

# FEDORA Digital Repository Implementation at UVa

or, You've All Heard About Fedora, Here's What We're Doing with It

Leslie Johnston, UVa Library
DLF Spring Forum
May 15, 2003



# Background

In 1999, The UVa Library's Digital Library Research and Development (DLR&D) group read an article in <u>D-Lib</u> about the Fedora (<u>Flexible</u> and <u>Extensible</u> <u>Digital</u> <u>Object and</u> <u>Repository Architecture</u>) system, designed by Carl Lagoze and Sandra Payette of the <u>Cornell Digital Library Research Group</u>.

By the summer of 2001 UVa built an "alpha" testbed that included 500,000 data objects and a variety of disseminators for electronic finding aids, TEI-encoded etexts of letters and books, and for XML-encoded structured collections of art, architecture and archeology images, and a set of social science data.



# Background

In late 2001, The UVa Library Digital Library Research and Development Group began collaboration with the Cornell Digital Library Research Group to develop Fedora under a \$1,000,000 <a href="Mailto:Andrew W. Mellon Foundation grant">Andrew W. Mellon Foundation grant</a>. A number of deployment partners signed on to test deploy the system during its development.

On May 16, 2003, release 1.0 of the <u>Fedora</u> digital object repository management system was made available under a Mozilla Public License through the <u>project web site</u>. The UVa Library is also launching its first phase production repository within the UVa domain.



### Fedora Architecture

- Fedora is written in Java 1.4.
- A Fedora repository is exposed as a Web service and is described using Web Services Description Language (WSDL).
- Digital object behaviors are implemented as linkages to distributed web services that are expressed using WSDL and implemented via HTTP GET/POST or SOAP bindings.
- Digital objects are encoded and stored as XML using the Metadata Encoding and Transmission Standard (METS).



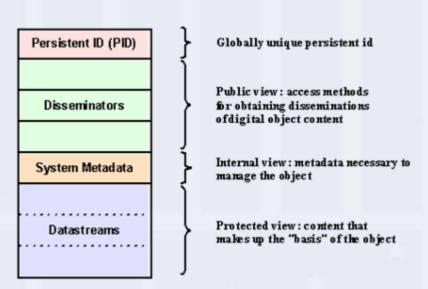
### Fedora Architecture

- The Fedora architecture is based on object models that represent data objects (units of content) or collections of data objects.
- The objects contain linkages between datastreams (internally managed or external), metadata (inline or external), and behaviors that are themselves code objects and link to disseminators (processes, mechanisms, and external software.
- Object models can be thought of as containers that give a useful shape to information poured into them; if the information fits the container, it can immediately be used in predefined ways.



# Object Models

- A Fedora object has four basic components:
  - A persistent identifier or PID. Handles are not currently used.
  - A set of disseminators that define a set of behaviors the object can perform.
  - A set of descriptive and administrative metadata about the object and its content.
  - One or more datastreams that define the content of the object.
- Objects "subscribe" to the same object model if they share the same basic object structure by having the same number and type of datastreams (content streams) and by having the same set of disseminators or behaviors.





### Design of Models and Disseminators

- Fedora developers met with image and text content and format specialists, application developers, and user service librarians to understand what media files we have and how our users expect to find them and use them.
- Specifications were set for:
  - Datastreams (formats, variation in deliverables [EAD vs. TEI vs. Ebooks, page images vs. documentary images])
  - Metadata
  - Discovery (metadata vs. full-text searching, presentation of results sets, etc.)
  - Delivery (must support static and on-the-fly file delivery, and varied end user download and printing requirements)



# Image Object Model

- The <u>General UVa Image Object Model</u> design includes four separate datastreams (or basis):
  - Preview- sized images
  - Screen-sized images
  - A MrSID version
  - The "Delivery Master" (the image that was used to derive the other datastreams)
- The disseminators that define the functionality of the object by describing a set of behaviors or Behavior Definitions. The uvaImage disseminator consists of five behaviors
  - getPreview retrieve the preview size of the image
  - getScreen retrieve the screen size of the image
  - getImageViewer retrieve the screen size of the image but with additional image manipulation tools
  - getSizedImage(x,y) retrieve the specified size of the image
  - getDeliveryMaster retrieve the delivery master (or ascertain its off-line location)

Persistent ID (PID)			
Disseminators			
Disseminator	Behavior Definition	Behavior Mechanism	
uvalmage	genImageBdef	genImageBmech	
	getPreview	HTTP GET	
	getScreen	HTTP GET	
	getlmageViewer	TBD	
		get_mrsid_url.pl	
	getDeliveryMaster	TBD	
uvaMetadata	uvaMetaBdef	TBD	
	getDescriptive	TBD	
	getTechnical(ds)	TBD	
	getDigiprov(ds)	TBD	
	getRights(ds)	TBD	
	getSource(ds)	TBD	
System Metadata			
uvalibadmin	Administrative metadata		
uvalibdesc	Descriptive metadata		
Basis			
D at astream(s)			
DS1	pointer to thumbnail size image		
DS2	pointer to screen size resolution image		
DS3	pointer to MrSID image		
DS4	TBD		



# Text Object Model

- All electronic texts in the Central Repository will be encoded as TEI in XML. Each digital object will also contain descriptive and administrative metadata about the object as a whole and about each of its datastreams. The uvaMetadata disseminator will be available on every object and will provide the capability to retrieve descriptive and administrative about the object and its content. The <a href="UVa General Text Object Model">UVa General Text Object Model</a> contains five datastreams:
  - Static XML version of text points to the raw XML version of the text.
  - Static XHTML version of text points to an XHTML version of the text or to a
    placeholder indicating that a static version does not exist and will be dynamically
    generated.
  - Static PDF version of text points to PDF version of the text or to a placeholder indicating that a static PDF version does not exist and will be dynamically generated.
  - Static PDB version of text points to PDB version of the text or to a placeholder indicating that a static PDB version does not exist and will be dynamically generated.
  - Static LIB version of text points to LIB version of the text or to a placeholder indicating that a static version does not exist and emails a request to have one externally generated.



