"Preservation Planning for Digital Information"

Final Report of the HVC^2 Digital Preservation Task Force

Report Prepared by:

Richard Fyffe
Deborah Ludwig
Mary Roach
Becky Schulte
Beth Forrest Warner

University of Kansas

November 11, 2004

I. Introduction

Digital content permeates every aspect of the academic enterprise. It includes both scholarly content (research papers, syllabi, experimental datasets, etc.) and administrative content (enrollment data, financial data, personnel handbooks, etc.). Not only is this information vital to the current operation of the 21st-century university; long-term preservation of this data is required to assure administrative integrity and meet our responsibilities for stewardship of intellectual and cultural heritage.

The challenges presented by long-term access to digitally encoded information are widely acknowledged. The key to understanding them lies in recognizing the mismatch between traditional information management practices – practices largely formed in a print-on-paper environment – and the characteristics of digital information. Analog forms of representation – "print" – endure with relatively modest upkeep and can be read without intermediary technology. We can set aside a paper document and reasonably expect to find it readable years later. As a consequence, formal preservation programs for print information can generally accommodate a wide range of user practices.[1]

Digital information is not so forgiving. Digital information resides on physical media (tapes, disks) that are fragile and susceptible to corruption. More important, digital information requires intermediary technology (software and hardware) to read it – technology created in a marketplace that prizes proprietary control and innovation, resists standardization, and is often inattentive to backward compatibility. In such an environment and without a planning program for continued access to and usability of digital information, there can be no reasonable expectation that digital information created today will remain usable in a few years.

Digital preservation is the **ongoing process of managing data for continuing access and use.** Digital preservation is an outcome of the organization's successful day-to-day management of its digital information. Digital objects are never preserved once and for all like pickles in a jar. Preserved information cannot simply be encapsulated and set aside. The Digital Preservation Task Force was charged to explore the implications of a University commitment to the preservation of digital assets, both academic and administrative. With a minimum of theory, our report emphasizes actions the University should take; it is not a primer on digital preservation. **We recommend, over a three-year timeline, implementation of the following components in a university-wide digital preservation program:**

- An integrated **technical architecture** designed around the **whole lifecycle** of digital information, from creation forward (see Section IV of this report for further discussion of lifecycle and S**ection V** for our recommendations).

- Definition and assignment of a set of specific **roles** or functions exercised by staff within the University, and development of a set of **policies** to guide those roles (**Section VI**).

- **Education** for faculty, staff, and administrators in the basic concepts and challenges in digital preservation and **training** in information management practices that will contribute to the ongoing availability of digital files (**Section VII**).

[1] We acknowledge that throughout this report the preservation of tangible objects is presented in an oversimplified fashion, and recognize that effective preservation strategies for all information formats also require a life-cycle management approach and physical and organizational infrastructure.

II. Method of Work

As part of its first stage of work – October 2003 through June 2004 – the Task Force surveyed international research and best practice. Although largely un-cited in this report, this research significantly informed our work and is summarized in Appendix C, "Digital Preservation Best Practice: Working Draft Summaries." In addition, the Task Force constructed a working definition of "university digital asset" to help define the scope of its work (Appendix D), made a first attempt at a methodology for inventorying KU's digital assets (Appendix E), and took the first steps toward an outreach program to educate the KU community on digital preservation issues. Subsequent work between June and October, 2004, focused on two areas: (1) outlining an outreach program to educate the community about preservation issues and good information management practices and to train Information Services staff in preservation practices; and (2) developing a technical model for a digital preservation program with specific reference to systems under development, in planning, or not yet under development at KU, and recommending possible organizational structures and responsibilities for a preservation program.

III. Scope of Investigation: University Digital Assets

A digital asset is an electronic object that has value for some purpose. It may have been created digitally or it may have been digitized from a non-digital original source. Examples of digital assets include word processing documents, databases, websites, organizational records, digital recordings of musical performances, etc.

To become part of the University's digital preservation program, a digital asset must support (directly or indirectly) the University's fundamental instructional, research, or public service missions. A digital object should be considered a University Digital Asset if it satisfies one or more of the following criteria:

- It was created in fulfillment of the research, teaching, or creative work of University faculty, staff, or students;
- It is relevant to the planning, managing, operating, controlling, or auditing of administrative functions of an administrative or academic unit of the University;
- It was purchased or licensed by the university in fulfillment of an academic or administrative function under a contract that permits continuing use of the asset (for example, certain electronic journals licensed through the University Libraries).

Not all University digital assets will have equal priority for preservation. This working definition of "University digital asset" establishes minimum conditions for eligibility and does not determine any additional conditions that an object may need to meet to be included in the preservation program or set priorities for preservation.

