CHAPTER 4

ARCHITECTURE AND SYSTEMS

Tryg Ager

INTRODUCTION

This chapter summarizes, compares, and evaluates the digital library architectures, systems and software components the WTEC team learned about in Japan. The baseline for comparison is a reference architecture derived from a set of core capabilities of digital libraries. The team was not charged to investigate digital library implementations in detail, so both architecture and systems will be described from a general or reference model standpoint. The team did find, however, that the concept of digital library is frequently explicit in Japan, both in the plans of technology providers, where numerous point solutions, components, and products were seen, and in operational information management approaches in businesses, universities, and museums. The WTEC teams found many different architectural approaches and system designs. In this chapter, a few examples are discussed in order to develop the general findings. Not all sites the WTEC panel visited are represented, but most of the main themes observed on the trip are covered. The sites that will be covered are the following:

• Nara Institute of Science and Technology (NAIST)
• Nihon Nikkei
• National Museum of Ethnology
• IBM Global Digital Library
• Fujitsu Multimedia Library Vision
• Japan's National Digital Library II Project

PRINCIPLES OF DIGITAL LIBRARIES

The purpose of a digital library is to provide coherent organization and convenient access to typically large amounts of digital information. The following principles provide working definitions of a digital library from both a conceptual and a practical standpoint:

• A digital library is an integrated set of services for capturing, cataloging, storing, searching, protecting, and retrieving information.
• Digital library services bring order where data floods and information mismanagement have caused much critical information to be incoherent, unavailable, or lost.
• Digital library architecture emphasizes organization, acquisition, preservation, and utilization of information.
• Digital library systems are realizations of an architecture in a specific hardware, networking, and software situation.

**Core Capabilities of Digital Library Systems**

Digital library systems compose a family of automated systems that together provide a comprehensive capability to manage the digital content of an enterprise. It is useful to divide the capabilities of digital library systems into the following areas:

• capture or creation of content
• indexing and cataloging (metadata)
• storage
• search and query
• asset and property rights protection
• retrieval and distribution

Content exists in multiple sizes, formats, and media, each with accompanying technical challenges. Content may be structured or unstructured. It may have exact, precise meaning; or it may be fundamentally ambiguous. Content may directly or indirectly support a business process or function.

A digital library architecture shows how capabilities are realized and related, and does this at several levels. Digital library architectures show how business processes or functions are enhanced; they show how technology components fit together and how, in detail, components interoperate with each other.

Such functions and relationships, when reduced to a particular software and hardware implementation, lead to operational digital library systems.

**Digital Libraries and Traditional Libraries**

Digital library functions, insofar as they purport to organize information, may be compared with traditional library functions. Consider digitization, which technically is the conversion of analog to digital formats. A common human artifact, such as a bound book, loses value when simply scanned into bits. In a library context, where organization, access, protection, and preservation are important business functions, digitization technologies are starting points for a complicated set of computational processes that in the first instance reconstruct the cultural, conventional, and intuitive significance, structure, and external relationships that defined the original artifact. Additionally, digitization and other processes may be able to add value and support certain fiduciary responsibilities that resemble functions of traditional libraries.

In a similar way, other core capabilities of traditional libraries can be transposed to the digital domain. Cataloging is transposed to the generation of metadata, and is an area where much work needs to be done to develop automated, multidimensional indexing and cataloging procedures. Just as the public card catalog is a gateway to the holdings of a conventional library, search of content and metadata is the gateway to a digital library. Circulation in a conventional library transposes to network access, retrieval and delivery.

The fiduciary responsibilities of traditional libraries are related to issues of copyright protection and intellectual property rights. Table 4.1 relates digital library capabilities to well-known capabilities of traditional libraries. The point is that traditional libraries have established uniform business processes and highly interoperable data formats which support especially bibliographic catalogs, item ordering, and interlibrary loan. Although many of these procedures pre-date "digital" libraries, digital library design can benefit from the comparisons.
Table 4.1
Comparison of Digital and Traditional Library Capabilities