# Text Object Model, continued

- It is desirable to have a single behavior definition for all electronic text, but with multiple implementations (behavior mechanisms) for each different type, sharing the same behavior definition. The General Text Behavior Definition defines eight basic behaviors for all electronic texts:
  - getPreview display a "preview" representation of the text that represents a bibliographic citation.
  - getTreeView(level) display an XML DOM node tree representation of the text down to the specified level in the text.
  - getChunk(idref) get a chunk of XML specified by the idref; this behavior retrieves the specified XML fragment from the text.
  - getChunks(XPath) get multiple chunks of XML specified by the XPath expression; this behavior may retrieve multiple XML fragments from the text.
  - getStaticView display static HTML view of text; the static view would be a view just of the text itself.
  - getDynamicView display the text in an interactive HTML form; the dynamic view would include the full set of support tools available for the particular type of text.
    - getPrintable(format) download a version of the file in the specified format; at present, three formats would be allowed including .pdf, .pdb, and .lib.
    - getDeliveryMaster download raw XML text of delivery master

Persistent ID (PID)			
Disseminators			
Disseminator	Behavior Definition	Behavior Mechanism	
uvaText	uvaTextBdef	uvaTextBmech	
	getPreview	TBD	
	getTreeView(level)	TBD	
	getChunk(idref)	TBD	
	getChunks(XPath)	TBD	
	getStaticView	TBD	
	getDynamic∀iew	TBD	
	getPrintable(format)	TBD	
	getDeliveryMaster	HTTP GET/FTP	
uvaMetadata	uvaMetadataBdef	uvaMetadataBmech	
	getDesc(ID)	TBD	
	getTech(ID)	TBD	
	getDigiprov(ID)	TBD	
	getSource(ID)	TBD	
	getRights(ID)	TBD	
System Metadata			
uvalib:admin	Administrative metadata		
uvalib:desc	Descriptive metadata		
Datastreams			
D at astream(s)			
DS1	pointer to XML text		
DS2	pointer to static XHTML version of text		
DS3	pointer to PDF version of text/placeholder		
DS4	pointer to PDB version of text/placeholder		
DS6	pointer to LIB version of text/placeholder		



### Effect on Production

- Development of new file naming conventions across the Library.
- Documentation of deliverable file formats and sizes.
   <a href="http://iris.lib.virginia.edu/dl/reports/best\_practices.html">http://iris.lib.virginia.edu/dl/reports/best\_practices.html</a>
- Development of <u>UVa Metadata</u>, documentation of minimum metadata standards, and mappings to TEI, EAD, VRA Core, and Dublin Core.
- Development and introduction of the <u>General Descriptive</u> <u>Modeling Scheme (GDMS)</u> and a GDMS Tool to create structured representations of collections of objects.
- Development of a workflow and tools for batch processing of deliverable files and loading into repository directory trees.
- Development of batch tools to create Fedora objects, creating the linkages between media files, metadata, and disseminators.



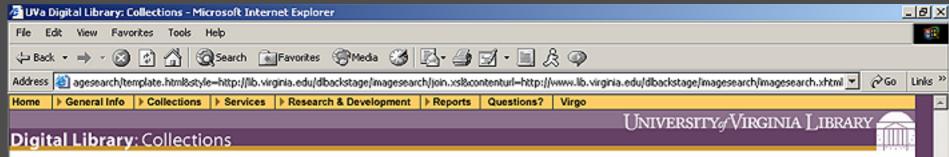
### Discovery

- Indexing and searching are handled outside of Fedora.
- The current UVa implementation uses Tamino and OpenText. OpenText will be updated to XPAT, and Tamino has proved to be a poor fit for use with full-text, so other products (such as <u>Ipedo</u>) are under review.
- The web-based discovery interface uses external indexes to build a results set; the interface uses XSLT to format results, in combination with JavaScripts that build menus on-the-fly to display metadata and available functions.



# Delivery

- When objects are selected for viewing or downloading, calls are sent to Fedora via URL parameters to retrieve the objects, which are formatted using XSLT.
- The majority of the current Fedora disseminators use a combination of Perl programs and XSLT to format the objects for display.
- Other disseminators include wrapping a zooming and panning Java applet around an image; generating a downloadable JPEG from the MrSID file; and delivering downloadable versions of XML EAD files for partnering institutions.



### Search All Image Collections

all terms. C any term. SEARCH

Present results as:

Titles only

Images only

Enter a word or phrase in the search box to find images through keywords anywhere in the image description.

In the title display, the first line is a link to the description and all related images and viewing options; additional lines refer to each related image file. The extent of the available title and description varies by image source and collection.

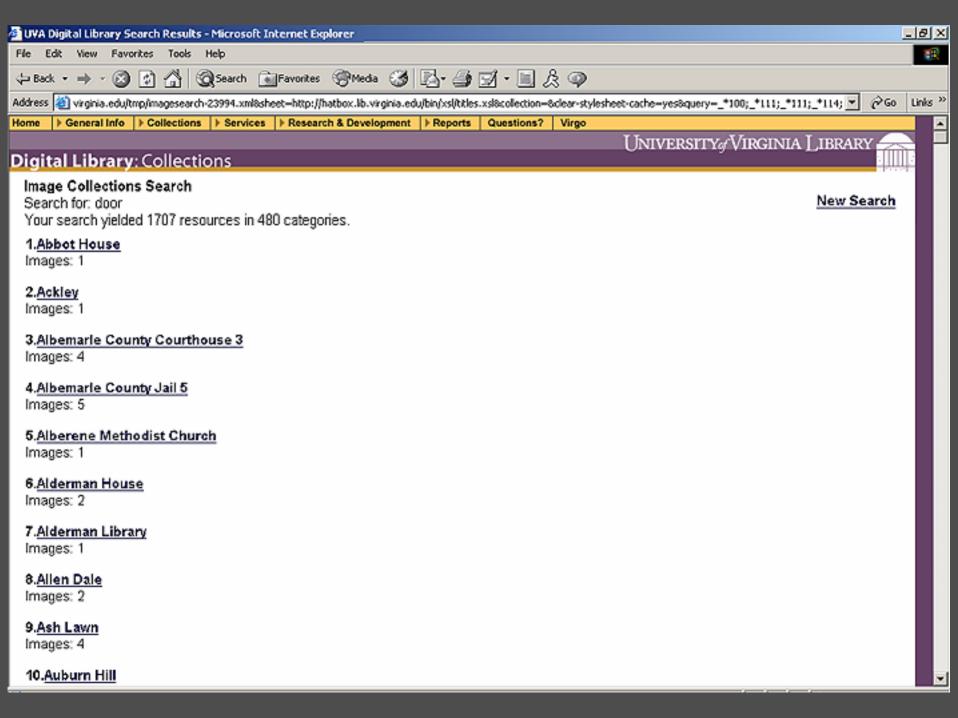
In the image and description displays, roll your cursor over the image. to see titles and options for viewing the image files.

