilitated with knowledge of the task in which the user is engaged or on a user needs analysis. While the goal of user interface design is to make the interface as intuitive as possible, users still need training about most systems. This training can be based on generic educational principles such as developing conceptual models. In addition, people may learn about a system simply by watching and talking with other users.

Analogies for Explaining Interfaces

How the user may think about the interface?

Analogy can help to simplify complexity (6.2.3). Thus, a group of directories and files on a computer may be explained as being "similar to a filing cabinet". Or, we might explain atoms by saying they are like little planetary systems. On the other hand, can a word processor be compared to a typewriter in the same way? While metaphor can be helpful, it must be applied carefully; thinking of a word processor as a typewriter will not give a user a complete idea of the word processor's capabilities (6.3.4). The type of conceptual models given to students can impact on their ability to think abstractly.

Figure 7.59: Application documentation example.

In documentation, it is desirable to identify typical tasks the user is likely to complete. One strategy is known as "minimalist" instruction. This attempts to provide just the right amount of focused support [83]: (a) Choose an action-oriented approach, (b) anchor the tool in the task domain, (c) support error recognition and recovery, and (d) support reading to do, study, and locate.

The documentation of system features provides support to users. It is more effective when applied to actual work that a user is trying to complete — task-oriented documents perform this function. There is a question as to whether task instruction is preferable to documentation. Documentation and user needs/tasks. The need for documentation can vary depending on whether the system is being used for routine tasks or more unstructured tasks. Sometimes documentation is developed according to system functions rather than features. Increasingly, help for users is available from posting on blogs or on question-answering sites.

Interactive Help Systems and Interactive Performance Support

Queries to help systems. Imagine that you were writing a message with a word processor but forgot how to check the spelling. Ideally, an information system might track the actions of the user, and, with the task and system models, it may be able to figure out what the user is trying to do. So, when the user requests help from the computer, hat help is enhanced by the computer's understanding (even if it is quite limited) of the user's aims. The system typically knows the choices available to the user; it may have some history of what the user has been doing and knowledge of the range of tasks. Task models (7.9.3) can provide a framework for this kind of context-sensitive help. The method shares aspects of tutoring and training, such as assessing how to intervene with tutoring (5.11.3). This can be thought of as a plan recognition problem (3.7.2). The trick is to identify the "root cause" of the activity. Beyond providing explanations, the system might attempt to correct user errors.

Information Technology System Acceptance

There are many cases of failures to have information systems accepted. This is of information systems in businesses as well as libraries and hospitals. Ultimately, an information system has to be accepted by the users. Acceptance is often not trivial and has been explained as based on two factors: ease of use and usefulness [46]. Such perceptions are characterized as being based on attitudes and beliefs (4.5.0). Indeed, the likelihood of technology adoption can be indicated by statements of behavioral intention (4.5.2). Many attempts to integrate information system into an organizations have meet with failure — either the new technology does not reflect the organization's overall strategy, or it does not reflect the organization's established processes and the technology is not accepted (Fig. 7.60) [49]. The Technology Acceptance Model (TAM) proposes that acceptance of technology depends on two factors: Perceived ease of use and perceived usefulness. Self-efficacy and TAM. TAM and individual learning. Generational effects of TAM are related to the notion of "digital natives". TAM and group learning (culture). Factors contributing to Perceived usefulness include: norms, image, relevance, quality, and demonstrability. Ultimately, TAM is an application of the broader Theory of Reasoned Action (TRA (4.5.2).

Acceptance Factor	Description
Relative advantage	How much of an improvement is the new system over the previous one?
Compatibility	Does the new system support existing processes and structured?
Complexity	How difficult is it to understand the use of the system?
Trialability	Can it easily tested?
Observability	Is it transparent?

Figure 7.60: Factors in information system acceptance[49].

7.10. Organizational Information System Management

Once the information system has been introduced, it needs to keep it running well. Management of information systems, setting policies and making decisions about them. Changing environments often mean revisions to requirements. Managing the operation and effectively managing the entire software lifecycle (7.9.3). This is sometimes termed IT governance [10]. It is not unusual for maintenance costs of an information system to greatly exceed the development cost. Some of these issues also relate to management of content (1.5.2). IT groups are increasingly influential in organizations. Preservation planning as IT governance. Digital preservation and ITgovernance.

7.10.1. Organizational Frameworks for Information System Governance

There are many structures for organizations. We can break down a complex information system in components to deal with them effectively. One breakdown of information system components which can be used for management is shown in Fig. 7.61. This is a technologist's view of the organizational sub-system rather than a breakdown into subsystems based on business-function structure

Subsystem	Description of Subsystem	Example of Attributes
Input	Data Entry	Accurate input
Communication	Networking	Encryption
Processing	Operating systems	Accurate calculations
Storage	Database, File Management	Referential integrity
Output	Printing	Report printing capability
Boundary	Interface between system and outside world	Passwords

Figure 7.61: Subsystems of a typical information system (adapted from[85]).

Beyond the system view, there is a context in which it works. An information system should work seamlessly with organizational needs and goals. Information systems have many components and the following areas need to be managed[85]. There are many threats for the integrity of information systems and the information they contain. Human organization issues (Fig. 7.62). The implementation of those policies is essential. A policy must be enforced to be effective.

