

Opportunities for Collaboration:

The HEARTH Project



Joy Paulson and
Nathan Rupp

Cornell University



Digital Library Federation
Spring Forum

New Orleans, Louisiana
April 2004

Outline

- Introduction
- Metadata
- Tying metadata together

The HEARTH Project

- Home Economics Archive: Research, Tradition, History
- Subjects covered
- Bibliography development
- Reformatting
- Technical details
- Funding

Vendor Metadata

TIFF headers contain:

- Bibliographic information
- Descriptive information
- Date and time of scan
- Source description
- Image description

Image Description

- Monographs and serials
 - Shipment number
 - ID number
 - Image sequence number (padded to 8 digits)
 - Technical description information
- Serials only
 - Volume and issue number

Image Description

Image width

Resolution unit

Image length

Bits per sample

X resolution

Compression

X position

Orientation

Y resolution

Y position

Descriptive Metadata: Traditional Cataloging

- All titles included in OPAC
 - Titles available at Cornell
 - Titles from other libraries supporting home economics
- MARC 856 field

Voyager Cataloging - [Bib 4400471 : Syllabus and laboratory manual of household chemistry [electronic resource]]

File Edit Record Display Options Window Help

New Open Save Search Headings Titles Save to DB Save Plus New Hldgs Get Hldgs New Items Get Items Hierarchy Help

MARC System History

Leader 01400cam 22002891a 4500 006 m: d

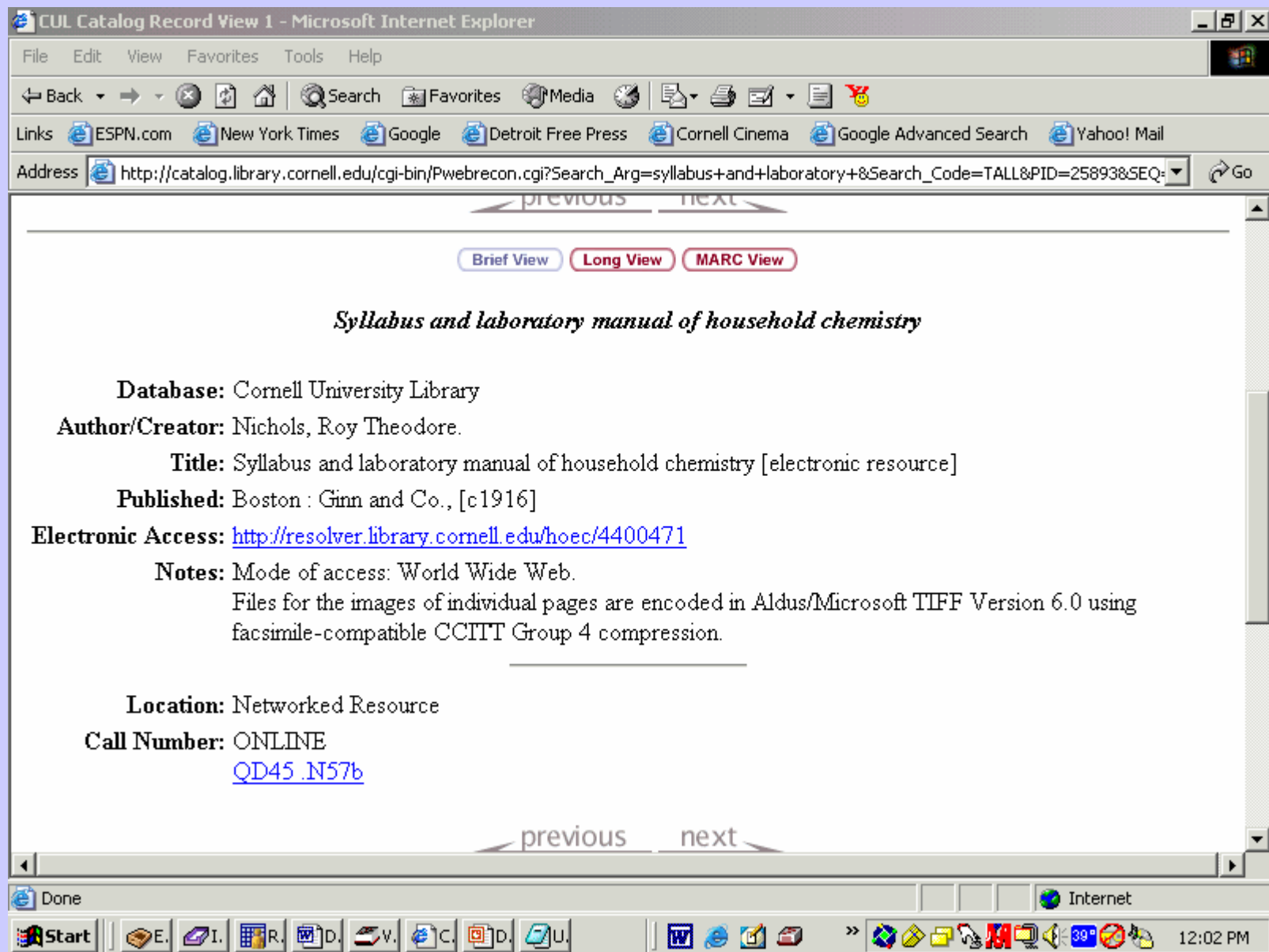
005: 20040206110444.0 007 c:r bn 001 apadp

008 830630 s 1916 ___ mau a _ s _ 0 0 0 _ 0 _ eng _ d

Tag	I1	I2	Subfield Data
→ 040			#a NIC #c NIC
050		4	#a QD45
100	1		#a Nichols, Roy Theodore.
245	1	0	#a Syllabus and laboratory manual of household chemistry #h [electronic resource]
260			#a Boston : #b Ginn and Co., #c [c1916]
440		0	#a Home economics archive--research, tradition and history
538			#a Mode of access: World Wide Web.
538			#a Files for the images of individual pages are encoded in Aldus/Microsoft TIFF Version 6.0 using facsimile-compatible CCITT Group 4 compression.
500			#a Digitization funded by Institute of Museum and Library Services, 2001-2003. Preserving the Core historical literature of home economics before 1950.
533			#a Electronic reproduction. #b Ithaca, N.Y. : #c Cornell University, A.R. Mann Library, Preservation Unit ; #b Montreal, Qu'ebec : #c Trigonix Inc., #d 2002. #e 121 image files.
650		0	#a Chemistry #x Laboratory manuals.
650		0	#a Sanitary chemistry.

