E-book Collections -- Guide

This is a brief overview of some of the electronic books sources freely available to you on the Internet as well as those accessible courtesy of Watson Library subscriptions and purchases.

Resources covered:

- Eighteenth Century Collections Online
- ACLS Humanities E-Book
- Project Gutenberg
- Online Books page (University of Pennsylvania)
- Google Books
- Amazon 'Search Inside the Book'
- Open Content Alliance
- Internet Archive Text Archive
- Universal Digital Library or the Million Books Project
- Watson Library del.icio.us account: E-Books

PURCHASED/SUBSCRIPTION ELECTRONIC BOOK COLLECTIONS
Eighteenth Century Collections Online (ECCO)
http://libmma.org/portal/ecco-eighteenth-century-collections-online/
Note that you will be prompted to log in to your library account if using ECCO outside the Museum.

Overview
ECCO, Eighteenth Century Collections Online, contains a fully searchable and digitized variety of materials including books, auction catalogs, artists' biographies and handbooks, treatises on painting and architecture, pamphlets, broadsides, and ephemera. ECCO includes works printed between 1701 and 1800 in the United Kingdom, the Americas and around the world.

Access
ECCO is a closed collection and is accessible remotely to Museum staff via our library portal or WATSONLINE. Individual ECCO titles are searchable in WATSONLINE by keyword, author, or title (search for "Eighteenth Century Collections Online").

Watson Library has purchased the Fine Arts and Social Sciences collections within ECCO -- these comprise a full-text digital collection of over 30,000 English language and foreign language titles. These items have been digitized from microfilm, and note that the quality of the images varies, depending upon the state of the original. Most content comes from the British Library. ECCO combines full citations, including English Short Title Catalog numbers, with complete facsimiles of the individual works allowing scholars to compare different editions of the same work. The links to tables of contents and lists of illustrations are especially valuable features.

Searching for and viewing works
The simple search screen includes keyword, author, title, and full text. You may search within a particular sub-collection, Fine Arts or Social Sciences.

The default search is a **proximity** search -- you must use AND between two words that may not occur within a few words of each other. Compare the full text search results for

journey through Spain (no operators, 17 results) to **journey AND through AND spain** (AND operators, 2785 results).

Search for **Zurbaran** to compare results for keyword (0 results) to full text (several hits) results. It's important to note that the keyword search does not search the full-text of all documents -- only the author, title, and tables of contents.

You may also use the other traditional Boolean operators OR and NOT, as well as the proximity operators W (within) and N (near). Briefly, W will find documents containing the specified words in the specified order within the number of words you indicate. N will find items containing the specified words within a certain number of words, but the words can be in any order. For further explanation, please consult the very good ECCO Search Tips page.

Results list
Returning to the full-text results for **Zurbaran**, you may sort search results by author, title, or publication year ascending or descending. ECCO may provide a variety of a work's editions, enabling scholars to compare versions. The
links to contents, and illustration lists are especially valuable features.

Select **Palomino, "An account…," 1739**. On the left are links to the pages including your search terms, or you may search within the selected work. Your search terms on the relevant pages are highlighted. There are several other functions available:
- Page through using arrows.
- Skip to a specific page or image number using the appropriate boxes.
- Scale view -- changes the zoom level. Makes the page view smaller or larger.
- Rotation -- if you have a landscape-orientation page and need to rotate the page, use this function.
- Reformat for reading -- good for on-screen reading, but not for printing.
- Printing – you may print one page at a time, or open several pages in PDF to print or save.

If you search this work, you search the full text of the work and may set the "fuzzy level" as you wish. For example, let's look for **color** and set fuzzy to medium. Again, you see the links to search results on the left and the term is highlighted on that page. On page 80 we can see it retrieved **colour'd**, and on 115 you see that the search engine has retrieved the name **Corner**, providing a hint of how the fuzzy search casts a wider net.

The search history is always available on the search pages: you may always re-do searches within the same session with two clicks.

Advanced search
The advanced search screen adds other indexes to the basic search. The drop-down menu includes front matter, publisher and place of publication, among others. There are is also the fuzzy search setting as well as limiting features by language, and illustration type.

Let's do a sample search for **Rembrandt AND catalogue** in full-text, limiting to illustration type **portrait**.

