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
“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
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)
• 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)
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.

lish the money they paid to the Italians for silk. He also tried to raise many valuable tropical plants. There was hardly any good thing which needed doing that A GEORGIA ROAD was undertaken by the new colony. Such an enterprise! appealed strongly to the benevolent, and many thousands of pounds were given to help on this good work. Parliament also voted a donation to Georgia. No one was allowed to make any profit out of the new colony, on the seal of which was a device of silk-worms spinning, with a motto in Latin which meant, "Not for themselves, but for others" (Non sibi sed aliis). In 1732 Oglethorpe took out his first company of a hundred and sixteen people, with whom he began the town of Savannah. Many others were added, among whom were a regiment of Scotch Highlanders.
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.

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>[Mansfield Park: a novel.], Single Works</td>
</tr>
<tr>
<td>Authors</td>
<td>{'creator': ['Austen, Jane']}</td>
</tr>
<tr>
<td>Year</td>
<td>1875</td>
</tr>
<tr>
<td>Issuance</td>
<td>Monographic</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
</tr>
<tr>
<td>Place</td>
<td>London</td>
</tr>
<tr>
<td>Publisher</td>
<td>Groombridge &amp; Sons</td>
</tr>
<tr>
<td>Shelfmark</td>
<td>British Library HMNTS 12618.h.4.</td>
</tr>
<tr>
<td>Shelfmark Number</td>
<td>12618</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Silas Marner ... By the author of “Adam Bede”, etc, Silas Marner</td>
</tr>
<tr>
<td>Authors</td>
<td>{'creator': ['Eliot, George']}</td>
</tr>
<tr>
<td>Year</td>
<td>1863</td>
</tr>
<tr>
<td>Edition</td>
<td></td>
</tr>
<tr>
<td>Issuance</td>
<td>Monographic</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
</tr>
<tr>
<td>Place</td>
<td>S. H. Goetzel</td>
</tr>
<tr>
<td>Publisher</td>
<td>Mobile</td>
</tr>
<tr>
<td>Shelfmark</td>
<td>British Library HMNTS 12623.g.27.</td>
</tr>
<tr>
<td>Shelfmark Number</td>
<td>12623</td>
</tr>
</tbody>
</table>
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.
Number of Volumes by Year in English Language Corpus (19th Century)

Publication Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1815</td>
<td>Battle of Waterloo</td>
</tr>
<tr>
<td>1837</td>
<td>Coronation of Queen Victoria</td>
</tr>
<tr>
<td>1845</td>
<td>The Great Famine</td>
</tr>
<tr>
<td>1859</td>
<td>Darwin's Origin of Species</td>
</tr>
<tr>
<td>1861</td>
<td>American Civil War</td>
</tr>
<tr>
<td>1885</td>
<td>First Automobile</td>
</tr>
<tr>
<td>1896</td>
<td>First Modern Olympics</td>
</tr>
</tbody>
</table>
Publication Locations

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

<table>
<thead>
<tr>
<th>Location</th>
<th>Number of Texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>23,853</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>1,219</td>
</tr>
<tr>
<td>New York</td>
<td>1,137</td>
</tr>
<tr>
<td>Boston</td>
<td>620</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>448</td>
</tr>
<tr>
<td>Dublin</td>
<td>383</td>
</tr>
<tr>
<td>Glasgow</td>
<td>222</td>
</tr>
<tr>
<td>Chicago</td>
<td>191</td>
</tr>
<tr>
<td>Oxford</td>
<td>187</td>
</tr>
<tr>
<td>Calcutta</td>
<td>182</td>
</tr>
<tr>
<td>Manchester</td>
<td>180</td>
</tr>
<tr>
<td>Cambridge</td>
<td>168</td>
</tr>
</tbody>
</table>
Book Titles

- The title metadata of the texts reflect the diversity of the corpus, with 29,072 unique words appearing in the titles.
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.

91.5% of volumes contain < 200k words
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.

