Kansas State University Digital Library Program

Recommendations to:
the Steering Committee

Provost Coffman, Vice-President Rawson, Vice-Provost Unger and Dean Hobrock

May 30, 2001

Digital Library Task Force

David Allen             Donna Schenck-Hamlin
Karen Cole              Charlene Simser
David Hillier           Roger Terry
Hani Melham             Harvard Townsend
Sonny Ramaswamy         Mike Wonderlich
Table of Contents

Executive Summary

Introduction
  Understanding the Need/Defining the Program
  Defining Concepts

Goals/Objectives

Implementing a Digital Library Prototype:
  Conceptualization
  Functional Requirements
  Organizing Collections
  Roles and Responsibilities

Recommendations
Executive Summary

In March, 2000, a task force was formed to explore the particular dimensions of a digital library at KSU and develop a working definition of such a program. This task force investigated issues central to information access and content development at KSU and identified key “defining concepts” for a digital library initiative. These defining concepts included –

- **Collaboration.** Digital collections will be the product of collaborative efforts by virtual development teams of researchers and content creators in distributed locations.
- **Content.** Collections will be diverse in content, format, location and authorship. Digital library content issues will include selection and funding policies, quality control and management of access rights.
- **Access.** While enhancing access to established print collections, the digital library must provide the networking capacity required in the knowledge creation activities of KSU’s research and learning communities. Access capabilities should include search engine software to simultaneously search multiple repositories through a user-customized portal.
- **Durability.** For digital content, achieving persistent access over time will require accommodation of both changing content formats and evolving information technology. Digital library developers must assist in establishing industry best practices in this area.
- **Management Focus.** The digital library service environment must support the full range functions required by a library to manage, administer, monitor engagement with, preserve and ensure fair use of its collections.

In August, 2000, after receiving its final report, Dean Brice Hobrock and Vice Provost Beth Unger expanded the task force membership into a “Core Task Force to Develop the Kansas State University Digital Library.” This enlarged task force was charged to develop a working prototype by June, 2001, and provide recommendations for full implementation of a digital library at KSU.

The task force determined that the manner in which information is currently collected – distributed across numerous independent “silos” or repositories – is the university’s fundamental information access challenge. Consequently, the capacity to integrate or bridge across these silos was a high priority as the task force searched for existing or emerging software to support a digital library. Other digital library system requirements included:

- a unified user interface,
- accommodation of emerging standards and protocols for displaying, describing and accessing disparate data in multiple formats,
- capacity to organize information using metadata,
- facilitation of distributed or centralized production input,
- user authentication and rights management, and
- provision of a portal type presence for all KSU information.

A commercial product was sought that could deliver this holistic vision of a digital
library. After several product reviews and vendor demonstrations it was clear that no single commercial system in the current market meets all of these requirements. However, Endeavor Information Systems’ ENCompass product addressed the essential metadata organizer and search engine needs. Further, Endeavor agreed to a development partner relationship with KSU. Since January, 2001, Endeavor has worked very closely with the task force in building the digital library prototype.

Because the prototype is intended to model and test recommendations for a full production system, the “collections” to be included reflect a range of content, format, data conversion and workflow alternatives. Among the prototype collections are Extension publications housed on hardware in College of Agriculture; translated research documents from the China Agricultural University hosted on a KSU server; an index of the Collegian and K-STATER converted from an ACCESS database; a sampling of Landon Lecture video files; an Information Support Services for Agriculture database; KSU Libraries’ online catalog and some commercially licensed content.

The development process also demonstrates the use of “content development teams,” comprised of content owners/creators (faculty and librarians), metadata support staff (catalogers), technical support staff (CNS and ISO) and applications support staff.

The task force recommends that the university proceed toward full implementation of a KSU Digital Library along the lines of the prototype. Central among its recommendations is the creation of a digital library management department that can function as a steering group with both policy and implementation responsibilities. Other recommendations include –

• identify software for a university portal that can be integrated with ENCompass system software;
• implement university-wide authentication/authorization/rights management;
• centrally managed hardware for housing content repositories.
Introduction

The term “digital library” is commonly used to describe initiatives ranging from the digitization of a specific collection of materials to a formal digital library organization representing the full set of administrative, policy and management functions traditionally associated with a library. The Digital Library Federation, a consortium involving the nation’s leading academic and research libraries, seeks to accommodate this wide range of contexts with the following comprehensive definition:

> Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities. (http://www.clir.org/pubs/annual/anrpt97/libraries.html)
> – Digital Library Federation (DLF)

Understanding the Need and Defining the Program

To explore the particular dimensions of a digital library at KSU and develop a working definition of such a program, a task force was formed in March, 2000. This task force investigated issues central to information access and content development at KSU and identified key “defining concepts” for a digital library initiative.

