Preparing Future Digital Curators

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

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)

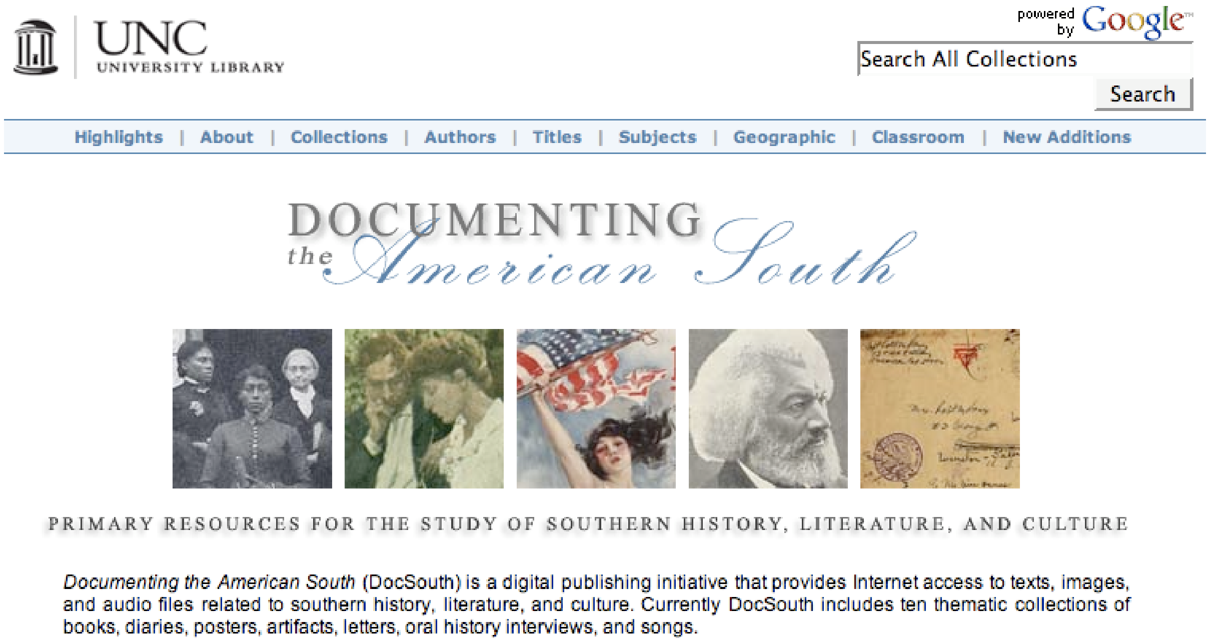
PART II

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


UNC
UNIVERSITY LIBRARY
powered by Google
Search All Collections
Search
Highlights | About | Collections | Authors | Titles | Subjects | Geographic | Classroom | New Additions

DOCUMENTING the American South

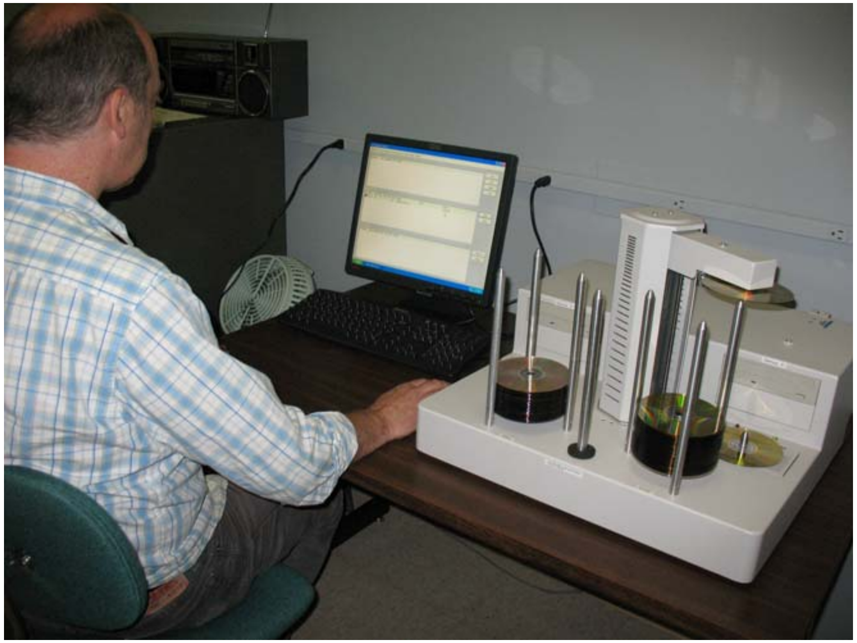
PRIMARY RESOURCES FOR THE STUDY OF SOUTHERN HISTORY, LITERATURE, AND CULTURE

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

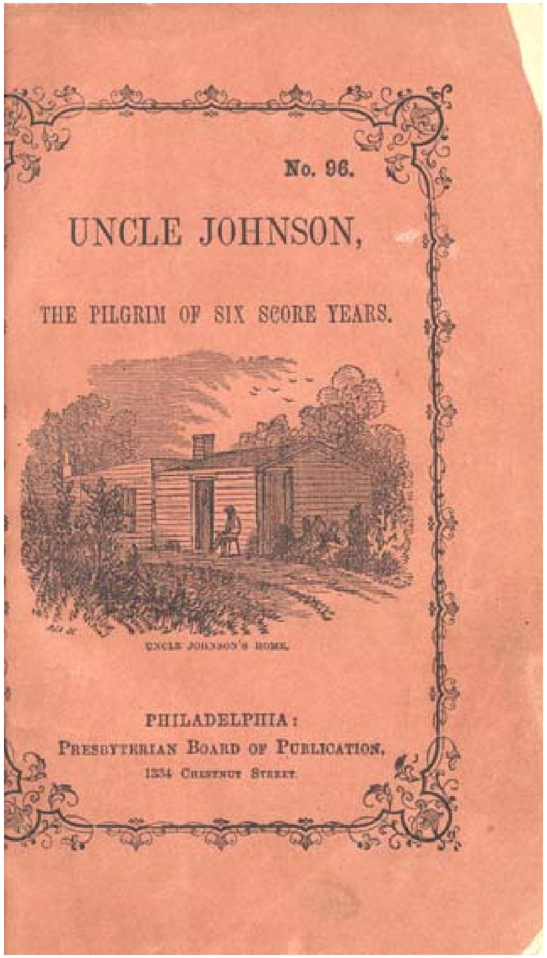
FROM CD to DARK ARCHIVE

- 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



No. 96.