IV. Establishing a Lifecycle Management Program for Digital Assets through Interaction of Technology and Organizational Culture

"Digital preservation" is the process by which access to usable digital information is managed and maintained over time. Effective digital preservation requires that we understand and attend to the full lifecycle of digital objects. This lifecycle has 2 key dimensions:

- *a chronological dimension,* which highlights the various stages of an object's life from the point of creation forward. Examples of these stages include being filed awaiting further work; being copied by a new user into a new version; being consulted by a user authorized to read it but not change it; being copied by a systems administrator as part of a nightly backup; being copied into slow-access storage; being transformed so as to be usable with new versions of software.

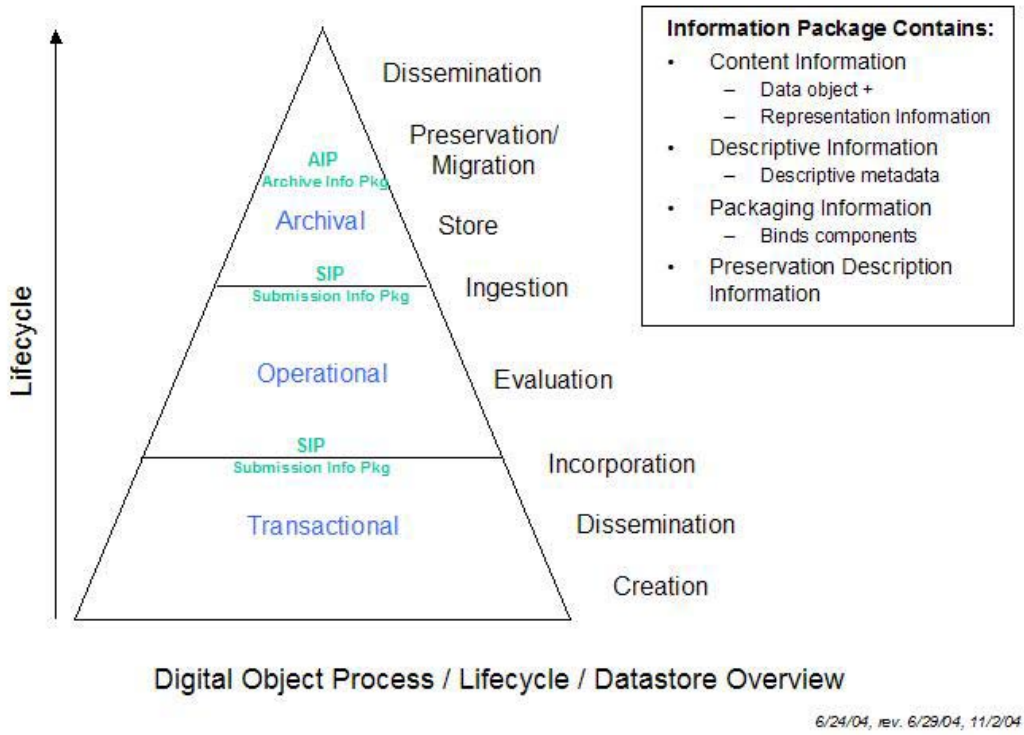

Figure 1: The Chronological Dimension of Information Lifecycle

- *a functional dimension,* which highlights the different *roles* that are occupied through an object's interaction with creators, editors, users, stewards, policymakers, and others over the course of its lifetime. See Appendix F for further discussion.

