Working around Walled Gardens

The Princeton Prosody Archive as Workflow

PRINCETON PROSODY ARCHIVE

Mary Naydan, Rebecca Sutton Koeser, Meredith Martin

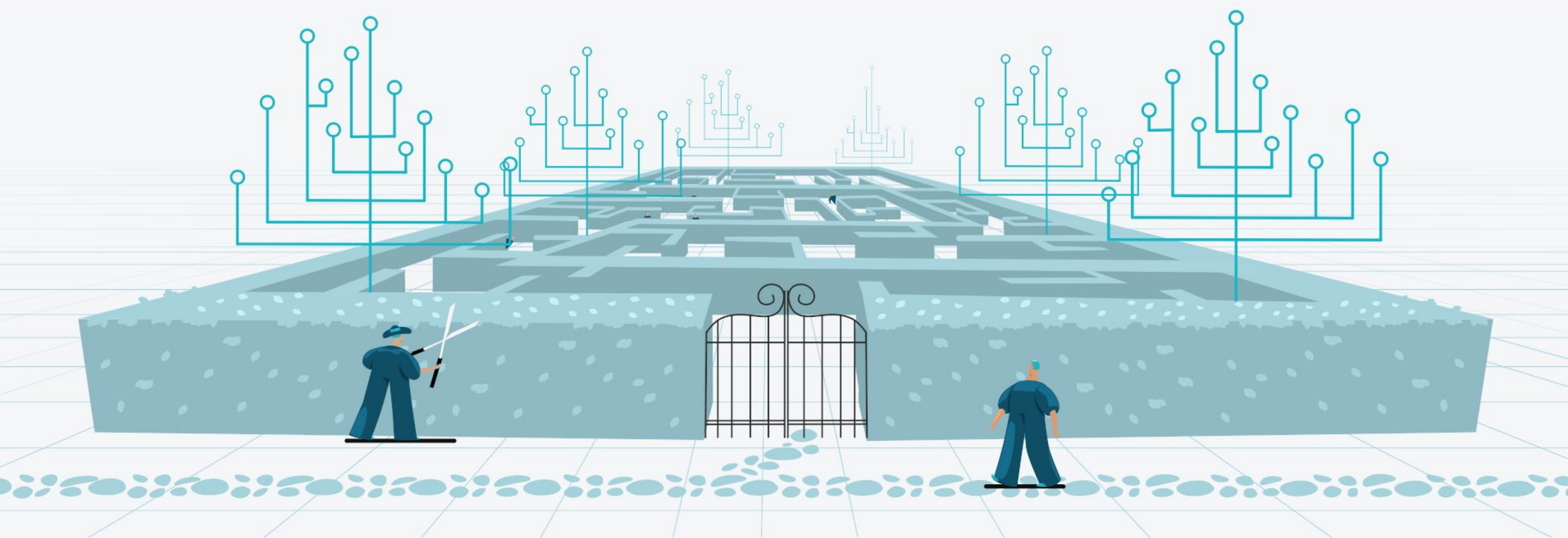
DARIAH Annual Event | June 20, 2024

THE CENTER FOR DIGITAL HUMANITIES @PRINCETON

Overview

1. What are Walled Gardens?
    a. The Digital Research Landscape: How We Got Here
    b. Limitations of Existing “Workflows”
    c. Additional Challenges with TDM
    d. Collaborative DH Projects
2. Case Study: The Princeton Prosody Archive
    a. Overview
    b. Data Sources
    c. Technical Architecture as Workflow
    d. “Dead Ends” & Roadblocks
    e. The Path through the Hedge Maze
3. Futures Beyond Walled Gardens

What are Walled Gardens?



THE CENTER
FOR DIGITAL
HUMANITIES
@PRINCETON

***Since* 1995–**

A **closed platform**, **walled garden**, or **closed ecosystem** is a software system wherein the carrier or service provider has **control** over applications, content, and/or media, and **restricts** convenient access to non-approved applicants or content. This is in contrast to an open platform, wherein consumers generally have unrestricted access to applications and content.

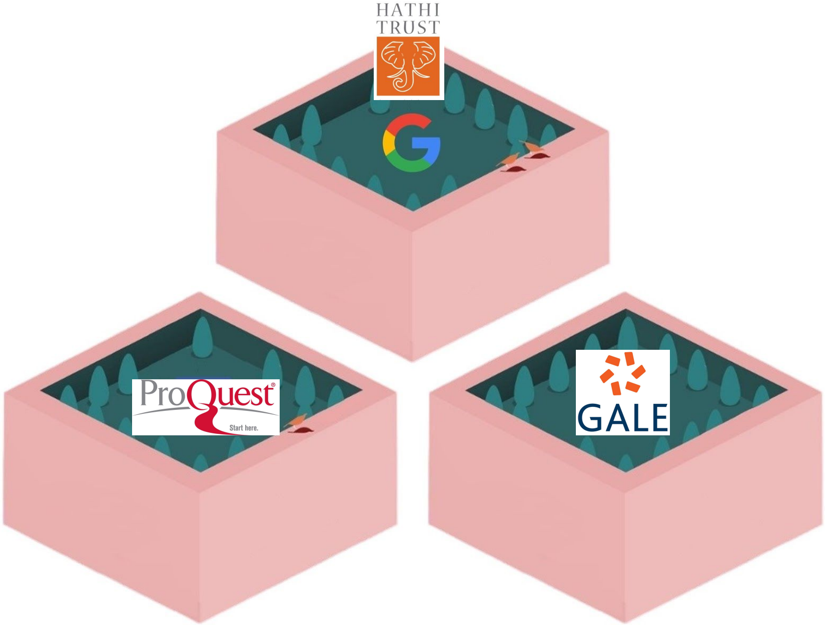
– *Wikipedia*, accessed 2024

The Digital Research Landscape: How We Got Here

- 2008 Google Books project & lawsuits → "propertization and privatization of culture" (Frosio, 2011)

- "Uneven digitization" of texts privilege works from the US and UK (Risam, 2019)

