

# Sustaining Digital Scholarship

---

Mike Furlough, Bradley Daigle,  
Thornton Staples, Madelyn Wessel

DLF Fall Forum: November 9, 2005

UNIVERSITY *of* VIRGINIA LIBRARY

# Outline for Today

---

- Mike: Definitions and framework for digital scholarship
- Bradley: Parameters of the Sustaining Digital Scholarship project
- Thorny: Collection and aggregation models
- Madelyn: Legal and policy implications

# What is “digital scholarship?”

---

- Scholarship—especially in the humanities—which elaborates new interpretive methods for research, teaching, and learning through the use of digital resources and tools.

# Digital Scholarly Projects

---

- These are not illustrated articles, monographs, or electronic theses
- Networks of heterogeneous digital content
- Virtual exhibitions, thematic archives & scholarly editions
- Elaboration of new scholarly genres
- Tend to cross research/teaching boundary



## THE TIBETAN & HIMALAYAN DIGITAL LIBRARY

[THDL](#)[Popup](#)[Collections](#)[Reference](#)[Community](#)[Education](#)[Tools](#) 

[status](#)  
[first time visitors](#)  
[introduction](#)  
[spot light](#)  
[searching](#)  
[site map](#)



### REFERENCE

The Tibetan and Himalayan Digital Library is an international community using Web-based technologies to integrate diverse knowledge about Tibet and the Himalayas for free access from around the world.

Serving a wide range of communities, we publish multilingual studies, multimedia learning resources, and creative works concerned with the area's environments, cultures, and histories.



### EDUCATION



### COMMUNITY



### COLLECTIONS



### TOOLS

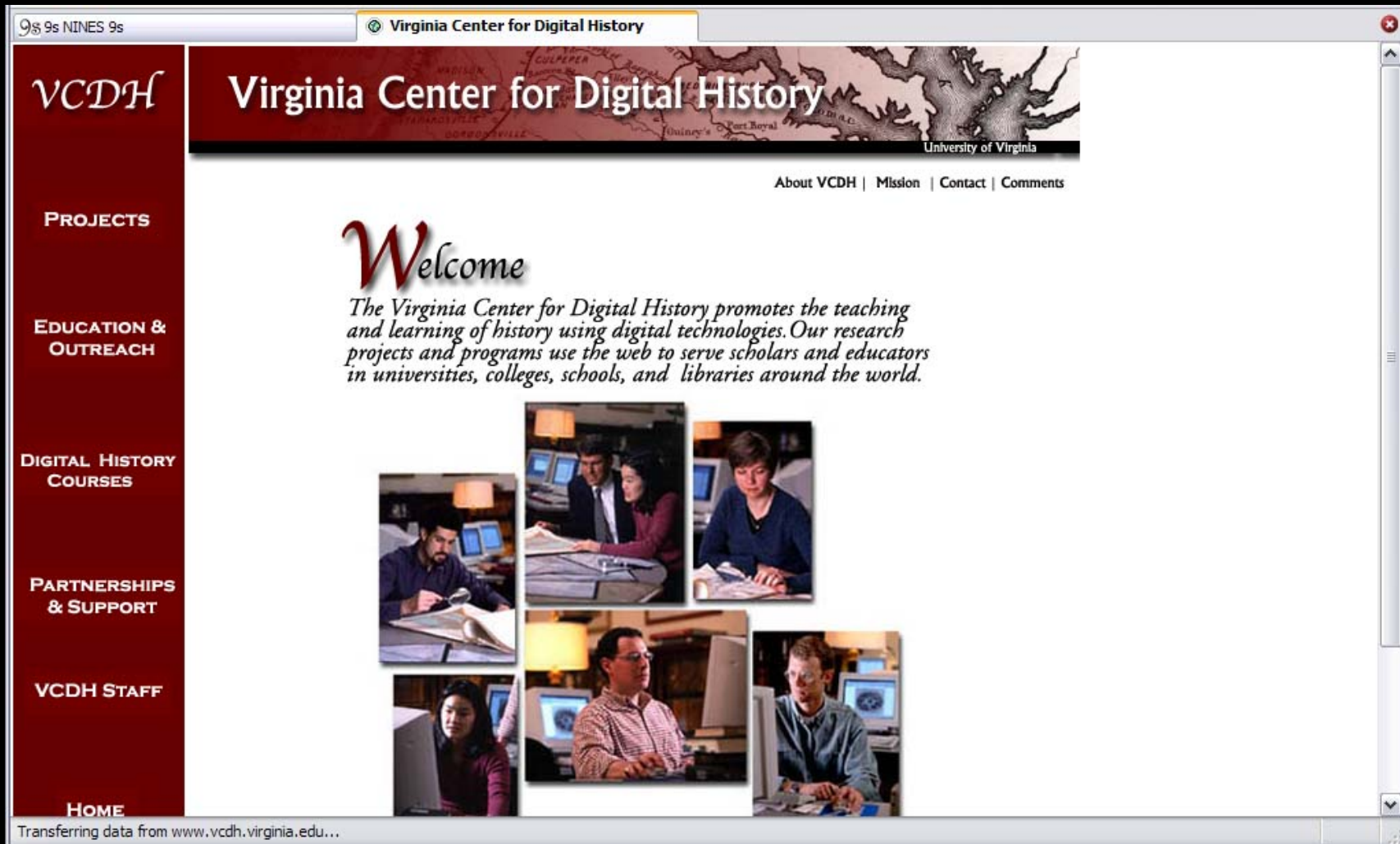
[English](#)[Tibetan \(བོད་ཡིག\)](#)[Nepali \(नेपाली भाषा\)](#)[Chinese \(汉语\)](#)