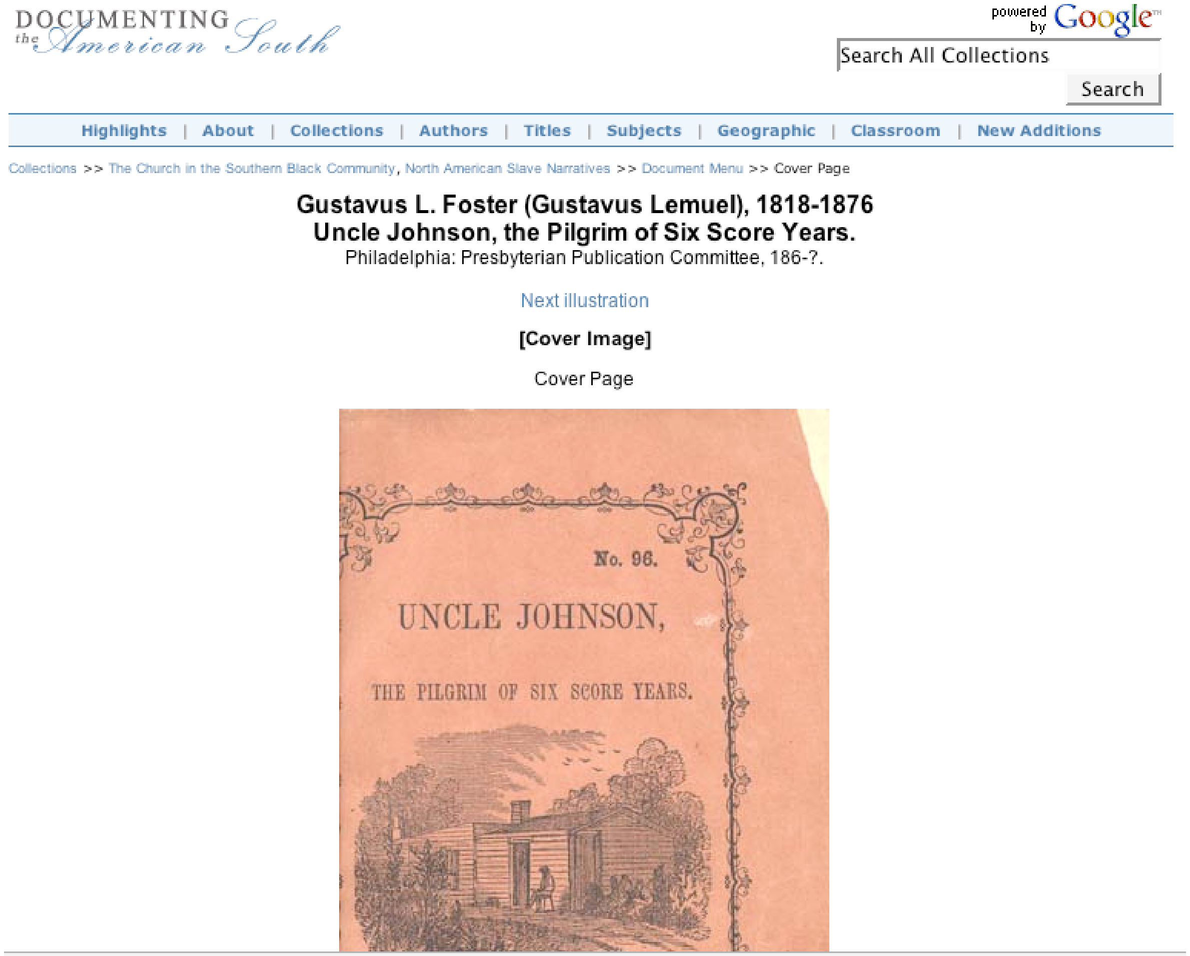
UNCLE JOHNSON,

THE PILGRIM OF SIX SCORE YEARS.

UNCLE JOHNSON'S HOME.

PHILADELPHIA:
PRESBYTERIAN BOARD OF PUBLICATION,
1334 CHESTNUT STREET.

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


DOCUMENTING the American South
powered by Google
Search All Collections
Search
Highlights | About | Collections | Authors | Titles | Subjects | Geographic | Classroom | New Additions
Collections >> The Church in the Southern Black Community, North American Slave Narratives >> Document Menu >> Cover Page

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

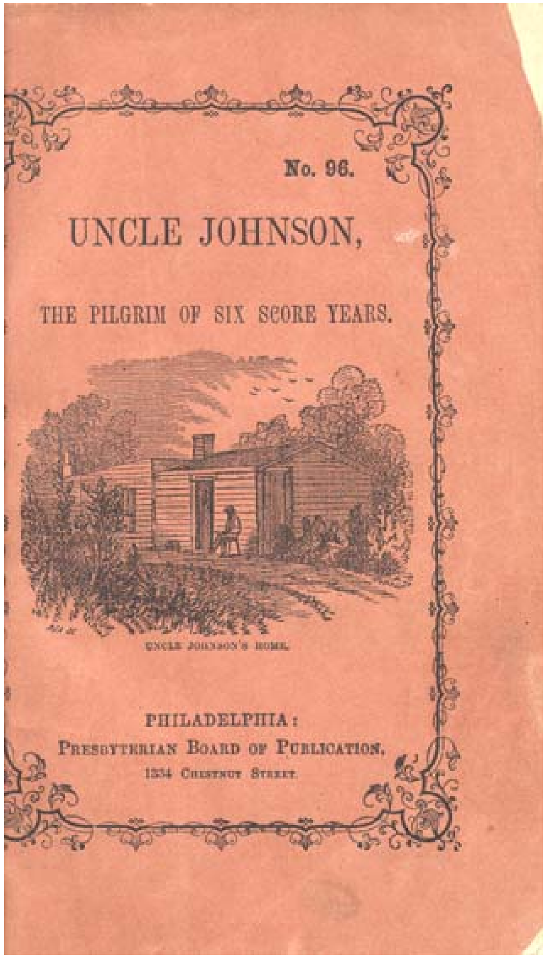
No. 96.

UNCLE JOHNSON,

THE PILGRIM OF SIX SCORE YEARS.