Organization	Examples
Top management	Staff, Leadership.
Systems development management	Planning, Design, Testing.
Programming management	Software quality and updates.
Data resource management	Content management.
Security management	Prevention and recovery from damage.
Operations management	Maintenance, workflow.

Figure 7.62: Organizational units involved in system management (adapted from[85]).

7.10.2. System Metrics, Audits, and Evaluation

There any many types of information systems and applications so evaluation of covers many perspectives. Many level of evaluation are needed. Because of the complexity of the information systems, there are many aspects which need to be evaluated. We have discussed evaluation in many places; here we summarize those issues. Quality. Prevention of defects, and correction.

An audit is a systematic evaluation of a system. Audit log for access and use. Identify subsystems which can be monitored: Understand the controls, Test the controls, Examine the results of the controls. Audits could inspect procedures and transactions. Information systems provide access control and better record-keeping about access. An audit trail provides a trace who has accessed which records.

System Metrics

Evaluation from successful usage but this is fuzzy. So, we emphasize, metrics for providing a standard for what is happening. Do the metrics actually test what intending to test? Construct validity. Need for organizational infrastructure as in trusted repositories. Evaluation. Outcome-based metrics. Ultimately, metrics should be ground in measurement theory (7.10.2).

Complex systems are so difficult to manage that they almost inevitably have failures. We have considered techniques for managing complexity in distributed systems (7.7.1). Thus, handling such failures should be built into the system. Accidents are common even in carefully designed systems[70] They occur from multiple failures and cascading failures. Often this is as much an organizational failure as much as a system failure. The system should meet the requirements. Ideally, we should refer back to design goals and the requirements to system performance. Scope of a metric. Goal, Question, Measures (GQM)[31] (Fig. 7.63).

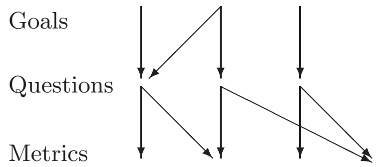

Figure 7.63: System design goals suggest questions and those questions suggest metrics.

System attributes: Number of users supported at one time. System metrics include availability (e.g., 24 X 7), the amount of down time, and security. Response time for user queries. Meeting a business model. Cost-effectiveness.

Organizational Environment and Content Metrics

Is the organization stable? Does it have reasonable policies? Disaster recovery plan.

The content metrics need to be matched to the tasks the users will complete. Many aspects of content management have already been discussed with respect to library services (7.2.2). We may evaluate the scope of the collection. The simplest approach is to count the number of items. How many records, what speed of access, preservation, quality of telephone reproduction? Database integrity. The "coverage" of my Agatha Christie collection can be measured by the proportion of how many of Christie's books I own out of all that she wrote. More generally, I am interested in the breadth and depth of coverage of a collection[89]. Coverage, quality, exhaustiveness. Implementing policies such as those for archives.

Attributes of quality: having few errors but also extensible to new contexts and new services.

Metadata quality.

Usability and Usage Metrics

Systems should allow users to be effective for accomplishing tasks. This means that they should not only be useful for the tasks for which they are designed, they should also be "easy" to use (Fig. 7.64). Checklist of features. Design and evaluation of interactive systems (4.8.0).

Aspect	Description
Learnability	Is it easier to learn how to use this information system than other systems?
Memorability	Once a person has learned to use an information system, can he or she remember how it works when they come back to it later?
Satisfaction	Does the information system change the user's job satisfaction?

Figure 7.64: Aspects of usability for an end-user (adapted from[67]).

ISO definition. Effectiveness, Efficiency, and Satisfaction

Many users show strong preferences for aspects of systems that do not seem to have any real impact on their effectiveness. Although there is no strong performance enhancement with color monitors rather than black-and-white, but today, almost all monitors are color. Information retrieval system and perceived effectiveness (3.3.3). In any event, user satisfaction is an important component of long-term system acceptance. Satisfaction may also increase motivation which is critical to use of educational systems. Attitudes, long-term usage.

Many approaches have been proposed for evaluating usability (Fig. 7.65) and there are tradeoffs about when these methods should be used. There is considerable debate about the value of interviews. In the "thinking-aloud" evaluation method, users talk about their experiences; while verbal protocols can be useful for determining a person's state, they may be biased (4.3.2). "Discount usability testing" can give quick feedback on design. As few as two or three trial users can provide sufficient feedback to detect major design problems[67].

Label	Description
Expert reviews	Analysis by experts of system components.
Heuristic evaluation	Apply rules of thumb.
Experiments	Systematic examination of variables. Typically, these are rigorous but may be difficult to generalize to complex situations
User surveys, interviews, and observation	Observation of the use of the system in a natural environment. These are less systematic than experiments but may reveal hidden clues.

Figure 7.65: There are several techniques for evaluating the effectiveness of user interaction with information systems.

Web usage metrics. Server logs (-A.14.2).

