Are We Ready for the Digital Humanities?
What are the Digital Humanities?

How can libraries support DH?
Humanities

• Philosophy, literature, religion, art, music, history and language.
• Core library users.
• Core print users.
Digital Humanities

Opening up new knowledge and new ways of learning through the application of digital technologies to any humanities subject.

DH is about creating digital toolsets that allow users to undertake new forms of research in the humanities.

Growing interest and investment.
George Mason: CHNM
PROJECTS

SEMANTIC FIELDS

“ABSTRACT VALUES IN THE 19TH CENTURY ENGLISH NOVEL: DECLINE AND TRANSFORMATION OF A SEMANTIC FIELD,” BY LONG LE-KHAC AND RYAN HEUSER

[FOR MORE INFORMATION, EMAIL LLEKHAC@STANFORD.EDU.]

Our paper analyzes the historical behavior of several semantic fields of “abstract values” in a corpus of 2,822 19th-century British novels. In his classic study, Culture and Society , Raymond Williams claimed that a group of keywords that arose and/or changed dramatically in the nineteenth century
Digital Research & Scholarship in Alderman Library

The Scholars’ Lab caters to the digital research and scholarly analysis needs of faculty and advanced students in the humanities and social sciences. Staffed with friendly, expert consultants from the U.Va. Library’s Digital Research and Scholarship unit, the Scholars’ Lab is the perfect place to take your work to the next level. More...
Digital Humanities in Canada

- U Alberta
- U Toronto
- U Victoria
- UNB
- U Montreal
- McMaster
- McGill
- York
Digital Humanities Projects
Thematic Digital Archives

The Eve of War
- 1860 Statistics
- Church Records
- Census & Tax Records
- Site Map

Letters & Diaries
Reference Center
Newspapers

Spring 1859 to Spring 1861

The War Years
- Images
- Official Records
- Battle Maps
- Soldiers’ Records

Letters & Diaries
Reference Center
Newspapers

Spring 1861 to Spring 1865

The Aftermath
- 1870 Statistics
- Freedmens Bureau
- Census & Veteran Records
- Memory of the War

Letters & Diaries
Reference Center
Newspapers

Spring 1865 to Fall 1870
Virtual Anthologies

Gravures, c. 1630–c. 1715

À partir de 1630, l'invective contre les femmes et le mariage passe du texte à l'image.
[Remarques introductives.]

Galerie (1 – 15 / 71) ➞
Aggregate, Annotate, Review

<table>
<thead>
<tr>
<th>Featured Object</th>
<th>Browse Recent Tags</th>
<th>Recent News</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sarah Langhorne Cadwallader</strong>&lt;br&gt;&lt;br&gt;Date: about 1850-1865&lt;br&gt;Site: The Willa Cather Archive</td>
<td>catalogue_raisonne • francis • slavery • new_orleans • roman_de_la_rose • etiquette_manuals • art-history • fashion • portrait • byron • mardi_gras • american-civil-war • music • piano_forte • emerson • eulogy • lincoln • patriotism • elections • map • longfellow • daguerreotype • disease • medicine • new_york • hayes • mayor • suffrage • native-americans • emancipation • bourgeois</td>
<td>• Photo(s) of the Week: Abraham and Mary Todd Lincoln&lt;br&gt;• Music and Digital Humanities&lt;br&gt;• Photo of the Week: A Special Wedding Cake&lt;br&gt;• Photo of the Week: St George and the Dragon&lt;br&gt;• Photo of the Week: Coney Island Baby</td>
</tr>
</tbody>
</table>
GIS & Mapping
Clustering & Visualization
The Proceedings of the Old Bailey, 1674-1913

A fully searchable edition of the largest body of texts detailing the lives of non-elite people ever published, containing 197,745 criminal trials held at London's central criminal court. If you are new to this site, you may find the Getting Started and Guide to Searching videos and tutorials helpful.

To search the Proceedings use the boxes on the right or go to the Search Pages.

New Features: Research and Study Guides and Personal Workspaces
Edition Comparisons
The Digital Michelangelo Project:
3D Scanning of Large Statues

A computer rendering made from a 3D model of Michelangelo's David. The model was built by scanning the statue using a laser triangulation rangefinder and assembling the resulting range images to form a seamless polygon mesh. The mesh contains 8 million polygons, each about 2.0 mm in size. The raw data from which the mesh was built contains 2 billion polygons, representing range samples spaced 0.25 mm apart on the statue surface. Although we also digitized the statue's color, the veining and reflectance shown here are artificial. The rendering includes simulated subsurface scattering, but with arbitrary parameters. Thanks to Henrik Wann Jensen for computing this image.