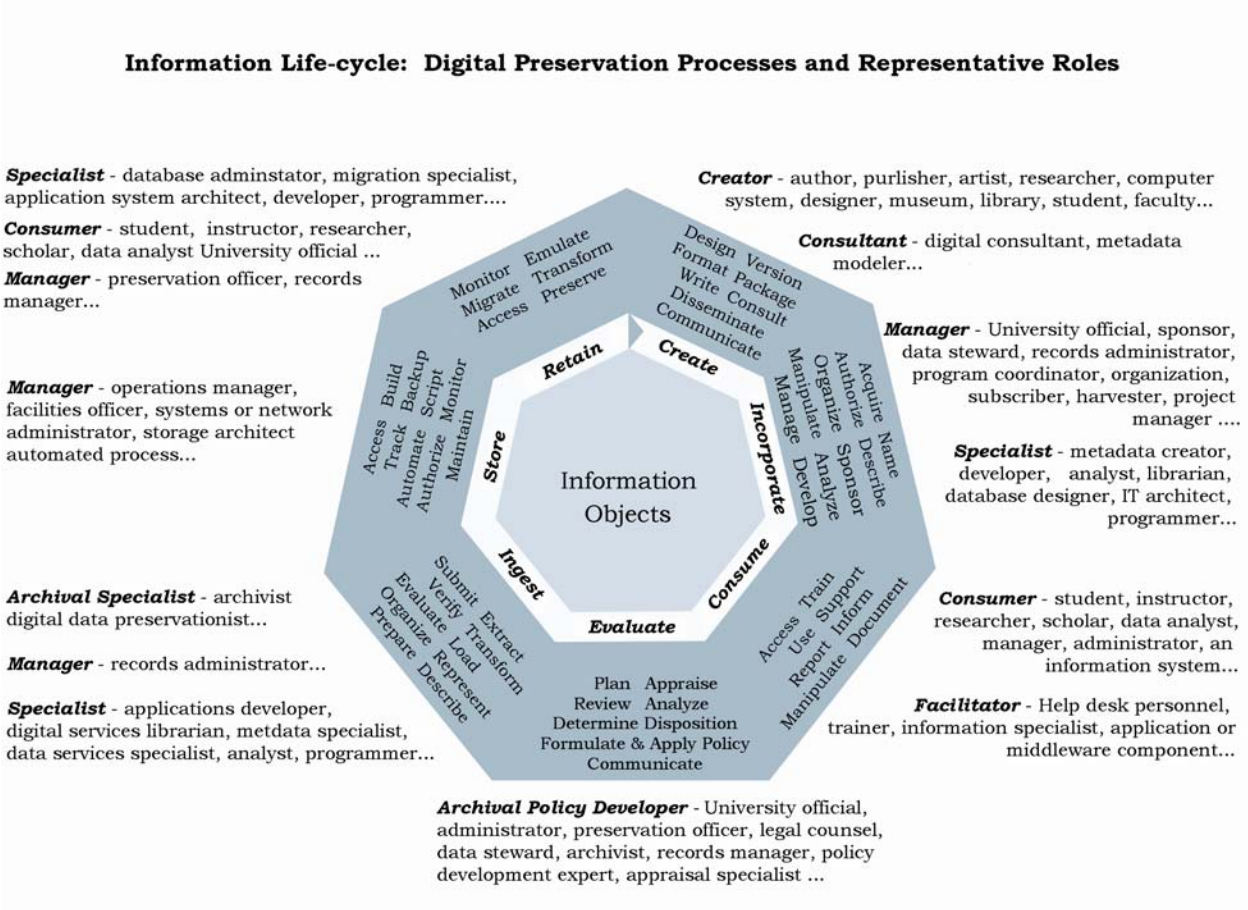

Figure 2: The Functional Dimension of the Information Lifecycle

Special Focus on the Upstream Side of the Cycle

No single step in an object's lifecycle is "the preservation step" and no single role is the "preserving role." Rather, digital preservation is the whole complex of roles and operations designed around management of information for long-term accessibility and usability. Nevertheless, we want to give particular emphasis to the impact that creators' choices (and the choices of other agents close to the creator) in the early stages of an object's lifecycle may have on the likelihood that it will remain accessible and usable over time. There are four key areas in which those "upstream" choices can be exercised:

- **Appraisal:** Early recognition that the object will (or will not) be used in the future will encourage creators and stewards to make preservation-oriented choices.

- **Choice of file format:** Some file formats are more likely to remain usable over long periods than others. In general, proprietary formats – formats that are owned by companies – are less likely to be usable in the future than "open" formats.

- **Choice of storage location:** In general, files stored on networked storage media maintained by professional technology administrators adhering to best practices are more likely to remain usable than files held on isolated disks (including workstation hard drives) if other preservation steps are followed, as well.

- **Description (Metadata):** Good documentation of the content of an asset and the circumstances under which it was created (name of creator, date, data source, file format, etc.) will enhance the likelihood that the asset can be discovered, rendered, and used effectively over time.

Additional information regarding these issues will be found in Appendix G. Sections V (Technical Infrastructure) and VI.b (Policy) provide more specific recommendations for developing the mechanisms that will help to integrate these choices into the everyday work of content creators at the University.

Digital Preservation: Three Converging Paths

In implementing a digital preservation program, three equally important tracks must be followed:

- A technical **infrastructure** must be created that includes all the systems and services necessary to manage and access information over time;

- A set of functional **roles** and institutional **policies** must be defined and assigned within the organization, to insure that these systems and services are implemented and maintained;

- An **educational program** must be created to encourage good stewardship of institutional digital assets by creators and users of information objects, and to encourage support for digital preservation by administrators and resource-allocators.

Interaction of all three of these elements creates a digital preservation program. Our recommendations for developing each aspect of the program follow.