site supported and managed by the  
UNIVERSITY of VIRGINIA LIBRARY

Transferring data from iris.lib.virginia.edu...

Daigle/Furlough/Staples/Wessell:  
Sustaining Digital Scholarship

UNIVERSITY of VIRGINIA LIBRARY





# IATH

The Institute for Advanced Technology in the Humanities



Michael Satlow, "Inscriptions from the Land of Israel: the Second Temple Period to the Moslem Conquest"

## ▶ ABOUT IATH

As a research unit of the [University of Virginia](#), IATH's goal is to explore and expand the potential of information technology as a tool for humanities research. To that end, we provide our Fellows with consulting, technical support, applications programming, and networked publishing facilities. We also cultivate partnerships and participate in humanities computing initiatives with libraries, publishers, information technology companies, scholarly organizations, and others interested in the intersection of computers and cultural heritage. [To find out who supports our work \(or how you might see our credits page.](#)

## ▶ FELLOWSHIPS

## ▶ OUR RESEARCH

## [IATH NEWS](#)

Google Search of IATH

Intranet Login

# NINES

a networked interface for nineteenth-century electronic scholarship

[About 9s](#) | [People & Projects](#) | [Tools & Interfaces](#) | [Events & Activities](#) | [Join 9s](#)



## ~ nineteenth-century scholarship ~

Over the past ten years a growing body of digital scholarly work has been undertaken, much of it put online, nearly all of it executed without peer review processes, none of it integrated (except by hyperlinking). NINES is a project to found a publishing environment for integrated, peer-reviewed online scholarship centered in nineteenth-century studies, British and American.

## ~ and a scholarly collective ~

NINES believes it is clearly in the interest of scholars to coordinate our work. We know that the migration of scholarship from paper-based to digital platforms and networks, already underway, will only grow apace. Scholars and educators must act on our own behalf if we are to help shape the form and result of this migration. To that end, NINES is promoting the means and a way for excellent work in digital scholarship to be produced, vetted, (eventually) published, and recognized by the discipline. [\[read more\]](#)

a networked interface for nineteenth-century electronic scholarship

Done



# The Library supports digital scholarship through

---

- Creation of digital content
- Services that support its use in teaching and research
- Research and development in digital library tools and applications
- Unique copyright and IP expertise