Fundamental among KSU’s information access issues is the manner in which that information is currently collected, i.e., distributed across numerous independent “silos” or repositories. These repositories are diverse in their formats, search interfaces and target audiences. Tools for accessing and managing digital collections are specialized for a single type of content and not integrated. Consequently, research in this digital environment frequently requires that searchers know that an item exists, know where it exists and then know the navigational nuances of a specific access system to complete their task.

Defining Concepts

In seeking to understand the relevant components of a digital library initiative, the task force recognized the following to be key “defining concepts.”

**Collaboration.** It is clear that the scope of a digital library program must extend beyond the existing organization and resources of KSU Libraries. Digital collections will be the product of collaborative efforts by virtual development teams of researchers and content creators in distributed locations. Building a digital library infrastructure will require the integration of IT resources and talent as well as information management expertise. Hence a digital library service environment must promote and support collaboration.

**Content.** Collection content will be diverse in all aspects. Collections will include commercially acquired content as well as locally created/published resources. Collections will be virtual with some resources residing locally and some distributed among geographically dispersed owning repositories. Multiple content formats will
be represented as well as multiple metadata types. Content selection issues will be among the most challenging issues to be addressed in the digital library service environment. These include establishing policies and processes for content selection, collection development funding, quality control and management of access rights.

**Access.** The KSU digital library must respond to the university’s information needs in a time of transition, when both established systems and emerging electronic enterprises must be supported. While enhancing access to long-established print collections, it must surpass the limitations of the print world and enable the knowledge creation activities of KSU’s research and learning communities through the power of electronic communications and networking. The digital library access environment should be characterized by:

- search engine capacity to discover and “mine for information in multiple repositories and varying “depths;”
- accommodation of emerging standards and protocols for displaying, describing and accessing disparate data in multiple formats,
- capacity to organize information using metadata,
- integration of instructional, research and administrative information;
- user authentication and rights management, and
- provision of a portal type presence for all KSU information with user defined access tools.

**Durability.** For digital content, the task of preservation must address issues of both storage and access. Achieving persistent access over time will require accommodation of changing content formats and information technology. Insuring adequate migration of content as technologies evolve must be at the center of any decisions concerning selection of storage media and software. Similarly, consideration for long-term access should influence decisions about adopting/supporting the diverse array of access tools potentially available in a media-rich digital library environment. Choice and flexibility should be balanced with sustainability. Digital library developers must promote and assist in establishing industry best practices in this area.

**Management Focus.** Persistent access, on a broader level, depends on the digital library’s long term organizational and budgetary viability. If KSU is to do more than continue its past pattern of independent digital collection development, a management structure will be required to establish and maintain a digital library service environment. The Digital Library Federation describes this essential program aspect in the following statement: “The digital library service environment is not simply about access to and use of information. It also supports the full range of administrative, business and curatorial functions required by a library to manage, administer, monitor engagement with, and ensure fair use of its collections . . . whether located locally or off site.”

**Goals and Objectives**
Based on its understanding of the preceding concepts and its evaluation of KSU information issues, the task force proposed the following goals.

**Goal 1: Establish an oversight and management structure based on broad community participation to guide the funding, implementation and ongoing support of this initiative.**

**Objectives:**
1. Seek support of the digital library concept document by key representatives of the KSU Academic Community including:
   - President, Vice-President of Administration and Finance, Provost, Vice-Provosts;
   - Selected committees from key campus entities (such as IT departments, KSU Libraries, Extension, Continuing Education, academic colleges and departments);
   - Deans and Academic Department Heads.
2. Establish a Steering Committee consisting of division heads from key campus entities and a development Working Group for 2000/2001. Dedicate a minimum of 8 FTE to the project Working Group.

