

Contextualizing Subject Access Across Digital Collections

The "See Also" Problem

Joseph B. Dalton
The New York Public Library
Research Libraries
Digital Library Program

DLF Fall Forum 2006 - Nov. 9, 2006

What We'll Cover



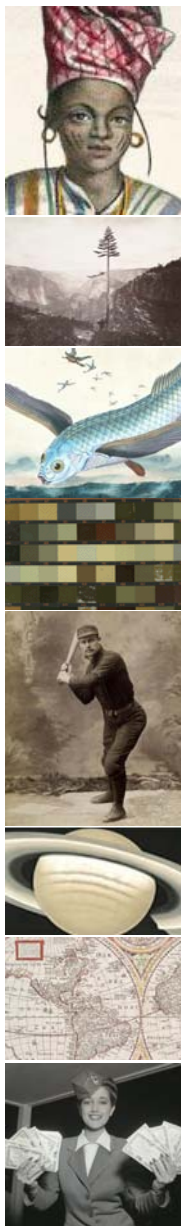
- Overview of Problem
- Some Approaches
- Expanding Subject-Access
- Problems, Opportunities and Challenges

Background: Some Numbers



- **NYPL Digital Gallery collections**
 - 524,000 images
 - 318,000 bibliographic (item-level) records
- **NYPL Digital Library Program uses several subject thesauri**
 - LCTGM, LCSH, LCNAF, AAT, GMGPC, etc.
- **Number of records containing at least 1 subject heading:**
 - 260,000 (81 % of total)

The Problem



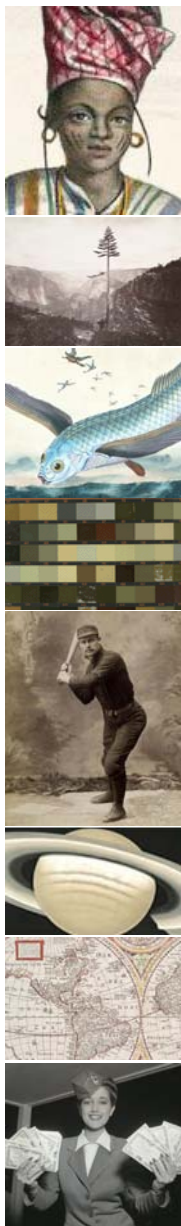
- **58,000 subject headings indexed for searching NYPL Digital Gallery**
- **Browsing a list this size is like trying to find a needle in a haystack**
 - ❑ **Lincoln:**
‘Lincoln, Abraham’
 - ❑ **Posters depicting circuses:**
‘Posters - - Circus’ or ‘Circus Posters’?

Possible Approaches for Subject Browsing



- **Parse or index list by subjects' facets**
 - 19th century - French - Posters
 - French - Posters - 19th century
 - Posters - French - 19th century
- **Map subjects to selected thesauri and provide cross-references**
- **Build hierarchical browsing on top of some taxonomy**
 - **Printing & Graphics** > Circus Posters
 - Culture & Society** > Posters - French - 19th century

Our Approach



- Create separate index of subjects in Lucene
- Index the pointers from those subjects to their associated objects or containers
- Provide front-end context lists, like “subjects in [Collection|Title],” by filtering on the objects
- Provide free-text retrieval of subjects through some kind of “subject finder”

Indexing Field Relationships at Object-Level



- **First Goal: provide object-based lists**
 - **Gather all subjects and some associated bibliographic identifiers: item-level, title-level, collection-level, etc.**

query					
	HITS	IMAGE_ID	PARENT_ID	STRUC_ID	SUBJECT
1	176			306136	Book jackets Quaker Oats Company
2	176			569411	Quakers Physicians
3	176			569413	Quakers Abolitionists
4	176			604621	Abolitionists Women's rights -- United States Mott, Lucretia, 1793-1880 Quake
5	176			580961	Quakers Woolman, John, 1720-1772
6	176			610436	Reformers Abolitionists Antislavery movements -- United States Mott, Lucretia,
7	176			138222	Baseball Philadelphia Quakers (Baseball team)
8	176			138226	Baseball Philadelphia Quakers (Baseball team)
9	176			138228	Baseball Philadelphia Quakers (Baseball team)
10	176			138230	Baseball Philadelphia Quakers (Baseball team)

Subject Display by Parent-Title Object

Browse Source

1-12 of 26 items | Page 1 of 3

1 2 3

< Previous | **Next** >

Source: Catlin's North American Indian portfolio. Hunting scenes and amusements of the Rocky mountains and prairies of America.