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



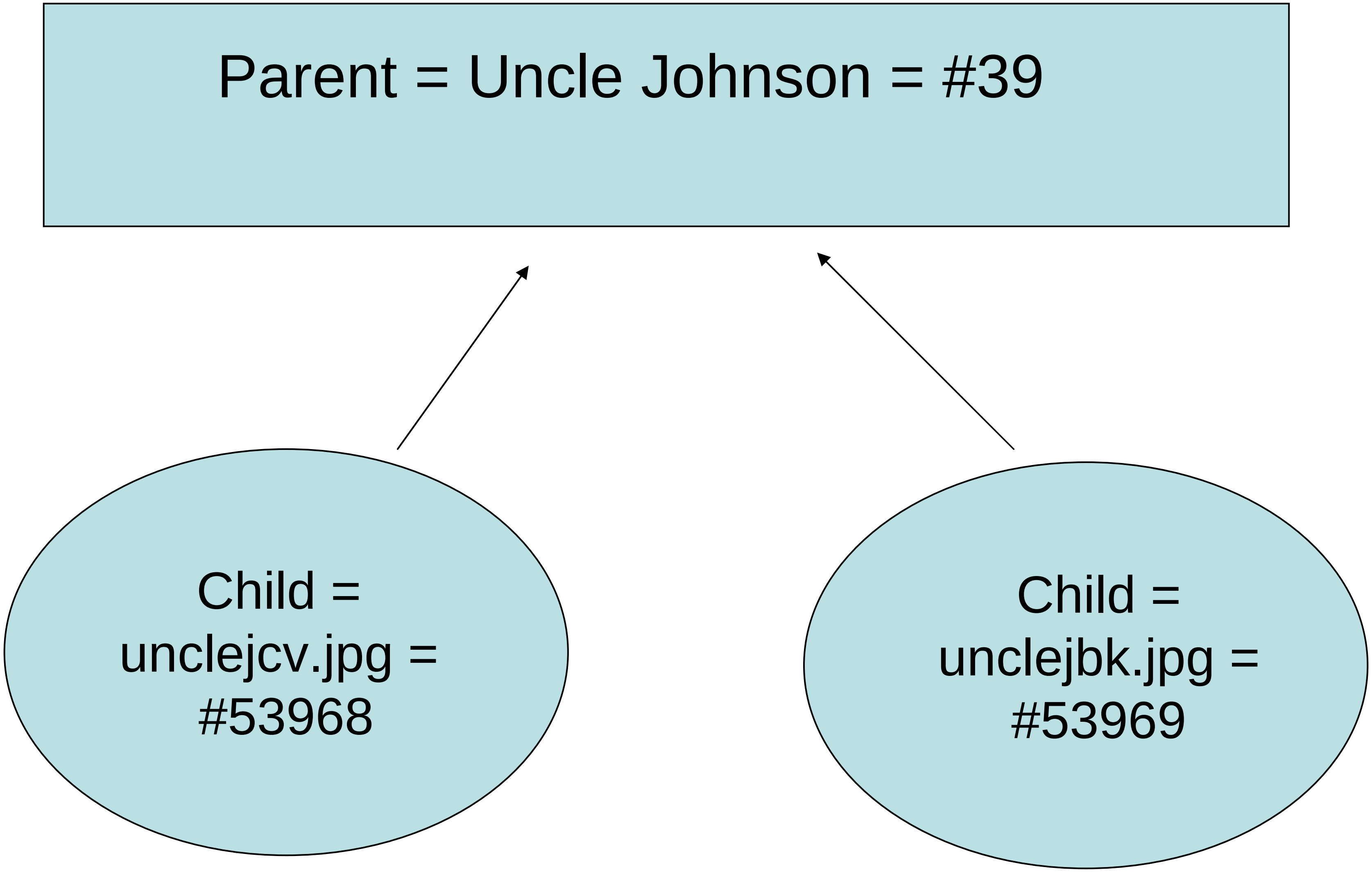
No. 96.

UNCLE JOHNSON,

THE PILGRIM OF SIX SCORE YEARS.

UNCLE JOHNSON'S HOME.

PHILADELPHIA:
PRESBYTERIAN BOARD OF PUBLICATION,
1334 CHESTNUT STREET.



Parent = Uncle Johnson = #39

Child = unclejcv.jpg = #53968

Child = unclejbk.jpg = #53969

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

Specific 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

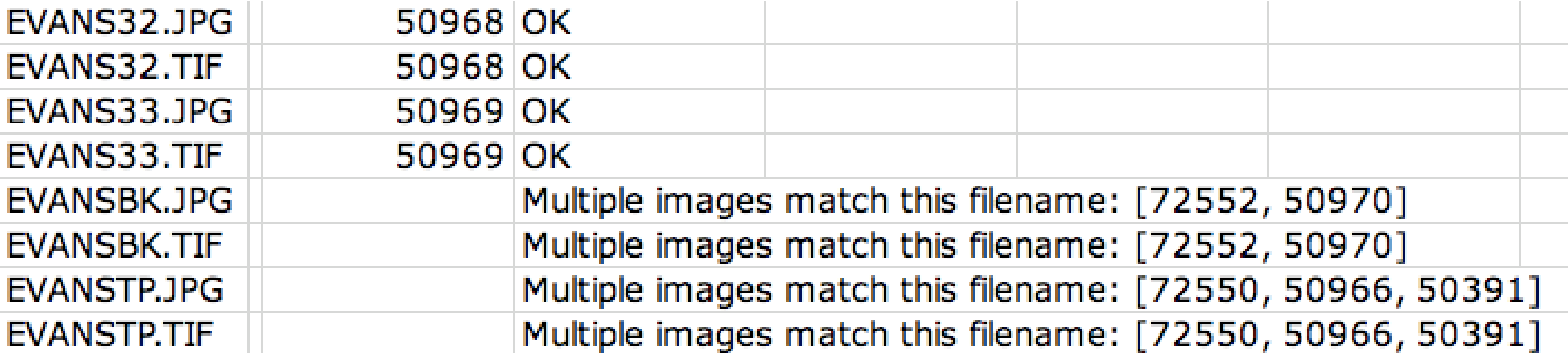
- Establish Parent_ID by searching *Item* table for name of publication
- Using Parent_ID search *Item_child* for range of *Item_IDs* that belong to Parent
- Use *Item_ID* of children to search *Illustration_item* for specific file name associated with that ID.

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

Multiple Matches: Making Assumptions



EVANS32.JPG	50968	OK
EVANS32.TIF	50968	OK
EVANS33.JPG	50969	OK
EVANS33.TIF	50969	OK
EVANSBK.JPG		Multiple images match this filename: [72552, 50970]
EVANSBK.TIF		Multiple images match this filename: [72552, 50970]
EVANSTP.JPG		Multiple images match this filename: [72550, 50966, 50391]
EVANSTP.TIF		Multiple images match this filename: [72550, 50966, 50391]

