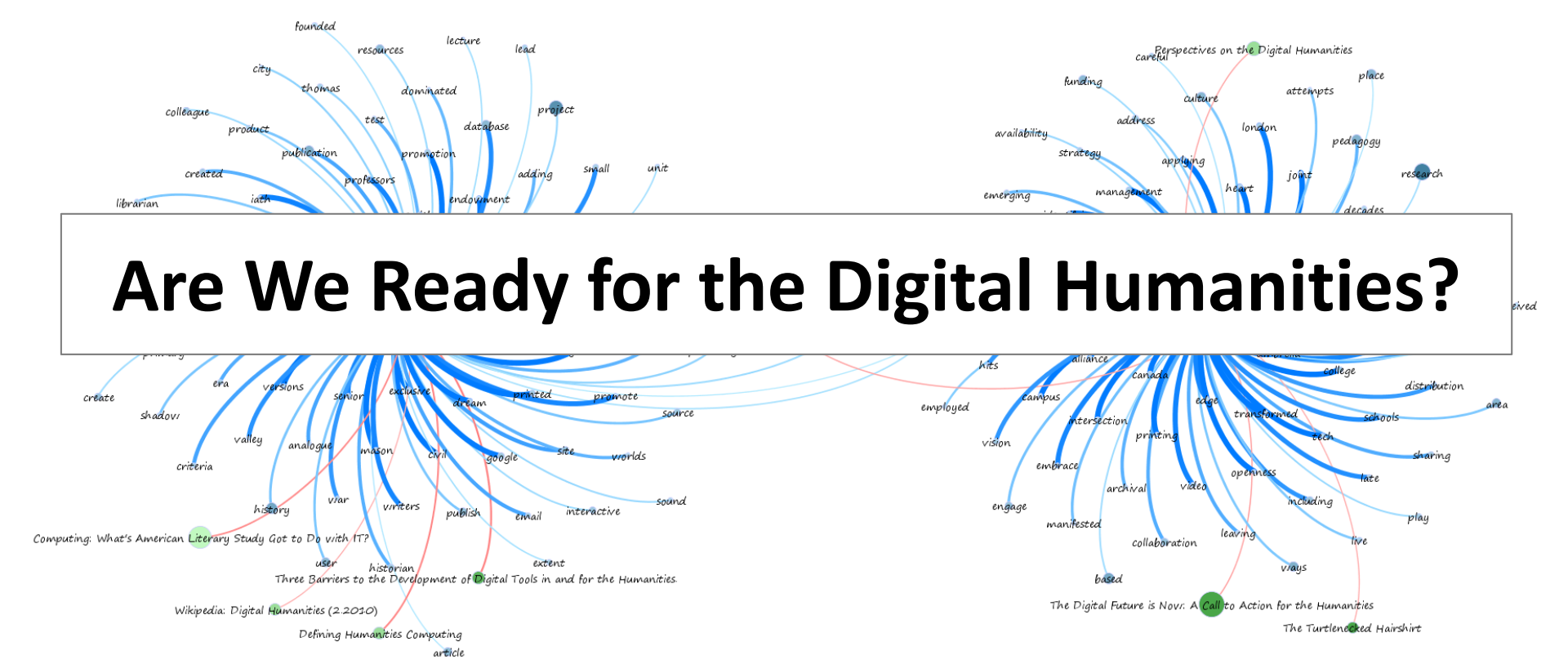



# Are We Ready for the Digital Humanities?

Atlantic Provinces Library Association
Lisa Goddard
Memorial University Libraries
May 2011

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.

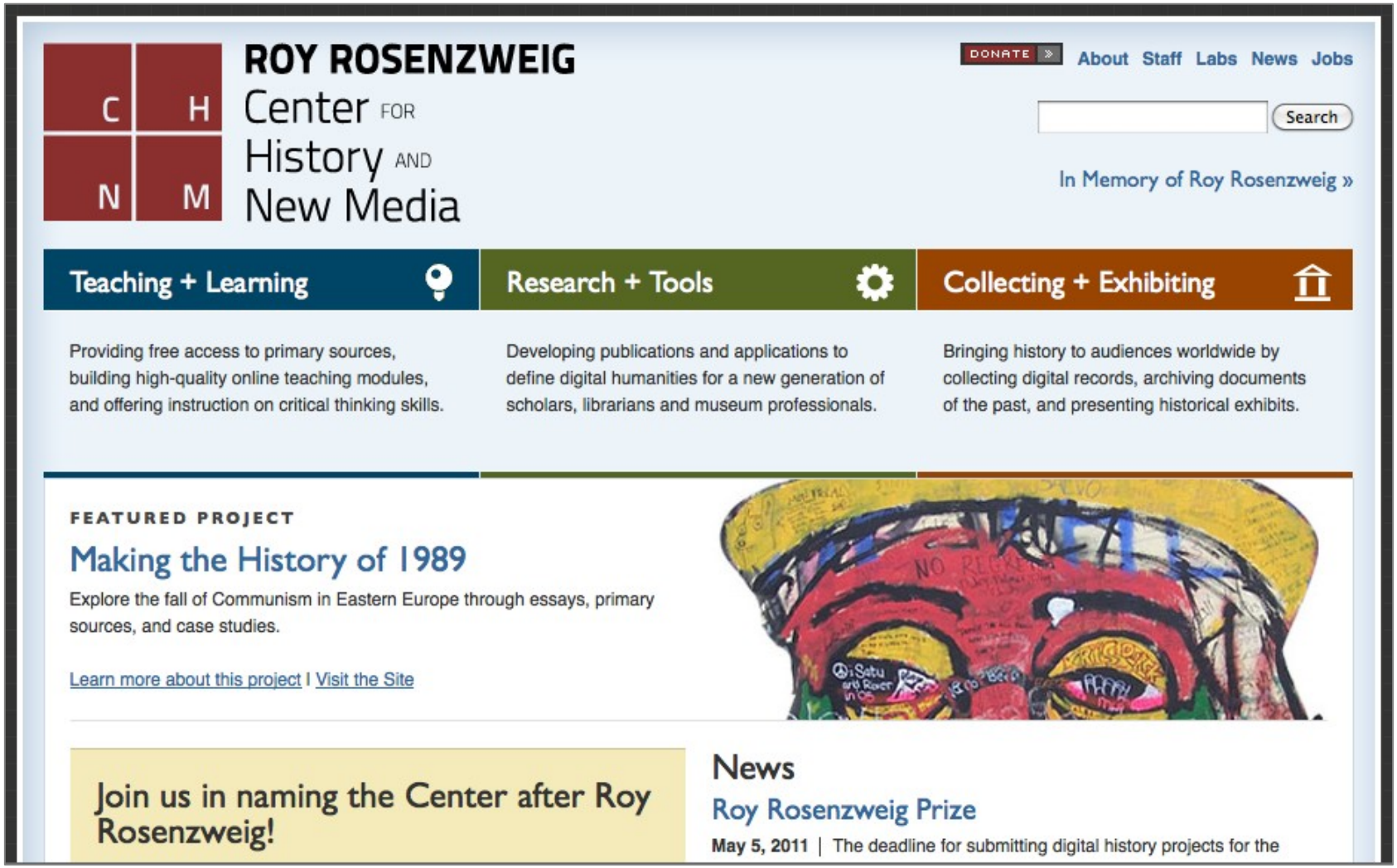
# Digital Humanities

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



# Stanford Literary Lab

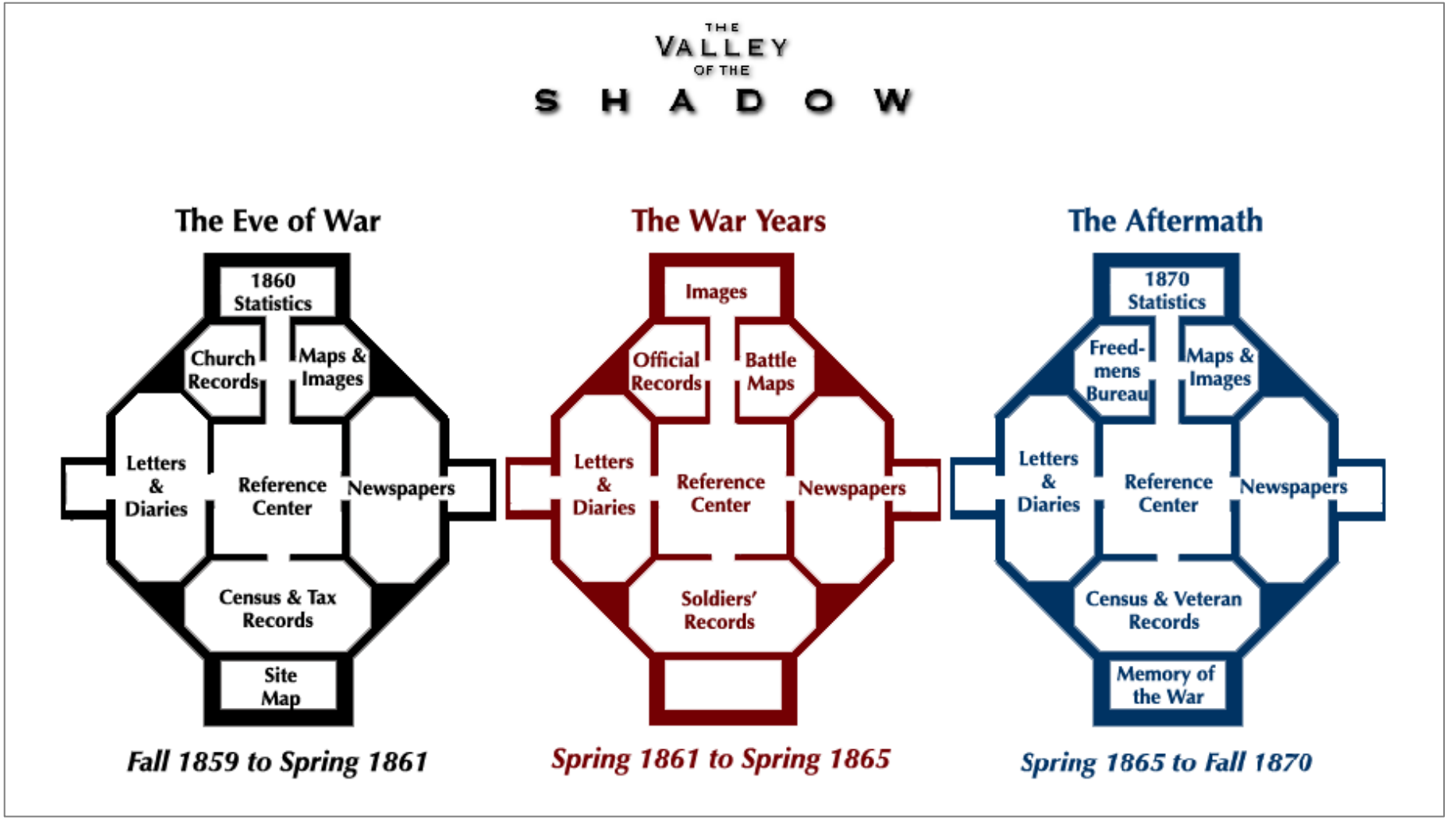
# U Virginia Scholar's Lab

# Digital Humanities in Canada

- U Alberta
- U Toronto
- U Victoria
- UNB
- U Montreal
- McMaster
- McGill
- York