<< Back to thumbnail images

☒ This Source ☐ All Gallery Images

- ☐ Animal behavior
- ☐ Antelopes
- ☐ Archery
- ☐ Assiniboine Indians
- ☐ Bear dance
- ☐ Bear hunting
- ☐ Bison
- ☐ Bison hunting
- ☐ Buffalo dance
- ☐ Disguise
- ☐ Horses
- ☐ Indian dance -- North America
- ☐ Indians -- Kings and rulers
- ☐ Indians of North America -- Games
- ☐ Indians of North America -- Hunting
- ☐ Indians of North America -- West (U.S.)
- ☐ Mandan Indians
- ☐ Wolves

GO

[clear all]

Subject Display by Group of Collections

NYPLDigitalGallery

SEARCH: **GO**

Advanced Search | Help

Explore:
Printing & Graphics **GO**

Browse:
Names A-Z | Subjects | Libraries

My Digital:
Search History | Selections

All Collections > Explore Printing & Graphics

Advertising, book binding, dust jackets, friendship books, menus, post cards, posters, trade and greeting cards.

Search Collections

Collections | Subjects **BETA**

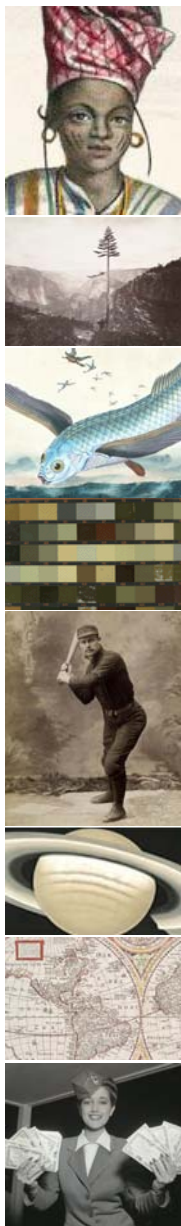
Subjects in these Collections

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z

Aaron (Biblical priest)
Abbeys
Abraham (Biblical patriarch)
Absecon (N.J.)
Absecon Lighthouse (Atlantic City, N.J.)
Abū Sunbul (Egypt)
Academic costume
Acadia -- Maps -- Early works to 1800
Acanthi
Accidents
Acoma (N.M.)
Acoma Indians -- Clothing & dress
Acorns
Acrobats
Acropolis (Athens, Greece)
Actaeon (Greek mythology)
Acting
Actors
Actors -- France
Actors -- Great Britain
Actresses
Adam (Biblical figure)
Adams, John, 1735-1826 -- Birthplace
Adelbert College
Adirondack Mountains (N.Y.)
Admirals -- German
Adobe buildings
Adobe buildings -- Arizona -- Tucson
Adobe buildings -- New Mexico
Adobe buildings -- New Mexico -- Albuquerque
Adobe buildings -- New Mexico -- Santa Fe
Adobe houses -- California -- Monterey



Indexing Subjects A-Z



- **Early Assumption:**
1 row = 1 subject 'field' per object ID
- **Reality:**
subjects are indexed as 'multiple fields'

STRUC_ID	SUBJECT
306136	Book jackets Quaker Oats Company
569411	Quakers Physicians
569413	Quakers Abolitionists
604621	Abolitionists Women's rights -- United States Mott, Lucretia, 1793-1880 Quaker women
580961	Quakers Woolman, John, 1720-1772
610436	Reformers Abolitionists Antislavery movements -- United States Mott, Lucretia, 1793-1880 Quaker women
138222	Baseball Philadelphia Quakers (Baseball team)

```

nypldg>
<structure struc="610436" image="1257540">
- <title level="2" struc="610430" catalog="b2707620" seq="1" type="parent">
  Personal reminiscences of the anti-slavery and other reforms and reformers. By Aaron M. Powell.
</title>
<title level="1" struc="610436" seq="6" type="main">Lucretia Mott.</title>
- <title type="manual_filing_title" level="2">
  Personal reminiscences of the anti-slavery and other reforms and reformers. By Aaron M. Powell.
</title>
<identifiers catalog="b2707620" bibl="" barcode="33433021591189" call="Sc 326.973A-P (Powell, A. M. Personal reminiscences)"
sref=""/>
<location division="General Research and Reference Division" center="Schomburg Center for Research in Black Culture" div_id="s
- <obj image="1257540" seq="1">
  <medium width="201" height="300">http://images.nypl.org/?id=1257540&t=r</medium>
  <small width="100" height="150">http://images.nypl.org/?id=1257540&t=t</small>
  <large width="510" height="760">http://images.nypl.org/?id=1257540&t=w</large>
</obj>
<subject thesaurus="Library of Congress Subject Headings" subjectType="Topical Subject Heading"
subjectId="88421">Reformers</subject>
<subject thesaurus="(Library of Congress) Thesaurus for Graphic Materials I: Subject Terms" subjectType="Topical Subject Heading"
subjectId="68100">Abolitionists</subject>
<subject thesaurus="Library of Congress Subject Headings" subjectType="Topical Subject Heading" subjectId="79611">Antislavery
movements -- United States</subject>
<subject thesaurus="Library of Congress Name Authority File" subjectType="Personal Name Subject" subjectId="87930">Mott, Luc
1793-1880</subject>
<subject thesaurus="Library of Congress Subject Headings" subjectType="Topical Subject Heading" subjectId="87932">Quaker
women</subject>
<physical medium="Halftone photomechanical prints" specific="Prints" parent_description="xx, 279 p. illus. 21 cm." root_descripti
279 p. illus. 21 cm."/>
<publisher date="1899" name="Caulon Press" place="Plainfield, N. J., A. R. Powell; New York"/>
<name name="Powell, Aaron M. (Aaron Macy)" nameId="21377" role="Author" dates="1832-1899" level="Additional"/>
<pagination page="126" seq="6"/>
<date type="published" level="main">1899</date>
<date type="published" level="parent">1899</date>
<date type="published" level="root">1899</date>
<collection id="147" name="[Africana & Black History]"/>
<extract_info extract_date="2/16/2006"/>
<oai_info unique_id="610436"/>
</structure>
nypldg>