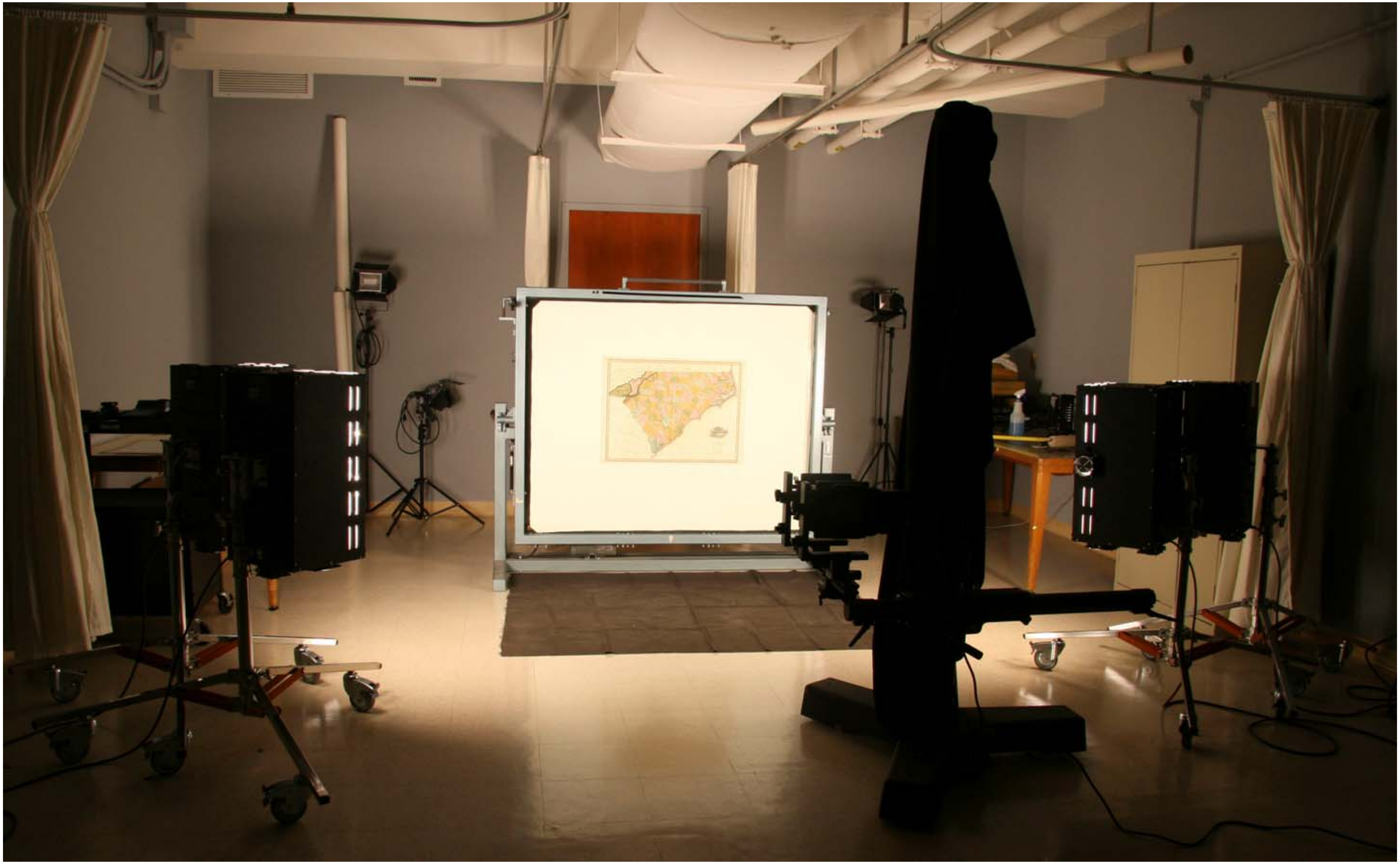
Next Steps

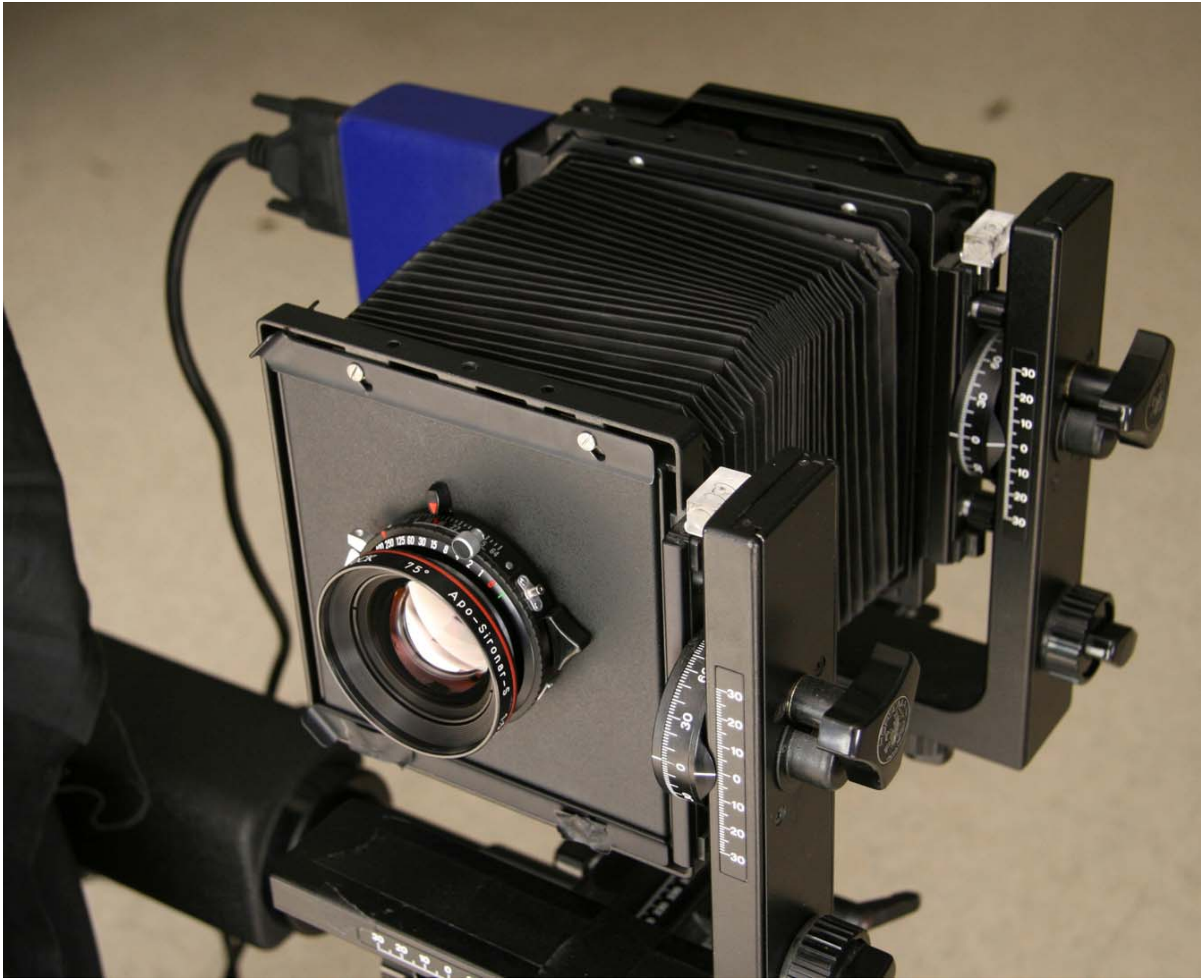
- Strip punctuation from folder names in dark archive and names in *sort_by* field of *Item* table
- Reduce multiple *Item_ID* possibilities by creating auto process to evaluate *Item_IDs* based on their numerical proximity to each other.
- Hire a student

