

Preparing Future Digital Curators



PART I: A Summary Report on the Digital Curation Curriculum (DigCCurr) Project
Christopher (Cal) Lee and Carolyn Hank

PART II: Four Perspectives on Applying Academic Understanding in a Practice Setting
John Blythe, Lisa Gregory, Samantha Guss, and Jennifer Mantooth

School of Information and Library Science
University of North Carolina at Chapel Hill
<http://www.ils.unc.edu/digccurr/>

LAUNC-CH Research Forum, Chapel Hill, NC: 21 May 2008

PART I:

A Summary Report on the Digital
Curation Curriculum (DigCCurr)
Project

Professional Evolution

- Advances in management, preservation & dissemination of digital resources
- Many streams of activity (e.g. computer scientists, archivists, records managers, librarians, scientific data engineers, museum curators, organizational IT staff)
- Increasing recognition in past decade of common challenges & opportunities
- Recent adoption of term “digital curation” – pulling together many previously distinct research communities

Digital Curation

- “The active management and preservation of digital resources over the life-cycle of scholarly and scientific interest, and over time for current and future generations of users.”
- Widely used by scientists & those responsible for data sets
- Seen by many as more inclusive – in disciplinary scope & coverage of lifecycle -- than “digital preservation”

*Digital Curation Centre. “What is Digital Curation?” <http://www.dcc.ac.uk/about/what/>

Education & Professional Development

- Many valuable components of a digital curation curriculum
 - Individual courses & components within graduate programs (most in LIS programs, but also e.g., computer science, business, public policy, history)
 - Professional workshops (usually 1-5 days)
- Training in specific disciplines generally doesn't address issues such as long-term access, integrity, contextual information
- LIS students would benefit from more understanding of specific digital environments & resource types

DigCCurr Project

- IMLS Grant # RE-05-06-0044
- Collaboration of School of Information & Library Science (SILS), University of North Carolina at Chapel Hill (UNC-CH) & U.S. National Archives & Records Administration (NARA)
- Runs July 1, 2006 – June 30, 2009

DigCCurr Components/Goals

Curriculum: To prepare students for digital curation with wide variety of organizations, contexts & types of resources:

- Graduate-level curricular framework
- Course modules
- Experiential components

Two International Symposia:

- First was held April 18-20, 2007 in Chapel Hill - <http://ils.unc.edu/digccurr2007/>
- Second to take place early April 2009 (near end of project)

Carolina Digital Curation Fellowship program

Practical Field Experience

- Should engage in at least two different field experiences in different institutional contexts
- Should involve some hands-on work with digital objects with actual consequences, rather than just conceptual or policy work
- Importance of partnering with sites that already actively engage in digital curation

Carolina Digital Curation Fellows

- 5 Digital Curation Fellows pursuing degrees at SILS - began fall 2007
- UNC partners providing practical experience opportunities: ibiblio, ITS, Odum Institute, University Library
- Specialized introductory seminar held Fall 2007
- Overseeing & learning from their practical engagement work
- Advising on course selection
- Plan for future practical engagement opportunities

Matrix of Digital Curation Knowledge & Competencies

- Tool for thinking about, planning for, identifying & organizing curriculum
- Each unit of curriculum content can address one or more dimensions
- Helping to address fundamental issue: All digital curation students should get some aspects of curriculum, but other aspects only necessary for students planning to work in particular types of places or jobs (i.e. balancing core vs. specialized knowledge)



DigCCurr Progress Report: Development of a Graduate-Level Digital Curation Curriculum

Christopher A. Lee, Carolyn Hank, Helen R. Tibbo

The Digital Curation Curriculum (DigCCurr) project (IMLS RE-05-06-0044) is developing a graduate-level curricular framework, course modules & experiential components to prepare students for digital curation professions in a variety of information environments. The two tables below summarize the contents of our two main conceptual products & tools for organizing our work. More detailed elaboration of the components & sources from which we've drawn will be available at the DigCCurr web site.