# Digital Humanities Projects

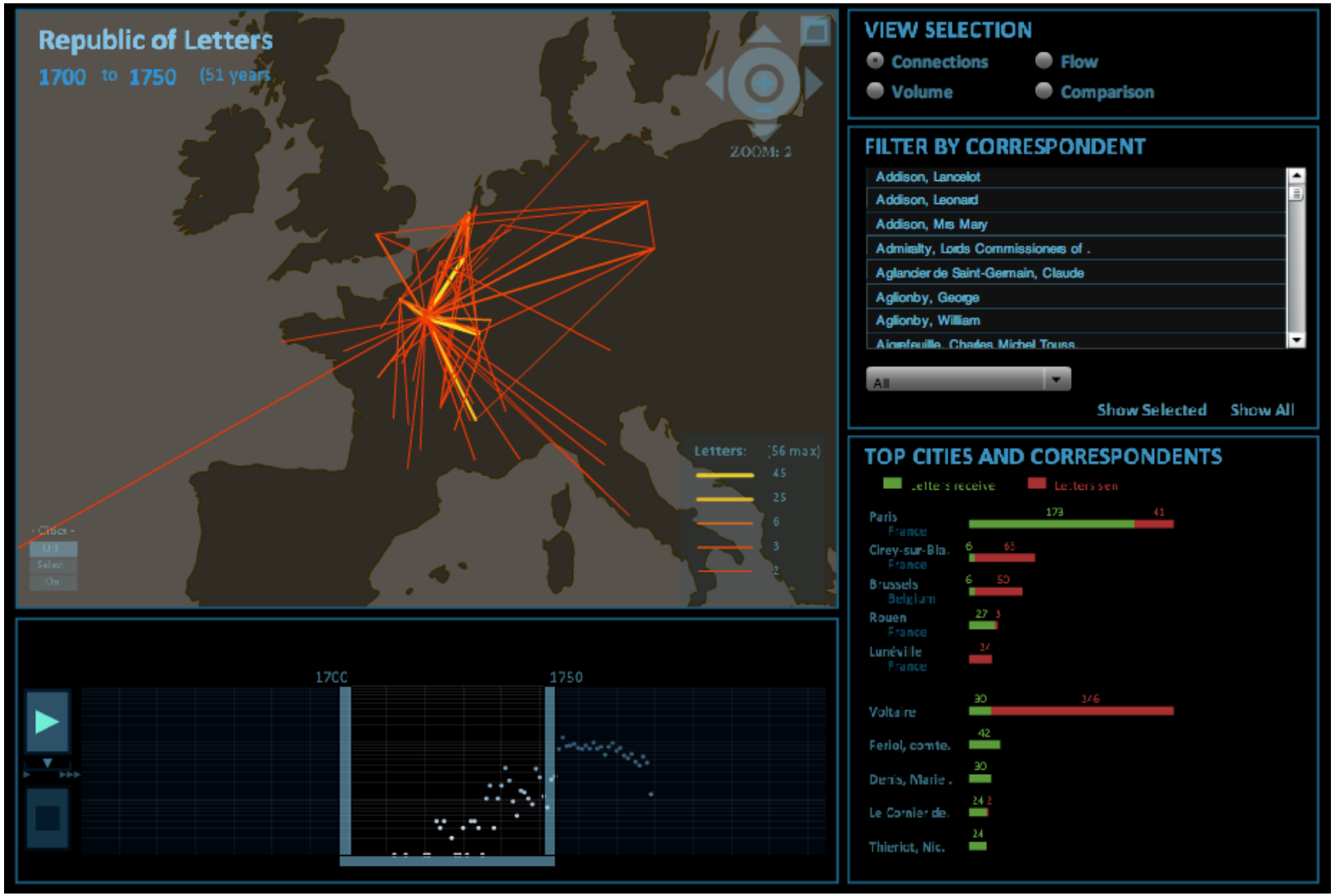
# Thematic Digital Archives



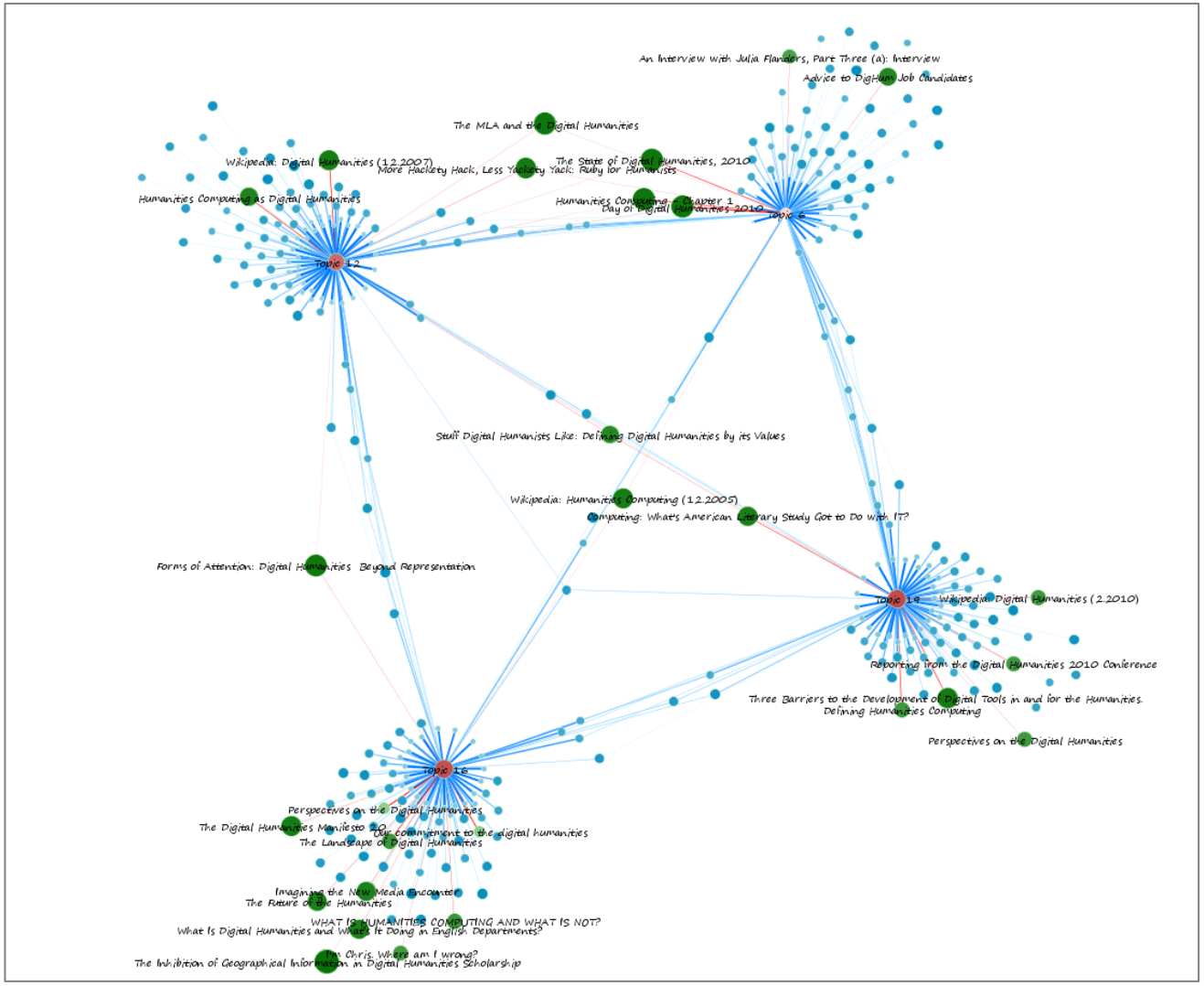
# Virtual Anthologies

# Aggregate, Annotate, Review

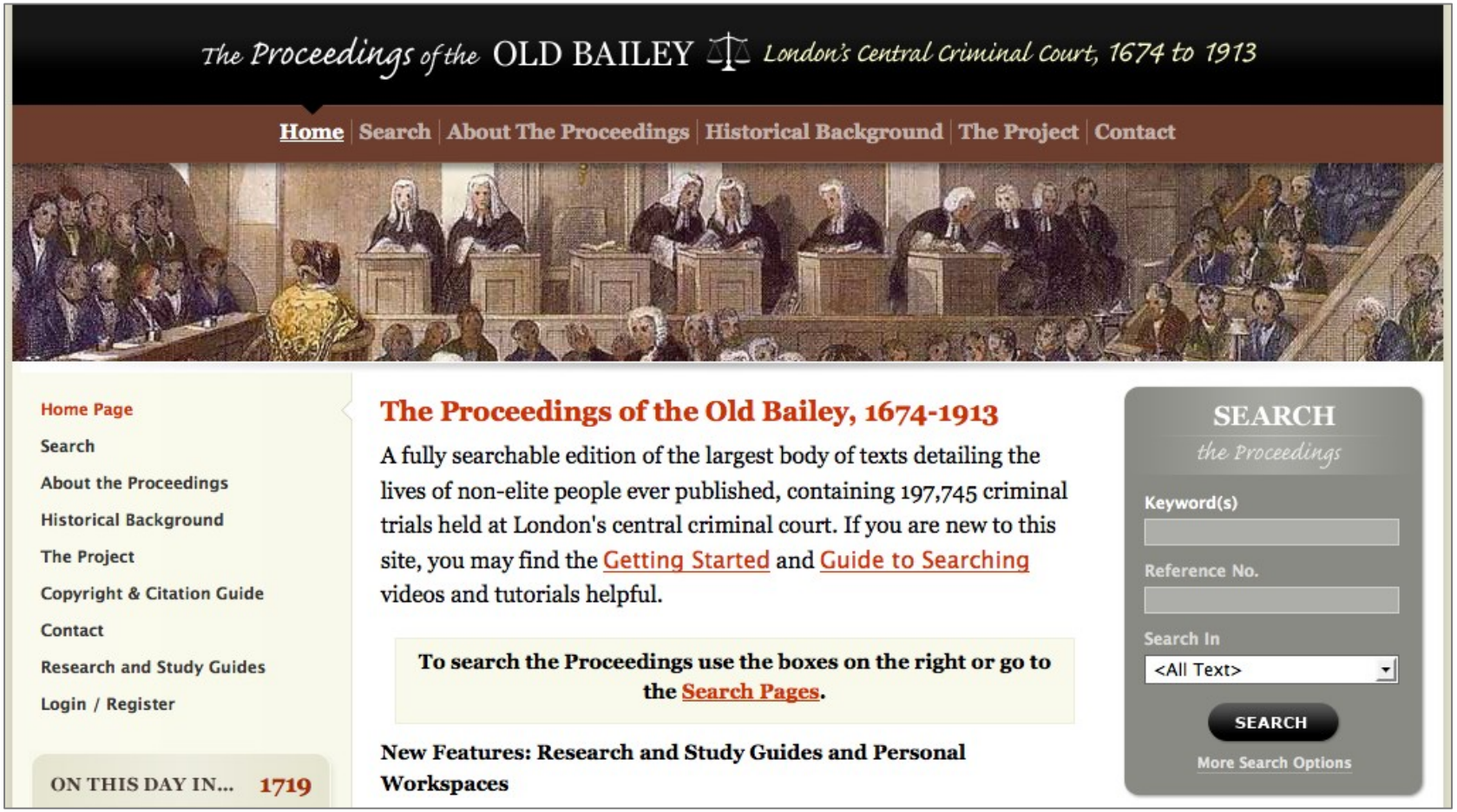
# GIS & Mapping


Republic of Letters
1700 to 1750 (51 years)
ZOOM: 2
Letters: (56 max)
45
25
6
3
VIEW SELECTION
Connections
Flow
Volume
Comparison
FILTER BY CORRESPONDENT
Addison, Lancelot
Addison, Leonard
Addison, Mrs Mary
Admiralty, Lords Commissioners of .
Aglancier de Saint-Germain, Claude
Aglionby, George
Aglionby, William
All
Show Selected
Show All
TOP CITIES AND CORRESPONDENTS
Paris
France
173
41
Cirey-sur-Bla.
France
6
65
Brussels
Belgium
6
50
Rouen
France
27
3
Lunéville
France
34
Voltaire
30
246
Feriol, comte.
42
Denis, Marie .
30
Le Cornier de.
24
2
Thieriot, Nic.
24
1700
1750