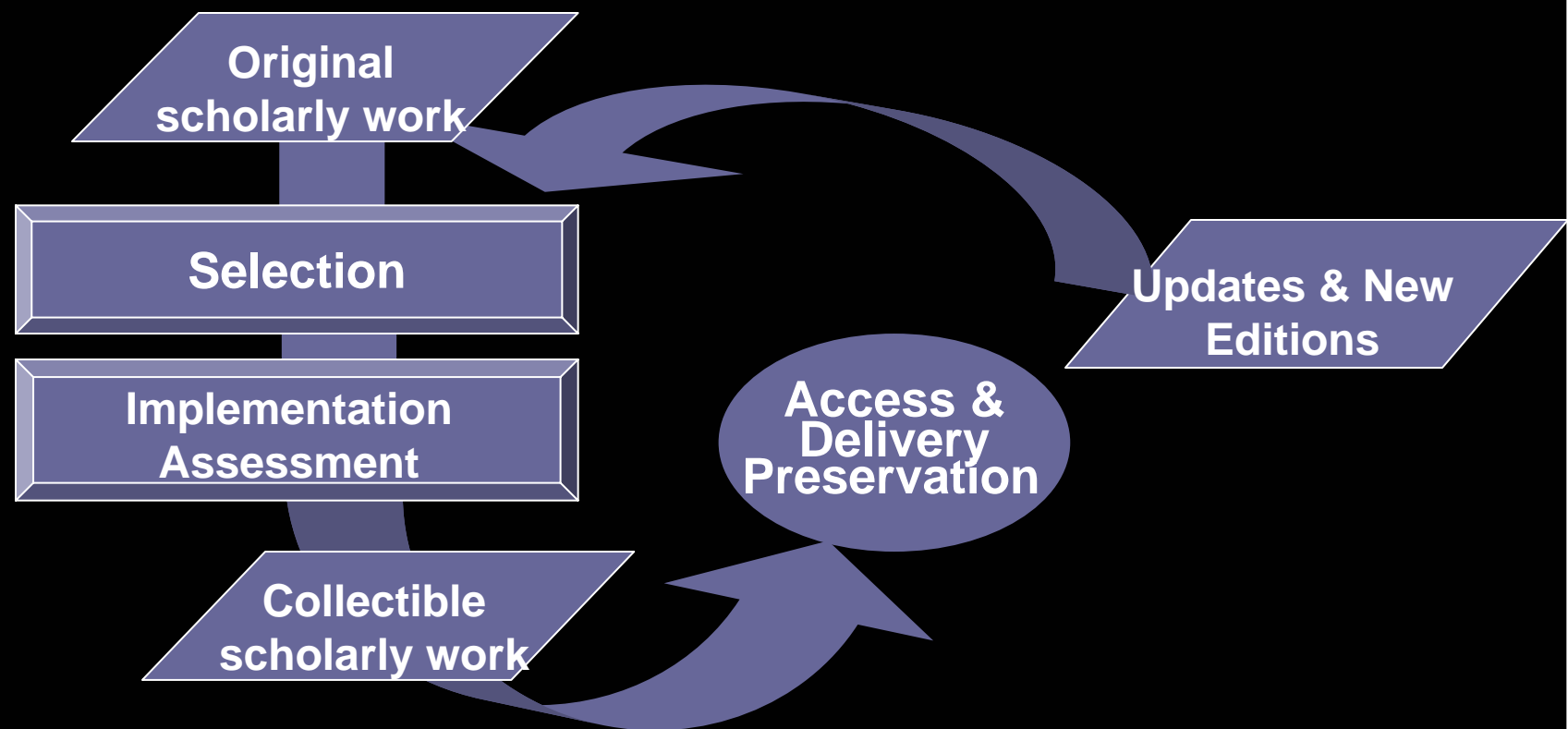
# Collection policy framework

---

- Collect digital projects through a defined selection process
- Manage them with our digital library infrastructure
  - Provide for multiple “views” of the content
  - Collect new “genres” and formats as needed
- Provide open-access to the extent possible
- Develop formal agreements with authors
  - Define the level of effort to sustain the work

# Sustaining Digital Scholarship: UVA's model

---



# Parameters of the Sustaining Digital Scholarship project

---

# Herding cats...

---

- Defining the SDS experiment
- Roles and responsibilities
- Process and workflow
- Assessment objectives
- Establish reasonable timelines

SDS_Timeline.xls							
	A	B	C	D	E	F	G
4	<b>Current SDS Overview Timeline</b>						
5							
6	<b>Week</b>	<b>Rossetti</b>	<b>Owner</b>	<b>THDL</b>	<b>Owner</b>	<b>Library</b>	<b>Owner</b>
7							
8	<b>18-22 July</b>	analyze stylesheet	Bill	--		Create SDS timeline	Bradley
9	<b>25-29 July</b>	Implement & debug stylesheet	Bill	--		Aggregate model meeting THDL	Thorny/Bradley
10						deliver image software to extract metadata	Thorny
11						Disseminator discussion	Thorny
12	<b>01-05 Aug</b>	Objects in proper directory structure	Bill	--		deliver image software to extract metadata	Thorny
13						Aggregate model meeting Rossetti	Thorny
14	<b>08-12 Aug</b>	Image metadata extraction	Duane	create XML Pagebook objects with TIBBIBL catalogue record	Than	Search Interface discussion	Thorny/Bradley
15	<b>15-19 Aug</b>	--		--		Create ingest scripts	Library
16						Pld substitutions	Library
17	<b>22-26 Aug</b>	--		--		2nd Disseminator discussion	Thorny
18	<b>29-02 Sep</b>	Work on agreement stucture	Jerry	Create image objects by creating new tiff derivatives	Than	Work on agreement stucture	Library/Bradley
19	<b>05-09 Sep</b>			--		--	--
20	<b>12-16 Sep</b>			Objects in proper directory structure	Than	Ross begin creating FEDORA 2.0 instance	Ross
21	<b>19-23 Sep</b>			Work on agreement stucture	David / Than	Work on agreement stucture	Library/Bradley
22	<b>26-30 Sep</b>	Begin building aggregate content models	Beth/Jerry	Begin Building aggregate content models	Than	Begin building aggregate content models	Ross/Thorny
23	<b>03-07 Oct</b>						
24	<b>10-14 Oct</b>						
25	<b>17-21 Oct</b>						
26	<b>24-28 Oct</b>						



# Sustaining Digital Scholarship: an experiment

---

- Test our policy frameworks
- “Ingest” subsets of Rossetti and THDL into our digital library system
- Assume a three-year commitment
- Aim to complete first round by summer 2006
- Responsibilities somewhat negotiable

# SDS Participants

---

- University Library staff (sr. administrators, lawyer, managers, subject librarians, technologists)
- Rossetti Archive and NINES staff (faculty member, post-doc, technologists, project managers)
- Tibetan-Himalayan Digital Library staff (faculty member, project managers, technologists)

# Author responsibilities

---

- Prepare “camera ready copy”
- All editorial review, including peer review
- Meet intellectual property standards
- Define an “edition” of the work to be collected-- “snapshot” in time.