Click above for a 700 x 700 pixel version
Click here for a 2K x 2K version
(388 KB)
(frontispiece of Siggraph 2000 proceedings)
Welcome to TAPoR @ UAlberta!

TAPoR is the Text Analysis Portal for Research, a collaboration by six Canadian universities to build a centralized gateway to representative texts and sophisticated text analysis tools. The computing infrastructure is available to host and support research projects using text, text encoding, text transformation, and XML technologies. For more information on the TAPoR in Canada initiative, please consult our TAPoR Canada page.

TAPoR @ UAlberta is a hub for text-based computing research in the Faculty of Arts at the University of Alberta. Associated projects
Anatomy of a DH Project

Lancaster Newsbooks Corpus: Query launcher

Word Search
Search for all instances of a word in the corpus by entering the word-form into the search-box below

- Use no wildcards
- Use SQL wildcards
- Use regular expressions

Run the search  Reset the search

Create a concordance
Word search  Tag search  Word-and-tag search  Phrase search

Dunning et al, “Freeing up digital content with text mining”, *Serials*. 22(2), July 2009
Text Corpus

• 17\textsuperscript{th}C English news pamphlets
• Dec 1653 to May 1654
• British Library electronic texts
• 312 documents, 800 000 words
CLAWS: Part of Speech Tagging

In the Patrick of Liverpool — which we lately recovered from the Brest men of War — was one Walter Roche — who was to carry her to Brest — and he

- Grammar parsing tool (Linguistics)
- Identifies proper nouns, common nouns, plural nouns, adjectives, prepositions...
- 95–97% accuracy
The **World Gazetteer** provides a comprehensive set of *population data* and related *statistics*.

**Population data arranged by country**

- Current population figures and area size for all *countries* and important territories
- Population figures and area size of *administrative divisions*
- Population figures for the largest *cities, towns* and *places* as well as for metropolitan areas
- Historical population data (census or estimates; mostly from last or last two censuses)
- Overview maps that show the position of the country or the administrative divisions

**Related data**

- **Summary statistics** for countries and territories as well as for the largest cities and agglomerations
- Current *national flags*
- Summary maps for various subject
- A pronunciation guide for some languages
- There is a downloadable file for offline research
- A population clock that show the current population of the world
- Comprehensive meta informationen about statistics and the project

*Last change:* 2010-01-01
Geographical Analysis
USAS: Semantic Parsing

... two_N1 ships_M4 from_Z5 Dunkirk_Z2 have_Z5 brought_M2 Men_S2.2m Arms_B1 ,_PUNC and_Z5 Ammunition_G3 to_Z5 Middleton_Z1mf

Ships = M4 (shipping, swimming, etc.)
Men = S2.2m (People:-Male)
Ammunition = G3 (warfare, defence and the army; weapons)
3. Place-name map generated in response to a search for words related to the concept of ‘war’
GIS Mapping: Topic ‘Money’

Place-name map generated in response to a search for words related to the concept of ‘money’
Where’s the Beef? Does Digital Humanities Have to Answer Questions?

May 12th, 2010 · 11 Comments

The criticism most frequently leveled at digital humanities is what I like to call the “Where’s the beef?” question, that is, what questions does digital humanities answer that can’t be answered without it? What humanities arguments does digital humanities make?

Concern over the apparent lack of argument in digital humanities comes not only from outside our young discipline. Many practicing digital humanists are concerned about it as well. Rob Nelson of the University of Richmond’s Digital Scholarship Lab, an accomplished digital humanist, recently ruminated in his THATCamp session proposal, “While there have been some projects that have been developed to present arguments, they are few, and for the most part I sense that they haven’t had a substantial impact among academics, at least in the field
DH Library Fears

The Humanists are leaving us for the Computer Scientists.
Digital Humanities Sources
Since I am likely to be some days in this hospital for treatment, I am going to make use of my enforced leisure, and fill in the time by writing out roughly some of the things that have happened in my life which are a bit out of the common. I shall scumble them roughly and hope that at some future day of leisure - not in hospital for treatment, I hope - I shall be able to put them in proper, or at least more presentable, form.

