



# Empowering Humanities Scholars to Better Explore Our Cultural Past

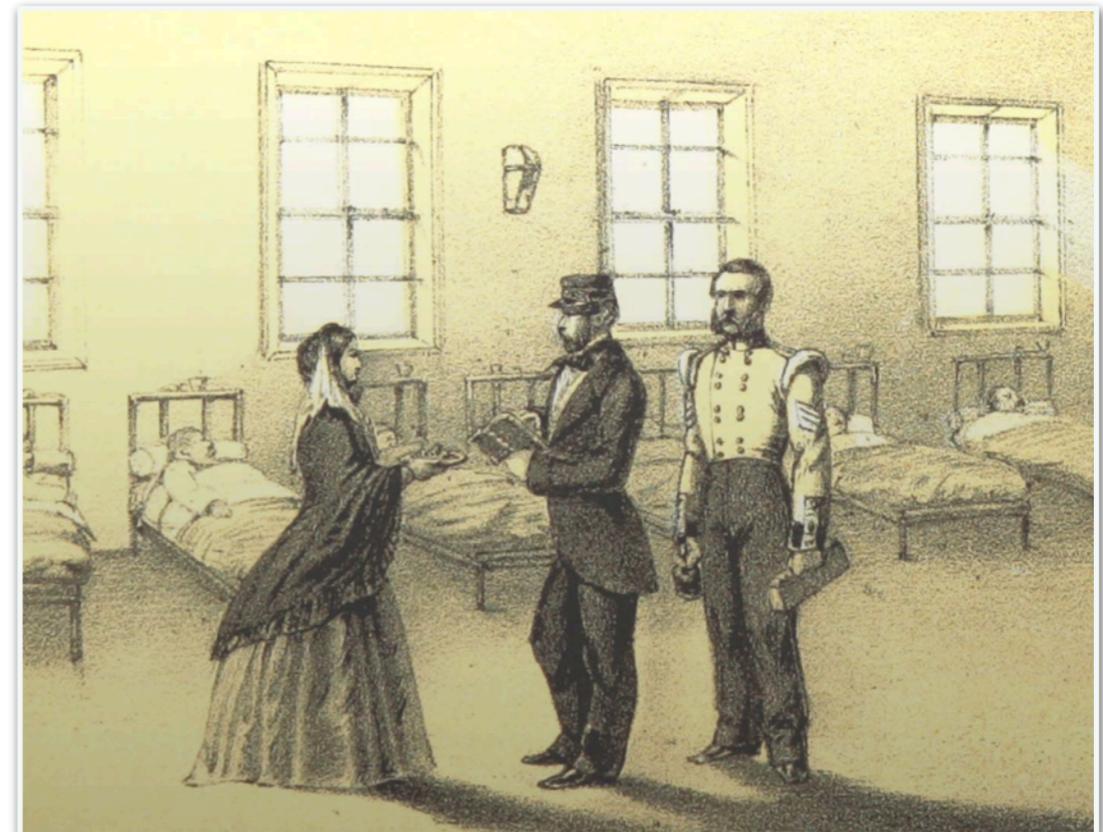
Derek Greene

School of Computer Science  
University College Dublin



# IRC Contagion Project

- *Contagion, Biopolitics and Cultural Memory*: Inter-disciplinary project between UCD School of English and SFI Insight Centre.
- Funded under Irish Research Council *New Horizons* scheme.
- Focuses on using large scale historical datasets to understand contemporary attitudes and improve public health outcomes.



<http://contagion.ie>

# Digital Humanities

“Technology and computing are used to reexamine traditional humanities research and explore new modes of scholarship.”

- *Timothy Lepczyk*

“Computers allow for doing things with texts and other cultural artefacts that could not be done feasibly without the computational power and storage modern computers provide. Computers should be considered an extension of the scholar's mind - very useful tools indeed.”

- *Philip R. Burns*

“Incorporating things like text mining, data visualisation, community building and curating online exhibitions and collections, Digital Humanities is the emerging field concerned with how studies of History, Literature, Music, Philosophy, Sociology, Performing Arts and other traditional humanities disciplines are utilising new digital tools.”

- *David Nelligan*

<https://whatisdigitalhumanities.com>

# The British Library

- The largest national library in the world by number of items catalogued, containing over 150 million items dating back to 2000BC.
- Total of over 800km of shelving, between the main library and a separate storage facility.



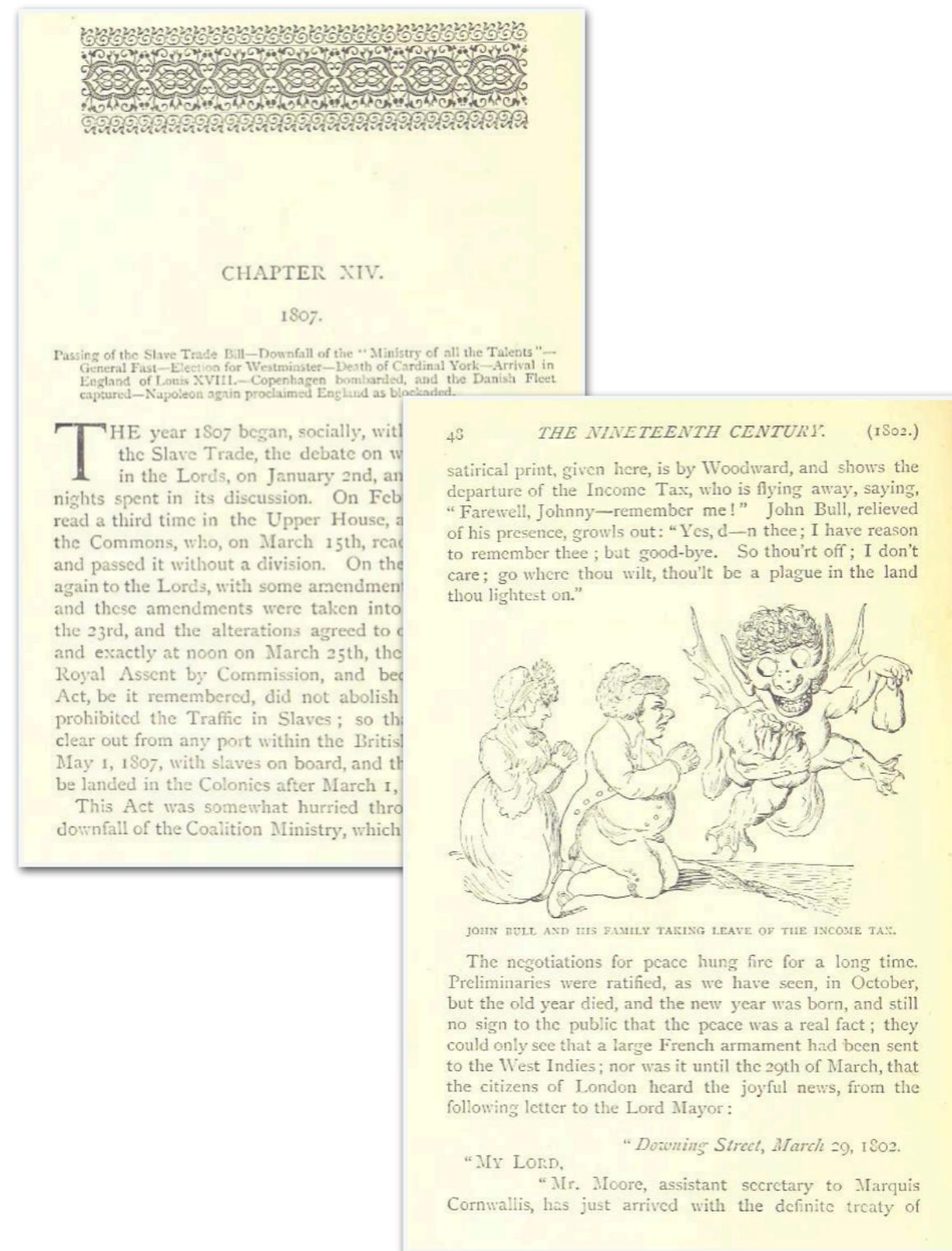
The British Library



The King's Library Tower

# British Library Digital Corpus

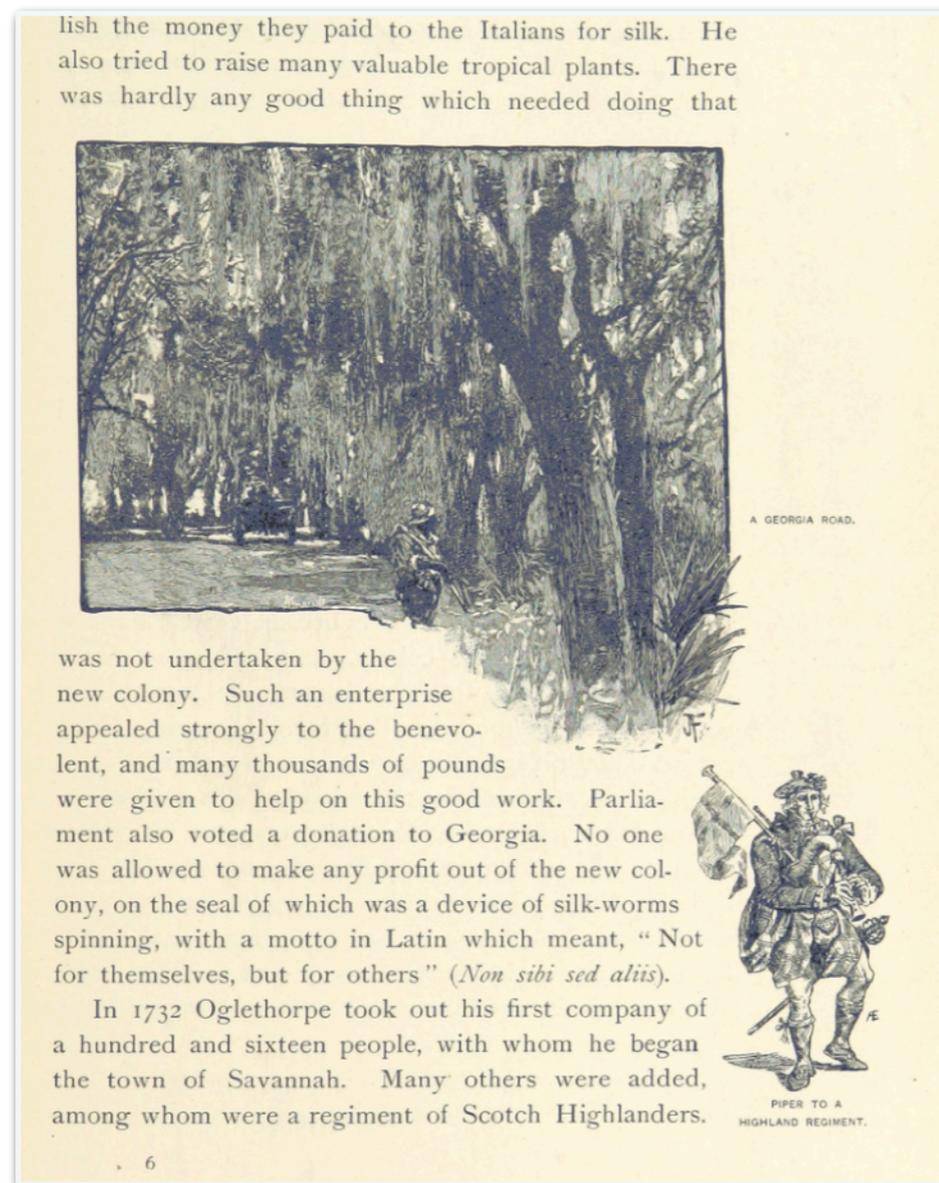
- Our collaborators at the *British Library Labs* in London provided a collection of digitised out of copyright 18th and 19th century books.
- Originally scanned by Microsoft in 2007-2008. The project was never fully completed.
- The digital corpus has 63,985 volume files in total (around 25 million pages).
- When we do not count duplicates or multi-volume books, there are 49,509 unique books in the corpus.



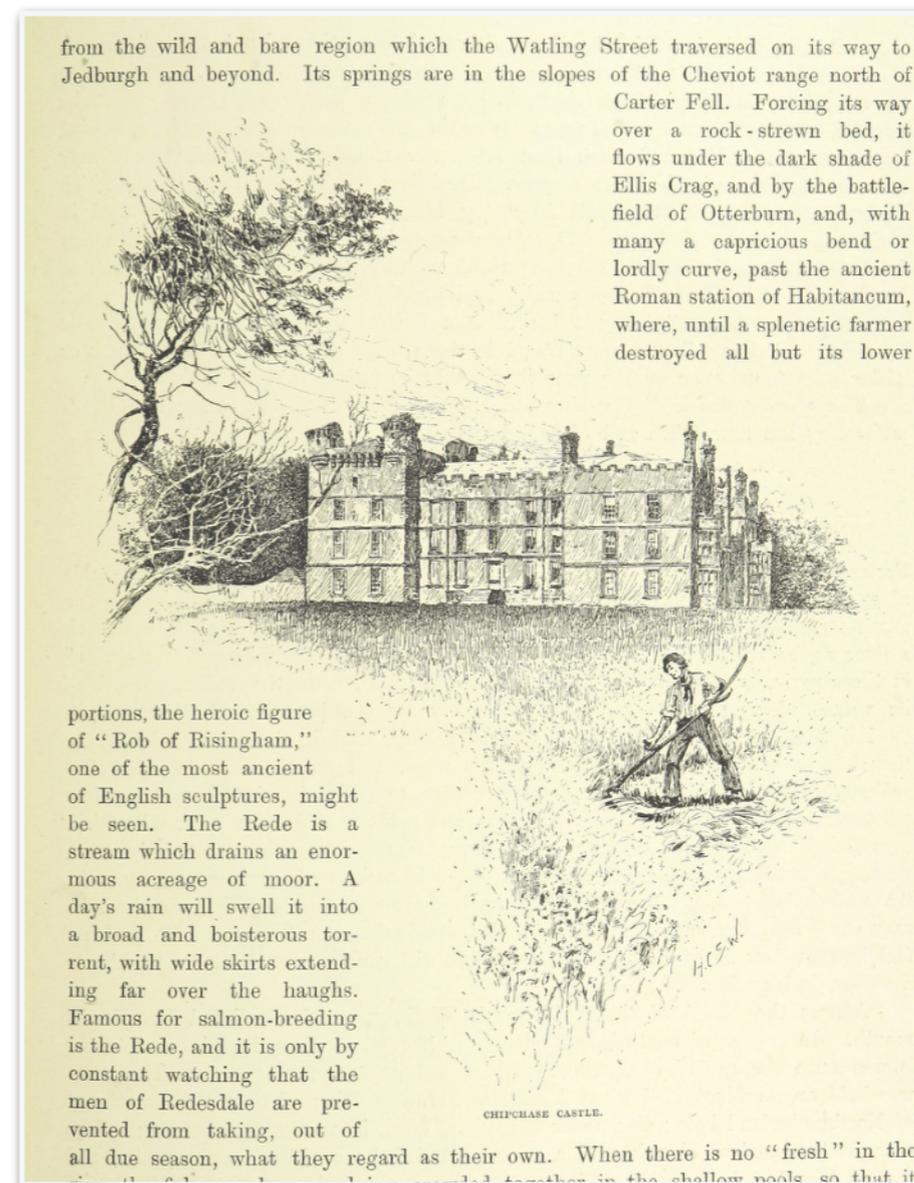
The Dawn of the XIXth Century in England.  
John Ashton (1886)

# British Library Digital Corpus

- The original book scans are stored as high resolution image files. However, we work with plain text versions produced by Microsoft via Optical Character Recognition (OCR).