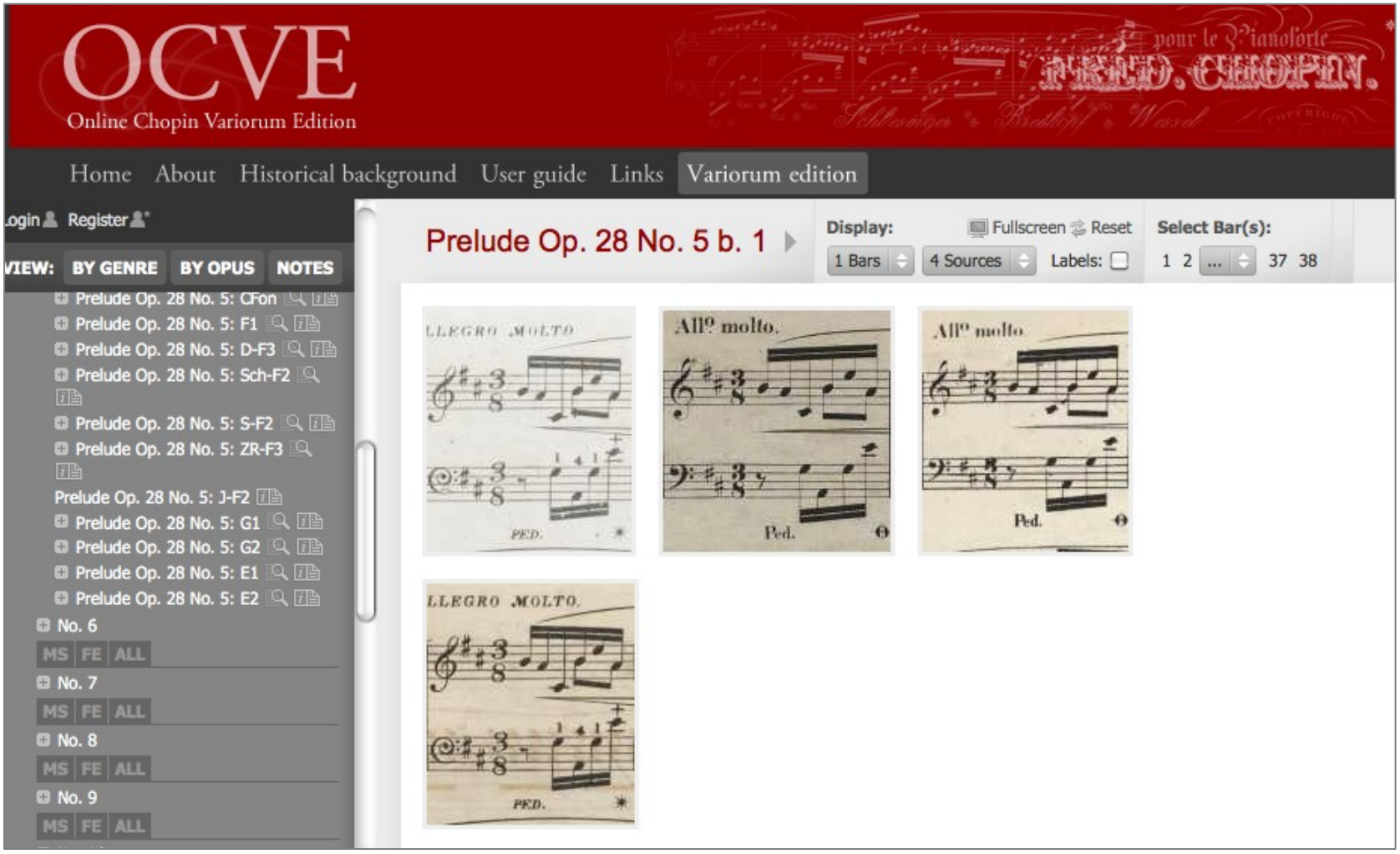
# Clustering & Visualization



# Text Transcription & Markup

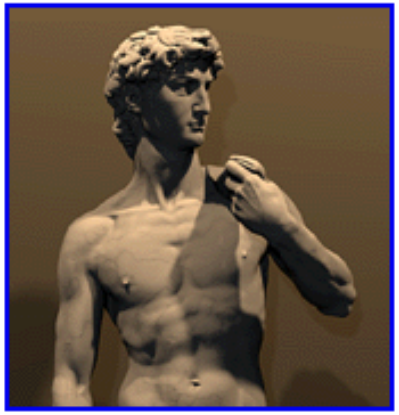


# Edition Comparisons



# 3D Sculpture Modeling

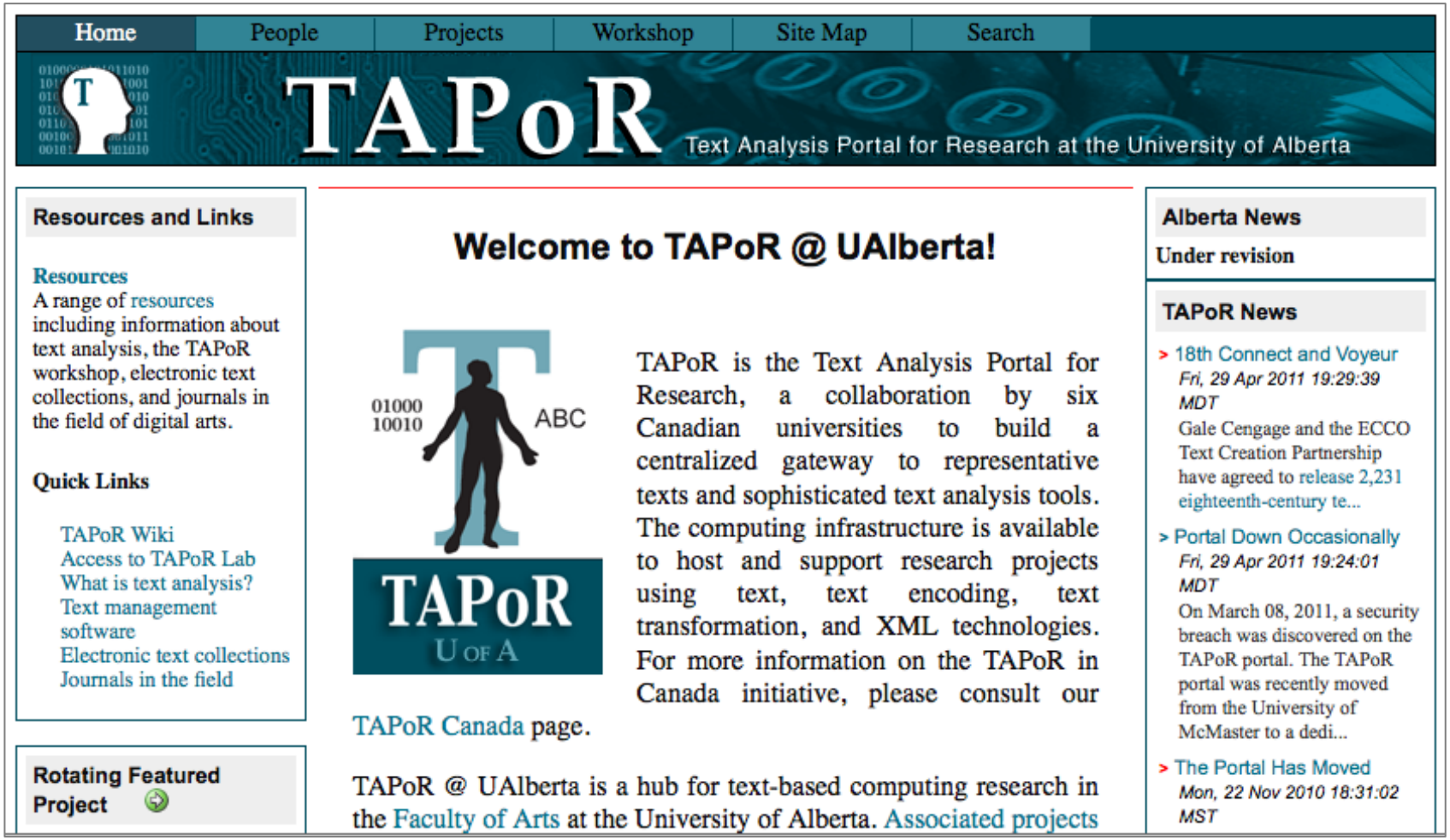
## The Digital Michelangelo Project: 3D Scanning of Large Statues



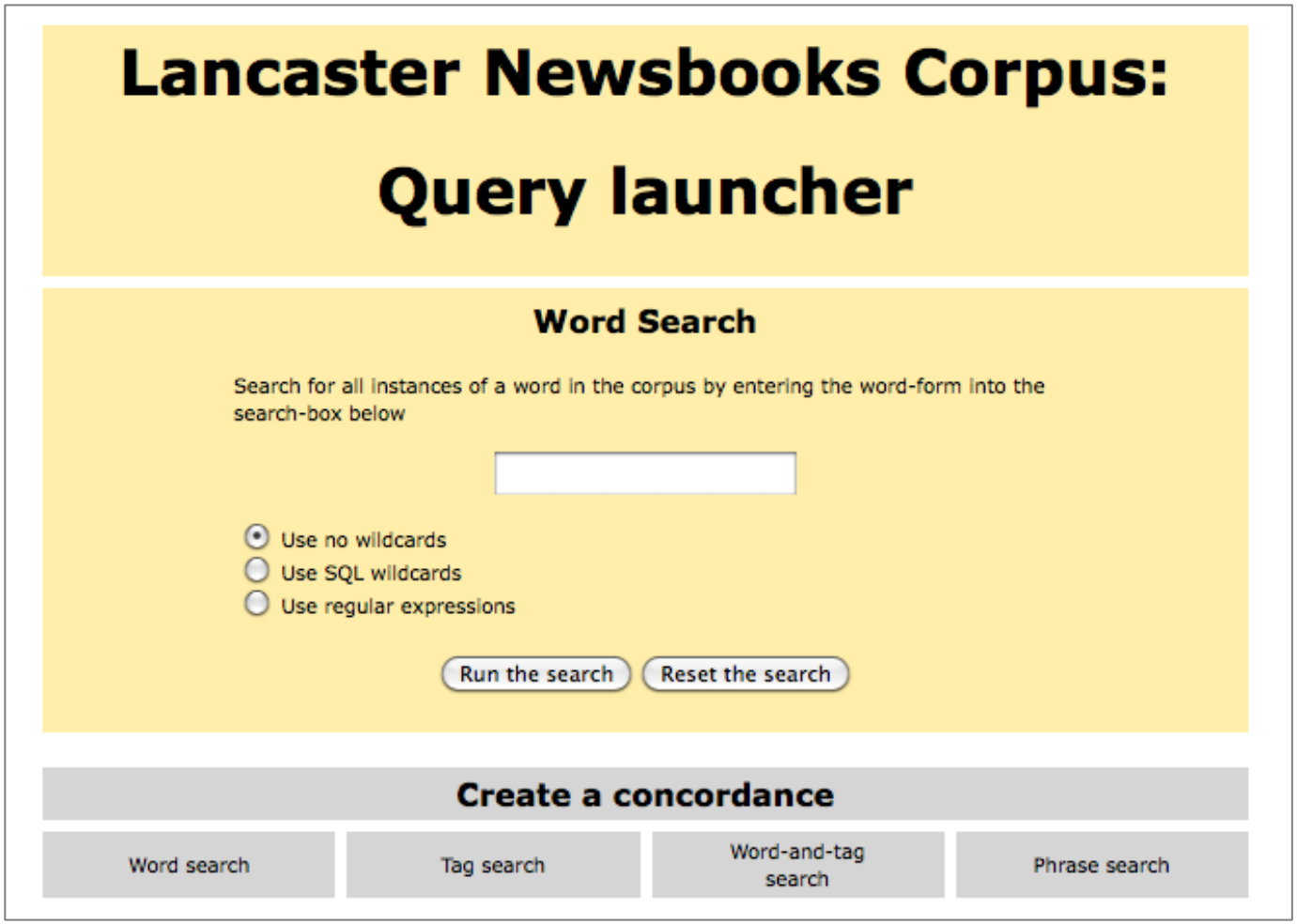
A computer rendering made from a 3D model of Michelangelo's David. The model was built by scanning the statue using a laser triangulation rangefinder and assembling the resulting range images to form a seamless polygon mesh. The mesh contains 8 million polygons, each about 2.0 mm in size. The raw data 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)*

# Tool Building



# Anatomy of a DH Project




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

# Text Corpus

- 17$^{th}$C English news pamphlets
- Dec 1653 to May 1654
- British Library electronic texts
- 312 documents, 800 000 words

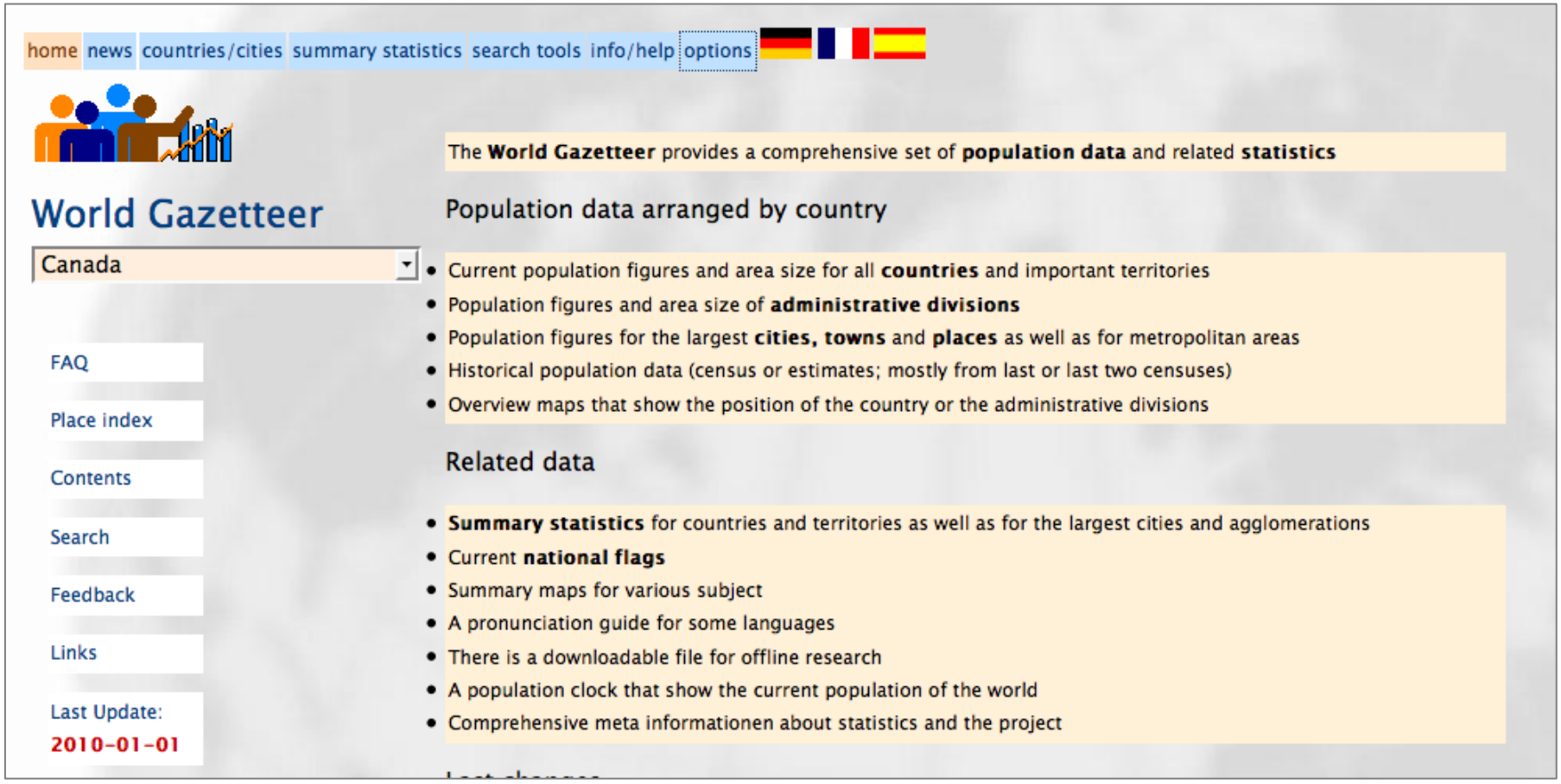
# CLAWS: Part of Speech Tagging

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

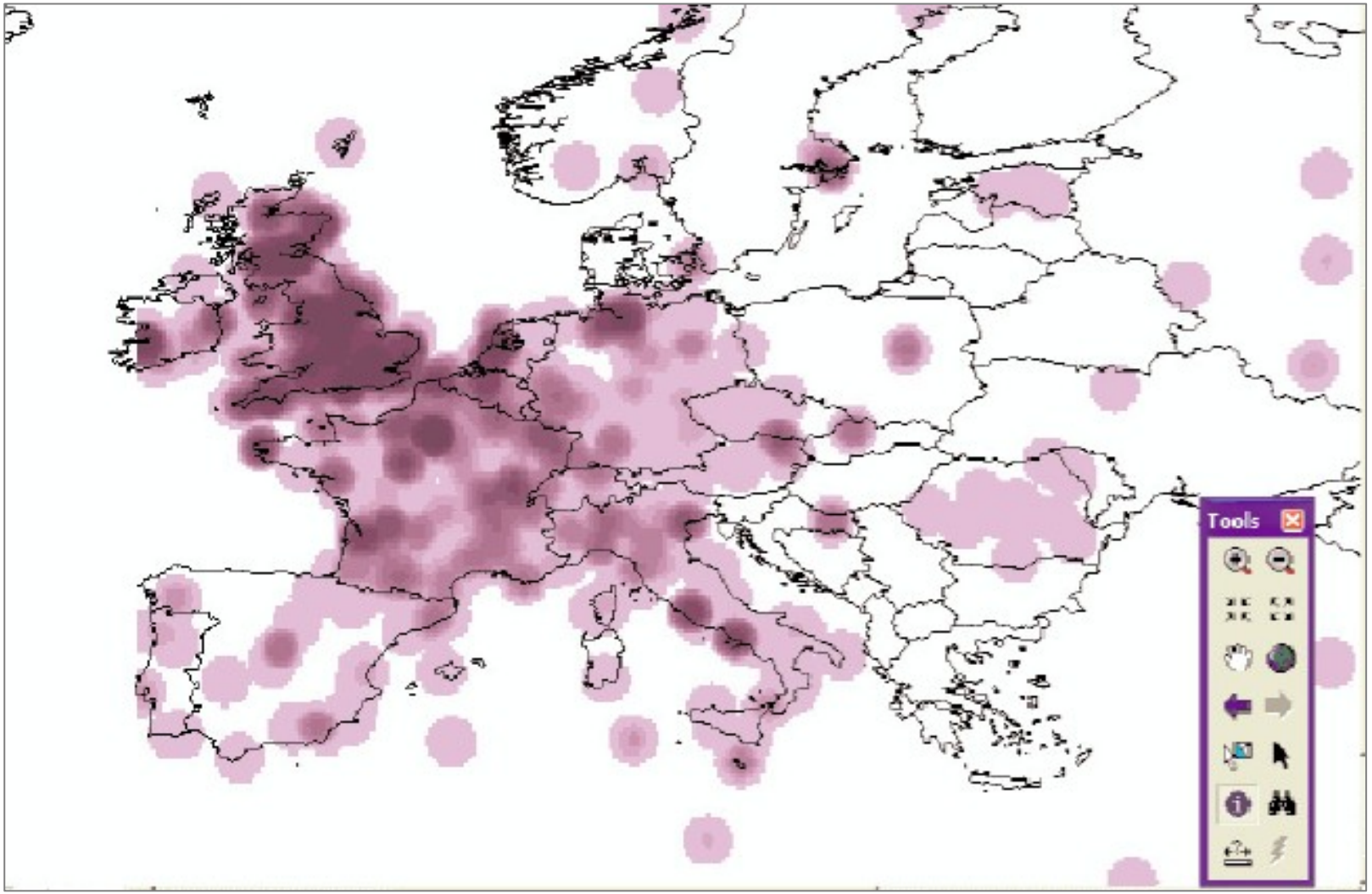
```
<p> In_II the_AT <em> Patrick_NP1 </em> of_IO <em>
<reg orig="Liverpoole"> Liverpool_NP1 </reg> </em> ,_,
which_DDQ we_PPIS2 lately_RR recovered_VVN from_II
the_AT <em> Brest_JJT </em> men_NN2 of_IO War_NN1 ,_,
was_VBDZ one_MC1 <em> Walter_NP1 Roche_NP1 </em> ,_,
```