# Library responsibilities

---

- Define and develop the infrastructure based on Fedora
- Provide technical assistance and guidance in our specifications
- Provide legal/policy assistance as needed

# Collection and aggregation of digital scholarly projects

---

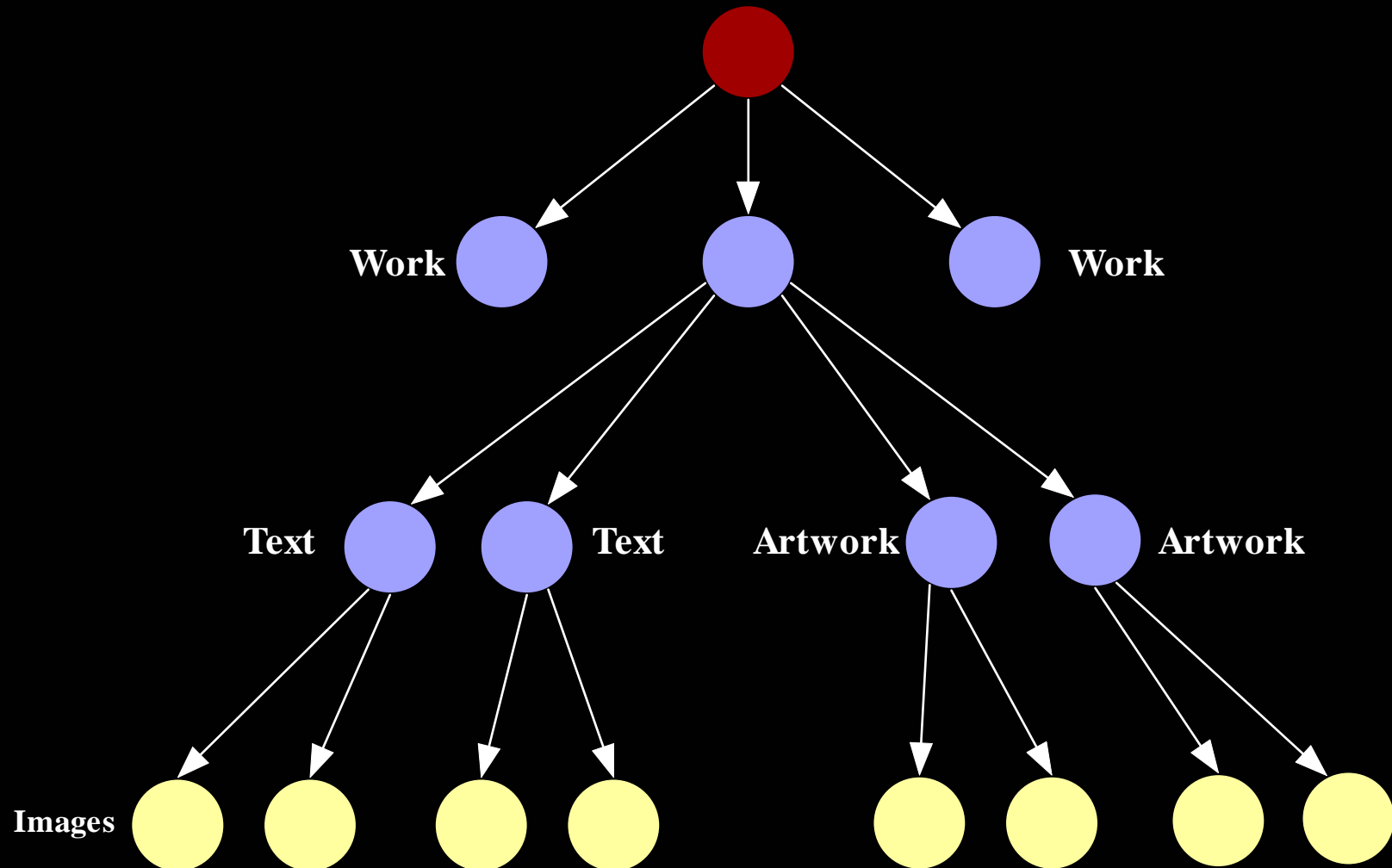
# The Rossetti Project

---