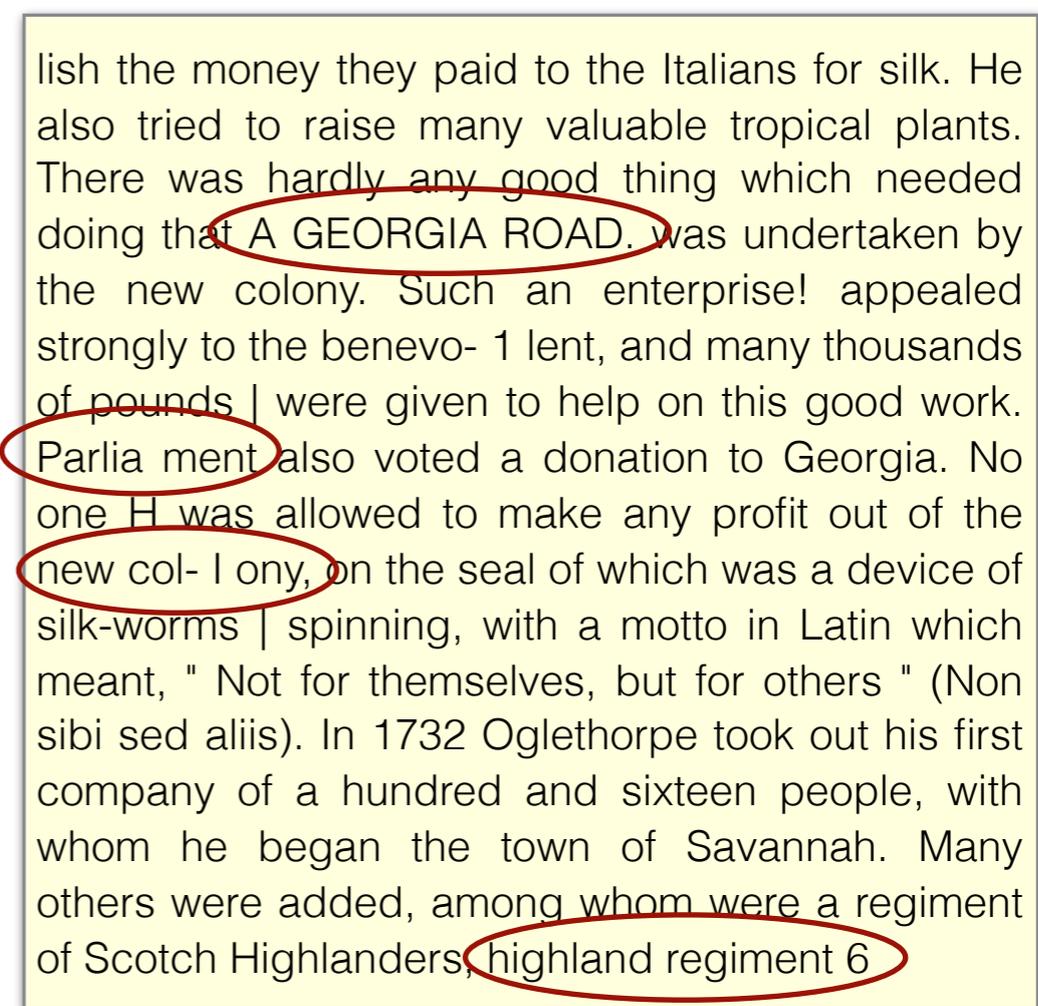
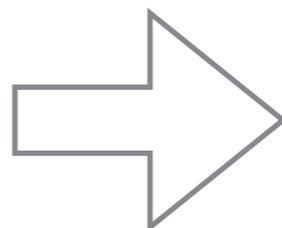
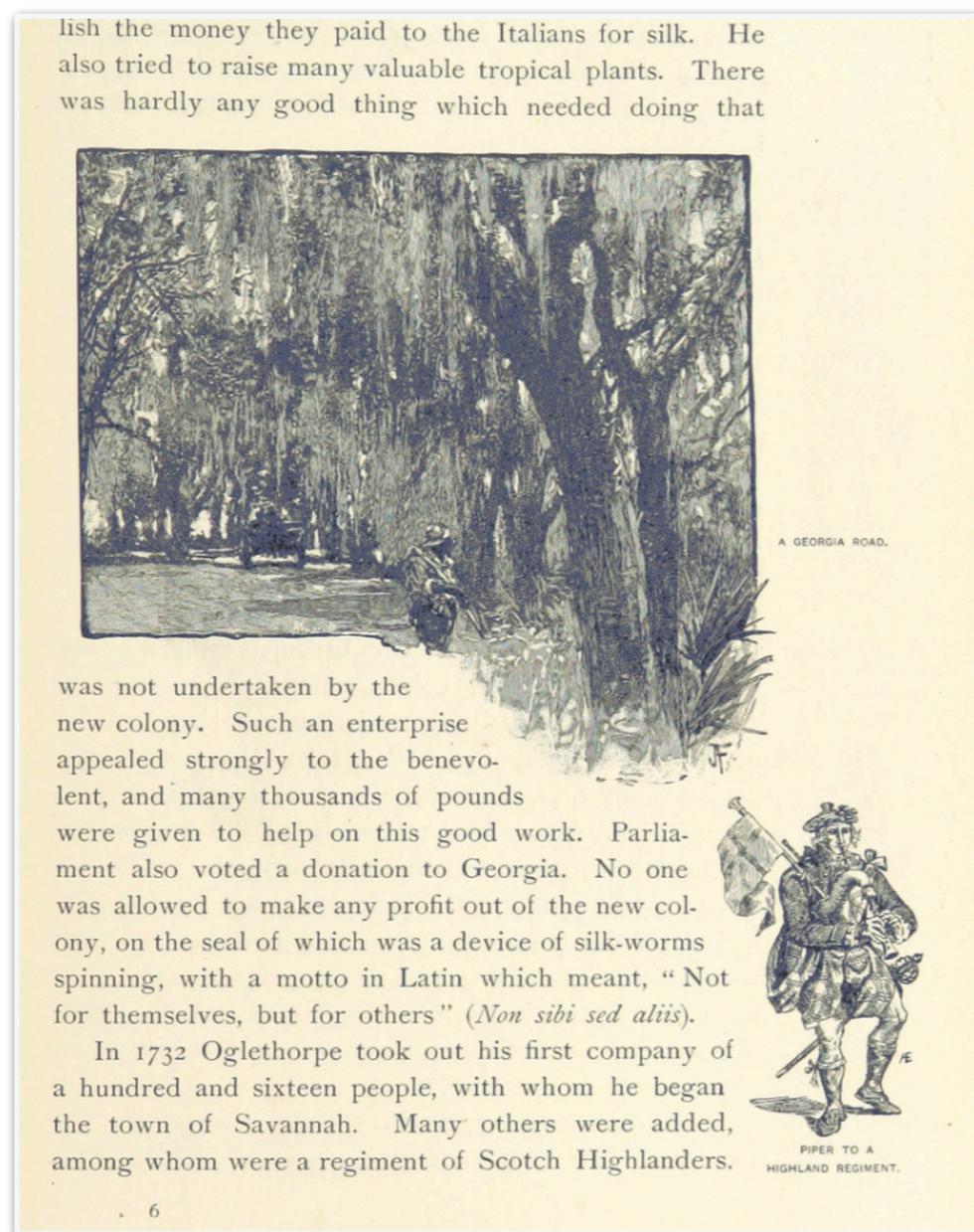
The Household History of the United States and its people. Edward Eggleston (1889)



Rivers of Great Britain, the Thames, from Source to Sea. Author Unknown (1891)

# British Library Digital Corpus

- The quality of the OCR output varies from book to book. Some books are close to the original, others are more difficult to read. The original formatting has been lost in all cases.



The Household History of the United States and its people.  
Edward Eggleston (1889)

# Corpus Metadata

- Each book in the corpus has associated metadata to describe it.
- Some metadata fields contain erroneous or missing values, particularly for year and publication location. Author names are also sometimes missing or contain errors.

Field	Value
<b>Title</b>	[Mansfield Park: a novel.], Single Works
<b>Authors</b>	{'creator': ['Austen, Jane']}
<b>Year</b>	1875
<b>Edition</b>	[Another edition.] Illustrated from drawings by A. F. Lydon.
<b>Issuance</b>	Monographic
<b>Language</b>	English
<b>Place</b>	London
<b>Publisher</b>	Groombridge & Sons
<b>Shelfmark</b>	British Library HMNTS 12618.h.4.
<b>Shelfmark Number</b>	12618

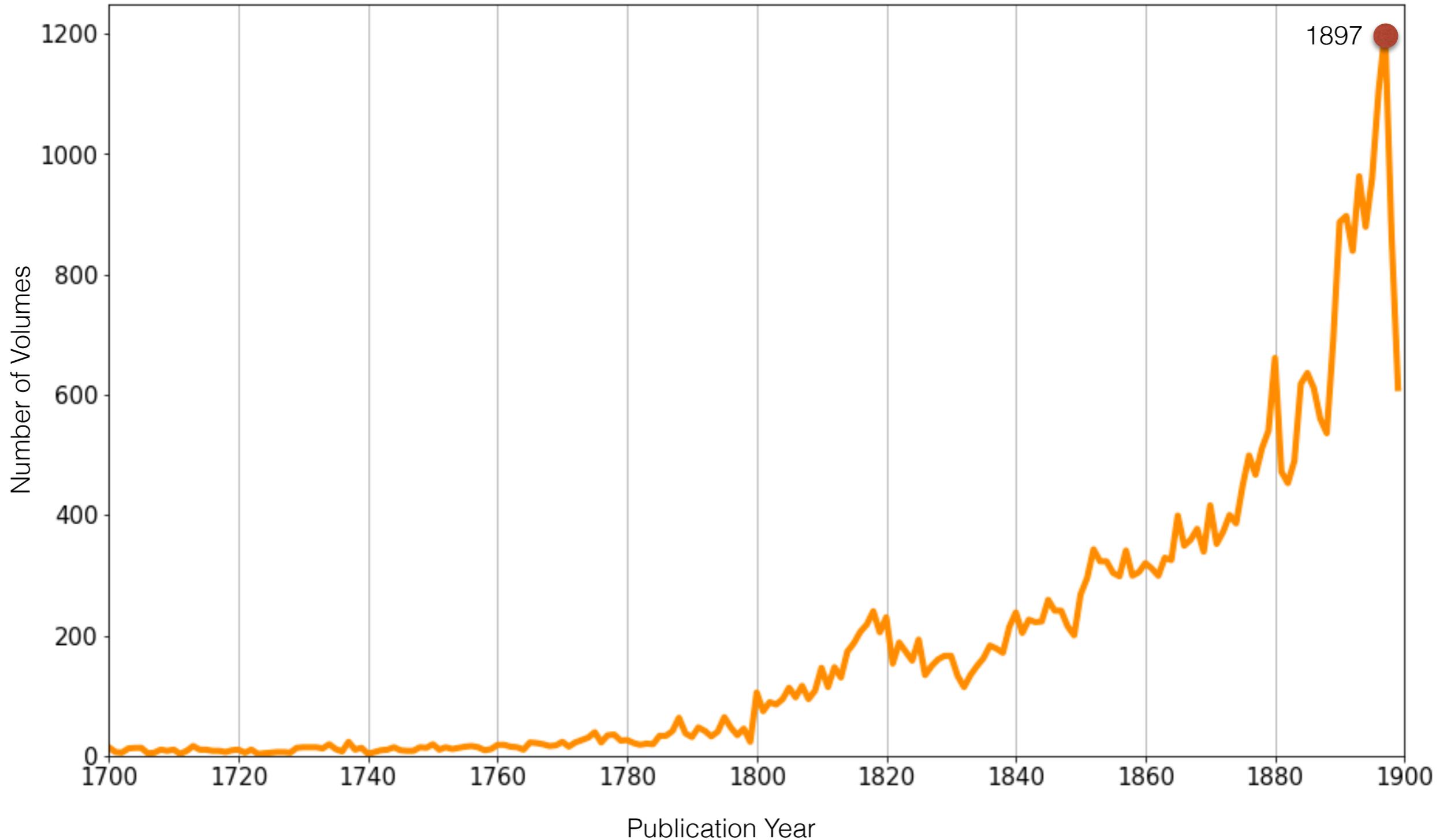
Field	Value
<b>Title</b>	Silas Marner ... By the author of "Adam Bede", etc, Silas Marner
<b>Authors</b>	{'creator': ['Eliot, George']}
<b>Year</b>	1863
<b>Edition</b>	
<b>Issuance</b>	Monographic
<b>Language</b>	English
<b>Place</b>	S. H. Goetzel
<b>Publisher</b>	Mobile
<b>Shelfmark</b>	British Library HMNTS 12623.g.27.
<b>Shelfmark Number</b>	12623

# Corpus Filtering

- Before performing any detailed analysis of the data, we filtered the full corpus using the following steps:
  1. We removed books where key metadata was missing or invalid - i.e. year or title. We retained books where the author was anonymous.
  2. We focused on 18th and 19th century, so filtered books where the year was pre-1700 or post-1899.
  3. We applied language detection to remove non-English texts. This required a further detection step to remove classical Greek and Roman books.
- This filtering process reduced the original set of 49,509 unique books to a subset of 35,918 English language books.
- Of these, 5,484 books consist of two or more volumes, yielding 46,438 "documents" in total.

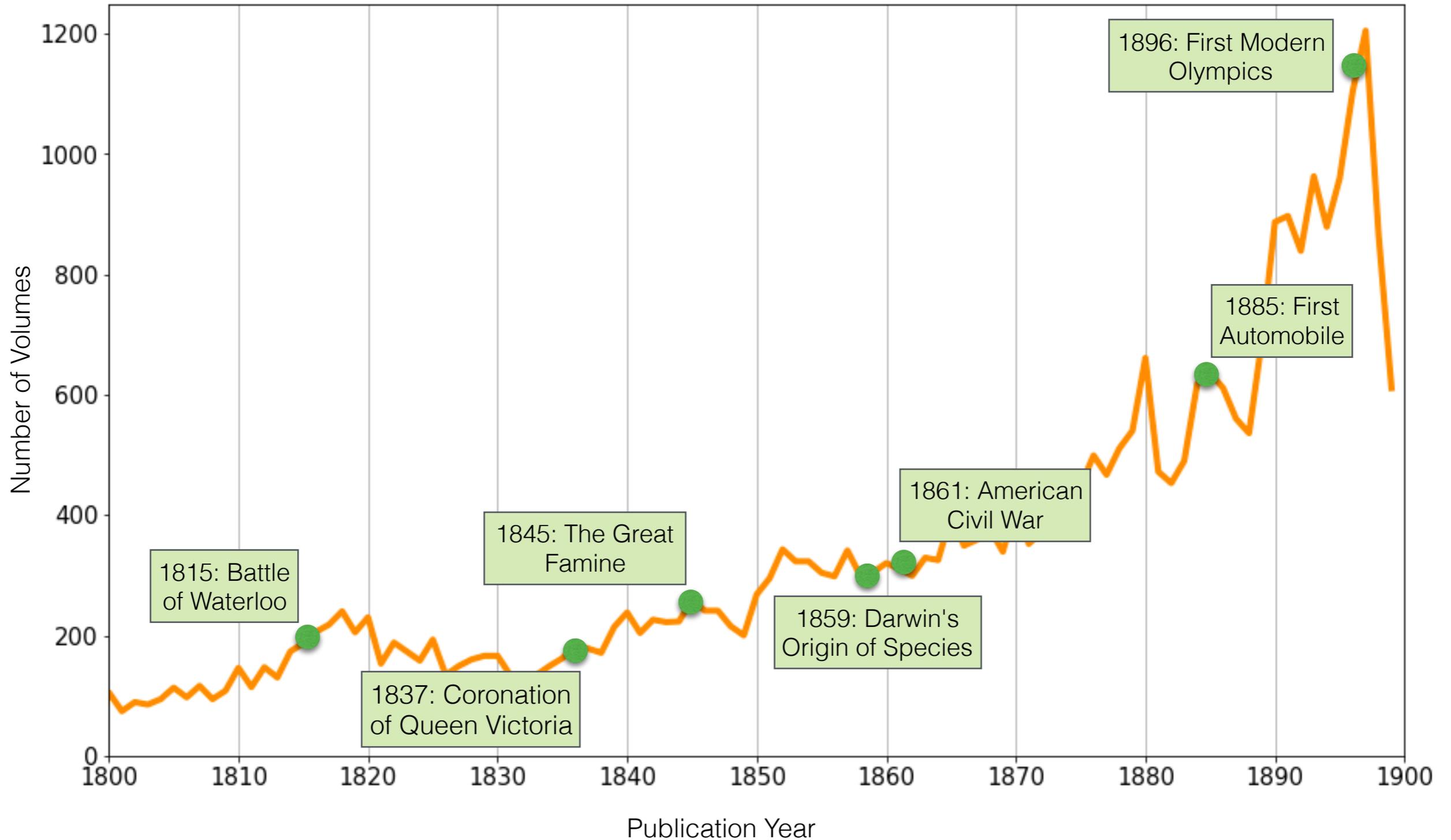
# Volumes by Year

Number of Volumes by Year in English Language Corpus



# Volumes by Year

Number of Volumes by Year in English Language Corpus (19th Century)