```

New Opportunity: “Related Subjects”



As we examined this problem, some “opportunities” emerged:

- **Results might be expanded by each subject’s association to its “related” (item-level) subjects**
- **These results resemble a reverse-mapping of NYPL Digital Gallery subjects, derived not from an applied top-level taxonomy but from the objects’ “descriptions” of themselves**

Example of Basic Term-Matching Results



25 subjects for “Posters”

Your basic search for **posters** found these subjects:

Cine Rivoli (Buenos Aires, Argentina) -- Posters

Cine Teatro Standard (Buenos Aires, Argentina) -- Posters

Circus posters

Motion picture posters

New York (N.Y.) -- Posters

Performing arts posters

Political posters

Political posters -- Soviet Union -- History

Political posters, Russian

Posters

Posters -- 1790-1799

Posters -- 1860-1869

Posters -- 1870-1879

Posters -- French -- 19th century

Posters -- United States

Posters -- United States -- 1910-1919

Salon Garibaldi (Buenos Aires, Argentina) -- Posters

Explore Subjects (BETA)

Exact Match:

posters

Matches single word or phrase

Expanded Search:

Expands search to find more

GO [clear all options]

? [More about expanded](#)

This alphabetical list of subjects includes people, institutions, Library of Congress Subject Headings, and Thesaurus for Graphic Materials.

NOTE: Many Gallery records have descriptions. These records are indexed. Try the [names index](#) or [advanced searching](#) for more results.

“Related” Subjects Query Expands Results



781 subjects for “Posters”

Actresses

Adventure and adventurers

Advertisements

Advertisements -- New York (State) -- New York -- 1870-1879

Afforestation

African American clergy

African American farmers

African American men

African American theater

African American women

African Americans

Africans

Africans -- United States

Agricultural implements

Alcoholic beverages

Alexander, the Great, 356-323 B.C.

Alliances

Alligators

Expanded Search:

posters

Expands search to find mc

GO

[clear all option

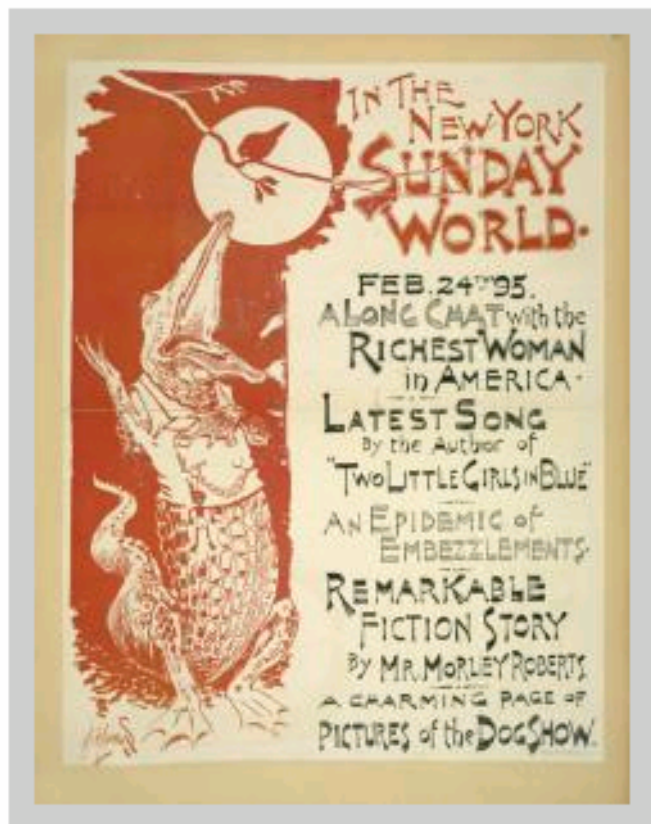
? More about expanded s

This alphabetical list of over 781 subjects includes people, institutions, and places. The Library of Congress Subject Thesaurus for Graphic Materials is the source of the list.