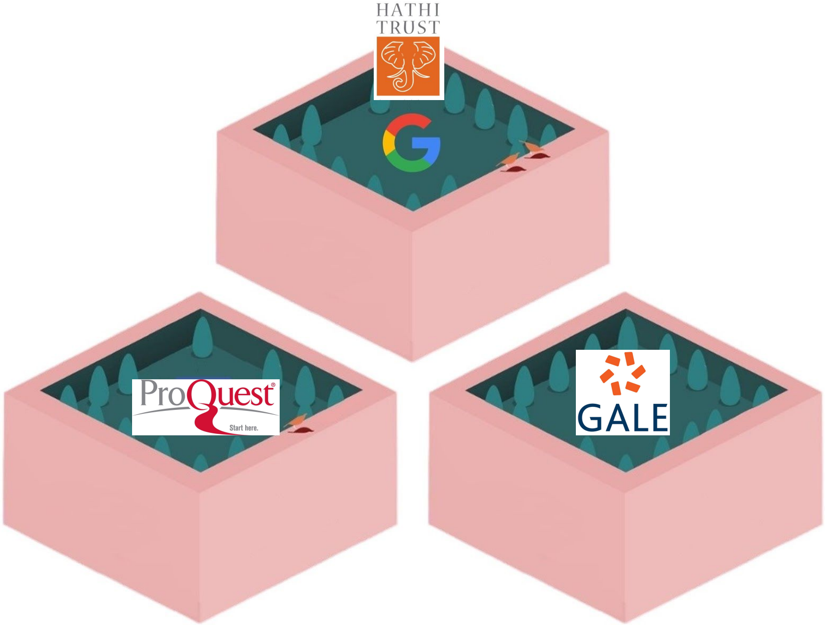
- Prohibited, costly, or unclear access for TDM since 2015 (McCracken & Raub, 2022)



THE CENTER FOR DIGITAL HUMANITIES @PRINCETON

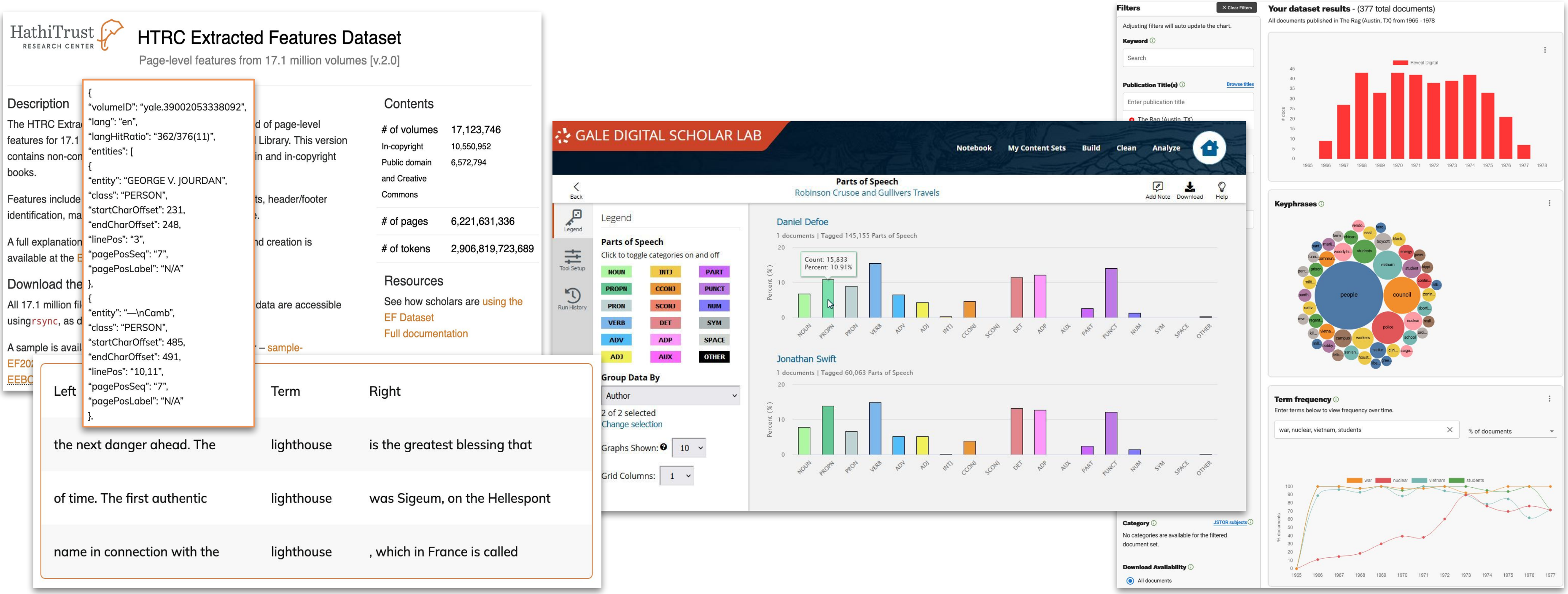
The Digital Research Landscape: How We Got Here

"[Researchers] seek materials that are the most relevant to their theses, which likely come from multiple vendors. While acceptance of TDM by vendors is growing, one of the current problems with vendor-permitted TDM is an **uncooperative, siloed approach**."



Peter McCracken and Emma Raub, "Licensing Challenges Associated with Text and Data Mining," *JLSC* (2023)

Limitations of Existing "Workflows"



HathiTrust RESEARCH CENTER

HTRC Extracted Features Dataset

Page-level features from 17.1 million volumes [v.2.0]

Description

The HTRC Extra… d of page-level features for 17.1 … Library. This version contains non-con… in and in-copyright books.

Features include … ts, header/footer identification, ma… .