# Publication Locations

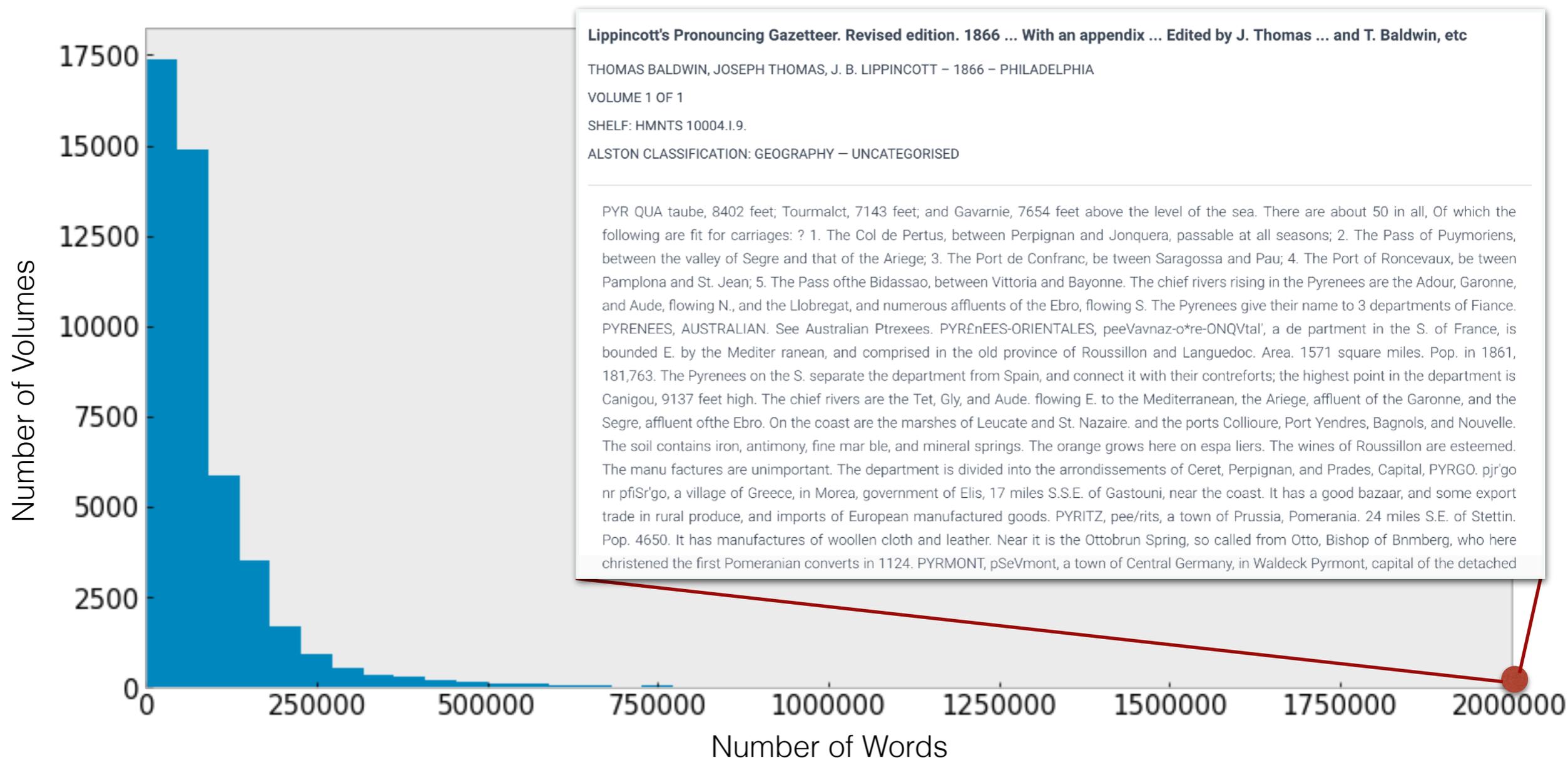
- The corpus metadata lists 1,462 distinct publication locations, with the vast majority of texts listed as published in London.

Location	Number of Texts
London	23,853
Edinburgh	1,219
New York	1,137
Boston	620
Philadelphia	448
Dublin	383
Glasgow	222
Chicago	191
Oxford	187
Calcutta	182
Manchester	180
Cambridge	168



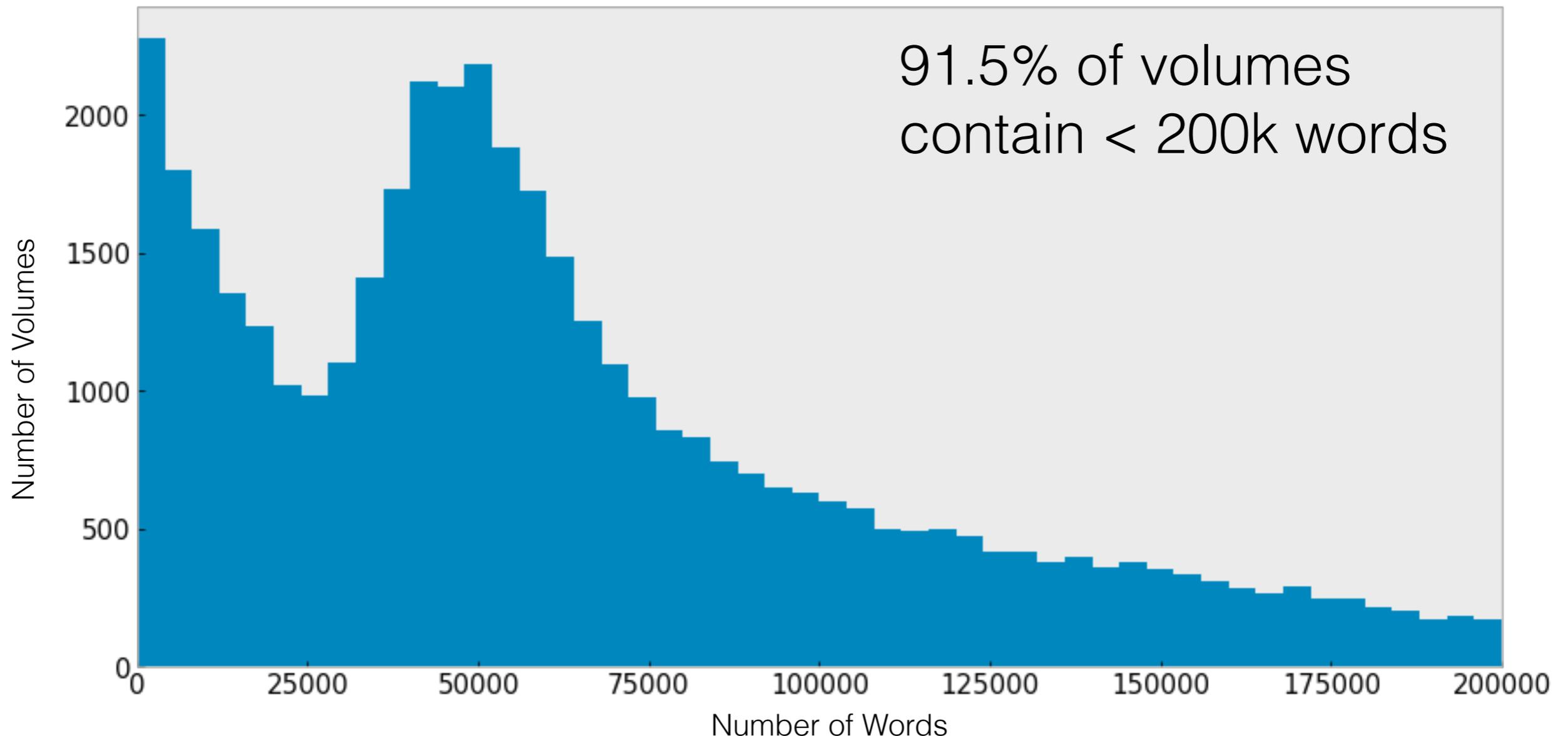
# Word Counts

- By splitting the text of each volume into individual words, we see that the complete English corpus contains 4,092,977,777 words.
- The average volume length is 88k words, but the distribution of volume word counts is highly skewed.



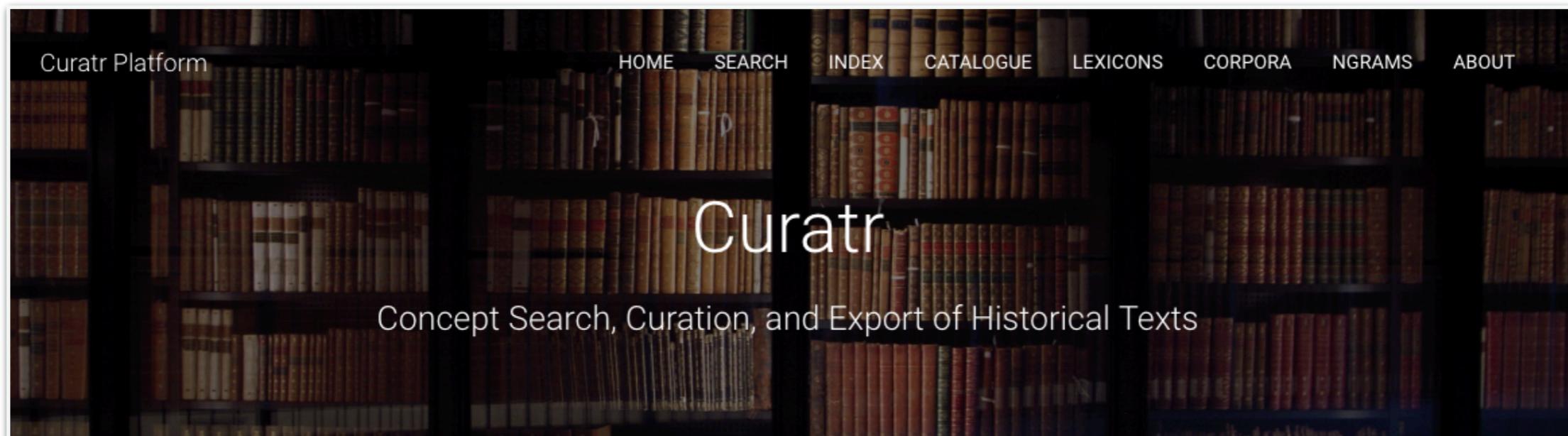
# Word Counts

- By splitting the text of each volume into individual words, we see that the complete English corpus contains 4,092,977,777 words.
- The average volume length is 88k words, but the distribution of volume word counts is highly skewed.



# Curatr Platform

- To make the British Library Corpus accessible to humanities researchers, we developed the online *Curatr* platform.



## Curatr Overview

*Curatr* is an online platform for the exploration and curation of the British Library Corpus, comprising 35,918 unique English language texts, and a searchable interface of 12.3 million individual pages of text, which can be filtered and sorted by author, title and year. The interface also incorporates the topical classification index of volumes used by the British Library from 1823-1985.

*Curatr* enables keyword and lexicon search functions, which help locate your research topics within little known or very long unwieldy texts. It also supports the development and export of sub-corpora, defined thematically, chronologically and by classification. The development of this system addresses the requirements for online document curation without the need for extensive technical training, but the quality of the results depends on the accuracy and focus of the search terms or lexical seed words provided by the researcher. The use of chronological filters is strongly recommended.

There are inevitable variations in the legibility of earlier texts and those in non-standard formats. A key use of *Curatr* is to assist researchers to identify original texts relevant to their work for consultation in situ in the library. The system is built upon a corpus curation workflow that incorporates elements of information retrieval techniques and the use of word embeddings from the field of natural language processing.

# Corpus Search

- A key requirement for the web interface was to make the entire English language corpus easily searchable by researchers.
- We consider two different document representations:
  1. Complete text of each **volume** (46,438 volumes).
  2. Short text **segments** of length 2,000 characters. Equivalent of approximately one page in a modern paperback (12.3m segments).

...to tell what he's suffering from if you saw him." I at once went below, and found the lad who waited upon us in the cabin preparing the table for breakfast.

6 THE PHANTOM DEATH. I asked him if the captain was alone. He answered that Mr. Stroud, the chief mate, was with him. On this I went to the door of Captain Joyce's cabin and lightly knocked. The mate looked out, and, seeing who I was, told me in a soft voice to enter. Captain Joyce lay in his bunk dressed in a flannel shirt and a pair of white drill trousers. All his throat and a considerable portion of his chest were exposed, and his feet were naked. I looked at him scarcely crediting my sight : I did not know him as the man I had parted with but a few hours before. He was swelled from head to foot as though drowned : the swelling contorted his countenance out of all resemblance to his familiar face ; the flesh of him that was visible was a pale blue, as if rubbed with a powder of the stuff called "blue" which the laundresses use in getting up their linen. His eyes were open, but the pupils were rolled out of sight, and the " whites," as they are called, were covered with red blotches. I had no knowledge of medicine, and could not imagine what had come to the poor man. He was unconscious, and evidently fast sinking. I said to Mr. Stroud, " What is this ? " The mate answered, " I'm afraid he's poisoned him self accidentally. It looks to me like poison. Don't it seem so to you, sir? See how his fingers and toes are curled." I ran my eye over the cabin and exclaimed, "Have you searched for any bottles containing poison ?"

THE PHANTOM DEATH. 7 " I did so when he sent for me at four o'clock, and complained of feeling sick and ill. He was then changing colour, and his face was losing its proper looks. I asked him if he thought he had taken anything by mistake. He answered no, unless he had done so in his sleep. He awoke feeling very bad, and that was all he could tell me." I touched the poor fellow's hand, and found...

The Phantom Death, and other stories, etc.

William Clark Russell (1895)

Segment 5 of 260

# Corpus Search

- The corpus can be searched and filtered based on a range of criteria, include both textual content and metadata.

## Corpus Search

We currently index content and metadata for 46,438 volumes and 12,334,075 short text segments from 35,918 different English language texts dating from 1700 to 1899. You can use the form below to search the entire corpus or a specific subset of the corpus.

Search Keywords  
vaccination

Search Field	Document Type
Full Text	Volumes

Classification	Sub-classification
All Classifications	All Sub-classifications

Start Year	End Year
1850	1890

Publication Location  
London

SEARCH

# Corpus Search

- The corpus can be searched and filtered based on a range of criteria, include both textual content and metadata.
- Researchers can use all of the standard search query operations found in popular online search engines.

```
vaccine vaccination smallpox
```

Search for texts containing one or more keywords

```
"Queen Victoria"
```

Search for the exact phrase contained in quotes

```
irish histor*
```

Use \* to perform a wildcard search matching any characters

```
smallpox +ireland -vaccination
```

Use + for mandatory words and use - for prohibited words

# Corpus Search

- Search results are ranked by relevance to the query words and filtering criteria, and are presented with extracts of matching text.

## Search Results

558 matching volumes were found for **vaccination** from 1850 until 1890 published in London

  
EXPORT

---

**Miscellanies: chiefly addresses, academical and historical (moral and religious, political and social) – Volume 5**  
Francis William Newman – 1869 – London

- Science Eeligious in the long run - 275 On the Weakness of Eoman Empire - - 281 On the Eelation of the Supply of Food to the Laws of Landed Tenure - - - 304 The Political Side of the **Vaccination** System - 321 Religion at the Bar of Ethics - - - 331 Preamble to Shakespear's Eichard II. - - 343 From the...

**Miscellanies: chiefly addresses, academical and historical (moral and religious, political and social) – Volume 3**  
Francis William Newman – 1869 – London

- Disestablishment and Disendowment ? . 424 From Fraser's Magazine, August 1877. On the War Power, ..... 433 From Tlee Contemporary Review, Nov. 1877. The Barbarisms of Civilization, . . . 449 From T/te Contemporary Review, June 1879. Compulsory **Vaccination**, . . . 471 Written by request for a Public Meeting. The...

**The Children of Lutetia, Single Works – Volume 1**  
William Blanchard Jerrold – 1864 – London

- conducted— The Census of the Paris Poor – Private Œuvres— The Man who introduced **Vaccination** into France— De Gerando's " Visitor of the Poor" – Jules Simon on Habitual Beggars, and False Paupers— The Vices of the Hospital System, . . . . . 1 CHAPTER III. The Rue Traversine and the Rue Tirechape— The...
- Statistics— The Bureau of Benevolence of the First Ar- rondissement – A Sale of Bridal Bouquets— Groups of Happy Couples –The Bureau of the Fifth Arrondissement and its Approaches – Management of the Doctors of the Poor – Administrators and their Helpers ; their Mode of giving Relief – **Vaccination** Tickets...
- **Vaccination** Bribes— The Rue Vert-Bois – The Exterior of a House of Help, ..... CHAPTER VI. The Great Paris Nursery for Poor and Abandoned Children – St. Vincent de Paul – The Creche – Rows of Cradles— Business-like Nursing— The Sick-Wards— The Weaning Department— A Primitive Toilette – Bri- gades of Children...

**Lord Bantam. By the author of "Ginx's Baby" i.e. J. E. Jenkins – Volume 1**  
Author Unknown – 1872 – London

- child, after a careful examination to decide Avhether he Avas strong enough to bear it. "Sept. 24. Alfred's **vaccination** took : he is very feverish and restless. I asked Mr. BelleAA', and he tells me he never kneAV it to be fatal. "Sept. 25. Alfred's arm very much inflamed. SAVanston says he is a...