DigCCurr Research Activities:

- Extensive review & analysis of literature
- Interviews with 17 expert advisory board members
- Survey of participants in major international conference in April 2007
- Planned wider-scale survey of digital curation needs & expectations
- Analysis of syllabi
- Analysis of job postings

Draft Matrix of Topics for Digital Curation Curriculum

I. Mandates, Values & Principles			
1.1. Ethics			
1.2. Core Digital Curation Principles & Values	<ul style="list-style-type: none"> • Abstraction • Accountability • Adaptability & Robustness • Authority • Automating & Informating of Tasks 	<ul style="list-style-type: none"> • Collection • Content • Continuum & Lifecycle Orientations • Critical Inquiry • Diversity 	<ul style="list-style-type: none"> • Evidence • Long Term • Openness • Interoperability • Provenance & Chain of Custody
1.3. Legal Requirements			
1.4. Standards			
1.5. Interoperability & Sustainability Requirements			
2. Functions & Skills			
3. Professional, Disciplinary, Institutional, or Organizational Context			
3.1. Professional Context			
3.1.1. History of Professional Activities	<ul style="list-style-type: none"> • Careful preparation of physical media • Hardware and software interoperability 	<ul style="list-style-type: none"> • Long-term management of institutional archives & personal papers • Social science data archives 	<ul style="list-style-type: none"> • Natural & physical science data archives (e.g. north & south science) • Management & provision of access to digital library collections
3.1.2. Professional Development			
3.2. Disciplinary Context			
3.3. Institutional/Organizational Context			
3.3.1. Characteristics of information & record creating environments			
4. Type of Resource			
4.1. Level of aggregation	• Component	• Object	• Collection
4.2. Level of abstraction	• Work	• Expression	• Manifestation
4.3. Medium - optical, magnetic			• Item
4.4. Format			
4.5. Genre			
5. Prerequisite Knowledge			
5.1. Terminology			
5.2. Characteristics of Technologies			
5.2.1. Definitions of Technology			
5.2.2. History & evolution of ICT			
5.2.2.1. General Patterns & Lessons			
5.2.2.2. Specific Developments & Orientations of Technology	<ul style="list-style-type: none"> • From purpose-built computers to mainframes, time sharing, minicomputers, personal computers, client-server, mobile devices 	<ul style="list-style-type: none"> • Dominant storage technologies from given era 	<ul style="list-style-type: none"> • History of the Internet
5.2.3. Essential Characteristics & Elements of Current & Emerging ICT Landscapes	<ul style="list-style-type: none"> • Abstraction and visualization (e.g. emulation, virtual machines, virtual name spaces) • Compression • Data structures and data types • Databases • Encryption • File formats • File Systems 	<ul style="list-style-type: none"> • Identifiers for digital objects <ul style="list-style-type: none"> ◦ Commonly used identifiers (e.g. file names, URIs) ◦ Promising approaches for persistent identifiers • Networks • Operating Systems • Programming • Differences between local and global identifiers • Images (vector, raster) • Markup languages (including XML) • Models and modeling 	<ul style="list-style-type: none"> • Entity-relationship and object-oriented approaches • Types of models (e.g. functional, data flow) • Security • Storage media and systems - operation and characteristics (magnetic and optical) • Standards • Roles of standards - both within the computer industry and for digital curation • Types of standards • Standards development process
6. Transition Point in Information Continuum			
6. Pre-Creation Design and Planning	• Transfer to Archives	• Transfer Copies or Surrogates to Secondary Use Environment	
6. Creation	• Archives (Preservation Environment)	• Secondary Use Environment	
6. Primary Use Environment (Active Use)			

