Digitized Collections Online
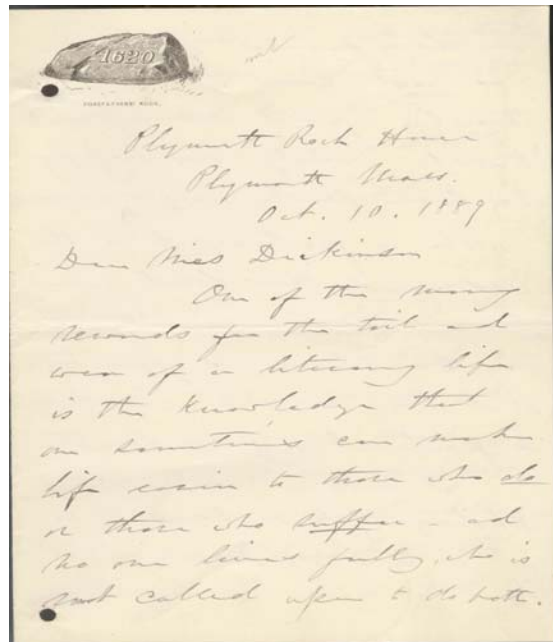
New York Society Library
April 15, 2008
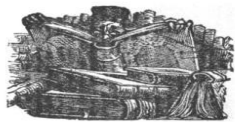
By Keren Fleshler

kfleshler@nysoclib.org

1. What is digitizing or the process of digitization?

Digitizing is the process of converting data to digital format for processing by a computer.
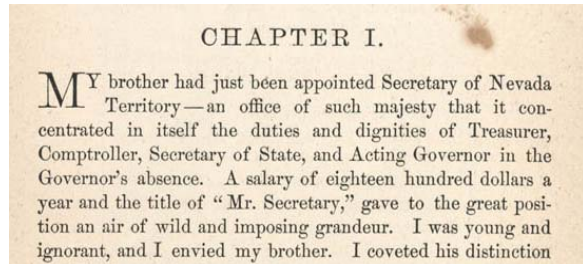
Digitized texts and images are scanned and then uploaded to the web in their original structure or design, or they may undergo a process involving **OCR**, or Optical Character Recognition.

OCR uses computer software designed to translate images of handwritten or typewritten text, usually captured by a scanner, into machine-editable text (such as Notepad, Microsoft Word, etc.). OCR is helpful because it allows users to search for words in the text, but is problematic because words are sometimes transcribed incorrectly in the process.

OCR version   Scanned-only digitized version

My brother had just been appointed Secretary of Nevada Territory -- an office of such majesty that it concentrated in itself the duties and dignities of Treasurer, Comptroller, Secretary of State, and Acting Governor in the Governor's absence. A salary of eighteen hundred dollars a year and the title of "Mr. Secretary," gave to the great position an air of wild and imposing grandeur. I was young and ignorant, and I envied my brother.

Some materials are "born digital" and do not to be digitized. This refers to such resources as Word documents, e-journals, photos taken with a digital camera, or websites.

Why digitize?

Institutions around the world are focusing on digitization projects in order to share their resources with users around the globe, to promote their collections, and to preserve older or rare material for posterity. The digitized item often becomes the surrogate for material that is often too brittle to be used for research.

2. What is a digital collection?

Digital collections are the products of digitization projects, and are composed of materials converted to machine-readable format for preservation or to provide electronic access, usually by libraries, museums, or archives. A digital collection consists of digital objects that are selected and organized to facilitate their access and use.

Digital collections often include books, manuscripts, letters, photographs, journals, magazines, newspapers, documents, and audio and visual clips. They also frequently complement ongoing or past exhibitions.

Copyrights and permissions for digital material

The majority of books digitized by libraries and through projects, such as Google Books and Project Gutenberg, are **public domain** books. These are books that are out of copyright, and therefore can be distributed to the public and copied freely.

Books published in the United States prior to 1923 are public domain books, and may be digitized without obtaining any permissions and put online in full-text. Works of the United States Government and various other governments are also excluded from copyright law and may therefore be considered to be in the public domain.

In some cases, permissions for public use of digitized material must be obtained from the holding institution. The terms of use for a particular institution/collection are normally found on the website for the collection.

3.   How would I search for digitized resources in my area of interest?

1. If you are aware of an institution that holds or collects in your area of interest, then there is a good chance it may have a digital collection on the subject. To find digital collections at a university, go to the library homepage and look for links for "Digital xxx" or something similar. Sometimes, a larger library collection may have its own page and the digital collections may be found there. Checking the rare books and special collections department page is also helpful.

2. If your research topic is geographically based, there is a good chance a library, university, or historical society located in the area of interest may have a digital collection on the topic. The same hold true for research on influential people who lived or flourished in specific geographical areas.

3. A simple internet search engine, using sites such as Google, may pull up digital collections by a combination of name, subjects, or institutions.