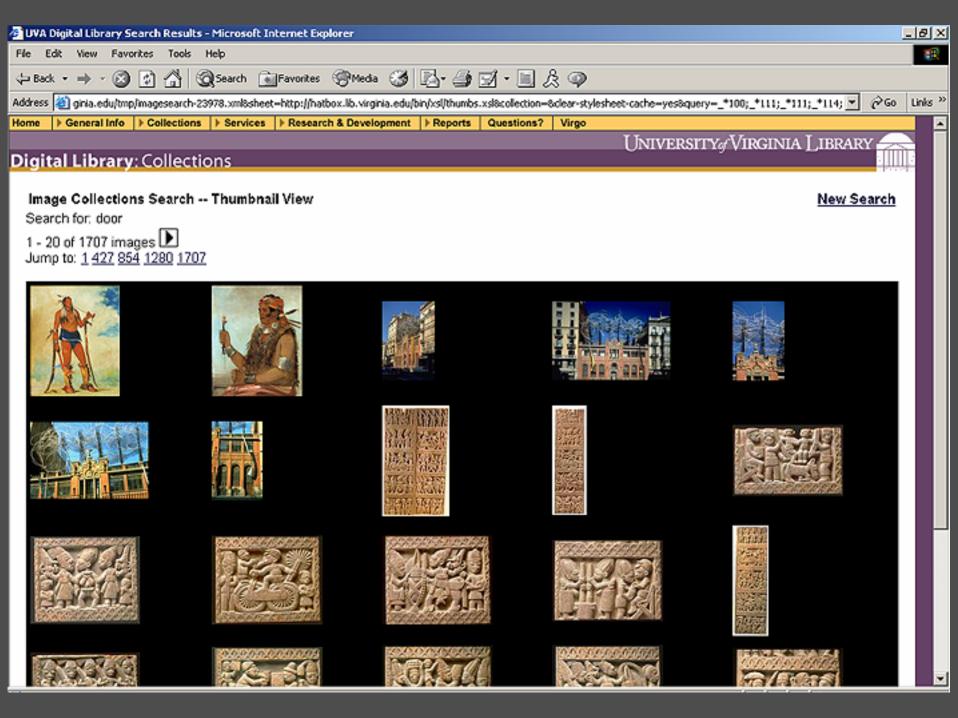
The Architecture of Jefferson Country < Central Virginia architectural inventory compiled by the University of Virginia>

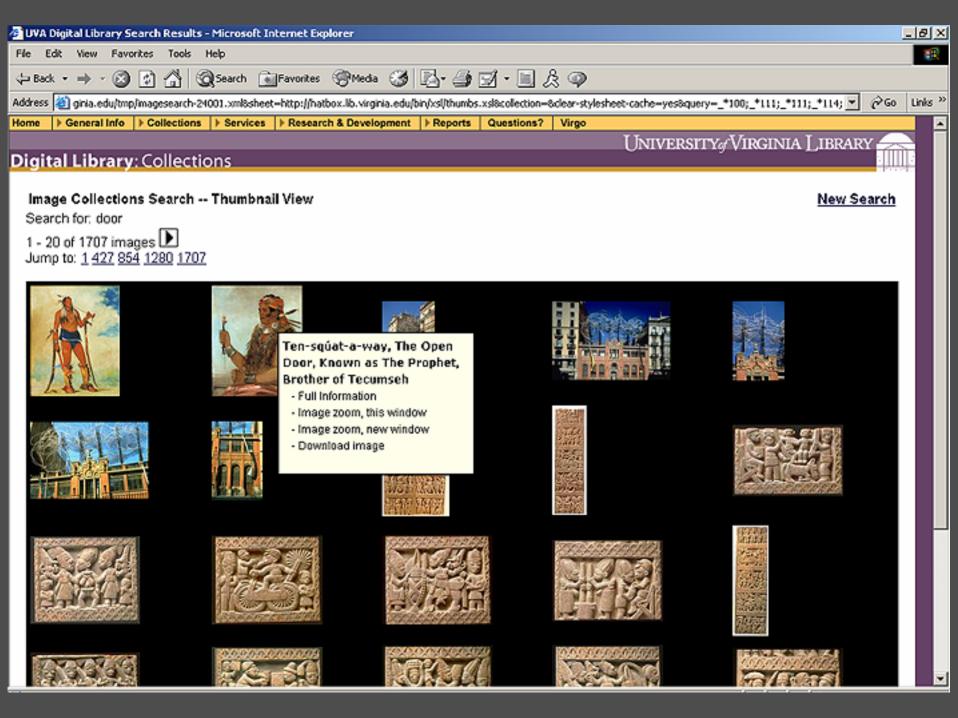
The Barcelona Collection <University of Virginia image collection for sites and buildings in Barcelona, Spain>

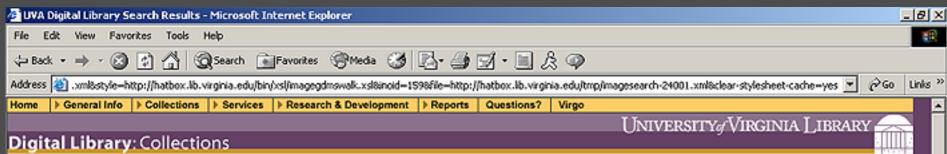
The UCLA Fowler Museum Collection of African Art

The Smithsonian American Art Museum Catlin Indian Paintings Collection









Ten-squat-a-way, The Open Door, Known as The Prophet, Brother of Tecumseh

artist: George Catlin

owner: Smithsonian American Art Museum

subject: portrait medium: oil

extent: 29 x 24 in.

credit: Smithsonian American Art Museum, Gift of Mrs. Joseph Harrison, Jr.