Cluny Macpherson

Woolavington Wing
Middlesex Hospital
London, W1

Friday Oct. 30, 1936
Primary source materials
Primary source materials
Digital Images of Yale’s Vast Cultural Collections Now Available for Free
posted: May 10, 2011

Scholars, artists and other individuals around the world will enjoy free access to online images of millions of objects housed in Yale’s museums, archives, and libraries thanks to a new "Open Access" policy that the University announced today. Yale is the first Ivy League university to make its collections accessible in this fashion, and already more than 250,000 images are available through a newly developed collective catalog.

The goal of the new policy is to make high quality digital images of Yale's vast cultural heritage collections in the public domain openly...
Large text corpora
Large text corpora

ECCO-TCP

The University of Michigan, the University of Oxford, and Gale have cooperated in a Text Creation Partnership to create 2,231 accurately keyed and fully searchable SGML/XML text editions from among the 150,000 titles available in the Eighteenth Century Collections Online (ECCO) corpus. These texts are now fully available to the public in two formats: plain text editions can be requested from 18thConnect, an online resource initiative in 18th century studies, and sgml/xml editions can be requested directly from our office. ECCO is an important research database that includes every significant English-language and foreign-language title printed in the United Kingdom during the
Please note that some of our resources do not appear in this table. Reasons for this might include that they were deposited with us for preservation only, and not dissemination, or that our legal right to disseminate them has been questioned.

<table>
<thead>
<tr>
<th>ID</th>
<th>Availability</th>
<th>Title</th>
<th>Language</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2541</td>
<td>free</td>
<td>VU Amsterdam Metaphor Corpus</td>
<td>English</td>
<td>Gerard J Steen; Aletta G Dorst; J Berenike Herrmann; Anna A Kaal; Tina Krennmayr</td>
</tr>
<tr>
<td>2540</td>
<td>free</td>
<td>Speech, Thought and Writing Presentation Corpus (STWP)</td>
<td>English</td>
<td>Culpeper, Jonathon; Semino, Elena; Short, Mick; Wynne, Martin</td>
</tr>
<tr>
<td>2539</td>
<td>restricted</td>
<td>British Academic Written English Corpus</td>
<td>English</td>
<td>Nesi, Hilary; Gardner, Sheena; Thompson, Paul; Wickens, Paul</td>
</tr>
<tr>
<td>2537</td>
<td>free</td>
<td>GerManC. A Historical Corpus of German Newspapers 1650-1800</td>
<td>German</td>
<td>Durrell, Martin; Ensslin, Astrid; Bennett, Paul (ed.)</td>
</tr>
<tr>
<td>2531</td>
<td>free</td>
<td>The Lancaster Newsbooks Corpus</td>
<td>English</td>
<td>Thomason, George, d. 1666</td>
</tr>
<tr>
<td>2530</td>
<td>restricted</td>
<td>Language convergence and grammatical borrowing database</td>
<td>English</td>
<td></td>
</tr>
</tbody>
</table>
The digital repository HathiTrust, in partnership with technology centers at Indiana University and the University of Illinois at Urbana-Champaign, recently launched the HathiTrust Research Center, which will leverage the two academic institutions' considerable computing power, and HathiTrust's millions of digitized words, for text-mining projects serving the humanities.
Large text corpora

[Google Books]

Proposed settlement

In-copyright works owned by universities

Non-consumptive purposes

One or two centres
Linguistic Corpora

About the BNC

The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of current British English, both spoken and written. [more]

Search the Corpus

Type a word or phrase in the search box and press the Return key on your keyboard to see up to 50 random hits from the corpus.

Look up: Enter text [Go]

You can search for a single word or a phrase, restrict searches by part of speech, search in parts of the corpus only, and much more.

The search result will show the total frequency in the corpus and up to 50 examples. [more information]
Historical Linguistic Corpora

PennParsed Corpora of Historical English

The Penn Historical Corpora, including the Penn-Helsinki Parsed Corpus of Middle English, second edition (PPCME2), the Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME), and the Penn Parsed Corpus of Modern British English (PPCMBE), are syntactically annotated corpora of prose text samples of English from the indicated time periods. Their syntactic annotation (parsing) permits searching not only for words and word sequences, but also for syntactic structure. The corpora are designed for the use of students and scholars of the history of English, especially the historical syntax of the language, and they are publicly available.

The Penn Parsed Corpora of Historical English are available on CD-ROM. Site licenses for departments/research groups and libraries are also available. See the corpus Order Form for charges. Corpus license fees go toward improving the corpora and increasing them in size. Upgrades, when completed, are available to corpus license holders at a modest cost.

If you currently hold a license for the two older corpora, the PPCME2 and the PPCEME, and wish to obtain the PPCMBE, please contact akroch AT gma i l DOT com for upgrade information.

