

# Digitized newspapers and digital research: What is the library's role?

**Juan Garcés & Julia Meyer**

Sächsische Landesbibliothek - Staats- und Universitätsbibliothek, Dresden

# Three points

1. Libraries are only slowly coming to terms with the potential of the (digital) use of digitized newspapers.
2. Newspapers in digital format don't want to be read on a screen only. They also want to be analysed with digital methods.
3. One of the library's core services will be to acquire the competency and provide research support in the area of digital research.

# Newspaper research in the digital era

- exaggerated reports of its death / difficulties in adapting to the new dispensation
- primary source for historical research - a „first draft of history“
- mass-retrodigitization
- interfaces are mainly catering to traditional usages and separation of concern

All	Books	Pictures, photos, objects	Journals, articles and data sets	<b>Digitised newspapers and more</b>	Government Gazettes	Music, sound and video	Maps	Diaries, letters, archives	Archived websites (1996 – now)	People and organisations	Lists
-----	-------	---------------------------	----------------------------------	--------------------------------------	---------------------	------------------------	------	----------------------------	--------------------------------	--------------------------	-------

## Advanced search - Digitised newspapers and more

Enter search terms in at least one of the fields below [?](#)

All of these words

The phrase

Any of these words

Without these words

**Search**

### Limit your results to (optional)


#### Search Scope

Limit keywords to only match in the field(s) selected

- Search anywhere in article (including tags and comments)
- Search anywhere in article (excluding tags and comments)
- Search headings only
- Search headings, author or 1st 4 lines (news articles only)
- Search captions only

#### Publication Date

Return only items published between dates (yyyy-mm-dd)


   
 From pub date - to pub date

#### Places and Titles

To search all newspapers, do not make a selection

Only show titles in the list below that... [\[SHOW FILTERS\]](#)

- National** - 3 title(s) [\[SHOW TITLES\]](#)
- ACT** - 8 title(s) [\[SHOW TITLES\]](#)
- New South Wales** - 438 title(s) [\[SHOW TITLES\]](#)
- Northern Territory** - 8 title(s) [\[SHOW TITLES\]](#)
- Queensland** - 97 title(s) [\[SHOW TITLES\]](#)

## Erweiterte Suche

---

VOLLTEXTSUCHE ERWEITERTE SUCHE ARTIKELSUCHE 

**Zeitungsunternehmen** 

Auswählen

**Zeitungstitel**

Auswählen

**Seitentext**

...

**Datum**

von **Jahr**

**Monat**

**Tag**

bis **Jahr**

**Monat**

**Tag**

**Erscheinungs- oder Verbreitungsort**

...

SUCHEN

ZURÜCKSETZEN

# Beyond discovery

## Allen & Sieczkiewicz (2011)

(Proceedings of the American Society for Information Science and Technology 47:1 - <http://onlinelibrary.wiley.com/doi/10.1002/meet.14504701131/pdf>)

„Historians will clearly benefit from improved search tools and interfaces which would support richer ways of interacting with collections. Two specific additional search services would appear to be particularly useful.“

1. searches restricted by a variety of attributes
2. visualization tools

# Distant reading



franco moretti

...the study of national bibliographies made me realize what a minimal fraction of the literary field we all work on: a canon of two hundred novels, for instance, sounds very large for nineteenth-century Britain (and is much larger than the current one), but is still less than one per cent of the novels that were actually published: twenty thousand, thirty, more, no one really knows— and close reading won't help here...



'A great iconoclast of literary criticism.'  
JOHN SUTHERLAND, *Guardian*

A network diagram on a dark blue background showing connections between literary figures. The nodes include: Horatio, Ophelia, Polonius, Laertes, Reynaldo, Priest, Norwegian Captain, Fortinbras, Ambassadors, Sailor, and Gentleman. Lines connect these nodes, representing relationships between them.