tribe: Shawnee creation: 1830

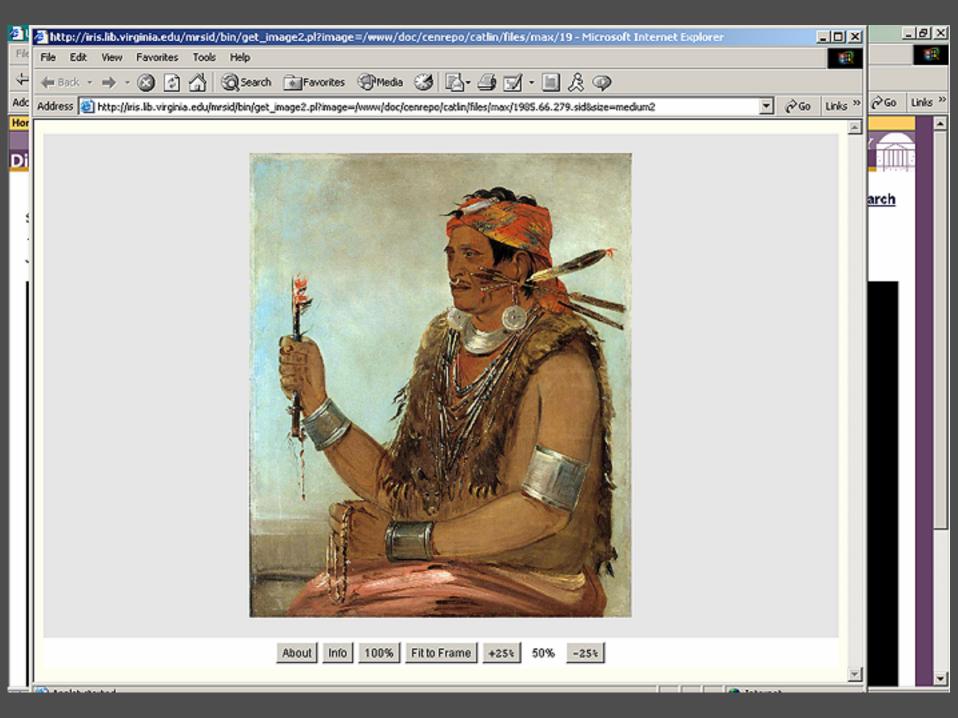
identifier: 1985.66.279

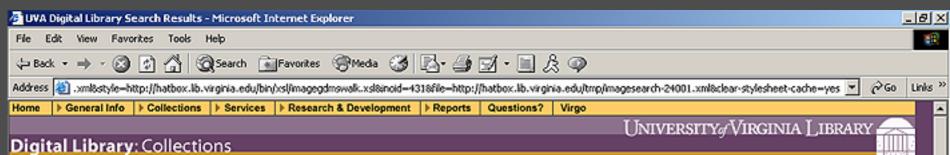


### Related Texts

William H. Truettner, The Natural Man Observed; A Study of Catlin's Indian Gallery (Washington, D.C.: The Smithsonian Institution Press in cooperation with the Amon Carter Museum and The National Collection of Fine Arts, 1979).

"The 'Shawnee Prophet,'; is perhaps one of the most remarkable men, who has flourished on these frontiers for some time past. This man is brother of the famous Tecumseh, and quite equal in his medicines or mysteries, to what his brother was in arms; he was blind in his left eye, and in his right hand he was holding his 'medicine fire,'; and his 'sacred string of beads'; in the other. With these mysteries he made his way through most of the North Western tribes, enlisting warriors wherever he went, to assist Tecumseh in effecting his great scheme, of forming a confederacy of all the Indians on the frontier, to drive back the whites and defend the Indians' rights; which he told them could never in any other





Chair and Cloud

alternate title: Editorial Montaner y Simón. alternate title: Fundació Antoni Tàpies.

creator: Tàpies, Antoni,

contributor: Domènech i Montaner, Lluís.

subject: Office buildings., Modernist., Public art, Outdoor sculpture, Sculpture

notes: Building originally a publishing house for Montaner y Simón; designed by Lluís Domènech i Montaner, 1879.

identifier: S-20-Tap-3.4

location: Barcelona, Spain ; C/Arago, 255, Eixample 41.4172.167

creation: 1990

medium: Steel and aluminum mesh.