7.10.3. Information Security

An information system is not useful if we can't trust its contents. Corrupted information can have many implications. One example of lax security is shown in Fig. 7.66. This involves both content management (1.5.2) and systems. Basic procedures. Personnel files. Work product. Protecting an archival or master copy (7.5.1). These are system and policy issues. Security is a result of simple user choices and of group interaction. Trust (5.2.3). Trusted archival repositories (7.5.6).

Confidentiality and Encryption

There are many times when information needs to be kept confidential. Confidentiality of information, integrity without compromising usability. Encryption (-A.13.1). Both social and technical issues.

Data breaches. Some serious and some not.

Why Security Matters

Thus, security cuts across many levels from policy to implementation. Thus, security must address both social and technical dimensions. That is, the technology can protect the content in some points in its management but security at other points in fundamentally a human activity. We can apply the sub-division of the sub-systems described earlier (7.10.1). Scope of the system to include communication. Organizational policies which promote security. We have also seen these issues in the context of Trusted repositories (7.5.6). One type of attack in cyber-war (7.11.4) is to corrupt, or at least create distrust in the accuracy of information. Principles for developing secure systems. Advanced persistent threat (APT). Spear-phishing. Developing attacks through personal reconnaissance.

The tendency to protect all information but it makes more sense to prioritize information to be protected and the threats to it. Securely deleting data from digital media. There are real threats to information that is exposed out on the network. Authorization Managing policies: Metadata and descriptions for information security such as (2.4.0).

Forensics examines artifacts for evidence. Digital forensics plays a role in examining the authenticity of documents held by archives. It is also applied in criminal and security proceedings. For instance, a disk may be recovered as part of a police raid and its contents need to be examined for evidence relating to criminal activity.

Relevant data may include the context of other files on a disk. Digital forensics and its flip-side data cleaning need to consider that there may be several layers of pointers to data on a disk. Cleaning a disk needs to consider that simply destroying pointers to data is different from destroying the data records themselves.

Security within the firewalls.[87]

Although still in high school in 1968, Schneider started a company called Creative Systems Enterprises and began selling electronic telecommunications gadgets he invented. Each day as he passed the Pacific Telephone and Telegraph Company office, he scavenged the firm's dumpster for discarded equipment that could be used to build his gadgets. He also collected a wide variety of documents, ranging from invoices to training manuals. Within just a few years, he became an expert on telephone company technology and business, and reportedly knew more about Pacific Telephone's telephone equipment supply procedures than any of its employees.

In June, 1971, Schneider set into motion an elaborate plan to steal new telephone equipment from Pacific Telephone and resell it as refurbished equipment through Creative Systems. Eventually the scam would net him hundreds of thousands of dollars worth of Pacific Telephone equipment. Schneider accessed Pacific Telephone's computerized ordering system and by using a telephone card dialer succeeded in placing orders for equipment. To complete the scam, he needed to learn the telephone equipment budgets for individual telephone company's sites, equipment inventory levels and other key pieces of information. He gathered the required information by getting access codes to a commercial time-sharing service used by the telephone company for inventory control and parts distribution.

In January 1972, acting on information provided to them by one of Schneider's former employees, law enforcers raided Schneider's offices and a warehouse where they found equipment the district attorney said was worth $8,000. They also learned at that time that Schneider had stolen a total of $125,000 worth of equipment. Later, Schneider would admit that he had taken close to $900,000 worth of goods.

Figure 7.66: Effects of a security breach[18]. (check permission)

tavi	coo	elisabeth	mail	news	nobody	operator
calvin	elissa	pam	ftp	games	pcap	sshd
konchog	cobra	pass	adm	bin	rpc	smmsp
jflores	apache	lp	mysql	squid	mail	carloscota

Figure 7.67: Fragment of an actual security report for a Linux machine. Some hacker is probing the machine through the network trying to find valid user accounts. If one of these accounts were found to be active then the attacker would test a list of potential passwords against it.

Black hat, white hat.

Privacy depends on data security. Security breaches. Corruption of files versus obtaining copies of sensitive data. Loss of credit card information. Loss of health information.

Human versus Technical Solutions for Security

The major difficult for password security are the users rather than the encryption. Users often don't bother with firewalls or backups of their own files. Incentives for protecting one's own files with firewalls and backups. Organizational procedures

Vulnerabilities for Information Security

The cost of failure of an information system can be very large. When a society relies heavily on information systems, it might even be possible to disrupt that society by disrupting the information systems (7.11.4). Essential information can be corrupted; incorrect information can be deliberately planted.

Risks are a combination of technical and human aspects. Human aspects of information assurance. Some of the most serious threats to information systems come from insiders. These problems may range from simple insider abuse of resources to malicious attacks by disgruntled employees.

Attacks on different sub-systems described in Fig. 7.61. Attacks from outside the system. Connectivity to the broad public network has greatly increased the need for security. Some specific mechanisms for attacks include. Viruses (attached to other software) and worms (standalone). Trojan horse is hidden in something else. Entry into system via peripherals. Packet sniffing. Snooping on the network. Denial of service.

Attributes of security: confidentiality, authorization.