<table>
<thead>
<tr>
<th>Search Keywords</th>
<th>Document Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>vaccination</td>
<td>Volumes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Field</th>
<th>Sub-classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Text</td>
<td>All Sub-classifications</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classification</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All Classifications</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Start Year</th>
<th>End Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1850</td>
<td>1890</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Publication Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
</tr>
</tbody>
</table>
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

Miscellanies: chiefly addresses, academical and historical (moral and religious, political and social) — Volume 5
Francis William Newman — 1869 — London
  • Science Egregious 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 Shakespeare's Richard 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 T'lee Contemporary Review, Nov. 1877. The Barbarisms of Civilization, ..... 449 From T'lee 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 (Ouvres— 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 Traverse and the Rue Tirechape— The...
  • Statistics— The Bureau of Benevolence of the First Arrondissement — 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 — Brigades of Children...

Lord Bannam. By the author of "Gin'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.

---

...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 twice. 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 cow's. He made many experiments, and found that this came from a disease called the cow pox, and that the vaccine lymph could not only be taken direct from the cow, but also by transmission from the patients Avho had been inoculated shrith 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 Awritten pro and con. "Vaccination Vindicated," Ed. Jones; "A Reply to the Anti-Vaccinists," Jas. Moore; "The Vaccine Contest," Wm. Blair; "Cow Pock Vaccination," 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 illustration. Here Dr. Jenner (a very good likeness) is attending 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 exhibiting the different phases of their vaccination. As a rule, they seem to have "taken" too well.
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").
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.

Rachel Fanny Antonina Lee – 1808 – London
Classification: Aristotle – Philosophy, particularly logic
Top 15 Top Level Classifications in English Language Corpus

- Fiction
- Topography
- Poetry
- History
- Plays
- Drama
- Geography
- Voyages & Travels
- Collected Works
- Geology
- Hydrography
- Geography & Topography
- Travels
- Geography & Ethnology
- Collections of Travels

Number of Texts

0  2000  4000  6000  8000  10000  12000
Top 20 Sub-Classifications for Top Level Classification "History"
The subject classification index can be used as filtering criteria when searching the corpus.
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.
• *Curatr* search can also act as a discovery tool for the physical collection of the British Library.

• Example: Finding lesser known poetry

---

**Curatr** search can also act as a discovery tool for the physical collection of the British Library. Example: Finding lesser known poetry
<table>
<thead>
<tr>
<th>Year</th>
<th>Title</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1700</td>
<td>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 Danier. In verse.</td>
<td>Author Unknown</td>
</tr>
<tr>
<td>1700</td>
<td>Friendship improvd, or the Female Warriour; A tragedy in five acts and in verse</td>
<td>Charles Hopkins</td>
</tr>
<tr>
<td>1700</td>
<td>Amasia, or, The works of the Muses. A collection of poems</td>
<td>John Hopkins</td>
</tr>
<tr>
<td>1700</td>
<td>A Journey to Hell; or, a visit paid to the Devil: a poem. By Edward Ward.</td>
<td>Edward Ward</td>
</tr>
<tr>
<td>1700</td>
<td>An ode performed at the anniversary feast of the Gentlemen, natives of the County of Kent, at Marchant-Taylors-Hall, Nov. 21, 1700, Appendix</td>
<td>Author Unknown</td>
</tr>
<tr>
<td>1700</td>
<td>The Generous Choice. A comedy in five acts, and in prose, with songs</td>
<td>Francis Manning</td>
</tr>
<tr>
<td>1700</td>
<td>An Elqaiack Essay upon the Decease of the Groom-Porter i.e. T. Neale, and the Lotteries</td>
<td>Thomas Neale</td>
</tr>
<tr>
<td>1700</td>
<td>Sacellium Honoris. A congratulatory poem to ... the Marquis of Tavistock, on his happy return from travel</td>
<td>Elkanah Settle</td>
</tr>
<tr>
<td>1700</td>
<td>The Fate of Capua; a tragedy in five acts and in verse</td>
<td>Thomas Southern</td>
</tr>
<tr>
<td>1700</td>
<td>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.</td>
<td>John Taylor</td>
</tr>
<tr>
<td>1700</td>
<td>The Venetian Marriage</td>
<td>Author Unknown</td>
</tr>
<tr>
<td>1700</td>
<td>The Plain-Dealer ... The fifth edition. . Single Works</td>
<td>William Wycherley</td>
</tr>
<tr>
<td>1700</td>
<td>Carmen saeculare; for the year 1700. To the King. By Matthew Prior.</td>
<td>Author Unknown</td>
</tr>
<tr>
<td>1700</td>
<td>The Perjur'd Husband: or, the Adventures of Venice. A tragedy</td>
<td>Susanna Centlivre</td>
</tr>
<tr>
<td>1701</td>
<td>The Bath; or, The Western Lass. A comedy in five acts and in prose</td>
<td>Thomas D’urfey</td>
</tr>
</tbody>
</table>
## 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.