Cataloging Source 11:59 AM

Start E. I. R. D. V. Y. D. U. W. 11:59 AM

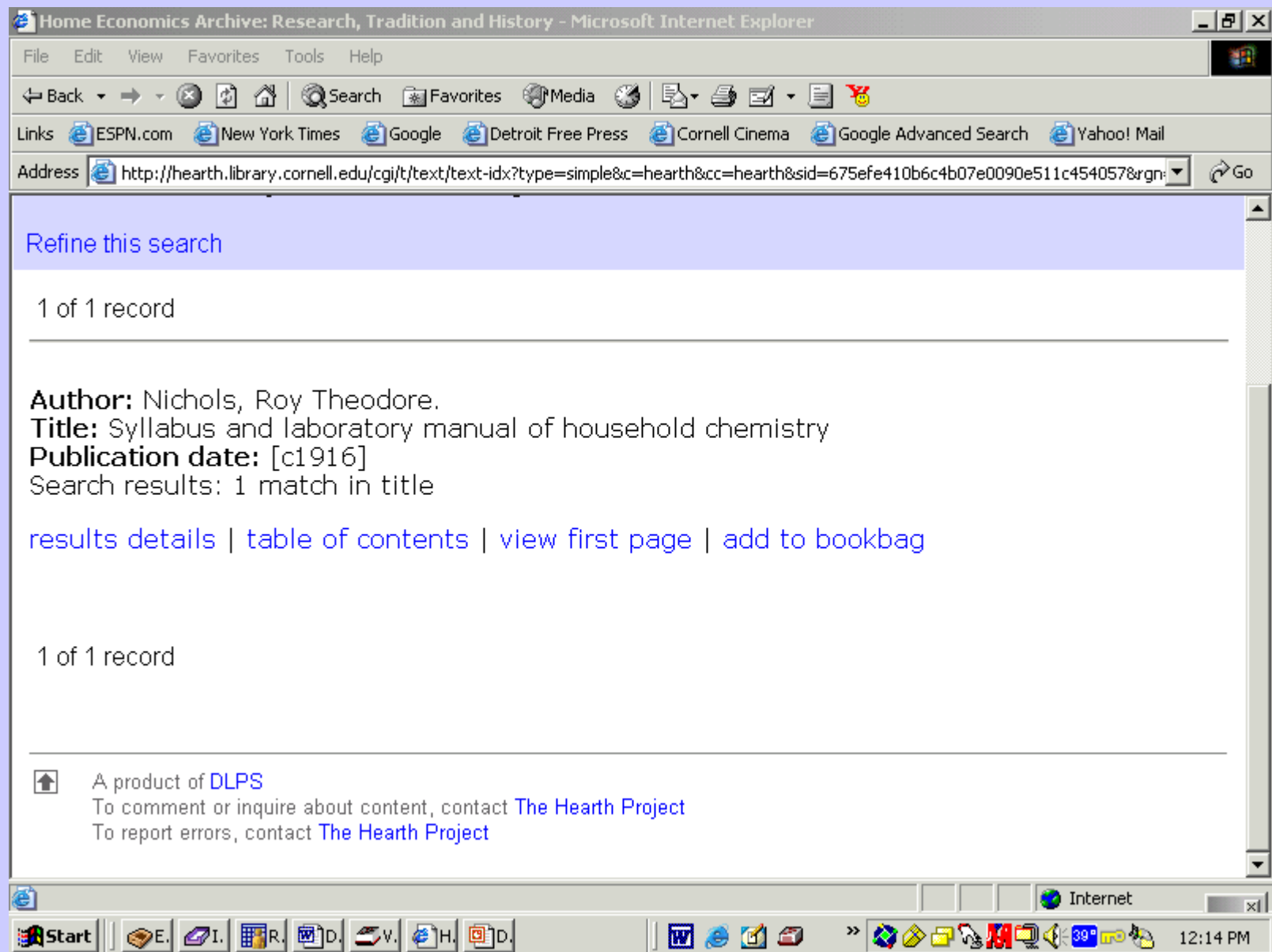


Descriptive Metadata: Repurposing MARC

- Descriptive metadata from MARC records
- MARC to TEI Lite conversion scheme
 - 100ad → Author
 - 300a → Pagination
 - 245ab → Title
 - 6XX → Keywords
 - 260abc → Publication
- Problems
 - Dates
 - Edition statements

```
Untitled - Notepad
File Edit Format Help

<HEADER>
<FILEDESC>
  <TITLESMT>
    <TITLE TYPE="245">Syllabus and laboratory manual of household chemistry</TITLE>
    <AUTHOR>Nichols, Roy Theodore. </AUTHOR>
  </TITLESMT>
  <EXTENT>121 600dpi TIFF page images</EXTENT>
  <PUBLICATIONSMT>
    <PUBLISHER>Cornell University, Mann Library</PUBLISHER>
    <PUBPLACE>Ithaca, New York</PUBPLACE>
    <IDNO TYPE="dlps">4400471</IDNO>
  </PUBLICATIONSMT>
  <NOTESMT>
    <NOTE> Digitization funded by Institute of Museum and Library Services, 2001-2003. Preserving
  </NOTESMT>
  <SOURCEDESC>
    <BIBL>
      <TITLE TYPE="main">Syllabus and laboratory manual of household chemistry</TITLE>
      <AUTHOR>Nichols, Roy Theodore. </AUTHOR>
      <PUBLISHER>Ginn and Co.,</PUBLISHER>
      <PUBPLACE>Boston :</PUBPLACE>
      <DATE>[c1916]</DATE>
    </BIBL>
  </SOURCEDESC>
</FILEDESC>
<ENCODINGDESC>
  <EDITORIALDECL N="2">
    <P>This electronic text was created by Optical Character Recognition (OCR). No corrections have
  </EDITORIALDECL>
</ENCODINGDESC>
<PROFILEDESC>
  <TEXTCLASS>
    <KEYWORDS>
      <TERM>Chemistry</TERM>
      <TERM>Sanitary chemistry.</TERM>
      <TERM>Laboratory manuals.</TERM>
    </KEYWORDS>
  </TEXTCLASS>
</PROFILEDESC>
```



Structural Metadata

Document Structuring Tool

- Designed in house
- Provides project management
- Manages copyright information
- Structures the document
- Performs quality control of images

Administrative System (1)

HEARTH Admin: Users - Netscape

File Edit View Go Bookmarks Tools Window Help

http://hearth.mannlib.cornell.edu/admin/users/ Search

Mail AIM Home Radio My Netscape Search Shop Bookmarks TheFreeSite.c... Netscape.com

[jp243](#) [logout](#)

HEARTH

[Home](#)

[Review Images](#)

[Publications](#)

[Administer Users](#)

[Collections](#)

HEARTH Users

Create New User View Inactive Users

ADMINISTRATOR

login	name	e-mail	expiration		
frances	Frances Webb	fbw4@cornell.edu	2004-11-16	Edit	Deactivate
jlh53	Johanna Hartnagel	jlh53@cornell.edu	2004-11-11	Edit	Deactivate
jp243	Joy Paulson	jp243@cornell.edu	2004-12-30	Edit	Deactivate*

USER

login	name	e-mail	expiration		
bad1	Barbara DiSalvo	bad1@cornell.edu	2005-02-23	Edit	Deactivate
bjl23	Brian J Lowe	bjl23@cornell.edu	2005-03-01	Edit	Deactivate
cag43	Corey Grassl	cag43@cornell.edu	2005-04-07	Edit	Deactivate

Document: Done (0.594 secs)

Start Joy Pauls... In-tray Eudora ... HEARTH ... Microsoft ... untitled - ... 10:45 AM

Administrative System (2)

HEARTH Publications - Netscape

File Edit View Go Bookmarks Tools Window Help

http://hearth.mannlib.cornell.edu/admin/pubs/ Search

Mail AIM Home Radio My Netscape Search Shop Bookmarks TheFreeSite.c... Netscape.com