A full explanation … nd creation is available at the E…

Download the …

All 17.1 million fil… data are accessible using `rsync`, as d…

A sample is avail… – sample-EF202… EEBC…

Contents

# of volumes	17,123,746
In-copyright	10,550,952
Public domain and Creative Commons	6,572,794
# of pages	6,221,631,336
# of tokens	2,906,819,723,689

Resources

See how scholars are using the EF Dataset

Full documentation

{
"volumeID": "yale.39002053338092",
"lang": "en",
"langHitRatio": "362/376(11)",
"entities": [
{
"entity": "GEORGE V. JOURDAN",
"class": "PERSON",
"startCharOffset": 231,
"endCharOffset": 248,
"linePos": "3",
"pagePosSeq": "7",
"pagePosLabel": "N/A"
},
{
"entity": "—\nCamb",
"class": "PERSON",
"startCharOffset": 485,
"endCharOffset": 491,
"linePos": "10,11",
"pagePosSeq": "7",
"pagePosLabel": "N/A"
},

Left	Term	Right
the next danger ahead. The	lighthouse	is the greatest blessing that
of time. The first authentic	lighthouse	was Sigeum, on the Hellespont
name in connection with the	lighthouse	, which in France is called

GALE DIGITAL SCHOLAR LAB

Notebook · My Content Sets · Build · Clean · Analyze

Parts of Speech

Robinson Crusoe and Gullivers Travels

Legend

Parts of Speech

Click to toggle categories on and off

NOUN INTJ PART PROPN CCONJ PUNCT PRON SCONJ NUM VERB DET SYM ADV ADP SPACE ADJ AUX OTHER

Group Data By: Author

2 of 2 selected

Change selection

Graphs Shown: 10

Grid Columns: 1

Daniel Defoe

1 documents | Tagged 145,155 Parts of Speech

Count: 15,833
Percent: 10.91%

Jonathan Swift

1 documents | Tagged 60,063 Parts of Speech

Filters

✕ Clear Filters

Adjusting filters will auto update the chart.

Keyword: Search

Publication Title(s) — Browse titles

Enter publication title

The Rag (Austin, TX)

Category — JSTOR subjects

No categories are available for the filtered document set.

Download Availability

● All documents

Your dataset results - (377 total documents)

All documents published in The Rag (Austin, TX) from 1965 - 1978

Keyphrases

Term frequency

Enter terms below to view frequency over time.

war, nuclear, vietnam, students — % of documents

Peter McCracken and Emma Raub, "Licensing Challenges Associated with Text and Data Mining," *JLSC* (2023)

THE CENTER FOR DIGITAL HUMANITIES @PRINCETON

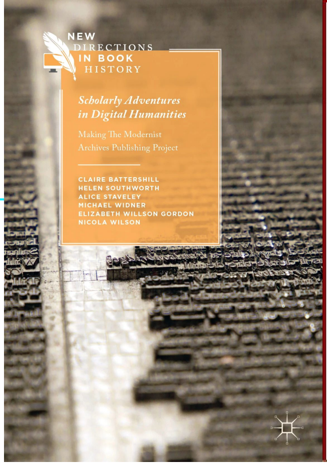
Additional Challenges of TDM

- Inhibitors to TDM research:
  - Strong copyright laws, privacy, and anti-hacking laws that vary country to country
  - Researchers' lack of legal knowledge
  - Private corporation and government grip on data, which impedes access/sharing & disincentivizes them from improving data as technology improves
- Effects:
  - Lessens cross-border/international collaboration
  - Places onus on researcher to transform data into suitable format for TDM
  - Leads US-based researchers to avoid certain projects more frequently than EU-based researchers due to litigious nature of US, despite fair use clause

Patricia Aufderheide, Brandon Butler, and Kimberly Anastacio, "The Chilling Effects of Obstacles to Accessing, Using, and Sharing In-Copyright Data for Quantitative Research," *Information & Culture* (2024)

Collaborative DH Projects

- Typically focus on one data source
- Often reinvent what libraries could be doing
- Digital editions and text corpora are both workarounds to the walled garden
- Front-end UX focus
- Where's the data?

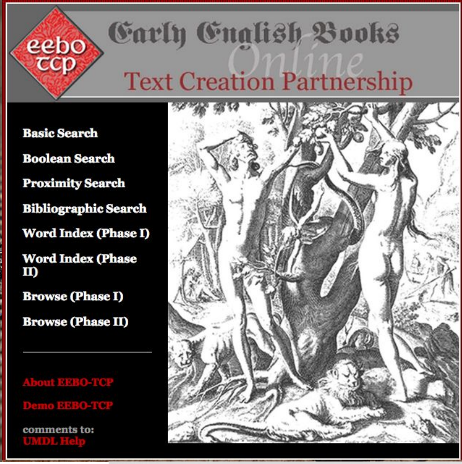


NEW DIRECTIONS IN BOOK HISTORY

Scholarly Adventures in Digital Humanities

Making The Modernist Archives Publishing Project

CLAIRE BATTERSHILL
HELEN SOUTHWORTH
ALICE STAVELEY
MICHAEL WIDNER
ELIZABETH WILLSON GORDON
NICOLA WILSON



eebo tcp

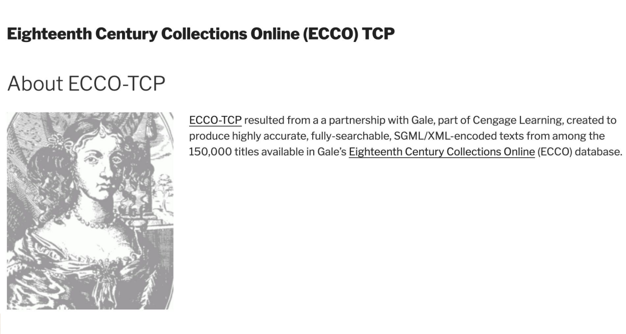
Early English Books Online

Text Creation Partnership

Basic Search
Boolean Search
Proximity Search
Bibliographic Search
Word Index (Phase I)
Word Index (Phase II)
Browse (Phase I)
Browse (Phase II)

About EEBO-TCP
Demo EEBO-TCP

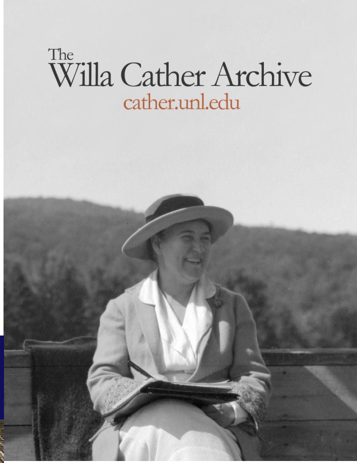
comments to:
UMDL Help



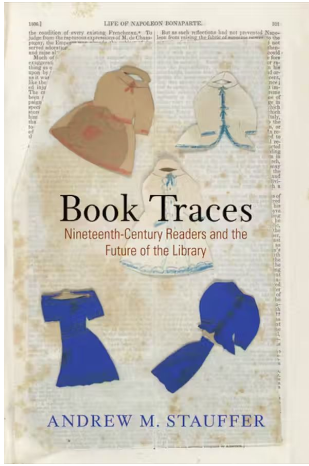
Eighteenth Century Collections Online (ECCO) TCP