V. Infrastructure: Recommended Technical Systems and Services

Just as no single step in an object's lifecycle is "the preservation step" and no single role is the "preserving role," the technical infrastructure required to support digital object preservation cannot be defined separately from the basic technical infrastructure of the enterprise. Rather, a successful digital preservation program must be seen as part of the day-to-day operations and basic systems of the organization as much as possible in order to ensure long-term accessibility and usability.

As a first step in building this technical infrastructure, it should be defined within a consistent technical architecture for the enterprise. The purpose of the architecture is to guide the development of the overall information systems infrastructure. It establishes consistency by helping to:

a. Provide a mechanism for a constant view of the information system infrastructure to serve as the basis from which the various groups of IT professionals develop and deliver information systems and services;
b. Provide business support services managers and staff with an understanding of the information systems infrastructure they are using;
c. Ensure that external development projects or application packages do not make incompatible changes to the infrastructure.

KU is in the process of defining its target enterprise architecture; infrastructure recommendations made in this report will continue to inform, and be informed by, that process.

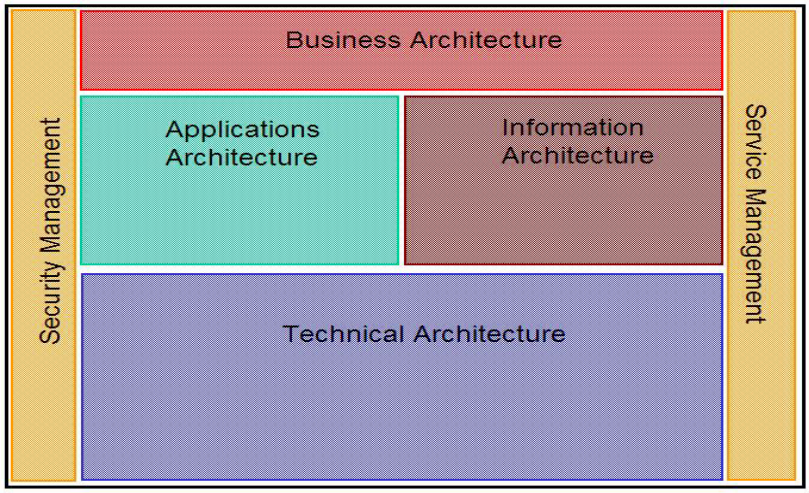

Figure 3: KU Enterprise Architecture Framework

As shown in the KU Enterprise Architecture Framework (Figure 3), a key component of the framework is the Business Architecture, or business processes that must be supported. In the case of digital preservation, these processes broadly include those described in Section IV: Create, Incorporate, Consume, Evaluate, Ingest, Store, and Manage (or "Preserve").

In turn, these processes can be mapped to a basic architectural view comprised of several layers that approximate KU's technical organization, including:

- Information (Metadata)
- Applications
- Data Management
- Middleware
- Platforms (Servers, Storage)
- Network / Utilities
- Security Management
- Service Management

Within this architectural framework, a number of infrastructure components that contribute to lifecycle information management can be identified; see Appendices H ("Information Lifecycle Management Infrastructure") and I ("Process, Architecture, Infrastructure") for graphic and tabular representations of these components.

These components, their status, and examples within the current KU information technology infrastructure, include:

- **Digital Objects:** Digital material submitted for registration and management. Objects can be physical or logical. They will have sufficient metadata (supplied by both the submitter and the system) associated with them for access, retrieval, and management. They should be identified via non-contextual URN(s) contained in the object's administrative metadata and linked to physical storage location(s) via *Name Resolution Service(s).*
  Status: Exist
  Example: scholarly papers, learning objects, administrative data, datasets, etc.

- **Registry Services:** provide the initial ingest and registration of digital objects into the overall management infrastructure. These services draw on additional services such as *User Profiles*, *Metadata Templates*, *Data Management Rules,* and *Name Resolution Service(s)* to lead

submitters through the processes and decisions required to collect descriptive, technical, and administrative information about the submitter/creator and the digital object, apply appropriate classification and disposition elements, determine the appropriate metadata and storage repositories, and assign the appropriate URN. These services should be available for both interactive and batch submissions. They should be web-based and/or include a desktop "drop-box" feature for ease of use.

Status: Minimal for selected specific systems; need a generic, standalone service that interacts with multiple systems

Example: Submission process for DSpace (KU ScholarWorks)

- **Data Management Rules Repository/Service:** provides information required for both short- and long-term management of metadata and objects. These rules determine such issues as appropriate *Metadata Templates* / metadata formats, acceptable object formats, classification / records groups, disposition of metadata and objects (methods, dates, etc.), access rights, etc. These rules are developed based on approved *standards, policies, and best practices*. Rule elements are associated with objects (either directly stored or via pointers, as appropriate) as administrative or technical metadata.