# Corpus Search

- The text of individual volumes or segments can be accessed for "close reading" via the search results.

Title:	The Dawn of the XIXth Century in England. A social sketch of the times ... With ... illustrations, etc.
Segment:	Volume 1 of 1, Segment 314 of 382
Author(s):	John Ashton
Published:	1886 – London (T. Fisher Unwin)
Edition:	Popular edition.
Classification:	History – English 18th, 19th Centuries
Physical Description:	xx, 476 pages
Shelfmark:	HMNTS 9525.b.6.
Related Links:	<a href="#">Volume Full Text</a> – <a href="#">British Library Catalogue</a> – <a href="#">Volume PDF</a> – <a href="#">Volume Images</a>

...f prevention then believed in, as the patient generally suffered but slightly from the inoculation, and it was then a creed, long since exploded, that small-pox could not be taken tAvice. This fact of their resistance to variolous inoculation set him thinking, and he came to the conclusion that they had absorbed into their systems, a counter poison in the shape of some infection taken from the coavs. He made many experiments, and found that this came from a disease called the cow pock, and that the vaccine lymph could not only be taken direct from the coav, but also by transmission from the patients Avho had been inoculated Avith that lymph, and Avhence the present system of so-called vacci nation – the greatest blessing of modern times. Jenner, of course, Avas opposed ; fools do not even believe in vaccination noAv, and great Avas the battle for, and against, in the medical profession, and many Avere the books Avritten pro and con. " Vaccination Vindicated," Ed. Jones ; "A Reply to the Anti-Vaccinists," Jas. Moore; "The Vaccine Contest," Wm. Blair; "Cow Pock Vacci nation," Rowland Hill ; " Birch against Vaccination," " Willan on Vaccination," &c, &\*c. Gillray could not, of course, leave such a promising subject alone, and he perpetrated the accompanying illus tration. Here Dr. Jenner (a very good likeness) is attend ing to his patients – vaccinating, rather too vigorously, one lady – the lymph, in unlimited quantity, being borne by a Avorkhouse boy, and receiving his patients who are ex hibiting the different phases of their vaccination. As a rule, they seem to have " taken " too well.

# Classification Index

- From 1823 to 1985 the British Library used a topical index of shelfmarks to record the physical arrangement of books on shelves in the library.
- The index is arranged in several levels, from the top level main classification (e.g. "Fiction", "Geography", "Medicine") to more fine-grained topics (e.g. "Geography": "Ireland & Scotland").

463-467	Bibles	
468	Service Books	
469-477	Theology	Fathers
478-480	English Theology	
481	Theology	Collected Works
482-484	Church History	
485	Lives of Saints	
486		Church History
487	Church History	
488-491		Religious Biography
492-496	Theology	Collected Works
497-504	Law & Jurisprudence	
505-519	English Law	
520-521	Aristotle	Philosophy particularly Logic
522-524	Politics	
525-528	Moral Philosophy	
529-530	Mathematics	
531-533	Astronomy	
534		Military Art
535-536	Natural Philosophy	
537-538	Mechanics	
539-549	Medicine	
550-560	Art	Illustrated Books
562-565	Art	Topography
566-571	Geography	
572	Topography	Greece & Rome

731-732	Education	Universities
733	Academies	
734-751	Encyclopaedias & Series	
752-761	Removed to Oriental Department	
762	Statutes	
763	Directories	
764	Boase Collection of Irvingite Books	
765-770	Fine Arts	Archaeology
771-772	State Papers	
773-777	Medicine	
778-779		Midwifery
780	Surgery & Anatomy	
781	Morbid Anatomy	
782-783	Surgery	
784	Physiology	
785	Music	
786	Art	
787	Archeology	& Art
788		& Architecture
789-792	Topography	Travels
793-795	Geography	Topography
796		England
797		Scotland & Ireland
798		General America
799-801	History	General
802		Ancient
803		Ancient & Medieval

793-795	Geography	Topography
796		England
797		Scotland & Ireland
798		General America

# Classification Index

- Previously no useful digital version of the index existed, only a scan of the 1987 version.
- Using a combination of OCR and manual annotation, we extracted the two top levels of the index.
- We then classified 98% of the English language texts from the British Library Digital Corpus, by linking the index pressmarks to the shelfmarks in the corpus metadata.

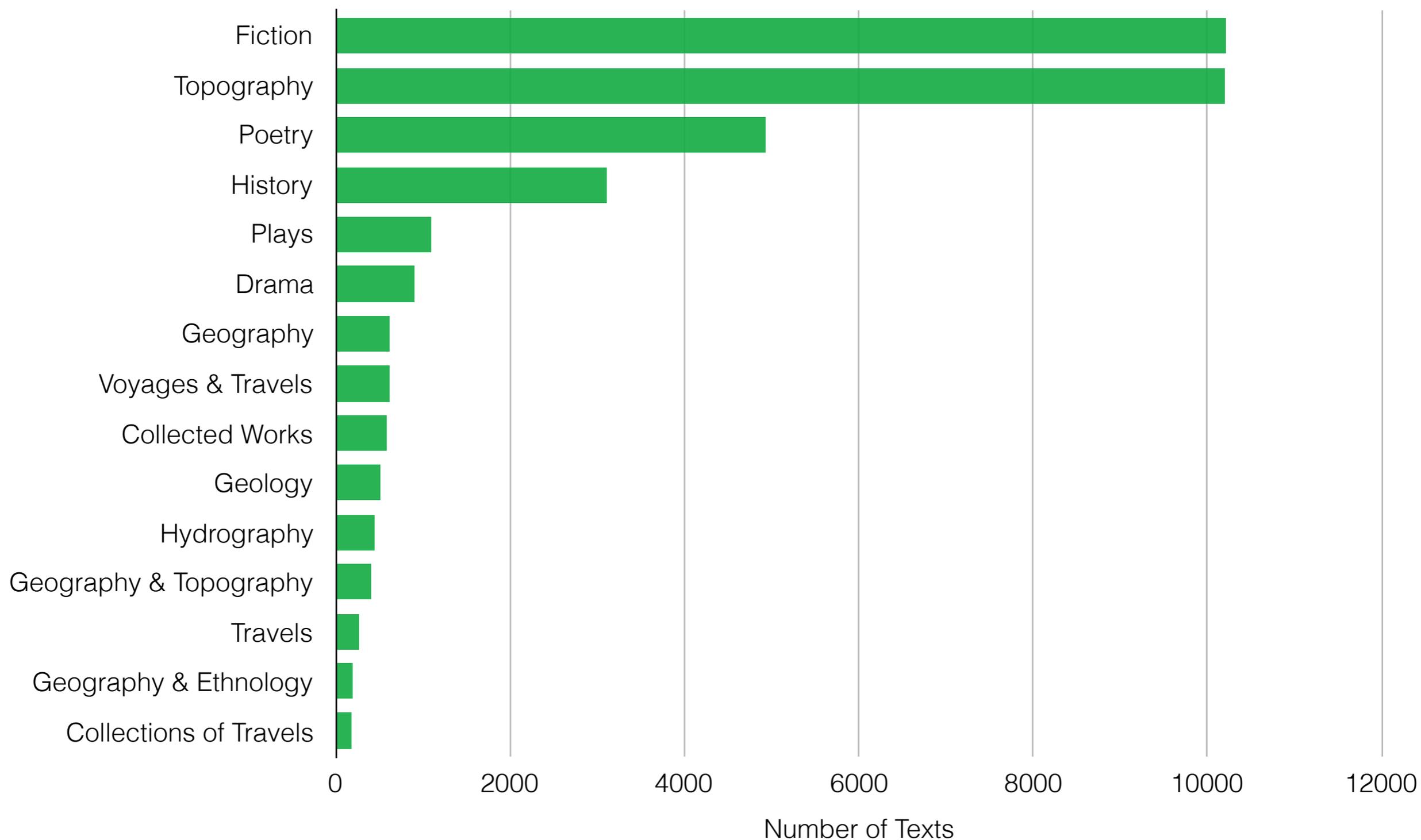
```
"id": "002111057_01_000029",  
"title": "An essay on Government. By Philopatria (R. F. A. L.)",  
"book_id": "002111057",  
"volume": 1,  
"segment": 29,  
"max_segment": 151,  
"year": 1808,  
"authors": ["Lee, Rachel Fanny Antonina"],  
"location": "London",  
"shelf": ["HMNTS 521.i.24."],  
"shelf_num": 521,  
"content": "with respect to Cassibelaunus, who fought against Cæsar",  
"_version_": 1611609354732044288}]
```

505-519	English Law	
520-521	Aristotle	Philosophy particularly Logic
522-524	Politics	
525-528	Moral Philosophy	
529-530	Mathematics	
531-533	Astronomy	
534		Military Art

An essay on Government. By Philopatria, (R. F. A. L.)  
Rachel Fanny Antonina Lee – 1808 – London  
Classification: Aristotle – Philosophy, particularly logic

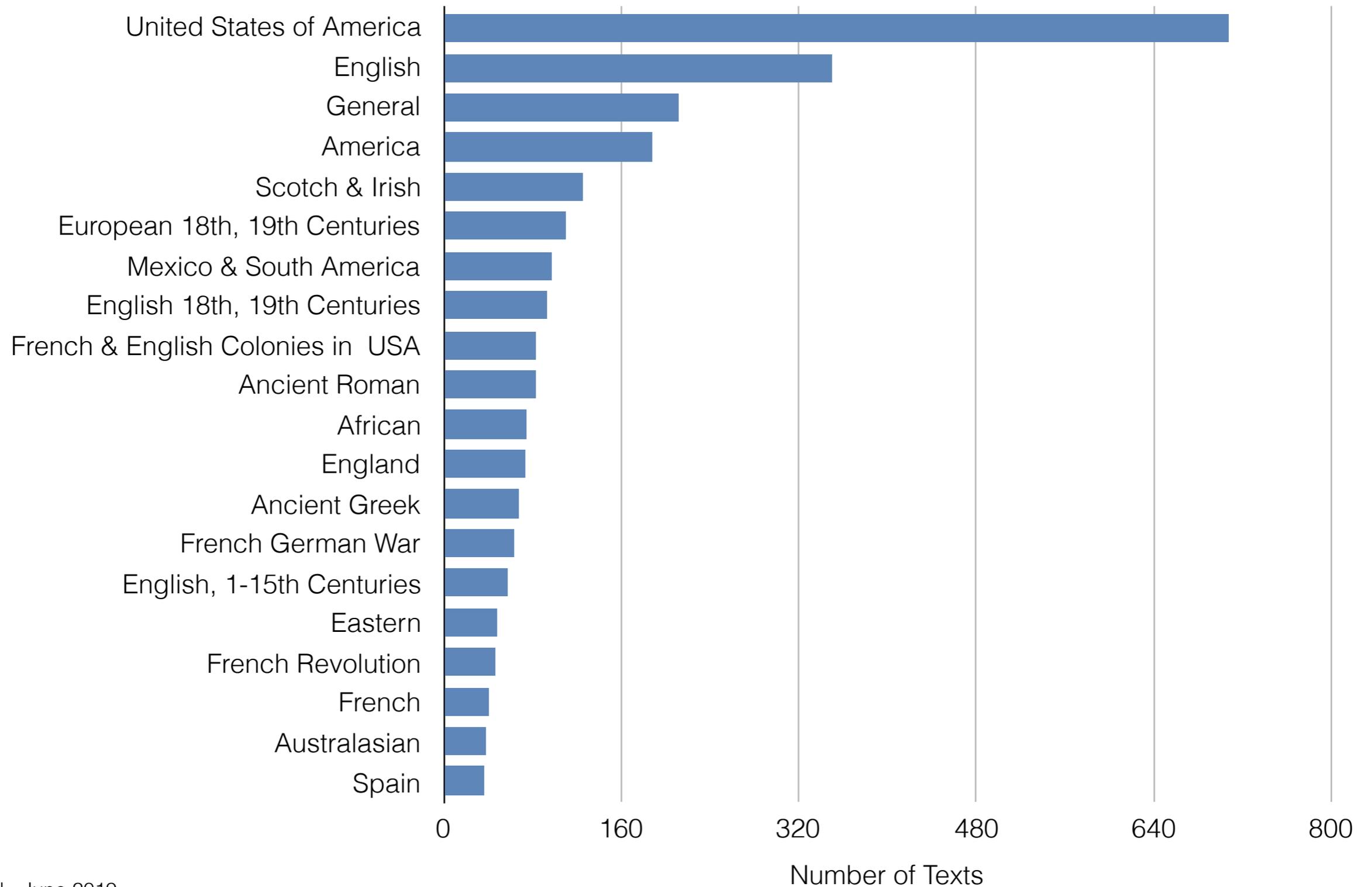
# Classification Index

Top 15 Top Level Classifications in English Language Corpus



# Classification Index

## Top 20 Sub-Classifications for Top Level Classification "History"



# Classification Index

- The subject classification index can be used as filtering criteria when searching the corpus.

## Corpus Search

We currently index content and metadata for 46,438 volumes and 12,334,075 short text segments from 35,918 different English language texts dating from 1700 to 1899. You can use the form below to search the entire corpus or a specific subset of the corpus.

Search Keywords

Enter search keywords

---

Search Field	Document Type
Full Text	Volumes
<b>Classification</b>	<b>Sub-classification</b>
Fiction	English 3 Volume Novels

---

Start Year	End Year
Start YYYY	End YYYY

---

Publication Location

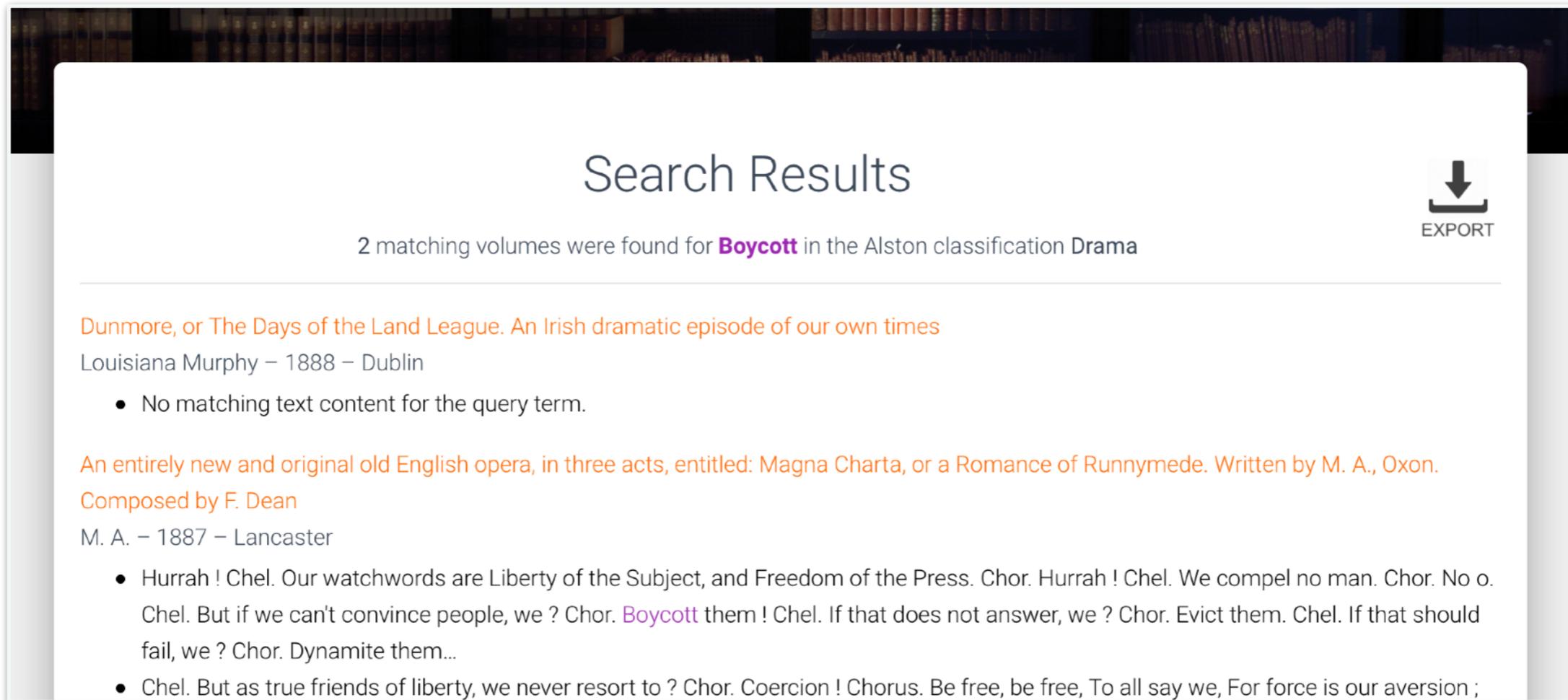
All Locations

---

**SEARCH**

# Rare Material

- Combining search queries with the classification index allows researchers to uncover rare material for research and teaching.
- Example: Searching for the keyword "boycott" in the "Drama" classification returns an opera about the Ladies Land League and another about Magna Carta.