Select: **Daulby, Daniel. *A descriptive catalogue of the works of Rembrandt…*** . On the first page you can see your terms highlighted, and under the list of illustrations you will see a link to a portrait of "Rembrandt van Ruyn."

When using Advanced Search, you need not fill out any of the text boxes on the top. Try using only these limits: year of publication **1748**, collection **Social Sciences**, and

illustration type **Coat of arms**. Select any of the links for "List of illustrations" in the result set and you will see that all these 1748 works include coats of arms illustrations.

Citations
You may mark works for saving and email yourself the citations. You may also mark works in the search results list by checking the box.

ACLS Humanities E-Book (HEB)
http://libmma.org/portal/acls-humanities-e-book/

Overview
Humanities E-Book is a growing digital collection of over 1700 full-text titles offered by the American Council of Learned Societies, nearly 90 contributing publishers, and librarians at the University of Michigan. The result is an online, fully searchable collection of high-quality books in the Humanities, recommended and reviewed by scholars. 1601 titles are cataloged individually in WATSONLINE and therefore accessible by title, author, subject, and keyword search. The oldest book I've seen was published in 1850 and there are even 2008 books in the collection. Most works are post-1930 and many are from the 1980s and 1990s.

Humanities E-Book has begun to work with the College Art Association to publish reprints in its series, *Monographs on archaeology and fine arts*, starting with those that are out of print. One example you can see in WATSONLINE is **Meyer Schapiro's *The Parma Ildefonsus : a Romanesque illuminated manuscript from Cluny and related works*** (click here for WATSONLINE search).

Access
You may access HEB via our E-Resources A-Z list or by using the Site Search. Here is also a link to browse the Humanities E-Book title list in WATSONLINE:
http://library.metmuseum.org/search/tHumanities+E-Book – which you may then post-limit if you like. But the most up-to-date searching is done on the HEB site itself.

Site layout
A brief description of the important sections.
- About – here you can find a complete title list organized by subject category and make recommendations.
- Search – search for works in the HEB collection.
- Help – a decent help file answering questions about searching and content.
- Browse – you may browse on these pages by author, title, or subject.

Searching for and viewing works
Several options exist for you – Basic, Boolean, Proximity, Bibliographic (restricts the search to a word or phrase contained in titles, authors' names or subject headings), and History (shows you your search history, allowing you to re-execute previous searches). Searching has a default AND for multiple terms and you may truncate words with the *.

Say I'm interested in more information about the Maasai, an African tribe in Kenya and Tanzania, since I came across this shield in the MMA galleries:
http://www.metmuseum.org/toah/ho/10/sfe/ho_1978.412.644.htm

I'm going to run a basic search for **maasai** in full-text.

Search results

You may sort by the traditional elements. Red icons indicate that the books were scanned as page images, blue icons are for those works with encoded text. The difference is that the encoded text works (blue) are more interactive and may not have traditional pagination, delivered to the user instead in paragraphs. The page image material is less flexible and hews to the same pagination as the print edition. I'll show you an example in a moment.

If you select "Results details" for one of the works you will see a listing of where your search terms appear in the text with links to those pages. You may also move to the next item in your results list or return to the list (end up at "Speaking with vampires"). You see that the number of matches are given and your terms are highlighted. Select "Urbanization In Kenya" (under Chapter 5) and take a look at the text – note also that footnotes are linked.

At the top are links to the title record, contents, reviews, etc. which are useful links. Let's go to the Reviews link – if there are reviews, a drop-down menu will allow you to choose the one you'd like to see. A new window will open and you see that the reviews are those available in JSTOR.

Looking back at the table of contents we can view the index, which is fully linked to the text. Finally, if we select "Siting vampires" under Chapter 1 we can see that illustrations for encoded works can be embedded in the text (map, paragraph 17-18) and are enlargeable.

Example: illustrations
For comparison, execute a basic search for byzantine art in title – this is a book Watson Library has in print, in 2 editions. I'm going to go to the table of contents and select the illustrations chapter (the first page is blank). This is a red-icon (page image) work, so you can see that the images aren't that great and zooming in to look more closely doesn't help much.

Citations
You may save citations for works either in the search results list or in the upper right corner of the page when viewing a contents page or a page within a work.