Franco Moretti

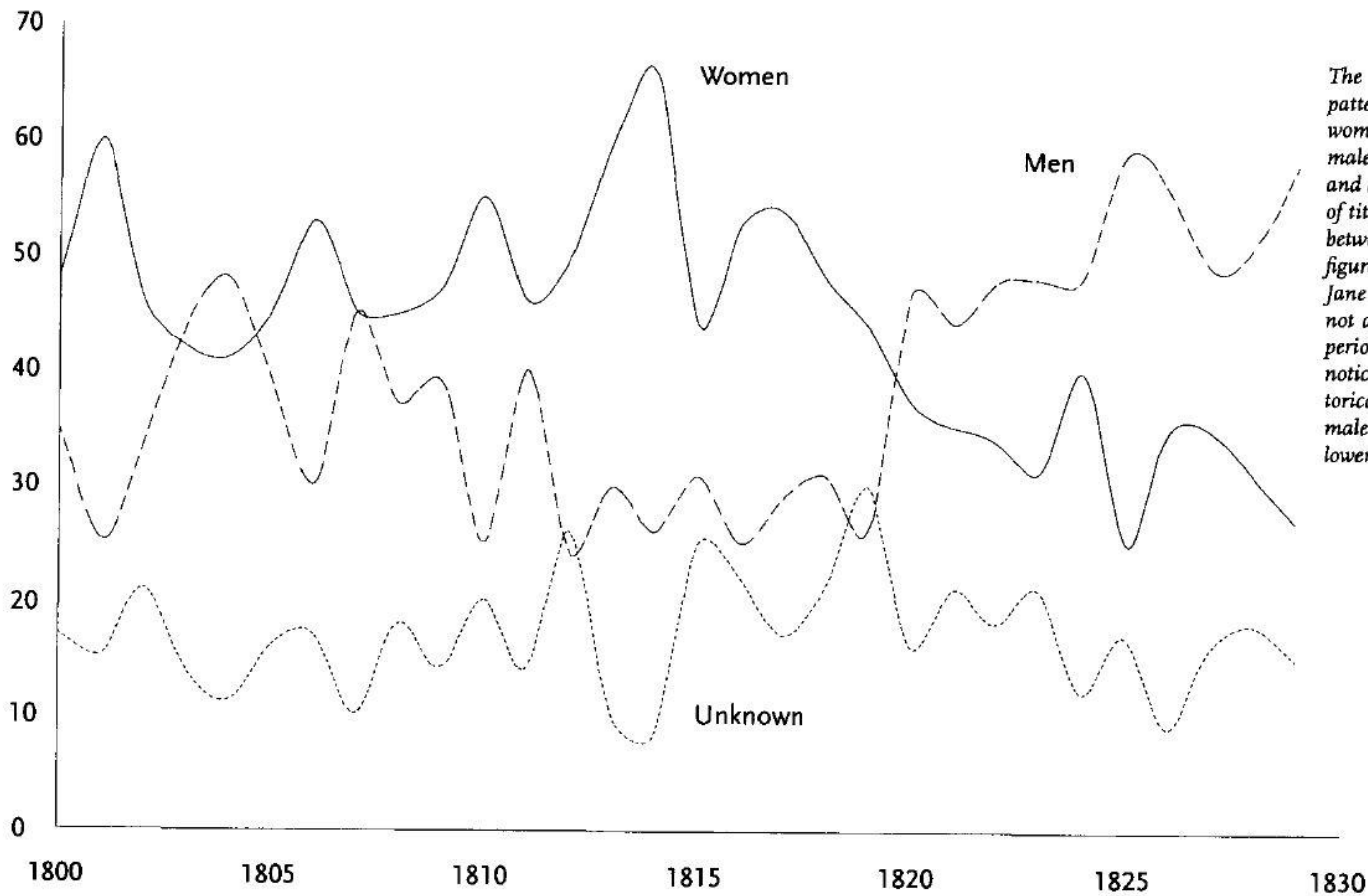
# Distant reading

- close reading
  - „the careful, sustained interpretation of a ... text”  
(Wikipedia)
- distant reading
  - abstraction
  - visualization (graph/map/tree)
  - interpretation



# Distant reading

FIGURE 12: *Authorship of new novels, Britain 1800–1829: gender breakdown (percentage)*



The 1810s show an even clearer pattern of female dominance, with women novelists out-producing their male counterparts in every year, and accounting for over 50 per cent of titles in six out of the eight years between 1810 and 1817 . . . As these figures indicate, the publication of Jane Austen's novels was achieved not against the grain but during a period of female ascendancy. It is noticeable that Scott's earliest historical novels were launched when male authorship of fiction was at a lower than usual ebb.