NOTE: Many Gallery record descriptions. These records are indexed. Try the [names index](#) or [advanced searching](#) for more results.

Example of “Related” Subjects Query

Searched for: the phrase "alligators" in Subject and "posters" in Subject



Digital ID: 1259144



Enlarge Image



Add to Selections



Printer-friendly Version



Purchase Print or Gift

Image Caption: The N. Y. Sunday World

In: Posters : arranged by artists. > M. de Lipman (published 1893-1924)

Published Date: February 1895

Library Division: Humanities and Social Sciences Library / Art and Architecture Collection, Miriam and Ira D. Wallach Division of Art, Prints and Photographs

Description: 10 v., 57 cm.

Specific Material Type: Prints

Subject(s): Alligators

Night

Posters

Collection Guide: Turn of the Century Posters

Digital Image ID: 1259144

Digital Record ID: 611905

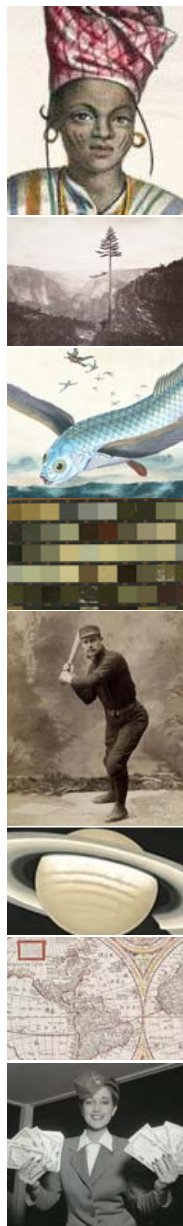
NYPL Call Number: 3-MDW+++ (Posters: arranged by artists)



View Library Catalog Record

[▶ Search For More Images](#)

The Scrapbook Problem



- **Frame-of-reference implicitly tied to single images**
- **In a single object (one image) a dog and cat are considered "related"**
- **In printer's proofs, scrapbooks of illustrations, multiple plates, etc. this notion can be problematic**
- **Subjects may share 1 bibliographic reference, but are they "related?"**

Related Subjects: "sailboats"

Chesapeake Bay (Md. and Va.) -- 1600-1609

Children blowing bubbles

Chimneys

China

Chivalry

Christianity

Christmas cards

Cigarette cards

City & town life -- New York (State) -- New York

City & town life -- New York (State) -- New York -- 1860-1869

Cityscapes

Clogs

Clothing & dress -- Egypt -- Jîzah

Clothing & dress -- Egypt -- Luxor

Clouds

Coastlines

Colossi -- Egypt -- Thebes (Extinct city)

Columns -- Egypt -- Philae

Connecticut

Costumes -- French

Croton Point (N.Y.)

Decks (Ships)

Dolphins

Dovecotes -- Egypt

Related Subjects: "sailboats"

Chesapeake Bay (Md. and Va.) -- 1600-1609

Children blowing bubbles

Chimneys

China

Chivalry

Christianity

Christmas cards

Cigarette cards

City & town life -- New York (State) -- New York

City & town life -- New York (State) -- New York -- 1860-1869

Cityscapes

Clogs

Clothing & dress -- Egypt -- Jizah

Clothing & dress -- Egypt -- Luxor

Clouds

Coastlines

Colossi -- Egypt -- Thebes (Extinct city)

Columns -- Egypt -- Philae

Connecticut

Costumes -- French

Croton Point (N.Y.)

Decks (Ships)

Dolphins

Dovecotes -- Egypt

Children blowing bubbles



Publisher's Proofs

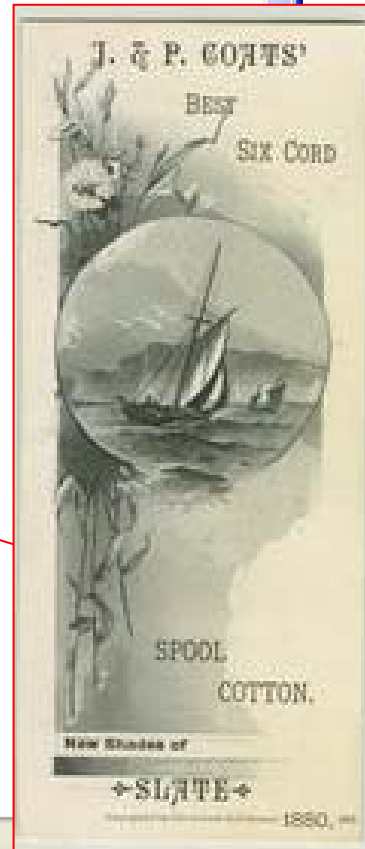
<http://digitalgallery.nypl.org> - Digital - Enlarged Image 487920 - Mozilla Firefox