[jp243](#) [logout](#)

HEARTH

[Home](#)

[Review Images](#)

[Publications](#)

[Administer Users](#)

[Collections](#)

Cottage economy, containing&

< Last Next >

Save Changes Reset

Publication ID: 715

Title: Cottage economy, containing&

Author: Cobbett, William

Pub. Type: MONOGRAPH

Dates:

Shipment: Shipment 10 or create new shipment: ☐

Call No.: S511.C65 1926

Voyid: 4391519

Voyidold: 486282

Review: 0 volumes (pages) are fully reviewed.
0 volumes (pages) are being worked on.
1 volumes (228 pages) are not yet being worked on.
[View full review history](#)

Copyright: OK

Document: Done (0.313 secs)

Start Joy Pauls... In-tray Eudora - ... HEARTH ... Microsoft ... 11:02 AM

Administrative System (3)

HEARTH Publications - Netscape

File Edit View Go Bookmarks Tools Window Help

http://hearth.mannlib.cornell.edu/admin/pubs/ Search

Mail AIM Home Radio My Netscape Search Shop Bookmarks TheFreeSite.c... Netscape.com

[jp243](#) [logout](#)

HEARTH

[Home](#)

[Review Images](#)

[Publications](#)

[Administer Users](#)

[Collections](#)

Review Status: Journal of Home Economics

Reviewed:

Volume	Pages	Reviewer	
ABF4553-0001	476	Johanna Hartnagel < jlh53@cornell.edu >	Reset Status

Date	Action	User	
Fri Jul 26, 08:26:24 AM	Open for Review	Scott Wiessinger < smw22@cornell.edu >	
Fri Jul 26, 12:29:36 PM	Save Changes	Scott Wiessinger < smw22@cornell.edu >	
Fri Jul 26, 12:59:44 PM	Save Changes	Scott Wiessinger < smw22@cornell.edu >	
Fri Jul 26, 03:12:34 PM	Save Changes	Scott Wiessinger < smw22@cornell.edu >	
Fri Jul 26, 03:15:04 PM	Save Changes	Scott Wiessinger < smw22@cornell.edu >	
Mon Jul 29, 01:37:08 PM	Save Changes	Scott Wiessinger < smw22@cornell.edu >	
Mon Aug 5, 02:54:30 PM	Save Changes	Scott Wiessinger < smw22@cornell.edu >	
Fri Dec 13, 12:05:37 PM	Unassign	Scott Wiessinger < smw22@cornell.edu >	by Shawn Milrad
Fri Dec 13, 12:06:19 PM	Open for Review	Shawn Milrad < smm43@cornell.edu >	
Thu Feb 20, 01:25:39 PM	Save Changes	Shawn Milrad < smm43@cornell.edu >	
Fri May 23, 11:34:20 AM	Unassign	Shawn Milrad < smm43@cornell.edu >	by Shawn Milrad
Fri May 23, 11:34:20 AM	Assign	Johanna Hartnagel < jlh53@cornell.edu >	by Shawn Milrad
Wed Aug 6, 09:20:24 AM	Save Changes	Johanna Hartnagel < jlh53@cornell.edu >	