Clicking "view my citations" (you must "save citation" first, even if you have already saved that citation) brings you to a page where you may email, download (as a text file – but you must change the file extension to .txt) or empty the contents of your saved citations list. This works very well and as you can see is quite intuitive.

The big problem
The big problem with HEB is that printing is extremely difficult and electronically saving a portion of the text is near impossible. The most you are able to print or save is one page at a time, though you may order a reprint of the entire text for a fee.

Project Gutenberg
http://www.gutenberg.org/wiki/Main_Page
Project Gutenberg was founded in 1971 and is the first producer and distributor of free electronic books (ebooks). All the work, including scanning, is done by volunteers. Books are captured in Plain ASCII text, which means they can be read easily by any machine, operating system or software. As of the end of 2007, Project Gutenberg has over 26,000 titles available, spread across four mirror sites in the United States, Canada, Europe and Australia.

Online Books Page
http://digital.library.upenn.edu/books/

If you are new to ebooks, we recommend first visiting Online Books Page. This site provides comprehensive, current indexes and lists of free digital books, serials and 'definitive collections' available on the Internet. Online Books Page is maintained by University of Pennsylvania Librarian John Mark Ockerbloom and hosted at University of Pennsylvania Libraries. To date it has indexed over 30,000 English works alone on its local site.

Of particular interest are indexes to online literature available outside the site. From the homepage, click on the link for "Archives and Indexes." This page covers general, large-scale repositories such as Google Books, Internet Archive and Project Gutenberg, to smaller, specialized archives such as Humanities Text Initiative at University of Michigan, and foreign language resources such as Perseus.

You may subscribe to their RSS feed to receive new listings and updated entries. You can also provide feedback, suggest other titles, even put a book online yourself: http://onlinebooks.library.upenn.edu/getinvolved.html.

Google Book Search
http://books.google.com/
Google Book Search is a Google beta project that allows users to use keywords to search the full text of scanned books. While book digitization projects have been around for decades, Google Book Search started the movement of "mass digitization", which differs from previous projects for its focus on massive scale, quantity and speed. Their aim is "to make all the world's books discoverable and searchable online."

Google Book Search consists of three parts: the **Partner Program**, or participating publishers, the **Library Project** and books still in copyright. For the Partner Program and Library Project Google pays for the scanning costs; libraries are given a copy of the digital file.

In the Partner Program, publishers owning rights to a book authorize Google to scan the full text of the book and add the content to Google's search results. For newer books, Google will usually display limited content. example:
http://books.google.com/
books?id=Dm8aXSwL0OgC&printsec=frontcover&dq=gandhara&ei=qwmrR8qrDpy8zASjr5GLDw&sig=xiscyl

Under the Library Project, Google will scan only books in the public domain. These books will then be made available for downloading in PDF form. As of January 2008, Google Books Library project has partnered with 28 libraries, including 2 art library collections:

- Kohler Art Library of University of Wisconsin, Madison : Approximately 6000 regular-sized art Cutter books are being digitized in the Google Book Project as a way to expand access to these important historical materials.
- Art, Architecture and Engineering Library of University of Michigan: example.

Over the next decade, Google expects to scan 15 million books from the Library Project.

For books still in copyright and not included in the Partner Program, Google will still scan the entire text into its search index, but show just 'snippets' of the scanned content. Snippet view example: http://books.google.com/
books?id=RXBQAAAAMAAJ&q=metropolitan+museum+of+art&dq=metropolitan+museum+of+art&lr=&ei=
npJp6SzQTvp4SJCQ&pgis=1

How to use Google Book Search
Go to http://books.google.com and search for "Metropolitan Museum of Art."

- **On screen viewing options:** Zoom, two-page layout vs. facing pages, scanned image or text only
- Find other titles of related interest
- Download public domain works
- Buy the book or find it in a library through **WorldCat**
- **Write reviews, assign ratings, attach labels/tags, share with others, locate on map**
- **Full text searchable**
- Search in **multiple languages,** including scripts: 石涛
- Tailor searches by limiting additional fields in <u>Advanced Book Search</u>.
- **Search feeds** automatically set up for new additions to your library. You may subscribe to others' feeds as well.
- **"My Library"** function gathers favorite digital books found via Google Books and creates online libraries (log-in required). You can share your collection and set up RSS feeds so subscribers will be know when new books are added.
- Embed clippings from **public domain books** directly in Google Notebook or Blogger (still in beta).
- **New API** provides previews of scanned books within library catalogs: <u>http://melvyl-test.cdlib.org:8164/F</u>. API in Open library: <u>http://openlibrary.org/b/OL5298255M</u>