**Goal 2: Develop a working prototype digital library by June 2001.**

**Objectives:**
1. Include in the prototype selected resources which reflect a range of content, format, data conversion and workflow alternatives, such as:
   - Digitized full-text Extension and Research Experiment Station publications,
   - Online catalog from KSU Libraries,
   - Commercially produced databases and/or full-text sources,
   - Image collections (e.g. Kansas wildflower, Architecture slides, etc.),
   - Indexes to *Manhattan Mercury* and *Collegian*, full-text of the latter,
   - Citation databases of Information Support Services for Agriculture,
   - Kansas topographic maps.
2. Procure or develop digital library system software with the capacity to:
   - integrate independent information repositories;
   - accommodate emerging standards/protocols for displaying, describing and accessing data in multiple formats;
   - organize information using metadata;
   - enable user authentication and rights management, and
   - provide a portal type presence for all KSU information.
3. Demonstrate within the development process the use of “content development teams,” comprised of content owners/creators (faculty and librarians), metadata support staff (catalogers), technical support staff (CNS and ISO) and applications support staff.

**Goal 3: Invest in software, hardware and training for key personnel involved in FY2001 Prototype.**
Objectives:
1. Investigate software such as Endeavor Corporation’s *Image Server* and *Citation Server* to determine their potential in conjunction with KSU Libraries’ ILS (integrated library system) for managing and developing specific digital library functions.

2. Fund Endeavor’s *Voyager* certification training for technical support staff from CNS, ISO, and KSU Libraries to develop common operational basis for future systems planning.

3. Procure server/file storage hardware for digital content files to be developed.

4. Investigate applications/search software to implement/enhance digital library content and access.

Goal 4: Upon successful demonstration of the Prototype, establish an independent organizational entity for the KSU digital library, including a director and associated staff.

Objectives:
1. Create an organizational unit that will provide a management focus and technical infrastructure for digital library content and access. Responsibilities will include:
   • central coordinating and management services for content development;
   • promoting integration of digital collections on campus and interoperability with other digital libraries;
   • providing and promoting use of centralized server and storage services;
   • providing and promoting centrally supported development tools and services;
   • encouraging content development partnerships and consortial activities.

2. Based on the roles identified and developed in the prototype phase, design the organizational structure to accommodate development and management staff with both short-term (content sponsorship) and continuing (database administration, metadata management, applications support) responsibilities.

3. Establish appropriate policy making and administrative structures governing participation, prioritization, funding, development, access, intellectual property rights and privacy.
Implementing a Digital Library Prototype

In August, 2000, after receiving its final report, Dean Brice Hobrock and Vice Provost Beth Unger expanded the initial task force membership into a "Core Task Force to Develop the Kansas State University Digital Library." This enlarged task force was charged to develop a working prototype by June, 2001, and provide recommendations for full implementation of a digital library at KSU. The Task Force was asked to report to the Steering Committee for the joint Libraries, Computing and Network Service and Information Systems Office project (Provost Coffman, Vice-President Rawson, Dean Unger, and Dean Hobrock).

Conceptualization

During Fall 2000 the task force investigated model digital library initiatives undertaken by other universities and research institutions. Many of these efforts were narrowly focused, consisting of scanning projects for special collections or organizing commercially licensed products. Very few digital library initiatives involved integrating these types of information along with the library catalog, administrative data and digitally born intellectual content.

In planning for the KSU prototype, the task force pursued a more holistic vision, i.e., “a knowledge discovery facility, a framework for organizing, accessing and delivering knowledge resources to a widely (liberally) defined KSU academic community.” (KSU Digital Library: a Definition for Development, 3/26/2000)

Further, the task force recognized that a successful digital library will not be an isolated “project” or “entity” that exists separate from the University’s total IT environment. Rather it must be part of a whole integrated information system for the university, providing the capacity and structure to share, organize and discover here-to-fore difficult to find information. It should provide the capacity to create “new synergies” in the discovery process between and among unrelated disciplines and content. It is critical that the digital library be implemented within the context of the following simple model:
**Functional Requirements**
As the task force determined functional requirements for a system that could deliver its holistic vision of a digital library, it considered both locally developed and commercial software possibilities. Key functional requirements included:
- capability to organize all types of data
- must utilize metadata
- ability to handle all types of formats
- provide for distributed development
- integrate into or be the University portal
- utilize/integrate administrative data
- cross data/format searching
- ability to integrate commercial products
- utilize standard plug-in software for viewing, listening
- flexible and customizable
- rights management
- authentication/authorization