- Grammar parsing tool (Linguistics)
- Identifies proper nouns, common nouns, plural nouns, adjectives, prepositions…
- 95–97% accuracy

# Geographical Analysis



# Geographical Analysis



# USAS: Semantic Parsing

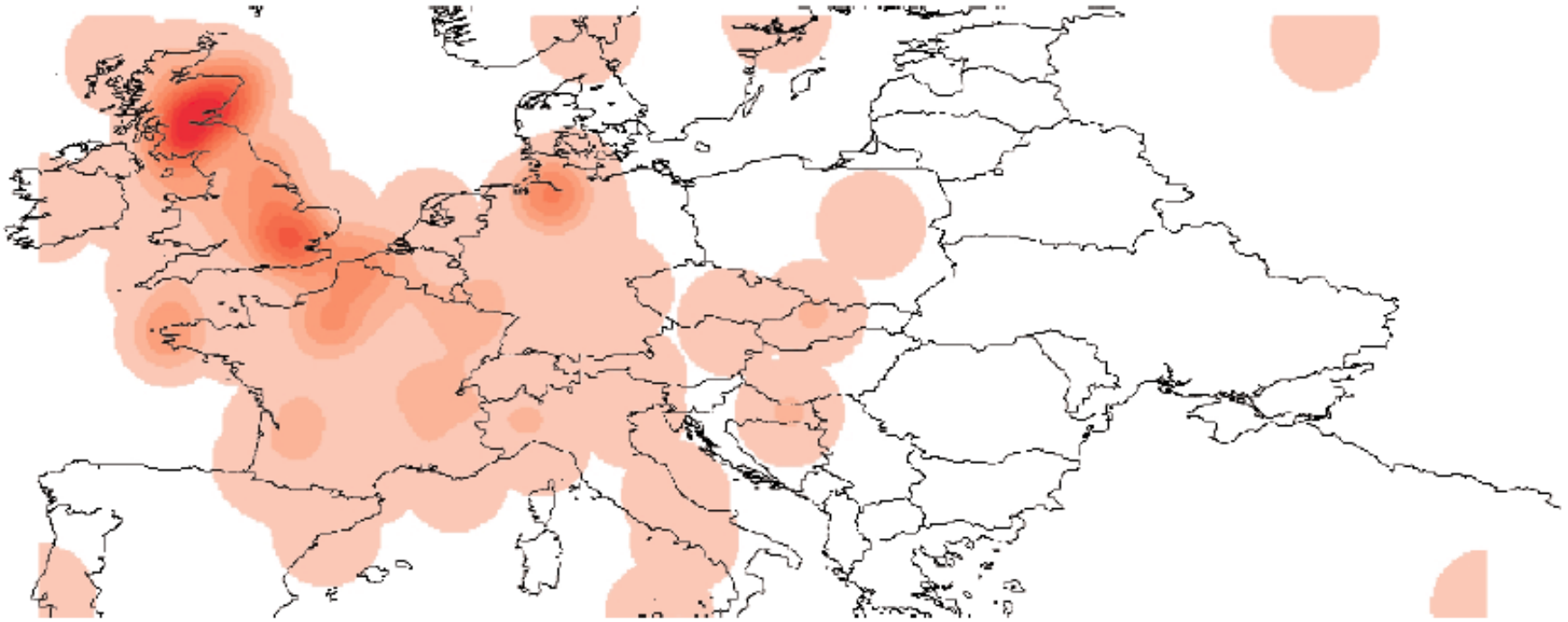
... two_N1 ships_M4 from_Z5 Dunkirk_Z2 have_Z5 brought_M2 Men_S2.2m Arms_B1 ,_PUNC and_Z5 Ammunition_G3 to_Z5 Middleton_Z1mf

Ships = M4 (shipping, swimming, etc.)

Men = S2.2m (People:-Male)

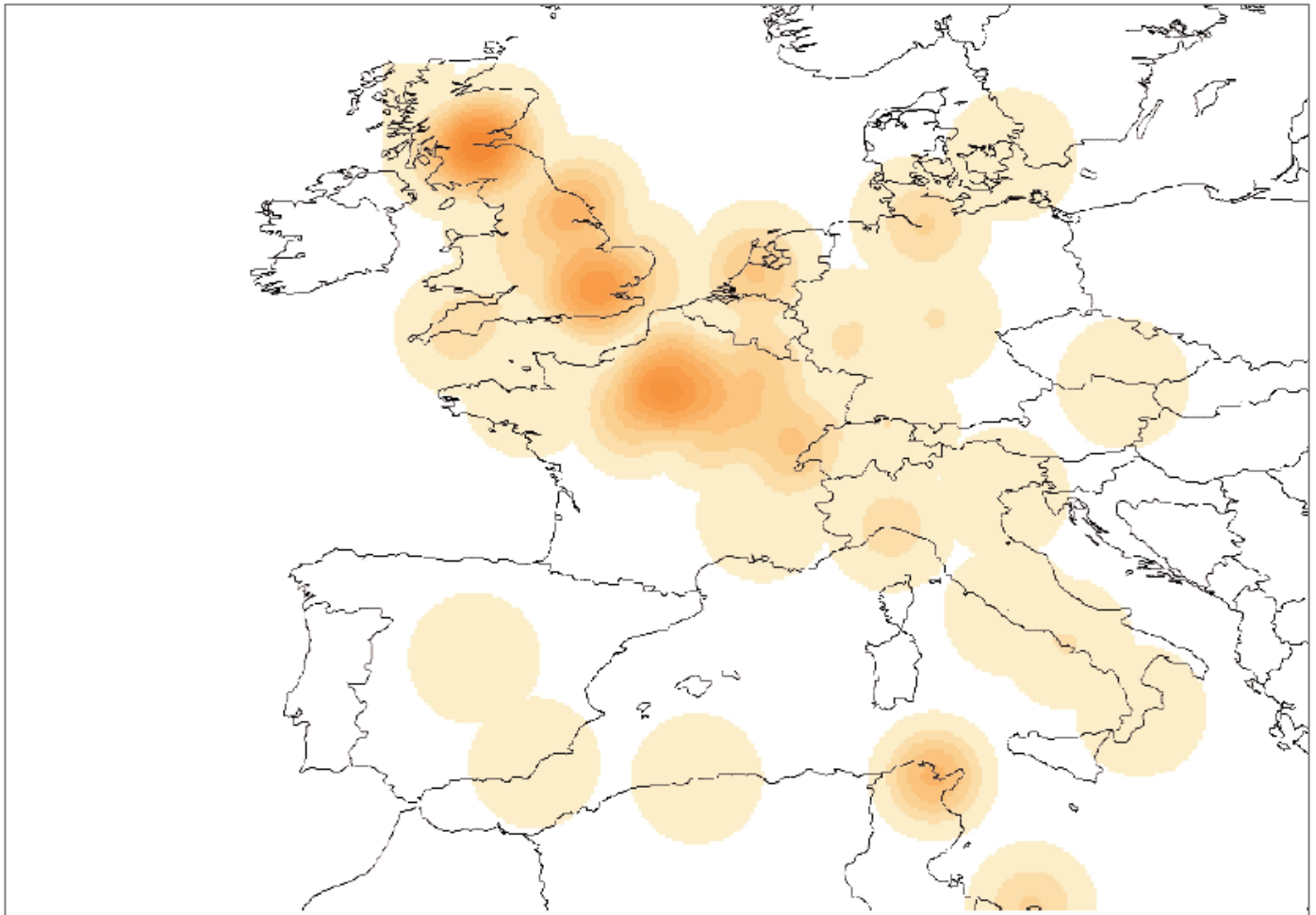
Ammunition = G3 (warfare, defence and the army; weapons)

# GIS Mapping: Topic ‘War’



*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’*

# DH Growing Pains

## ABOUT FOUND HISTORY

Found History explores public history and digital humanities in all their forms. It is produced by Tom Scheinfeldt, Managing Director at the Center for History and New Media at George Mason University.

## BEST OF FOUND HISTORY

- Omeka and Its Peers
- Lessons from One Week | One Tool – Part 3, Serendipity
- Lessons from One Week | One Tool – Part 2, Use
- Lessons from One Week | One Tool – Part 1, Project Management
- New Wine in Old Skins: Why the

← Briefly Noted for May 6, 2010 Briefly Noted for May 13, 2010 →

## 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?

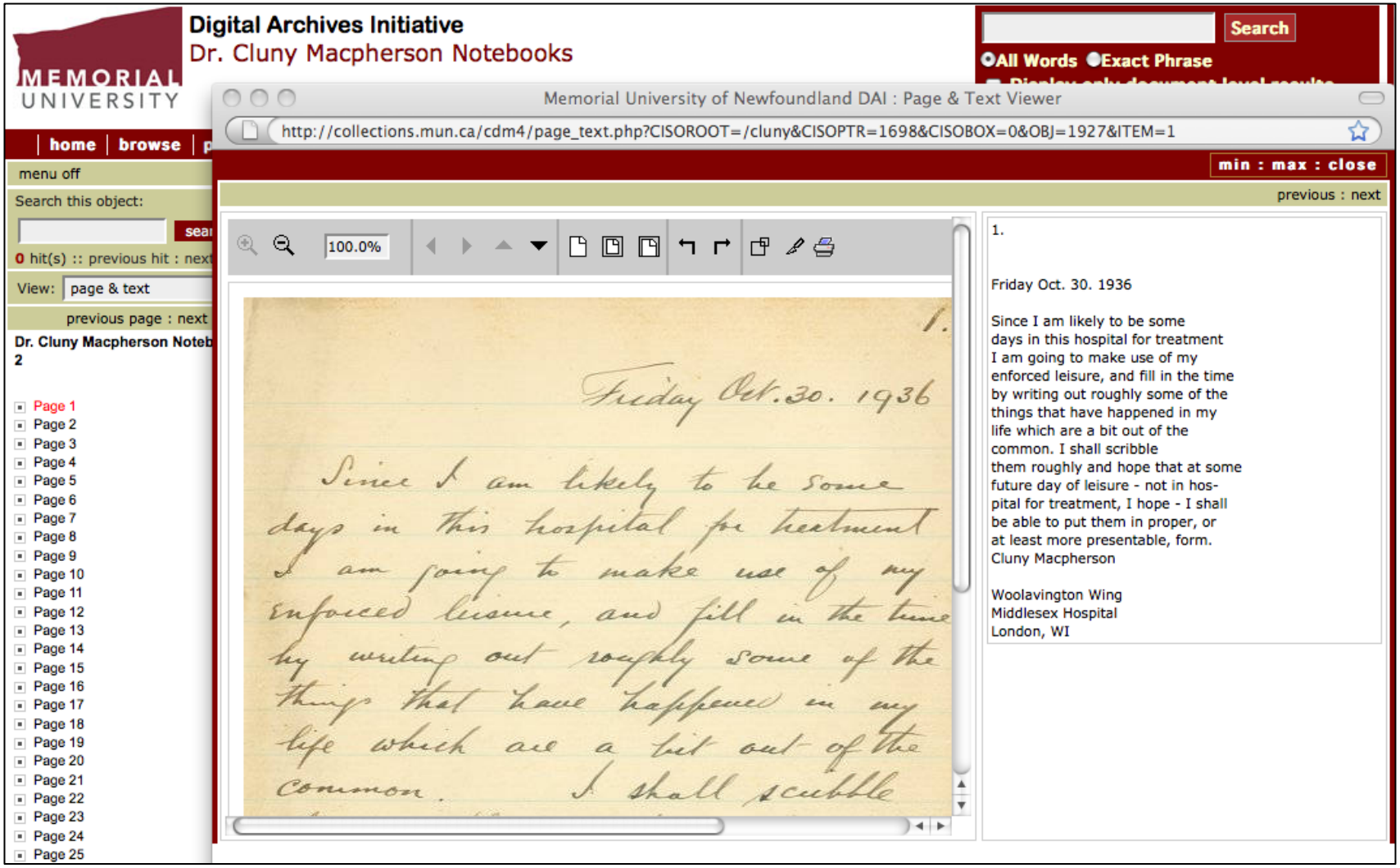
Concern over the apparent lack of argument in digital humanities comes not only from outside our young discipline. Many practicing digital humanists are concerned about it as well. Rob Nelson of the University of Richmond's Digital Scholarship Lab, an accomplished digital humanist, recently ruminated 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

# Primary source materials

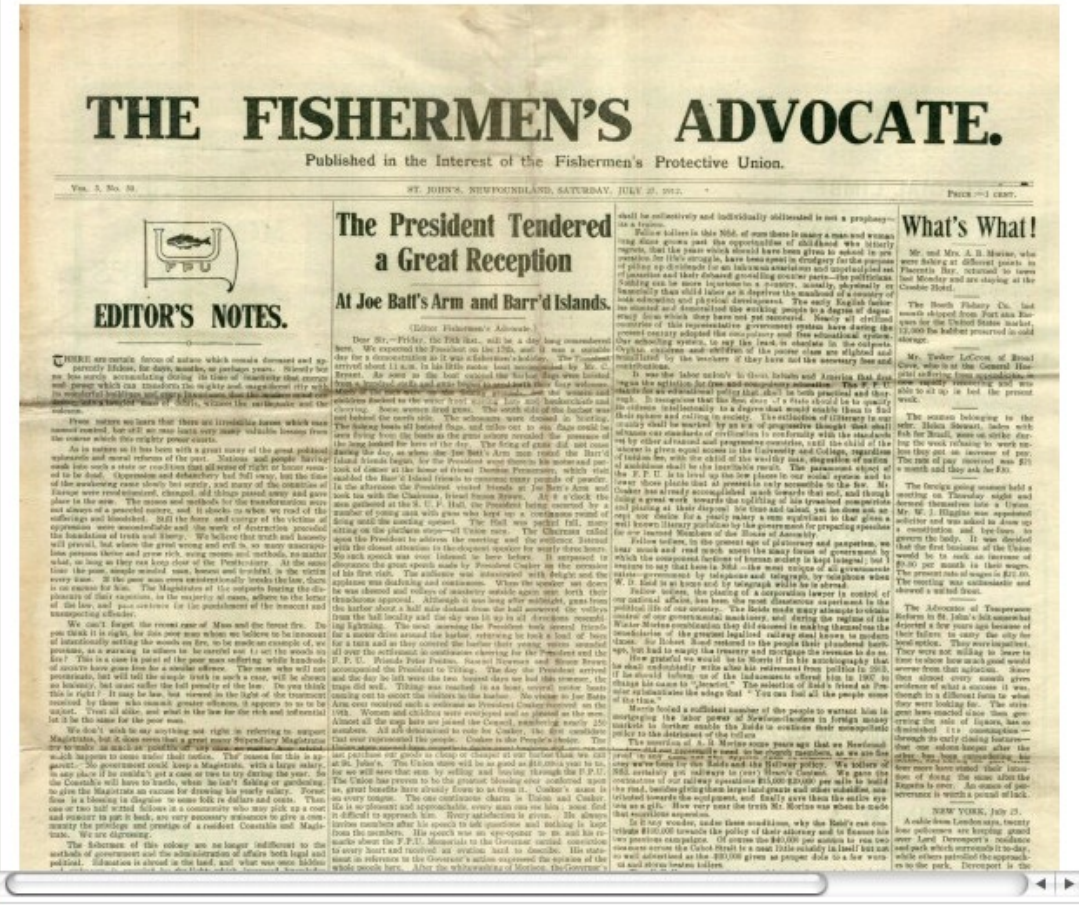


# Primary source materials

min : max : close

previous : next

13.3%



THE FISHERMEN'S ADVOCATE.