Draft High-Level Categories of Digital Curation Functions

Category	High-Level Sub-Functions	Key Tasks & Activities
1. Systems Engineering & Development	<ul style="list-style-type: none"> • Analysis • Design & development • Database analysis • Design & development • User requirements & usability studies 	<ul style="list-style-type: none"> • System design & specification • Design • Software design • User analysis
2. Production	<ul style="list-style-type: none"> • Design & development • User requirements & usability studies 	<ul style="list-style-type: none"> • System design & specification • Design • Software design • User analysis
3. Selection, Appraisal & Disposition	<ul style="list-style-type: none"> • Selection • Appraisal & disposition • Disposition & archiving of collections 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
4. Identifying, Locating & Harvesting	<ul style="list-style-type: none"> • Identifying records information for preservation, access & reuse • Identifying records for access & reuse 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
5. Transfer	<ul style="list-style-type: none"> • Transfer planning, preparation & execution • Transfer of records information • Transfer of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
6. Ingest	<ul style="list-style-type: none"> • Ingest planning, preparation & execution • Ingest of records information • Ingest of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
7. Data Management	<ul style="list-style-type: none"> • Data management planning, preparation & execution • Data management of records information • Data management of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
8. Description, Organization & Intellectual Control	<ul style="list-style-type: none"> • Description, organization & intellectual control planning, preparation & execution • Description, organization & intellectual control of records information • Description, organization & intellectual control of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
9. Archival Storage	<ul style="list-style-type: none"> • Archival storage planning, preparation & execution • Archival storage of records information • Archival storage of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
10. Management	<ul style="list-style-type: none"> • Management planning, preparation & execution • Management of records information • Management of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
11. Administration	<ul style="list-style-type: none"> • Administration planning, preparation & execution • Administration of records information • Administration of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
12. Preservation Planning & Implementation	<ul style="list-style-type: none"> • Preservation planning, preparation & execution • Preservation of records information • Preservation of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
13. Access	<ul style="list-style-type: none"> • Access planning, preparation & execution • Access of records information • Access of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
14. Use, Reuse & Adding Value to Accessed Information	<ul style="list-style-type: none"> • Use, reuse & adding value to accessed information planning, preparation & execution • Use, reuse & adding value to accessed information of records information • Use, reuse & adding value to accessed information of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
15. Reference & User Support Services	<ul style="list-style-type: none"> • Reference & user support services planning, preparation & execution • Reference & user support services of records information • Reference & user support services of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
16. Common Services	<ul style="list-style-type: none"> • Common services planning, preparation & execution • Common services of records information • Common services of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
17. Destruction & Removal	<ul style="list-style-type: none"> • Destruction & removal planning, preparation & execution • Destruction & removal of records information • Destruction & removal of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
18. Collaboration, Coordination & Constructing with External Actors	<ul style="list-style-type: none"> • Collaboration, coordination & constructing with external actors planning, preparation & execution • Collaboration, coordination & constructing with external actors of records information • Collaboration, coordination & constructing with external actors of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
19. Advocacy & Outreach	<ul style="list-style-type: none"> • Advocacy & outreach planning, preparation & execution • Advocacy & outreach of records information • Advocacy & outreach of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
20. Analysis & Evaluation of Producer Information Environment	<ul style="list-style-type: none"> • Analysis & evaluation of producer information environment planning, preparation & execution • Analysis & evaluation of producer information environment of records information • Analysis & evaluation of producer information environment of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
21. Analysis & Characterization of Digital Objects/Packages	<ul style="list-style-type: none"> • Analysis & characterization of digital objects/packages planning, preparation & execution • Analysis & characterization of digital objects/packages of records information • Analysis & characterization of digital objects/packages of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
22. Packaging & Licensing of Records	<ul style="list-style-type: none"> • Packaging & licensing of records planning, preparation & execution • Packaging & licensing of records of records information • Packaging & licensing of records of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
23. Transformation of Digital Objects/Packages	<ul style="list-style-type: none"> • Transformation of digital objects/packages planning, preparation & execution • Transformation of digital objects/packages of records information • Transformation of digital objects/packages of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
24. Packaging & Licensing of Records	<ul style="list-style-type: none"> • Packaging & licensing of records planning, preparation & execution • Packaging & licensing of records of records information • Packaging & licensing of records of records information 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
Meta-Level Functions (to be applied to one or more of the above)		
25. Analysis & Documentation of Curation Functions	<ul style="list-style-type: none"> • Analysis & documentation of curation functions 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
26. Evaluation & Audit of Curation Functions	<ul style="list-style-type: none"> • Evaluation & audit of curation functions 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
27. Research & Development to Support Curation Functions	<ul style="list-style-type: none"> • Research & development to support curation functions 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies
28. Education & Sharing of Expertise or Guidance on Curation Functions	<ul style="list-style-type: none"> • Education & sharing of expertise or guidance on curation functions 	<ul style="list-style-type: none"> • Select to reach • Select records information resources • Create selection, appraisal or disposition studies