NYPLDigitalGallery

Item 34 of 131

Add to Selections View Verso Purchase Print or Gift

34



Publisher's Proofs - Detail

Browse Source - Image Detail

Item 34 of 131

[<< Back to Thumbnail Images](#)

[< Previous](#) | [Next >](#)

Source: Scrapbooks of colored advertising cards in English and French : volume 2.



Image Caption: [Calendars and trade cards depicting a rowing race, jesters blowing bubbles, sail boats, thread, a horseshoe, Cleopatra's Needle, fans, kimonos, a man, a woman and a dog.]

In: Publisher's proofs of the publications of L. Prang & Co., and scrapbooks of colored advertising cards, published by L. Prang & Co. and other firms. > Scrapbooks of colored advertising cards in English and French : volume 2. (published [c1876-90])

Published Date: [c1876-90]

Library Division: Humanities and Social Sciences Library / Print Collection, Miriam and Ira D. Wallach Division of Art, Prints and Photographs

Item/Page/Plate Number: 34

Medium: Chromolithographs

Specific Material Type: Prints

Subject(s): Children blowing bubbles

Fans

Sailboats

Trade cards

Notes: Publishers include : John A. Lowell & Co. ; Wemple & Company.

Collection Guide: Publisher's Proofs and Related Work from L. Prang & Company

Digital Image ID: 487920

Digital Record ID: 303773

Digital ID: 487920



Enlarge Image

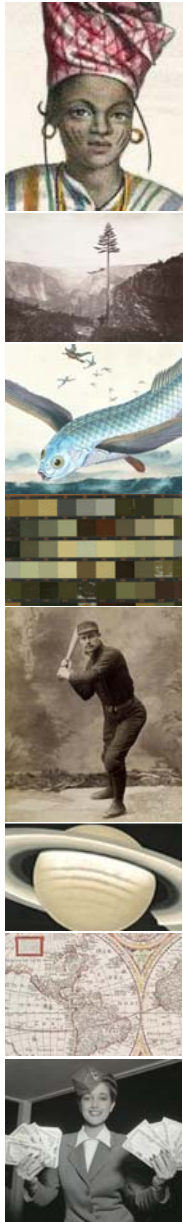


Add to Selections



View Verso

The 80/20 Problem



- How much metadata is enough metadata?
- 260,000 out of 318,000 images contain subject headings; however, 50,000 items are virtually invisible to our interface
- If user-experience proves the utility of leveraging subject headings, more staff and \$ could be allocated

Expanding Subjects: The UI Problem



- Expanded (“related”) subject list doesn’t fall easily into familiar categories:
 - Faceted browse display
 - “More like this” context
 - Top-down hierarchical menus
- Additional user-testing needed on GUI
- Front-end processing, though lightweight, is sometimes expensive: more can be done to optimize index and query

The Relevance Problem: Do Subjects Matter?



- Outside of specialized domains (Medline), do researchers still need subjects?
- The Big Indexers don't care so much about subjects, they want to index all of your data: scale is where the big gains are in search now, right?
- Good subject-analysis is expensive
- Folksonomies, tags, etc. attempt to describe things the way people think of them

Flickr: "Bad Day, 1445"

flickr GAMMA

You aren't signed in [Sign In](#) [Help](#)

[Home](#) [Learn More](#) [Sign Up!](#) [Explore](#) ▾

Search everyone's photos

[Search](#) ▾

Bad Day, 1445



Uploaded on [March 17, 2006](#)
by [pantufla](#)

[pantufla's photostream](#)



1,751
photos

[View as
slideshow](#)

[← more](#) | [browse](#) | [more →](#)

Tags

- medieval
- illustration
- book
- men
- gibbet
- violence
- torture
- blindfold
- mutilation
- hands
- 1445
- 1440s
- cunradus
- schlapperitzi
- german
- printed
- ephemera
- printedephemera
- design

NYPL: "Rubric and full-page miniature of..."