About ECCO-TCP

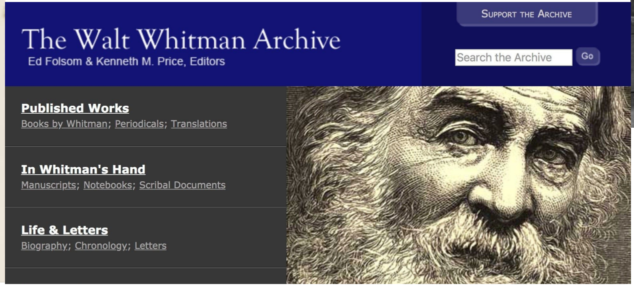
ECCO-TCP resulted from a a partnership with Gale, part of Cengage Learning, created to produce highly accurate, fully-searchable, SGML/XML-encoded texts from among the 150,000 titles available in Gale's Eighteenth Century Collections Online (ECCO) database.



The Willa Cather Archive
cather.unl.edu



Book Traces
Nineteenth-Century Readers and the Future of the Library
ANDREW M. STAUFFER



The Walt Whitman Archive
Ed Folsom & Kenneth M. Price, Editors

Support the Archive

Search the Archive | Go

Published Works
Books by Whitman; Periodicals; Translations

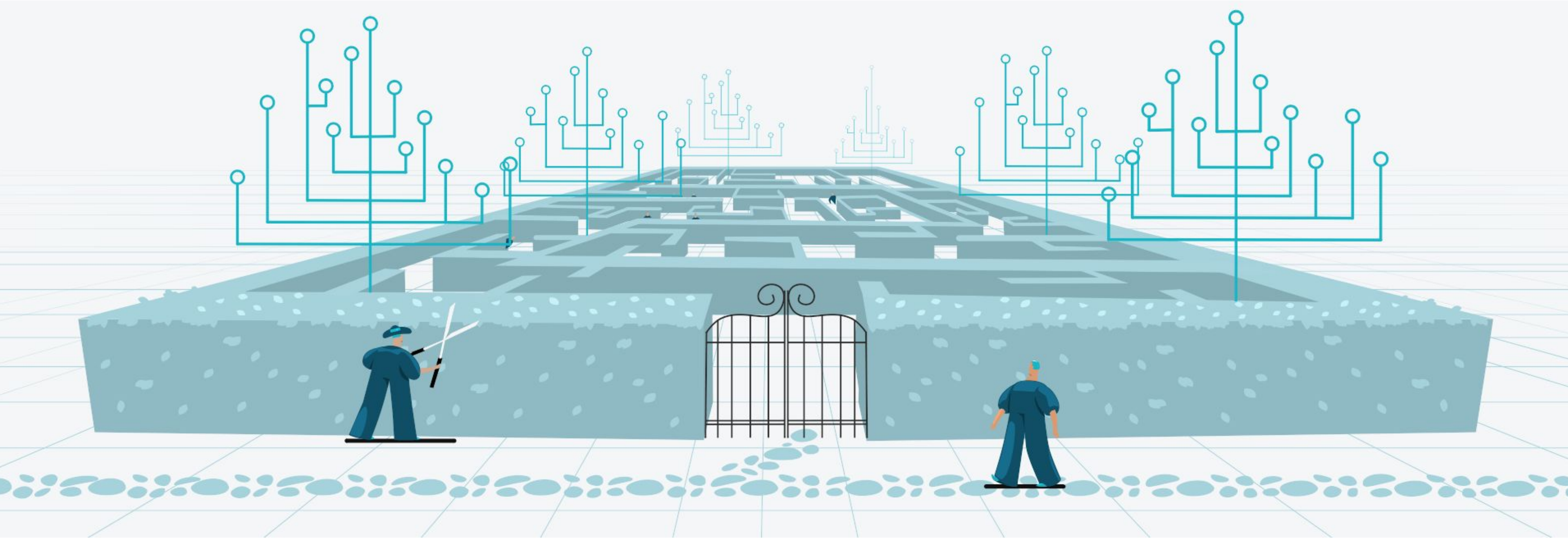
In Whitman's Hand
Manuscripts; Notebooks; Scribal Documents

Life & Letters
Biography; Chronology; Letters

TROVE

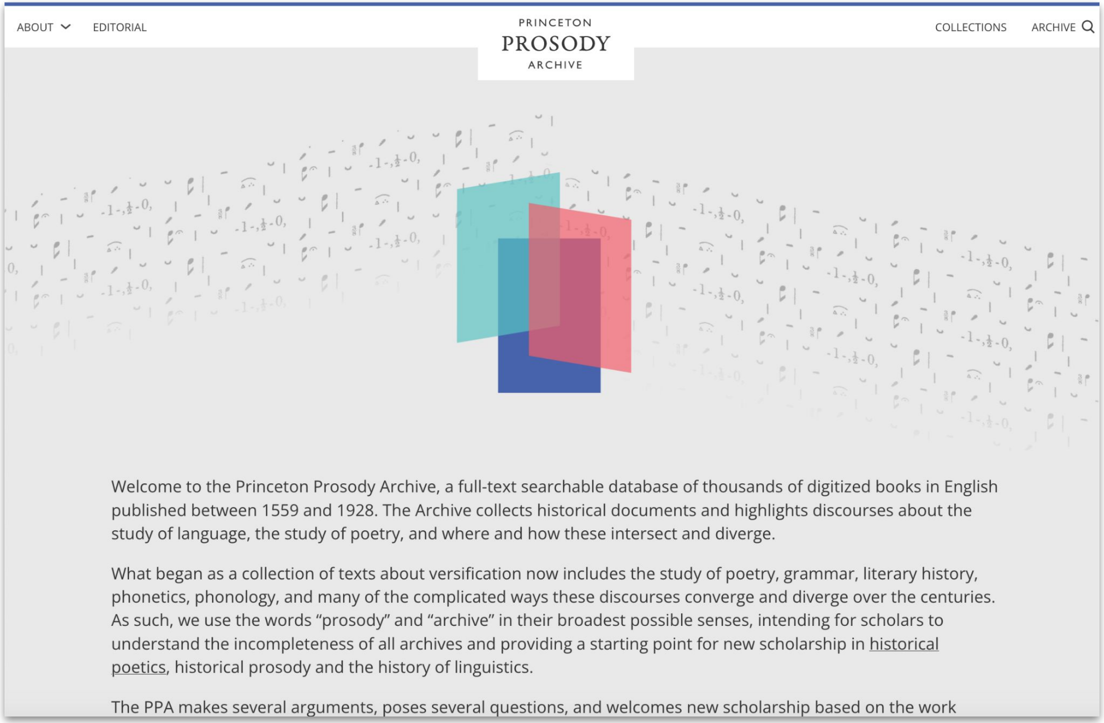
THE CENTER FOR DIGITAL HUMANITIES @PRINCETON