<table>
<thead>
<tr>
<th>Digital Library Capability</th>
<th>Traditional Library Capability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capture</td>
<td>Acquisitions and collection development</td>
</tr>
<tr>
<td>Catalog and Index</td>
<td>Cataloging rules and bibliographic control</td>
</tr>
<tr>
<td>Store</td>
<td>Stacks, inventory management and shelf lists</td>
</tr>
<tr>
<td>Search</td>
<td>Public card catalog</td>
</tr>
<tr>
<td>Protect</td>
<td>Patron privileges and circulation rules consistent with public law and policy</td>
</tr>
<tr>
<td>Retrieve</td>
<td>Loan management and interlibrary loans</td>
</tr>
</tbody>
</table>

Having made these comparisons, it must be emphasized that neither in the United States nor in Japan is the digital library regarded as a technology related to library automation or the provision of integrated library systems for operating traditional libraries. None of the digital library projects visited in Japan either utilized or were based on library automation technologies. On the contrary, the panel saw digital library technologies, which very strikingly enabled the digital library capabilities enumerated in this chapter, creating new lines of business in both public and private endeavors.

**DIGITAL LIBRARY ARCHITECTURE**

In the following paragraphs an architectural approach to the digital library will be developed, which is based on taking the fundamental capabilities, introduced above, as the fundamental requirements the architecture must satisfy. I begin with the following notional architecture for digital libraries:

- A digital library approach to information management depends fundamentally upon a distinction between data and metadata. Metadata provide external classifying and organizing relations for data that may be unstructured, complex, or very large.

- Middleware services such as search, asset protection, and retrieval processes depend on metadata. Since metadata refers to data, which may be stored in separate hierarchical storage subsystems, integrity of reference must be maintained between metadata and data.

Figure 4.1 illustrates these fundamental distinctions between data management and metadata management, and between ingestion and utilization of content.

**Operational Architecture**

Operational architecture is an information management system represented in terms of the business processes it supports, and how information related to conduct of the business processes passes through the system’s components.

The example shown in Figure 4.2 is an enterprise that conducts training by utilizing an extensive computer-based simulation system. The operational (business) processes, most obvious in the example, depend on the timely and well-organized capture of training information as it happens, and both contemporaneous and retrospective search and retrieval of information from a training event. Although the information is generated in several different enterprise domains (eight in the example), effective utilization of information often depends on cross-domain searches and retrievals. Therefore, digital library services must provide information interoperability in middleware.
Fig. 4.1. Notational architecture—building blocks of information to enhance existing functions and enable new operational capabilities.

Fig. 4.2. Operational architecture.

Technical Architecture

A technical architecture breaks down operational (business) processes into functional components and capabilities (Figure 4.3). Hardware and software implementations are still not resolved.

The utilization of digital library materials depends on the existence of metadata to give an efficient and accurate view of content. Metadata must be created as content is added to the digital library. Metadata and data must be bound together logically, and there must be a robust underlying technology to manage the
logical connection through time, across platforms, and over geographical separations, all on a networked, distributed system.

Systems Architecture

A systems architecture shows the technology enablers and their inter-relationships. In Figure 4.4, the digital library is a centralized subsystem that interacts with a variety of data producers and consumers within a complex distributed system.

A fully detailed systems architecture resolves into software and hardware systems. Desirable systems properties such as scalability and extensibility can be taken into account at the systems architecture level. The systems architecture is rationalized relative to the operational and technical architectures.

THE DIGITAL LIBRARY IN JAPAN

This section will show how specific examples of digital libraries in Japan align with the design principles and the reference model sketched. It will try to represent each case on its own terms, leaving general comparisons and evaluations until the end of the chapter.
This section begins with the digital library at the Nara Institute of Science and Technology (NAIST). NAIST’s digital library is functional today and exhibits all the characteristics of the generic digital library reference architecture. The presence and management of both data and metadata is especially apparent in the NAIST technical and operational architectures (see Figure 4.5). The reader is directed to the site visit reports (Appendix C) for descriptive information about NAIST.

**Capture and Catalog**

The acquisition or capture process at NAIST is based on local conversion of print materials to digital form. Semi-automated means are used to prepare journals for scanning. For journals, article-level bibliographic data are produced; the articles are scanned and passed through optical character recognition systems for both English and Japanese texts. NAIST uses external cataloging services from NACSIS, a national service, which provides bibliographic records for many Japanese libraries.

**Store**

NAIST conceptualizes storage as a multimedia database, but clearly separates the “primary content” from metadata, which, as the NAIST concept of operations diagram (Fig. 4.5) indicates, are a mixture of bibliographic, presentation, inventory, and navigation data.