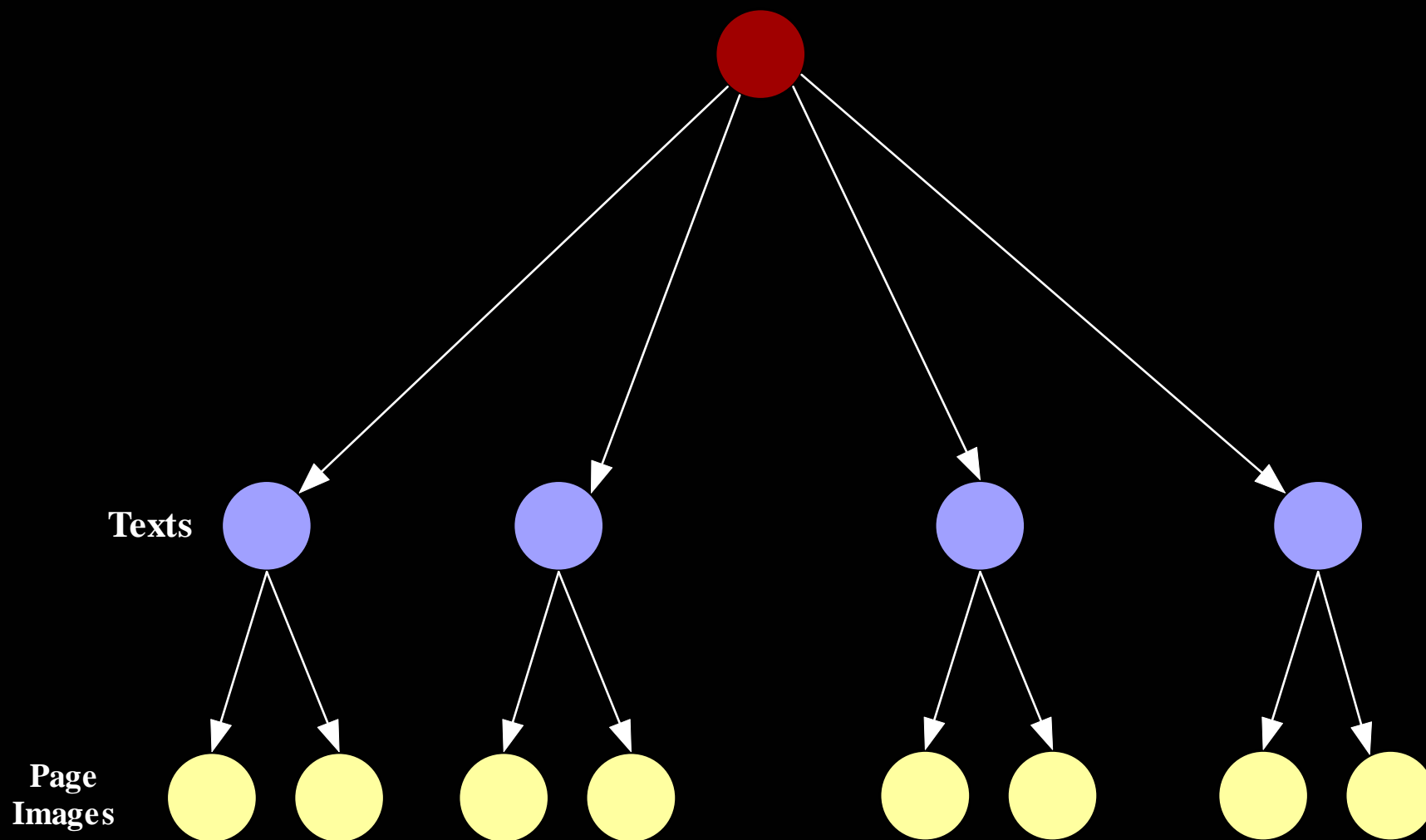
- Started in 1992 by Jerome McGann
- Intended to be a model for digital critical archives that combine critical and facsimile editions
- An edition of all of Rossetti's written and graphic works
- ~5,000 XML files, ~5,000 image files
- Involves many different rights agreements