Case Study: The Princeton Prosody Archive



THE CENTER
FOR DIGITAL
HUMANITIES
@PRINCETON

The Princeton Prosody Archive - Overview



Welcome to the Princeton Prosody Archive, a full-text searchable database of thousands of digitized books in English published between 1559 and 1928. The Archive collects historical documents and highlights discourses about the study of language, the study of poetry, and where and how these intersect and diverge.

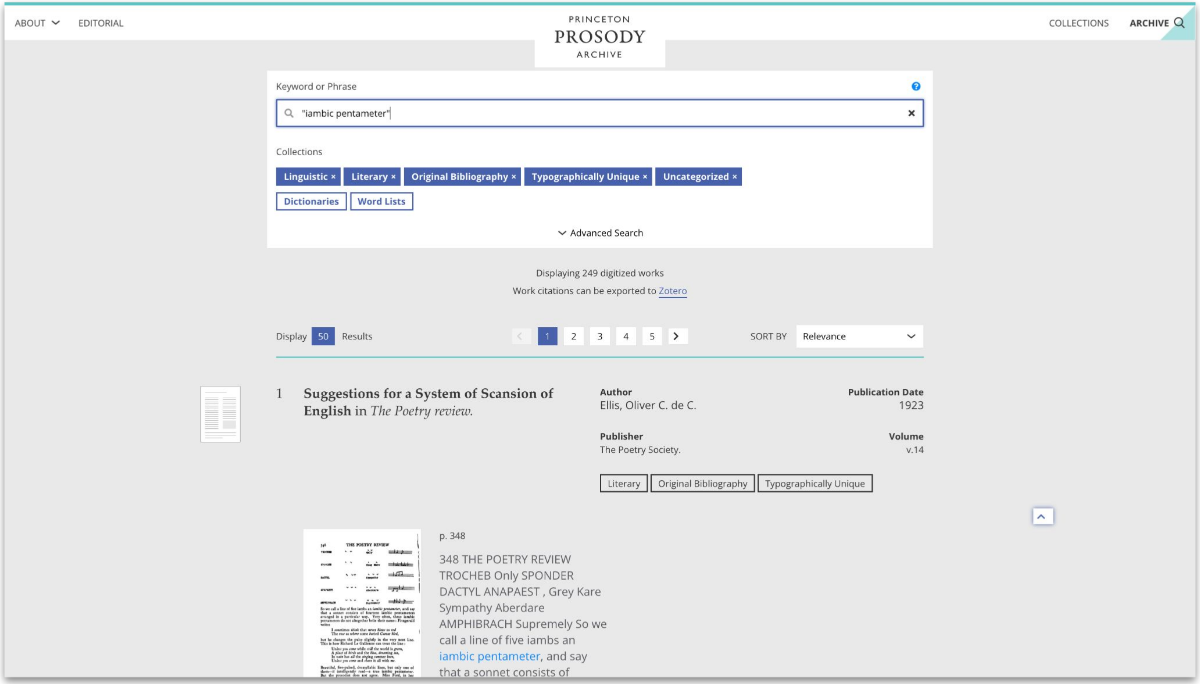
What began as a collection of texts about versification now includes the study of poetry, grammar, literary history, phonetics, phonology, and many of the complicated ways these discourses converge and diverge over the centuries. As such, we use the words "prosody" and "archive" in their broadest possible senses, intending for scholars to understand the incompleteness of all archives and providing a starting point for new scholarship in historical poetics, historical prosody and the history of linguistics.

The PPA makes several arguments, poses several questions, and welcomes new scholarship based on the work

- Full-text searchable database of ~7,000 English-language works from 1559 to 1928 about the study of versification and pronunciation

THE CENTER FOR DIGITAL HUMANITIES @PRINCETON

The Princeton Prosody Archive - Overview



Keyword or Phrase

"iambic pentameter"

Collections

Linguistic × | Literary × | Original Bibliography × | Typographically Unique × | Uncategorized ×

Dictionaries | Word Lists

Displaying 249 digitized works

Work citations can be exported to Zotero

Display 50 Results

1 2 3 4 5

SORT BY Relevance

1 **Suggestions for a System of Scansion of English** in *The Poetry review.*

Author
Ellis, Oliver C. de C.

Publication Date
1923

Publisher
The Poetry Society.