PART II

Four Perspectives on Applying
Academic Understanding in a
Practice Setting

**Current Status of
DocSouth CD Migration:**
A Report from the
Carolina Digital Library & Archives

John Blythe

DOCUMENTING the *American South*



PRIMARY RESOURCES FOR THE STUDY OF SOUTHERN HISTORY, LITERATURE, AND CULTURE

Documenting the American South (DocSouth) is a digital publishing initiative that provides Internet access to texts, images, and audio files related to southern history, literature, and culture. Currently DocSouth includes ten thematic collections of books, diaries, posters, artifacts, letters, oral history interviews, and songs.

- The Church in the Southern Black Community
- The Colonial and State Records of North Carolina
- The First Century of the First State University
- First-Person Narratives of the American South
- Library of Southern Literature
- North American Slave Narratives
- The North Carolina Experience
- North Carolinians and the Great War
- Oral Histories of the American South
- The Southern Homefront, 1861-1865
- True and Candid Compositions: The Lives and Writings of Antebellum Students at the University of North Carolina

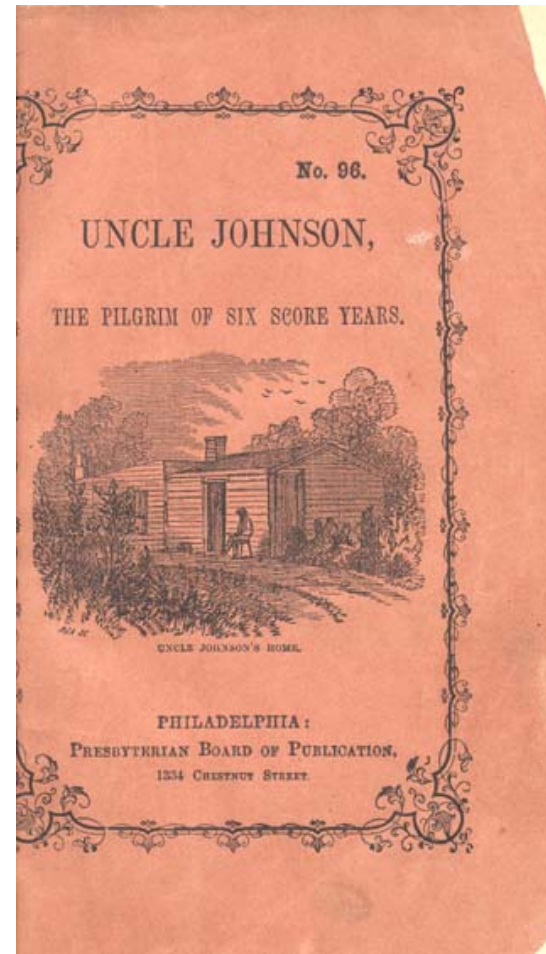
FROM CD to DARK ARCHIVE

- MF Digital Scribe 9000 allows data extraction from two CDs simultaneously
- 1,642 CD's
- 44,787 files
- 3,296 folders
- 903 GB
- Approx 100 hours



darkark/Imported Data/Discs 11-49b/Disc16/Uncle Johnson, the Pilgrim of Six Score Years/unclejcv.jpg

- *Discs 11-49b* = group of discs extracted during one session
- *Disc 16* = individual disc. Named manually
- *Uncle Johnson, the Pilgrim of Six Score Years* = name of folder on CD
- *unclejcv.jpg* = filename



<http://docsouth.unc.edu/neh/foster/cover.html>

Gustavus L. Foster (Gustavus Lemuel), 1818-1876
Uncle Johnson, the Pilgrim of Six Score Years.
Philadelphia: Presbyterian Publication Committee, 186-?.