**Protect**

NAIST does not use technologies beyond user login to protect content. NAIST relies on publisher agreements to manage copyright issues.
Search and Distribution

The NAIST interface for query and retrieval is mainly Web-based, but there are specialized video presentation and editing workstations.

NAIST’s digital library mission subsumes traditional research library missions, and practically all library capabilities are partially or fully realized with digital technologies. NAIST includes all core capabilities of a digital library and was the most complete example of a digital library that we saw.

From an architectural standpoint, NAIST’s digital library aligns with the reference model, with emphasis on multimedia as the content rather than an abstract content type such as “library holdings” or “library objects.” Implementation is largely through integration of off-the-shelf components running on commercially available systems, servers, and networks.

NAIST’s systems include five main subsystems and interconnections to campus and external networks (Figure 4.6):

- The input system includes capabilities to digitize materials, especially journals and technical literature. There are color and monochrome scanners, digital cameras, and microform conversion equipment, and capability to convert already digitized materials for storage and retrieval in the NAIST digital library.
- A video subsystem supports special conversion and delivery requirements for quality of service for video ingestion and playback. Commercial (SGI Challenge XL) equipment is used.
- The main storage subsystem is controlled by an SGI Challenge XL. A hierarchical storage system is used with hard disk, optical, and tape capacity.
The retrieval subsystem includes World Wide Web serving, and some special client-server capabilities that utilize SGI Indy workstations.

A business support subsystem manages the administrative computing for the digital library.

NAIST adopts mainline technologies and builds most digital library services upon a readily available commercial base. The panel did not find, however, that the NAIST digital library, or any other Japanese example was either built upon or extended the capabilities of conventional, commercial automated library systems. NAIST’s operational concept is very advanced. Remarkably, NAIST’s system implementation is very conventional.

Nikkei (Nihon Keizai Shimbun)

Nikkei demonstrates how digital library technologies provide support for new and expanded lines of business where information management is central to core business functions. Nikkei’s systems illustrate the following principles:

- Nikkei has designed and built a comprehensive system for organizing digital information.
- The architecture and system design are driven by business processes.
- Well-managed digital information permits creation of new lines of business
- The recent history of Nikkei online services illustrates commercially viable scaling and market segmentation

The reader is directed to the site visit report (Appendix C) for further details about the scope of Nikkei’s newspaper and information businesses.

The Nikkei concept of operations is shown in Figure 4.7, taken from a 1997 corporate overview. News gathering corresponds to the acquisition function of a digital library, but in the Nikkei situation, one sees that capture technologies, such as scanning, are secondary. Instead, the emphasis is on information gathering by
a worldwide staff of reporters. The Nikkei automated systems are sharply focused on converting
information into information-intensive products and services that are distributed via print, broadcast, and
online media.

Fig. 4.7. Nikkei operational concept.

The operational concept diagram (Figure 4.7) shows Nikkei’s fundamental information acquisition resource:
a worldwide team of reporters and news-gathering offices (the blocks at the top).

The ovals represent automated information management systems that are very good examples of the
application of digital library approaches to a commercial information service. The systems shown are an
editorial system, a publishing system, and three network-based businesses that distribute various blends of
information, most of which has passed through the editing and production processes.

Nikkei has mastered the problem of building a system for management of digital information that can very
easily adapt to new technologies (e.g., Internet). This achievement appears to derive from an operational
architecture that is explicitly designed to re-purpose and leverage information that is derived from or
complementary to the newspaper production system. But in leveraging its core capabilities, Nikkei is
pursuing new lines of business, such as being the Japanese supplier of AOL.