Oblique view of building



View of sculpture

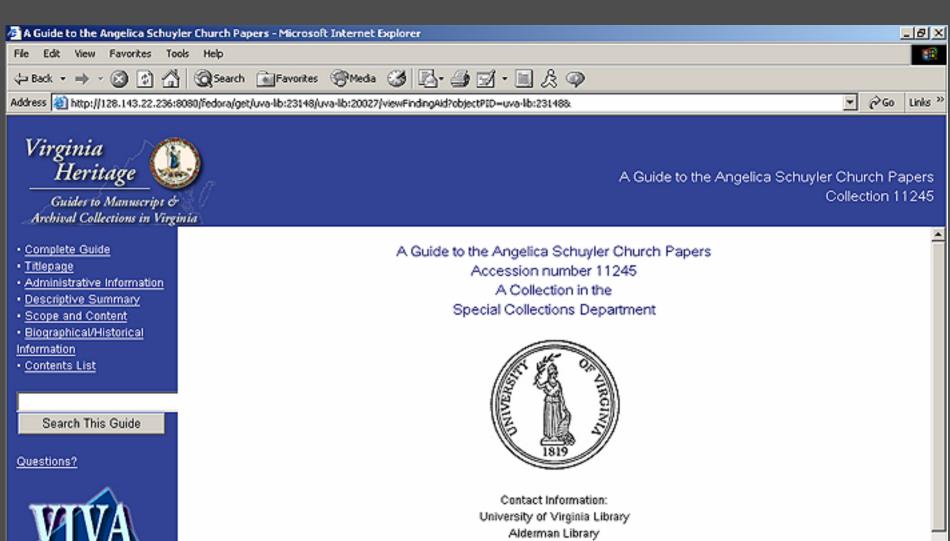


Front view with rooftop artwork





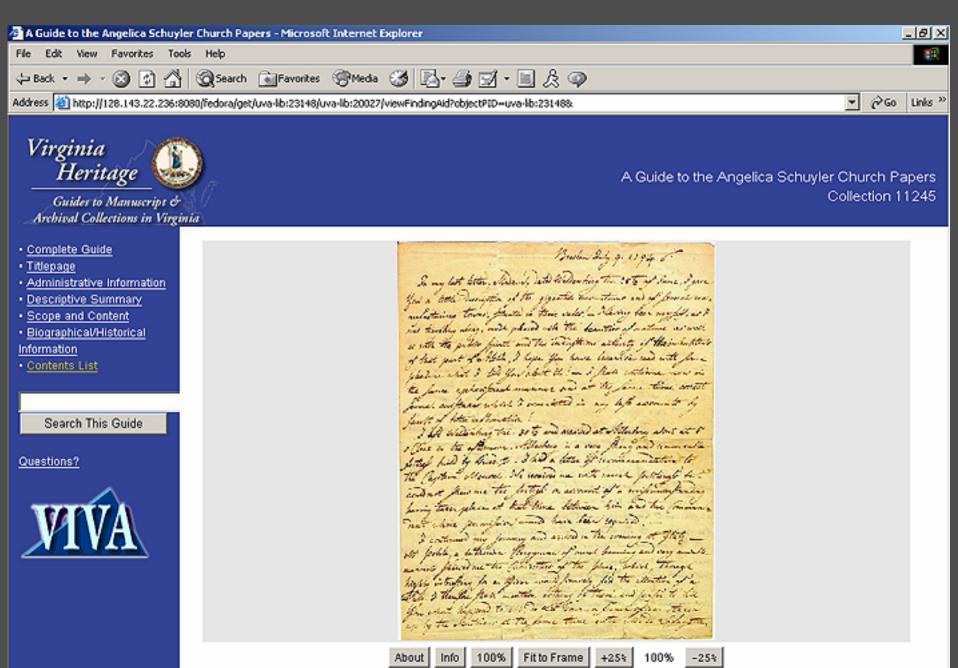
Detail, three bays of building.

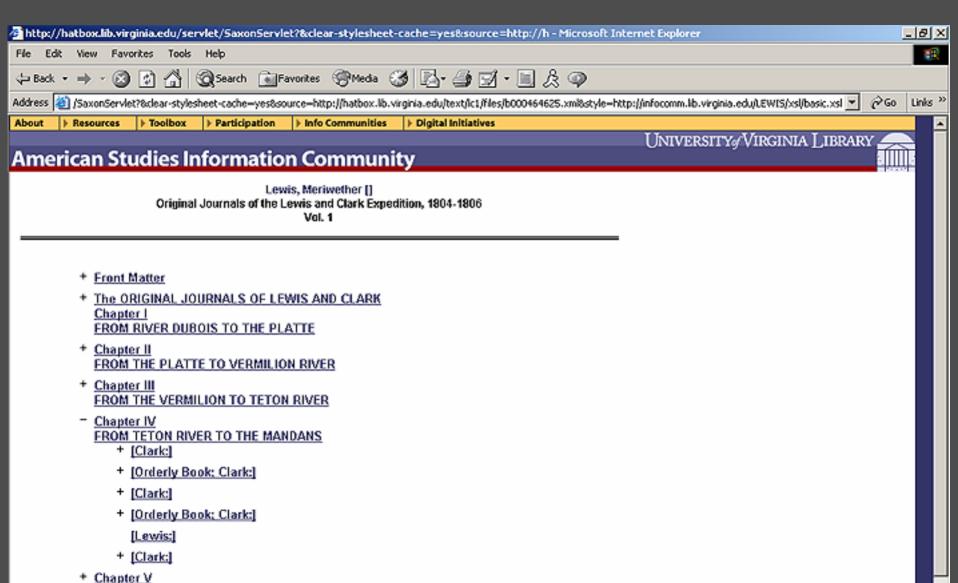


University of Virginia Library
Alderman Library
University of Virginia
Charlottesville, Virginia 22903
USA
Phone: (804) 924-3025

