

**CUL Digital Image Collections Inventory
IMLS Project**

MAKING OF AMERICA 2

**THE CORE HISTORICAL LITERATURE OF AGRICULTURE. [INCOMPLETE INFORMATION]
..... 4**

DIGITAL TO MICROFILM CONVERSION 5

**JOINT STUDY IN DIGITAL PRESERVATION: DIGITAL CAPTURE, PAPER FACSIMILES,
AND NETWORK ACCESS (MATH / CLASS) 8**

NEW YORK STATE HISTORICAL 10

HANA..... 12

SAGANET 14

UTOPIA 16

HILL ORNITHOLOGICAL ARTWORK 23

HUMAN ECOLOGY HISTORICAL PHOTOGRAPHS 25

FUERTES ARTWORK..... 28

HOME ECONOMICS PHOTO COLLECTIONS 25

INVENTION AND ENTERPRISE: EZRA CORNELL 29

MAKING OF AMERICA 2 31

DISSERTATIONS AVAILABLE ON INTERNET SYSTEMS (DAISY) 36

UTOPIA 41

Making of America

Project description

- year the collection was created 1996 - 97
- project leaders/coordinators, team members Anne Kenney, Steve Worona, Sal Gurnani, Pela ???, Lee Brink, University of Michigan, Northern Micrographics (outsourced scanning)
- subject matter: 19th century journals and monographs

Source type and characteristics

- document type:
black and white text-based material, lightly-heavily illustrated text
- physical dimensions : size varies, but no greater than 11 x 17, or reduced in size
- scanned: from original

Collection size

- total file size of the collection:
- number of images:
910,000 scanned images
728,857 pages of journal literature, representing 991 volumes
181,463 pages of monographs, representing 242 volumes
- rough file size and number estimate of other accompanying files

Storage media

SCSI disks (library 2)

Scanning information

- resolution 600 dpi
- bit depth 1 bit
- file format and version bitonal TIFF 5.0
- compression technique and ratio ITU group 4 lossless compression
- scanner used
- vendor vs. in-house scanning outsourced to Northern Micrographics

Processing information

- any image enhancements on the master copy
special halftone processing on illustrated text
- derivatives created on-the-fly and static thumb
A portion of the collection was pre-converted to 300dpi in order to provide high-speed translation into screen resolution, especially of heavily illustrated text for which the UM TIF2GIF conversion proved unsatisfactory. 10dpi TIFF images were produced to facilitate visual page-browsing.

Metadata

- BIB and STRUCTURE files:

Bib file:

BIB-VERSION:: CS-TR-v2.0
ID:: MOA-JOURNALS2//.-00000001
ENTRY:: July 12, 1996
TITLE:: The Atlantic monthly Volume 1 (November 1857 - May 1858)
AUTHOR:: Monthly Co., Atlantic
ABSTRACT:: Galaxy,
END:: MOA-JOURNALS2//.-00000001

Structure file (partial):

0	ROOT		0
0	PAGES		1
0	CONTENTS		2
1	0302	308	1308
1	0009	15	1015
1	0143	149	1149
1	0436	442	1442
1	0729	735	1735
1	0570	576	1576

- any external metadata files? **Yes, STRUCTURE file, Dienst configuration**
- internal structuring information
page numbers, page sequence, page type, bibliographic information
- availability of a NOTIS record and extent of cataloging (MARC fields used)

Journals, bib record (Living Age):

899/1: 0: |a MOAProj

856/1:40: |u <http://moa.cit.cornell.edu/MOA/MOA-JOURNALS2/LIVN.html>

Journals, mhd record:

843/1: : |a Reproduction from digital master. |m v.1-227 (1844-1900). |b Ithaca, NY : |c Cornell University Library, |d 1995. |e 23-31 cm.

866/1: 1: |8 0 |a v.1-360 (1844-1941)

Monographs, bib record (Trappers of New York, Simms):

899/1: 0: |a MOAProj

856/1:40: |u <http://moa.cit.cornell.edu/dienst/oa/browse/fly/MOA-JOURNALS2:SIMM-0023/5/1:TIFF2GIF:100>

Monographs, mhd record:

843/1: : |a Computer file. |b Ithaca, N.Y. : |c Cornell University Library, |d 1995. |e [290] image files.

852: : |z Files for the images of individual pages are encoded in

Aldus/Microsoft TIFF Version 5.0 using facsimile-compatible CCITT Group 4 compression.

- manual indexing and how it was recorded
All but two journals (Scientific American and Living Age) were indexed to the article level. Over 102,800 articles for 703 volumes were indexed at Cornell. Data was entered into Excel spreadsheets (one for each journal), with a maximum of 16 data points per entry. Preservation Assistant II staff performed this work.
- descriptive, administrative, and structural metadata elements used (how and where were they recorded)

Access mechanisms

- Web addresses <http://moa.cit.cornell.edu>
<http://www.umdl.umich.edu/moa>
- Cornell implemented a prototype digital library (CDL), based on the Dienst system from the Computer Science Department
- Cornell and Michigan undertook subsequent processing to enhance access to the collection via OCR/SGML

Hardware/software dependencies **HTML, Dienst, Perl scripts**

Refreshing/migration history:

Copyright restrictions: **none**

Quality of image and metadata (will be determined by the IMLS group)

The Core Historical Literature of Agriculture. *[incomplete information]*

The Core Historical Literature of Agriculture is a representative collection of agricultural texts published during the period 1850-1950. Full-text materials cover agricultural economics, agricultural engineering, animal science, crop protection, food science, forestry, human nutrition, rural sociology, and soil science.

For technical specifications, see MOA and Digital to Microfilm Conversion projects

- year the collection was created **1996 - 97**
- project leaders/coordinators, team members
Rich Enlich, ?
Northern Micrographics and (outsourced scanning)
- subject matter:

Source type and characteristics

- document type: black and white text-based material, lightly-heavily illustrated text
- physical dimensions : size varies, but no greater than 11 x 17, or reduced in size
- scanned: from original

Collection size

- total file size of the collection:
- number of images:
300,000 (scanned with MOA) and 450,000 (Digital to Microfilm Conversion)
- rough file size and number estimate of other accompanying files

Storage media

SCSI disks (library 2)

Access mechanisms

Web addresses **<http://chla.mannlib.cornell.edu/>**

- Cornell implemented a prototype digital library (CDL), based on the Dienst system from the Computer Science Department
- Cornell and Michigan undertook subsequent processing to enhance access to the collection via OCR/SGML

Digital to Microfilm Conversion

Project description

- year the collection was created **1994 - 96**
- project leaders/coordinators, team members
Anne Kenney, Sam Demas
Dept of Preservation & Conservation
Mann Library
CIT
Image Graphics Recording Center (COM producer)
- subject matter: **core agricultural historical literature** (19th & 20th century)

Source type and characteristics

- document type: black and white text-based material, lightly illustrated text
- physical dimensions : size varies, but no greater than 8.5 x 11
- scanned: from original

Collection size

- total file size of the collection:
1270 volumes scanned
599 reside on 'library2' in two directories:
/epoch/TULIP/xdodfinal
/epoch/TULIP/xdodstor
- number of images: 1270 volumes scanned represents 450,000 scanned images
- rough file size and number estimate of other accompanying files
Each monograph that has been loaded to the server has 5 - 6 descriptive files (metadata) and one directory of tiff images. Because of the way the file systems are set up we cannot easily calculate, at this point, exactly how much space these volumes require. It is somewhere between 10 and 20 gigabytes.