Volume
v.14

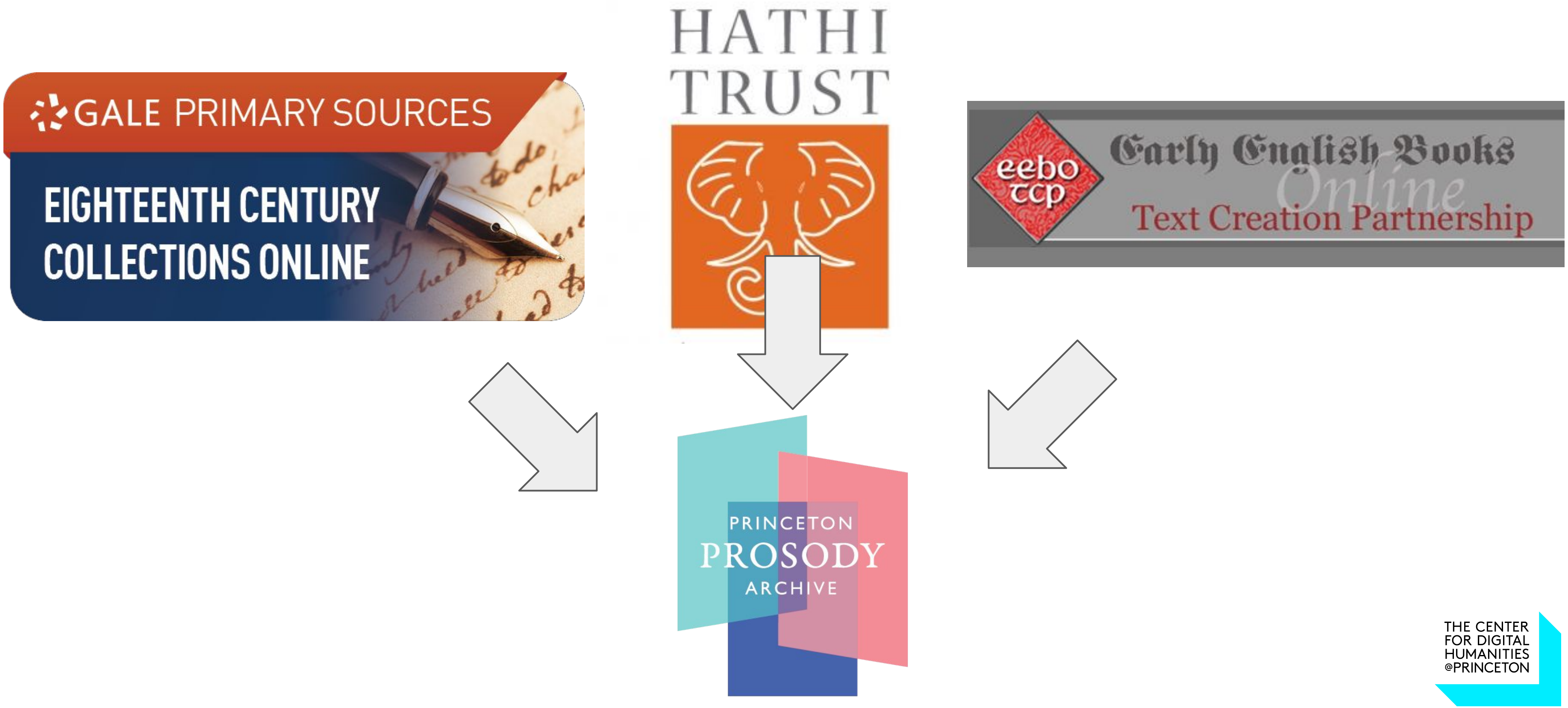
Literary | Original Bibliography | Typographically Unique

p. 348

348 THE POETRY REVIEW TROCHEB Only SPONDER DACTYL ANAPAEST , Grey Kare Sympathy Aberdare AMPHIBRACH Supremely So we call a line of five iambs an iambic pentameter, and say that a sonnet consists of

- Includes "how-to" poetry handbooks, textbooks, grammar books, dictionaries, scholarly articles, elocution manuals, pronunciation guides…
- Curated into 6 collections