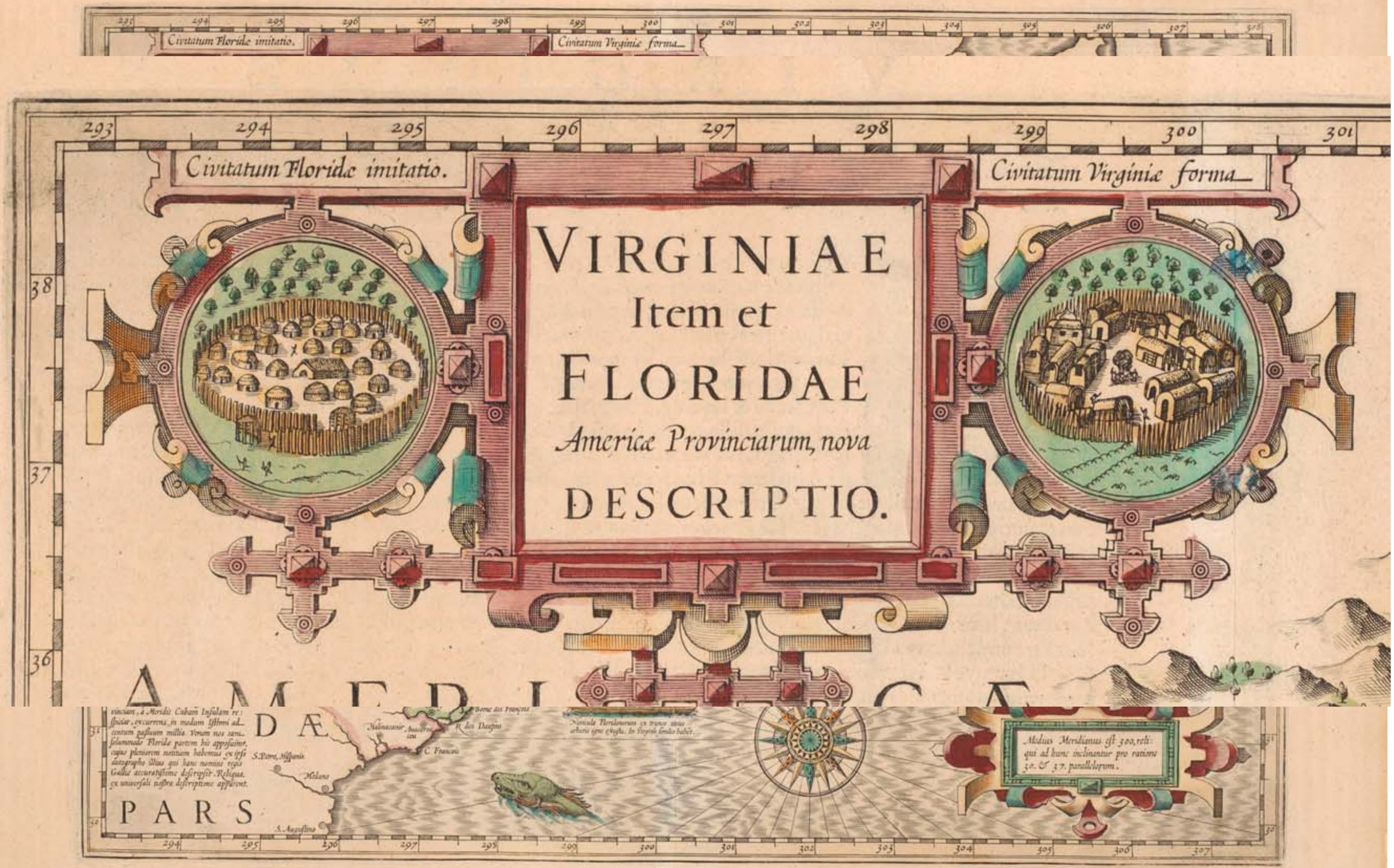
The Intersection of Image Needs and Expectations:

A Look from the Carolina Digital Library & Archive's Digital Production Center

Lisa Gregory




75°
Apo-Sironar-S


VIRGINIAE
Item et
FLORIDAE
Americæ Provinciarum, nova
DESCRIPTIO.
Civitatum Floridæ imitatio.
Civitatum Virginiæ forma
AMERICÆ
DÆ
PARS
vinciam, à Meridie Cubam Inſulam re-
ſpiciat, excurrens in modum Iſthmi ad
centum paſſuum millia. Verum nos tan-
ſolummodo Floridæ partem hic appoſuimus,
cujus pleniorem notitiam habemus ex ipſo
autographo illius qui hanc nomine regis
Galliæ accuratiſſime deſcripſit. Reliqua,
ex univerſali noſtra deſcriptione apparent.
Borne des François
R. des Dauphin
Halmacanir
Auacertocou
C. François
S. Petri, Hiſpanis
Melano
S. Auguſtino
Navicula Floridanorum ex trunco unius arboris igne exſcalpta. In ſiccum fundo habet.
Medius Meridianus eſt 300, reli-
qui ad hunc inclinantur pro ratione
30. & 37. parallelorum.
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
38
37
36
31
30

mercator 1633

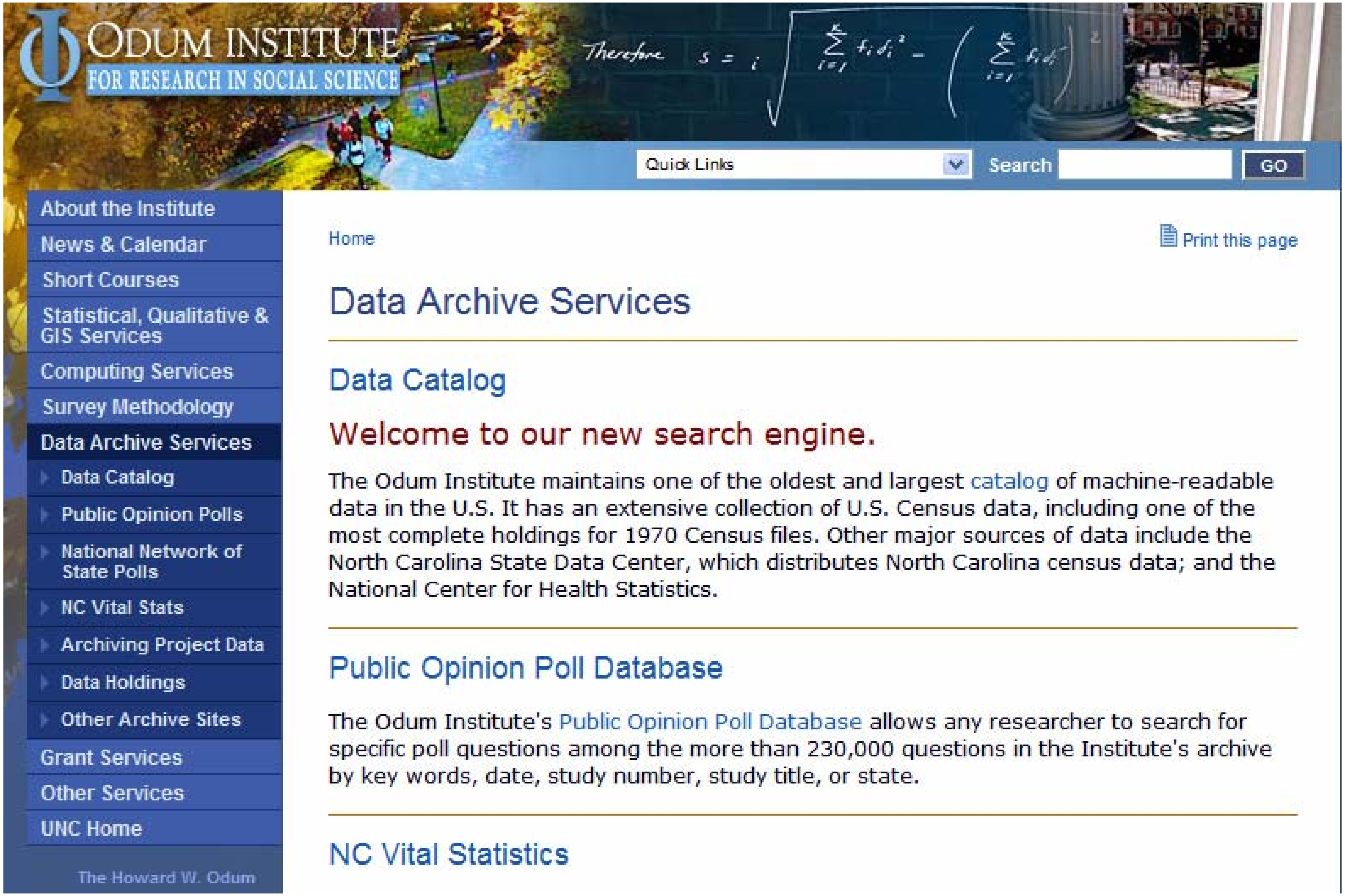
no 81

55-
[illegible]

3 P

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

Samantha Guss



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.