NYPLDigitalGallery

SEARCH: **GO**
[Advanced Search](#) | [Help](#)

Explore:

Arts & Literature

GO

Browse:
[Names A-Z](#) | [Subjects](#) | [Libraries](#)

My Digital:
[Search History](#) | [Selections](#)

Search Results - Image Detail


Item 175 of 200

<< [Back to Thumbnail Images](#) [Previous](#) | [Next](#) >

Searched for: "Schlapperitzi, Cunradus" as Name



Digital ID: 426562

 Enlarge Image

 Add to Selections

 Printer-friendly Version


 Purchase Print or Gift

Image Caption: Rubric and full-page miniature of men with their hands and feet being cut off.

In: Renaissance and medieval manuscripts collection, ca. 850-ca. 1600. > [Bible History] (created 1445)

Library Division: Humanities and Social Sciences Library / Manuscripts and Archives Division

Description: ff. 470, 28 x 21 cm.

Item/Page/Plate Number: f. 372

Specific Material Type: mss. text

Additional Name(s): Schlapperitzi, Cunradus -- Author

Collection Guide: Medieval and Renaissance Illuminated Manuscripts from Western Europe

Digital Image ID: 426562

Digital Record ID: 249634

NYPL Call Number: MA 104

 [View Library Catalog Record](#)

 [Search For More Images](#)

"Bad Day, 1445"



Image Caption:

Rubric and full-page miniature of men with their hands and feet being cut off.

In:

Renaissance and medieval manuscripts collection, ca. 850-ca. 1600. > [Bible History]
(created 1445)

Library Division:

Humanities and Social Sciences Library / Manuscripts and Archives Division

Description:

ff. 470, 28 x 21 cm.

Item/Page/Plate Number:

f. 372

Specific Material Type:

mss. text

Additional Name(s):

Schlapperitzi, Cunradus -- Author

Collection Guide:

Medieval and Renaissance Illuminated Manuscripts from Western Europe

Digital Image ID:

426562

Digital Record ID:

249634

NYPL Call Number:

MA 104

BOOK
men
gibbet
violence
torture
blindfold
mutilation
hands
1445
1440s
cunradus
schlapperitzi
german
printed
ephemera
printedeephemera
design

Additional Information

© All rights reserved

◇ Taken on [March 17, 2006](#)

◇ Viewed **147** times

 This photo is **public**

"Bad Day, 1445"

Comments



[facecrunch](#) pro says:

Man, that would suck.

Posted 8 months ago. ([permalink](#))



[facecrunch](#) pro says:

Do you know what language it is written in?

Posted 8 months ago. ([permalink](#))



[pantufla](#) pro says:

Looks like German..see more pages by this guy at

digitalgallery.nypl.org/nypldigital/dgkeysearchresult.cfm...

Posted 8 months ago. ([permalink](#))



[greentea flute](#) says:

I really like what you are doing here with your stream on flickr.... that said, this is pretty horrific, even if fascinating.... I came across a similar book in Brittany some few years back... I guess this is still going on, if you read the news, all over the world... état humaine

Posted 8 months ago. ([permalink](#))



[sadalit](#) pro says:

I was searching for a photo of a gibbet, came across this - your title is too funny, but did they HAVE to? Do we still have to?

Posted 23 hours ago. ([permalink](#))

Would you like to comment?

[Sign up](#) for a free account, or [sign in](#) (if you're already a member).

Search: violence or

Stringed instruments

Stroud, William

Suffolk, William de la Pole, Duke of, 1396-1450

Suriname

Swindlers & swindling

Synagogues -- Netherlands -- 1700-1799

Süss-Oppenheimer, Joseph, 1698- or 9-1738

Tacks

Tarring & feathering -- Massachusetts -- 1860-1869

Tarring & feathering -- Massachusetts -- Boston -- 1770-1779

Tarring & feathering -- Pennsylvania -- 1790-1799

Tarring & feathering -- United States

Tarring & feathering -- United States -- 1760-1769

Tax collection -- United States -- 1760-1769

Thomas, Clément

Throwing

Torture

Torture of self -- Great Plains -- 1800-1899

Torture victims

Tourniquets

Towers -- England

Tramps

Treadwheel

Trees



BUNCO-STEERERS AND CONFIDENCE MEN.

"I mounted to the summit of a small precipice, and looked over."

"No. 862 (1892)"

Heaven + hell

N.Y. PUBLIC LIBRARY
ASTOR LENOX TILDEN FOUNDATION
FUTURE COLLECTION


Digital ID: 833484

Creator(s): Young, Art, 1866-1943 -- Artist

Bunko-steerers and confidence men.