Fax: (804) 924-3143 Email: mssbks@virginia.edu

Processed by: Robin Wear Transcribed and translated by: Robert Dennome





AMONG THE MANDANS

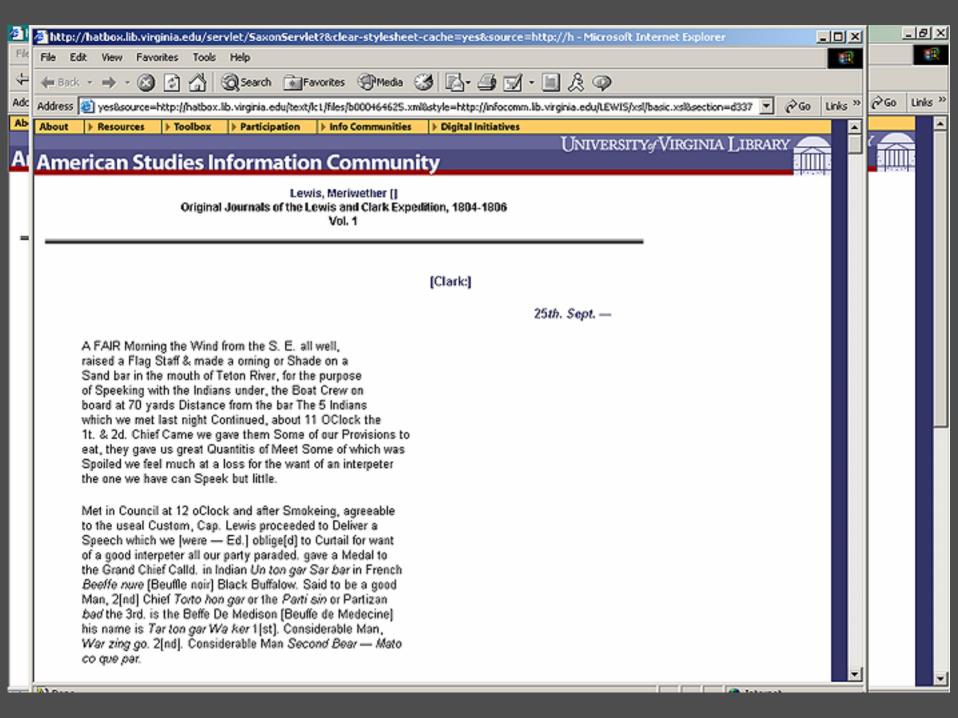
AMONG THE MANDANS

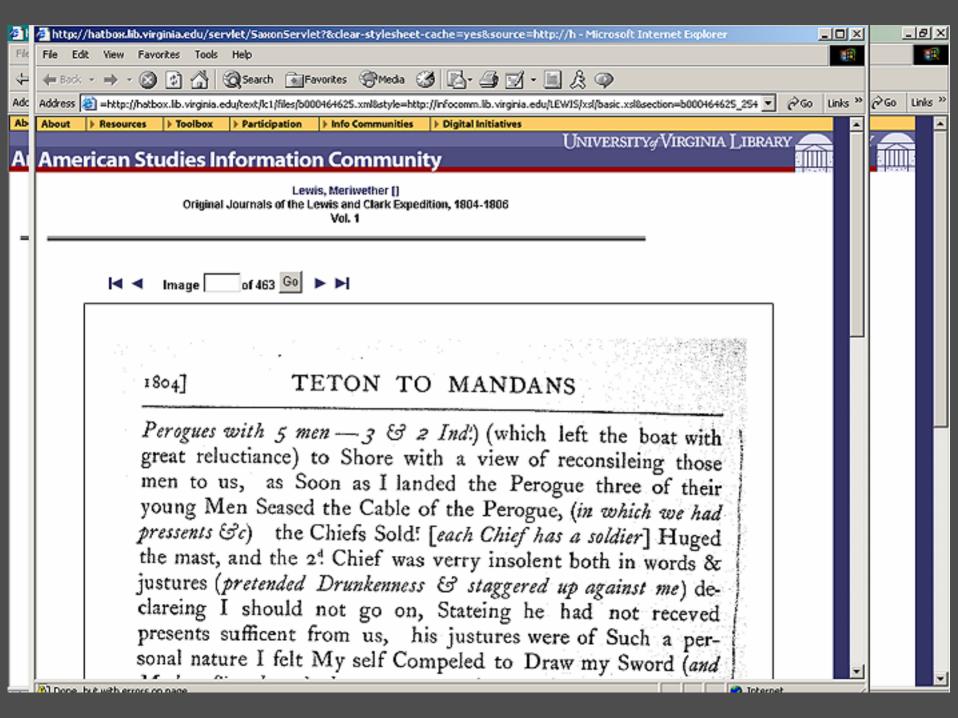
+ Chanter VIII DADT I

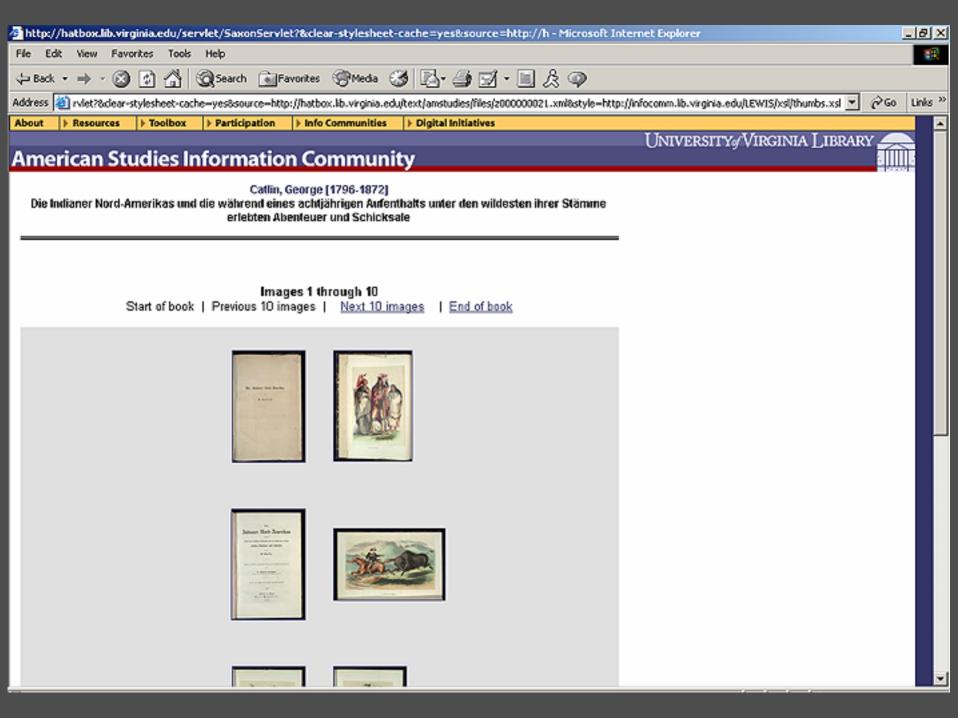
FROM FORT MANDAN TO THE YELLOWSTONE

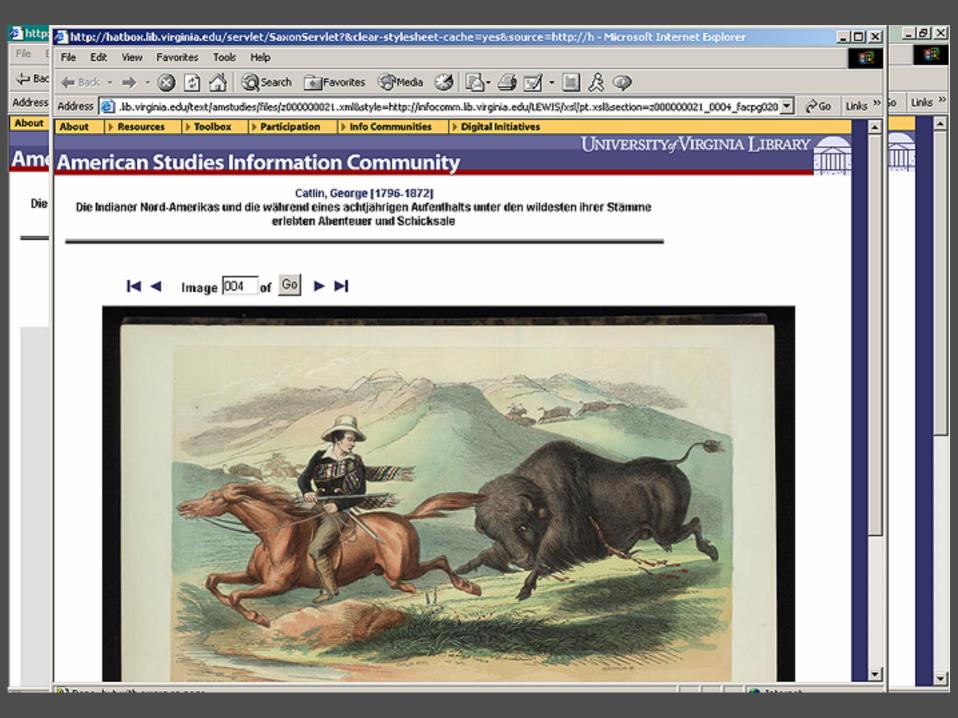
Chapter VI

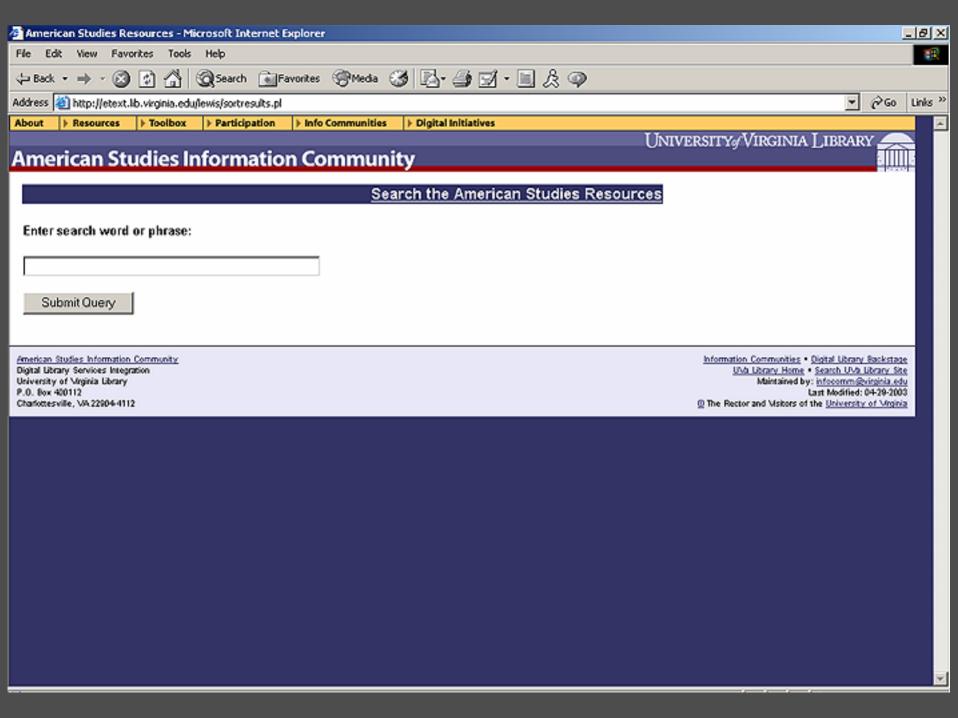
+ Chapter VII

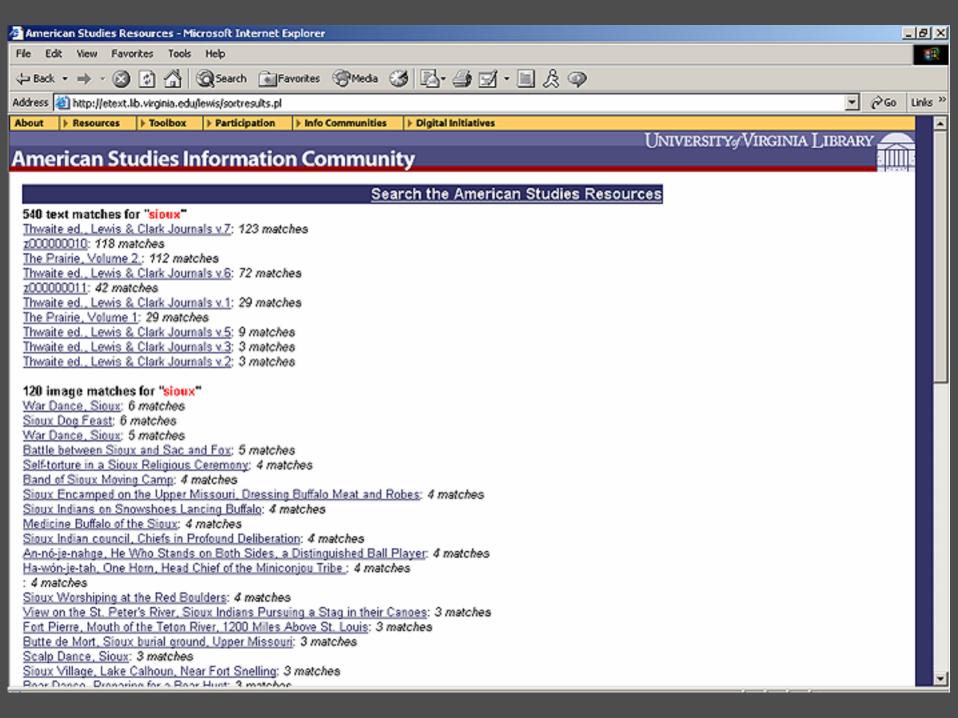


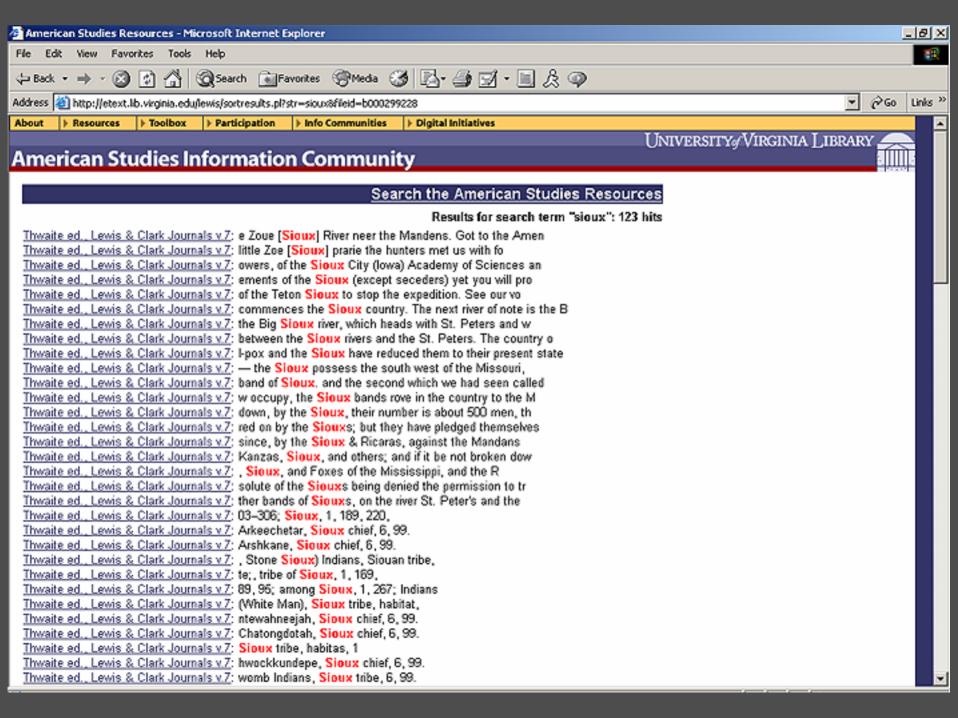


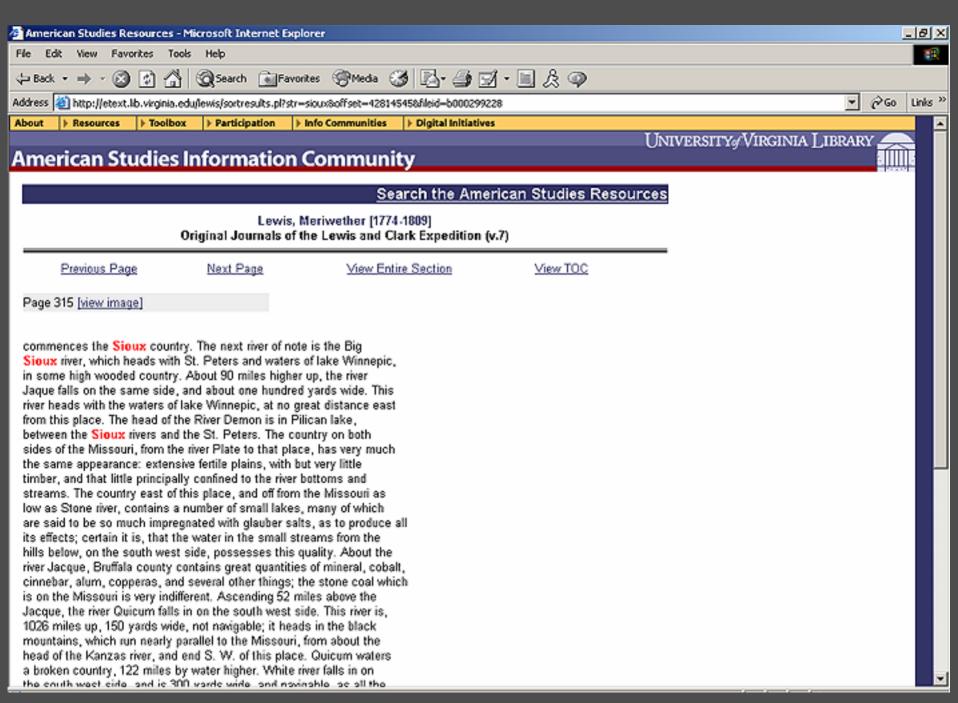


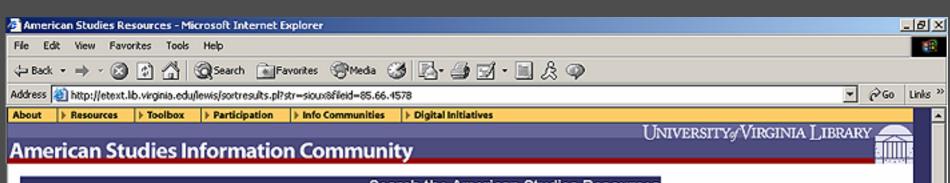












### Search the American Studies Resources

Stán-au-pat, Bloody Hand, Chief of the Tribe

subject portrait medium; oil extent: 29 x 24 in.

credit: Smithsonian American Art Museum, Gift of Mrs. Joseph Harrison, Jr.

artist: George Catlin

owner: Smithsonian American Art Museum

tribe: Arikara tribe: Sahnish creation: 1832

identifier: 1985.86.123

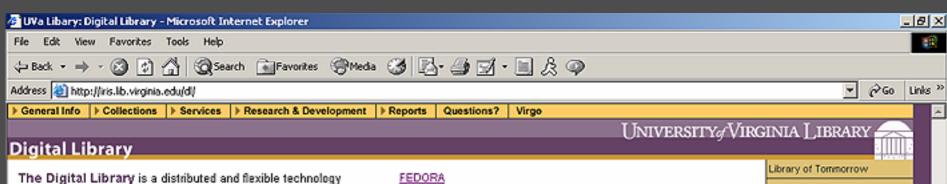


### Related Texts

William H. Truettner. The Natural Man Observed: A Study of Catlin's Indian Gallery (Mashington, D.C.: The Smithsonian Institution Press in cooperation with the Amon-Carter Museum and The Intrinsit Callection of Fine Arts, 1979).

Bloody Hand is described as having A"his face painted with red vermilion, scalping-knife in his hand,A" and A"wearing a beautiful dressA" (1848 catalogue, p. 18).