Storage media **SCSI disks (library 2)**

Scanning information

- resolution **600 dpi**
- bit depth **1 bit**
- file format and version **bitonal TIFF 5.0**
- compression technique and ratio **ITU group 4 compression**
- scanner used **XDOD**
- vendor vs. in-house scanning **in-house scanning**

Processing information

- any image enhancements on the master copy **special halftone processing**
- derivatives created **no**

Metadata

- file header (if possible tags used)
00000006.tif: Warning, unknown field with tag 292 (0x124) ignored.
TIFF Directory at offset 0x8
Subfile Type: (0 = 0x0)
Image Width: 5408 Image Length: 3552

Resolution: 600, 600 pixels/inch
Bits/Sample: 1
Compression Scheme: CCITT Group 4 facsimile encoding
Photometric Interpretation: min-is-white
Thresholding: bilevel art scan
FillOrder: msb-to-lsb
Orientation: row 0 rhs, col 0 top
Samples/Pixel: 1
Rows/Strip: 3552
Planar Configuration: single image plane
Group 4 Options: (0 = 0x0)

- any external metadata files? **Yes, XDOD RDO** **see below**
- internal structuring information

The XDOD scanning process produced a directory containing scanned images and several descriptive files:

armsby.con	directory containing scanned images
armsby.rdo	structure information (binary)
armsby.fof	ascii representation of .rdo file - file of files
armsby.log	ascii representation of .rdo file - conversion log
armsby.ort	ascii representation of .rdo file - orientation of images
armsby.phy	ascii representation of .rdo file - physical structure
armsby.str	ascii representation of .rdo file - logical structure

- availability of a NOTIS record and extent of cataloging (MARC fields used) (Diane can help us to gather this information)

Index screen

- 2 .ECONOMIC AND SOCIAL ASPECTS OF FEDERAL RECLAMATION <1930> (CU 100)
- 3 .ECONOMIC AND SOCIAL ASPECTS OF FEDERAL RECLAMATION <1930> **datafile**
- 4 .ECONOMIC AND SOCIAL ASPECTS OF FEDERAL RECLAMATION <1930> **microfilm**

Datafile

- 538/2:** : |a Files for the images of individual pages are encoded in Aldus/Microsoft TIFF Version 5.0 using facsimile-compatible CCITT Group 4 compression.
- 500/3:** : |a Titles for the NEH digital to microfilm conversion project were selected from the Core historical literature of agriculture, Agricultural economics and rural sociology. The titles are also part of the series: Literature of the agricultural sciences.
- 590/4:** : |a Digital master, c.1: ?????????? |5 cons, opt
- 899/2:** 0: |a CHLA
- 830/3:** 0: |a NEH digital to microfilm conversion project.
- 830/4:** 0: |a Core historical literature of agriculture. |p Agricultural economics and rural sociology.
- 830/5:** 0: |a Literature of the agricultural sciences.

Microfilm

- 533/2:** : |a Computer output microfilm. |b Ithaca, N.Y. : |c Cornell University Library ; |b Shelton, Ct. : |c Image Graphics, Inc., |d 1995. |e 1 microfilm reel : negative ; 35 mm.
- 899/2:** 0: |a CHLA

- manual indexing and how it was recorded

After scanning the technician paginated the book in the RDO file, and indicated those portions of the book included with the process sheets.

- descriptive, administrative, and structural metadata elements used (how and where were they recorded)

Access mechanisms

- online/offline ??????
- Web address

Hardware/software dependencies

Refreshing/migration history: **moved from EPOCH 1997**

Copyright restrictions: **none**

Quality of image and metadata (will be determined by the IMLS group)

Joint Study in Digital Preservation: Digital Capture, Paper Facsimiles, and Network Access (Math / CLASS)

Project description

- year the collection was created **1990 - 1991**
- project leaders/coordinators, team members (60+ people)
 - Anne Kenney, Lynne Personius
 - Library computing staff
 - CIT staff (Docutech printer)
 - Librarians and faculty selectors
 - Xerox Corporation (28)
 - Commission on Preservation and Access
- subject matter: **Math books, 1850 - 1916**, including works by significant authors and those individual titles that have contributed substantially to the development of the discipline

Source type and characteristics

- document type: black and white text-based material, lightly illustrated text
- physical dimensions : size varies, but no greater than 8 1/2 x 11
- scanned: from original

Collection size

- total file size of the collection:
- number of images: 572 math volumes
171,500 digital files (pages)
- rough file size and number estimate of other accompanying files

Storage media

SCSI disks (library 2)

Scanning information

- resolution **600 dpi**
- bit depth **1 bit**
- file format and version **Aldus/Microsoft TIFF 5.0**
- compression technique and ratio **CCITT group 4 compression**
- scanner used **Xerox developmental scanner (CLASS)**
- vendor vs. in-house scanning **in-house**

Processing information

- any image enhancements on the master copy
special care taken to reproduce formulas; threshold settings adjusted
- derivatives created **static**

Metadata

- BIB and STRUCTURE files:
 - Bib file:**
library2% more 00000257.BIB
BIB-VERSION:: CS-TR-v2.0
ID:: MATH/MATH2-00000257
ENTRY:: May 31, 1995
TITLE:: How to Draw a Straight Line; A Lecture
AUTHOR:: Kempe, Alfred Bray
ABSTRACT:: How to Draw a Straight Line; A Lecture
END:: MATH/MATH2-00000257
 - Structure file (partial):**
|0|ROOT||0|
|0|PAGES||1|

0	CONTENTS		2
1	Production Note	1	1001
1	2	1002	
1	3	1003	
1	4	1004	
1	Title Page	5	1005
1	6	1006	
1	Notice	7	1007
1	8	1008	

- any external metadata files? **Yes, STRUCTURE file, Dienst configuration**
- internal structuring information
 - page numbers, page sequence, page type, bibliographic information
- availability of a NOTIS record and extent of cataloging (MARC fields used)

538/2: : |a Files for the images of individual pages are encoded in Aldus/Microsoft TIFF Version 5.0 using facsimile-compatible CCITT Group 4 compression.

590/3: : |a Digital master, c.1: 0207010567B3AF819F84420000000100000010 |5 cons,opt

- manual indexing and how it was recorded
- descriptive, administrative, and structural metadata elements used (how and where were they recorded)

Access mechanisms

- Web addresses <http://moa.cit.cornell.edu>
<http://moa.cit.cornell.edu/dienst-data/cdl-math-browse.html>

Hardware/software dependencies **HTML, Dienst, Perl scripts**

Refreshing/migration history: **moved from EPOCH 1997**

Copyright restrictions: **none**

Quality of image and metadata (will be determined by the IMLS group)

New York State Historical

Project description

test and evaluate network access to Cornell's Digital Library (CDL)

- year the collection was created **1993 - 94**
- project leaders/coordinators, team members
11 comprehensive research libraries in New York State
Anne Kenney, Michael Friedman, Sue Poucher
Bill Turner, Dave Fielding, Chris Stuart
Lydia Pettis
- subject matter: **NYS & Local history materials**
-

Source type and characteristics

- document type: black and white text-based material, heavily illustrated text
- physical dimensions : size varies, but no greater than 8.5 x 11, or reduced in size
- scanned: from original

Collection size

- total file size of the collection: 691 monographs
- number of images:
- rough file size and number estimate of other accompanying files

Storage media

SCSI disks (library 2)

Scanning information

- resolution **600 dpi**
- bit depth **1 bit**
- file format and version **bitonal TIFF 5.0**
- compression technique and ratio **ITU group 4 lossless compression**
- scanner used **CLASS**
- vendor vs. in-house scanning **in-house**

Processing information

- any image enhancements on the master copy
special halftone processing on illustrated text
- derivatives created **none**

Metadata

- any external metadata files? **no**
- internal structuring information
 - original page numbers and self-referencing sections of the original book (table of contents, lists of illustrations, indices, etc)
- availability of a NOTIS record and extent of cataloging (MARC fields used)

Monographs, bib record (A geographical history... Mather, Joseph H.):

256: : |a Computer data (432 image files)

538/2: : |a Files for the images of individual pages are encoded in Aldus/Microsoft TIFF Version 5.0 using facsimile-compatible CCITT Group 4 compression.

590/3: : |a Digital master, c.1:020701053A76AF64A59742000000300000121 |5 cons,opt

899/1: 0: |a NYDigit

Monographs, mhd record:

583/1: : |a reformatted |b paper reproduction |c 1993

843/1: : |a Computer file. |b Ithaca, NY : |c Cornell University Library, |d 1993. |e 432 image files.

852: : |z Available on demand as hard copy or computer file from Cornell University Library.

866/1:41: |8 0 |z <020701053A76AF64A59742000000300000121>

- manual indexing and how it was recorded
- descriptive, administrative, and structural metadata elements used (how and where were they recorded)
 - The document structure reflects the organization of each book. It associates the raw images captured during the scanning process into logical arrangement for presentation to the user. This document structure was built at the time of scanning by the technicians. It accurately describes the original text and is stored with each digital book.