Example searches:
Women's history digital collections - pulls up sites with women's history digital collections at various institutions.
UCLA digitization projects – pulls up link all of UCLA's online collections

These web searches may also help you pull up other digital collections that may include your subject in some of the documents.

4. Even if an institutional website does not offer digital collections, often times a helpful "links" page can point you in the right direction.

Example: a small historical society in Vermont may have a links page that will point you to a major Vermont history digitization collaboration being administered by a network of Vermont libraries and universities, even if the historical society itself is not taking part.

5. You may also use the one of the more comprehensive Guides to Digital Collection in this handout to search several collections at once.

System requirements

Java, download from: http://java.com
Adobe Reader (PDF viewer), Adobe Flash Player, Adobe Shockwave Player: Download all three from www.adobe.com

For video players: Windows Media Player, Real Player, and Quicktime Player are most commonly used.

A few sites may require you to download specific reader software. Some may also allow you to download the e-resources to your PDA or iPod.

Guides to Digital Collections

Institute of Museum and Library Services
Digital Resources from Libraries, Museums, and Archives
http://imlsdcc.grainger.uiuc.edu
Searches the collections of institutions with digital collections that have been funded by IMLS.

Internet Archive
http://www.archive.org
Searches e-books from American and Canadian libraries, as well as from the Million Books Project, Project Gutenberg, Open Source Books, and classic children's books. It also includes 85 billion archived web pages.

Digital Library Federation
http://quod.lib.umich.edu/i/imls
Portal to the University of Michigan Digital Library Production Service project, with over 2 million records from over 60 contributing institutions.

OAIster
http://www.oaister.org
Searches the collections of institutions that are part of the OAI initiative, and includes digitized books, audio and video files.

Registry of U.S. Government Publication Digitization Projects
http://www.gpoaccess.gov/legacy/registry
The Registry contains records for projects that include digitized copies of publications originating from the U.S. Government.

E-Book Collections

Google Books
http://books.google.com
Search full-text for books out of copyright, and selections of texts in copyright.
http://zuhause.org/dp/gfound1.html
Third-party website that has compiled a listing of books found on Google Books.

Project Gutenberg
http://www.gutenberg.org
First to offer free digitized book access, now offers over 25,000 ebooks, as well as a selection of audio books and sheet music.

Project Gutenberg (Australia)
http://gutenberg.net.au
The Australia Gutenberg Project offers books out of copyright in Australia, and includes a collection of Australiana etexts.

Early American Fiction (1789-1875)
http://etext.lib.virginia.edu/eaf
The University of Virginia allows restricted access to 158 volumes by 52 authors, plus 199 manuscript items and 124 non-text images.

American Verse Project
http://www.hti.umich.edu/a/amverse
The American Verse Project is a collaborative project between the University of Michigan Humanities Text Initiative (HTI) and the University of Michigan Press. The project is assembling an electronic archive of full-text volumes of American poetry prior to 1920.

Early Books

British Library: Turning the Pages
http://www.bl.uk/onlinegallery/ttp/ttpbooks.html
A selection of interactive books and manuscripts, including Lewis Carroll's original Alice manuscript, the first atlas of Europe, and Mozart's musical diary.

British Library: Treasures in Full
http://www.bl.uk/treasures/treasuresinfull.html
Includes the Magna Carta, as well as a collection of 253 books about Renaissance festivals in Europe during the period of 1475-1700.

Gutenberg Digital
http://www.gutenbergdigital.de/gudi/start.htm
View all 1282 pages of the two volumes of the Göttingen Gutenberg Bible.

Renaissance in Print
http://www.lib.virginia.edu/rmds/collections/gordon
Sixteenth century French Books in the Douglas Gordon Collection, including three printed books of hours, travel narratives, books for women, etc.

Periodicals

Note: The National Endowment for the Humanities and the Library of Congress have recently embarked on the National Digital Newspapers Program (NDHP), a 20-year long project intended to digitize significant American newspapers published between the years 1836-1922. http://www.neh.gov/projects/ndnp.html

Library of Congress: Chronicling America Newspaper Guide
http://www.loc.gov/chroniclingamerica
This is the prototype for the NDHP project.

The Nineteenth Century in Print
http://memory.loc.gov/ammem/ndlpcoop/moahtml/snchome.html
Includes over twenty 19th century periodicals from the Library of Congress

International Coalition on Newspapers (ICON)
http://icon.crl.edu
http://icon.crl.edu/digitization.htm
http://icon.crl.edu/guides.htm
Provides an extensive list of international newspapers available in the United States, with links to digitized newspapers, as well as bibliographies and union lists, and is based at the Center for Research Libraries.

Making of America
http://library5.library.cornell.edu/moa   - Cornell University
http://quod.lib.umich.edu/m/moagrp   - University of Michigan
Collaboration between Cornell University and the University of Michigan, with American books and serials from the 19th century.