Data Sources



GALE PRIMARY SOURCES

EIGHTEENTH CENTURY COLLECTIONS ONLINE

HATHI TRUST

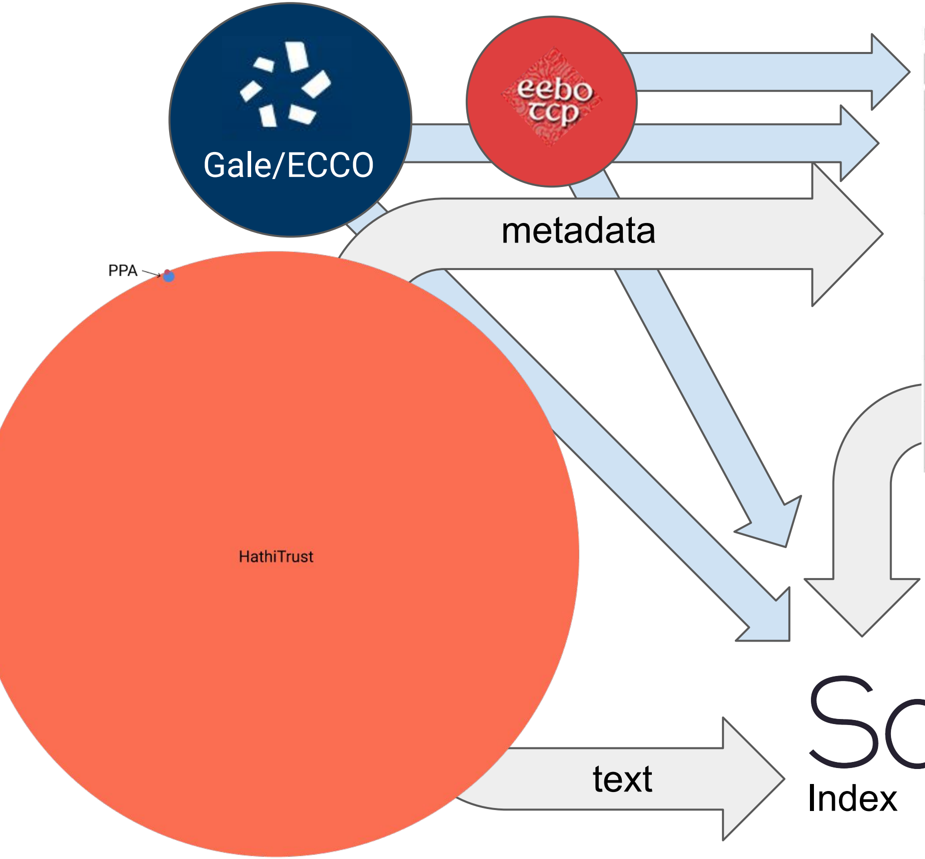
eebo tcp

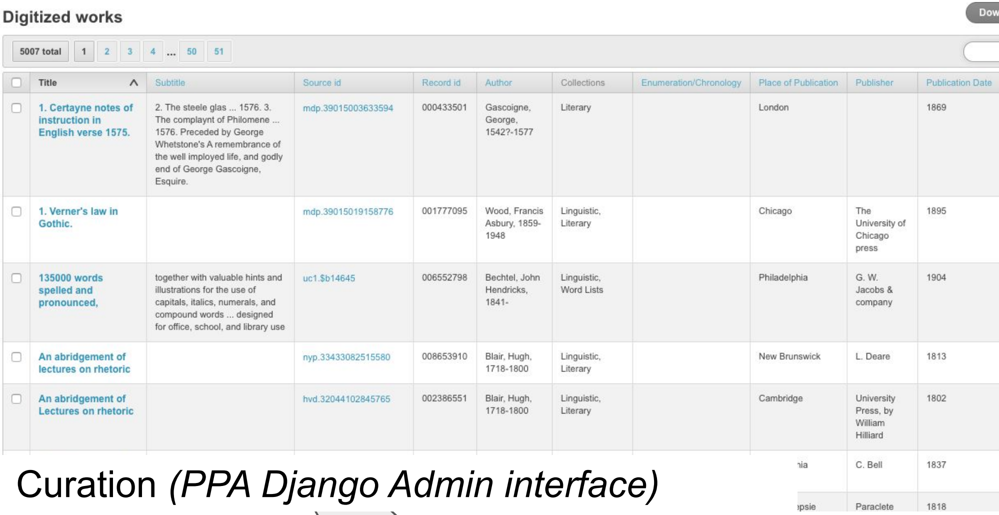
Early English Books Online

Text Creation Partnership

PRINCETON PROSODY ARCHIVE

Technical Architecture as Workflow



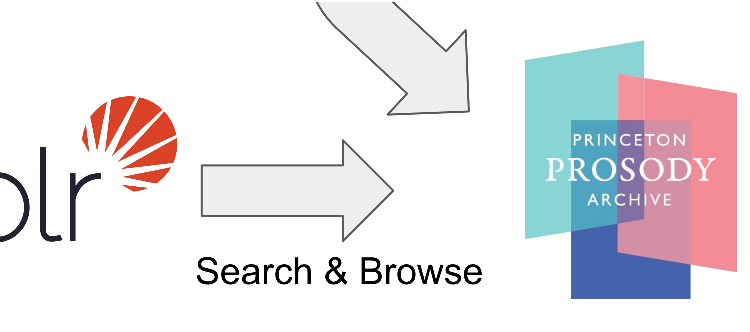


Digitized works

5007 total | 1 | 2 | 3 | 4 | ... | 50 | 51

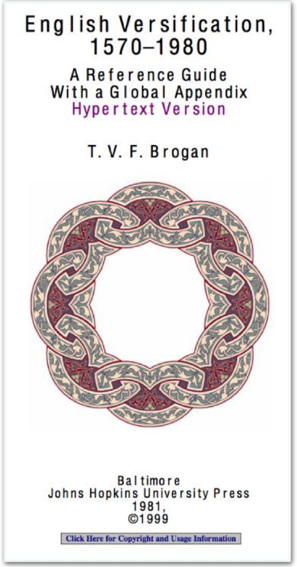
☐	Title	Subtitle	Source id	Record id	Author	Collections	Enumeration/Chronology	Place of Publication	Publisher	Publication Date
☐	1. Certayne notes of instruction in English verse 1575.	2. The steele glas ... 1576. 3. The complaynt of Philomene ... 1576. Preceded by George Whetstone's A remembrance of the well imployed life, and godly end of George Gascoigne, Esquire.	mdp.39015003633594	000433501	Gascoigne, George, 1542?-1577	Literary		London		1869
☐	1. Verner's law in Gothic.		mdp.39015019158776	001777095	Wood, Francis Asbury, 1859-1948	Linguistic, Literary		Chicago	The University of Chicago press	1895
☐	135000 words spelled and pronounced,	together with valuable hints and illustrations for the use of capitals, italics, numerals, and compound words ... designed for office, school, and library use	uc1.$b14645	006552798	Bechtel, John Hendricks, 1841-	Linguistic, Word Lists		Philadelphia	G. W. Jacobs & company	1904
☐	An abridgement of lectures on rhetoric		nyp.33433082515580	008653910	Blair, Hugh, 1718-1800	Linguistic, Literary		New Brunswick	L. Deare	1813
☐	An abridgement of Lectures on rhetoric		hvd.32044102845765	002386551	Blair, Hugh, 1718-1800	Linguistic, Literary		Cambridge	University Press, by William Hilliard	1802
								ia	C. Bell	1837
								psie	Paraclete	1818

Curation *(PPA Django Admin interface)*



THE CENTER FOR DIGITAL HUMANITIES @PRINCETON

“Dead Ends” & Roadblocks



English Versification, 1570–1980

A Reference Guide With a Global Appendix

Hypertext Version

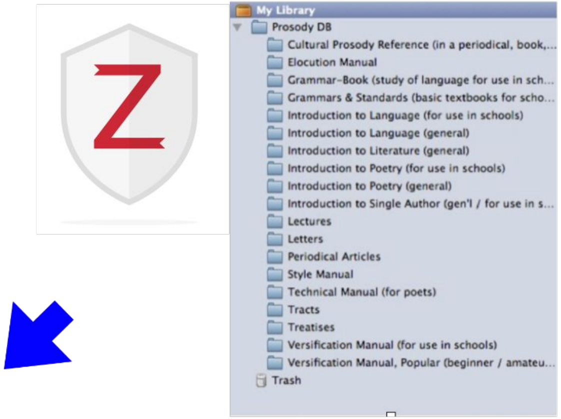
T. V. F. Brogan

Baltimore
Johns Hopkins University Press
1981,
©1999

Click Here for Copyright and Usage Information

ProQuest®
Start here.

ProQuest®
Start here.

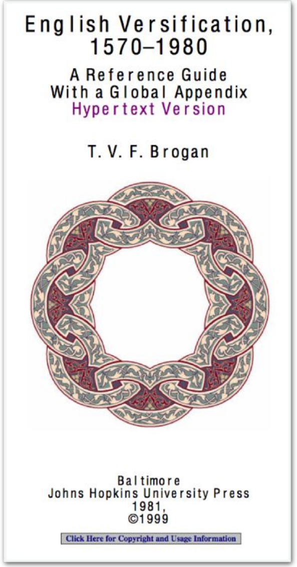


- My Library
  - Prosody DB
    - Cultural Prosody Reference (in a periodical, book,...
    - Elocution Manual
    - Grammar-Book (study of language for use in sch...
    - Grammars & Standards (basic textbooks for scho...
    - Introduction to Language (for use in schools)
    - Introduction to Language (general)
    - Introduction to Literature (general)
    - Introduction to Poetry (for use in schools)
    - Introduction to Poetry (general)
    - Introduction to Single Author (gen'l / for use in s...
    - Lectures
    - Letters
    - Periodical Articles
    - Style Manual
    - Technical Manual (for poets)
    - Tracts
    - Treatises
    - Versification Manual (for use in schools)
    - Versification Manual, Popular (beginner / amateu...
  - Trash