# The Rossetti Archive



# Modern English Collection



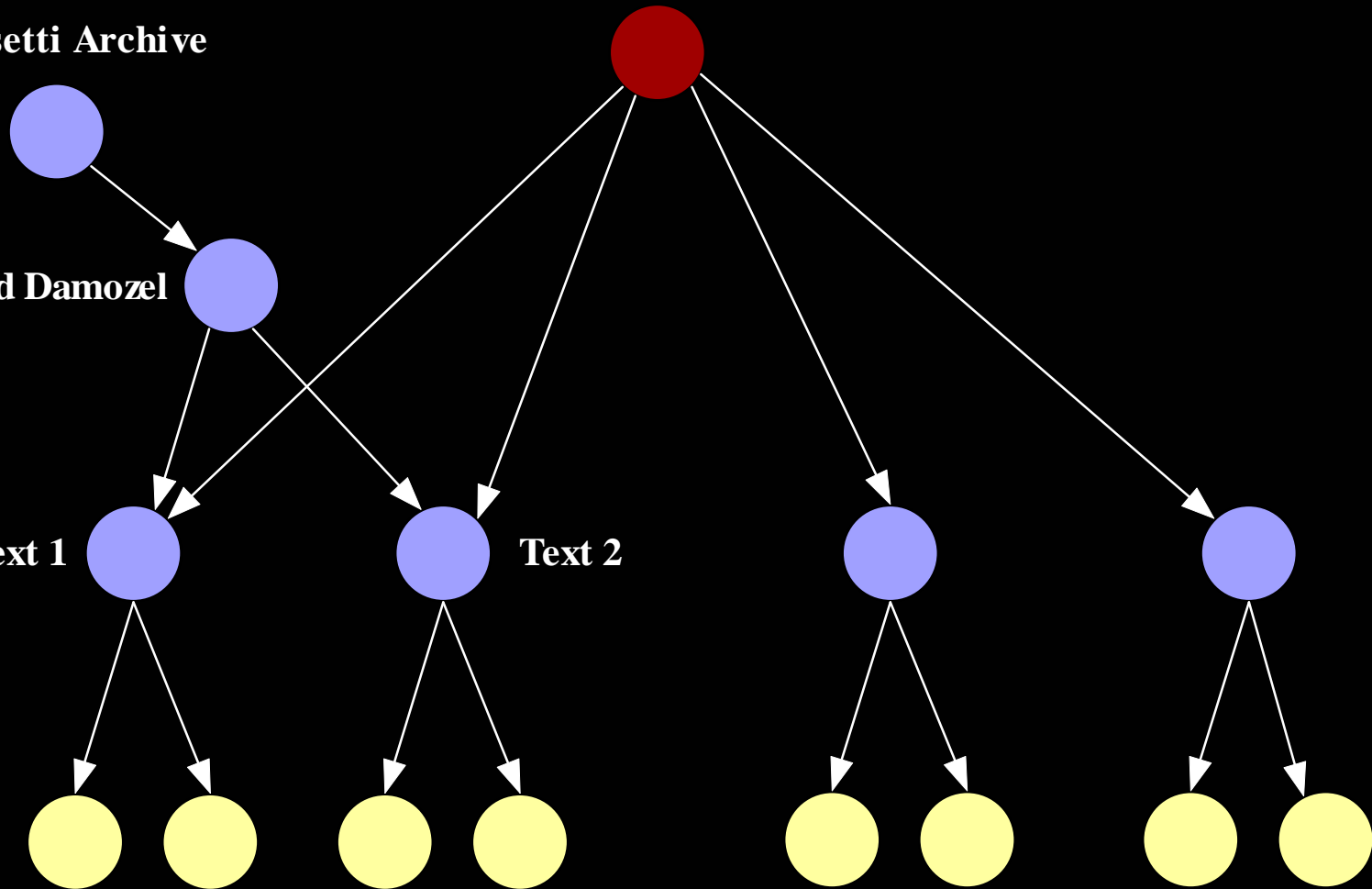
## Modern English Collection

The Rossetti Archive

The Blessed Damozel

Text 1

Text 2



# Assumptions

---

- The primary assumption is that we are bringing the projects into the Library's management stream
- They often include large numbers of digital surrogates of traditional resources
- Increasingly, they will include born digital primary resources

### General Image Object

Persistent ID (PID)
Web-default
Web-image
Admin
Metadata
thumbnail image file
med res. image file
high res. image file
max res. image file

### General Image Behavior Mechanism

get-thumbnail-sized-image
get-med-sized-image
get-high-res-image
get-max-sized-image

### Web-Image Behavior Description

get-thumbnail-sized-image
get-med-sized-image
get-high-res-image
get-max-sized-image