[Next illustration](#)

[Cover Image]

[Cover Page](#)

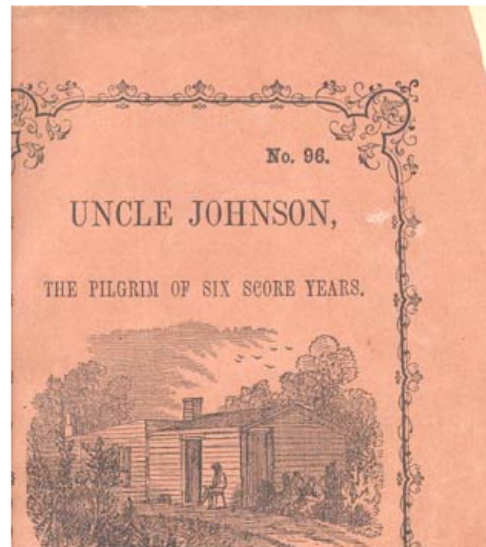
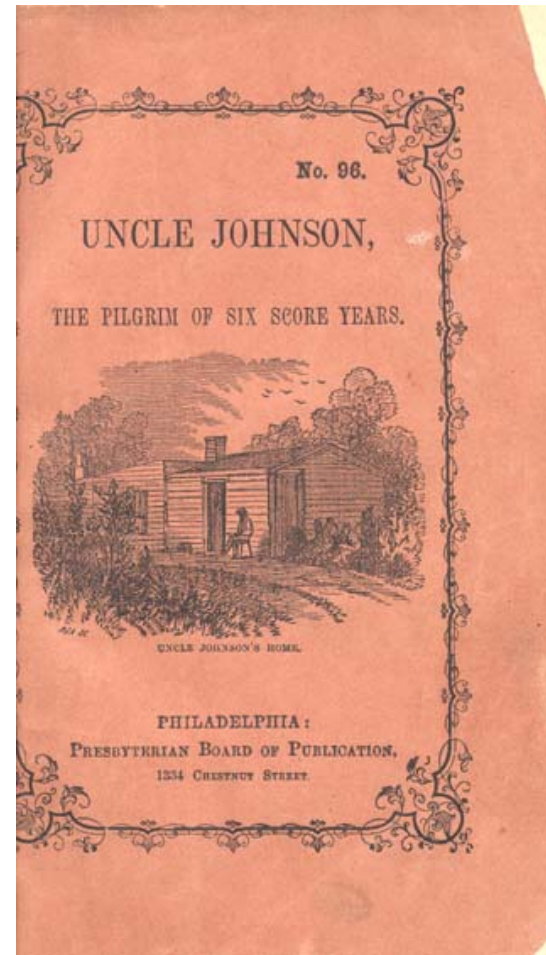


Image cataloged in DocSouth database

- Source publication and each image assigned *Item_ID*
- *Item_ID* for *Uncle Johnson, the Pilgrim... = 39*
- *Item_ID* for cover image = 53698



The Ideal: Programming Magic

- Script matches file name in dark archive with *item_id* in DocSouth database
- Script rebuilds dark archive directory to match database structure.

The Reality: Auto-matching is not perfect

- 16,847 file names matched and assigned appropriate *item_id*
- 1,257 file names matched more than one *item_id*
- 14,800 file names did not match

Why no match?