  Status: Not yet started locally; need to review for external options

  Example: None

- **User Profiles:** include basic information about the object submitter. Elements could include correct form of name, department affiliation(s), address, default repository(ies), default object classifications / groupings, default subject terms / keywords, default access rights designations, etc. Selected elements could be linked to authoritative databases such as directory services and authority files for name / address information, *Data Management Rules* for current object classifications / groups, a list of available repositories, etc.

  Status: Minimal development

  Example: Directory services and authority files for some elements

- **Metadata Templates:** provide forms for collecting standardized descriptive (Dublin Core, MARC, EAD, IMS, etc.), technical (file format, file relationships, software / version, scanner settings, etc.), and administrative (access rights, classification/disposition, etc.) information. Information is collected through direct input by the submitter, by system analysis of the submitted object(s), or by system assignment. Metadata records are stored in managed *Metadata Repositories*.

  Status: In place for selected descriptive formats and systems; primarily for staff use; need standalone versions of templates to link to *Registry Service(s)*

  Example: Dublin Core format for DSpace, ENCompass; MARC format for Voyager; EAD; Geospatial format for DASC; LUNA

- **Metadata Repositories:** managed repositories for storing descriptive, technical / structural, and administrative metadata. Managing digital files for ongoing accessibility will be highly dependent on preservation metadata. Metadata will be stored in a standardized format such as METS. Administrative and/or technical metadata elements will include standardized URNs/URIs linking to objects in object repositories via the *Name Resolution Service(s)*. Metadata is maintained (or harvested) and indexed for end-user searching and *Rules Processing* for long-term disposition and/or preservation management of objects.

  Status: In place for selected descriptive formats and systems; primarily for staff use

  Example: DSpace, Voyager, ENCompass, LUNA, PeopleSoft, DASC

- **Object Repositories / Managed Storage Facilities:** are managed repositories of digital objects. Repositories can be either centrally managed or distributed. They can be online, near-line, or offline (disk, tape, etc.). Objects are maintained as Masters and Access Derivatives (as

appropriate). Objects may include bundled metadata for preservation purposes (self-definition). Repositories should be registered and certified based on compliance with approved / standardized storage facility management policies and procedures including backup procedures, refresh / migration management, disaster recovery plans, security procedures, etc. *Note: backup procedures and disaster recovery plans alone do not equate to a complete digital preservation plan and process*[2]. File management procedures must tie into *Name Resolution Services* to maintain accurate links to *Metadata Repositories* and records for objects.

Status: In place at a basic level for selected formats and systems
Example: DSpace, Voyager, ENCompass, PeopleSoft, DASC. BEsafe project

- **Name Resolution Service(s):** The persistence of links between resources and resource discovery services is essential to ensure long-term access to materials. A persistent identifier is a name (URN: Uniform Resource Name) for an object which will remain the same regardless of where the object is located. Use of a persistent identifier will ensure that when an object is moved, or its ownership changes, the links to it will remain actionable. It is important to note that a *Name Resolution Service* (NRS) will only be effective if it is maintained. When objects are moved, the current location must be associated with the persistent identifier through use of a resolver database. A resolver database is used to translate / map the name (URN) to a current location (URL: Uniform Resource Locator).

  Status: Minimal implementation; need to review for decision on option(s)
  Example: Handles (CNRI) for DSpace, NRS (Harvard), Purls (OCLC)

- **Rules Processing:** processes / utilities that can be invoked for a variety of management tasks such as applying record disposition rules (deleting expired files, changing access rights, etc.), identifying object normalization processes/utilities, identifying objects in specific formats that should be converted to function with new software versions, identifying objects that can be migrated from online to offline storage, etc. *Rules Processing* must interact with the *Data Management Rules Service*, *Metadata Repositories*, file management utilities, conversion utilities, and *Name Resolution Services*. Audit trails detailing actions taken should be maintained as part of the object's administrative / technical metadata.

  Status: Not yet started locally; need to review for external options
  Example: None

- **Indexing Service(s):** index descriptive, technical, and administrative metadata and object content (as appropriate) stored in managed, registered *Metadata Repositories*. These services will generally be distributed and associated with specific applications. Object identifiers should be based on URNs; resolution of URNs should be through the *Name Resolution Service(s)*. Indexes can be used by end-users and by management utilities such as *Rules Processing*.

  Status: In place for selected descriptive formats and systems; primarily for staff use
  Example: DSpace, Voyager, ENCompass, DASC, campus website Google service

- **Search Service(s):** provide the ability to search indexes and discover object metadata and content. Search services will generally be distributed and associated with specific applications.