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

<table>
<thead>
<tr>
<th>Topic</th>
<th>Lexicon Word List</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>agriculture, agriculturist, farming</td>
</tr>
<tr>
<td>Education</td>
<td>education, school, training</td>
</tr>
<tr>
<td>Smallpox</td>
<td>smallpox, pox, cowpox</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Lexicon Word List</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>agriculture, agriculturist, farming, agri, landholder, landowner, farmed, cultivators, husbandry, tillage</td>
</tr>
<tr>
<td>Education</td>
<td>education, school, training, tuition, instruction, curriculum, day-school, grammar-school, discipline</td>
</tr>
<tr>
<td>Smallpox</td>
<td>smallpox, pox, cowpox, erysipelas, epidemic, inflammation, fever, jenner, vaccination, vaccine</td>
</tr>
</tbody>
</table>
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:

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

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

**Initial Seed Words**
- education
- school
- training

**Word Embedding Model**

**List of Recommended Words**
1. tuition
2. teaching
3. instruction
4. schools
5. curriculum

**List of Recommended Words**
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:

<table>
<thead>
<tr>
<th>Recommended Word</th>
<th>Explanation</th>
<th>Accept</th>
<th>Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>tuition</td>
<td>Similar to existing lexicon words education and training</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>teaching</td>
<td>Similar to existing lexicon words education and training</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>instruction</td>
<td>Similar to existing lexicon words education and training</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>schools</td>
<td>Similar to existing lexicon word school</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>educational</td>
<td>Similar to existing lexicon words education and training</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>educations</td>
<td>Similar to existing lexicon words education and training</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>attainments</td>
<td>Similar to existing lexicon words education and training</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>curriculum</td>
<td>Similar to existing lexicon words education and school</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>educated</td>
<td>Similar to existing lexicon word education</td>
<td>✔️</td>
<td>✗</td>
</tr>
</tbody>
</table>
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:

<table>
<thead>
<tr>
<th></th>
<th>cancer</th>
<th>diphtheria</th>
<th>fever</th>
<th>miasma</th>
<th>pestilence</th>
<th>pneumonia</th>
<th>pox</th>
</tr>
</thead>
</table>

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

<table>
<thead>
<tr>
<th>Recommended Word</th>
<th>Explanation</th>
<th>Accept</th>
<th>Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>typhus</td>
<td>Similar to existing lexicon words <em>diphtheria</em> and <em>fever</em></td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>typhoid</td>
<td>Similar to existing lexicon words <em>diphtheria</em>, <em>fever</em>, <em>pneumonia</em></td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>scarlet fever</td>
<td>Similar to existing lexicon words <em>diphtheria</em> and <em>pox</em></td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>brain fever</td>
<td>Similar to existing lexicon words <em>diphtheria</em> and <em>fever</em></td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>bronchitis</td>
<td>Similar to existing lexicon words <em>diphtheria</em> and <em>pneumonia</em></td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>cholera</td>
<td>Similar to existing lexicon words <em>diphtheria</em>, <em>fever</em>, <em>pestilence</em></td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>influenza</td>
<td>Similar to existing lexicon words <em>fever</em> and <em>pneumonia</em></td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>
• 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