Published in the Interest of the Fishermen's Protective Union.

EDITOR'S NOTES.

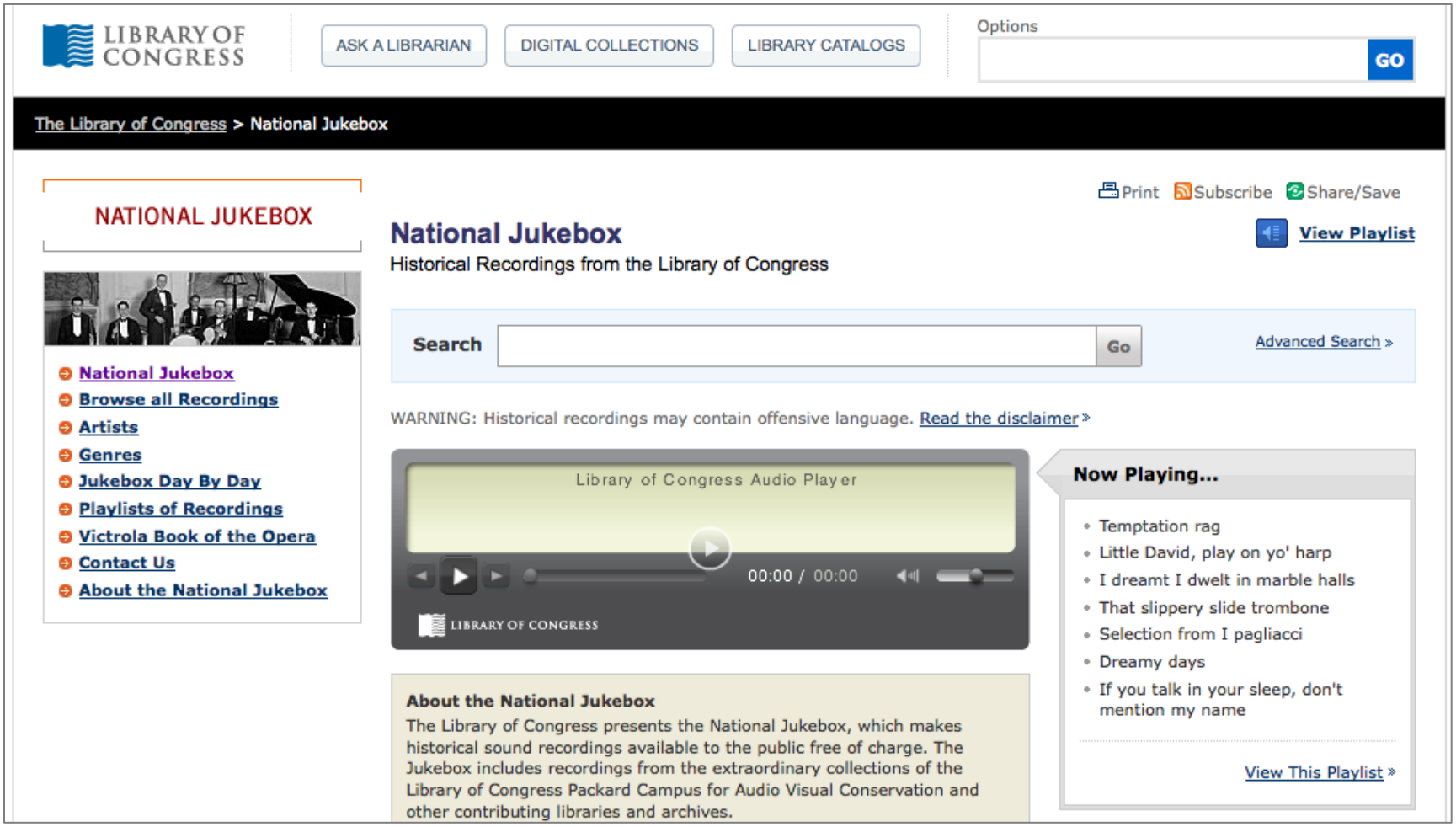
The President Tendered a Great Reception

At Joe Batt's Arm and Barr'd Islands.

What's What!

THE FISHERMEN'S ADVOCATE. Published in the Interest of the Fishermen's Protective Union. Voi,. 3, No. 30. ST. JOHN'S, NEWFOUNDLAND, SATURDAY, JULY 27. 1912. Price:"1 cent. EDITORS NOTES. -o- 7THERE are certain forces of nature which remain dormant and apparently lifeless, for days, months, or perhaps years. Silently but no less surely accumulating during its time of inactivity that energy and power which can transform the mighty and magnificent city witjt ' its wonderful buildings and every luxuriance that the-modern mind can . [illegible] devise, into a tangled mass of debris, witness the earthquake aud the volcano. From nature we learn that there are irresistible forces which man cannot control, but still we may learn very many valuable lessons from the course which this mighty power exerts. As in nature so it has been with a great many of the great political upheavels and moral reforms of the past. Nations and people having sunk into such a state or condition that all sense of right or honor seemed to be dead. Oppression and debauchery had full sway, but the time of the awakening came slowly but surely, and many of the countries of Europe were revolutionized, changed, old things passed away and gave place to the new. The means and methods for the transformation were not always of a peaceful nature, and it shocks us when we read of the sufferings and bloodshed. Still the force and energy of the victims pf oppression were uncontrollable and the work of destruction preceded the foundation of truth and liberty. We believe that truth and honesty will prevail, but where the great wrong and evil is, so many unscrupulous persons thrive and grow rich, using means and methods, no matter what, so long as they can keep clear of the Penitentiary. At the same time the poor, simple minded man, honest and truthful, is the victim every time. If the poor man even unintentionally breaks the law, there is no excuse for him. The Magistrates of the outports fearing the displeasure of their superiors, in the majority of cases, adhere to the letter of the law, and pa^s sentence for the punishment of the innocent and unsuspecting offender. We can't forget the recent case of Moss and the forest fire. Do you think it is right, for this poor man whom we believe to be innocent of intentionally setting the woods on fire, to be made an

# Primary source materials



# Primary source materials

Yale DAILY BULLETIN

GO Y

Yale News Yale University

NEWS, EVENTS AND CONVERSATIONS FOR THE YALE COMMUNITY

May 12, 2011

Yale responds to Japan CLICK HERE

Home
Calendar
Bulletin Board
Awards & Honors
Classifieds
Send Us Feedback

Sign up for e-news
Submit an event
Contact us

## Digital Images of Yale's Vast Cultural Collections Now Available for Free

posted: May 10, 2011 Print E-Mail Share

Enlarge +

George Stubbs, *Zebra*, 1763, oil on canvas, Yale Center for British Art, Paul Mellon Collection. From the collections of Yale University.

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 and freely available

Login to the Yale portal

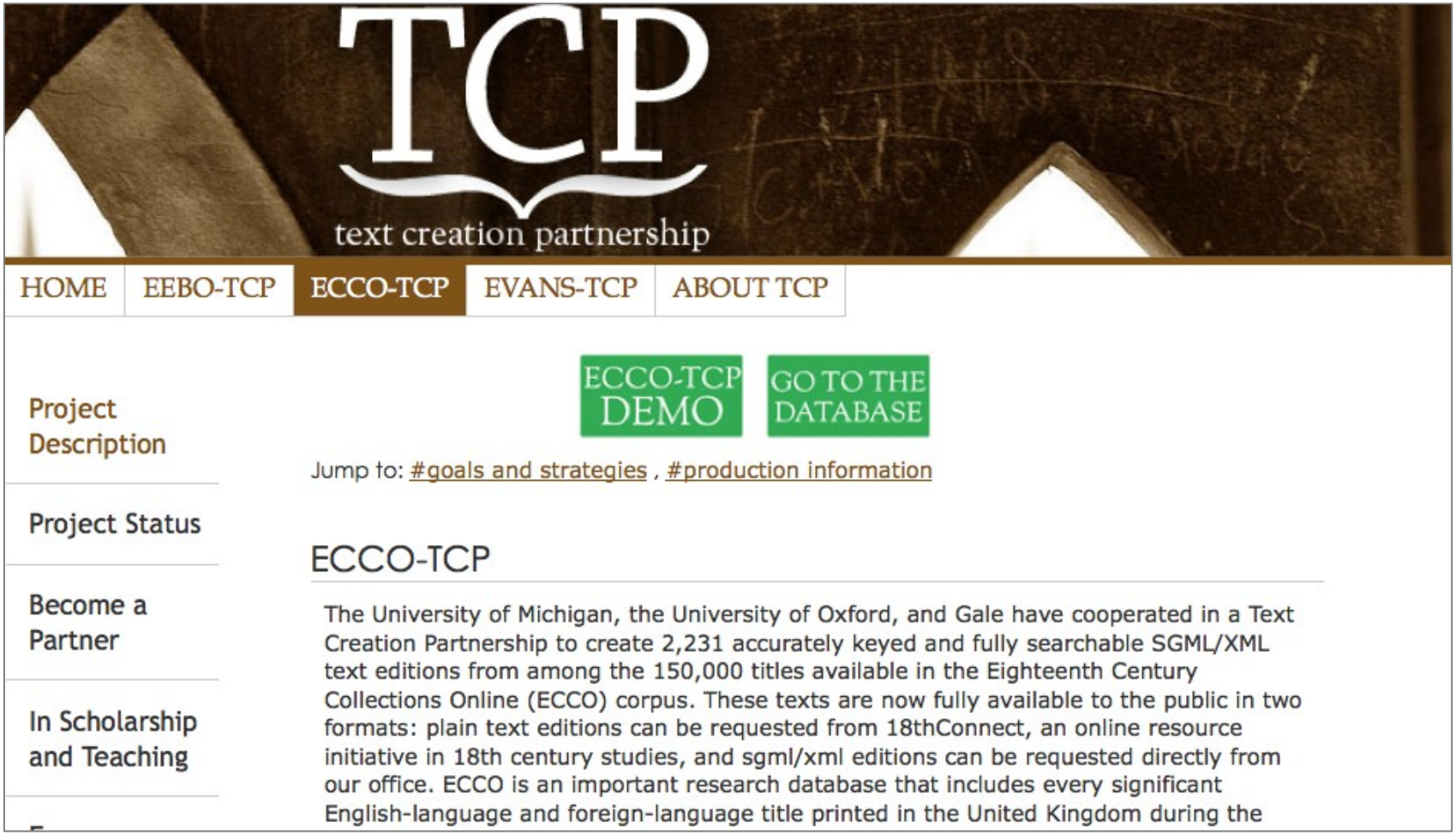
YOUR FEEDBACK

I'd like more of this type of story.
I'd like less of this.
Send us your suggestions:

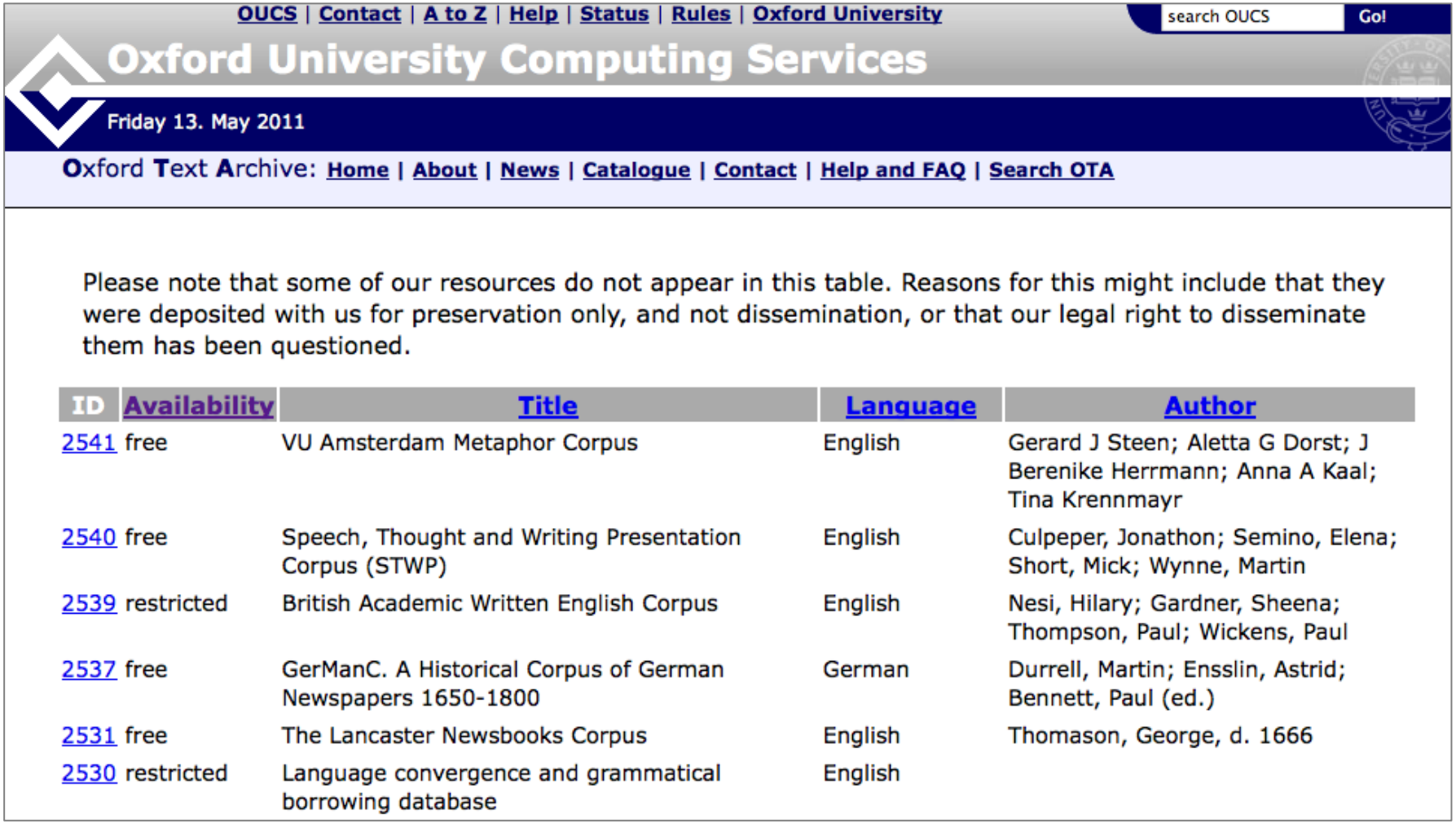
Send

# Large text corpora

# Large text corpora



# Large text corpora



OUCS | Contact | A to Z | Help | Status | Rules | Oxford University

search OUCS Go!