  Status: In place for selected descriptive formats and systems; primarily for staff use
  Example: DSpace, Voyager, ENCompass, DASC, campus website Google service

- **Retrieval / Transformation / Display Services:** processes / utilities that retrieve metadata and objects from managed repositories; dynamically render metadata / objects into usable formats

[2] For additional details on management requirements see RLG/OCLC Working Group on Digital Archive Attributes, *Trusted Digital Repositories: Attributes and Responsibilities*, Research Libraries Group: Mountain View, CA, May 2002, **http://www.rlg.org/longterm/repositories.pdf** and the section on *Storage* at the PADI site, **http://www.nla.gov.au/padi/topics/10.html**.

through either transformation or emulation processes; and display rendered objects to the user. These services work in conjunction with *Indexing* and *Search Services*.

Status: some utilities in place for selected transformation services; Web browsers for display

Example: XML to HTML (ENCompass); batch processes for conversion to MrSID and JPEG2000 (GIS); IE, Firefox/Mozilla, Safari

- **Identification / Authentication / Authorization Services:** services that identify and verify users and their membership in a specific community (i.e. university faculty / staff / students, general public, etc.) and indicate what resources they are allowed access to and/or what actions they are allowed to perform on those objects.

  Status: Basic development in place; need better integration across systems

  Example: Argus/AIMS, Shibboleth

- **Standards / Policies / Best Practices:** The use of standards is one strategy used to assist in preserving the integrity of and access to digital information. Use of standards can also help ensure best practice in the management of digital information. There are standards for many different aspects of storing and accessing digital information, including standards for interoperability, data formats, resource identification, resource description, data archiving, and records management. Digital preservation policies give structure and general direction for specific actions.

  Status: Under development, various stages

  Example: Digital Initiatives recommended standards; DASC standards; DSpace policies

To create this infrastructure the following steps are recommended:

- Review existing major digital asset systems on campus to determine compliance with the Information Lifecycle Management (ILM) Infrastructure Components outlined above.
- Develop specifications for and begin development or purchase of components that are currently missing in the overall ILM Infrastructure. **We recommend that initial work be focused on *Registry Service(s), Data Management Rules Service, and Name Resolution Service(s).***
- Identify responsible units for management of specific ILM components, procedures, and policies based on staffing roles outlined below (Section VI).

VI. Organization: Recommended Roles and Policies

1. Priorities for Roles

The ongoing reorganization of Information Services staff to reflect overall enterprise architecture will significantly facilitate realization of the vision proposed in this report. **The following recommendations focus on selected critical roles that we believe should be strengthened within the University** and are intended to be consistent with the overall organizational trend**.** In some cases, these are new roles that will need to be assigned to one or more individuals. In other cases the roles are already filled but require greater capacity (more staff, or staff with greater authority).

a. ***Consultants for Data Creation:***

- Metadata specialists able to model metadata in standard ways that lead to long term understanding and preservation of the context of preservable information.
- Database and repository designers who understand the principles of constructing interoperable and forward-migratable containers for information
- Content specialists who understand the specialized formats for a variety of types of digital media and can advise creators in developing media and digital documents of enduring quality.

- Trainers able to educate creators in best practices for digital creation.

b. Applications and Data Services Specialists:

- Data specialists able to create interoperability between streams of data so that it can be consumed in ways that make our backend processes invisible to the end user.
- Data analysts and programmers able to transform information as needed for forward migration as older databases, applications and systems are replaced.
- Analysts and programmers able to write code to develop submission systems, move data from system to system, etc.
- Metadata specialists to create metadata.

c. Policy Developers:

- Administrators responsible for developing the policy framework and analysts able to develop rules for implementing policy.
- One policy-developing role of particular importance to the digital preservation program is that of records manager.

d. Operational Support:

- Staff to perform resource and capacity planning, ensure compatibility with the existing technical environment during product selection, install and configure new components, and provide ongoing production support (disk management, system and data backup services, problem response and resolution, etc.).

e. Administrative Support and Oversight:

- *Oversight:* Although digital preservation should not be seen as a discrete activity but rather as the outcome of successful day-to-day management of working digital assets, it is nonetheless important that someone in the organization be charged with remaining current with best practices, auditing local practices for consistency with established architecture and best practices, identifying needs for new resources, staff roles, and infrastructure, and keeping senior administration informed. We believe that could be assigned either to a single individual or to a small steering committee. Consideration should be given to assigning this role to the Preservation Officer position currently vacant within the Libraries, in cooperation with the IS Program Office currently being developed.