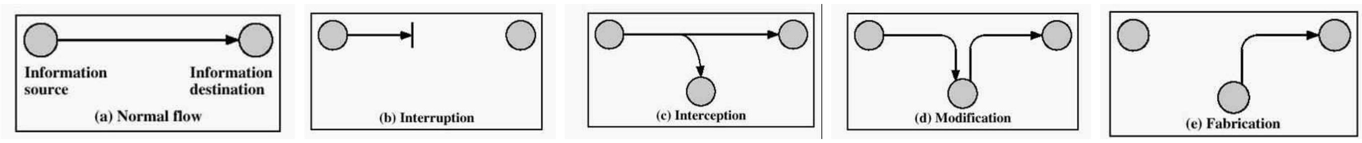

Figure 7.68: Types of attacks on network transmissions. (redraw) (check permission)

A large part of the security is the human procedures behind it. Security failures are often human rather than technical. Honeypots are sites set out by investigators to draw individual who may inclined to engage in unauthorized or illegal activities. This lure can also be useful for assessing the strengths and strategies of the attackers. It is difficult to control distributed networked services. Access to children. Policy and privacy issues. Do programs actually do what you expect them to? Operating system security.

Security Risk Analysis and Management

Given that there are risks, how do we access them for a given student? Absolute security for information is probably impossible; rather, it is more useful to analyze these problems with a cost-benefit analysis (7.10.3). When there is the possibility of a threat, the effort of risk-prevention activities need to be prioritized. More systematic vulnerability analysis. There are many possible failure points. Risk analysis and calculating the probabilities of threats. Particularly, for analysis of attacks. Analyzing and predicting risks. Scenario models. Difficulty of risk evaluation in the context of "group think" and conformity of opinions. Game theory (3.4.1) analysis

Security policies. Security policies and argumentation (6.3.5). Risk and argumentation.

Security audit as with other audits (7.10.2). Perception of risk[78]. Argumentation and risk assessment (6.3.5).

The human aspect of security risks can be modeled as adversarial situations and actions determined by game theory (7.11.0). Risk analysis may include likely strategies for attack. Once the risks have been determined, then protection strategies can be prioritized and developed. Includes cost matrix.

Security Risk Management Once the risks have been identified they need to be addressed. When the threat comes from a person or group we can consider what strategies they may adopt by applying game theory (3.4.1).

Security in financial transactions.

There are multiple levels at which security needs to be managed. Analysis, system and network, physical protection. Organizational processes. Security requires decisions. Best practices for IS management.

Developing systems that resist attacks. Encryption. Secure servers. Physical security for information is also important. Media should be protected from vibration and heat. Prevent users from accessing information.

Seemingly simple precautions are often effective. Passwords. Two-factor autentication. Multiple back-ups and out-of-state.

Prevention. Investigation procedures.

Cyber-army against the pro-democracy movements.

7.11. Highly Adversarial Situations

Information is often crucial in adversarial situations. Information security (7.10.3). Politics and economic self-interest are also adversarial but here we emphasize more extreme crime and war.

Offense and defense. Protection of evidence. Coordinated law-enforcement databases.

False information as a systematic attack. Attacking the information infrastructure.

Malware infections. Organize infected computers into botnets. This botnets can create coordinated attacks.

Phishing.

Espionage. Systematic attempts to uncover protected information. Commercial and governmental espionage. For instance, email among members of a corporate negotiating team may be attacked.

Increasingly, cybercrime and information warfare overlap.

Distinguishing police from military.

Defense: Assessing Terrorist Networks Information security (7.10.3). Social networks (5.1.0).

Collaborative emergency management.

Defense and prevention of attacks. Decision making for determining attacks.

Tracing the source of cyber-attacks.

Cyber-reconnaissance provides intelligence about individuals.

Systematic Disinformation

Disinformation (5.3.3). Attacks on history ((sec:attackhistory)).

Soft Power

Peoples and cultures often influence each other indirectly. When the influence of one culture on another becomes pervasive, it becomes a type of power[69]

Public opinion on the home front, in the opponent, and in third parties. This is an extreme form of attitude change. Persuasion (4.5.2). Culture (5.8.2). Battle for hearts and minds.

“Winning hearts and minds”. Brainwashing.

7.11.1. Signals Intelligence and Intelligence Analysis

Could be competitive intelligence (7.3.5) but the consequences here can be greater. Collecting information by surveillance of different media. [**?**]. This is related to competitive intelligence (7.3.5).

Much of this is data mining.

Decoding messages. Traffic analysis.

Extracting a hidden story line (VAST).

Satellite images and image processing (9.10.2).

Critical analysis must consider the substantial chance of deception.

Many applications of sensors (-A.19.0). Dragon fly sensors (Fig. 7.69). This also raises privacy issues (8.3.1) and ethical issues.

Dangers of autonomy of lethal robots. This often minimize responsibility. Various battle robots.

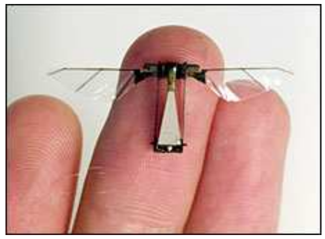

Figure 7.69: Artificial dragon-fly sensor. Eventually, this device might carry microphones and cameras[1]. (check permission)

Providing Decision support systems to support security analysis.

Complexity of intelligence information.

Figure 7.70: Complexity of intelligence information.