The screenshot shows a search results interface. At the top, the title "Search Results" is centered. Below it, a message states "2 matching volumes were found for **Boycott** in the Alston classification Drama". In the top right corner, there is a download icon and the word "EXPORT". The results are listed below a horizontal line. The first result is "Dunmore, or The Days of the Land League. An Irish dramatic episode of our own times" by Louisiana Murphy, dated 1888, from Dublin. It includes a bullet point: "No matching text content for the query term." The second result is "An entirely new and original old English opera, in three acts, entitled: Magna Charta, or a Romance of Runnymede. Written by M. A., Oxon. Composed by F. Dean" by M. A., dated 1887, from Lancaster. It includes two bullet points: "Hurrah ! Chel. Our watchwords are Liberty of the Subject, and Freedom of the Press. Chor. Hurrah ! Chel. We compel no man. Chor. No o. Chel. But if we can't convince people, we ? Chor. **Boycott** them ! Chel. If that does not answer, we ? Chor. Evict them. Chel. If that should fail, we ? Chor. Dynamite them..." and "Chel. But as true friends of liberty, we never resort to ? Chor. Coercion ! Chorus. Be free, be free, To all say we, For force is our aversion ;".

# Discovery Tool

- *Curatr* search can also act as a discovery tool for the physical collection of the British Library.
- Example: Finding lesser known poetry

Title:	The Botanical and Horticultural Meeting, or Flora's and Pomona's Fete; a poem. ... By a Lady Mrs. E. S. Perkins . From notes by John-Quik. Second edition
Volume:	Volume 1 of 1
Author(s):	Elizabeth Steele Perkins, John Quill
Published:	1834 – Birmingham
Edition:	N/A
Classification:	Poetry – English, Separate Works, 19th Century
Physical Description:	N/A
Shelfmark:	HMNTS 11641.e.20.
Related Links:	<a href="#">British Library Catalogue</a>

FLORA AND POMONA'S FETE.

THE BOTANICAL AND HORTICULTURAL MEETING, on FLORA'S AND POMONA'S FETE. A POEM, IN HUMBLE IMITATION OF THE BUTTERFLY'S BALL, fjc. \$c. AND RESPECTFULLY DEDICATED TO THE MEMBERS OF THE VARIOUS botanical and horticultural societies. BY A LADY. FROM NOTES BY JOHN-QUILL. And under a Flag, looking lanky and taper, Stood Jonquil, to write a report for the paper ! SECOND EDITION. The Profits arising from the Sale of this little Book will be entirely devoted to Charity. BIRMINGHAM : BE1LBY, KNOTT, AND BEILBY. 1834. PR1CL TWO SHILLINGS AVD SIXPENCE.

# Corpus Catalogue

## Catalogue

The catalogue below contains a subset of the metadata for 35,918 different English language texts dating from 1700 to 1899, where each row corresponds to a unique text. You can use the search box on the right-hand side to filter the texts.

Search:

Year ▲	Title	Author(s) ▼
1700	Caledonia; or, the Pedlar turn'd merchant. A tragi-comedy, as it was acted by His Majesty's subjects of Scotland, in the King of Spain's Province of Darien. In verse.	Author Unknown
1700	Friendship improv'd, or the Female Warriour; A tragedy in five acts and in verse	Charles Hopkins
1700	Amasia, or, The works of the Muses. A collection of poems	John Hopkins
1700	A Journey to Hell: or, a visit paid to the Devil: a poem. By Edward Ward.	Edward Ward
1700	An ode performed at the anniversary feast of the Gentlemen, natives of the County of Kent, at Marchant-Taylors-Hall, Nov. 21, 1700, Appendix	Author Unknown
1700	The Generous Choice. A comedy in five acts, and in prose, with songs	Francis Manning
1700	An Elegaic Essay upon the Decease of the Groom-Porter i.e. T. Neale , and the Lotteries	Thomas Neale
1700	Sacellum Honoris. A congratulatory poem to ... the Marquis of Tavistock, on his happy return from travel	Elkanah Settle
1700	The Fate of Capua; a tragedy in five acts and in verse	Thomas Southern
1700	The Olde, Old, Very Olde Man: or, The Age and long life of Thomas Par ... who was Borne in the Raigne of King Edward the 4th. and is now living in the Strand, being aged 152 yeares and odd Monethes, etc.	John Taylor
1700	The Venetian Marriage	Author Unknown
1700	The Plain-Dealer ... The fifth edition. , Single Works	William Wycherley
1700	Carmen sæculare, for the year 1700. To the King. By Matthew Prior.	Author Unknown
1700	The Perjur'd Husband: or, the Adventures of Venice. A tragedy	Susanna Centlivre
1701	The Bath; or, The Western Lass. A comedy in five acts and in prose	Thomas D'urfey

# Corpus Catalogue

## Authors

The table below contains a list of the 16,590 authors relating to English language texts dating from 1700 to 1899. You can use the search box on the right-hand side to filter the author list. Clicking a row in the list will search for volumes matching the author name.

Search:

Name	Count	Start Year	End Year
Byron, George Gordon Byron	129	1807	1898
Scott, Walter	100	1800	1899
Wood, Henry	78	1862	1899
Oliphant, Margaret	73	1849	1897
Marryat, afterwards Church, afterwards Lean, Florence	58	1865	1898
Goldsmith, Oliver	53	1765	1898
Pindar, Peter - pseud.	50	1783	1824
Shakespeare, William	49	1703	1893
Ainsworth, William Harrison	47	1847	1893
Burns, Robert	44	1797	1898
Norie, John William	41	1827	1859
Payn, James	41	1858	1898
Carey, Rosa Nouchette	40	1869	1899
Dickens, Charles	40	1836	1899
Fenn, George Manville	39	1867	1899
Riddell, J. H	39	1858	1897

# Word Lexicons

- When searching large text corpora, a common task involves creating **word lexicons** (i.e. lists of words of interest) relating to specific themes and topics.
- Typically this is a manual process, where a small initial set of "seed" words is expanded through a process of trial and error.

Topic	Lexicon Word List
Agriculture	agriculture, agriculturist, farming
Education	education, school, training
Smallpox	smallpox, pox, cowpox

# Word Lexicons

- When searching large text corpora, a common task involves creating **word lexicons** (i.e. lists of words of interest) relating to specific themes and topics.
- Typically this is a manual process, where a small initial set of "seed" words is expanded through a process of trial and error.

Topic	Lexicon Word List
Agriculture	agriculture, agriculturist, farming, agri, landholder, landowner, farmed, cultivators, husbandry, tillage
Education	education, school, training, tuition, instruction, curriculum, day-school, grammar-school, discipline
Smallpox	smallpox, pox, cowpox, erysipelas, epidemic, inflammation, fever, jenner, vaccination, vaccine

# Word Lexicons

- In the *Curatr* platform, word lexicons can be created and used to refine the corpus search and filtering process.

## Word Lexicons

Word lexicons, which consist of curated lists of keywords related to a given topic, allow scholars to further search the British Library corpus, and to define sub-corpora related to that topic. A lexicon can be expanded based on automatically recommended keywords, which frequently appear in the same context as existing words in the lexicon.

**Existing Word Lexicons:**

Lexicon Name	Description	Keywords	Edit	Delete	Search
<i>Contagion</i>	Lexicon related to fears around the transmission of contagious disease	contagion, contagious, epidemic, infection, infectious, plague			
<i>Disease</i>	General 19th century medical vocabulary related to various diseases	cancer, diphtheria, fever, miasma, pestilence, pneumonia, pox			
<i>Ethnic Identity</i>	Concepts around ethnic identity	federals, fenianism, hebrews, hibernian, jewish			
<i>Immorality</i>	Topics and themes related to immorality	debauchery, immoral, immorality, irreligion, profligacy, venality, vice, wickedness			

# Creating Word Lexicons

- Lexicons are created based on an initial user-specified set of seed words related to the topic or theme of interest.
- Creating appropriate lexicons for niche topics is often a time consuming manual process.

**Create New Lexicon:**

You can alternatively create a new lexicon by completing the form below, including one or more initial "seed" keywords:

Lexicon Name

Disease

Lexicon Description

General 19th century medical vocabulary related to various diseases

Alston Classification

Fiction

Initial Lexicon Keywords

cancer diphtheria fever miasma pestilence pneumonia pox

**CREATE**

# Word Embeddings

- Words that co-occur in similar contexts typically have similar meanings. We can often infer the meaning of a word based on the words which frequently surround it.
- “You shall know a word by the company it keeps” (Firth, 1957)
- We could count frequency of each pair of words appearing in the same paragraph in all books in our corpus - e.g. (Queen, Victoria)

The Reign of Queen Victoria, a survey of fifty years of progress. Edited by T. H. W. – Volume 2

Thomas Humphry Ward – 1887 – London

- REIGN OF QUEEN VICTORIA VOL. II THE REIGN OF QUEEN VICTORIA A SURVEY OF FIFTY YEARS OF PROGRESS EDITED BY THOMAS HUMPHRY WARD, M.A. LATE FELLOW OF BRITISH ASSOCIATION, OXFORD IN TWO VOLUMES VOL. II. LONDON SMITH, ELDER, & CO., 15 WATERLOO PLACE 1887 \AU rt fhti r. CONTENTS OF THE SECOND VOLUME...
- was much narrower than it is. Canada was a thinly7 settled and half-rebellious province, with little sign of the development of those possibilities of extension of popu- VOL. II. b THE REIGN OF QUEEN VICTORIA 2 lation and settlement in the Far West which now make it a great dominion. But apart...
- deficit which could not be grappled with were characteristic features of the early years of her Majesty's reign, and they reflected only too faithfully the general distress. B 2 THE REIGN OF QUEEN VICTORIA 4 It would be impossible in the short space at my disposal to describe at length the successive...

The Reign of Queen Victoria, a survey of fifty years of progress. Edited by T. H. W. – Volume 1

Thomas Humphry Ward – 1887 – London

- Policy and Progress The Editor 403 India Sir Henry Sumner Maine, K.C. S.I. 460 Ireland Sir Rowland Blennerhassett, Bart. . 52'J MAPS. Mai- of England To face Title- p Scotlani Qo flux page l " Ireland . h "29 THE REIGN OF QUEEN VICTORIA. INTRODUCTION. The fifty years which are surveyed in these...
- possibility of our mistaking the broad significance of the changes which have covered England and the world with railways and telegraph wires, have vol. r. B THE REIGN OF QUEEN VICTORIA 2 thronged the sea with steamships, have completed the transformation of our Constitution from a monarchical aristocracy...
- forests. The two counties in Ireland which have increased are those B i THE REIGN OF QUEEN VICTORIA 4 of Dublin and Antrim, examples which offer a further illustration of the tendency of the time towards the flocking of population to capital cities and 10 industrial centres. As to London itself, which...

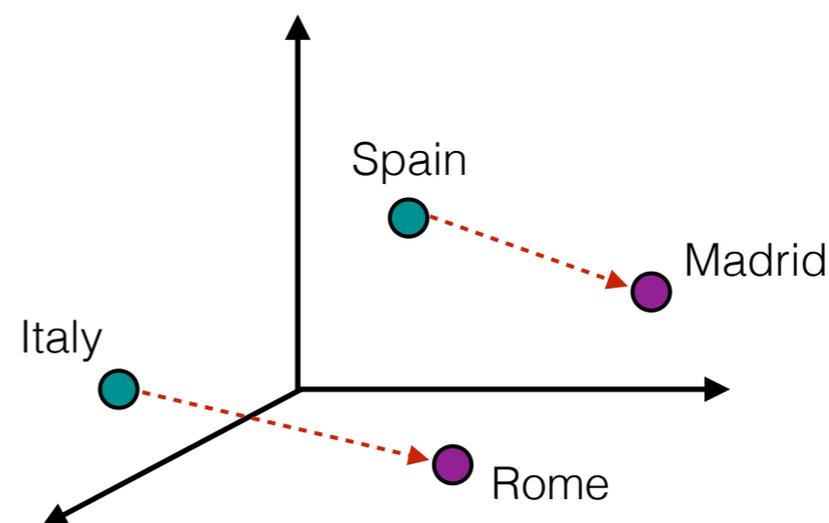
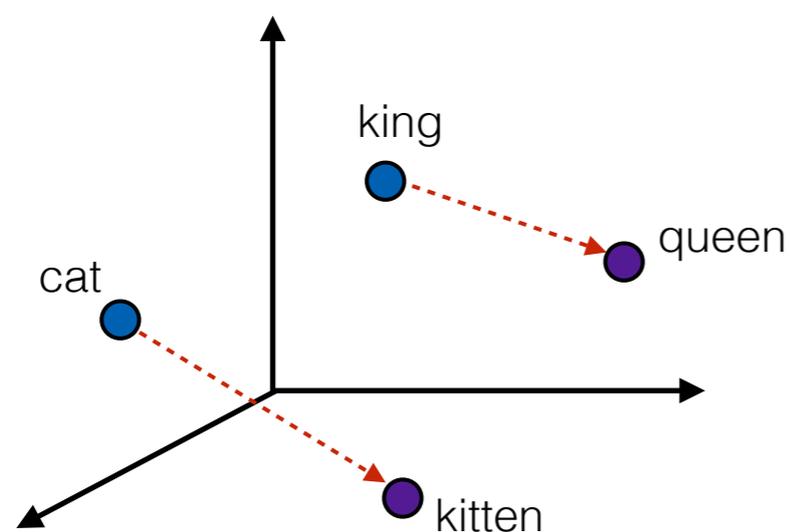
The First Year of a Silken Reign. 1837-8. With ten illustrations from contemporary original plates

Charles Edward Fagan, Andrew White Tuer – 1887 – London

- throne." – Lord Beaconsfield. INTRODUCTION. -[Ftftg gears fiave passed since Queen Victoria ascended the throne of these realms. Three other monarchs only of the thirty-five who have ruled England since the Norman Conquest completed the fiftieth year of their reign. But the half century of Queen...
- each. The following is a list of sovereigns who were reigning in Europe on the accession of Queen Victoria : – Great Britain – Alexandrina Victoria, born May 24, 1819. Holland – William Frederick, born August 24, 1772. Belgium – Leopold I., born December 16, 1790. Austria – Ferdinand I., born April 19...
- applause and shouts of " Long live Queen Victoria ! " Joined by the Lord Mayor and the City authorities, the procession halted again at the end of Wood Street to make the proclamation a fourth time. Here the procession was fur ther swelled by several of the City companies ; and it then moved on to the...

# Word Embeddings

- **Word embeddings** refer to a set of AI algorithms which "map" all the words in a corpus vocabulary to lists of numbers.
- In this new representation, words which frequently appear together in the original texts will appear to be **similar** to one another, while words which do not frequently appear together will be **dissimilar**.



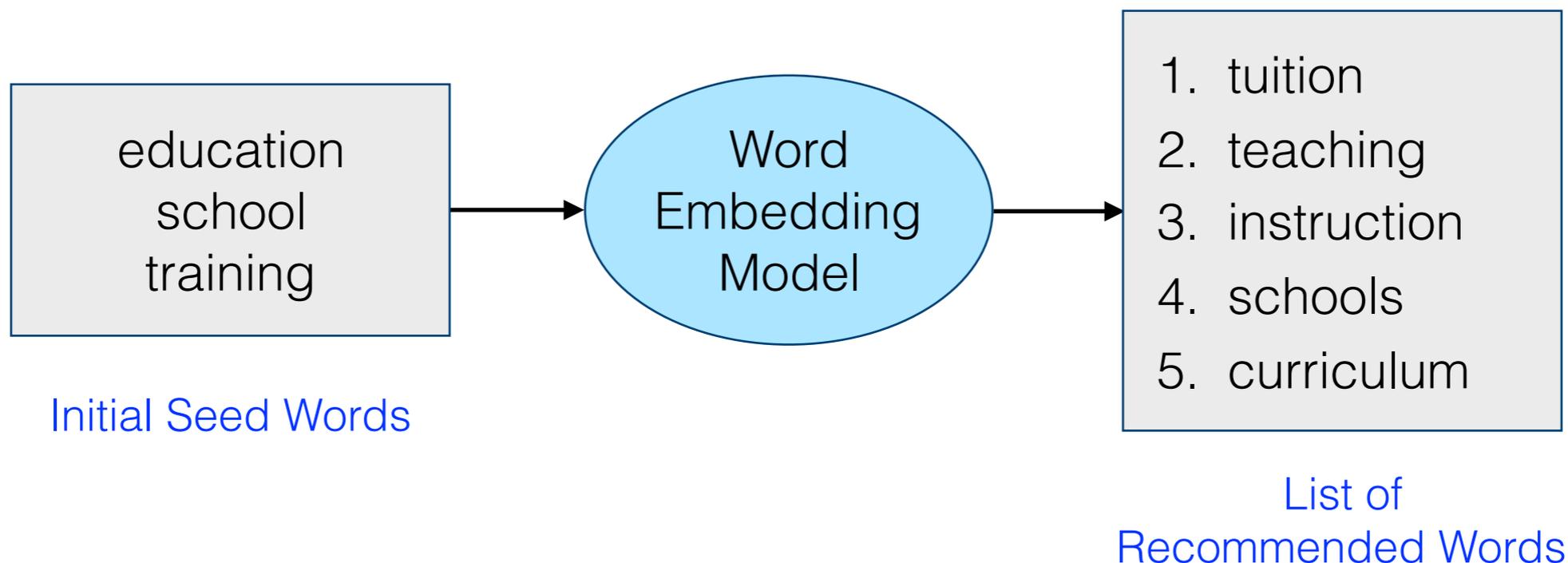
- We can create a word embedding model from the entire English-language British library corpus to "learn" which words are similar to one another.