Peter Garside, 'The English Novel in the Romantic Era'

Source: Garside, Raven and Schöwerling, eds, *The English Novel 1770–1829*.

# Google Ngram Viewer

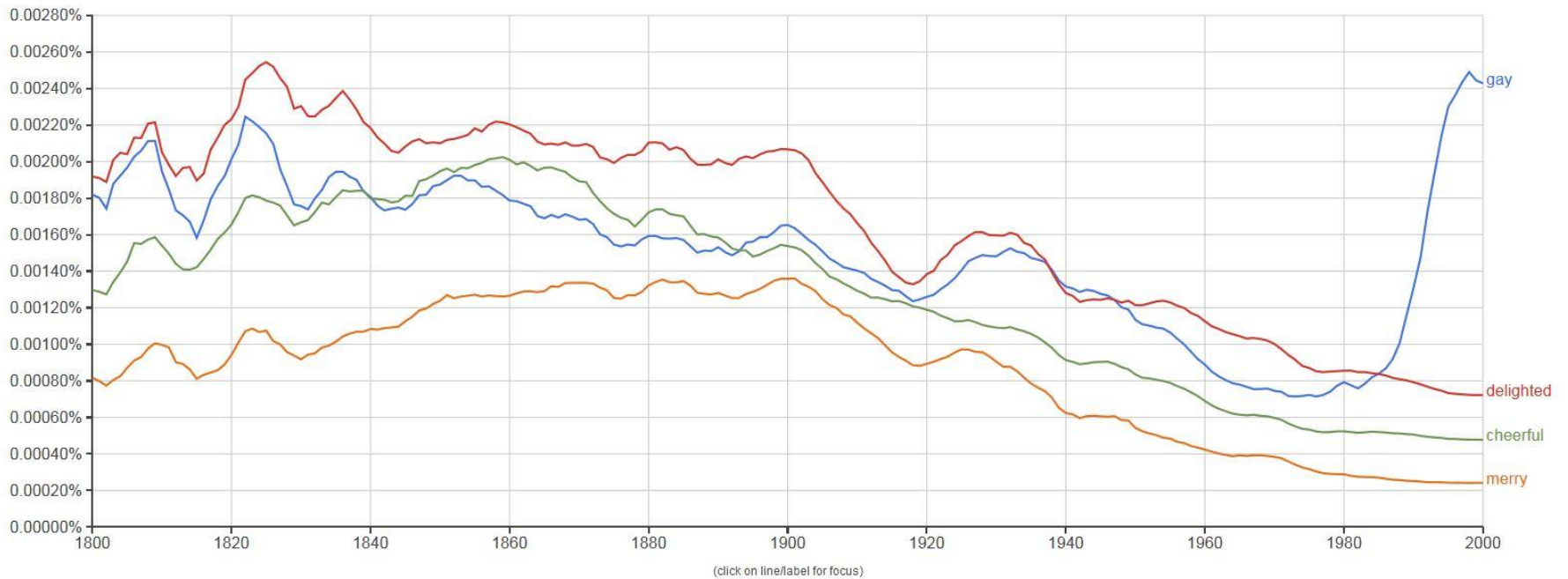
Google Books Ngram Viewer

Graph these comma-separated phrases:   case-insensitive

between  and  from the corpus  with smoothing of

[Tweet](#)

[Embed Chart](#)



# Text mining

“...the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning.” (Wikipedia)

Cheney (2013) (<http://library.ifla.org/id/eprint/233>)

challenges:

- access - licensing/copyright
- technology and software support
- storage and access issues

and

- **reference, instruction, and training needs**

CHRONICLING AMERICA  
Historic American Newspapers

Search America's historic newspaper pages from 1789-1924 or use the U.S. Newspaper Directory to find information about American newspapers published between 1690-present. Chronicling America is sponsored jointly by the [National Endowment for the Humanities](#) and the Library of Congress. [Learn more](#) »