PLES

PLES is the PaperLess Editing System that prepares material for publication (Figure 4.8). PLES processes
the information produced by the worldwide staff of 1,400 reporters, data gathered from wire services, and
other internal archival and current information resources. Interestingly, it uses a text-to-speech system for
copyediting. It is believed that multi-modal editing (both listening and reading) is more accurate than simply
reading copy. PLES also includes a complete graphics input system, including scanners and digital format
conversions. PLES corresponds to the capture and catalog capabilities in the digital library reference
architectures discussed at the beginning of this chapter. The PLES subsystem provides inputs to the
computer-based newspaper production subsystem, ANNECS, shown in Figure 4.9.
ANNECS

ANNECS is the computer-based publishing system. Not only does it perform typesetting and layout, but also routes its data to other Nikkei businesses (Figure 4.9). Nikkei’s approach to leveraging and reuse of information works because they have digital information which can inter-operate with a variety of systems, and which can be effectively reused in other lines of business besides publishing. One of the features of the digital library reference models discussed at the beginning of this chapter is data interoperability. Nikkei’s approach depends on data interoperability. Accordingly, data that support newspaper production are passed along to subsystems that support online services and broadcast media.
Figure 4.10 shows how capture and production facilities for news gathering and newspaper publishing pay off for Nikkei by supporting additional lines of business.

Fig. 4.10. Wire services and databases (NEWS, NETS and NEEDS).

*NEWS* is a distribution system (Nikkei Economic Data Wire Service) that feeds broadcast and online services.

*NETS* is a system to convert information originally in Japanese into English for resale or inclusion in Nikkei English-language products.

*NEEDS* is a database service and text search and retrieval system.

All of the above services that organize and manage digital information feed additional products and lines of business based on digital content, as shown in Figure 4.11.

Fig. 4.11. New business based on organized digital information.
**QUICK** is a customizable, personalizable online product that delivers high-end business information to select customers.

**NIKKEI NET** is an Internet, Web-based news service that charges users.

**AOL** services in Japan are provided by Nikkei. Nikkei information in Japanese is a value-added product for Japanese AOL customers.

**NSN** is an all-business television channel that is broadcast using digital satellite technologies.

**Nikkei Telecom** is another Internet service that features a hyper-linked online newspaper format that offers search and retrieval for specialized business information such as corporate strategies and management news items.

Nikkei illustrates very clearly how interoperability in data and middleware can leverage information assets into many lines of business with different market targets, different selections of information, and different application-level interfaces and capabilities. The architectural and systems approaches seen in Nikkei information systems are the clearest and most advanced examples of digital library approaches to the organization of information for commercial purposes that the panel saw in Japan.

**National Museum of Ethnology**

The National Museum of Ethnology is a leader in utilizing technology for many aspects of museum operations, which are detailed in the site report (Appendix C). The museum is a good example of systems and architecture because all technologies at the museum are specifically designed and implemented to automate or enhance internal museum procedures.

The museum utilizes 3D imaging and measurement technologies to partially automate the acquisition process. Figure 4.12 shows one of the scanners and also gives a sense of the exhibition space at the museum. Digital library technologies are extending the capabilities of the museum, and over more than two decades, an elaborate local system to support museum functions has evolved, which features the following:

- digital library technology used to manage holdings of a museum
- support for:
  - online asset management
  - multimedia support for exhibition
  - indexing, cataloging, and search of scholarship
  - virtual collections for Internet communities
- example of a local museum solution

The museum’s current systems design places technologies for individual museum functional capabilities such as video exhibits, kiosks, scholarship, asset management, and Internet on a high performance local network (Figure 4.13). This system is not designed as a single unified or comprehensive museum system. In that sense the design is conservative. However, the scale of the system relative to the museum’s overall mission is very impressive, because nearly everything at the museum is strongly supported by technologies that map clearly to the digital library reference models.
IBM Tokyo Research Global Digital Museum

One of the visions of digital library is global virtual collections. Working with the National Ethnographic Museum and the British Museum, IBM’s Tokyo Research Laboratory has designed and implemented a global virtual museum focused on problems of K-12 museum education. The virtual museum includes the following characteristics:
features architecture and system for worldwide virtual museum
creates shared abstractions of collections that are managed differently
focuses on K-12 museum education as a line of business

One of the problems a virtual collection must address is a data architecture that makes the different legacy systems of the various museums interoperable in the virtual collection space. The Global Museum Project defines data abstractions and user-level operations that allow teachers to create virtual collections for instructional purposes, and students to annotate, select, and present their own personal collections (Figure 4.14).

Fig. 4.14. IBM Tokyo Global Museum: operational concept.