Access mechanisms

Web addresses

<http://library2.library/epoch/olinsc2>

Hardware/software dependencies

Refreshing/migration history:

Moved from EPOCH 1997

Copyright restrictions:

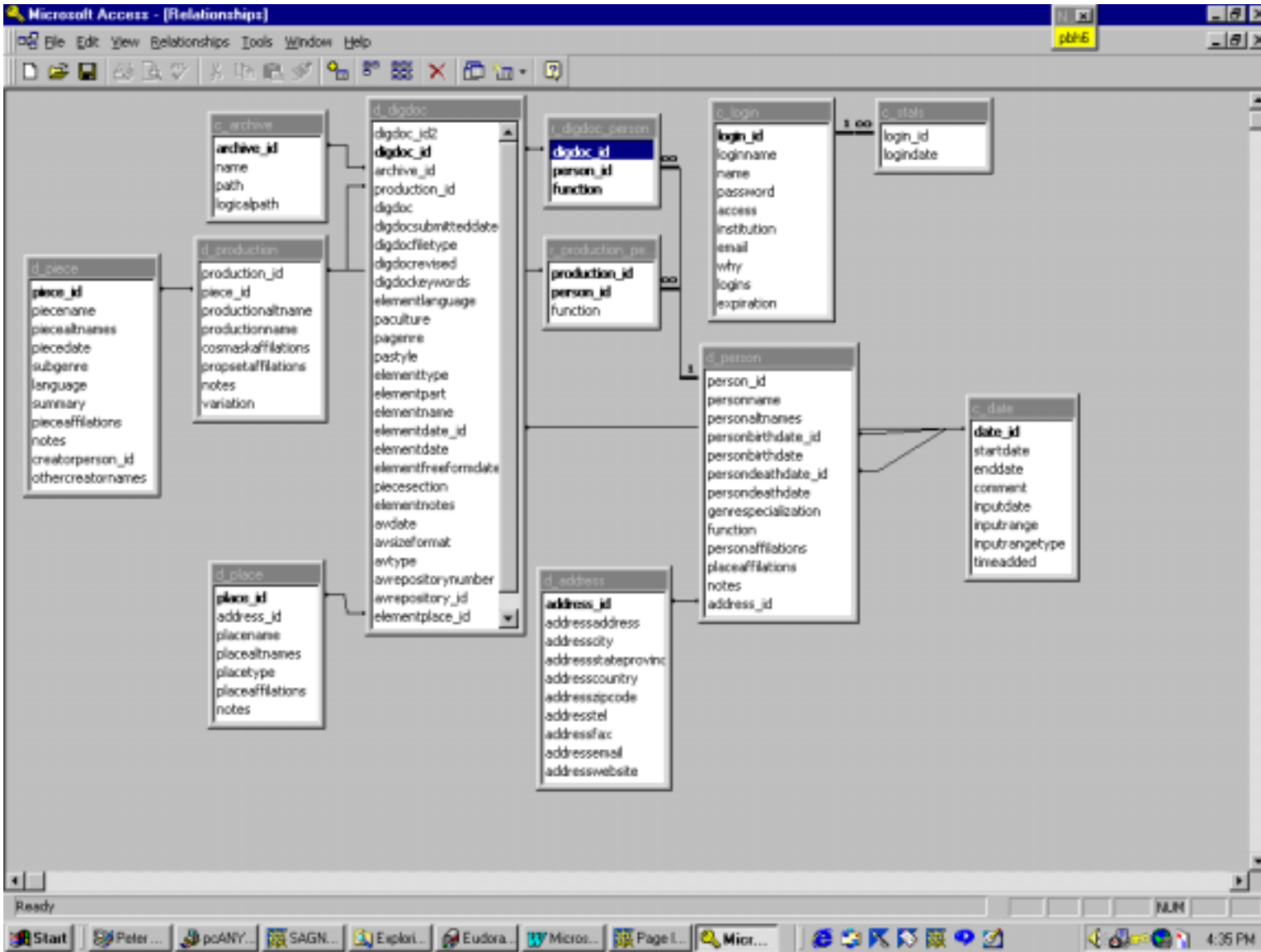
none

Quality of image and metadata (will be determined by the IMLS group)

Hana

<i>Project description</i>	Performing arts resources
• year the collection was created	1998- [ongoing]
• project leaders/coordinators, team members	Karen Brazell, East Asian Studies; Oliver Habicht, CIDC; Ray Wenderlich (student), CIDC; Ann Ferguson, RMC; Gertrude Stein Repertory Theater
• subject matter:	Representations of performances, especially in Japanese theater (but also Russian theater and Shaw)
<i>Source type and characteristics</i>	
• document type:	Primarily slides, though some documents will be added
• physical dimensions :	35 mm
• scanned:	Yes.
<i>Collection size</i>	
• total file size of the collection:	Currently 104 MB (including derivatives)
• number of images:	Currently about 1750 master images and 1750 derivatives.
• rough file size and number estimate of other accompanying files	15 MB
<i>Storage media</i>	Hard drives (faculty member has zip disks)
<i>Scanning information</i>	
• resolution	varies
• bit depth	24 bit
• file format and version	JFIF
• compression technique and ratio	JPEG
• scanner used	Film scanner - unknown brand
• vendor vs. in-house scanning	Students in East Asian Studies
<i>Processing information</i>	
• any image enhancements on the master copy	unknown
<i>Metadata</i>	
• BIB and STRUCTURE files:	Records created in Hana database
• any external metadata files?	Yes - Hana database
• internal structuring information	n/a
• availability of a NOTIS record and extent of cataloging (MARC fields used)	none
• manual indexing and how it was recorded	Students and faculty
• descriptive, administrative, and structural metadata elements used (how and where were they recorded)	ER diagram attached.

<i>Access mechanisms:</i>	
<i>Web addresses</i>	< http://cidc.library.cornell.edu/hana/ >
<i>Hardware/software dependencies</i>	System runs in Access on an NT box, with a Cold Fusion interface
<i>Refreshing/migration history:</i>	Not yet
<i>Copyright restrictions:</i>	Unclear. Currently restricted to people who have an account.
<i>Quality of image and metadata</i> (will be determined by the IMLS group)	



SAGANET

<i>Project description</i>	Icelandic literature from the 12th to the 14th centuries.
<ul style="list-style-type: none"> year the collection was created 	1998- [ongoing]
<ul style="list-style-type: none"> project leaders/coordinators, team members 	Porsteinn Hallgrímsson, The National and University Library of Iceland and H. Thomas Hickerson, Cornell, co-project leaders. Team members include: Patrick Stevens, Peter Hirtle, and the staff of the Preservation Division (Cornell) and Kristun ? and Bjorn Thordarson, Iceland (Bjorn also worked for CIDC).
<ul style="list-style-type: none"> subject matter: 	Icelandic Sagas and rímur poetry, both original material and secondary sources
<i>Source type and characteristics</i>	
<ul style="list-style-type: none"> document type: 	Manuscripts on vellum and paper, and printed texts. In some cases microfilm is scanned.