The Penn Historical Corpora are distributed with a search program CorpusSearch2, written by Beth Randall and released as open source software downloadable from its Sourceforge project website.

- The PPCME2 was created with the support of the National Science Foundation (Grants BNS89-19701 and SBR95-11368), with supplementary support from the University of Pennsylvania Research Foundation.
Welcome to NHGIS

The National Historical Geographic Information System (NHGIS) provides, free of charge, aggregate census data and GIS-compatible boundary files for the United States between 1790 and 2000. NHGIS offers the following options to obtain data:

- The NHGIS Shape Finder provides access to boundary files for mapping data in a GIS software package.

- The NHGIS Data Finder provides access to aggregate census data.

- Social Explorer allows online mapping of NHGIS data

Funding for the NHGIS provided by the National Science Foundation and the National Institutes of Health.
Historical Census Datasets

The Programme de recherche en démographie historique (PRDH), Département de Démographie, Université de Montréal, hosts several research and database construction projects based on the study and use of nineteenth-century Canadian historical censuses. Our current historical census database construction projects focus upon the Canadian censuses of 1852 and 1881. Together, these projects offer tremendous opportunities for historical, demographic and genealogical research.

The 1852 Canadian census was the first Canadian census to collect individual-level information on all household members. The 1852 Canadian census database based on this enumeration represents 20% of the population of Upper Canada and Lower Canada which was enumerated in 1852 and for which schedules of the manuscript census survive, an equivalent of 235,341 cases. The 1852 Census will also serve to close the period of observation covered by the Registre de la population du Québec ancien, a family reconstitution database which comprises Catholic diocesan records and provides identification which blocks identify individuals who were enumerated multiple times in the 1852 Census.
Etext Preferences: Cost

Free
Institution site-license
Paid for by grants
Same price as print
On-line subscription I pay for

Percent of Respondents

Important
Unimportant
Etext Preferences: Quality

- Reliable institution
- Peer reviewed
- Based on established edition
- Reputable publisher
- References to authoritative edition
- Available with original images
- References to student’s authoritative edition

Percent of Respondents

[Bar chart showing preferences with categories and data points]
Etext Preferences: Availability

- Available legally
- Stable form
- With documentation
- On web

Percent of Respondents

[Bar chart showing preferences with legends: Important (white) and Unimportant (black).]
Licensing & Rights

• Multi-layered sources with different licensing conditions.
• Copyright over annotations, data sets, other user-generated information.
• Need for clearly expressed rights info.
Copyright review project

Copyright Review Management System - IMLS National Leadership Grant

In 2008, the University of Michigan Library was awarded a National Leadership Grant from the Institute of Museum and Library Services to create a Copyright Review Management System (CRMS). The purpose of the project is to increase the reliability of copyright status determinations of books published in the United States from 1923 to 1963 in the HathiTrust Digital Library, and to help create a point of collaboration with other institutions. Hundreds of thousands of books were published in the United States between 1923 and 1963, and although many of these are likely to be in the public domain, individuals must manually check their copyright status to make a determination. If a work is not in the public domain, it may still be in copyright and can be licensed.
Library Role: Sources

- Acquire & preserve primary sources
- Digitization & transcription
- Open, flexible formats
- Large text aggregations
- Data sets
- Licensing & copyright
Digital Humanities Metadata
Traditional Metadata

- Access to print collections
- Authorities
- Thesauri
- Bibliographies
- Indexes
- Concordances
# Thomas Jefferson papers, 1606-1943

## Overview

<table>
<thead>
<tr>
<th>Title</th>
<th>Thomas Jefferson papers, 1606-1943</th>
</tr>
</thead>
<tbody>
<tr>
<td>Span Dates</td>
<td>1606-1943 (bulk 1775-1826)</td>
</tr>
<tr>
<td>Bulk Dates</td>
<td>MSS27748</td>
</tr>
<tr>
<td>ID No.</td>
<td>Jefferson, Thomas, 1743-1826</td>
</tr>
<tr>
<td>Extent</td>
<td>25,000 items: 225 containers plus 9 oversize; 90 linear feet; 65 microfilm reels</td>
</tr>
<tr>
<td>Language</td>
<td>Collection material in English</td>
</tr>
<tr>
<td>Location</td>
<td>Manuscript Division, Library of Congress, Washington, D.C.</td>
</tr>
<tr>
<td>Summary</td>
<td>United States president, vice president, and secretary of state; diplomat, architect, inventor, planter, and philosopher. Correspondence, official statements and addresses, including a rough draft of the Declaration of Independence, plantation and personal accounts, notebook, fee book, case book, garden book, farm book, calculations of interest, records of early Virginia laws and history and other writings on political, legal, educational, and scientific matters, newspaper clippings, and other papers.</td>
</tr>
</tbody>
</table>
Finding Aids: Data Sets
Name Authorities

Sources for London Lives

A fully searchable edition of 240,000 manuscripts from eight archives and fifteen datasets, giving access to 3.35 million names.