# Word Embeddings

Word	Top 10 Most Similar Words in Embedding
contagion	infection, contagious, disease, infected, contamination, distemper, pestilence, miasma, infectious, pestilential
famine	pestilence, famines, pestdence, starvation, pestilences, dearth, depopulation, decimating, desolating, decimated
wife	husband, daughter, widow, sister, mother, father, niece, married, avife, widower
husband	wife, mother, father, sister, niece, stepmother, daughter, avife, cousin, child
education	instruction, tuition, training, educated, educating, educations, teaching, educate, schools, schooling
god	heaven, almighty, jehovah, redeemer, creator, mercy, righteous, goodness, divine, righteousness
poverty	penury, indigence, beggary, wretchedness, destitution, misery, squalor, affluence, degradation, avretchedness
dublin	belfast, dubun, cork, londonderry, galway, waterford, limerick, athlone, clonmel, kilkenny

# Expanding Word Lexicons

- We use word embeddings to reduce the effort required to create word lexicons, while keeping the scholar "in the loop".
- Once a lexicon has been created, the word list can be expanded based on recommendations automatically surfaced from a word embedding built on the British Library corpus.
- Recommendations are based on finding the most similar words to all existing words in a lexicon.



# Expanding Word Lexicons

- On the *Curatr* platform, this recommendation and expansion process is accessed via the *Lexicon* page.

Word Lexicons:

Word lexicons, which consist of curated lists of keywords related to a given topic, allow scholars to further search the British Library corpus, and to define sub-corpora related to that topic. A lexicon can be expanded based on automatically recommended keywords, which frequently appear in the same context as existing words in the lexicon.

**Edit Word Lexicon - 'Education':**

There are currently 3 keywords in this lexicon:

education	✕	school	✕	training	✕
-----------	---	--------	---	----------	---

The following additional words were recommended based on word similarities in the corpus:

Recommended Word	Explanation	Accept	Reject
<i>tuition</i>	Similar to existing lexicon words <i>education</i> and <i>training</i>	✓	✕
<i>teaching</i>	Similar to existing lexicon words <i>education</i> and <i>training</i>	✓	✕
<i>instruction</i>	Similar to existing lexicon words <i>education</i> and <i>training</i>	✓	✕
<i>schools</i>	Similar to existing lexicon word <i>school</i>	✓	✕
<i>educational</i>	Similar to existing lexicon words <i>education</i> and <i>training</i>	✓	✕
<i>educations</i>	Similar to existing lexicon words <i>education</i> and <i>training</i>	✓	✕
<i>attainments</i>	Similar to existing lexicon words <i>education</i> and <i>training</i>	✓	✕
<i>curriculum</i>	Similar to existing lexicon words <i>education</i> and <i>school</i>	✓	✕
<i>educated</i>	Similar to existing lexicon word <i>education</i>	✓	✕

# Expanding Word Lexicons

- On the *Curatr* platform, this recommendation and expansion process is accessed via the *Lexicon* page.

Word Lexicons:

Word lexicons, which consist of curated lists of keywords related to a given topic, allow scholars to further search the British Library corpus, and to define sub-corpora related to that topic. A lexicon can be expanded based on automatically recommended keywords, which frequently appear in the same context as existing words in the lexicon.

**Edit Word Lexicon - 'Disease':**

There are currently 7 keywords in this lexicon:

cancer	✕	diphtheria	✕	fever	✕	miasma	✕	pestilence	✕
pneumonia	✕	pox	✕						

The following additional words were recommended based on word similarities in the embedding model *British Library Fiction*:

Recommended Word	Explanation	Accept	Reject
<i>typhus</i>	Similar to existing lexicon words <i>diphtheria</i> and <i>fever</i>	✓	✕
<i>typhoid</i>	Similar to existing lexicon words <i>diphtheria</i> , <i>fever</i> , <i>pneumonia</i>	✓	✕
<i>scarlet fever</i>	Similar to existing lexicon words <i>diphtheria</i> and <i>pox</i>	✓	✕
<i>brain fever</i>	Similar to existing lexicon words <i>diphtheria</i> and <i>fever</i>	✓	✕
<i>bronchitis</i>	Similar to existing lexicon words <i>diphtheria</i> and <i>pneumonia</i>	✓	✕
<i>cholera</i>	Similar to existing lexicon words <i>diphtheria</i> , <i>fever</i> , <i>pestilence</i>	✓	✕
<i>influenza</i>	Similar to existing lexicon words <i>fever</i> and <i>pneumonia</i>	✓	✕

# Searching with Word Lexicons

- Once a lexicon has been created, we can use it to directly search the corpus. The words in the lexicon provide the search query.

Search Results

31,677 matching volumes were found for the lexicon **Disease**

  
EXPORT

**San Francisco Municipal Reports for the fiscal year 1861-2 -1866-7, 1868-9-1870-71 – Volume 5**  
Author Unknown – 1863 – San Francisco

- of Decedents 190-193 Recapitulation 134 Remarks— Small-Pox 194-202 Small-Pox Hospital Statistics 202-206 Remarks— Disinfection 207-209 Vaccination 209-212 Still Born Infants 212-214 Phthisis— Mortality among Children 214-217 Sewerage, etc 217-221 Report of Health Inspector 221 HOSPITAL REPORT 222...
- Annual Licenses 363 POUND MASTER'S REPORT 447-448 Receipts 44? Dogs impounded, redeemed or killed, etc 448 PUBLIC ADMINISTRATOR'S REPORT 103-113 Estates Administered Upon 103-1 12 Effects deposited, of deceased Persons from Small-Pox Hospital. . 112-113 MUNICIPAL REPORTS. VI PAGE. PUBLIC SCHOOL REPORT..
- 2,400 00 Burials and Chemical Analyses- Burials 3,482 25 Chemical Analyses 550 00 4,032 25 Small Pox Expenses — Labor and Material for Small Pox Hospital 20,828 29 Employees at Small Pox Hos- pital 13,564 25 Extra Employees at Health Of- fice, and Vaccinators 14,460 56 Feed, Fuel and Lights 15,935 33...

**The ( twenty-second-twenty-eighth-thirty-first) Annual Report of the receipts and expenditures of the Town of Dorchester, etc**  
Author Unknown – 1860 – Boston

- all money received from all sources during that time. SCHOOLS. The appropriation for this department, with the income from the School Fund, has been sufficient to meet its ex penses, and leave a balance in the Treasury. The community have been somewhat startled by the spread of the small **pox** and...
- varioid, in the neighboring Cities and vicinity, and as several cases have occurred in this town, we have requested the Physicians to vaccinate all the children attending the public schools, who had not been recently vaccinated, with the tune **pox**. POOR IN THE ALMS-HOUSE. The appropriation for this...

**Argenina and the Argentines. Notes and impressions of a five years' sojourn in the Argentine Republic, 1885-90**  
Thomas A. Turner – 1892 – London

- XVIII. Argentine Politics. — A **Cancer** in the National Organism. — The Constitution Shaped from Many Models. — Government of the Provinces. — The Chambers. — The Executive. — Political Freedom a Myth. — Provincial and Federal Antagonism. — How the System Works. — Political Sects. — Literary Legislators...

**Thirty-five years in the East. Adventures, discoveries, experiments, and historical sketches, relating to the Punjab and Cashmere**  
John Martin Honigberger – 1852 – London

- pills, I found, as I expected, violent pur gatives, which may, however, be employed with ad vantage (?) by a judicious physician. I need scarcely observe, that they do not deserve the name of pana cea; neither can I advise any one, to take either of them in the beginning of a violent **fever**, having wit...
- nessed bad consequences from so doing. I have ad ministered the above-mentioned pills, in small doses ; also Warburg's **fever** drops, which are reputed good ; and the reader may find a description of their effects and composition in the second volume. I was pleased to see in a Report in the Bengal...
- considera- tion, my management was peculiarly fortunate ; for, during a period of two years from the foundation of that hospital, not a single patient died of an acute disease — such as inflammation, **fever**, cholera, serpent bites, &c. of which I had many cases ; some of them so dangerous, that I had to...

# Exporting Curated Sub-Corpora

- After a search has been performed, we can export some or all of the corresponding results for further study and close reading.

## Search Results

31,677 matching volumes were found for the lexicon **Disease**



**San Francisco Municipal Reports for the fiscal year 1861-2 -1866-7, 1868-9-1870-71 – Volume 5**  
Author Unknown – 1863 – San Francisco

- of Decedents 190-193 Recapitulation 134 Remarks— Small-Pox 194-202 Small-Pox Hospital Statistics 202-206 Remarks— Disinfection 207-209 Vaccination 209-212 Still Born Infants 212-214 Phthisis— Mortality among Children 214-217 Sewerage, etc 217-221 Report of Health Inspector 221 HOSPITAL REPORT 222...
- Annual Licenses 363 POUND MASTER'S REPORT 447-448 Receipts 44? Dogs impounded, redeemed or killed, etc 448 PUBLIC ADMINISTRATOR'S REPORT 103-113 Estates Administered Upon 103-1 12 Effects deposited, of deceased Persons from Small-Pox Hospital. . 112-113 MUNICIPAL REPORTS. VI PAGE. PUBLIC SCHOOL REPORT..
- 2,400 00 Burials and Chemical Analyses- Burials 3,482 25 Chemical Analyses 550 00 4,032 25 Small Pox Expenses — Labor and Material for Small Pox Hospital 20,828 29 Employees at Small Pox Hos- pital 13,564 25 Extra Employees at Health Of- fice, and Vaccinators 14,460 56 Feed, Fuel and Lights 15,935 33...

**The ( twenty-second-twenty-eighth-thirty-first) Annual Report of the receipts and expenditures of the Town of Dorchester, etc**  
Author Unknown – 1860 – Boston

- all money received from all sources during that time. SCHOOLS. The appropriation for this department, with the income from the School Fund, has been sufficient to meet its ex penses, and leave a balance in the Treasury. The community have been somewhat startled by the spread of the small pox and...
- varioloid, in the neighboring Cities and vicinity, and as several cases have occurred in this town, we have requested the Physicians to vaccinate all the children attending the public schools, who had not been recently vaccinated, with the tune pox. POOR IN THE ALMS-HOUSE. The appropriation for this...

**Argenina and the Argentines. Notes and impressions of a five years' sojourn in the Argentine Republic, 1885-90**  
Thomas A. Turner – 1892 – London

- XVIII. Argentine Politics. — A Cancer in the National Organism. — The Constitution Shaped from Many Models. — Government of the Provinces. — The Chambers. — The Executive. — Political Freedom a Myth. — Provincial and Federal Antagonism. — How the System Works. — Political Sects. — Literary Legislators...

**Thirty-five years in the East. Adventures, discoveries, experiments, and historical sketches, relating to the Punjab and Cashmere**  
John Martin Honigberger – 1852 – London

- pills, I found, as I expected, violent pur gatives, which may, however, be employed with ad vantage (?) by a judicious physician. I need scarcely observe, that they do not deserve the name of pana cea; neither can I advise any one, to take either of them in the beginning of a violent fever, having wit...
- nessed bad consequences from so doing. I have ad ministered the above-mentioned pills, in small doses ; also Warburg's fever drops, which are reputed good ; and the reader may find a description of their effects and composition in the second volume. I was pleased to see in a Report in the Bengal...
- considera- tion, my management was peculiarly fortunate ; for, during a period of two years from the foundation of that hospital, not a single patient died of an acute disease — such as inflammation, fever, cholera, serpent bites, &c. of which I had many cases ; some of them so dangerous, that I had to...

# Exporting Curated Sub-Corpora

- Exported sub-corpora can subsequently be downloaded as individual archives for further study and close reading.

## Sub-Corpora

The *Curatr* platform provides access to the *British Library Digital Corpus*, a collection of digitised texts, both fiction and non-fiction, largely from the 19th century. We work with the plain text version of the corpus, transformed using Optical Character Recognition (OCR) techniques. In total there are 49,509 unique texts across multiple languages. On this platform we focus on a subset of 35,918 English language texts dating from 1700 to 1899, for which key associated metadata is available. A subset of these are multi-volume texts, yielding a total of 46,438 English volumes.

### Existing Sub-Corpora

Smaller, topic-specific sub-corpora of the British Library Digital Corpus can be exported using the [Curatr search functionality](#). Currently, the following curated sub-corpora are available to download for further analysis and close reading:

Sub-Corpus Name	# Documents	Format	Sub-Corpus Summary	Download
<i>Biology</i>	55	Full Text	<ul style="list-style-type: none"><li>Date Range: All years</li><li>Query: *</li><li>Document Type: Volumes</li><li>Search Field: Full Text</li><li>Alston Classification: Biology</li></ul>	
<i>Contagion</i>	100	Full Text	<ul style="list-style-type: none"><li>Date Range: All years</li><li>Query: contagion contagious epidemic infection infectious plague</li><li>Document Type: Volumes</li><li>Search Field: Full Text</li><li>Alston Classification: All</li><li>Lexicon: Contagion</li></ul>	
<i>Vaccination</i>	283	Full Text	<ul style="list-style-type: none"><li>Date Range: From 1800 until 1830</li><li>Query: vaccine vaccination</li><li>Document Type: Volumes</li><li>Search Field: Full Text</li><li>Alston Classification: All</li></ul>	

# Examining Historical Trends

- We can study temporal trends in a large corpus by looking at changes in word frequencies over time (e.g. *Google Ngram Viewer*).
- We created a similar British Library Corpus Ngram Viewer tool as part of the *Curatr* platform.

## British Library Corpus Ngram Viewer

The form below can be used to search the frequency of all unique words in English language volumes from the British Library corpus texts. The plot shows the number of volumes containing the query keyword(s) in each year. Clicking on a point on the plot will show the corresponding set of search results. Also, click and drag the rectangle in the plot area to zoom in on a specific time range.