• physical dimensions :	varies
• scanned:	Yes.
<i>Collection size</i>	
• total file size of the collection:	Currently 36 GB
• number of images:	Estimated it will be 500,000 pages
• rough file size and number estimate of other accompanying files	Difficult to estimate
<i>Storage media</i>	DLT tape and CD-R
<i>Scanning information</i>	
• resolution	varies
• bit depth	24 bit; 8 bit; bitonal
• file format and version	TIFF - unknown version
• compression technique and ratio	Lossless (LZW?)
• scanner used	Eyelike digital camera, Minolta 3000 with grayscale board, Pres. Res. Film scanner
• vendor vs. in-house scanning	both
•	
<i>Processing information</i>	
• any image enhancements on the master copy	sharpening
<i>Metadata</i>	
• BIB and STRUCTURE files:	Created using BDST (also known as ImageTag)
• any external metadata files?	yes
• internal structuring information	
• availability of a NOTIS record and extent of cataloging (MARC fields used)	NOTIS records for Cornell's books; Icelandic records in their system for their manuscripts (copied into SagaNet system)
• manual indexing and how it was recorded	Catalogers in Iceland for manuscripts; document structuring by preservation staff
• descriptive, administrative, and structural metadata elements used (how and where were they recorded)	Too complicated to explain. Includes creation of a table of contents for Icelandic volumes.
<i>Access mechanisms:</i>	
<i>Web addresses</i>	< http://ts013146.bok.hi.is/saganet/ >
<i>Hardware/software dependencies</i>	System runs in Informix on an NT box, with catalog data extracted from the different catalogs and searched via ISITE on a Linux box.
<i>Refreshing/migration history:</i>	Not yet
<i>Copyright restrictions:</i>	none
<i>Quality of image and metadata (will be determined by the IMLS group)</i>	

Utopia

<i>Project description</i>	
• year the collection was created	1994
• project leaders/coordinators, team members	Tom Hickerson, Noni Korf Vidal, Rebecca Davidson, and professors and others at Museum, Art History Dept, RMC and Architecture College.
• subject matter:	Images of European Renaissance art objects, artifacts, architecture, book illustrations, gardens.
<i>Source type and characteristics</i>	
• document type:	35mm slides; originals vary [this is a collection of surrogate images]
• physical dimensions :	
• scanned:	
<i>Collection size</i>	
• total file size of the collection:	47 PhotoCDs
• number of images:	4575 images online
• rough file size and number estimate of other accompanying files	460 MB, 9,500 files (images and html files) 46 MB, (database)
<i>Storage media</i>	PhotoCD
<i>Scanning information</i>	All images were scanned to PhotoCD by Boston Photo
• resolution	
• bit depth	
• file format and version	
• compression technique and ratio	
• scanner used	
• vendor vs. in-house scanning	
<i>Processing information</i>	
• any image enhancements on the master copy	
<i>Metadata</i>	
• BIB and STRUCTURE files:	Master fielded data is in FilemakerPro.. Use copy in MS Access. Please see accompanying pages for field definitions and sample data.
• any external metadata files?	
• internal structuring information	
• availability of a NOTIS record and extent of cataloging (MARC fields used)	
• manual indexing and how it was recorded	
• descriptive, administrative, and structural metadata elements used	
<i>Access mechanisms:</i>	Images can be accessed via author, title, place, subject and keyword.
<i>Web addresses:</i>	http://rmc.library.cornell.edu/Utopia
<i>Hardware/software dependencies</i>	Current implementation: Windows NT server, Access97, ColdFusion; FilemakerPro for production updates
<i>Refreshing/migration history:</i>	Was CD standalone, then Mac website via Filemaker & cgi script, now WindowsNT above software.
<i>Copyright restrictions:</i>	Many restrictions. This site is restricted to the Cornell community.
<i>Quality of image and metadata</i>	

UTOPIA - FIELD DEFINITIONS

TOTAL NUMBER OF RECORDS = 4572

FIELD NAME	TYPE	FORMULA/OPTIONS	EXPLANATION	FORMAT
ENTRYNUM	Text	unique	Entry # is a unique identifier for each image in the Utopia database, last 4 digits of PhotoCD plus image#.	U+#### [PhotoCD ID] + _ + #### [image# on PhotoCD] ex. U2549_0025
FOLDER	Text		Derivative images are currently stored in folders that reflect a PhotoCD number. This is used only for web delivery.	U+#### [PhotoCD ID] ex. U2549
FILENAME	Text		the image filename for web delivery.	IMG####.JPG ex. IMG0025.JPG
CD	Numeric		The complete 16-digit PhotoCD number, 12-digit serial number plus 4-digit image#	ex. 6262302125490025
Cdspaced	Numeric		The same 16 digits with spaces to separate 4 digit groups.	#### #### #### [PhotoCD ID] #### [image# on PhotoCD] ex. 6262 3021 2549 0025
ImageCollection	Text	5 Values: Architecture Art & Planning History of Art Herbert F. Johnson Museum of Art Rare & Manuscript Collections, Kroch Library Private collection (CL)	Campus unit that owns or provided the image	
ImageLoc	Text		Call number or other location identifier, system different within each campus unit, to retrieve the surrogate or original item	
Donor	Text		Name of person who donated the work (used only by the Johnson Museum)	
ImageSource	Text		If image was purchased or loaned, the provider (e.g. Saskia) of the image	
Repository	Text		The museum where a work of art is presently held.	
RepLoc	Text		The name of the city, state (if in the United States), and country (if outside the United States) where the Repository is located.	
Site	Text		Site is the place where a work of sculpture, architecture, or painting integral to the site is located.	
SiteLoc	Text		The name of the city, state (if in the United States), and country (if outside the United States) where the Site is located.	

Title	Text		The Title field contains the name of the painting, sculpture, building, or other work of art. It may be a name given to it by the artist, or assigned later by a historian or cataloger.	
TitleVariant	Text		commonly known other titles (La Gioconda (Mona Lisa); Mona Lisa; Portrait of a Woman)	
Creator1	Text		names of person or group who created the work of art, including painters, architects, sculptors, and other artists.	
Creator2	Text		second known person or group who created the work	
Creator3	Text		third known person or group who created the work	
Creator4	Text		fourth known person or group who created the work	
CreatorRole1	Text	13 Values: Architect Architect and Painter Architect and Sculptor Artist Artist/Cartographer Artist/Printmaker Artist/Printmaker and Author Author Cartographer Delineator/Draftsman Painter Scribe Sculptor	Role of person or group who created the work	
CreatorRole2	Text	(same values as CreatorRole1)	Role of second person or group who created the work	
CreatorRole3	Text	(same values as CreatorRole1)	Role of third person or group who created the work	
CreatorRole4	Text	(same values as CreatorRole1)	Role of fourth person or group who created the work	
Patron	Text		Persons or groups who commissioned the work of art.	
Dates	Text		Dates included in Utopia are the beginning and ending dates of the creation of a work of art, when known.	ex. ca 1535, 1495-1580, 16th century
Dimensions	Text		Usually metric, although inches and explanatory text (e.g., "with base") are added, when available.	ht x width, most have inches in () ex. 25.4 x 51 cm (10 x 20 in)
Dimhtcm	Numeric		numeric, height in centimeters	ex. 25.4
Dimwdthcm	Numeric		numeric, width in cm	ex. 51

ObjectClass	Text	<p>30 (10) Values: Architecture Architecture; Painting Architecture; Painting; Sculpture Architecture; Sculpture Book Book; Architecture Book; Map; Print Book; Print Book; Print; Architecture Drawing Drawing; Architecture Drawing; Architecture; Sculpture Illuminated manuscript Illuminated manuscript; Painting Manuscript Manuscript; Drawing; Architecture Map Map; Architecture Map; Print Map; Print; Architecture Painting Painting; Architecture Painting; Drawing Print Print; Architecture Print; Manuscript Print; Map Sculpture Sculpture; Architecture Textile</p>	<p>This field is used to group together broad categories of works of art. 10 terms are used in combinations that result in 30 values. [Given the possibility of repeating field values, the 30 values listed should be broken out into the following 10: Architecture Book Drawing Illuminated manuscript Manuscript Map Painting Print Sculpture Textile]</p>
-------------	------	--	--

ObjectType	Text	<p>111 values: Abbeys; Altar tombs; Altarpieces; Anatomical studies; Architecture models; Baptisteries; Bas-reliefs; Basilicas; Birth trays; Bookplates; Books of hours; Cantoria; Casinos; Cassone panels; Cathedrals; Chapels; Chapterhouses; Chiaroscuro woodcuts; Churches; Cloisters; Copperplate engravings; Crucifixes; Cutaway drawings; Cutaways; Decorative arts; Detail drawings; Details; Diptychs; Drypoints; Engravings (facsimile); Equestrian monuments; Etchings; Facsimiles; Family portraits; Female portraits; Forts; Fountain sculptures; Fresco cycles; Frescoes; Frieze sculptures; Furniture; Garden pavilions; Genre; Government office buildings; Graduals; Grisaille paintings; Hospitals; Houses; Landscapes; Libraries; Lunettes; Male portraits; Martyria; Mausoleums; Medallions; Medals; Miniatures; Mints; Models; Monasteries; Museums; Niche sculptures; Orangeries; Orphanages; Palaces; Perspective drawings; Perspective sketches; Polyptychs; Portrait busts; Predellas; Preliminary drawings; Preliminary sketches; Preparatory drawings; Processional standards; Profile portraits; Proofs; Reliefs; Responsorials; Sacristies; Sarcophagi; Scuola cycles; Scuole; Self portraits; Sepulchral chapels; Sketches; Spalliera panels; Statuettes; Tabernacles; Tapestries; Temples; Theaters; Tomb sculptures; Tondi; Topographic maps; Town halls; Triptychs; Villas; Wall tombs; Warehouses; Woodcuts; City plans; Site plans; Architecture graphic representations; Axonometric drawings; Cross sections; Elevations; Longitudinal sections; Perspectives; Plans; Sectional elevations; Sections</p>	<p>An Object Type is a more specific kind of Object Class, such as building types in architecture (for example, Churches; Libraries; Villas). For paintings, the Object Type field contains information about such types as Portraits or Altarpieces; for sculpture, such types as Reliefs.</p> <p>Some Object Types refer to more than one Object Class. For example, <i>Predellas</i> is an Object Type that occurs for BOTH Painting and Sculpture Object Classes. Please see accompanying document "UtopiaObjectClass&Type" for a listing of the Class/Type fields.</p>
------------	------	--	---