### MrSID Image Object

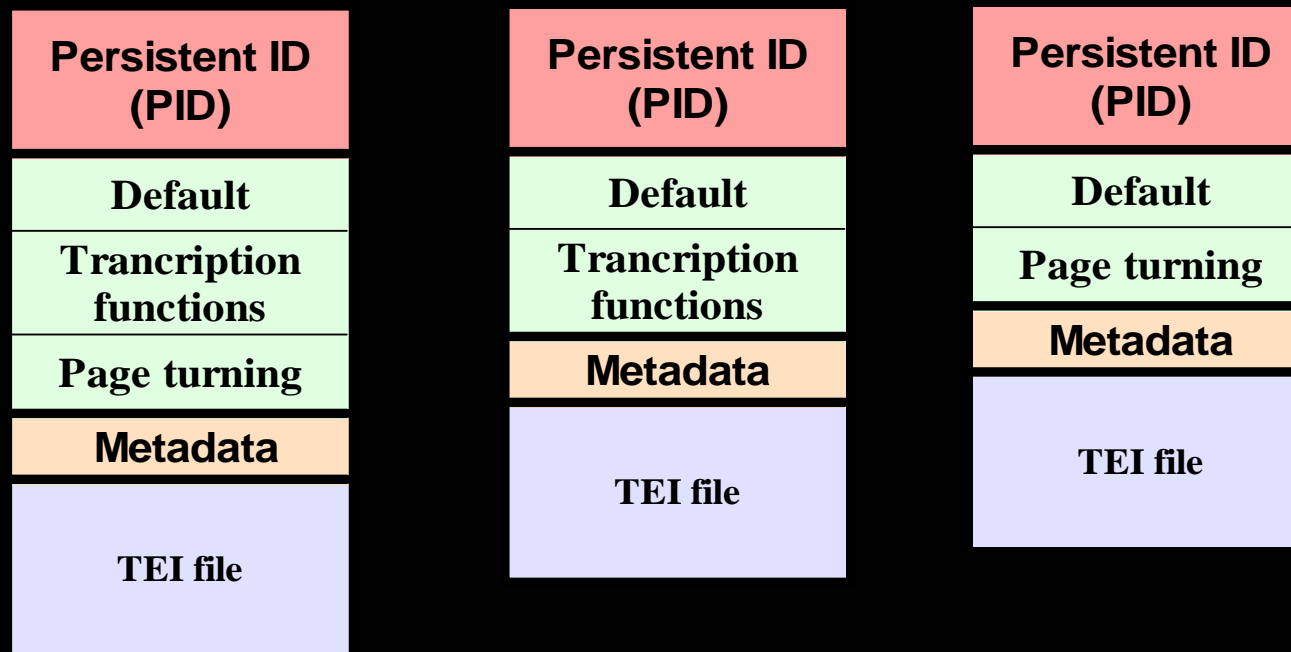
Persistent ID (PID)
Web-default
Web-image
Admin
Metadata
MrSID-encoded image file

### MrSID Image Behavior Mechanism

get-smallest-MrSID-size
get-midrange-MrSID-size
get-high-res-MrSID-size
get-max-MrSID-size

# Text Collections: three models

---





### Library Text Object

Persistent ID (PID)
Default
Trancription functions
Page turning
Metadata
Library TEI text

### Library trancription Behavior Mechanism

get-page
get-chapter
get-header

### Rossetti Text Object

Persistent ID (PID)
Default
Trancription functions
Page turning
Rossetti functions
Metadata
Rossetti DTD text

### Transcription Behavior Description

get-page
get-chapter
get-header

### Rossetti trancriptoin Behavior Mechanism

get-page
get-chapter
get-header

# Levels of Collecting

	Metadata	Files	Content Relationships	Original Delivery
1	Fully Compliant	Fully Compliant	Exact	Exact
2	Fully Compliant	Fully Compliant	Exact	Approximate
3	Fully Compliant	Fully Compliant	Exact	Representative
4	Fully Compliant	Fully Compliant	Exact	
5	Fully Compliant	Fully Compliant		
6	Fully Compliant	Binary Files Only		
7	Fully Compliant			

# The Future

---

- Collecting project that are created in the wild is not scalable
- An authoring environment that supports the creation of very complex projects
- Projects created into an institutional repository directly
- Move to digital library upon selection

# Legal and policy implications of collecting digital scholarly projects

---

# Legal Issues and Opportunities When Libraries “Publish”

---

- Libraries are not traditional publishers (peer review, editing, promotion, advertising)
- But publishers are not in the business of preservation, and few are willing to invest in technologies necessary to support complex digital scholarly works
- Libraries may be positioned to exploit Fair Use in innovative ways

# Structuring the Relationship

---

- As would be the case in the context of a formal publication, a written agreement is important to the scholar and the library
- We analogize these agreements to “deposit agreements” used in our special collections
- Agreements must be constructed around agreed project standards



# Agreements Must Address:

---

- Levels of collecting
- Access assumptions
- Duration of the commitment
- Financial issues
- Modifications of projects after collection
- Standards of preservation AND
- Copyright constraints

# Copyright Constraints

---

- Increasingly well-recognized as fundamental problems in digital scholarship
- Digital tools are facilitating a creative explosion in humanities scholarship through use of images, sound, music, film and other materials – but

# Copyright Restricts Use Of:

---

- Literary Works (including articles, books, software with written documentation)
- Musical Works (including score and lyrics)
- Dramatic Works (including any accompanying music)
- Choreography and Pantomimes
- Pictorial, Graphic and Sculptural Works
- Motion Pictures and other Audiovisual Works
- Sound Recordings
- Architectural Works (including plans and drawings)