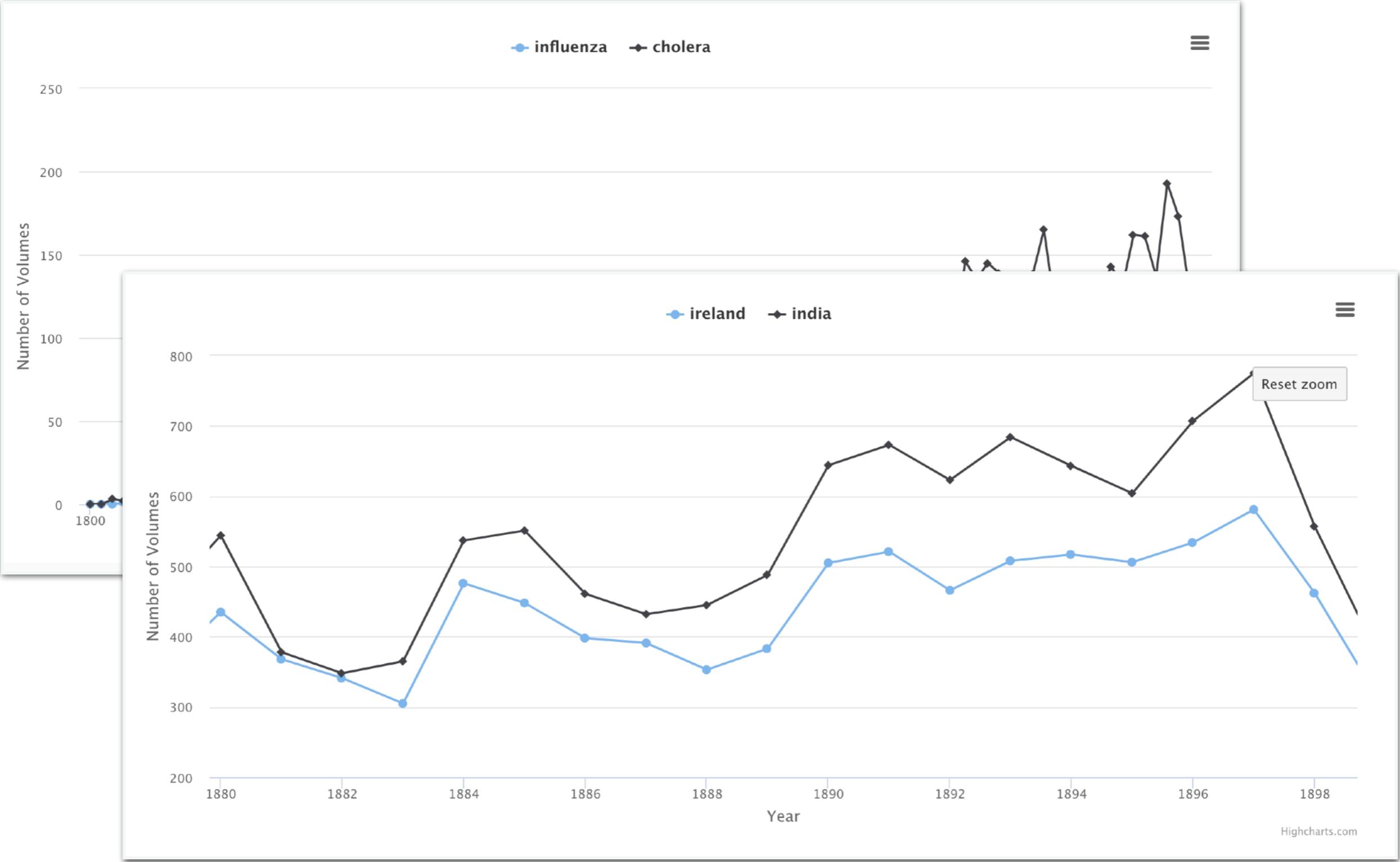
Enter one or more keywords (comma separated)

influenza, cholera

Start Year	End Year
1800	1899

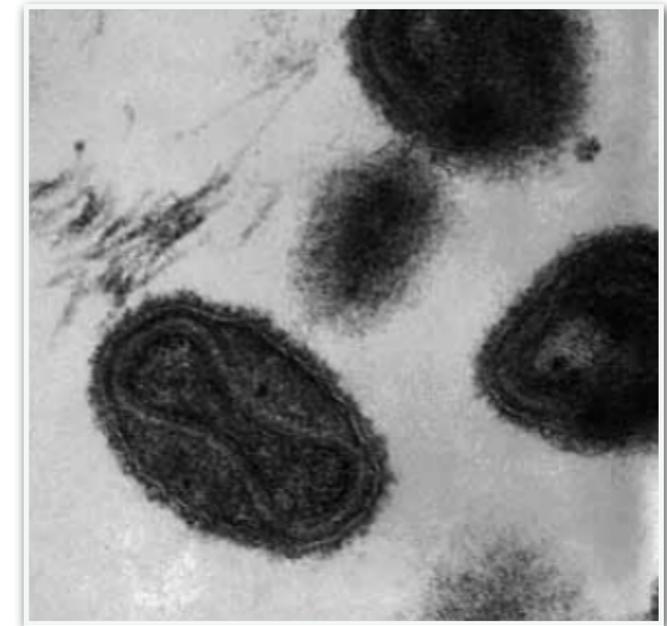
SEARCH

# Examining Historical Trends



# Case Study: Smallpox & Vaccination

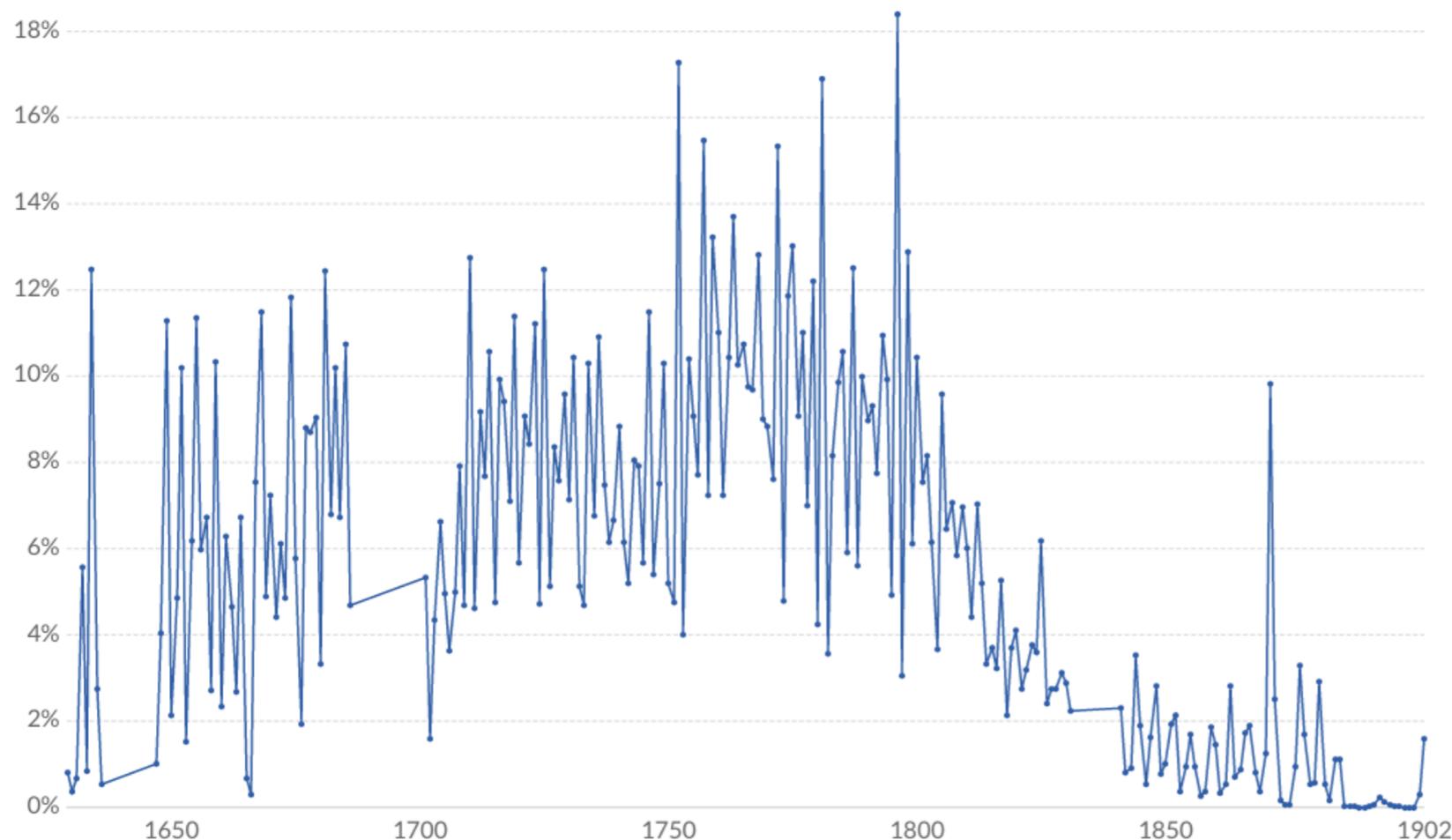
- Smallpox is caused by the *variola* virus, passed by coming into contact with droplets of a smallpox-infected patient.
- With no means of treatment, it was probably the single most lethal disease in 18th century Britain.



CDC

Deaths caused by smallpox as % of all deaths in London

In London in the 18th century, 1 in every 13 deaths was due to smallpox, the majority being young children.



# Case Study: Smallpox & Vaccination

- Edward Jenner, an English physician and scientist, was the pioneer of the smallpox vaccination, the world's first vaccine.
- Jenner proposed that the blisters which milkmaids received from cowpox protected against subsequent infection with smallpox.
- He successfully tested this theory in 1796, and by 1800 the use of vaccination spread rapidly in Britain.
- Jenner "may have saved more lives than any other person before or since".



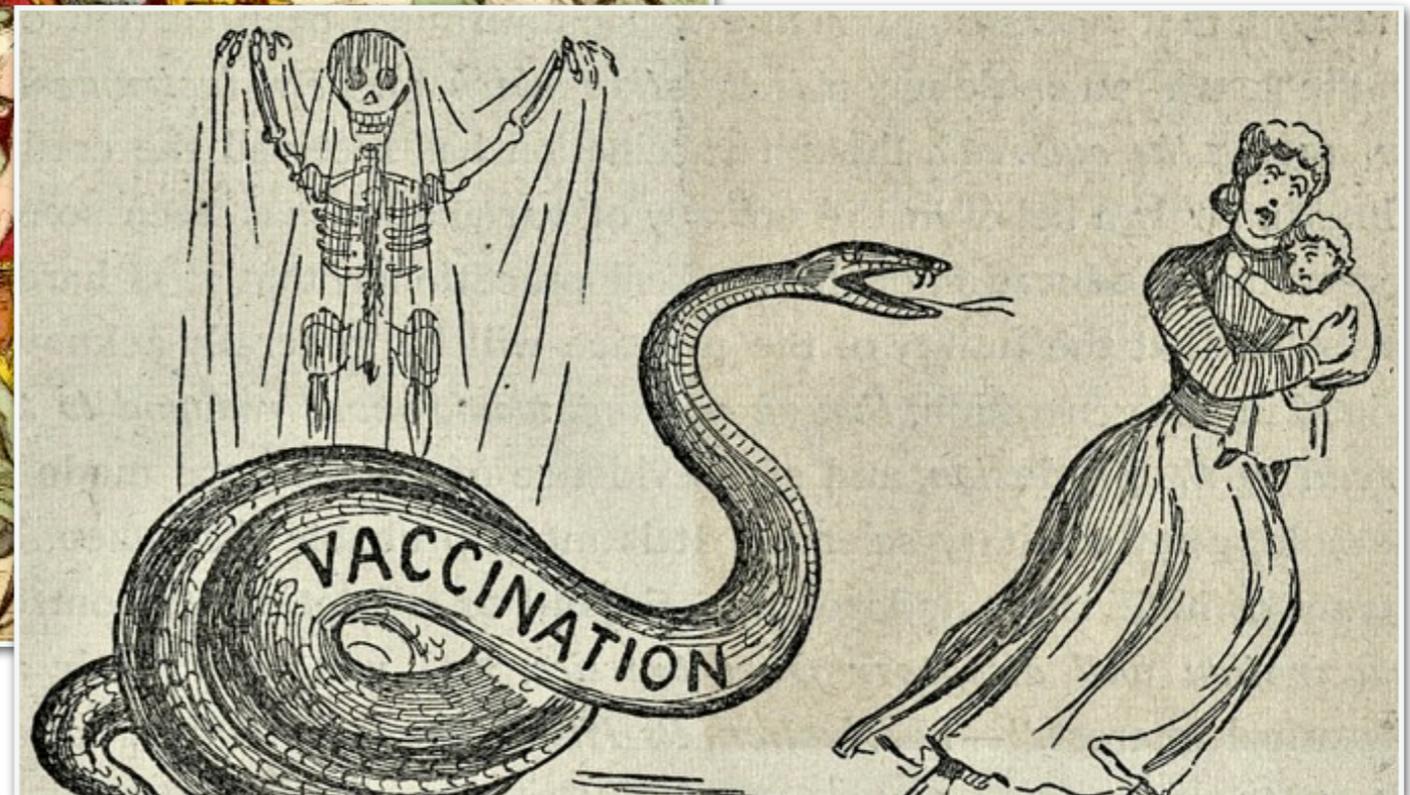
Edward Jenner (1749–1823)  
National Library of Medicine

# Case Study: Smallpox & Vaccination

- The use of vaccination lead to considerable public debate and the emergence of anti-vaccination movements in the 19th century.

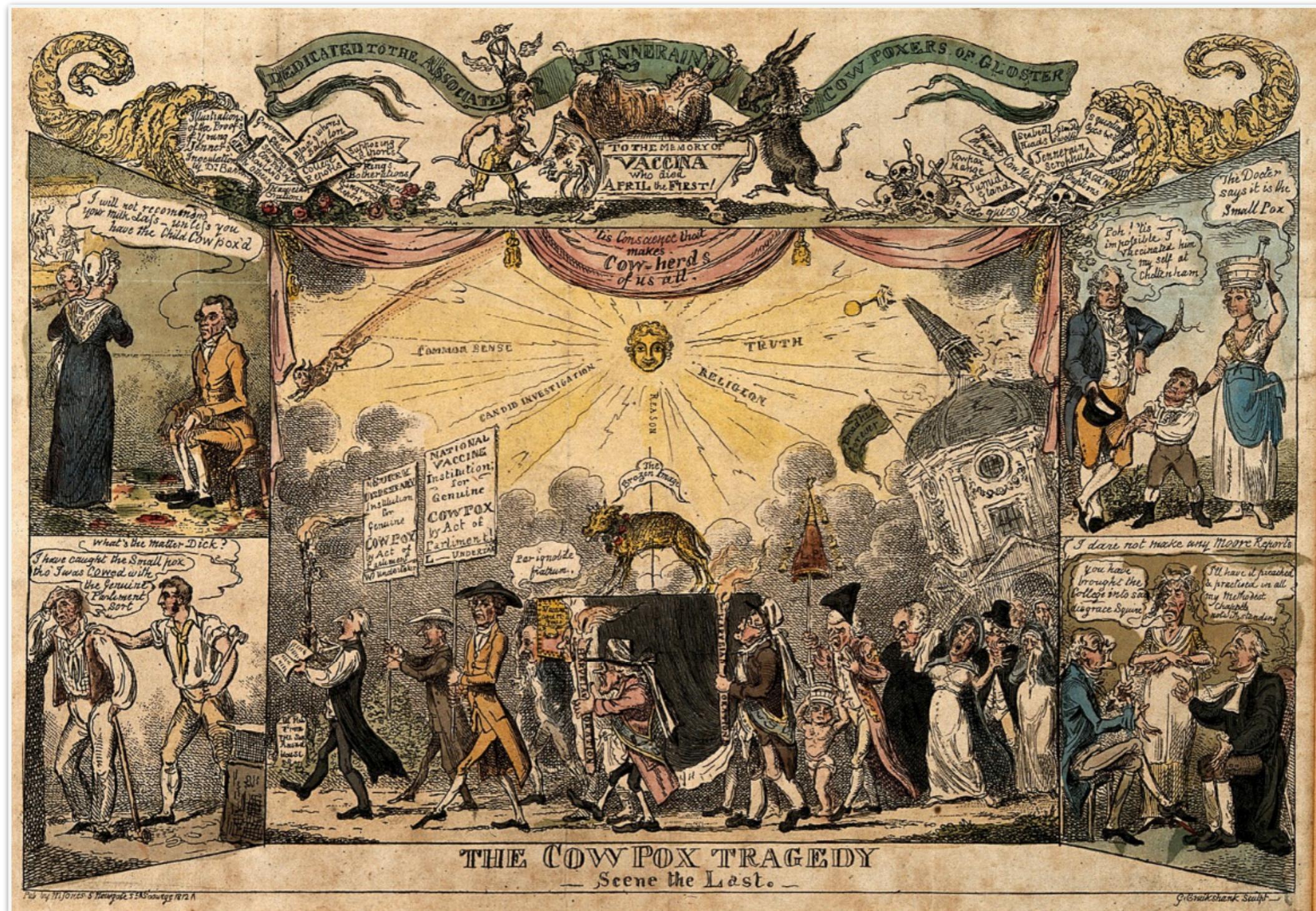


Publications of ye Anti-Vaccine Society (1802)



Library Of The College Of Physicians Of Philadelphia (1894)

# Case Study: Smallpox & Vaccination



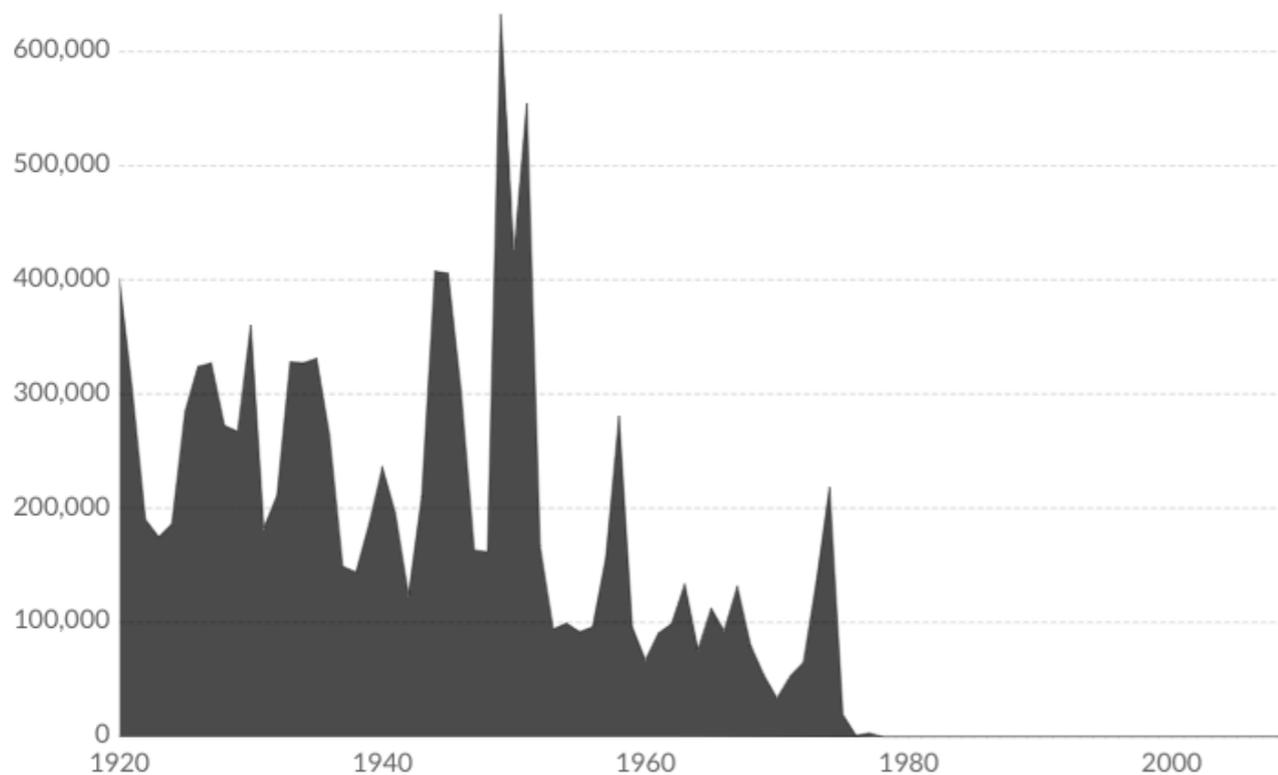
An attack on smallpox vaccination and on the Royal College of Physicians' advocacy of it.