Search Pages

Advanced Search

All Digitized Newspapers 1789-1924

US Newspaper Directory, 1690-Present

close X

## Select State(s):

- All states
- Alabama
- Alaska
- Arizona
- Arkansas
- California
- Colorado
- Connecticut

## Or Select Newspaper(s):

- All newspapers
- The Abbeville banner. (Abbeville, S.C.)
- The Abbeville bulletin. (Abbeville, S.C.)
- The Abbeville messenger. (Abbeville, S.C.)
- The Abbeville press and banner. (Abbeville, S.C.)
- Abbeville press. (Abbeville, S.C.)
- Abbeville progress. (Abbeville, Vermilion Parish, La.)
- The Aberdeen Democrat. (Aberdeen, South Dakota)

## Select Year(s)\*

Newspaper pages are available for newspapers published between 1789-1924\*

from  to

## Or Date Range

from  to

Limit Search:  only front page or  Specific page

Language

## Enter Search

...with **any** of the words:

...with **all** of the words:

...with the **phrase**:

...with the words:

within  words of each other

Clear

Search



## CHRONICLING AMERICA

### Historic American Newspapers

Search America's historic newspaper pages from 1789-1924 or use the U.S. Newspaper Directory to find information about American newspapers published between 1690-present. Chronicing America is sponsored jointly by the [National Endowment for the Humanities](#) and the Library of Congress. [Learn more](#) »

Pages Available: 12,226,675

Print Subscribe Share/Save Give Feedback



## About the Site and API

- [About Chronicing America](#)
- [About the Site and API](#)
- [Recommended Topics](#)
- [Help](#)

### More Resources

- [National Digital Newspaper Program](#)
- [NDNP Award Recipients](#)
- [Newspaper and Current Periodicals Reading Room](#)
- [Ask LC Newspaper & Current Periodicals Librarian](#)
- [Historic Newspapers on Flickr](#) (part of the LC Flickr Commons photostream)

### Introduction

Chronicing America provides access to information about historic newspapers and select digitized newspaper pages. To encourage a wide range of potential uses, we designed several different views of the data we provide, all of which are publicly visible. Each uses common Web protocols, and access is not restricted in any way. You do not need to apply for a special key to use them. Together they make up an extensive application programming interface (API) which you can use to explore all of our data in many ways.

Details about these interfaces are below. In case you want to dive right in, though, we use HTML link conventions to advertise the availability of these views. If you are a software developer or researcher or anyone else who might be interested in programmatic access to the data in Chronicing America, we encourage you to look around the site, "view source" often, and follow where the different links take you to get started. When describing Chronicing America as the source of content, please use the URL and a Web site citation, such as "from the Library of Congress, Chronicing America: Historic American Newspapers site".

For more information about the open source Chronicing America software please see the [LibraryOfCongress/chronam](#) GitHub site. Also, please consider subscribing to the [ChronAm-Users](#) discussion list if you want to discuss how to use or extend the software or data from its APIs.

### The API

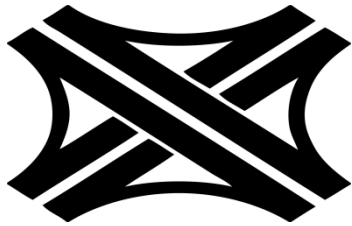
#### Jump to:

- [Search](#) the newspaper directory and digitized page contents using OpenSearch.
- [Auto Suggest](#) API for looking up newspaper titles
- [Link](#) using our stable URL pattern for Chronicing America resources.
- [JSON](#) views of Chronicing America resources.
- [Linked Data](#) views of Chronicing American resources.
- [Bulk Data](#) for research and external services.
- [CORS and JSONP](#) support for your JavaScript applications.

# Teaching Digital Humanities

- established history curricula (outside of DH) tend not to feature digital analysis methods
- there is a high demand for teaching and consultancy in digital methodological competency
- this aspect of DH fits in well within the expanding portfolio of libraries

# Areas



Technical **infrastructure**



Analytical **tools**



Methodological competency

# Remember

1. Libraries are only slowly coming to terms with the potential of the (digital) use of digitized newspapers.
2. Newspapers in digital format don't want to be read on a screen only. They also want to be analysed with digital methods.
3. One of the library's core services will be to acquire the competency and provide research support in the area of digital research.