MatTech	Text		The Materials and Techniques field contains information about what the work of art is made of and how it was made. At the present time, this includes primarily information about materials, for example, green marble or red chalk.
View	Text		View contains information specific to each individual image, for example, whether you are looking at a full view or a detail of a painting. In the case of architecture, the View field contains information about whether the image is a plan, section, elevation, etc., and/or whether the view is an exterior or interior photograph of the building. For medals and coins, obverse and reverse are entered here.
SubjectName	Text		people depicted in a work of art, for example, Hercules; Saint John the Baptist; Virgin Mary
SubjectPlace	Text		places depicted in a work of art, for example, the Garden of Eden; Florence, Italy; San Lorenzo
SubjectConcept	Text		ideas or meanings which may be represented in a work of art, for example, Death; Rape; Resurrection
SubjectDescriptor	Text		things depicted in a work of art, for example, Animals; Musical instruments; Corinthian columns
permissions	Text	4 Values: OK pending need to contact museum directly? DENIED	status of our permission to use images, only for those images for which there is an ImageSource outside of Cornell
ImageNotes	Text		additional information about the Image that amplifies the data in one of the other fields (or doesn't fit in any other field), e.g., further discussion of attribution, location, people or things in the image, etc.
Cataloging Notes	Text		internal notes, problems, etc. Also source of cataloging data, where useful.
BookMSNotes	Text		additional information about the Book that amplifies the data in one of the other fields (or doesn't fit in any other field), e.g., further discussion of attribution, location, people or information about printing or edition

Citation	(Text)	For images from books, we used this field to merge all bibliographic data about the Book together into one field.	Calculation= BookMsAuthor1 & BookMSTitle & BookMsPlace & BookMsPub & BookMsDates & BookMsPages & BookMsNotes.
BookMsAuthor1	Text	Author of the Book or Manuscript from which the image was taken	
BookMsAuthor2	Text	Second author of Book or Manuscript	
BookMsTitle	Text	Title of Book or Manuscript (if image itself has a title, it goes in Title field, above)	
BookMsTitleVariant	Text	Other titles by which Book or Manuscript is known	
BookMsPlace	Text	Place of publication/printing/origin of Book or Manuscript (because of Rare Book cataloging conventions and because geographical boundaries may have changed, this is usually the CITY only)	
BookMsPub	Text	Publisher/printer of Book	
BookMsDates	Text	Date(s) of publication/printing/origin of Book or Manuscript	
BookMsPages	Text	Folio/leaf/page on which image appears in Book or Manuscript. Also plate/figure numbers, if used	
BookMsNotisID	Text	Number identifying bibliographic record in Cornell Library Online Catalog	
tempauthcreator	(Text)	this temporary field combines authors and creators, used only for sorting	Calculation= Creator1 & BookMsAuthor1
tempsort	Numeric	this temporary numeric field is the result of a sort, it is used for current web use and allows us to keep full and details of works together	sorted by tempauthcreator, objectclass, title, view (reverse alpha)

STRUCTURAL RELATIONSHIPS

One-to-One:

Repository<-->RepLoc
Site<-->Site

One-to-Many

ENTRYNUM----->ObjectClass [At this point in no case are there more than 3 ObjectClasses for an EntryNum, this could change.]
ObjectClass----->ObjectType

NOTES:

Dimhtcm and Dimwdthcm are numeric fields that could be used to generate a field to replace the current Dimensions field. Dimensions as it is now is not standardized but it does contain information in text form not found in either Dimhtcm or Dimwdthcm.

Additional Field Definitions that we use but are not sending to Willoughby for conversion:

FIELD NAME	TYPE	EXPLANATION	FORMAT
Keywords	(Text)	used for web searching	Calculation= SubjectName & SubjectPlace & SubjectConcept & SubjectDescriptor & ObjectType & Title & TitleVariant & View & Site & SiteLoc
CreatorMerged	(Text)	used for web searching	Calculation= Creator1 & Creator2 & Creator3 & Creator4
CreatorRoleMerged	(Text)	used for web searching	Calculation= CreatorRole1 & CreatorRole2 & CreatorRole3 & CreatorRole4
TitleMerged	(Text)	used for web searching	Calculation= Title & TitleVariant & View & Site & SiteLoc

Hill Ornithological Artwork

<i>Project description</i>	
• year the collection was created	1999
• project leaders/coordinators, team members	Elaine Engst, David Corson, Noni Korf Vidal, Jennifer Conklin, Beth Fontana, and
• subject matter:	Ornithological artwork from books in the Hill Collection at RMC.
<i>Source type and characteristics</i>	
• document type:	35mm slides , originals are Plates from books
• physical dimensions :	
• scanned:	
<i>Collection size</i>	
• total file size of the collection:	This project is currently growing, cannot provide exact numbers at this time.
• number of images:	
• rough file size and number estimate of other accompanying files	
<i>Storage media</i>	PhotoCD, CD-R
<i>Scanning information</i>	
• resolution	Most of these 35mm slides were scanned by Luna Imaging.
• bit depth	
• file format and version	Most are PhotoCD, those that aren't are TIFF.
• compression technique and ratio	
• scanner used	
• vendor vs. in-house scanning	Mostly vended, some in-house scanning of larger transparencies.
<i>Processing information</i>	
• any image enhancements on the master copy	
<i>Metadata</i>	
• BIB and STRUCTURE files:	Data is stored in FilemakerPro database. Still evolving.
• any external metadata files?	
• internal structuring information	
• availability of a NOTIS record and extent of cataloging (MARC fields used)	
• manual indexing and how it was recorded	
• descriptive, administrative, and structural metadata elements used	
<i>Access mechanisms:</i>	Web users see the images via an online exhibition, or by searching.
<i>Web addresses:</i>	http://rmc.library.cornell.edu/ornithology
<i>Hardware/software dependencies</i>	Current implementation: WindowsNT server for images and html, Mac server running FilemakerPro for web access to database(s).
<i>Refreshing/migration history:</i>	Not yet.
<i>Copyright restrictions:</i>	
<i>Quality of image and metadata</i>	

Human Ecology Historical Photographs

<i>Project description</i>	Please see: http://rmc.library.cornell.edu/eDB-HEphotos/about.htm
• year the collection was created	1998
• project leaders/coordinators, team members	Noni Korf Vidal, Eileen Keating
• subject matter:	Historical photographs from the College of Human Ecology, part of Archives Collection #23-2-749
<i>Source type and characteristics</i>	
• document type:	Black & white photographic prints
• physical dimensions :	Vary, but none larger than approx. 9 x 12". Most are 8 x 10".
• scanned:	
<i>Collection size</i>	
• total file size of the collection:	22 CD-Rs (Mac-formatted)
• number of images:	1419
• rough file size and number estimate of other accompanying files	94MB, 3100 files (derivative images and html for web site) 6.6MB, database
<i>Storage media</i>	CD-R
<i>Scanning information</i>	
• resolution	Mostly @ 400 dpi
• bit depth	Most @ 8bit greyscale, some 16bit
• file format and version	TIFF
• compression technique and ratio	
• scanner used	(U-Max flatbed scanner, model?)
• vendor vs. in-house scanning	Vended to Media Services (CU)
<i>Processing information</i>	
• any image enhancements on the master copy	I understand that some minor level adjustments were made by scanning operator.
<i>Metadata</i>	
• BIB and STRUCTURE files:	Please see accompanying document with field definitions and sample data.
• any external metadata files?	
• internal structuring information	
• availability of a NOTIS record and extent of cataloging (MARC fields used)	
• manual indexing and how it was recorded	Fielded data was derived from finding aid (OCR on printed document).
• descriptive, administrative, and structural metadata elements used (how and where were they recorded)	
<i>Access mechanisms:</i>	Web users can browse collection by folder, or search by keyword, folder.
<i>Web addresses:</i>	http://rmc.library.cornell.edu/eDB-HEphotos
<i>Hardware/software dependencies</i>	Current implementation: Windows NT server for web pages and images & Mac server for web searching of FilemakerPro database.
<i>Refreshing/migration history:</i>	Some website changes for better access, no image changes.
<i>Copyright restrictions:</i>	Mostly © RMC
<i>Quality of image and metadata</i>	