Cyber-exploitation. Collecting information through the network [?]

Loss of intellectual property through industrial espionage.

7.11.2. Information and Warfare

Information has always been essential for gaining an advantage in warfare. Indeed, there are several dimensions to information operations. Information systems can be fragile. Danger of formal reasoning systems when human activities and inference is involved. Art of War[2] (Fig. 7.71).

> Victorious warriors win first and then go to war, while defeated warriors go to war first and then seek to win.
> Pretend inferiority and encourage his arrogance.

Figure 7.71: Quotations from the *Art of War*.

Warfare often provides confused situations then information is essential. The fog of war.

Defense and counter measures.

Plan recognition of actors with hostile intent.

Tactical Intelligence

Spying.

Intelligence analysis. Collecting evidence.

Ghostnet.

In military combat, intelligence is used in both direct and indirect ways. “Intelligence” requires developing an understanding of what is actually happening. Directly, military intelligence is responsible for acquiring as much information as possible regarding opponent troop strength, locations, and strategies. This may make use of direct, visual information gathering, or it may utilize satellite imagery, prediction models, or intercepted enemy communications.

Timely information and analysis of that information is often not available in the “fog of war”. Rapid decision making is critical. Reconnaissance. Information can assist the actions soldiers and their commanders take on and off the battlefield. Netcentric warfare is a catch-all term used to describe the new and advanced ways that the military is moving information amongst its parts. Military commanders may use decision evaluation software or even game theory to determine strategy and decisions using newly acquired intelligence. Logistical supply problems may be resolved using integrated supply chains. Personal battlefield computers are also being used as knowledge management systems designed to help soldiers make better battlefield decisions with all available information at hand.

Psychological Operations (PsyOps)

"Psy-ops," or "psychological operations" is a military strategy that attempts to discourage or disorient an opponent. This may include such far-flung tactics as dropping informational pamphlets from planes and playing loud, disheartening music.

Disrupting the social decision process. Information is also effectively being used in far more subtle ways during wartime. Hackers can attack the local cyber infrastructure, leading to enemy information isolation. Alternatively, disinformation warfare can be fed the same way. This is known as information warfare.

This is informational warfare on the cultural and sociological level. This tactic seeks to break down social boundaries between the two warring countries, though this is only perceived by one side, and make the cause of the war seem pointless, and the ideas and philosophies of the other side seem acceptable. In this regard, information and persuasiveness about social values is part of the point.

Intimidation, Terrorism, and Panic

Terrorist group organizations.

Complications of monitoring personal information.

Counter insurgency.

Categorization of content across media types. Sleep deprivation.

7.11.3. Netwar

Social networking facilitating coordination and infomation exchange by attackers.

7.11.4. Cyberwar: Attacks on Networked Information and the Information Infrastructure

Defense and offense. Strategic information warfare. Dominance of an opponent's information systems.

Cyber-security. Limitation of firewalls for defense. Problem of infection from consumer grade equipment. This can be a real problem but can also be exaggerated.

Defense by extensive encryption. Battlefield oriented cyber-war versus attack on military infrastructure. Disruption to command-and-control in the battlefield. Determining the source of attacks can be difficult. Swarm from multiple directions.

Attacking information infrastructure. This could mean scrambling data to make it unreliable or it could mean attacking the network itself such as knocking out the routers. Attack fundamental aspects of society: the financial system, the electric power grid, and traffic control in major cities.

Coordination of the elements in a non-hierarchical organizational structure. The principles of virtual organizations described earlier also apply to organized crime and terrorist groups [12]. Even the command-and-control structure of the traditional military is changing.

Insurgency versus revolution.

Terrorist and markets [33]

Many aspects of distributed information systems help terrorists. They may use web sites spread what is often incorrect information.

"Information war has no front line. Waging information war is relatively cheap. Boundaries are blurred in cyber-space. Opportunities abound to manipulate perception in cyberspace." [66]

Netwar is anti-hierarchical similar to the way that organizational hierarchies have been degraded (8.11.2)(Fig. 7.73). Pulsing attacks from the swarm.

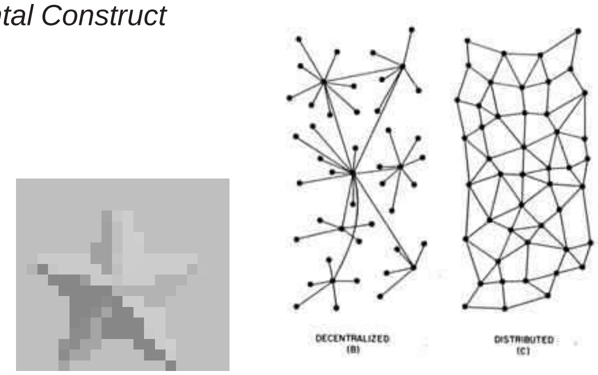

Figure 7.72: Centralized (left - to be drawn). Hierarchical organization versus situated context in a network. (redraw)(check permission)

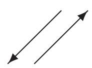

Figure 7.73: Centralized (left - to be drawn). Hierarchical organization versus situated context in a network.

Attacks by coordinated and often specifically teams of hackers can disrupt normal social interaction.