Some Applications for Google Book Search
- Research tool for primary sources
- Google Book Search is integrated into Google. When you enter a book title in Google.com, Google will automatically search against its Book Search engine and return matching results.
- Google Books may be valuable for finding books that are out of print and/or difficult to find. Example: <u>Handbook of the Benjamin Altman Collection,</u> 1910.
- Google books may provide more information about the book than the publisher, such as table of contents, sample pages. Compare *Cataloging Cultural Objects* as it appears on the <u>publisher's website</u> and as it appears in <u>Google Book Search</u>.
- Checking the authority of a book by looking at its citations.
- For books that are missing from a library catalog, you may look up the cover or dust jacket in Google Books
- Looking up different editions of a title.

Drawbacks to using Google Book Search
- Scan quality is uneven. Common problems include fingers showing, such as this: <u>http://books.google.com/books?id=vrw5mJphgp4C&printsec=frontcover&dq=inauthor:metropolitan+inauthor:museum+inau</u>
- Libraries that partner with Google must agree to restrict availability to other commercial search services.
- ongoing copyright litigation with 'snippet' views
- When logged into your Google account, Google will log your search history. Log records are kept on Google servers for up to 18 months before being expunged.

Print-outs provided by Google: <u>http://books.google.com/googlebooks/handouts.html</u>

Amazon: Search Inside The Book (SITB)
<u>http://www.amazon.com/</u>

Amazon was the first commercial vendor to pioneer new technologies in scanning, digitization and search discovery with their Look Inside! The Book (LIB) and Search Inside! The Book (SITB). The Search Inside database is integrated into Amazon's default search so that keywords are automatically run through the texts included therein. Users can search the content of books and view sample pages on Amazon's Online Reader, but only books for which Amazon has the publisher's permission to display copyrighted material.

In order to see all the features of SITB, you have to first log in to your Amazon account. Accounts are free.

Search for "Jacob Lawrence"

Citations: Amazon will link you to the citations in the book and
- Another function allows a user to search within a chosen title.
- Other functions include links to the **first sentence/first page and sample pages**;
- **Concordance** is an alphabetized list of the most frequently used words in a book, excluding common words such as "of" and "it." Hover your mouse over a word to see how many times it occurs, or click on a word to see a list of book excerpts containing that word.
- In short, Amazon created a fairly comprehensive tool that allows users to get to know a book without actually having to read it."

open *http://www.booksurge.com/*

In 2007, Amazon launched a new program to scan and sell rare and hard-to-find books from universities and public libraries. Readers can purchase hard copies of the digitized books from Amazon on a **print-on-demand basis**. Libraries keep digitial copies and share a portion of any print-on-demand profits.

Open Content Alliance (OCA)
http://www.opencontentalliance.org/
The Open Content Alliance (OCA) is a collaboration of "cultural, technology, nonprofit, and governmental organizations worldwide whose goal is to build a permanent archive of multilingual digitized text and multimedia content." The OCA was conceived by the Internet Archive and Yahoo! in early 2005 as a way to offer broad, public access to a rich panorama of world culture.
- non-profit
- over 80 contributing partners (40+ libraries)
- **total open access**: digital material can be crawled by any search engine
    ◦ **12,000** books/ month added to collection
    ◦ you don't need to login
    ◦ **costs shouldered by partners at $30/book**
    ◦ future plans for print-on-demand and scan-on-demand
- original scope was to scan only out-of-copyright books, but in October 2007 Internet Archive + Boston PL + others announced program to *scan out-of-print but in-copyright rights to be distributed through digital ILL*

Internet Archive Text Archive
http://www.archive.org/details/texts

The Internet Archive is building a digital library of Internet sites and other cultural artifacts in digital form. Like a paper library, they provide free access to researchers, historians, scholars, and the general public.

In the Text Archive section, over 300,000 books are now available and free to read, download, and/or print. Some have restrictions on bulk re-use and commercial use, depending upon the collection or the sponsor of a book. The entire Internet community may contribute, and contributors may upload any genre of text provided they own the rights or the text is in the public domain.