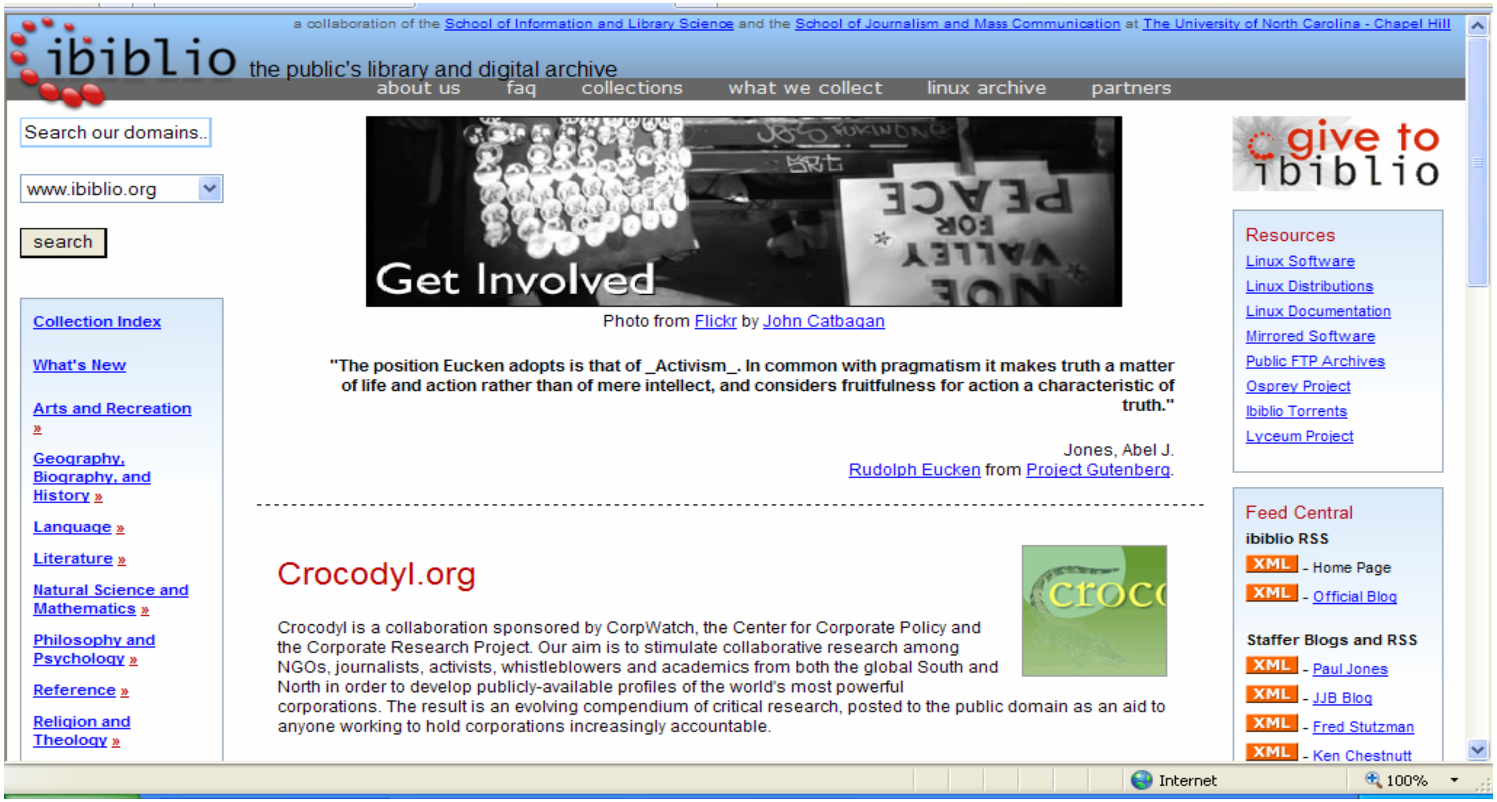
NC Vital Statistics

Determining Archival Value in Orphaned WebSpaces:

An Investigation within the ibiblio.org Domain

Jennifer Mantooth

ibiblio



a collaboration of the School of Information and Library Science and the School of Journalism and Mass Communication at The University of North Carolina - Chapel Hill

ibiblio the public's library and digital archive

about us faq collections what we collect linux archive partners

Search our domains..

www.ibiblio.org

search

Collection Index

What's New

Arts and Recreation »

Geography, Biography, and History »

Language »

Literature »

Natural Science and Mathematics »

Philosophy and Psychology »

Reference »

Religion and Theology »

Get Involved

Photo from Flickr by John Catbagan

"The position Eucken adopts is that of _Activism_. In common with pragmatism it makes truth a matter of life and action rather than of mere intellect, and considers fruitfulness for action a characteristic of truth."

Jones, Abel J.
Rudolph Eucken from Project Gutenberg.

Crocodyl.org

Crocodyl is a collaboration sponsored by CorpWatch, the Center for Corporate Policy and the Corporate Research Project. Our aim is to stimulate collaborative research among NGOs, journalists, activists, whistleblowers and academics from both the global South and North in order to develop publicly-available profiles of the world's most powerful corporations. The result is an evolving compendium of critical research, posted to the public domain as an aid to anyone working to hold corporations increasingly accountable.

give to ibiblio

Resources
Linux Software
Linux Distributions
Linux Documentation
Mirrored Software
Public FTP Archives
Osprey Project
Ibiblio Torrents
Lyceum Project