Oxford University Computing Services

Friday 13. May 2011

Oxford Text Archive: Home | About | News | Catalogue | Contact | Help and FAQ | Search OTA

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.

| ID | Availability | Title | Language | Author |
|---|---|---|---|---|
| 2541 | free | VU Amsterdam Metaphor Corpus | English | Gerard J Steen; Aletta G Dorst; J Berenike Herrmann; Anna A Kaal; Tina Krennmayr |
| 2540 | free | Speech, Thought and Writing Presentation Corpus (STWP) | English | Culpeper, Jonathon; Semino, Elena; Short, Mick; Wynne, Martin |
| 2539 | restricted | British Academic Written English Corpus | English | Nesi, Hilary; Gardner, Sheena; Thompson, Paul; Wickens, Paul |
| 2537 | free | GerManC. A Historical Corpus of German Newspapers 1650-1800 | German | Durrell, Martin; Ensslin, Astrid; Bennett, Paul (ed.) |
| 2531 | free | The Lancaster Newsbooks Corpus | English | Thomason, George, d. 1666 |
| 2530 | restricted | Language convergence and grammatical borrowing database | English | |

# Large text corpora

Home | LJ in Print | Newsletters | Libraries & Librarians | Reviews | Collection Development

Google™ Site Search | Any time | Search

## Text-Mining Ahead: HathiTrust Research Center to Open Corpus to Researchers

Share | | | | | RSS | Subscription | Reprint/License | Print | Email

By David Rapp Apr 28, 2011

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

**BRITISH NATIONAL CORPUS**

**About**

What is the BNC?
Creating the BNC
BNC Products
Copyright
Contact Us
Contents A-Z

**Using the BNC**

What can I do with the BNC?
Using BNC with Xaira
FAQ

**Obtaining**

**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]

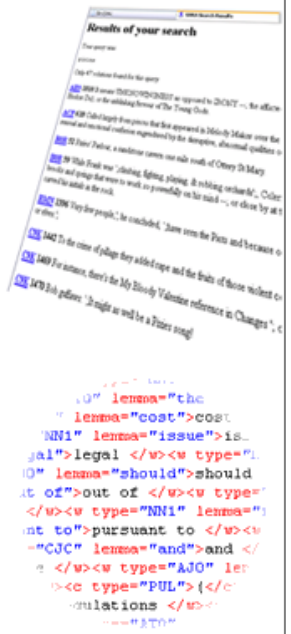
**Search the Corpus**

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

PPCME2

PPCEME

PPCMBE NEW

Corpus annotation

## Penn Parsed 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.

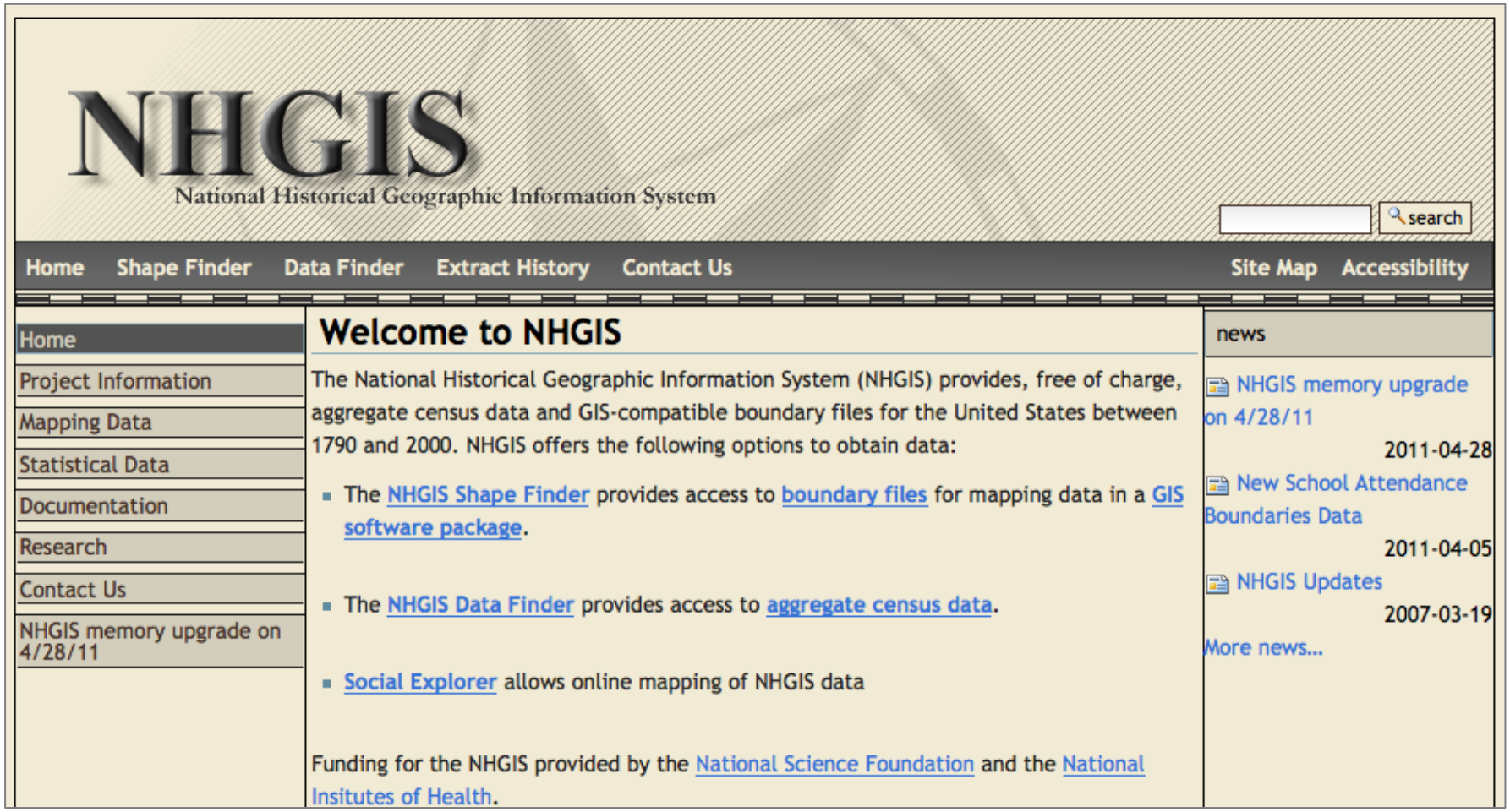
The Penn Parsed Corpora of Historical English are available on CD-ROM. Site licenses for departments/research groups and libraries are also available. See the corpus Order Form for charges. Corpus license fees go toward improving the corpora and increasing them in size. Upgrades, when completed, are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

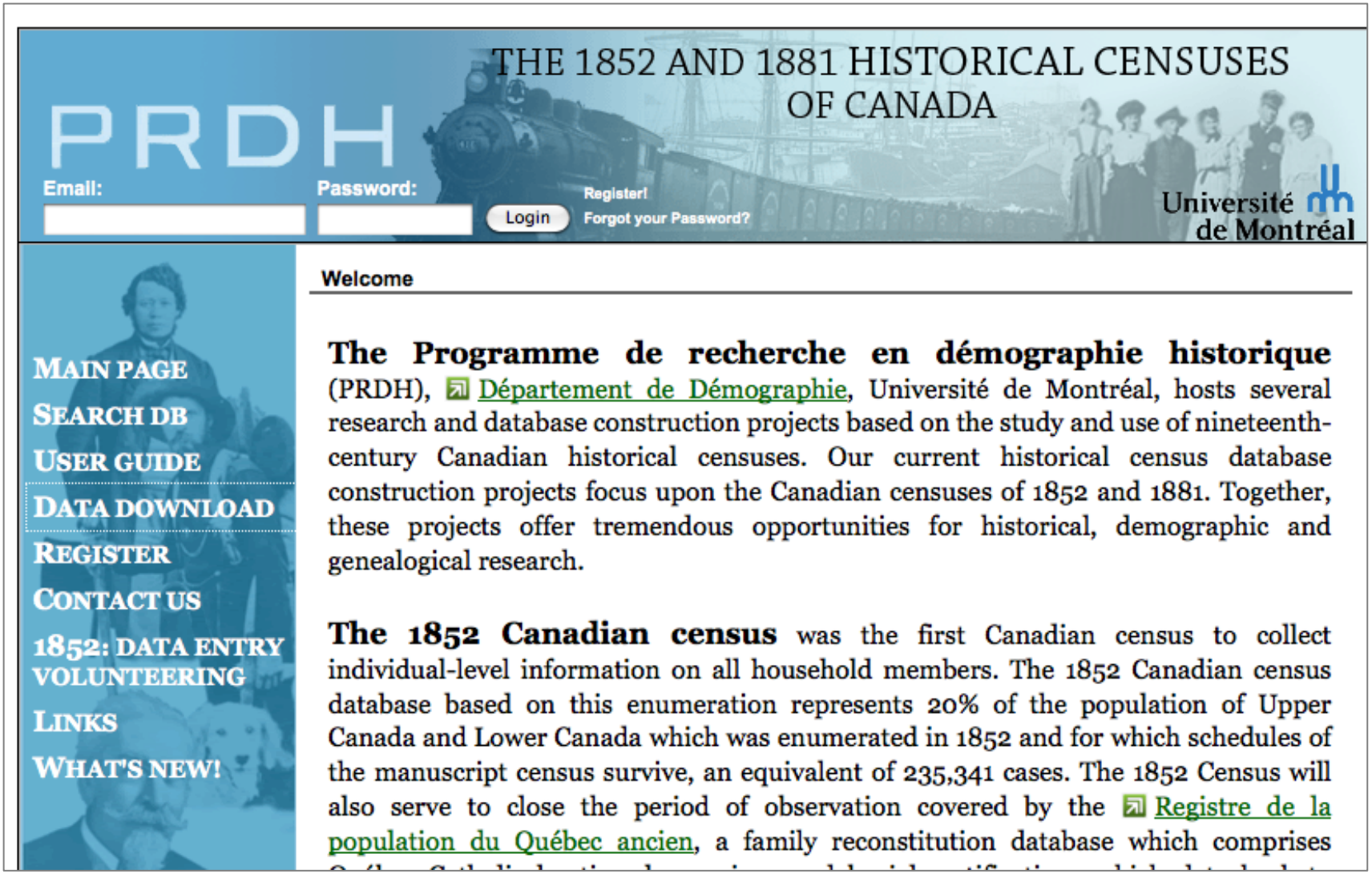
The Penn Historical Corpora are distributed with a search program CorpusSearch2, written by Beth Randall and released as open source software downloadable from its Sourceforge project web site.

- 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.