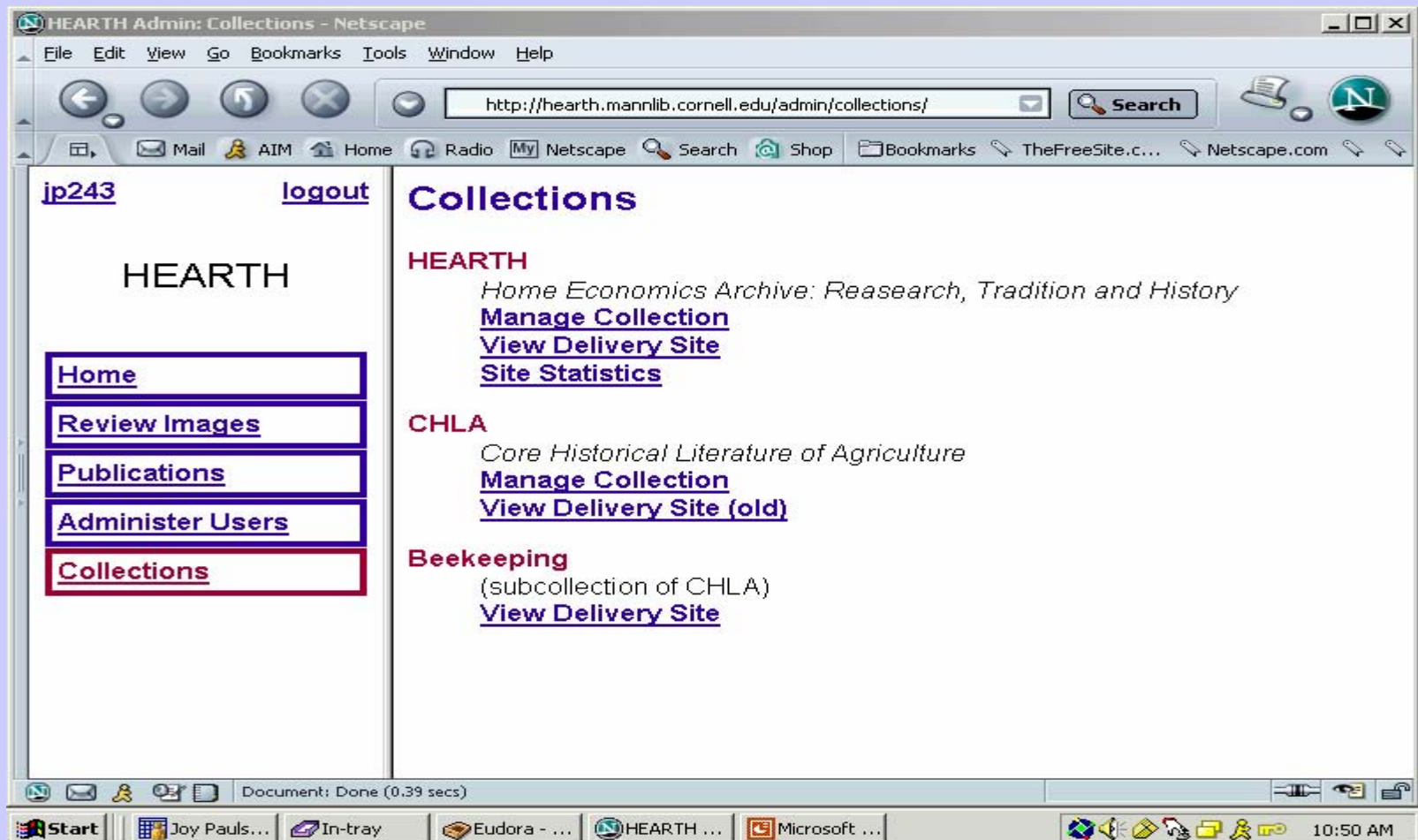
mailto:Shawn Milrad <smm43@cornell.edu>

Start Joy Pauls... In-tray Eudora - ... HEARTH ... Microsoft ... 10:53 AM

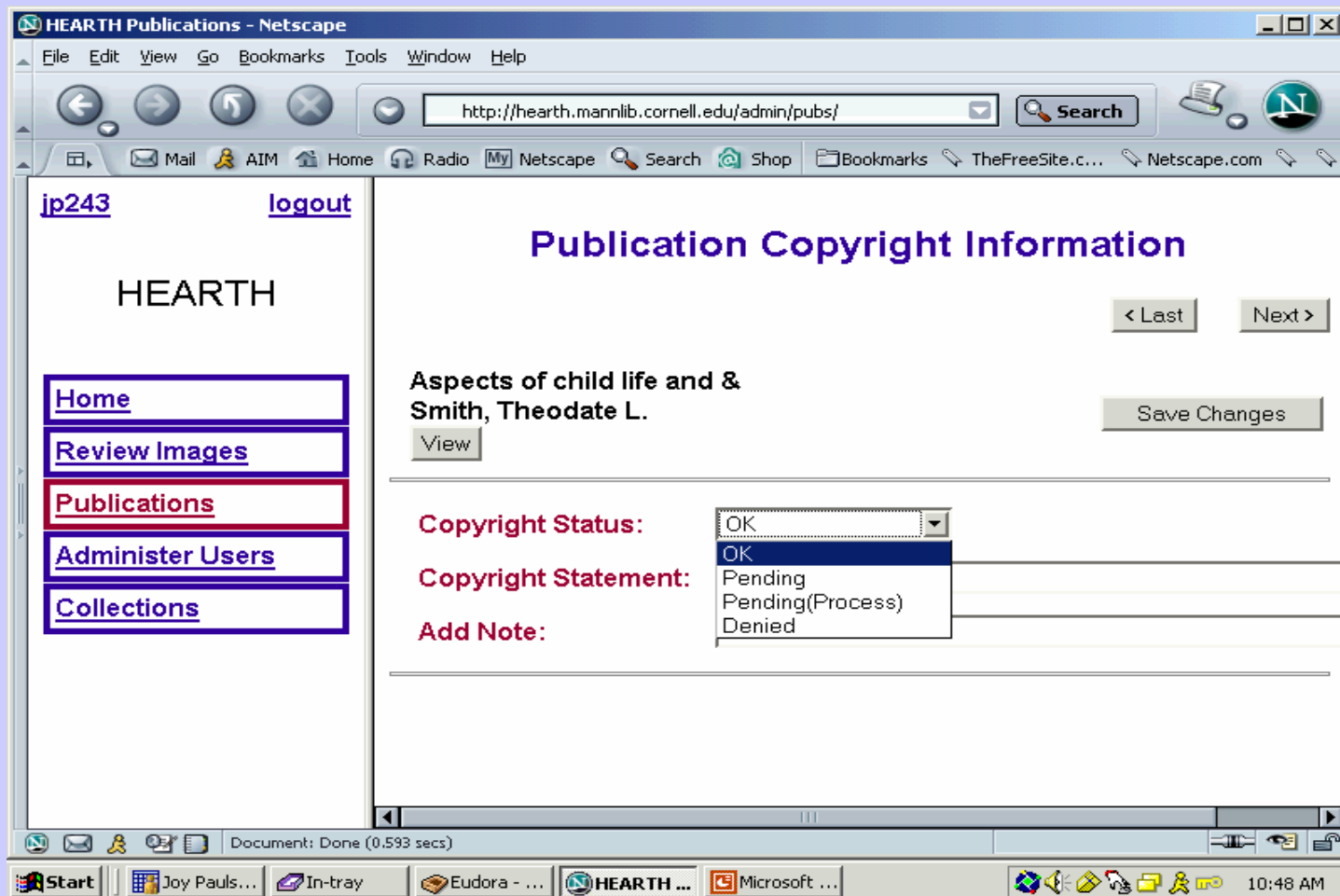
1/26/2007

16

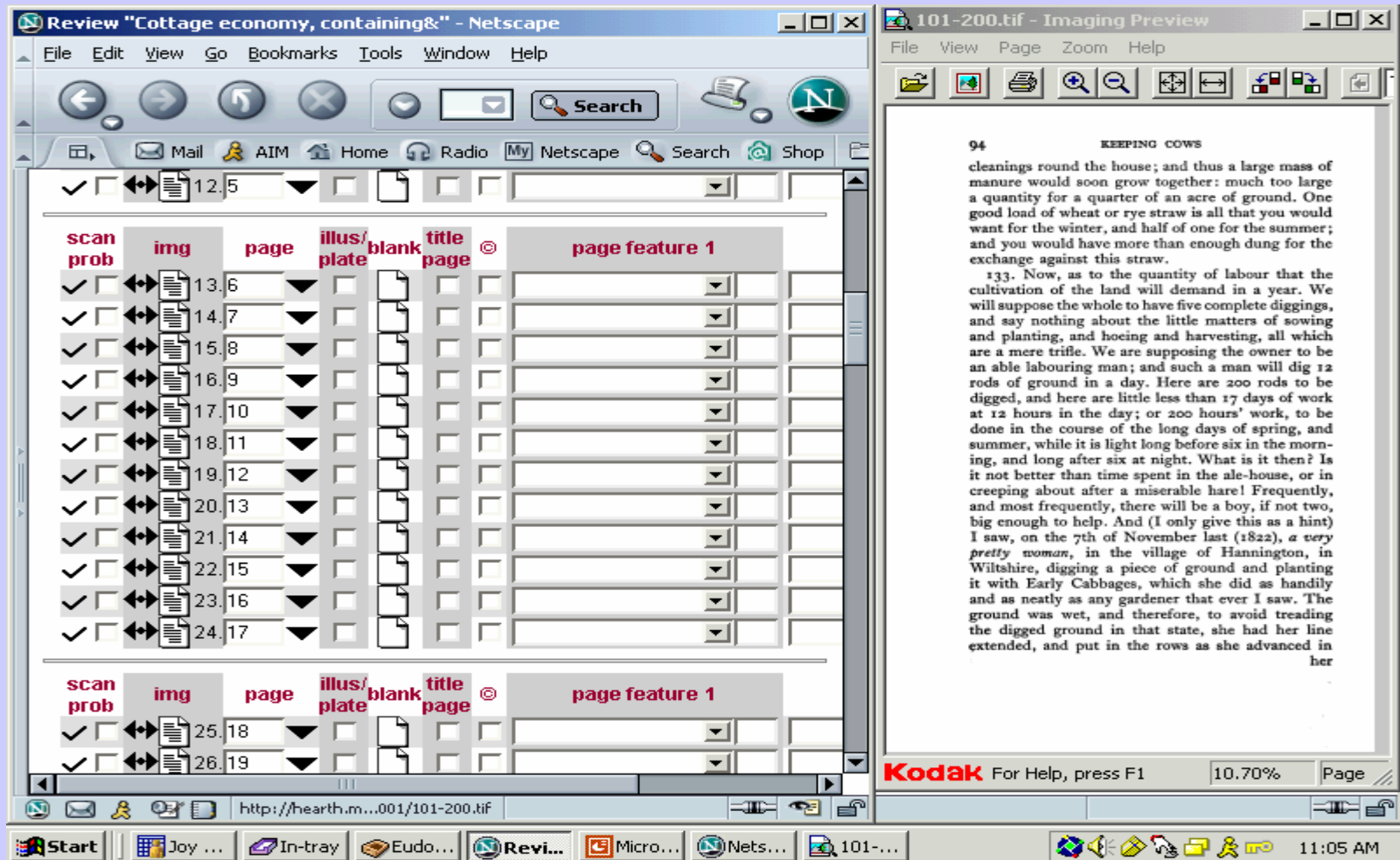
Administrative System (4)