Search: "mutilation"

NYPLDigitalGallery



SEARCH: **GO**

Advanced Search | Help

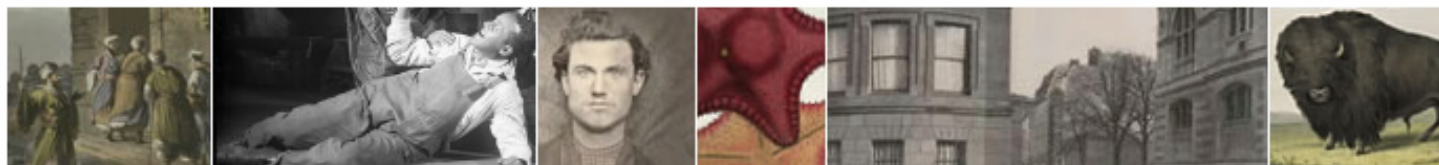
Explore:

Arts & Literature

GO

Browse:
[Names A-Z](#) | [Subjects](#) | [Libraries](#)

My Digital:
[Search History](#) | [Selections](#)



Subjects A-Z

[A](#) | [B](#) | [C](#) | [D](#) | [E](#) | [F](#) | [G](#) | [H](#) | [I](#) | [J](#) | [K](#) | [L](#) | [M](#) | [N](#) | [O](#) | [P](#) | [Q](#) | [R](#) | [S](#) | [T](#) | [U](#) | [V](#) | [W](#) | [X](#) | [Y](#) | [Z](#)

Your expanded search for **mutilation** found these related subjects:

Books -- Mutilation, defacement, etc.

New York Public Library

2 subjects found for this search

[View all images with mutilation in subject](#)

Explore Subjects (BETA)

BETA

Exact Match:

Matches single word or phrase

Expanded Search:

all words (AND) 

Expands search to find more related subjects



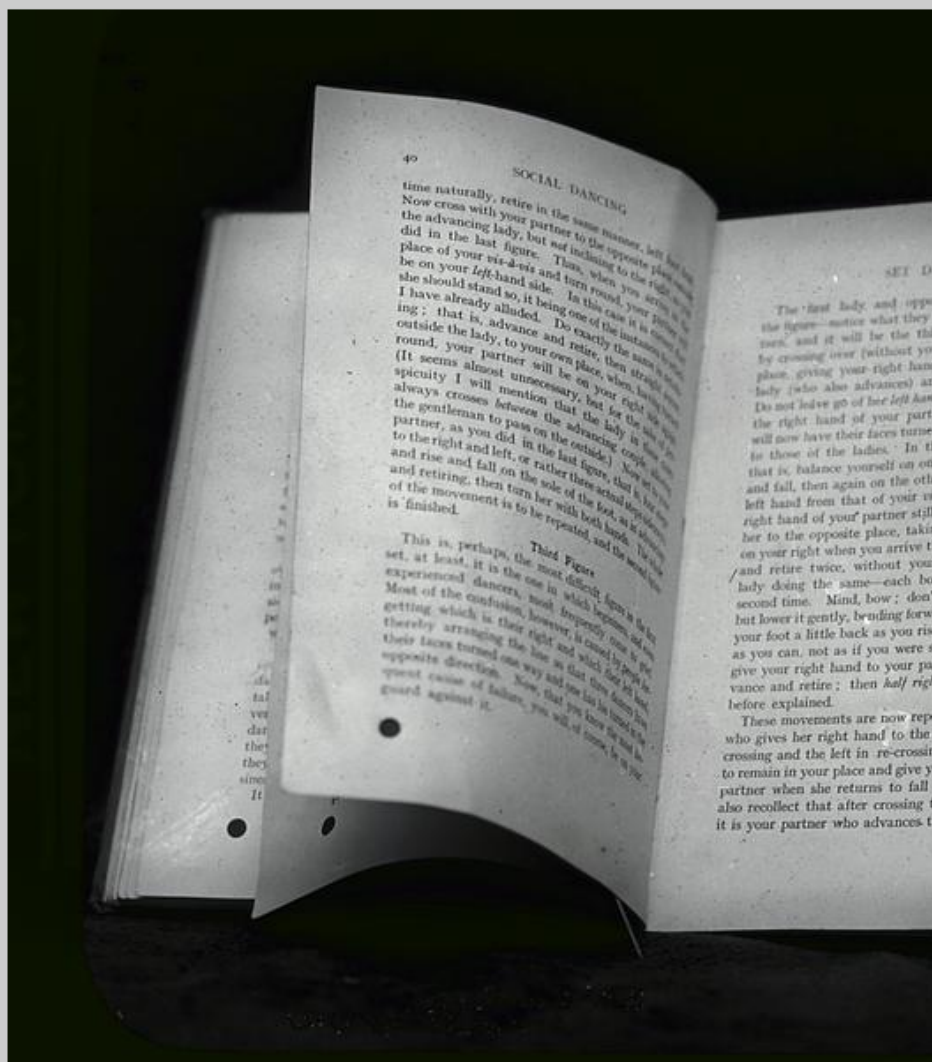
[clear all options]



More about expanded searching

This alphabetical list of over 58,000 subject terms includes people, institutions, places and topics from Library of Congress Subject Headings and the LC Thesaurus for Graphic Materials.