Home Economics Photo Collections

New York State College of Home Economics Records, Collection #23-2-749

The records of the New York State College of Home Economics at Cornell University document the history of the College, and its activities, events,

and personalities. The collection, housed in the Division of Rare and Manuscript Collections of the Cornell Library, includes administrative records and correspondence that document the College's involvement with government, educational, and private organizations at the national, state, and local levels; the College's involvement in World War I and II, defense and relief committees; student and alumni activities; extension and outreach work;

College committees and activities including the Farmers' Institute, Farm and Home Week, Honors, and Home Economics Institutes; and blueprints, sketches, and samples relating to the construction of Martha Van Rensselaer Hall. The collection also includes personal papers of Martha Van Rensselaer. Additionally over fourteen hundred photographs of students and faculty, buildings and scenes portray the College from its founding through 1969.

This Web site provides a database of selected images and information from the records of the College of Home Economics. The photographs are organized by general subjects: Resident Teaching; International, National, State, and Local Public Relations; All-College Activities; Administration and

Personnel; Departments and Divisions; Buildings and Equipment; History of the College; and Miscellaneous. Within any particular general subject category, there are more specific topics, e.g. "Resident Teaching" includes "Students and Student Life," "Homemaking Apartments," and "Apartment Babies." Descriptive information is provided for each image. Additional photographs may be found in this collection and in other holdings of the Division of Rare and Manuscript Collections.

NUMBER OF RECORDS = 1419

FIELD DEFINITIONS

FIELD NAME	TYPE	EXPLANATION	FORMAT
MsgID	Number	unique, serial id#	1-1419
Section	Text	name of category within which folders reside	ex. Resident Teaching
FolderNumber	Number	folder arrangement in physical collection	1-51
FolderTitle	Text	name of folder with section	ex. [Resident Teaching:] Students and Student Life
FolderCode	Text	code assigned to folder, reflects category and folder	ex. RT-S
ItemNumber	Number	number of photograph within folder	1-143
Description	Text	description of photograph	free text, no embedded returns
CD	Text	Letter assignment of CD storing Tiff files	A-U
ItemCode	(Text)	Calculation = FolderCode&-&ItemNumber	ex. RT-S-01
Filename	(Text)	Calculation = ItemCode&.&Fileformat	RT-S-01.JPG
Fileformat	Text	file extension (currently ALL in file are .JPG)	.JPG

SAMPLE RECORD

MsgID	1
Section	Resident Teaching
FolderNumber	1
FolderTitle	Students and Student Life
FolderCode	RT-S
ItemNumber	01

Description	Grinding valves, cleaning spark plugs and carbon from engine. Left side of car: Elizabeth Luxford of Hamburg, Carolyn Howland of New Orleans, LA, and Patricia Prescott of Sandy Creek. Right side of car: Marjorie Voorhees of Auburn, Mary Elizabeth Randolph of Niagara Falls and Lee English (?) of Rochester.
CD	A
ItemCode	RT-S-1
Filename	RT-S-1.JPG
Fileformat	JPG

STRUCTURAL RELATIONSHIPS

FolderTitle---one-to-one---->FolderCode
FolderTitle---one-to-many---->FolderNumber
Section---one-to-many---->FolderNumber
FolderNumber---one-to-many---->ItemNumber
CD---one-to-many---->ItemCode

Fuertes Artwork

<i>Project description</i>	
• year the collection was created	1994
• project leaders/coordinators, team members	Tom Hickerson, Elaine Engst, Noni Korf Vidal, Maggie Hale, and other RMC staff also Johnson Museum and Lab. Of Ornithology
• subject matter:	Ornithological artwork & photographs by L. A. Fuertes
<i>Source type and characteristics</i>	
• document type:	35mm slides, originals vary and include paintings, sketchbooks, and murals.
• physical dimensions :	
• scanned:	
<i>Collection size</i>	
• total file size of the collection:	28 PhotoCDs
• number of images:	2627
• rough file size and number estimate of other accompanying files	107 MB, 5400 files (images and html) 1.3MB, database
<i>Storage media</i>	PhotoCD
<i>Scanning information</i>	All images were scanned to PhotoCD by Boston Photo
• resolution	
• bit depth	
• file format and version	
• compression technique and ratio	
• scanner used	
• vendor vs. in-house scanning	
<i>Processing information</i>	
• any image enhancements on the master copy	
<i>Metadata</i>	
• BIB and STRUCTURE files:	Extent of fielded data does not extend beyond what is evident in web record for individual image.
• any external metadata files?	
• internal structuring information	
• availability of a NOTIS record and extent of cataloging (MARC fields used)	
• manual indexing and how it was recorded	
• descriptive, administrative, and structural metadata elements used (how and where were they recorded)	
<i>Access mechanisms:</i>	
<i>Web addresses:</i>	http://rmc.library.cornell.edu/Birds
<i>Hardware/software dependencies</i>	Current implementation: Windows NT server, Access97, IIS (.idc, .htx)
<i>Refreshing/migration history:</i>	Was CD standalone, then Mac website via Filemaker & cgi script, now WindowsN with above software.
<i>Copyright restrictions:</i>	Mostly © RMC
<i>Quality of image and metadata</i>	

Invention and Enterprise: Ezra Cornell

INCOMPLETE INFORMATION

Project description

- year the collection was created
- project leaders/coordinators, team members
Cornell's Interactive Media Group
Cornell University Library
- subject matter:
19th century letters, diaries, photographs, documents and publications that chronicle the life of Ezra Cornell. Aimed at the study of local history by elementary students in upstate New York.

Source type and characteristics

- document type:
- physical dimensions :
- scanned:

Collection size

- total file size of the collection:
- number of images: 30,000
- rough file size and number estimate of other accompanying files

Storage media

Scanning information

- resolution **600 dpi**
- bit depth **1 bit**
- file format and version **bitonal TIFF 5.0**
- compression technique and ratio **ITU group 4 lossless compression**
- scanner used
- vendor vs. in-house scanning
-

Processing information

- any image enhancements on the master copy

Metadata

- BIB and STRUCTURE files:
- any external metadata files? **Yes, STRUCTURE file, Dienst configuration**
- internal structuring information
- availability of a NOTIS record and extent of cataloging (MARC fields used)
- manual indexing and how it was recorded
- descriptive, administrative, and structural metadata elements used (how and where were they recorded)

Access mechanisms:

Web addresses

<http://cidc.library.cornell.edu/cidc6/Default.htm>

Hardware/software dependencies

html, Dienst and PERL scripts

Draft

Refreshing/migration history: **moved from EPOCH, 1997**

Copyright restrictions: **none**

Quality of image and metadata (will be determined by the IMLS group)

Ezra Cornell

Project title

Ezra Cornell

Project description

- year the collection was created
- project leaders/coordinators, team members
RMC
- subject matter: **The Ezra Cornell papers consist of correspondence, financial and legal records, court proceedings, and other documents pertaining principally to the Cornell family, the telegraph industry, and the founding of Cornell University. The papers also show Cornell's career as a farmer, New York State legislator, entrepreneur, and philanthropist, and include subsequently printed materials collected in support of his various affairs and enterprises. 19th century**