# Copyright Law is Reeling From

---

- From the impact of the web and digital technologies
- From an aggressive property rights culture that has been enabled through recent legislation (DMCA, Sonny Bono, TEACH)
- From the political impact of widespread infringement activities by kids
- From legally unresolved questions about the use of photographs, images, and sound recordings

# Because of Open Access Decision SDS Requires

---

That accepted projects meet legal requirements for web-based public access which means they are either:

- Self-created
- Public domain
- Used with permission, or
- Fair Use

# Permissions

---

- Museums and other cultural institutions see these as source of revenue
- Short-term and use-limited are the norm
- Contracts can limit what the law otherwise protects
- Maddening processes typical

# Publishers and Permissions

---

“Complete this form and deliver it to your acquisitions editor as soon as you have a complete MS that has been accepted for publication. We need this information to ensure that copyrighted and archival materials are properly acknowledged in your book. ...

Have any portions of your MS or its accompanying art program been created or written by someone other than yourself? ...

***Photocopies of ALL necessary letters of permission must accompany the final version of your MS.”***

# Fair Use

---

- **Purpose and character** of the use including whether the use is for educational vs. commercial purposes
- **The nature of the copyrighted work** – is it a factual or creative work
- **The amount and substantiality** of the portion to be used in relation to the work as a whole
- **The effect or impact** of the use upon the potential market for or value of the work.



# Fair Use is a flexible, fact-specific standard

---

- No single factor in the test is decisive.
- The fact that a use might be commercial, or that it might impair the marketability of a work, does NOT invalidate the use as fair (Cambell v Acuff Rose Music – 1994)
- That a use is educational does NOT guarantee that it IS fair.
- Fair Use is an affirmative defense, which means that some lawyers hate it.

# It IS Easier to Defend A Use As Fair...

---

- When no sale of materials is involved
- When a strong public interest argument is available
- When the uses are research, academic, educational
- In other words, when an academic library is involved, which is one among several important motivators to our SDS work

# Special issues With Images, Photographs, and Music

---

- Viewed by publishers as the thorniest areas
- Object owners have rights independent of copyright
- New case law is helpful as to thumbnails and images of 2-D objects
- Perplexing case law out of 6<sup>th</sup> Circuit on use of even tiny portions of sound recordings

# Changing Law and Limited Permissions

---

- Mean both sides need to be prepared to accommodate potential removal or shielding of materials or portions of sites
- Institutional support for IP stance of the library is critical so that projects meeting those standards are defended by university legal counsel

# Broader IP Issues in Academia - Intellectual Property Policies

---

## ■ The UVA Policy:

### **Employee Ownership**

... The University cedes copyright ownership to the author(s) of scholarly and academic works (such as journal articles, books and papers) created by academic and research faculty who use generally available University resources. **However, the University asserts its right of copyright ownership if significant University resources (including sponsor-provided funds) are used in the creation of such works, and: (a) the work generates royalty payments; or (b) the work is of commercial value that can be realized by University marketing efforts.**

# “Significant Resources”

---

## ■ “Significant” University Resources:

...The use of a computer in a faculty office, incidental supplies and occasional use of University personnel or shared facilities would typically not be considered significant use. **In contrast, utilization of University laboratories or special instrumentation, dedicated assistance by University employees, special financial assistance or extensive use of shared facilities would constitute significant use.**

# Science vs. Humanities

---

- Significant resources focus of current policies typically stem from an interest in commercially exploiting scientific inventions made possible over 20 years ago by Congress.
- Impact on digital humanities scholarship which also benefits from special resources, is not clear.

# Commercialization by Universities

---

- University patent policies are less implicated in humanities context
- But software treated under most policies as reportable and subject to commercial exploitation
- Can have negative impact on collaborative software developed for humanities projects and open source decisions



# Digital Scholarly Communities

---

- Need to develop common strategies and proposals to modify institutional IP policies to better address and protect interests of digital scholars
- Open source and dual license strategies may provide some solutions.

# Some elephants in the room

---

- Macro-economics of digital scholarship
  - Shifting subsidies for scholarly communication
- Mechanics of preservation
  - The first great migration
- Institutional structures
  - Roles within the academy
  - Peer review and promotion/tenure
  - Peer review and library selection

# Contacts

---

- Mike Furlough: [furlough@virginia.edu](mailto:furlough@virginia.edu)
- Bradley Daigle: [bradley@virginia.edu](mailto:bradley@virginia.edu)
- Thornton Staples: [thorny@virginia.edu](mailto:thorny@virginia.edu)
- Madelyn Wessel: [wessel@virginia.edu](mailto:wessel@virginia.edu)
  
- SDS Project:  
<http://www.lib.virginia.edu/digital/sds>