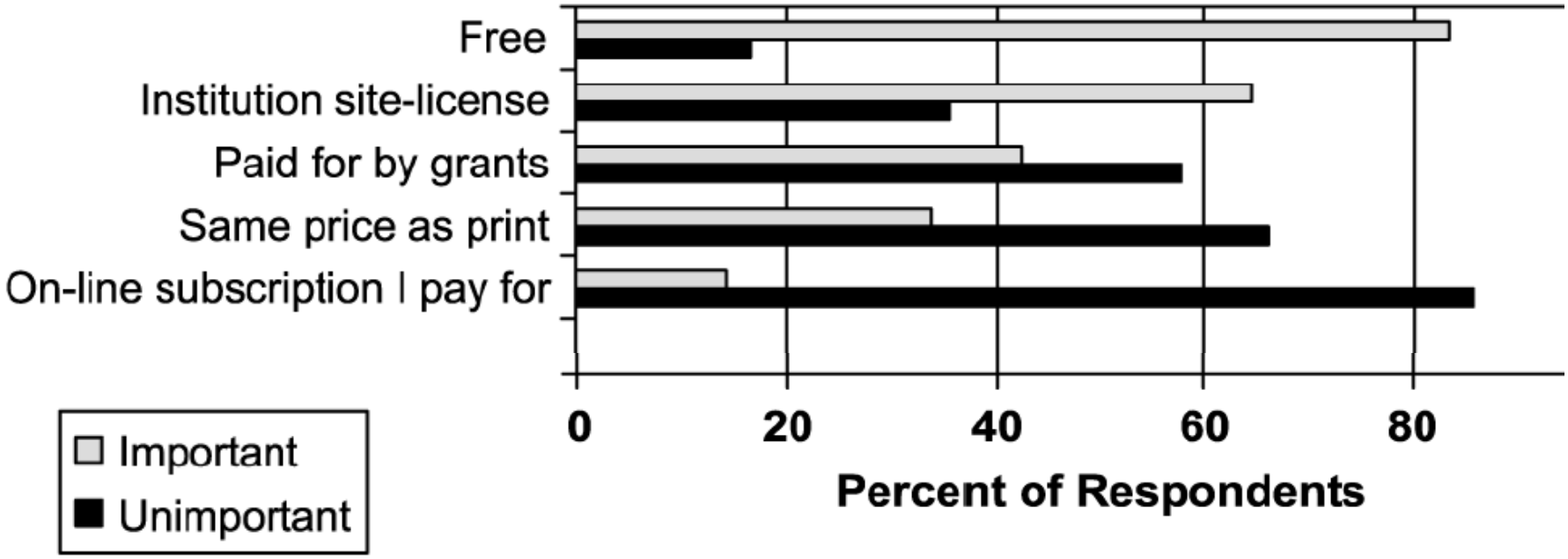
# Historical GIS Data



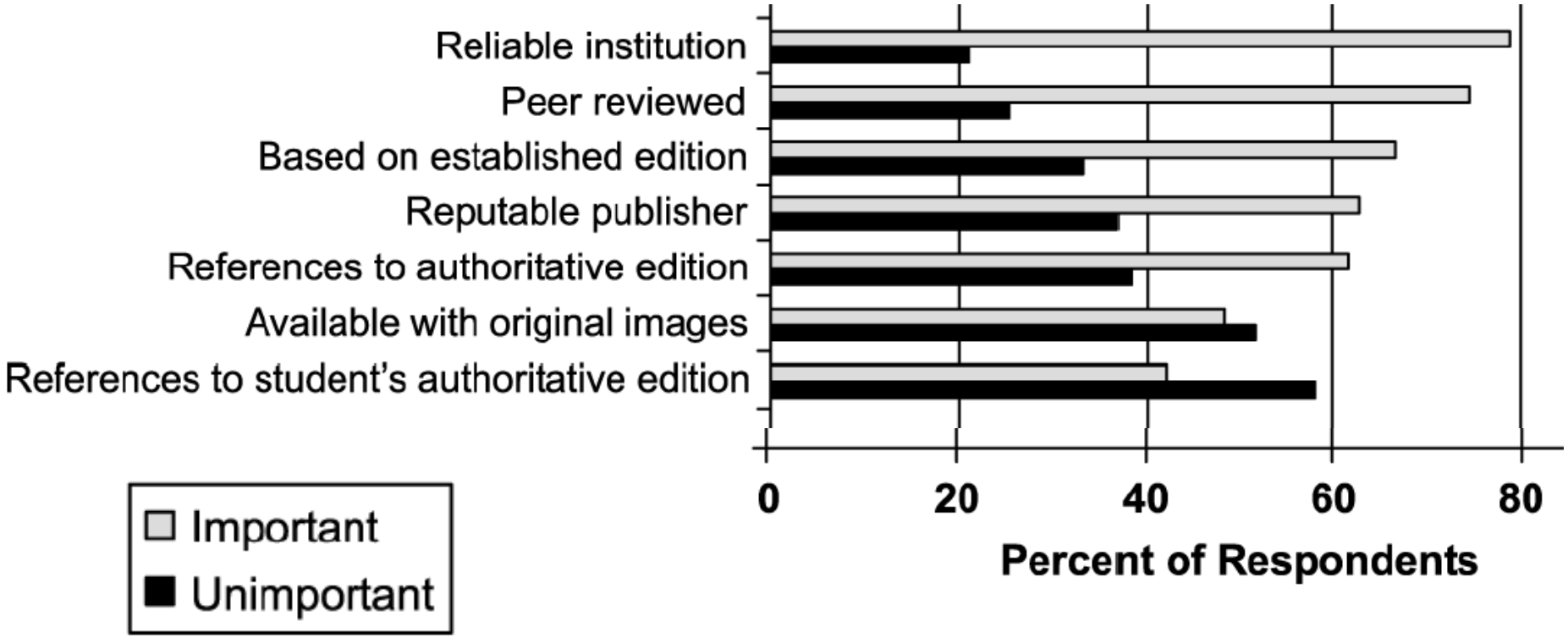
# Historical Census Datasets



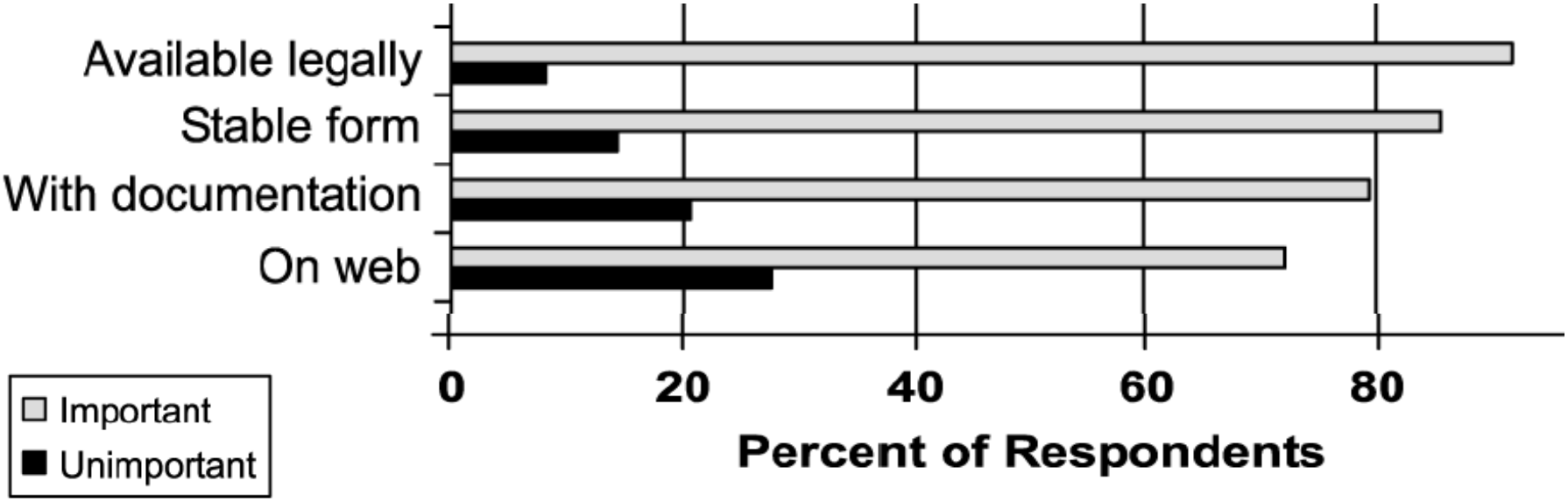
# Etext Preferences: Cost



# Etext Preferences: Quality



# Etext Preferences: Availability



# 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

# Library Role: Sources

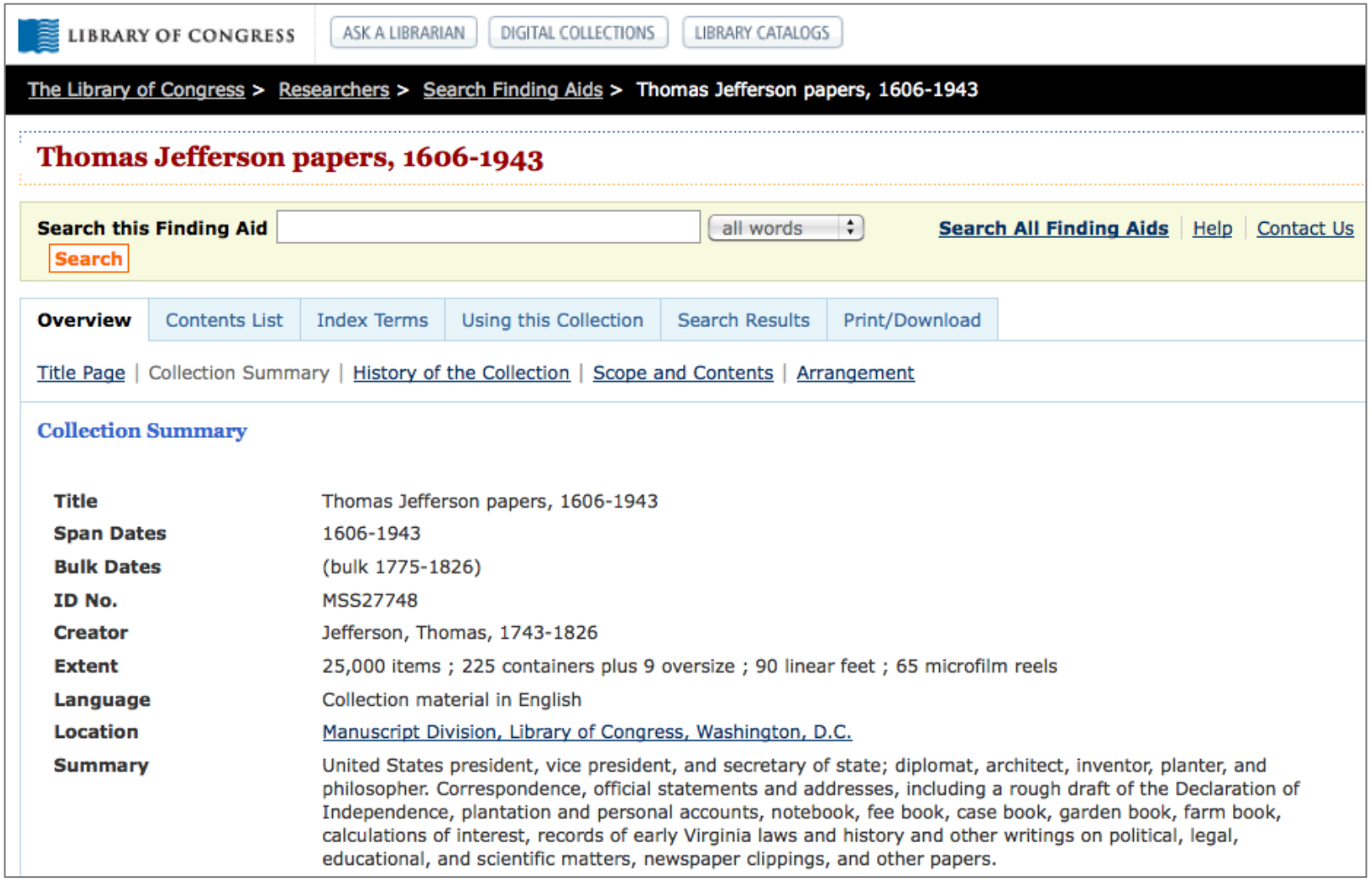
- Acquire & preserve primary sources
- Digitization & transcription
- Open, flexible formats
- Large text aggregations
- Data sets
- Licensing & copyright

# Digital Humanities Metadata

# Traditional Metadata

- Access to print collections
- Authorities
- Thesauri
- Bibliographies
- Indexes
- Concordances

# Finding Aids: Primary Sources



LIBRARY OF CONGRESS | ASK A LIBRARIAN | DIGITAL COLLECTIONS | LIBRARY CATALOGS

The Library of Congress > Researchers > Search Finding Aids > Thomas Jefferson papers, 1606-1943

## Thomas Jefferson papers, 1606-1943

Search this Finding Aid [ ] all words Search

Search All Finding Aids | Help | Contact Us

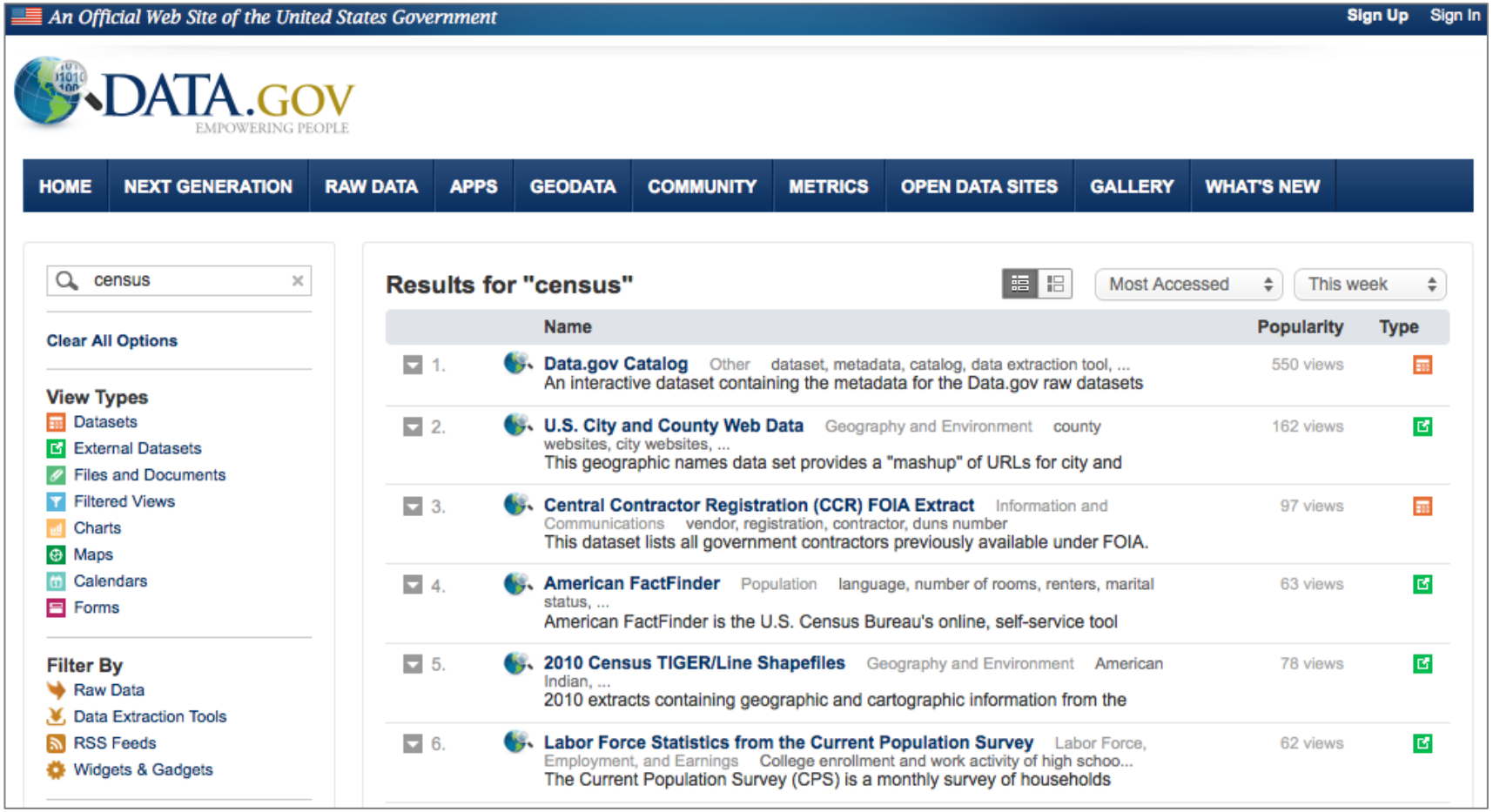
Overview | Contents List | Index Terms | Using this Collection | Search Results | Print/Download

Title Page | Collection Summary | History of the Collection | Scope and Contents | Arrangement