Source type and characteristics

- document type: black and white text-based material, lightly-heavily illustrated text, photographs
- physical dimensions : size varies, but no greater than 11 x 17, or reduced in size
- scanned: from original

Collection size

- total file size of the collection: boxes 1 - 38
- number of images:
- rough file size and number estimate of other accompanying files

Storage media

SCSI disks (library 2)

Scanning information

- resolution **600 dpi**
- bit depth **1 bit**
- file format and version **bitonal TIFF 5.0**
- compression technique and ratio **ITU group 4 lossless compression**
- scanner used
- vendor vs. in-house scanning **in-house**

Processing information

- any image enhancements on the master copy
- derivatives created **static**

Metadata

- BIB and STRUCTURE files:
- any external metadata files? **Yes, STRUCTURE file, Dienst configuration**
- internal structuring information
- availability of a NOTIS record and extent of cataloging (MARC fields used)
856/1:7 : |u http://rnc.library.cornell.edu/Cornell-findaid/ECFa-top.html |2
- manual indexing and how it was recorded
- descriptive, administrative, and structural metadata elements used (how and where were they recorded)

Access mechanisms:

Web addresses <http://moa.cit.cornell.edu/MOA/EZRA.html>

Hardware/software dependencies **html, Dienst and PERL scripts**

Refreshing/migration history: **moved from EPOCH, 1997**

Copyright restrictions: **none**

Quality of image and metadata (will be determined by the IMLS group)

Making of America 2

Project description

- year the collection was created **1997 - present**
- project leaders/coordinators, team members
UC Berekely Library, New York Public Library, Stanford U Library, Penn State U Library, Cornell U Library
- subject matter: **Transportation in the United States, 1869 - 1900**
With a particular focus on the development of the railroads. Collection includes printed materials, broadsides, pictorial collections, manuscripts, diaries, letters, etc.

Source type and characteristics

- document type: archival manuscript materials
- physical dimensions : size varies, but no greater than 11 x 17
- scanned: from original

Collection size

- total file size of the collection: 1,197 as of 8/99, project continues
- number of images:
- rough file size and number estimate of other accompanying files

Storage media

SCSI disks (library 2)

Scanning information

- resolution 300 dpi
- bit depth 90% are 8 bit with the remainder scanned at millions of colors
- file format and version TIFF 6.0
- compression technique and ratio uncompressed
- scanner used HP Laser Jet flatbed scanner
- vendor vs. in-house scanning in-house

TYPE: Flatbed scanner

BRAND: HP

MODEL: Scan Jet 3c/T

SERIAL NUMBER: US00106105

BIT DEPTH: 90% of what I have scanned is 8 bit. The remainder is 24 bit.

ILLUMINATION: "single flourescent lamp"

COLOR INFO: ? **Their example is Gamma 1.8; do we know the Gamma for the scanner?**- how does this differ from the gamma question below?

COLOR SPACE: RGB

COLOR PROFILE: ?

FILE FORMAT: tiff

GAMMA:the only "gamma" we can tap into relates to the monitor, not the scanner

COMPRESSION: none

Processing information

- any image enhancements on the master copy
no image enhancement except occasional stitching for large images
- derivatives created

six derivatives

Metadata

The metadata represents a major research component of this project white paper: (<http://sunsite.berkeley.edu/moa2>). The intent is to develop community metadata standards and best practices for creating and encoding digitized versions of primary resources, as well as establishing naming conventions that can be implemented by multiple sites on projects of national scope.

Metadata will include a union catalog of USMARC collection-level catalog records accessible via the web, and Encoded Archival Descriptions (finding aids).

- BIB and STRUCTURE files:
- any external metadata files?
- internal structuring information
- availability of a NOTIS record and extent of cataloging (MARC fields used)
- manual indexing and how it was recorded
- descriptive, administrative, and structural metadata elements used (how and where were they recorded)

Access mechanisms

- Web addresses
temp site for images: <http://www.library.cornell.edu/preservation/moa2/MOAcamp.htm>

Hardware/software dependencies

Refreshing/migration history:

Copyright restrictions: **none**

Quality of image and metadata (will be determined by the IMLS group)

Dissertations Available on Internet Systems (DAISY)

Oya- Here is my best try for filling out your form with help from George Kozak and Steve Worona. Also include fyi is a report I started writing about the project years ago but never finished. The report is listed below and just for your info. It is not complete in any way. Let me know if I can do more- cheers- john

George Kozak says:

"I was able to add the filesize, but determining the number of images is not easy. The images are dispersed over a lot of directories since the old epoch system only allowed 99 files to a directory. I thought I had a good SQL dump of the collection, but found that it only included the first 44 records. I'd roughly guess that the image count is 22,000 images."

Project description

year the collection was created 1992-1995
project leaders/coordinators, team members Steve Worona, John Saylor, Debbie Brady (replaced by Ron Watkins, then Minnie Empson), Jim Doolittle, Stuart Lynn
subject matter PhD Dissertations from Computer Science, Electrical Engineering, Civil and Environmental Engineering

Source type and characteristics

document type (e.g., black and white text-based material, heavily illustrated text, color photographs, special collections materials, slides, etc.) black and white text based material
physical dimensions (category: regular, oversize - if possible exact size, or "size varies" statement with min and max measurements) size varies, but no greater than 8.5 x 11 or some such) 8.5x11
scanned from original or film intermediate scanned from original

Collection size

total file size of the collection 114 Dissertations
number of images 22,000 (approx.)
rough file size and number estimate of other accompanying files (metadata, processing, etc.) 1.7Gigabytes

Storage media

Scanning information

resolution 600dpi
bit depth bitonal
color space or CLUT (Color Lookup Table) information for color documents)
file format and version group 4 tiff
compression technique and ratio group 4
scanner used Xerox Xdod wg 30
vendor vs. in-house scanning in-house

Processing information

any image enhancements on the master copy? E.g., how were halftones handled? Any special treatment? None

derivatives created (access, processing; such as scaled/reformatted copies for Web delivery, OCR'ed images, etc.) None

Metadata

file header (if possible tags used)

any external metadata files? Full text ascii abstract

availability of internal structuring information (to navigate internal organization of a publication) and structuring labels used (to identify structuring information for pages, such as title page, index, bibliography, etc.)

availability of a NOTIS record and extent of cataloging (MARC fields used) (Diane can help us to gather this information) fully -cataloged records

manual indexing and how it was recorded (such as MOA1 descriptive indexing) None

descriptive, administrative, and structural metadata elements used (how and where were they recorded)

Access mechanisms

online/offline originally on EPOCH server

Web address none

Hardware/software dependencies

Refreshing/migration history Currently, this collection is queued to be migrated to the existing CDL format. It was moved on Nov.7, 1997 from the EPOCH jukebox server to library2.library.cornell.edu.

Copyright restrictions copyright owned by Author

DAISY (Dissertations Available on Internet SYstems) - The Cornell Proposal

Introduction

Our Goal: A PhD candidate begins here graduate program at Cornell. Her field of study advisor suggest she look at other dissertations done at Cornell and other universities in her area. She looks up her subject are in the Cornell Online Catalog from her personal workstation or a public workstation. She limits her search to other Phd Dissertations and finds several at Cornell and several at other institutions that look interesting by reading the abstract of the thesis in the catalog. She clicks on the icon that indicates the the table of contents and bibliography of the theses are available for viewing from her workstation. This helps her to narrow her search further. When she finds one she wants to see in more detail she clicks on the icon to bring up the full text of a particular page. She browses and reads for awhile decides she wants a copy of a portion of a dissertation, so she issues a print command that initiates a procedure by which she is charged for the printing and any copyright royalties if applicable. A high quality printed copy of her request is then delivered to her campus address.

What procedures do we need in place to accomplish this goal?

Cornell University, the University of Michigan, and the Pennsylvania State University have been discussing immediate steps that we may take to begin experimenting with the electronic capture, storage, search, and dissemination of dissertations since July of 1992. Our initial investigation suggests that technology exists today to allow each of our institutions to produce and store its own dissertations in printable/displayable image form, as well as to provide a searchable bibliographic database linked to the stored images. Moreover, the expanding infrastructure of high-speed Internet connections and the ever-increasing acceptance of such network protocols as gopher and WAIS and world-wide web allow easy network-wide access to both the image and text repositories. While emphasizing our ability to produce single-institution dissertation collections, we recognize the need to perform global searches throughout the universe of such collections. Again, existing technology appears to offer the means to this end.