- *Sponsorship/Advocacy:* Especially in the early stages of the digital preservation program, it is important that a senior administrator with major authority for resource allocation and policy-setting act as a sponsor for digital preservation and advocate for the program at the administrative level.

2. Policy Priorities

A successful digital preservation program depends equally on an appropriate technical *architecture* and on preservation-oriented *practices* being adopted within the KU community. Faculty, administrators, and staff will need to know how to act and will need incentives to act in preservation-oriented ways. One component of an effective preservation program is therefore a set of policies and guidelines. **We recommend that development or completion of the following policies receive priority, and that they be incorporated into educational programs:**

a. Preservation Policies Re: Appraisal

Not all university digital assets can be preserved; appraisal policies need to be developed to set priorities.

- *Administrative record appraisal and schedules:* Many of the administrative records generated by the university can be designated in advance for discard, discard after a set term, or permanent retention. Some of these rules are set by the state of Kansas. Work is already underway to develop a records schedule for the University that will include electronic records.

- *Non-administrative digital assets held by University units* (including academic research products): Appraisal of digital assets not covered by administrative record schedules should be the responsibility of the custodial units, with guidance and consultation provided from a central authority.

    - Develop a template by which units can identify and inventory classes of digital assets under their stewardship (the Task Force has developed an imperfect prototype; see Appendix E, "Data Collection Manual").

    - Develop guidelines to help determine their relative priority for long-term retention.

b. Preservation Policies Re: Creation of New Content

- *Recommended file formats:* We recommend that a set of guidelines or recommendations be developed for common preferred file formats (for word processing and spreadsheets for example), and that specific support levels be associated with them (e.g., for certain formats we would commit to preserving appearance and functionality, while for other formats we would commit only to preserving the bitstream. See Appendix G for further information.) Initial recommendations made for Digital Initiatives projects and KU Scholarworks should be reviewed for possible incorporation or extension.

- *Intellectual property for academic work by KU authors and deposit of academic work in KU ScholarWorks:* We recommend that conversations begin toward a university or governance resolution calling upon all faculty to (a) deposit scholarly work of enduring value in the KU ScholarWorks repository and (b) attempt to retain certain rights in their published scholarship, including the right to disseminate it through a university repository.

- *Permission to copy work on KU servers for preservation administration:* In a digital environment, the preservation process is inherently one of *copying* – creating new copies of an original file onto new media or transforming files into new versions. Preservation of locally hosted academic work could be hampered by lack of clarity regarding university rights and responsibilities for work hosted on university equipment. We recommend that policy be developed affirming the University's expectation that preservation processes such as migration and file transformation will be conducted on materials hosted on university equipment for the purpose of maintaining enduring access.

c. Preservation Policies Re: Resources and Infrastructure

- *Best practices for Technical Liaisons and System Administrators outside the Information Services structure:* We recommend that a set of guidelines be developed to define best practice in the administration of servers and information processes regarding back-up protocols, security, etc.

- *Resource allocation:* A commitment to a university-wide digital preservation program cannot be undertaken without a clear understanding of funding sources. Some elements of the program will be funded through central Information Services, while others may be the responsibility of individual units (sometimes though a contracted-services model). We recommend that a policy be developed to identify those different elements and funding responsibilities.

VII. Creating a Culture of Digital Stewardship: Outreach and Education

A. Training Model for Digital Preservation

Key to the success of digital preservation planning on the University of Kansas campus is the recruitment and involvement of staff at all levels of the University. Content creators, policy developers, and data stewards all have responsibilities for the preservation of digital materials that are generated as part of the University's mission.

An education program should be developed to provide Information Services staff and the greater KU community with the information needed to make decisions regarding preservation issues. The program will include a series of workshops designed to provide basic information to engender a general awareness for everyone in addition to more advanced instruction for those whose specific job responsibilities include the ongoing maintenance of systems and data.

The suggested curriculum includes five primary focus areas:

1. General Awareness: Recommend for all University staff, beginning with Information Services staff
2. Information Life Cycle Management
3. Information Storage Management and System Maintenance
4. Best Practices and Standards
5. Legal Issues and University Policies

Recommendation: We recommend that appropriate staff be commissioned to fully develop and implement this educational program. An outline of topics to be covered, at both basic and advanced levels, and the staff groups that should be targeted is in Appendix J.

B. Website for Digital Preservation Resources

In further support of the education program, a website for digital preservation resources should be developed. This website would disseminate selected information on current standards, research and pragmatic approaches to digital preservation. The site is intended to primarily support the needs of the KU Information Services community and KU community at large. Secondarily, it will serve as an information resource for the library and information technology communities outside of KU.