You must search the collection on the Text Archive -- we do not yet have a way of efficiently bringing records for relevant works into WATSONLINE so they may be searched alongside the rest of our collections.

Sub-collections
There are several sub-collections of the text archive: American Libraries, Canadian Libraries, Universal Library, Open Source Books, Project Gutenberg, Children's Library, Biodiversity Heritage Library, and Additional Collections. One of the collections holding particular interest to us are the works uploaded by the Getty Research Institute's Research Library:
http://www.archive.org/details/getty

Searching for and viewing works
I'm going to execute simple searches for **chippendale** (choose "**Hengrave hall...,**" an auction cat not in MMA collection) -- note that you are searching the metadata (title, subject, keywords, etc.) of each text file, *not* the actual full-text of the Text Archive. There is also an advanced search page allowing you to construct more complex searches. [alternate search: **hogarth** (choose "Hogarth's works : with life and anecdotal… vol. 3")]

For the list of search results the default sorting is by relevance, which is a combination of weighting terms from words in the title, author, subject and keywords fields. You will initially see "thumbnails" rotating through the first 20-30 pages of a work, but I recommend turning these off (link at bottom right -- preference is retained in a browser cookie) since it really slows down page loading.

On the right you have several ways to sort or refine the results.

File formats
Choose one of your results and you see on the left that you have several options for viewing the text -- PDF, Flip Book, DjVu, TXT, and FTP. Once you have one of these viewing formats open, you may search within the text for terms important to you.

Chippendale catalog – search within for **Holbein**.
- The PDF (black and white or color) is the best option. Printing, navigating and searching the work is easy and familiar. You may save the file to your computer. If you're viewing it through the Archive's web site it's a little slow, but performance improves if you save it to your computer and open it from there.
- For viewing, the Flip Book is a really nice option, though printing is impossible and there is no way to go to a particular page outside of clicking to it. The search function works well and provides you with clickable page markers that include contextual words upon hovering over them.
- Even though evidence points to its being a better file format for scanned books than PDF, DjVu crashes after turning a few pages in Firefox and doesn't work particularly well in Internet Explorer. Plus, it requires a browser plugin.
- The FTP link provides you with a directory list of files to be downloaded for that particular work -- if you know you want to save a particular type of file from the outset, go here first.

My advice is to use PDF until another format proves to be better.

Evaluative comments
- A rich resource but it's hard to know what you'll find or what's contained in the text archive.
- There can be a lot of junk to wade through – all sorts of people are uploading all sorts of content and it's often hard to distinguish the scholarly from the non-scholarly.
- Both of those reasons are good ones to simply take some time to explore the archive and try out a variety of searches.

The Open Library
http://demo.openlibrary.org/
- Funded by the Internet Archive, emulated after a British Library book reading machine. This is the reading interface for OCA, and very much still in beta. **OL's mission is to build the world's greatest library, then put it up on the Internet free for all to use and edit.**
  - Non-commerical
  - completel open source (code included) **http://demo.openlibrary.org/b/South_Indian_goddesses?m=edit**
  - *open http://demo.openlibrary.org/b/Flowers_from_val_history*
  - Unlike Google Book Search, Open library aims to create a reading experience.

Universal Digital Library or the Million Books Project
http://tera-3.ul.cs.cmu.edu/index.html
Founded in 2001, one of the earliest projects to make e-books freely available.
Supported by Carnegie Mellon University, National Science Foundation, governments and libraries in China, India and Egypt.
- can browse by century: http://tera-3.ul.cs.cmu.edu/ULIBOurCollections.htm
- requires plug-in for Dejavu application to launch

CONCLUSION
Honorable mentions – there are plenty of other electronic book collections out there and we'd like to mention a few of them. All may be found on our Watson Library del.icio.us page, which is openly accessible to anyone on the Internet. Simply click the tag "E-Books" to get a list of resources we've added – and we'll continue to add more as we find ones relevant to our collections.

Watson Library del.icio.us page for E-Books
http://del.icio.us/watson_library/E-Books -- some of the best here are:
- Gutenberg E (Columbia U)
- Wisconsin's decorative arts materials
- Perseus digital library
- Digital Book Index
Thanks for reading!

http://libmma.org/digital_files/EBooksGuide.pdf
Last updated June 18, 2008.