Copyright Management



Structuring Documents (1)



Structuring Documents (2)

- Quality control of images
- Match page numbers to image sequence
- Highlight important structures:

Title pages

Indexes

Table of contents

Illustrations

Other tables

Bibliographies

Front matter

Back matter

Notes

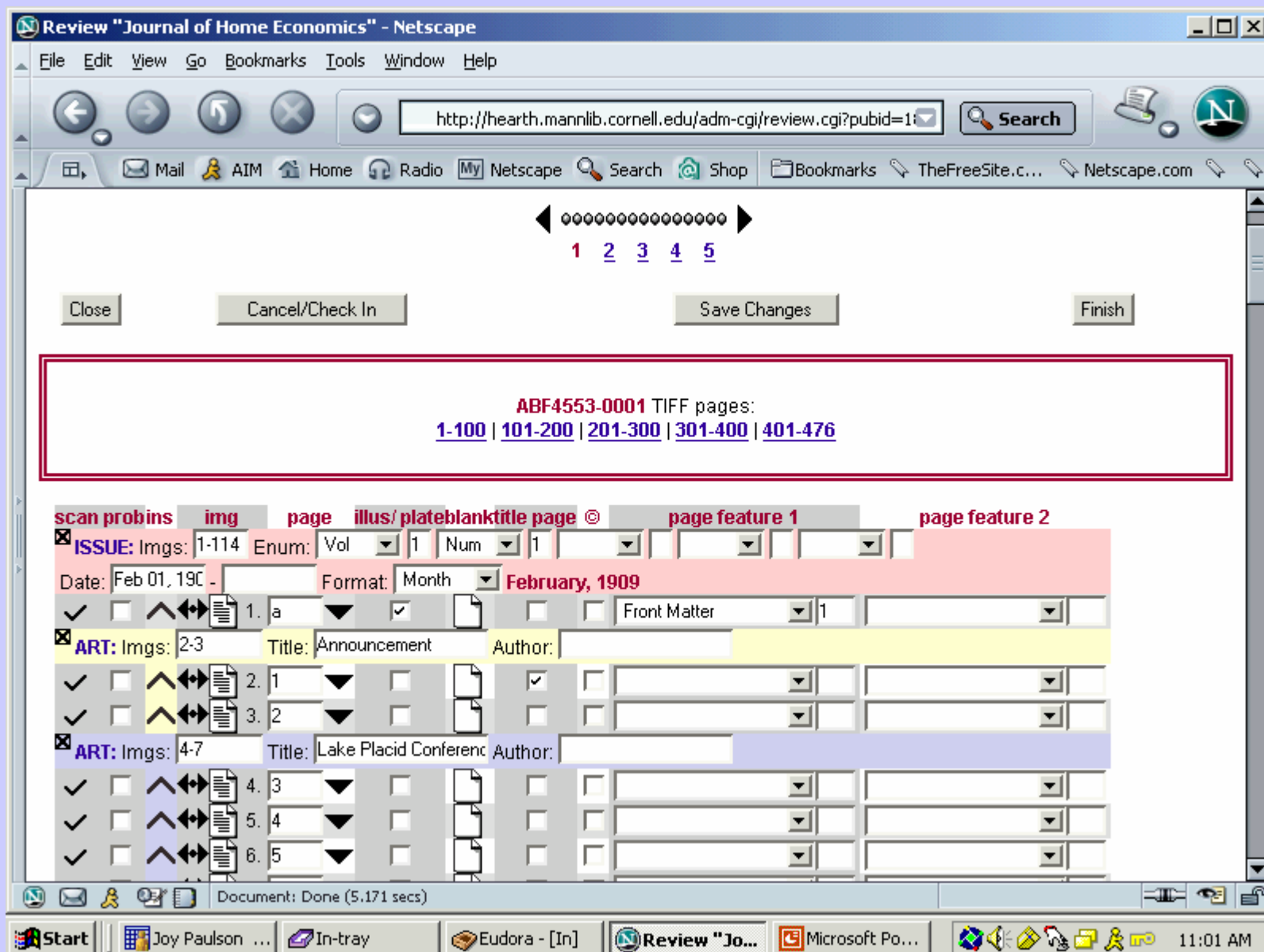
Errata

Providing Access

- Article level
 - Browse by author
 - Search for author or title
- Volume level
 - Structuring Process

Other Functionalities

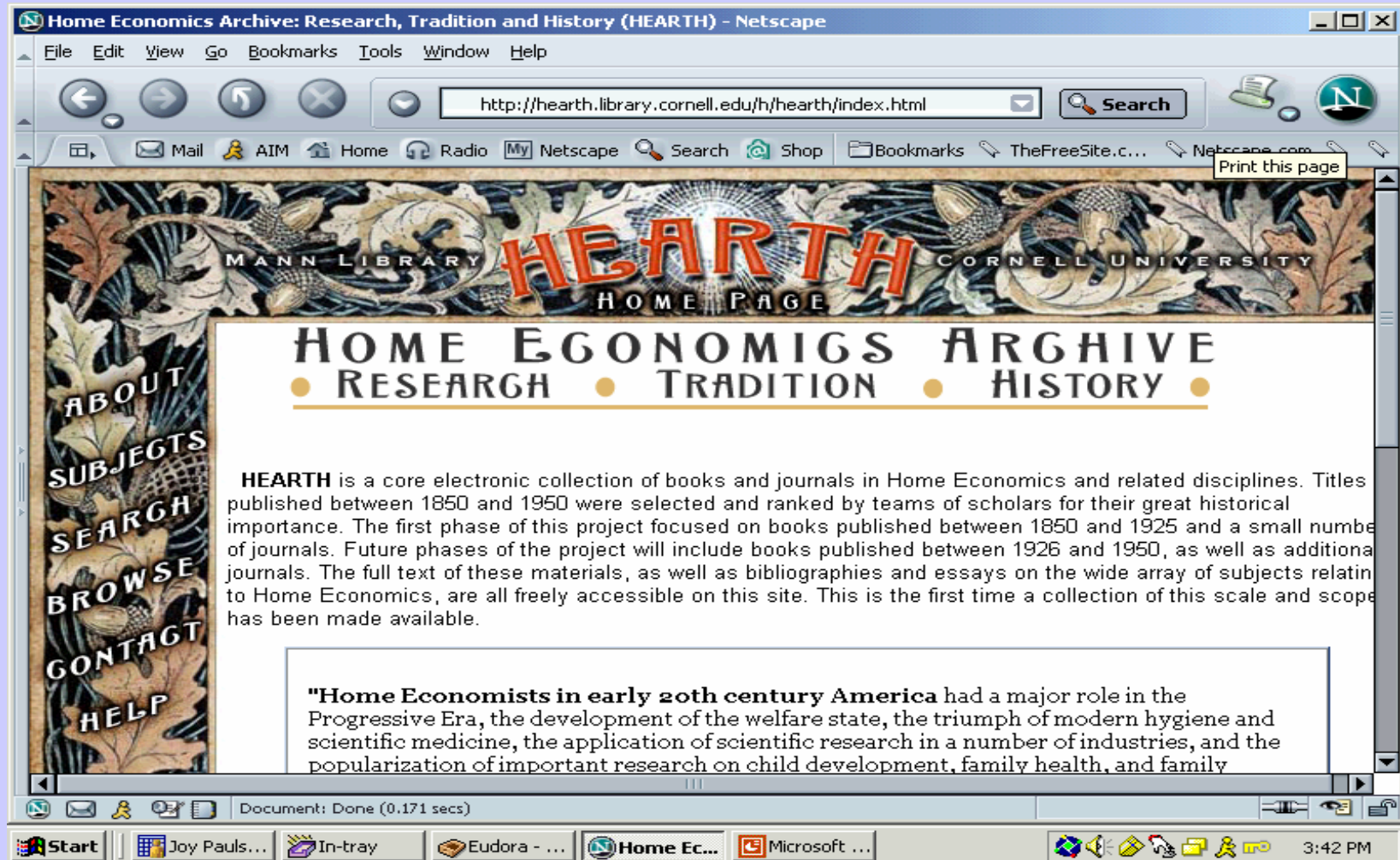
- Insert blank pages
- Insert pages missing note
- Reorder pages out of sequence
- Mark pages with scanning problems—
description placed in notes field