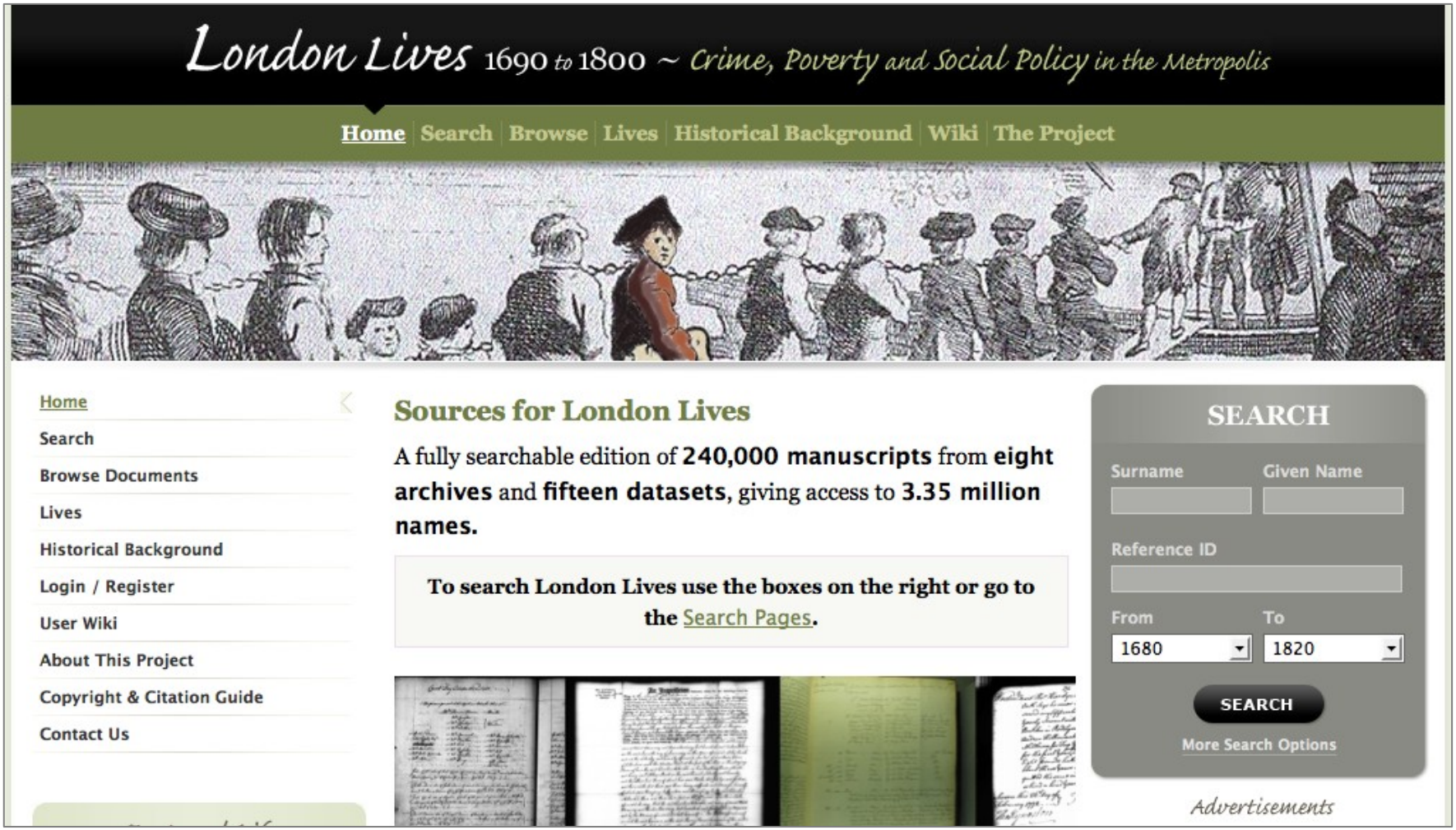
### Collection Summary

| | |
|---|---|
| **Title** | Thomas Jefferson papers, 1606-1943 |
| **Span Dates** | 1606-1943 |
| **Bulk Dates** | (bulk 1775-1826) |
| **ID No.** | MSS27748 |
| **Creator** | Jefferson, Thomas, 1743-1826 |
| **Extent** | 25,000 items ; 225 containers plus 9 oversize ; 90 linear feet ; 65 microfilm reels |
| **Language** | Collection material in English |
| **Location** | Manuscript Division, Library of Congress, Washington, D.C. |
| **Summary** | 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. |

# Finding Aids: Data Sets

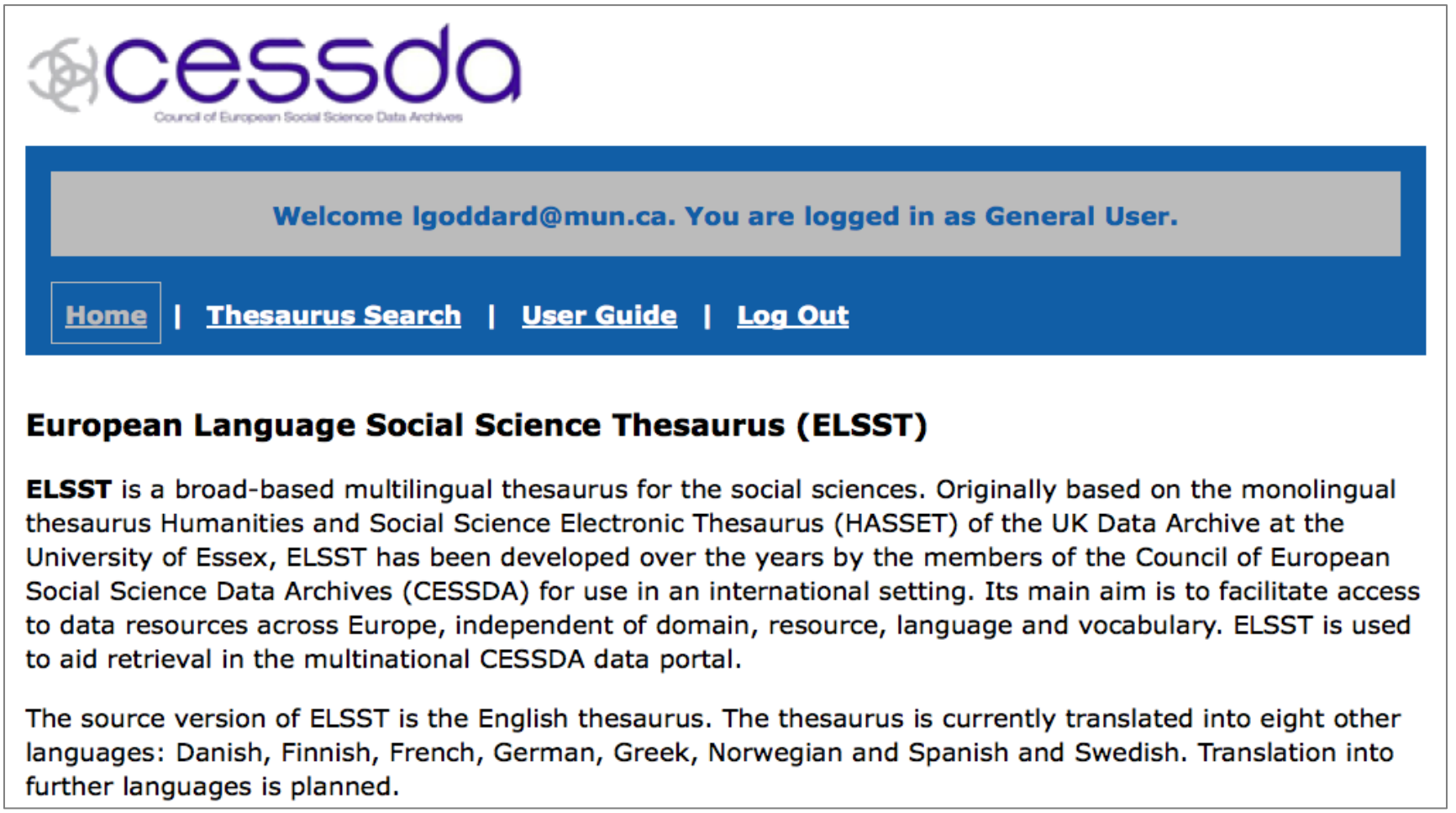


# Name Authorities

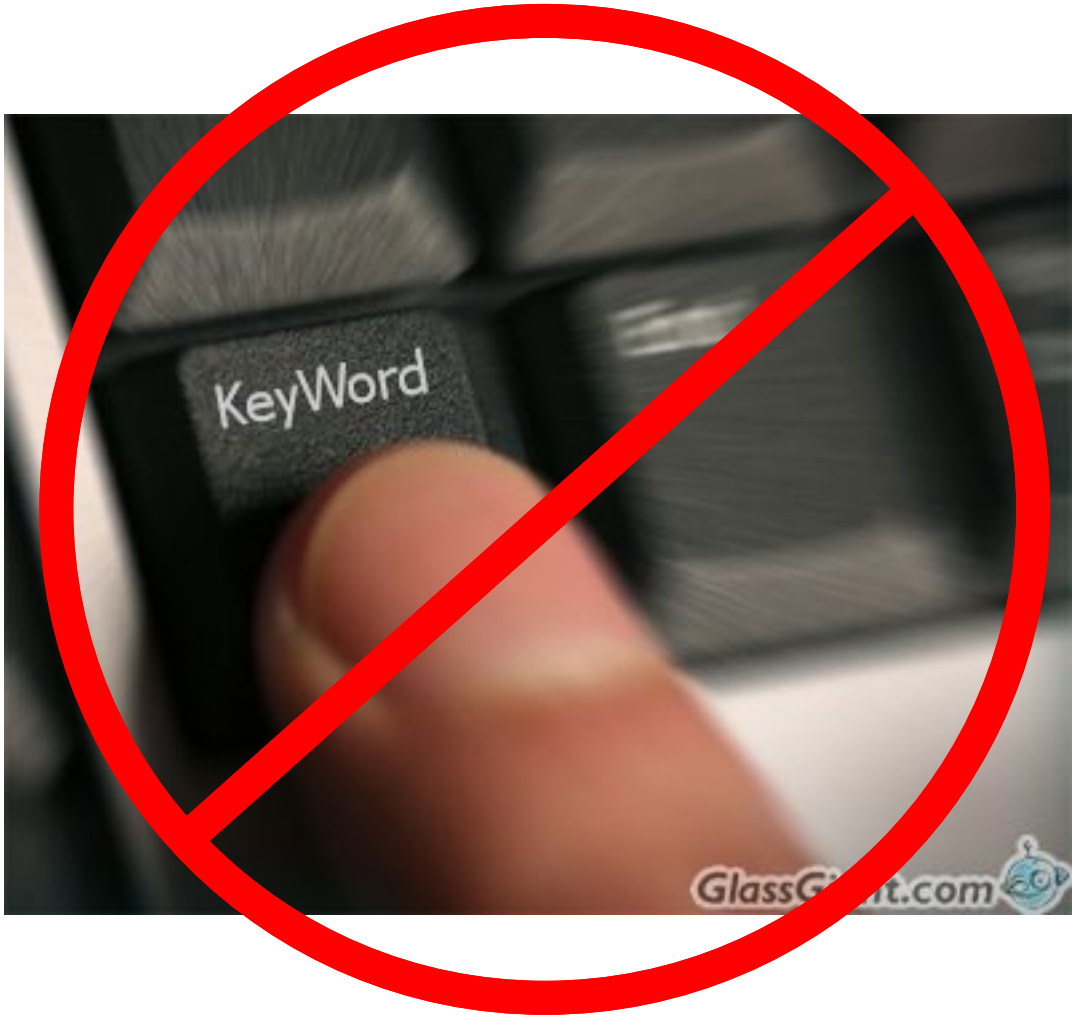


# Name Authorities

# Specialized Thesauri



# Full Text Markup
# Text Encoding Initiative (TEI)



# Metadata: 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>
```

# Metadata: TEI

```
<sp who="wag">
   <lb xml:id="l204"/>
    <p>For is he not
        <foreign xml:lang="la">
        Corpus naturale
        </foreign>?
    </p>
</sp>
```

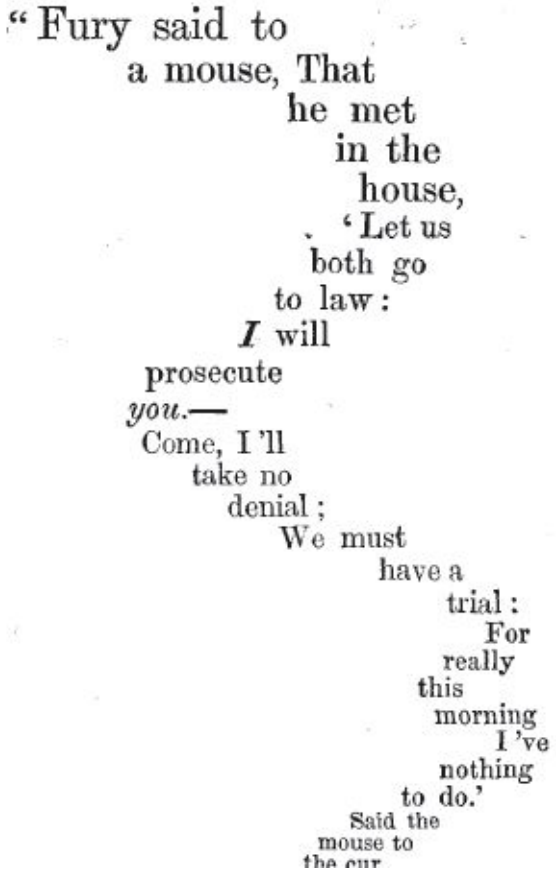
# Metadata: TEI

```
<l xml:id="l26">Nothing so sweet as
<choice>
<orig>Magicke</orig>
<reg>magic</reg>
</choice> is to him;</l>
```

# Metadata: TEI

```
<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>
```

# Metadata: TEI


"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

```
<l rend="font-size(100%) indent(5px)">both go</l>

<l rend="font-size(100%) indent(-7px)">to law:</l>

<l rend="font-size(100%) indent(-23px)"><hi
rend="italic">I</hi> will</l>

<l rend="font-size(100%) indent(-
26px)">prosecute</l>

<l rend="font-size(90%) indent(-40px)"><hi
rend="italic">you.</hi> —</l>
```

# Controlled Vocabularies

SUNDAY, 20. Battle
May the Lord grant that I may never
spend another day, li...
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 ammun-
ition was gone, with great credit, but much
loss to the 121st O.V.I. God alone spared me.

Chickamauga, Battle of, Ga., 1863

# Controlled Vocabularies

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>**

# 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