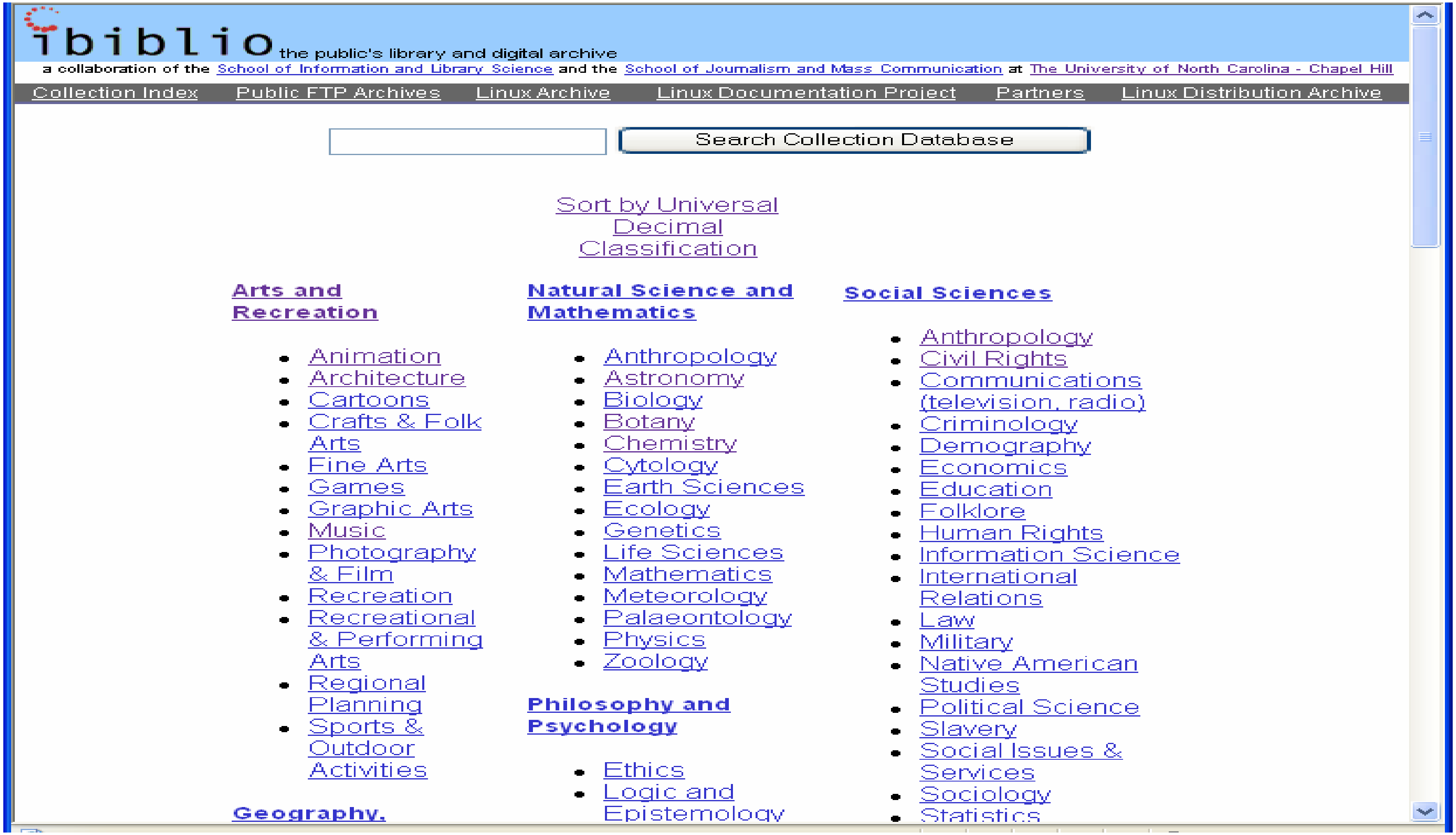
Feed Central
ibiblio RSS
XML - Home Page
XML - Official Blog

Staffer Blogs and RSS
XML - Paul Jones
XML - JJB Blog
XML - Fred Stutzman
XML - Ken Chestnutt

Internet 100%

Websites and Open Source Software

ibiblio Collection Index



ibiblio the public's library and digital archive

a collaboration of the School of Information and Library Science and the School of Journalism and Mass Communication at The University of North Carolina - Chapel Hill

Collection Index | Public FTP Archives | Linux Archive | Linux Documentation Project | Partners | Linux Distribution Archive

Sort by Universal Decimal Classification

Arts and Recreation

- Animation
- Architecture
- Cartoons
- Crafts & Folk Arts
- Fine Arts
- Games
- Graphic Arts
- Music
- Photography & Film
- Recreation
- Recreational & Performing Arts
- Regional Planning
- Sports & Outdoor Activities

Geography,

Natural Science and Mathematics

- Anthropology
- Astronomy
- Biology
- Botany
- Chemistry
- Cytology
- Earth Sciences
- Ecology
- Genetics
- Life Sciences
- Mathematics
- Meteorology
- Palaeontology
- Physics
- Zoology

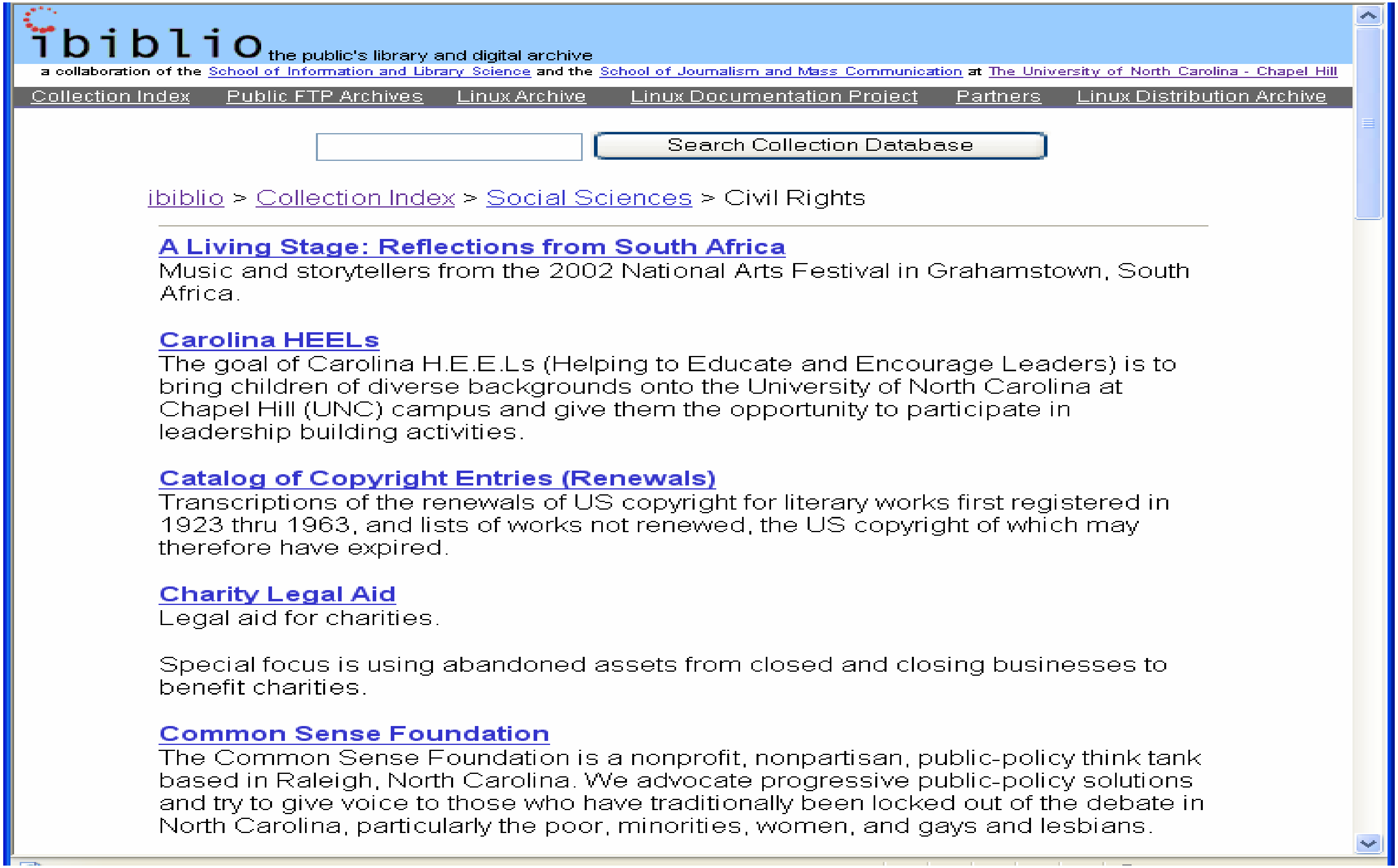
Philosophy and Psychology

- Ethics
- Logic and Epistemology

Social Sciences

- Anthropology
- Civil Rights
- Communications (television, radio)
- Criminology
- Demography
- Economics
- Education
- Folklore
- Human Rights
- Information Science
- International Relations
- Law
- Military
- Native American Studies
- Political Science
- Slavery
- Social Issues & Services
- Sociology
- Statistics