- *small.jpg* (a thumbnail) not in database
- Additional info added to end of file name (i.e. - *thumb, -1, -50, -75, -150, _100, _150, _at_75, _at_150*)
- Slight variations in file names used on website (database) and archived on CD (dark archive)
- Typos in file names - mistyped letter, extra space, extra letter (i.e. *circltp.jpg* becomes *circlrtp.jpg*)
- Images not in database

More examples

- Variance in one number between file name on website (database) and on CD (dark archive), i.e. *hicks23.jpg* in database is *hicks22.jpg* in dark archive
- Files have two different names: *unclecv.jpg* in dark archive is *fostecv.jpg* in database

Auto-matching: Take 2

- Establish Parent_ID by searching *Item* table for name of publication
- Using Parent_ID search *Item_child* for range of *Item_IDs* that belong to Parent
- Use *Item_ID* of children to search *Illustration_item* for specific file name associated with that ID.

The Results: 550 matches

The Reasons

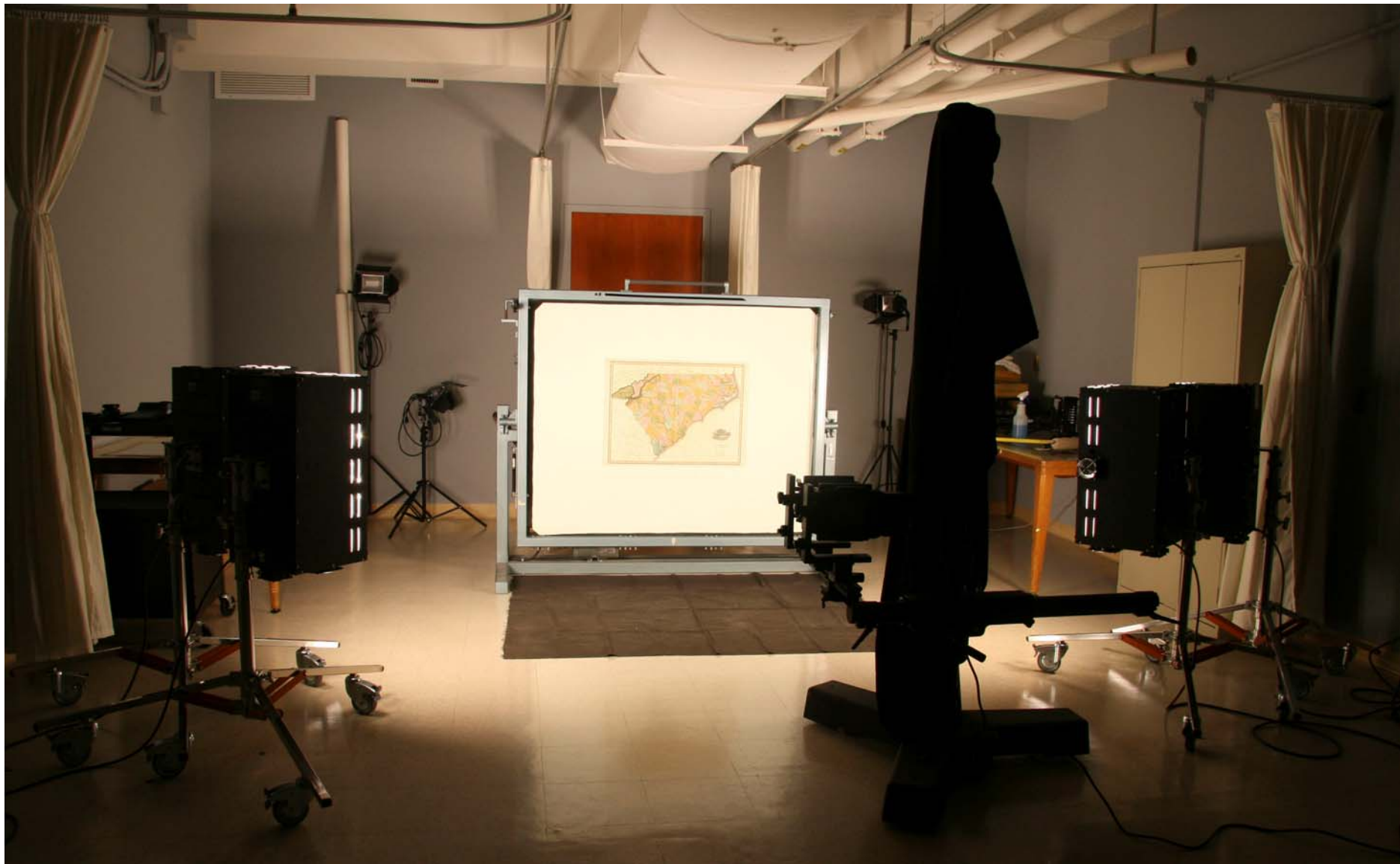
- Variations in names assigned to folders in dark archive (I.e. *Uncle Johnson, the Pilgrim...I*) and names listed in *Item* table of database
- Slight variation in punctuation