Internet Library of Early Journals
http://www.bodley.ox.ac.uk/ilej
20-year consecutive runs of: Gentleman's Magazine, The Annual Register, Philosophical Transactions of the Royal Society, Notes and Queries, The Builder, Blackwood's Edinburgh Magazine

Utah Digital Newspapers
http://www.lib.utah.edu/digital/unews/
Over 100,000 pages and over 15 titles.

Washington State Historical Newspapers
http://www.secstate.wa.gov/history/newspapers.aspx
Over a dozen Washington State newspapers

British Library Online Newspaper Archive
http://www.uk.olivesoftware.com/
Selections from Manchester Guardian, Daily News, News of the World, and Weekly Dispatch.

Freedom Journal
http://www.wisconsinhistory.org/libraryarchives/aanp/freedom/
First African-American periodical, 1827-1829, from Wisconsin.

Northern New York Historical Newspapers
http://news.nnyln.net/
Thirty New York newspapers from the 19th and 20th centuries.

Brooklyn Daily Eagle
http://www.brooklynpubliclibrary.org/eagle
Ongoing project offering digitized access for the years 1841-1902.

National and Local Histories and Periodicals

NARA (The National Archives) Archival Research Catalog
http://www.archives.gov/research/arc
The ARC is the online catalog of NARA's nationwide holdings in the Washington, DC area, Regional Archives and Presidential Libraries, including 127,785 digitized items.

American Memory, Library of Congress
http://memory.loc.gov
Comprising more than 9 million items that document U.S. history and culture, American Memory is organized into more than 100 thematic collections based on their original format, their subject matter, or who first created, assembled, or donated them to the Library.

British History Online
http://www.british-history.ac.uk
Includes local records, maps, from 11-19th centuries.

Digital Library of Wales: Digital Mirror
http://www.llgc.org.uk/
Manuscripts, maps, images, sound and video

Historic Missouri Newspaper Project
http://newspapers.umsystem.edu
A dozen 19th century newspapers from the Missouri area, collected from universities and public libraries.

Historic Pittsburgh
http://digital.library.pitt.edu/pittsburgh/
Books, images, censuses, maps relating to Pittsburgh, from a collaboration between Pennsylvania-area universities, libraries, archives, and museums.

A Selection of Other Collections

Cooperative Digital Resources Initiative (CDRI)
http://www.atla.com/digitalresources/
The CDRI is the digital repository of the American Theological Library Association and Association of Theological Schools. The CDRI database provides access to digital images of woodcuts, photographs, slides, papyri, coins, maps, postcards, manuscripts, lithographs, sermons, shape-note tune books, and various forms of Christian art, architecture, and iconography.

National Library of Medicine: History of Medicine Digital Collections
http://www.nlm.nih.gov/hmd/collections/digital
Online books, Islamic manuscripts, and anatomical illustrations.

Columbia Papyrus Project
http://www.columbia.edu/cu/lweb/projects/digital/apis
Repository of papyrological materials from various institutions, with over 18,000 digital images, along with English translations.

Complete Work of Charles Darwin Online (in process until 2009)
http://darwin-online.org.uk
Contains Darwin's complete publications, manuscripts, bibliographies, and supplementary works.

American Women's History: A Research Guide
http://www.mtsu.edu/~kmiddlet/history/women/wh-digcoll.html
An extensive collection of links to digital collections in the field of women's history.

Ad Access
http://library.duke.edu/digitalcollections/adaccess
Duke University offers a collection of 7,000 ads printed in US and Canadian magazine and newspapers between 1911-1955.

Department of Transportation (DOT) Special Collections
http://dotlibrary.specialcollection.net
Collections of interest to the transportation community and researchers, with digitized historical manuals, reports, orders, etc.

British Library: Guide to European digital collections
http://www.bl.uk/collections/epcdigital.html
Links to electronic text sites: http://www.bl.uk/collections/wider/etexts.html

David Rumsey Map Collection
http://www.davidrumsey.com/
The collections contains over 17,400 maps online, with focuses on rare 18th and 19th century North American and South American maps and other cartographic materials.

Image, Sound, and Video Collections

Bodleian Library Broadside Ballads
http://www.bodley.ox.ac.uk/ballads
Over 30,000 of the Library's ballad and ballad sheet holdings from between the 16th to early 20th centuries are available, with some audio files.

Steven Spielberg Jewish Film Archive
http://www.spielbergfilmarchive.org.il
Access to over 300 Jewish history and Zionist films in English and Hebrew, offered in collaboration with the Hebrew University of Jerusalem.

Vanity Fair Caricatures
http://www.lib.virginia.edu/rmds/collections/vanity_fair
Nine hundred illustrations of Victorian era personalities, available through the University of Virginia.

Library of Congress: Picture Catalogs Online
http://www.loc.gov/rr/print/resource/223_piccat.html
Links to 50 online catalogs of image collections.

ADDITIONAL HELPFUL RESOURCES:

LIBDEX
http://www.libdex.com
Index of 18,000 libraries worldwide.

Copyright Term and the Public Domain in the United States
http://www.copyright.cornell.edu/public_domain/