Botnets. Bot attacks.

Virus delivery by USB. Stuxnet worm attacking industrial control systems. Possibility of destruction in the physical world caused by illicit control of these systems.

Information security (7.10.3).

Cyber-war attacks the information infrastructure [29]. Disabling essential systems. More subtly, the attacks could make information less trustworthy.

Defense for cyber-attacks on individuals.

Stuxnet.

Many possible targets such as the electric power grid and the Internet itself. Defense against cyber-attacks. Security (7.10.3).

7.11.5. Adversarial Planning

Planning (3.7.2).

Game theory (3.4.1).

Exercises

Short Definitions:

API
Archive
Audit trail
Best practices (preservation)
Best practices (security)
Bureaucracy
Business rule
CASE tools
Chain of custody
Client-server system
Competitive intelligence
Conspectus
Dark archive
Data-flow diagram
Disaster recovery
Distributed searching
Emulation
Encapsulation
Federated system
Finding aid
Free and Open Source Software (FOSS)
Functionality
Habitable language
Heuristic evaluation
Honeypot
Information audit
Lessons learned
Load Balancing
Middleware
Micro-service
Migration
Peer-to-peer
Permanence rating
Persistence (documents)
Protocol
Provenance
Botnet
Cyber-crime
Data mining
PsyOps
Signals intelligence
Redaction
Regression testing
Repositories
Root cause
Stovepipes and silos
Strategic intelligence
Subject matter expert
Tactical intelligence
Tacit knowledge
Transaction (database)
Trusted repository
Walk-through
Web services

Review Questions:

1. How is a library similar to or different from an information system? (7.0.0)
2. What procedures does a library typically employ to demonstrate S.O.A.P.? (7.1.2)
3. In what way is a "digital library" similar to or different from a traditional library? Hint: Think S.O.A.P. ((sec:digitallibraries))
4. In what ways are records essential for modern society? (7.4.1)
5. What are the advantages and disadvantages of moving records from the organization that created them to a separate organization for managing records them. (7.4.1)
6. Explain what is needed for stored documents to be considered reliable evidence especially for them to be accepted as legal evidence. (7.4.2)
7. How are electronic records useful in business process management and re-engineering? (7.4.1, 8.11.2)
8. Distinguish between "libraries," "archives," and "enterprise management systems" ((sec:traditionallibraries), 7.4.4, 7.5.1)
9. How do libraries differ from archives? (7.2.1, 7.5.1)
10. Explain the difference between "preservation" and "conservation". (7.5.1)
11. Give examples of the strategies for handling complexity described in Fig. 3.46. (7.7.1)
12. Explain how peer-to-peer systems might create indexes. (7.7.1)
13. Describe the reasons distributed collections might be preferred to centralized collections. (7.8.0)
14. Why are repository servers better than Web servers for managing collections. (7.8.0)
15. Provide examples of three levels of training on an information system. (7.9.6)
16. What is an information system audit? How might an audit differ between a database and a digital repository. (7.10.2)
17. What are some criteria by which an information system might be evaluated? (7.10.2)
18. Distinguish between "risks" and "costs" (7.10.3)
19. What is intelligence analysis? (7.11.1)

Short-Essays and Hand-Worked Problems:

1. Is the Web a "collection?" Is it a "library?" Why or why not? (7.2.1)
2. Should community libraries be replaced by community technology centers? Explain. ((sec:traditionallibraries))
3. What are the similarities and differences in services provided by a video rental store and a public library? ((sec:traditionallibraries))
4. State a collection policy for a collection dealing with (7.2.2): a) animals of Australia, b) The Silk Road, c) information systems.
5. What is the business model for a public library? Who are the competitors for a public library. (7.2.2, 8.11.4)
6. Identify a situation in an organization with which you are familiar where information would be helpful. (7.3.1)
7. What are some metrics for effective knowledge management? (7.3.1)