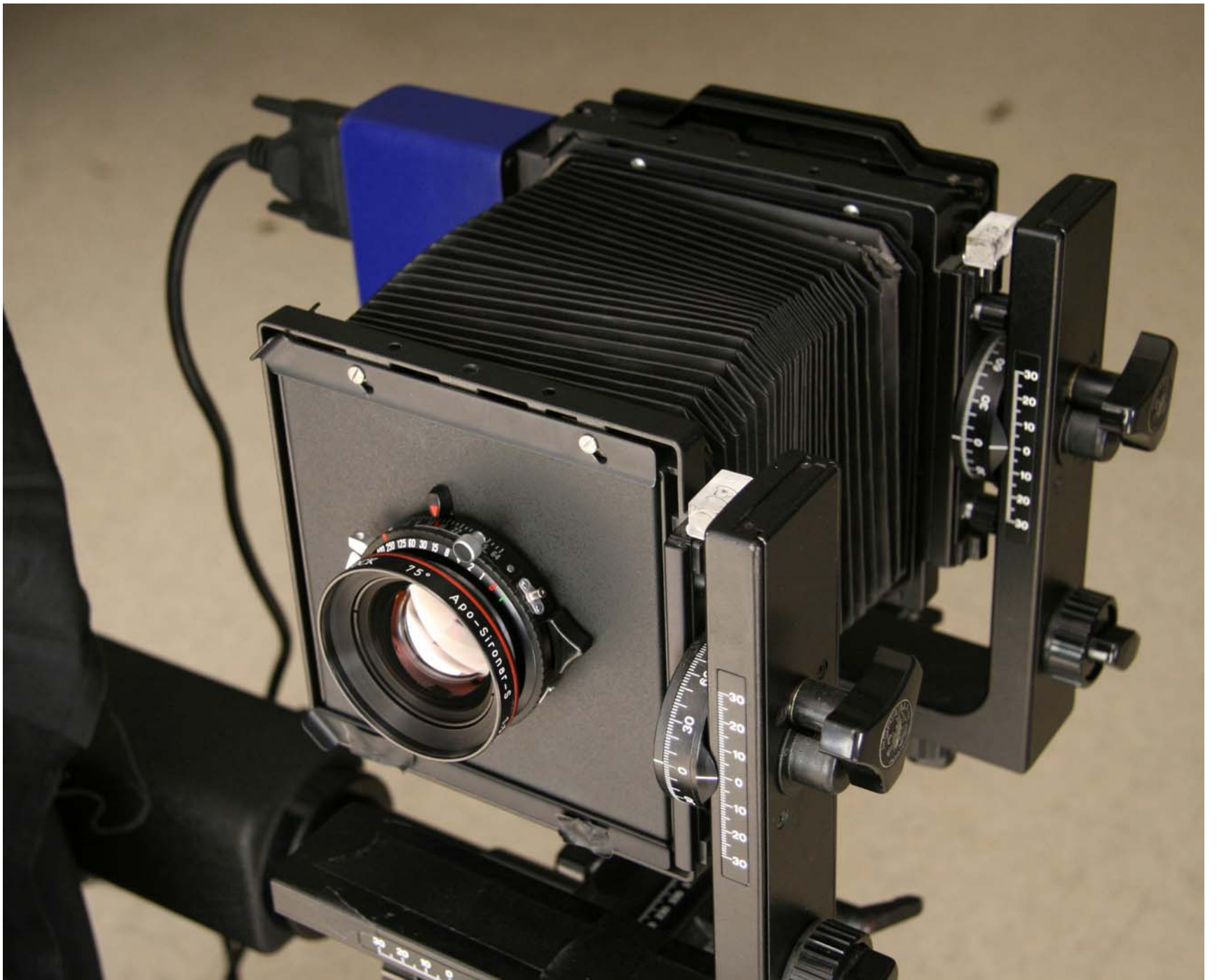
Next Steps

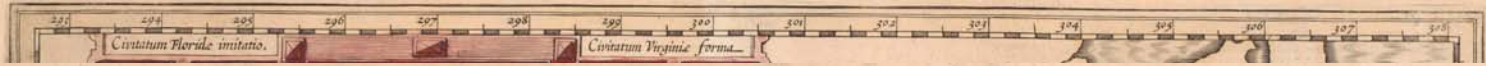
- Strip punctuation from folder names in dark archive and names in *sort_by* field of *Item* table
- Reduce multiple *Item_ID* possibilities by creating auto process to evaluate *Item_IDs* based on their numerical proximity to each other.
- Hire a student

The Intersection of Image
Needs and Expectations:
A Look from the
Carolina Digital Library & Archive's
Digital Production Center

Lisa Gregory







...invenit. a Nordi Cakam Infulam re-
 fluit, exarrens in modum iflimi ad-
 currens, quibus nullis vicinis non ara-
 solennibus Florida partem hic appropin-
 quat, pluviam notatam habemus ex ip-
 si datographo illius qui hanc numeris regi-
 strata accuratissime descripsit. Reliqua
 ex universali nostra descriptione apparet.

reductio 1683

no 81

55
 11
 84

Metadata Issues at a
Large Social Science Data Repository:
A Snapshot of Digital Curator Needs
in the Practice Setting

Samantha Guss

Therefore $s = i$

$$\sqrt{\sum_{i=1}^k f_i d_i^2} - \left(\sum_{i=1}^k f_i d_i \right)^2$$

Quick Links

Search

GO

About the Institute

News & Calendar

Short Courses

Statistical, Qualitative &
GIS Services

Computing Services

Survey Methodology

Data Archive Services

▶ Data Catalog

▶ Public Opinion Polls

▶ National Network of
State Polls

▶ NC Vital Stats

▶ Archiving Project Data

▶ Data Holdings

▶ Other Archive Sites

Grant Services

Other Services