Represented schematically, the task force’s desired digital library system looked like the following:

Given the time constraints of the prototype implementation deadline, only commercial software applications were investigated. After several product reviews and vendor demonstrations it was clear that no single commercial system in the current market meets all of these requirements. However, Endeavor Information Systems' ENCompass product addressed the essential metadata organizer and search engine needs. Further, Endeavor agreed to a development partner relationship with KSU.

The task force also investigated and purchased Virage’s Internet Video Application Platform software which provides powerful tools for indexing video and audio content and making it accessible to researchers.

**Organizing Collections**
Between late January and mid-May, 2001, content development teams worked under
the general direction of the task force to create the following digital collections:

- **Video/audio collections**
  - Public Service Media
  - KSU Instructional Media
  - Public Lecture Series (Landon, Huck Boyd)

- **Image collections**
  - Kansas Wildflowers and Grasses by Time of Flowering
  - Elevator Explosions

- **Text/Graphics collections**
  - Information Support Services for Agriculture
    - Precision Agriculture
    - Grain Economics
    - Stored Product Protection
    - Grain and Oilseed Quality
    - China Agricultural Conference Proceedings
  - KSU Cooperative Extension Current Publications

- **Externally owned collections**
  - Ezra Cornell Papers
  - Getty Research Institute image files

Data Type Descriptions (DTDs) and associated metadata for each of these collections was entered by staff working in separate locations into a central Oracle database using ENCompass running on a test Unix server. Metadata was also created for KSU Libraries’s online catalog and some commercial content to enable users to search all of the prototype repositories with a single query. Throughout this process, KSU development staff beta tested and advised on improvements to the ENCompass software. Additionally, they received training in the use of the Virage video application software and used it to create metadata for some of the video content in the prototype.

Task force objectives in this phase of the prototype were to demonstrate –

- the capacity of metadata to manage distributed collections;
- the flexibility of DTDs in handling multiple objects/formats;
- the feasibility of central metadata support with one-time entry from disparate locations in a central database;
- the capacity of the Virage software in creating searchable digital video content;
- the potential for searchers to discover “new” information relationships through metadata.

In implementing the prototype, the task force has attempted to model fundamental processes, procedures and structures which can be used in full implementation. Further, a number of issues have been identified, which, though not insurmountable, remain unresolved at present. These issues are inherent in the task force recommendations.
Roles and Responsibilities
In bringing to fruition the "collections/data" that comprised the content of the prototype, the task force identified the need to create Content Development Teams which are responsible for each collection. Each team consists of persons who bring to the group the following skills or expertise: content sponsorship/coordination, database management, metadata management, and software applications support.

These initial prototype teams represent in some manner the roles and responsibilities that will be critical to the success of the KSU digital library. Persons involved in the full implementation will need to assume responsibilities for the needed roles/functions. The critical roles fall into the following major categories:

- Management
- Analysis & Design
- Technical Support
- User Support
- Legal and Policy

It is important to recognize that there is not necessarily a one-to-one relationship between named people and specific roles. Further, individuals from a number of "University organizations" will have to work together in cross-functional work groups and potentially assume responsibilities and working relationships that don’t exist today.

Management
Management roles include: 1) user sponsorship, 2) program coordination, 3) project/collection management, 4) content selection/publication, 5) archiving and perpetual access.

Analysis & Design
Analysis & Design roles include: 1) metadata modeling and management, 2) technical/security architecture.

Technical Support
Technical support roles include: 1) conversion/integration, 2) quality assurance/data integrity, 3) database administration (run load scripts, monitor query and database performance, backups, administer user access and security, monitor hardware capacity), 4) programming/application support (evaluate software, develop and test application integration), 5) data loading, 6) hardware support.

User Support
User support roles include: 1) consultation with data providers, i.e. faculty/students, 2) training, 3) hands-on work with data/content providers.