Fujitsu Multimedia Library

Japan’s technology providers are very focused on multimedia systems. Fujitsu’s vision represents some of the architectural and systems issues that must be faced in order to build full-service, scaleable multimedia digital libraries, as indicated below:

- vision of modern system for scaleable multimedia management
- representative of technology companies seeking to provide core technologies at middleware level

Figure 4.15 summarizes challenges for digital library technology providers. From the top down, heterogeneous media requires changes in search and storage subsystems. Improvements in database technology for managing metadata must be complemented by advances in multimedia object stores.

Multimedia distribution raises quality of service issues and requires resource management at the systems level.

Finally, many device improvements will drive the multimedia library, especially devices that extend information management to new areas of internetworking, consumer electronics, home devices, and collaborative workspaces.

Japan’s Second Generation Digital Library Project

In Japan, a second national digital library project is underway, funded by the Ministry of International Trade and Industry (MITI), and conducted by the Information Technology Promotion Agency (IPA) and Japan Information Processing Development Center (JIPDEC).
The purpose is to develop a reference architecture that will drive development and utilization of advanced technologies for information management. Only an overview of a preliminary version of the next generation architecture is considered here. The following bullets highlight main themes of the project:

- operate a private and public consortium
- build scaleable, distributed multimedia information management systems
- use advanced, standards-based technologies
- proceed as a consensus project (U.S. DLI-2 is competitive)

The reference model reflects modern multi-tier distributed systems architectures (Figure 4.16). It features messaging middleware, agent technology, multimedia databases, mobile agents, and CORBA distributed object management. The project is practical, and the plan calls for a prototype system in the next two years.

Up-to-date information may be found at the Next Generation Digital Library Web site, http://www.dlib.jipdec.or.jp.

Figure 4.17 is representative of the project’s approach, indicating utilization of the three-tier model, CORBA, and Internet standards.

FINDINGS

The following represent key elements of digital library systems in Japan:

- Advanced, mission-specific, and compelling digital library systems are operational in some Japanese public institutions and commercial enterprises.
- Technology companies are exploring advanced digital library components and computing paradigms. Commercial emphasis is on custom integrations.
- Replicable solutions, services, and outsourcing were not observed except as envisioned by the Second Generation Digital Library project.
Fig. 4.16. Reference model for basic architecture (Japanese DLII initiative).

Fig. 4.17. Structure of messaging platform (Japanese DLII project).
Japanese businesses and public institutions are engaged in a second-generation digital library project with a comprehensive architecture.

Most libraries and museums do demonstration projects, while some are aggressive “early adopters” of digital library approaches.

The following points summarize the findings of the trip regarding systems and architecture:

- Advanced, mission-specific digital library systems are driving certain public institutions and commercial enterprises forward. These systems are usually part of a business process reengineering effort caused by rapid changes in the nature of information creation, consumption, and distribution technologies that affect core business functions and create new business opportunities. In these cases, architecture and systems are driven by business processes.

- Individually, technology companies are doing exploratory work on components for integration into current systems, but also looking forward to new computing paradigms that imply fundamental changes in systems, particularly with respect to information access and distribution. In these cases, it is clear that the interest is in providing advanced core technologies that support many information technology market segments.

- Cooperatively, Japanese business and institutions are engaged in a second-generation digital library project, which does provide a comprehensive architectural and systems point of reference. This is CORBA-based, takes advantage of advances in networking, features metadata management and many flavors of search and query.

- Libraries and museums, whose approach to information management in many ways inspires the technologies we saw, are split between those that are doing interesting, but limited demonstration or exploratory projects, and those that are aggressive “early adopters” with comprehensive visions and commitments to change the ways information is acquired, managed, and utilized in their environments. The library and museum sector featured many digitization projects, and offered numerous insights into the capture and cataloging subcomponents of an overall digital library system.

**COMPARISONS**

A comparison of digital library systems and capabilities in Japan and the United States suggests the following:

- In the United States and Japan, digital libraries are custom integrations with emphasis on a common set of core capabilities.

- In the United States and Japan, compelling examples of digital library systems and architecture solve specific business problems.

- Scaling of digital library capacity depends on internationally shared data, metadata, and distributed systems standards.

- For the next decade, the digital libraries will be based on common core technologies designed, implemented, and integrated to support certain business processes.

- Japan and the United States are comparable regarding architectures and systems as such.

- Japanese technology companies stress mainline multimedia middleware, while U.S. startup companies are more innovative.