UNC Home

The Howard W. Odum

[Home](#)

 [Print this page](#)

Data Archive Services

Data Catalog

Welcome to our new search engine.

The Odum Institute maintains one of the oldest and largest [catalog](#) of machine-readable data in the U.S. It has an extensive collection of U.S. Census data, including one of the most complete holdings for 1970 Census files. Other major sources of data include the North Carolina State Data Center, which distributes North Carolina census data; and the National Center for Health Statistics.

Public Opinion Poll Database

The Odum Institute's [Public Opinion Poll Database](#) allows any researcher to search for specific poll questions among the more than 230,000 questions in the Institute's archive by key words, date, study number, study title, or state.

NC Vital Statistics

Determining Archival Value in
Orphaned WebSpaces:
An Investigation within the
ibiblio.org Domain

Jennifer Mantooth

ibiblio



Websites and Open source software

Sample View of Context-less Files

Re-Appraisal Questions

1. Does the website contain content?
2. Is the material accessible?
3. Is the material easy to negotiate?
 - a. Are the file names meaningful?
 - b. Do the files have value on their own?

Ongoing Issues

- Migrating abandoned database driven websites
- Alerting the users of defunct links/
outdated sites
 - Many users don't go through ibiblio's collection index (Google)

Questions?

Thank you

<http://ils.unc.edu/digccurr>