Legal & Policy
Legal and policy roles include: 1) rights procurement (copyright permissions), 2) intellectual property concerns/ issues, 3) licensing.
Recommendations

Based on its experience in planning and implementing the prototype, the task force offers the following recommendations. These recommendations are somewhat general in nature and are not accompanied by budget recommendations. The task force believes that if its recommendations concerning a management structure are adopted, it will be these groups working with the university administration who will develop an long-term implementation plan and budget.

1. Establish an oversight and management structure to guide funding, policy review/implementation, implementation and ongoing support of the Digital Library initiative. It would be recommended that some members of the Core Digital Task Force reporting to the CNS/ISO/KSUL Steering Committee assume that role.

2. Establish a Digital Library management department to provide a management focus and coordinate technical infrastructure. This unit has access to and works closely with cross-University work groups from CNS/iTAC/DIA/ISO/DCE. It should function as a group of “movers and shakers” to promote and expedite full development. It should have both policy and implementation responsibilities including:
   - overseeing the KSU digital library program;
   - assuming roles of Management and Legal & Policy;
   - developing the content publication process;
   - monitoring national standards;
   - coordinating “projects” and collection implementation;
   - working with Endeavor and other vendors in development of software;
   - identifying archival needs;
   - promoting integration of digital collections on campus and interoperability with other digital libraries;
   - providing and promoting use of centralized server and storage services;
   - providing and promoting centrally supported development tools and services;
   - encouraging content development partnerships and consortial activities.

3. The Digital Library department in cooperation with the broader campus community should establish appropriate policy making and administrative structures governing participation, prioritization, funding, development, access, intellectual property rights and privacy. Recommendations from these groups would be reviewed and actualized by the “oversight steering committee” in #1.

4. Develop a university portal that can be integrated with the digital library interface as well as other university applications. This need is closely associated with the need to develop a central system for user authentication, authorization and rights management. Without this piece, the integration of commercial content within the digital library, planned for Fall, 2001, cannot proceed.
5. Provide centralized IT support for priority content development projects including:
   • server facilities and storage services for object repositories;
   • database administration;
   • metadata-creation tools.

6. Initial content development projects should include:
   • completion of prototype collections (listed on page 11);
   • development of the Kansas Weather Data collection ($12,500 grant funding has been awarded);
   • continuation of projects which have existing development partnerships in place/funding in process, such as:
     ▶ Advanced Manufacturing Institute,
     ▶ Krider Center images (College of Architecture),
     ▶ Historical costume images (College of Human Ecology – grant application in process),
     ▶ Teaching videos (College of Veterinary Medicine);
   • development of enhanced access to commercial content with cross-publisher, inter-document linking.

7. External partnerships should be pursued with institutions developing digital library content and interoperability including:
   • Cornell University (access systems, content collaboration in Precision Agriculture, Veterinary Medicine teaching videos, Historical Costumes);
   • Ohio State University (content collaboration in Precision Agriculture);
   • Getty Research Institute (access systems, content collaboration in Historical Costumes).

8. Establish a user support group who works in cooperation and collaboration with the Digital Library Department providing consulting support for faculty/students/staff in the initial stages of a "digital" project and training for constituents. Further the user support group works with content providers and application developers. Legal support, i.e. copyright issues and intellectual property issues, would be integral to the user support group.
9. Specific “video” recommendations have been provided by the working group and include:
   - a central video repository. By centralizing resources, a large video repository could be sponsored by the Digital Library. Provide both Real Video and Windows Media services and for the future of MPEG streaming technology. Purchase enough hardware to store a significant video collection and provide all of the server processes necessary (Virage, Real, Windows Media, etc.).
   - research conducted into the advantages/disadvantages of placing the Virage indexing in an Oracle database
   - all future (and past) lecture series be close-captioned. This will meet ADA requirements and greatly enhance the indexing capability of Virage.
   - Hire an encoder for indexing video collections in Virage. With the volume of video production that exists and the extensive indexing features available, we believe a full-time person is justified.
   - Produce video content delivery for both high-bandwidth and low-bandwidth users.

10. Continue the investment begun with the FY2001 Prototype including: software, hardware and training for key personnel.
    - Continue to develop ENCompass software
    - Fund appropriate training for technical support staff from CNS, ISO, DCE, KSU Libraries, etc. to develop common operational basis for future systems planning.
    - Develop Virage and other software applications/search software to enhance digital library content and access.