Painted at the Arikara village in 1832. The broad modeling and strong features are typical of the Upper Missouri portraits. Bloody Hand also appears, full length, in cartoon 35, with his wife.



infrastructure that enhances access, manipulation, storage, distribution, and integration of information and services throughout the University of Virginia.

The vision for the Digital Library initiatives come in part from the work of a series of committees that participated in the Library of Tomorrow (Loff) planning process in 2001, guiding a five-year program to transform the Library into the model university research library for the twenty-first century.

Digital Image Collections Search Demo

Information Communities

Electronic Centers

University of Virginia

Give to the UVa Library

### Spotlight







The integrated image collection is UVA's first repository that contains images from Library collections, licensed resources, and faculty projects where the resources are available through a single search for use by the UVa community for teaching and research.

### Announcements

Fedora Release: The first version of the <u>Fedora</u> digital object repository management system will be released under a public license on May 16, 2003, through the <u>project web site</u>.

Presentations: Thorny Staples gave a presentation on the Fedora release at the <u>CNI Task Force meeting</u> on April 28; Leslie Johnston gives a presentation on services developed with Fedora and Chris Ruotolo gives a presentation on the Information Communities at the <u>DLF Forum</u> on May 15.

Digital Library University of Virginia PO Box 400112 Charlottesville, VA 22004-4112 University of Virginia Library Home

Search the Library Site •UVa Home Maintained by: di@virginia.edu

@ The Rector and Visitors of the University of Virginia



### Future

- Further development of a heterogeneous Collection Object Model, with core behaviors that should be applicable across entire collections (getInventory, getPIDs, etc.)
- Migration of the image and text collections into Fedora, and transformation of the prototype delivery applications into Fedora disseminators.
- Development of object models and disseminators for data sets.
- Future authoring and delivery services will be developed on top of the repository for users:
  - An improved GDMS XML authoring tool that is usable by faculty.
  - Distributed repository storage in affiliated user's "home directories" elsewhere in the University environment.
  - A "shopping cart" for the repository that allows users to create and retain persistent sets.
  - Linkages between repository disseminators and commercial tools.