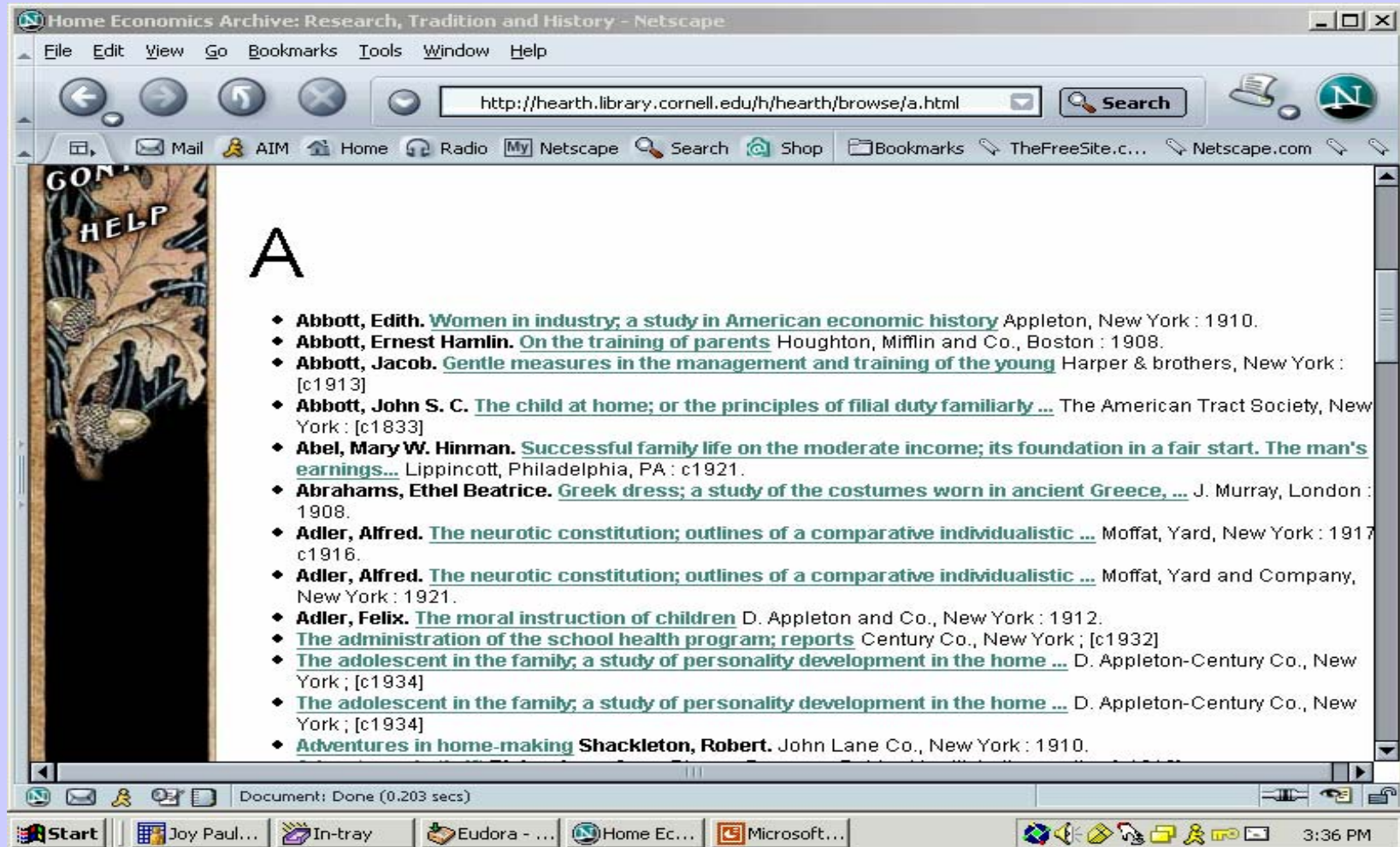
Tying It All Together

- Vendor-supplied metadata
- Descriptive (MARC) metadata
- Preservation metadata
- Structural metadata
- MARC record information
- OCR
- Images on server