G. Cruikshank (1812)

# Case Study: Smallpox & Vaccination

- Thanks to the roll out of Jenner's vaccine across Europe and North America, smallpox was almost entirely eliminated in these regions by the first half of the 20th century.
- In 1966 the WHO launched an intensified global plan to eradicate smallpox. By 1980 the disease was declared as eradicated.

Global number of reported smallpox cases (1920-)



Source: World Health Organization (2011)

SCIENTIFIC AMERICAN

nature

HEALTH

**Vials of Smallpox Virus Found Unsecured at NIH**

Sixty-year-old ampoule contains smallpox DNA, and it is unclear whether the virus is viable.

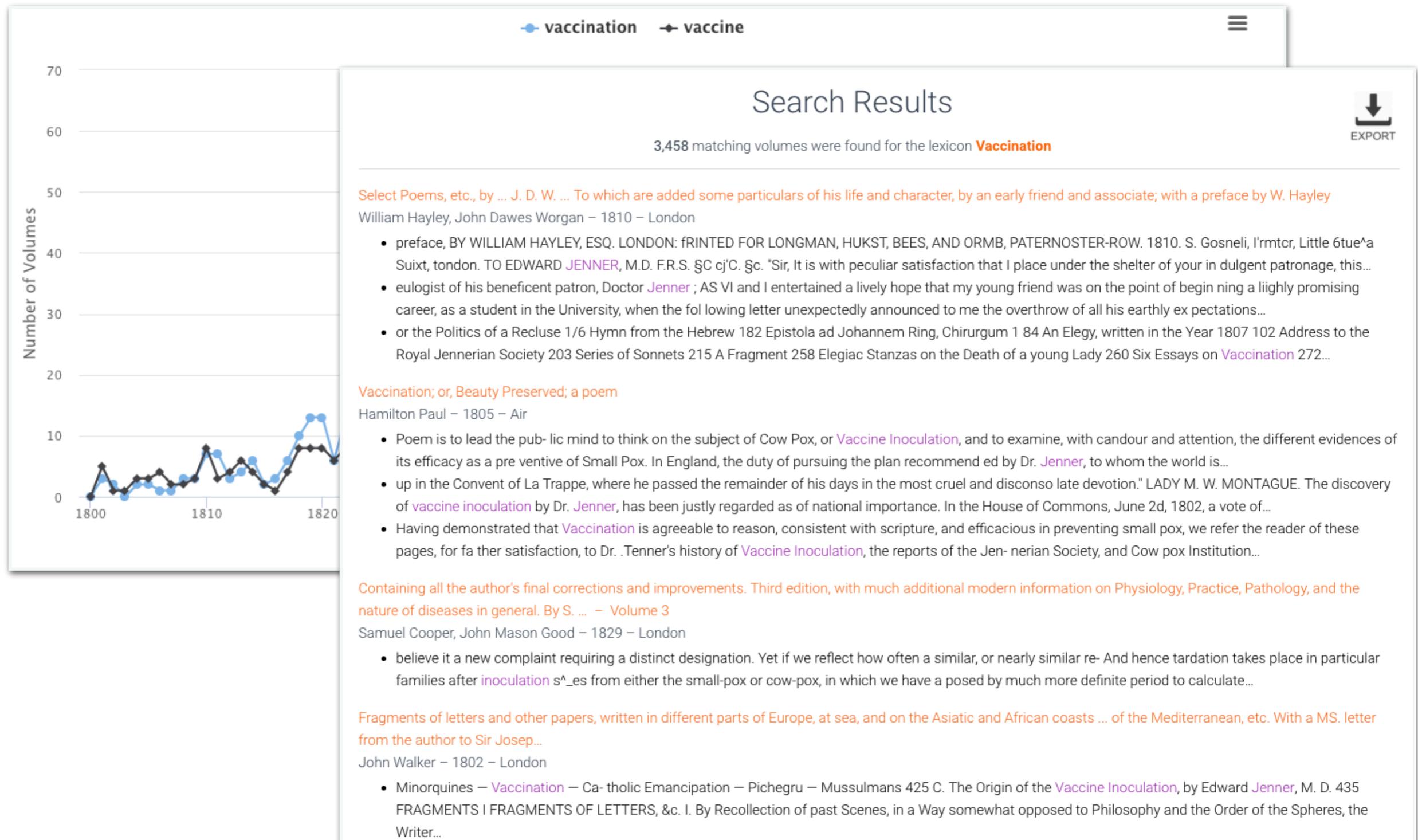
By Sara Reardon, Nature magazine on July 9, 2014

Smallpox, officially preserved in two repositories worldwide, may have been sitting alive and well in an unsecured US government refrigerator. On July 8, the US Centers for Disease Control and Prevention (CDC) announced that vials containing the deadly virus had been discovered in a cardboard box in the refrigerator, located on the National Institutes of Health (NIH) campus in Bethesda, Maryland.

scientificamerican.com

# Case Study: Smallpox & Vaccination

- Over 3,000 volumes in the corpus refer to vaccination in some way.



# Case Study: Smallpox & Vaccination

- On closer inspection, the majority of texts appear to be either "neutral" or pro-vaccination.

...n ten. For the attempts to alleviate this scourge, see Inoculation, introduced into England in 1722, and Vaccination, announced by Dr. Jenner in 1798. **Smallpox** Hospital, established 1746. **Smallpox** raged in parts of London, and thousands died, 1870-1 ; a temporary hospital was established at Hampstead (which see). The Anti- Vaccination society has been active, and many parents have been fined for opposing the vaccination of their children, 1870-6. In Sept. and Oct. 1862, a great many sheep died of **smallpox** in the West of England, till successful preventive measures were resorted to. Many cases in London, 1876-8 ; deaths principally of an vaccinated persons. **Smallpox** prevalent in London, 88 deaths 1 – 7 May; 103 deaths 15 – 21 May; diminishing July, 1881. Deaths, June, 1880 – June, 1881, 1532 deaths, 637 not Vaccinated ; in N.E. London, May, deaths, about 36 a week, June, 1884 ; deaths decreasing reported, 24 July ; reappears, but subsides, Dec. 1884. Severe epidemic of **smallpox** at Sheffield and neighbourhood, March, 1887 ; still prevalent, Jan. 1888 ; disappearing, April, 1888. Severe at Manchester and Leicester, about 16 Jan., 1893; at Gloucester, 1896. SLEEP, see Seven Sleepers. M. Chauffat, awoke after a nineteen days' sleep at the Alexandra Palace, 28 Nov. 1888 ; many similar cases are reported. Johann Latus, at Myslowitz, in Silesia, is said to have slept four and a-half months, being kept alive by imbibing milk, reported 31 Jan. ; he awoke about 9 Feb., 1892. The case was ascribed to catalepsy. SLESWIG, see Holstein. SLIDING-SCALE, see Corn Laws. SLING- In Judges xx. 16, is mentioned the skill of the Benjamite slingers (about 1406 B.C.), and with a sling David slew Goliath 1063 B.C. (1 Sam. xvii.) The natives of the Balearic isles (Majorca, Minorca, and Ivi\*a) were celebrated slingers, and served as mercenaries in the Carthaginian and Roman armies. Slings are said to have been used by the Huguenots at the siege of San eerre, in 1672, to economise thei...

Haydn's Dictionary of Dates, 22nd Ed.  
Joseph Haydn (1898)

# Case Study: Smallpox & Vaccination

- On closer inspection, the majority of texts appear to be either "neutral" or pro-vaccination.

Vaccination met with great and powerful support from the first. It was taken up voluntarily by the better classes in England, and its practice was widely extended amongst the poor by the labours of an enthusiastic band of workers, of whom John Ring and George Pearson were the foremost. The advance of the method was also furthered by the remarkable diminution in the number of cases of **smallpox** which followed the great epidemics in 1796, 1798, and 1800. This diminution was in part to be explained by the increasing number of persons who were vaccinated, but in part, no doubt, it was due to the enormous number of persons who had been protected by taking the disease itself during these epidemics, for one attack of small pox usually renders a person immune more effectually than the most perfect vaccination with cowpox. **Smallpox** re asserted itself in the course of a few years with its original

REVOLUTION AND REACTION. 429 802] virulence almost unimpaired by vaccination. It was then ascertained that an attack of cowpox only conferred temporary immunity from **smallpox**, and that it was necessary to repeat the operation of vaccination if the fullest amount of protection was to be obtained. From its very commencement vaccination has been surrounded by the most bitter controversy. It has been attacked upon scientific grounds ; it has been attacked from the lowest popular standpoint; statistics have been employed to demonstrate that it is useless. There remain, however, the stubborn facts that since it has been in general use the deaths from **smallpox** have sunk until they have almost reached a vanishing point in vaccinated, and especially in re-vaccinated, communities, whilst amongst the unvaccinated the disease still flourishes luxuriantly unless the most careful methods of isolating not only the sick, but also the members of invalid households, are enforced at an outlay which is economically prohibitive

Social England. A record of the progress of the people  
Henry Duff Traill (1894)

# Case Study: Smallpox & Vaccination

- In the *fiction* category, we see Victorian moral fables used as a way of inculcating both good public health practice (vaccination) and social responsibility (take care of a dying child).

...vaccination had made it less fatal among the upper classes, this frightful scourge still decimated the poor, especially children. Great was the obstinacy in refusing relief; and loud the outcry in Norton Bury, when Mr. Halifax, who had met and known Dr. Jenner in London — finding no practitioner who would do it, persisted in administering the vaccine virus himself to his own children. But still, with a natural fear, he had kept them out of all risk of taking the **smallpox** until now. " John, do you think " " No ; I will not allow myself to think. Not a word of this at home, mind. Good-bye ! "

JOHN HALIFAX, GENTLEMAN 309 He walked away, and I returned up the path heavily, as if a cloud of terror and dole were visibly hanging over our happy Longfield. The doctor appeared ; he went up to the sickly lad ; then he and Mr. Halifax were closeted together for a long time. After he was gone, John came into the kitchen, where Ursula sat with Walter on her knee. The child was in his little white night-gown, playing with his elder brothers, and warming his rosy toes. The mother had recovered herself entirely; was content and gay. I saw John's glance at her, and then — then I feared. " What does the doctor say ? The child will soon be well ? " " We must hope so." " John, what do you mean ? I thought the little fellow looked better when I went up to see him last. And there — I hear the poor mother upstairs crying." " She may cry; she has need," said John bitterly. " She knew it all the while. She never thought of our children ; but they are safe. Be content, love — please God, they are quite safe. Very few take it after vaccination." " It — do you mean the **smallpox** ? Has the lad got **smallpox** ? O God, help us ! My children — my children ! " She grew white as death ; long shivers came

John Halifax, Gentleman  
Dinah Maria Mulock Craik (1898)

# Case Study: Smallpox & Vaccination

- However, a smaller number of texts express anti-vaccination views, questioning the safety of the procedures, using words such as ‘pollution’, ‘infect’, ‘poisons’, and ‘horror’.

...f prevention then believed in, as the patient generally suffered but slightly from the inoculation, and it was then a creed, long since exploded, that small-pox could not be taken tAvice. This fact of their resistance to variolous inoculation set him thinking, and he came to the conclusion that they had absorbed into their systems, a counter poison in the shape of some infection taken from the coavs. He made many experiments, and found that this came from a disease called the cow pock, and that the vaccine lymph could not only be taken direct from the coav, but also by transmission from the patients Avho had been inoculated Avith that lymph, and Avhence the present system of so-called vacci nation – the greatest blessing of modern times. Jenner, of course, Avas opposed ; fools do not even believe in vaccination noAv, and great Avas the battle for, and against, in the medical profession, and many Avere the books Avritten pro and con. " Vaccination Vindicated," Ed. Jones ; "A Reply to the Anti-Vaccinists," Jas. Moore; "The Vaccine Contest," Wm. Blair; "Cow Pock Vacci nation," Rowland Hill ; " Birch against Vaccination" ; "Willer on Vaccination" &c. &c. Gilbey could not, of course, leave such a promising subject alone, and he perpetrated the accompanying

illus tration.  
being borne  
too well.

physicians were admitted to practise in many other countries besides their own. The enemies of Dr. Taylor met upon these grounds, and commenced shortly afterwards a war of persecution against the more prominent anti-vaccinators. Our friend replied vigorously to the assault with his pen, and afterwards in the meetings of the Colonial Council. Here he was as resolute as he was indefatigable. It was a grave question of public hygiene and of individual liberty. The vaccination law obliging parents who refused to poison their children with the virus which the vaccinators transmit from arm to arm, with the possibility of transmitting a host of maladies as well as syphilis, to pay a heavy fine, or, in default, to be levied upon or imprisoned. Dr. Taylor, who is actually paying at this moment six dollars fine every month to escape the action of this fatal law upon his children, remained valiantly at his post. His last work in this direction has been the presentation to the Colonial Council of a petition signed by the principal inhabitants of the island, praying the King of Denmark for an extenuation of the fines of those who have no longer faith in the protective power of vaccination. Unfortunately, it is the same in medical questions as it is in religious – Might makes Eight, and the strongest wins. Nevertheless, Dr. Taylor succeeded, by his persuasive eloquence, in

Leaflets from the Danish West Indies.  
P. Linet (1888)

# Conclusions

- The *Curatr* platform allows scholars to explore, search, and curate the British Library Corpus via an online interface. Full volumes and short text extracts are also available for close reading.
- Our work suggests that a significant opportunity exists to apply methods from data science in a range of other disciplines, where rich interesting datasets are available.
- In some cases, appropriate tools are already freely available "off the shelf". In other cases, the use of customised data analysis techniques may be required to solve the problem.
- Often the most significant challenge involves fully understanding the requirements of the researcher in the field.
- In all cases, the input of experts in the field is crucial - we need strong interdisciplinary teams with complementary skillsets.

# Acknowledgments

This project is supported by the **Irish Research Council** and the **SFI Insight Centre for Data Analytics** at UCD.

- Prof. Gerardine Meaney, UCD School of English, Drama and Film
- Dr. Derek Greene, UCD School of Computer Science
- Dr. Susan Leavy, UCD School of Computer Science
- Dr. Karen Wade, UCD School of English, Drama and Film
- Dr. Cormac O'Brien, UCD School of English, Drama and Film
- Katherine Campbell, Insight Centre @ UCD
- Mahendra Mahey, British Library Labs