To search London Lives use the boxes on the right or go to the Search Pages.
Name Authorities

The Getty Research Institute

Cultural Objects Name Authority™ Online

Copyright information

Accessing CONA
The CONA online search module is under development. It is scheduled to go live in early 2012. In the meantime, please contact us at vocab@getty.edu if your institution wishes to contribute to CONA. As with the AAT, TGN, and ULAN, CONA grows through contributions from the user community.

About CONA
Learn about the purpose, scope, and structure of CONA. CONA comprises authority records for cultural works, including architecture and movable works such as paintings and sculpture.

F.A.Q. about CONA
Here you will find a list of the most frequently asked questions about CONA.

CONA@getty.edu
European Language Social Science Thesaurus (ELSST)

**ELSST** is a broad-based multilingual thesaurus for the social sciences. Originally based on the monolingual thesaurus Humanities and Social Science Electronic Thesaurus (HASSET) of the UK Data Archive at the University of Essex, ELSST has been developed over the years by the members of the Council of European Social Science Data Archives (CESSDA) for use in an international setting. Its main aim is to facilitate access to data resources across Europe, independent of domain, resource, language and vocabulary. ELSST is used to aid retrieval in the multinational CESSDA data portal.

The source version of ELSST is the English thesaurus. The thesaurus is currently translated into eight other languages: Danish, Finnish, French, German, Greek, Norwegian and Spanish and Swedish. Translation into further languages is planned.
Full Text Markup
Text Encoding Initiative (TEI)
<div n="castlist" type="DramatisPersonae"
     org="uniform" sample="complete"> <castList>
    <head>Dramatis Personae</head>
    <castItem type="role">
      <role xml:id="wag">Wagner</role>
    </castItem>
</castList>
...</div>
For is he not Corpus naturale?

For is he not Corpus naturale?

For is he not Corpus naturale?
Nothing so sweet as choice
<orig>Magicke</orig>
<reg>magic</reg>
(choice) is to him;
<metDecl pattern="((+|-)+\|/??)*"> 
  <metSym value="trochee" terminal="false">+-</metSym> 
  <metSym value="iamb" terminal="false">-+</metSym> 
  <metSym value="spondee" terminal="false">++</metSym> 
  <metSym value="pyrrhic" terminal="false">--</metSym> 
  <metSym value="amphibrach" terminal="false">-+-</metSym> 
  <metSym value="anapaest" terminal="false">--+</metSym> 
  <metSym value="+">metrical prominence</metSym> 
  <metSym value="-">metrical non-prominence</metSym> 
  <metSym value="|">foot boundary</metSym> 
  <metSym value="/">metrical line boundary</metSym> 
</metDecl>
“Fury said to a mouse, That he met in the house, 'Let us both go to law: I will prosecute you.— Come, I'll take no denial; We must have a trial: For really this morning I've nothing to do.' Said the mouse to the cat, ...
SUNDAY, 20. Battle of Chickamauga, Battle of, Ga., 1863

May the Lord grant that I may never spend another day, like this, in danger. We changed position and at 1 P.M. we went into heavy action and all the dread and horrid scenes of Battle and fought till night and ammunition was gone, with great credit, but much loss to the 121st O.V.I. God alone spared me.
On the march A splendid fire of Cedar rails afforded us to make a
<seg type="food">cup of coffee and fry fresh mutton</seg></p>
Metadata: Library Role

- Metadata crosswalks
- Controlled vocabularies
- Thesauri & taxonomy
- Name authorities
- Text mark-up
Digital Humanities Preservation
Preservation: Library Role

• Storage infrastructure
• Digital objects & texts
• User generated annotations
• User generated data sets
• Software environments & tools
• Preservation metadata
• Long term access
Library Support for Digital Humanities

• Sources
• Licensing
• Digitization
• Metadata
• GIS & Data
• Sharing
• Preservation
Thank you.

Lisa Goddard
Scholarly Communications Librarian
lgoddard@mun.ca
http://www.library.mun.ca/goddard.php