Initial content for the website has been developed for the following components (see Appendix K):

- Good information practices: Tips for authors and creators of digital files
- Information management practices and standards
- Selected digital preservation research and initiatives
- Selected resource materials

Recommendation: The Task Force recommends that:

- Responsibility for the creation and maintenance of the website be assigned to one or more individuals
- Responsibility for overseeing the addition of new content be assigned to one or more individuals
- A website implementation date be determined

VIII. Concluding Observation: New Opportunities for the Digital Academy

Continuing access to digital content is fundamental to the University's fulfillment of its mission as an administrative agency, a custodian of the record of scholarship, and a creator of new knowledge. Without a program of lifecycle management, the University will not be able to guarantee future access to financial data, student records, or faculty research – guarantees that are legitimately demanded by the University's sponsors and stakeholders.

However, implementation of this program will yield benefits that exceed the merely custodial. The technical, staff, policy, and educational structures that we recommend are precisely those needed to make possible the new practice of scholarship and teaching in the digital academy – an academy where scholarly authors can readily share research reports and primary data with appropriate access controls; where teachers can easily re-purpose research material for beginning and advanced learners, and students learn the practices of digital authorship as part of their curriculum; and where business services support the needs of its members. Digital preservation, or lifecycle management, can thus be used as a framework for guiding KU's evolution in the early years of the 21st century.

For data to be used effectively by decision-makers, scholars, researchers, staff, and students, it must flow through smart pathways that conform to a service-oriented architecture. We believe that as the amount of information "born digital" or transferred to digital format grows, there will be an accompanying need for more staff who work in the broad discipline of data services and who provide robust support for both current and future operations. By "data services" we refer to those staff who design our data architecture, staff who administer databases and data repositories for many different administrative and academic applications, staff who create metadata models and populate repositories with metadata, and staff who design mechanisms for moving data between a variety of information systems or transform it for use by consumers. Although pockets of data services specialization may continue to exist in areas like statistical and research computing or library bibliographic cataloging, a horizontal data services team able to serve the University broadly in creating, maintaining, and moving the spectrum of University data from system to warehouse, system to system, system to user, and ultimately from present systems to future systems, provides the best opportunity for understanding, accessing, and preserving the University's information assets.

IX. Summary of Recommendations

A. Technical Infrastructure:

- Review existing major digital asset systems on campus to determine compliance with the Information Lifecycle Management (ILM) Infrastructure Components outlined above.
- Develop specifications for and begin development or purchase of components that are currently missing in the overall ILM Infrastructure. **We recommend that initial work be focused on *Registry Services, Data Management Rules Service, and Name Resolution Services.***
- Identify responsible units for management of specific ILM components, procedures, and policies based on staffing roles outlined in section VI.

B. Strengthen the Following Staff Roles:

- ***Consultants for Data Creation:***
  - Metadata specialists
  - Database and repository designers
  - Content specialists
  - Trainers

- ***Applications and Data Services Specialists:***
  - Data specialists
  - Data analysts and programmers
  - Application analysts and programmers
  - Metadata specialists

- ***Policy Developers***

- ***Operational Support***

- ***Administrative Support and Oversight***

C. Develop or Complete the Following Policies:

- Administrative record appraisal policy and retention schedules

- Template for units to identify and inventory classes of non-administrative digital assets under their stewardship (see prototype at Appendix F, "Data Collection Manual").

- Guidelines to help custodial units determine the relative priority of non-administrative assets for long-term retention.

- Recommendations for common preferred file formats

- Resolution calling upon all faculty to (a) deposit scholarly work of enduring value in the KU ScholarWorks repository and (b) attempt to retain certain rights in their published scholarship, including the right to disseminate it through a university repository.

- Right to copy work on KU servers for preservation administration

- Best practices for technical liaisons and system administrators outside the Information Services structure

- Model for resource allocation for digital preservation services

D. Educational Program: Commission appropriate staff to fully develop and implement an educational program. An outline of topics to be covered, at basic and advanced levels, and the staff groups that should be targeted is in Appendix J.

E. Website: Assign responsibility for the creation and maintenance of a digital preservation website. Suggested homepage text and links for the site are in Appendix K.

X. List of Appendices

A. Task Force Charge
B. Task Force Membership
C. Digital Preservation Best Practices: Working Summaries and Background Information
D. Definition of University Digital Assets
E. Draft Data Collection Manual
F. Lifecycle Management Roles and Processes
G. Good Information Management Practices
H. Lifecycle Management Infrastructure
I. Process, Architecture, Infrastructure
J. Digital Preservation Curriculum
K. Website Text and Links
L. Glossary of Terms: Tech to English