8. Describe the electronic records needed to be kept by (a) a chemical company, (b) the marketing department of a manufacturing company? (7.3.1)
9. Pick an industry such as automobile manufacturing. Do an environmental scan. Pick a company and do a competitive intelligence (CI) analysis from the viewpoint of that company. (7.3.3)
10. Pick an industry such as automobile manufacturing. Describe the organizational culture of different companies. (7.3.3)
11. Tacit knowledge. (7.3.4).
12. What is the role of information in creating a "learning organization"? (7.3.4)
13. Describe a taxonomy that would be useful to describe lessons learned for a manufacturing company. (2.2.2, 7.3.4).
14. Suppose your company proposed to start a new product line. What information resources would you make available to the employees in the company. (7.3.5)
15. How should we value information? (7.3.5).
16. Find market share for three companies in the laptop business. (7.3.5).
17. Describe the business model adopted by each of the companies in the previous question. Describe the competitive landscape for them. (7.3.5).
18. If your company was planning to introduce a new product, what information would be most helpful to collect about a competing company? (7.3.5)
19. Identify information that is under-utilized in an organization. It is possible that XML could solidify existing organizational stovepipes. How might that be minimized? (2.3.3, 7.3.6)
20. What are some strategies for combining the organizational knowledge of when two business when they merge? (7.3.6)
21. Take an organization with which you are familiar. Identify the major information services and systems. Are they centralized? If not, what are the major barriers to that? ((sec:knowledgesources))
22. What are some of the reasons organizations develop stovepipes? (7.3.6).
23. When are silos appropriate for managing information in organizations? (7.3.6)
24. If you were an archivist and every 6 months you could select 1000 Web pages that would be saved forever, which ones would you select? (7.5.1)
25. To what extent does the integrity of a document depend on maintaining its original appearance? (2.3.1, 7.2.1, 7.5.1)
26. If you could archive 10 Web sites for posterity, which ones would you choose and why? (7.5.1)
27. Describe some steps that can be taken to improve the understanding of the context in which archival records are stored. (7.5.1)
28. Should a library ever remove any content? (7.5.1, 8.3.1)
29. If we could save all information, would we would we be able to find anything? (1.6.1, 7.5.3)
30. What policies would you suggest for keeping digital copies of (7.5.3):
    a) your personal photographs and movies,
    b) your local newspaper, and
    c) email messages among government officials?
31. In the course of your education, you will generate many documents such as notes, exams, bills, hopefully a diploma. Describe a retention schedule for these documents. (7.5.3)
32. What should you do if you want to archive a Web site? (7.5.3)
33. Suppose your job was to archive old newspapers but you had space to keep just three newspapers a month. What criteria would you use in deciding which ones to keep? (7.5.3)
34. Develop a policy for the preservation of a) university and b) corporate Web sites. (7.5.5)
35. Describe a complete museum exhibition you would like to develop. (7.6.1)
36. Explain some of the problems of encapsulation as a strategy for preservation. (7.5.5)
37. Describe several threats to the trustworthiness of digital repositories? Why might attackers want to delete or change records? (7.5.6) company that wanted to create a new position for knowledge management.
38. Your university student database probably runs on a single mainframe computer. Under what circumstances would it make sense to have the university database be a distributed system? (7.7.1)
39. What are some ways to defeat peer-to-peer systems? (7.7.1)
40. Peer-to-peer services. (7.7.1)
41. Imagine that you worked for a software design. Specify the design for a new information system that could be used by your university. (7.9.0)
42. What job description would you write for that position? (7.9.0)
43. Explain the limitations of the following requirements specifications a) A motor vehicle inspection system should handle 100 people per hour. b) The CPU should be at least 4 GHz.
44. Read about the Shuttle Challenger accident. Describe what went wrong. (7.9.1)
45. What requirements would you specify for an evoting system? (7.9.1, 8.4.3)

46. Develop task models for (7.9.3): a) Composing letters. b) Copy editing of a letter which has been marked up.
47. Identify a software library. Determine who are the target users. Evaluate whether their needs are being served. (7.9.3)
48. Describe what information an aircraft mechanic would have available when working on an airplane engine. How could interactive performance support help them? (3.2.1, 7.9.6)
49. Examine the documentation for an information system. Explain its approach. Describe its strengths and weaknesses. (7.9.6)
50. Collect and evaluate the training materials for an online information system. (7.9.6)
51. Pick a commercial ecommerce site. Describe how you would evaluate it. (7.10.2)
52. Suppose you had introduced a course-ware management system for the teachers in your local public school system. How would you evaluate its effectiveness? (7.10.2)
53. Design a plan that would protect a Web server from natural disasters and communication and power failures. Estimate the cost for your proposal. (7.10.2)
54. Evaluate the usability of a piece of software or an interactive Web site. Describe the intended users and tasks. (7.10.2)
55. What privacy controls should be in place for an information system audit? (7.10.2, 8.3.1)
56. What factors should be considered in determining whether a health information system is cost effective? (7.10.2, 9.9.0)
57. Explain security precautions for (a) a publisher's database, (b) a television production company's records, and (c) a Web-based catalog ecommerce records. (7.10.3)
58. Describe how you would conduct a "security audit" for an information system such as (a) a web-based auction site, (b) for a library. (7.10.3)
59. What audit procedures would be useful for bank transactions and accounts? (7.10.3)
60. If you were managing the data processing center of a bank credit card system, what precautions would you make sure that: the account balances had not been tampered with? (7.10.3)

Practicum:

1. Mini ethnography of system use.
2. Develop a preservation plan for a set of digital photographs.
3. Records management system.
4. Security audit
5. Organizational audits.

Going Beyond:

1. How should a public library determine its users' information needs? (3.2.1, 7.2.1)
2. Systematically observe patrons in a library of 30 minutes and describe what they are doing. (3.2.1, 7.2.1)
3. Find out what percentage of the people in your town have public library cards. (7.2.1)
4. Describe the workflow involved in cataloging a book for a library. ((sec:traditionallibraries))
5. Many libraries are not funded directly by public taxes. There are school libraries and corporate libraries. Explore the business model for one of these libraries. (7.2.1)
6. Describe what content you would select for (7.2.1):
   a) A library to serve just your family.
   b) Resources for a course on information science and systems.
7. To what extent is war a good metaphor for business? (1.2.1, 7.3.1)
8. When is a company liable for the actions of its employees in dealing with information? (7.3.1)
9. Suppose your job was to manage the information resources for a digital video distribution company. What approaches would you suggest for managing (a) external or (b) internal information? (7.3.1).
10. In Chapter 1, we emphasized the importance of representations for learning. What are some effective representation of information for organizational learning? (1.1.2, 7.3.4)
11. Is there a conflict between the notions of best practices and learning organizations? ((sec:tpractices), 7.3.4)
12. In what ways do metadata exchange standards facilitate the development of virtual organizations. (2.4.4, 5.7.3, 7.3.4)
13. How does information help an organization to learn? (7.3.4)
14. What issues should be considered in keeping strategic knowledge on a company's intranet? (7.3.6).
15. List the types of information stored in the finance department of a business. (7.4.1, 8.13.5)
16. The papers of important government figures such as Supreme Court justices are kept in the archives but sealed until five years after their death. Why are they kept sealed? Some people feel this is too long, other feel it is tool little. What are some reasons for each these positions? (7.5.1)
17. Develop a plan for the selection and preservation of a) public records and b) corporate records. (7.5.1)
18. Propose a preservation plan for a university professors papers which include so early word processing files. (7.5.1)

19. Who should be responsible for ensuring that digital materials are available far into the future? (7.5.3)
20. Should we preserve hate literature? (7.5.3)
21. In the U.S. government, the control of documents to be preserved is transferred from the originating organization to the U.S. National Archives and Records Administration. Why do you think this transfer is made? (7.5.5, 8.5.1)
22. How can peer-to-peer systems be used to create indexes on the content being held in the peer systems? (7.7.1)
23. How is Grid computing similar to or different from peer-to-peer computing? (7.7.1).
24. Imagine that you have been hired to lead a design team for an information system for a newly formed NASA project which is going to send astronauts to Pluto. Describe how you would approach this job. Describe some features the system you might build. (7.9.0)
25. Describe how the principles described earlier in this book could be applied to creating an effective JAD. (5.6.0, 7.9.1)
26. Can we develop requirements for a system which is intended to support novel tasks? (7.9.1)
27. Generate a requirements specification for an information system to be used by: (7.9.1)
    a) An organization tracking public health in a large city.
    b) An organization seeking to improve public school education.
28. In the development of services, what are the advantages and disadvantages of re-using existing software? (7.0.0, 7.9.3)
29. Create a task model for the steps involved in starting a car. (7.9.3)
30. Who should be responsible for software failures: a) the programmers, b) the employees, c) the firm which purchased the system, or d) the end-user? (7.9.5)
31. Diagram the states in an interactive performance support (i.e., sensitive to the context of user actions) system for a Web search application. (7.9.6)
32. Identify a group of computers which you believe has in security. Plan a security audit for them. (7.10.2)
33. What are the pros and cons of depending entirely on metrics? (7.10.2)
34. Develop a case study of a computer security incident. In your opinion, who is responsible for this incident? How could it be prevented from recurring in the future? (7.10.3)
35. Discuss the ethical issues in the use of honeypots for catching illegal information behaviors on the Web. (7.10.3)
36. Describe the security considerations in developing a certification authorization system for a Web site. (7.10.3, -A.13.4)
37. Netwar and cyber-war. (7.11.4)

Teaching Notes

Objectives and Skills: Design strategies. The students should be able to identify and discuss different types of distributed systems and services. Do a knowledge audit of an organization.

Instructor Strategies: This chapter could be divided and presented from two different perspectives: (a) systems and architectures. (b) services and the input of these services.

Related Books

- Allen, T.J., and Henn, G. *The Organizational Architecture of Innovation: Managing the Flow of Technology.* Elsevier, Amsterdam, 2007.
- Barabasi, A-L., *Linked: The New Science of Networks.*
- Beninger, J. *The Control Revolution: Technological and Economic Origins of the Information Society.* Harvard University Press, Cambridge MA, 1986.
- Brooks, F.P. *Mythical Man-Month: Essays on Software Engineering.* Addison-Wesley, Boston MA, 1995.
- Davenport, T.H. *Working Knowledge.* Harvard Business Press, Boston MA, 1998.
- DeSoto, H. *Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else.* Random House, New York, 2000.
- Evans, G.E. *Developing Library and Information Center Collections.* 4^{th} ed. Libraries Unlimited, Westport CT, 2000.
- Hoffer, J.A., George, J.F., and Valacich, J.S. *Modern Systems Analysis and Design.* 3^{rd} ed., Pearson Education, Upper Saddle River NJ, 2004
- Kahaner, L. *Competitive Intelligence.* Touchstone, Washington DC, 1996.
- Kelly, S. *The Book of Lost Books: An Incomplete History of All the Great Books You'll Never Read.* Random House, New York, 2006.
- Lerer, F.A. *The Story of Libraries: From the Invention of Writing to the Computer Age.* Continuum Publishing, London, 1998.
- Newman, M.E.J. *Networks: An Introduction* Oxford University Press, Oxford UK.
- Sussman, G.J., and Abelson, H. *Programming Languages.* 2^{nd} ed. MIT Press, Cambridge MA. 1996.
- Weber, R.A. *Information Systems Control and Audit.* Prentice Hall, Upper Saddle River NJ, 1998.