Collection Level – User View


ibiblio the public's library and digital archive
a collaboration of the School of Information and Library Science and the School of Journalism and Mass Communication at The University of North Carolina - Chapel Hill
Collection Index
Public FTP Archives
Linux Archive
Linux Documentation Project
Partners
Linux Distribution Archive
Search Collection Database

ibiblio > Collection Index > Social Sciences > Civil Rights

A Living Stage: Reflections from South Africa
Music and storytellers from the 2002 National Arts Festival in Grahamstown, South Africa.

Carolina HEELs
The goal of Carolina H.E.E.Ls (Helping to Educate and Encourage Leaders) is to bring children of diverse backgrounds onto the University of North Carolina at Chapel Hill (UNC) campus and give them the opportunity to participate in leadership building activities.

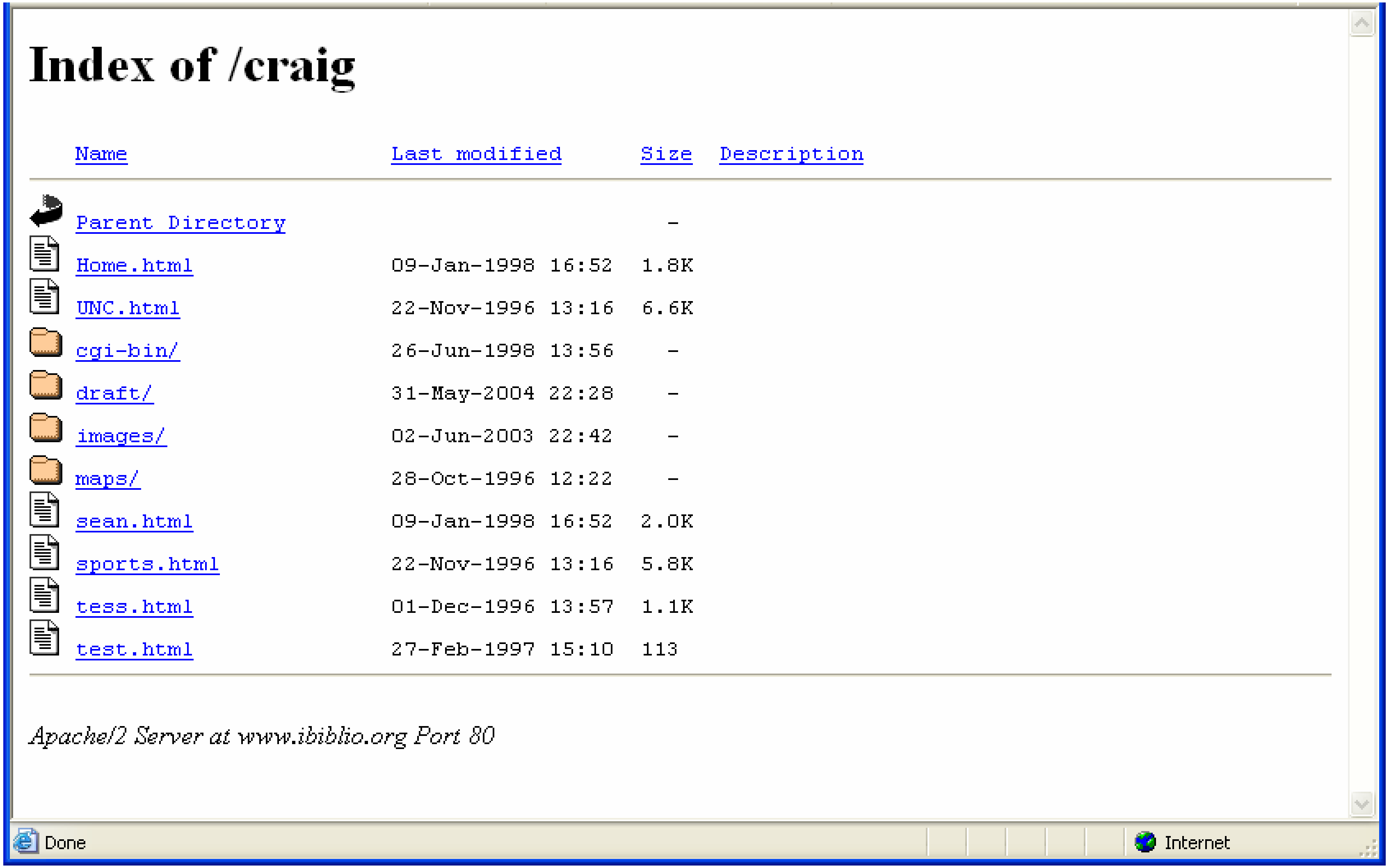
Catalog of Copyright Entries (Renewals)
Transcriptions of the renewals of US copyright for literary works first registered in 1923 thru 1963, and lists of works not renewed, the US copyright of which may therefore have expired.

Charity Legal Aid
Legal aid for charities.

Special focus is using abandoned assets from closed and closing businesses to benefit charities.

Common Sense Foundation
The Common Sense Foundation is a nonprofit, nonpartisan, public-policy think tank based in Raleigh, North Carolina. We advocate progressive public-policy solutions and try to give voice to those who have traditionally been locked out of the debate in North Carolina, particularly the poor, minorities, women, and gays and lesbians.

Sample Abandoned Webspace



Index of /craig

```
   Name                     Last modified       Size  Description
---------------------------------------------------------------------
   Parent Directory                               -
   Home.html                09-Jan-1998 16:52  1.8K
   UNC.html                 22-Nov-1996 13:16  6.6K
   cgi-bin/                 26-Jun-1998 13:56    -
   draft/                   31-May-2004 22:28    -
   images/                  02-Jun-2003 22:42    -
   maps/                    28-Oct-1996 12:22    -
   sean.html                09-Jan-1998 16:52  2.0K
   sports.html              22-Nov-1996 13:16  5.8K
   tess.html                01-Dec-1996 13:57  1.1K
   test.html                27-Feb-1997 15:10  113
---------------------------------------------------------------------
```

Apache/2 Server at www.ibiblio.org Port 80

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

- Preventing loss of context for future sites
- Preserving access to database driven sites
- Providing context to abandoned collections not accessed through the ibiblio collections index (Google)

Questions?

Thank you

http://ils.unc.edu/digccurr