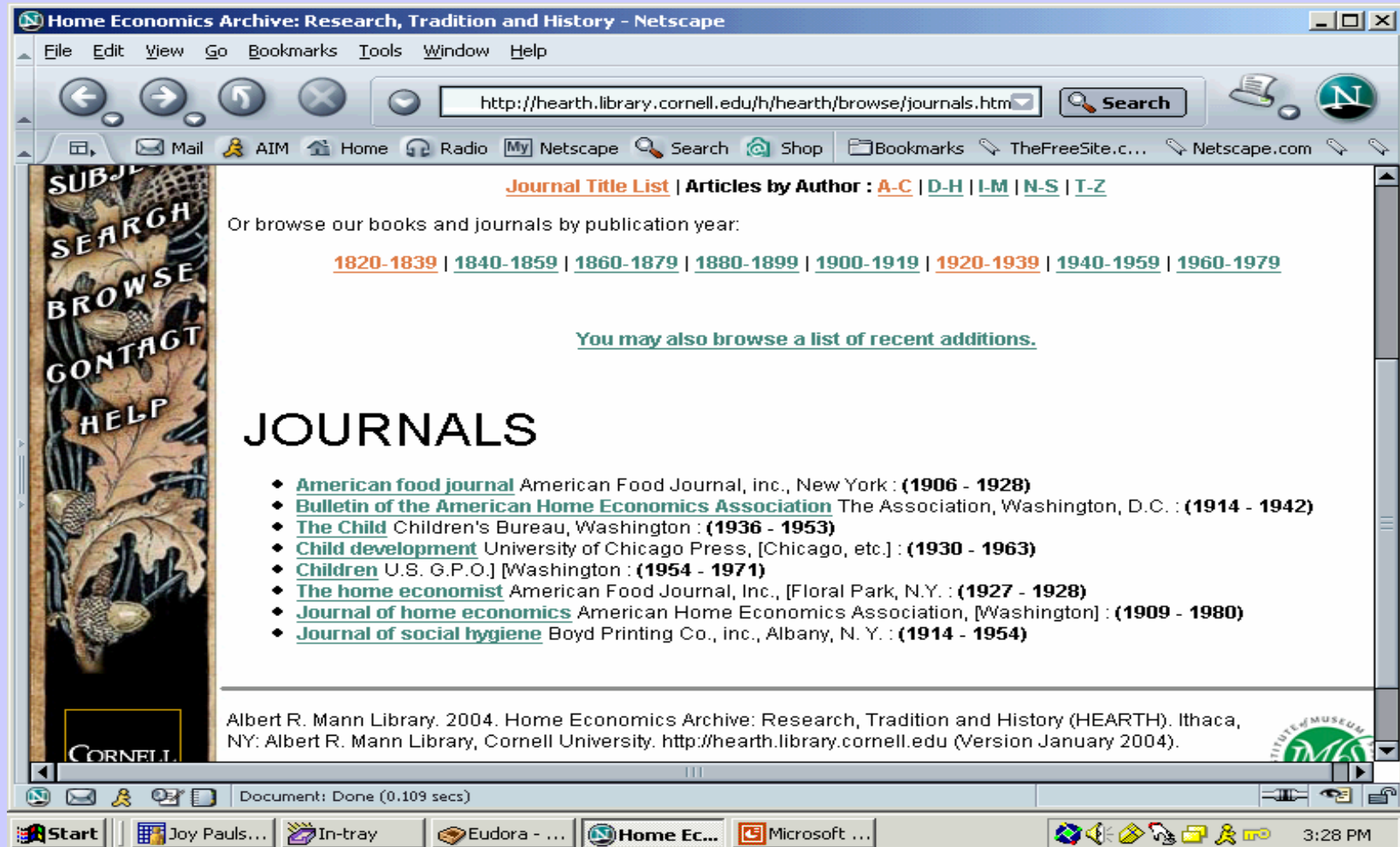
HEARTH



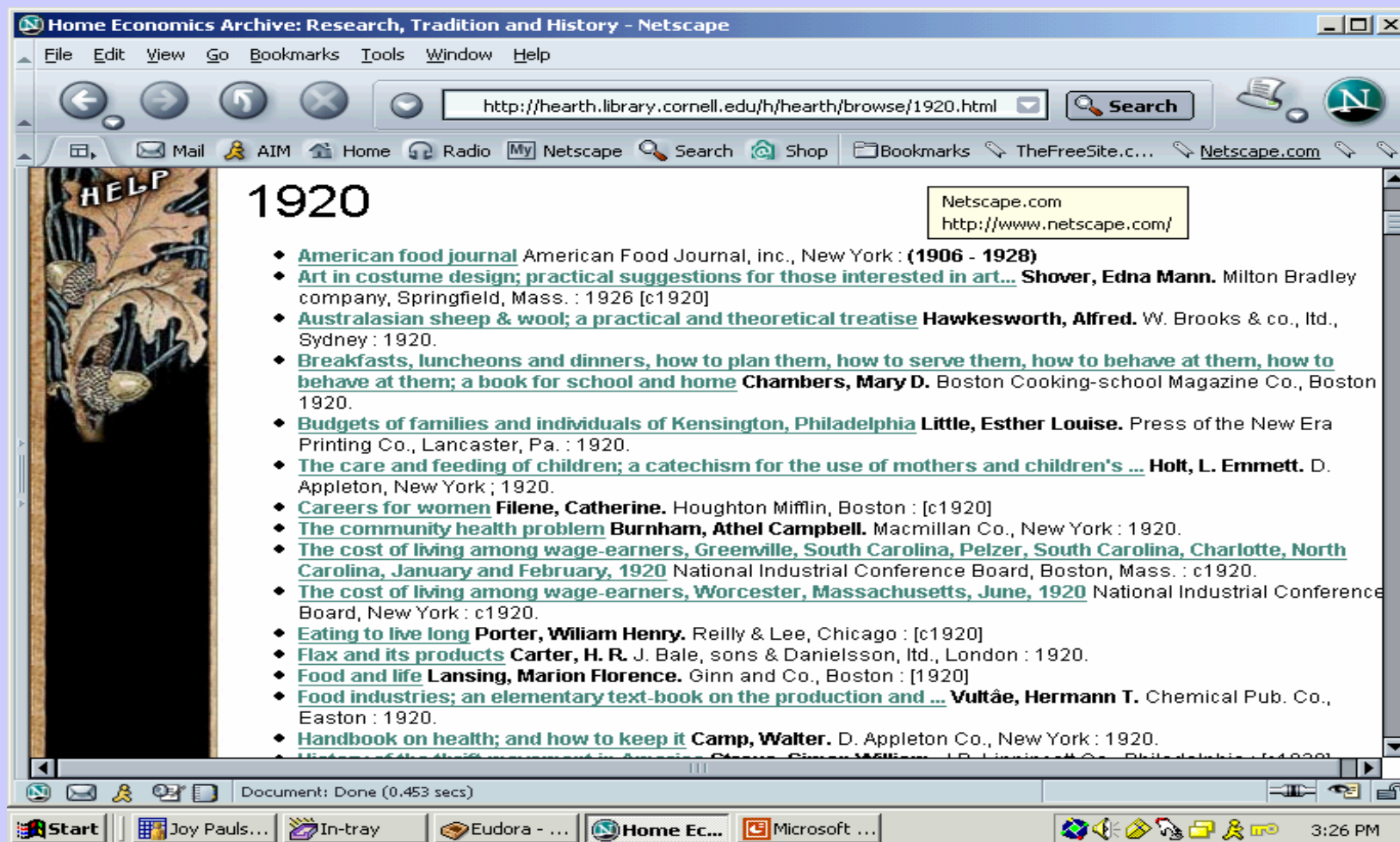
Browse by Author or Title



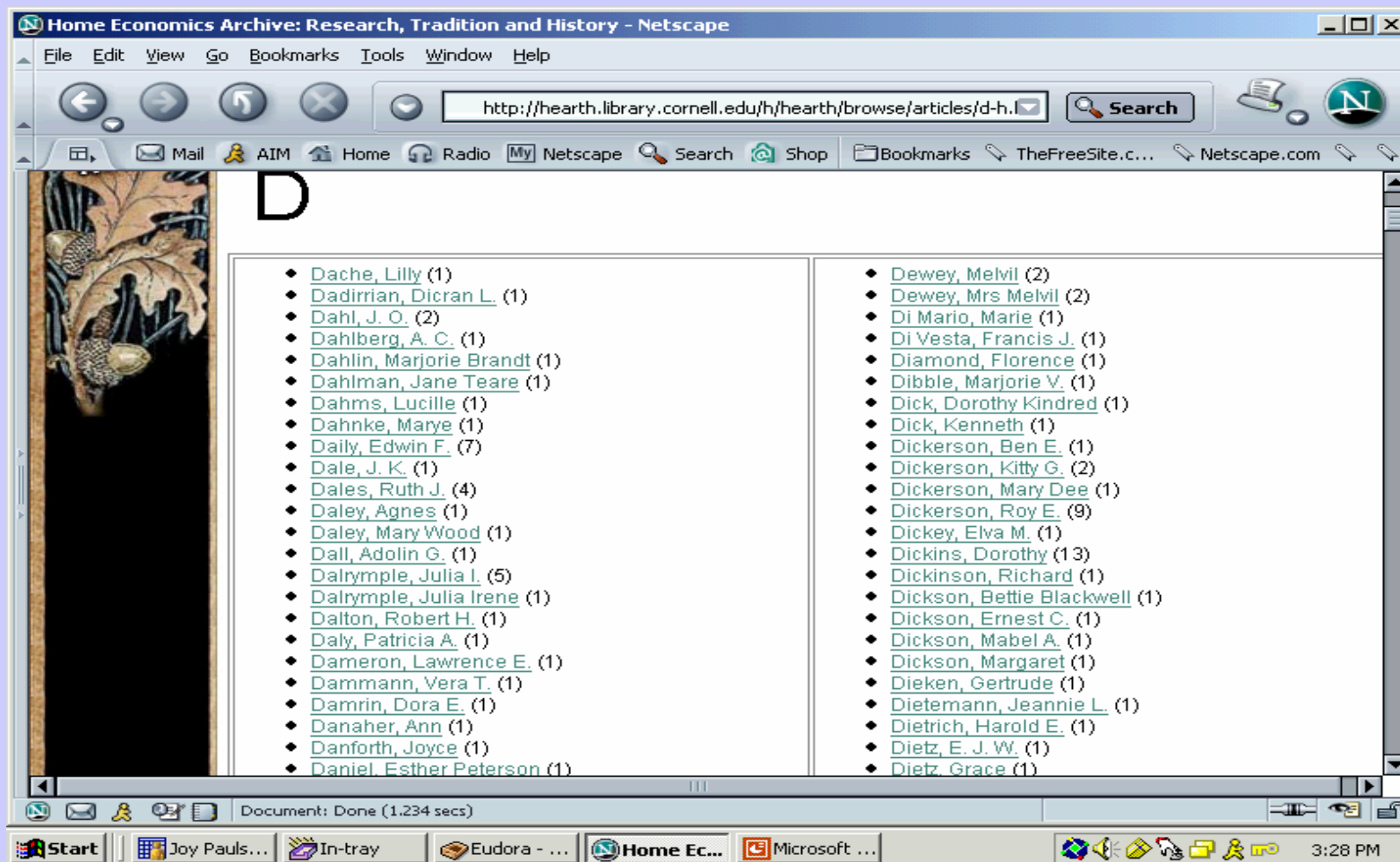
Browse Journals Page



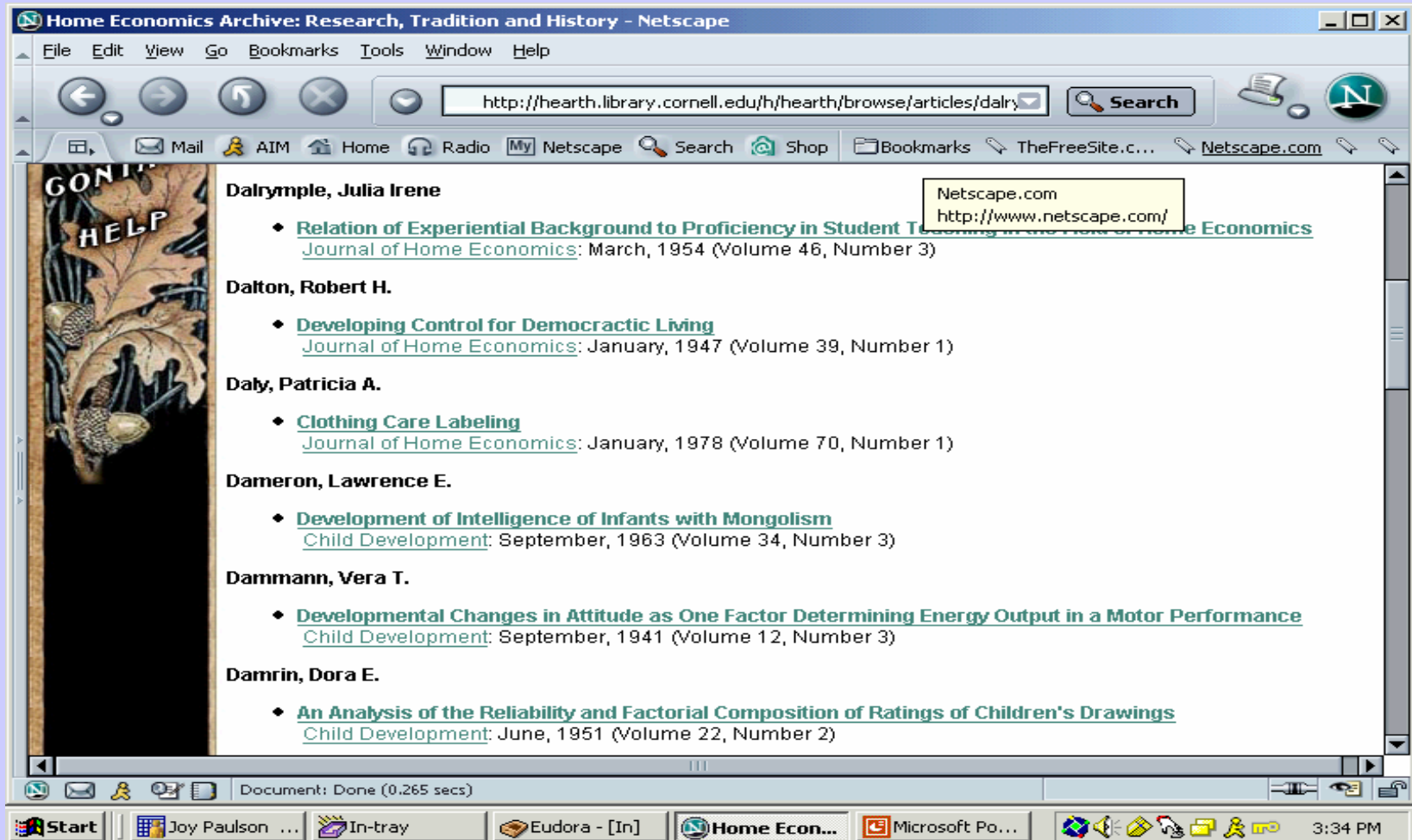
Journals and Monographs by Year



Article Author Browse Page (1)



Article Author Browse Page (2)



Questions?

Joy Paulson

(607) 255-7950

jp243@cornell.edu

Nathan Rupp

(607) 255-7943

nar25@cornell.edu