NOTE: Many Gallery records do not have subject descriptions. These records are not linked from this index. Try the [names index](#), keyword searching, or [advanced searching](#) for more results.



Digital ID: 115822

Drilled books, Mar. 27, 1913

In: New York Public Library Visual Materials > Lantern Slides > Office of Special Investigation

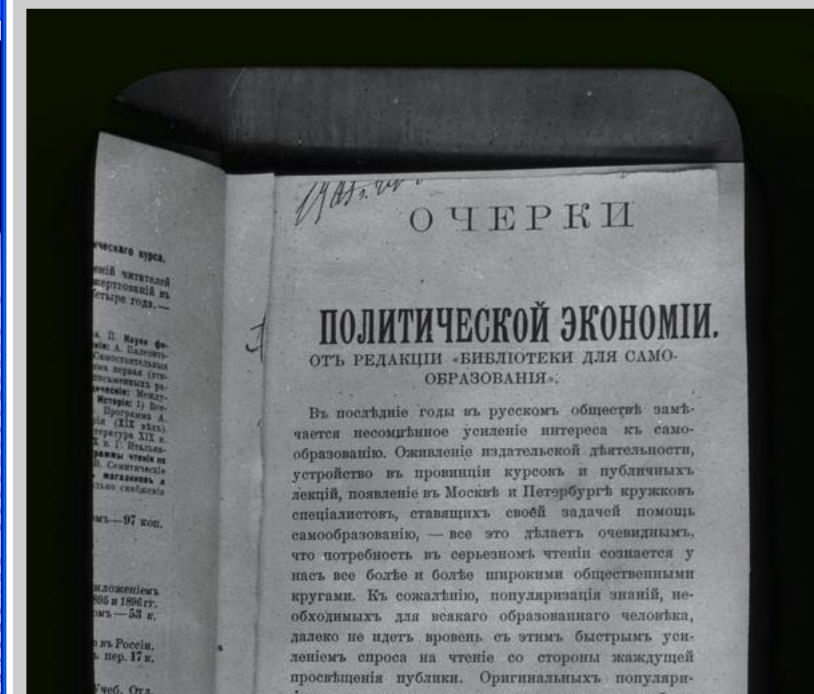
Done

NYPLDigitalGallery

Item 2 of 2

+ Add to Selections

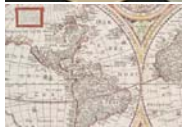
🛒 Purchase Print or Gift



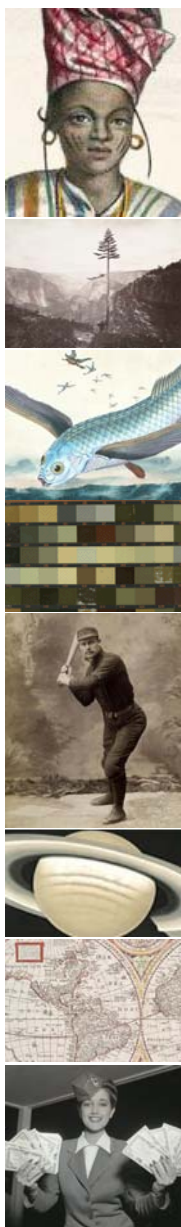
Done

These movements are now repeated by your partner, who gives her right hand to the opposite gentleman in crossing and the left in re-crossing, you having simply to remain in your place and give your right hand to your partner when she returns to fall in a line. You must also recollect that after crossing to the opposite place, it is your partner who advances twice and bows to her

The Answer: [Maybe|No|Yes] ?



- Images, at least, need description, and likely will for the foreseeable future(?)
- Subjects and other controlled vocabularies are good at describing while minimizing 'noise'
- Tags are subjects, just loosely typed?



Some Further References & Acknowledgments

- ❑ Krowne, Aaron and Martin Halbert. *An Initial Evaluation of Automated Organization for Digital Library Browsing* (JCDL '05). 2005.
- ❑ Lagoze, Carl, Dean B. Krafft, Sandy Payette, Susan Jesuroga. *What is a Digital Library Anymore, Anyway?* (D-Lib Magazine). November, 2005.
- ❑ NISO. *A Framework of Guidance for Building Good Digital Collections* (2nd Edition). 2004.
- ❑ Thanks to: Lee Horowitz (NYPL DGTL Oracle database consulting), Janet Murray (NYPL DGTL Metadata Coordinator), Tom Robertson (Lucene consulting, Stanford), Barbara Taranto (NYPL DGTL Director)
- **Contact: jdalton@nypl.org**