We have initiated pilot projects that explore this distributed method of dissertation storage and access, emphasizing common protocols and file formats, but changing as little as possible of the dissertation-creation process as perceived by students and their faculty advisors. We believe that short-term adoption of such an evolutionary approach must precede any of the more ambitious ventures in this area. The Coalition for Networked Information (CNI) sent out a "Call for Participation in design" meeting to explore the feasibility of a project to capture and store electronic dissertations in July of 1992. While the members of our group feel we have the interest and resources to carry out our projects ourselves, we'd prefer to coordinate our efforts with other similar ones under the CNI umbrella.

Each of our institutions has its own idiosyncratic methods for processing dissertations, and will therefore develop its own customized form of our joint pilot project. Our proposal for Cornell's new procedure follows a description of our experience so far..

Historical Poces

Cornell's procedures historically require students to submit two, single-sided copies of their dissertations on acid-free paper. These are ultimately bound and stored in our library, one (which never circulates) for archival purposes, and one available for on-campus and inter-library loan. At local commercial copy centers, per-page charges for these special-paper copies may be as high as 15 cents. Before binding, Cornell microfilms one of the copies and sends this microfilm to UMI. Local microfilming reduces overall costs to the student by lowering both UMI's direct charges and the charges for postage.

The student pays an \$85 fee at the time of final approval of the Graduate School Theses Advisor. This fee is usposed to cover the cost of microfilming, negative storage at UMI, and binding. The student has aready paid money to a local copy shop for the two unbound copies that she must submit.

Procedure for May 93

As a demonstration, we selected 5 dissertations from the May 1993 submission "crop" from the schools of Civil and Environmental Engineering, Computer Science and Electrical Engineering. After receiving permissions both of the suthors and their advisors and graduate filed representatives these 5 dissertations were scanned. The original paper copies that the student submitted were then microfimed, bound, cataloged by the library and shelved in the normal fashion. (STEVE_ CAN YOU TALK HERE ABOUT WHAT WAS DONE WITH THESE IMAGES)

Initial Prototype August 93 and January 94

Cornell's first pilot study had the student provide only a single dissertation copy on standard-quality paper. This copy was scanned locally into TIFF format and then sent to UMI. One-way postage was sufficient, since this paper copy need not be returned. Using our DocuTech printer and the scanned image, Cornell produced the two library copies on acid-free paper. Rough calculations show that the cost to the student of the on-campus scanning-based approach will be no greater than for the present model, and will likely be lower. See Appendix 1 for cost comparisons in the two models. See Appendix 2 for a diagram of the proposed process at Cornell.

The scanned images of all dissertations are suitable for storage and retrieval using any of a variety of mechanisms, and we are experimenting with several. Since dissertations become part of our library collection, we are already producing the relevant MARC records, and some link from MARC record to stored image is desirable. In addition, as part of this project, we required students to submit their abstracts in electronic (text only) form, which are WAIS indexed and linked to the dissertation images. Since UMI today requires that students produce separate paper copies of their abstracts, and since nearly all students use some form of word processing software for at least part of their dissertations, the extra requirement of a machine-readable abstract did not present a major problem. We orginally planned to look at ways of acquiring in electronic form such dissertation components as tables of contents and bibliographies, but have not done that yet.

The initial pilot project was and is still being run with threegraduate fields in the Engineering School- Civil and Environemntal Engineering, Computer Science and Electrical Engineering. We have written letters of support from each departments graduate field representative. These advisors provide the

faculty/committee chair perspective as we develop the procedures for the submission of the dissertation in the test project.

Proposed procedure for May, 1994 dissertations

- Student prepares final dissertation in form acceptable to Grad School. This may include several reviews of the dissertation by the Thesis Advisor, correcting format errors. Student brings final copy of Dissertation, printed single-sided on standard-quality paper in box to Grad School. As is the case today, the box contains a \$35 check if the student has opted to have UMI do copyright registration.

- At the same time, student sends (via e-mail) the text of the abstract to a mailing list including John Saylor and Library Cataloging Services. [There are two alternatives for having the student's completion of this requirement certified. 1: John acknowledges receipt with an e-mail form, which the student prints and brings along with the dissertation to the Grad School. 2: The Grad School is included among the recipients of the abstract and upon receipt the Grad School checks off an appropriate box in the student's file.]

- Assisted (and corrected) by Thesis Advisor, student fills out CIT copy specs form, which must specify at least 2 copies printed two-sided on archival paper. The total price for these two copies is \$0.20/page (where "page" refers to an original single-sided page). For example, the typical 200-page dissertation would cost \$40 for the two copies. (By comparison, the current requirement to provide 2 single-sided copies on archival paper costs \$60 at most local copy shops.) In addition, the student may order any number of additional copies at standard CIT "EZ-Publish" rates (currently \$0.025/impression plus \$0.03/page if archival paper is desired). A single \$10 handling fee will be added to the total.

- The total amount from the CIT copy specs form is entered (by student? by Thesis Advisor) on the Cashier coupon, and added to the \$85 Thesis Fee to produce a grand total to be paid to Cornell. CIT c/s form is put into box with dissertation. The box remains in the Grad School until...

- Student pays fees at Bursar Grad School where Thesis Advisor completes processing. Thesis Advisor verifies that payment amount matches specs sheet, signs specs sheet, puts spec sheet back in box, tapes box shut. If the student has not ordered extra copies of the dissertation, student is done; Grad School sends batches of no-extra-copy dissertations to CIT [how? on what schedule?]. - If the student has ordered extra copies of the dissertation, student brings boxed dissertation to CCC. CIT person opens box, takes specs form and compares it to student's copy of Bursar receipt. The CIT person tells the student when and how to return for those copies.

- CIT scans dissertation. Digital form gets put <where?>

- Original (in box, which might include \$35 check) is put on shelf awaiting pickup by Challenge. Challenge processing is exactly as today, except that Challenge must shred the original copy after UMI has confirmed satisfactory microfilming.

- CIT prints copies of dissertation based on spec sheet. Extra copies for student are retained for pickup.

- Two archival copies are held by CIT for pickup by library courier.

- Library Cataloging Services receives two copies of dissertations simultaneously and performs the same processing as today, including having the copies bound, shelved, and cataloged. In addition, Library Cataloging Services will have received the abstract via e-mail, and will merge this into the catalog entry.

Things that need to be done before the proposed procedure for May, 1994 dissertations can be put in place

- Grad School rewrites 3-Dept version of the "PhD Dissertation Submission Process"

- CIT prepares form on which specs for copies are written. This form also has a line for Grad School approval.

- Grad School produces special version of yellow payment coupon with space (on each section) to indicate total copying charge (from new form mentioned above).

- Note: The above material constitutes (part of) a special info packet for students in the 3 designated departments
- CIT makes arrangements w/Challenge for regular (on-demand?) pickup

Issues/questions/concerns

- Are we worried that the student could turn in for copying something other than what was approved?
- It would be nice to be able to mail dissertation copies to the student; how do we get that paid for? If we do find a way to provide for this, we can allow extra-dissertation-copy students to leave their dissertations with the grad school rather than taking them to CCC.
- Where does digital form of dissertation go?
- Now that we have an extra single-sided copy of the dissertation, we should re-examine the possibility of sending this to UMI, rather than doing the microfilming locally
- Now that we have an ascii form of the abstract, we should see if UMI will "pay" anything to avoid having to key it themselves.
- We need to check whether the Grad School is willing to use option 2 for certifying student submission of abstract

Utopia

Utopia is a digital resource, in development at Cornell University, designed to support undergraduate use of images from the Renaissance period. Included in Utopia are over 4500 digital surrogates of artifacts held in the collections of the Johnson Art Museum, the Division of Rare and Manuscript Collections, and the slide libraries in the Department of Art History and the College of Art, Architecture, and Planning. Original cataloging information is a key element of this resource.

??All the images for Utopia were photographed if they weren't already photographed and then scanned to PhotoCD by Boston Photo. Derivatives have gone through a variety of iterations but I'm guessing you aren't interested in that. It's not clear to me who is in charge of Utopia at this point, probably Peter H.?
Noni