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
    HOSPITAL REPORT 222...
  • Annual Licenses 363 POUND MASTER'S REPORT 447-448 Receipts 447 Dogs impounded, redeemed or killed, etc 448 PUBLIC ADMINISTRATOR'S
    REPORT 103-113 Estates Administered Upon 103-112 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 Hospital 13,564 25 Extra Employees at Health Office, 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 expenses, 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 have not recently been vaccinated, with the vaccine pock. POOR IN THE ALMS-HOUSE. The appropriation for this...

Argentina and the Argentines: Notes and impressions of a five years' sojourn in the Argentine Republic; 1885-90
Thomas A. Turner — 1892 — London
  • XVIII. Argentinian Politics. — A Cancer in the National Organism. — The Constitution Shaped from Many Models. — Government of the Provinces. — The
    Legislators...

Thirty-five years in the East: Adventures, discoveries, experiments, and historical sketches, relating to the Punjab and Cashmire
John Martin Honychurch — 1852 — London
  • pills, I found, as I expected, violent purgatives, 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...
  • dressed 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 1861-2-1865-7, 1868-9-1870-71 – Volume 5
Author Unknown – 1863 – San Francisco
  • Annual Licenses 363 POUND MASTER’S REPORT 447-448 Receipts 447 Dogs impounded, redeemed or killed, etc 448 PUBLIC ADMINISTRATOR’S REPORT 103-113 Estates Administered Upon 103-112 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-ice, 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 expenses, 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 variola pox. POOR IN THE ALMS-HOUSE. The appropriation for this.

Argentina and the Argentines. Notes and impressions of a five years’ sojourn in the Argentine Republic; 1883-90
Thomas A. Turner – 1892 – London

Thirty-five years in the East. Adventures, discoveries, experiments, and historical sketches, relating to the Punjab and Cashmire
John Martin Honegger – 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:

<table>
<thead>
<tr>
<th>Sub-Corpus Name</th>
<th># Documents</th>
<th>Format</th>
<th>Sub-Corpus Summary</th>
</tr>
</thead>
</table>
| Biology         | 55          | Full Text | • Date Range: All years  
|                 |             |         | • Query: *  
|                 |             |         | • Document Type: Volumes  
|                 |             |         | • Search Field: Full Text  
|                 |             |         | • Alston Classification: Biology |
| Contagion       | 100         | Full Text | • Date Range: All years  
|                 |             |         | • Query: contagion contagious epidemic infection infectious plague  
|                 |             |         | • Document Type: Volumes  
|                 |             |         | • Search Field: Full Text  
|                 |             |         | • Alston Classification: AI  
|                 |             |         | • Lexicon: Contagion |
| Vaccination     | 283         | Full Text | • Date Range: From 1800 until 1830  
|                 |             |         | • Query: vaccine vaccination  
|                 |             |         | • Document Type: Volumes  
|                 |             |         | • Search Field: Full Text  
|                 |             |         | • Alston Classification: AI |
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](image-url)

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: 1800
End Year: 1899
Examining Historical Trends

Graph 1: Comparison of influenza and cholera trends over the years 1880 to 1898.

Graph 2: Comparison of Ireland and India's number of volumes over the years 1880 to 1898.
**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.

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.

[Graph showing deaths caused by smallpox as % of all deaths in London from 1650 to 1902]
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.
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-)**


---

**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.
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.

...In 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, 1870-71; 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 Vaccinatcd; 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 neighbour hoood, March, 1887; still prevalent, Jan. 1888; dis- appearing, 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 ... (1 Sam. xvii.) The natives of the Balearic isles (Majorca, Minorca, and Ivit*a) were celebrated slingers, and served as mercenaries in the Cartha gian arid Roman armies. Sling is 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 smallpox usually renders a person immune more effectually than the most perfect vaccination with cowpox. Smallpox reasserted 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 at tacked 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’.

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