HATHI TRUST + GALE

PRINCETON PROSODY ARCHIVE

“Dead Ends” & Roadblocks



English Versification, 1570–1980

A Reference Guide With a Global Appendix

Hypertext Version

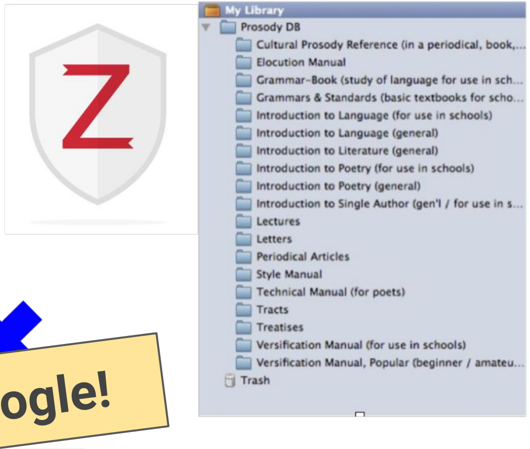
T. V. F. Brogan

Baltimore
Johns Hopkins University Press
1981,
©1999

Click Here for Copyright and Usage Information

ProQuest®
Start here.

ProQuest®
Start here.



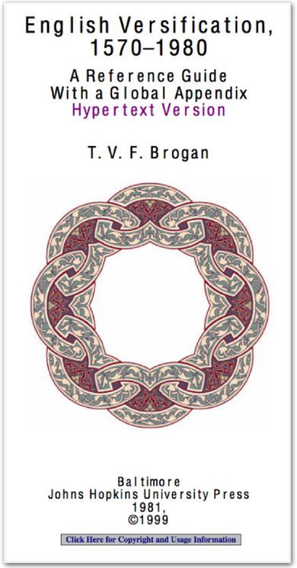
- My Library
  - Prosody DB
    - Cultural Prosody Reference (in a periodical, book,...
    - Elocution Manual
    - Grammar-Book (study of language for use in sch...
    - Grammars & Standards (basic textbooks for scho...
    - Introduction to Language (for use in schools)
    - Introduction to Language (general)
    - Introduction to Literature (general)
    - Introduction to Poetry (for use in schools)
    - Introduction to Poetry (general)
    - Introduction to Single Author (gen'l / for use in s...
    - Lectures
    - Letters
    - Periodical Articles
    - Style Manual
    - Technical Manual (for poets)
    - Tracts
    - Treatises
    - Versification Manual (for use in schools)
    - Versification Manual, Popular (beginner / amateu...
  - Trash

non-compete w/ Google!

HATHI TRUST + GALE

PRINCETON PROSODY ARCHIVE

“Dead Ends” & Roadblocks



English Versification, 1570–1980

A Reference Guide With a Global Appendix

Hypertext Version

T. V. F. Brogan

Baltimore
Johns Hopkins University Press
1981,
©1999

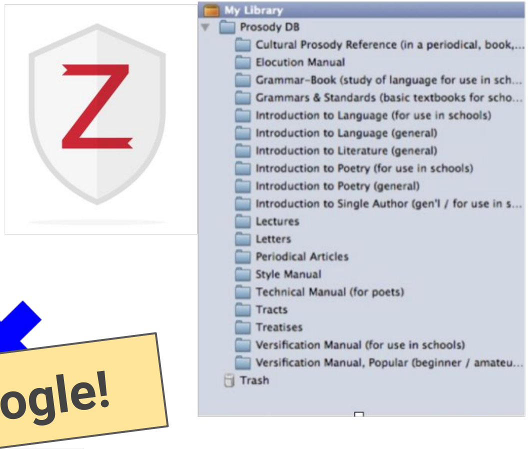
Click Here for Copyright and Usage Information

ProQuest®

Start here.

ProQuest®

Start here.



My Library

- Prosody DB
  - Cultural Prosody Reference (in a periodical, book,...
  - Elocution Manual
  - Grammar-Book (study of language for use in sch...
  - Grammars & Standards (basic textbooks for scho...
  - Introduction to Language (for use in schools)
  - Introduction to Language (general)
  - Introduction to Literature (general)
  - Introduction to Poetry (for use in schools)
  - Introduction to Poetry (general)
  - Introduction to Single Author (gen'l / for use in s...
  - Lectures
  - Letters
  - Periodical Articles
  - Style Manual
  - Technical Manual (for poets)
  - Tracts
  - Treatises
  - Versification Manual (for use in schools)
  - Versification Manual, Popular (beginner / amateu...
- Trash

ete w/ Google!

lots of restrictions!

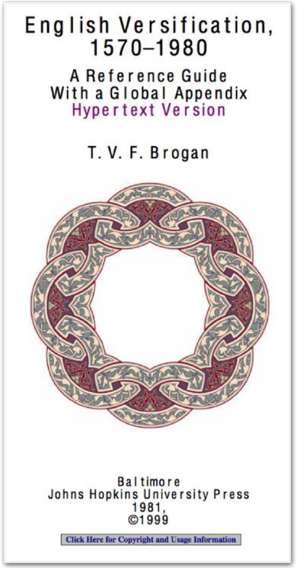
TRUST

+

GALE

PRINCETON PROSODY ARCHIVE

“Dead Ends” & Roadblocks



English Versification, 1570–1980

A Reference Guide With a Global Appendix

Hypertext Version

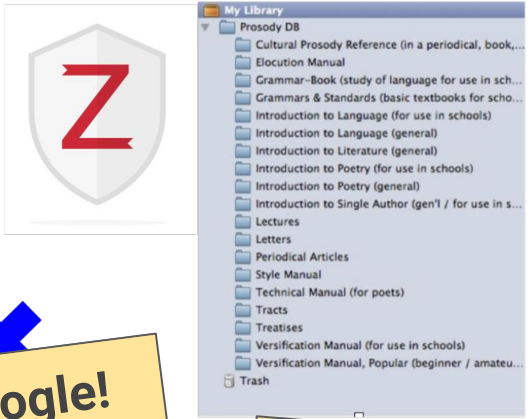
T. V. F. Brogan

Baltimore
Johns Hopkins University Press
1981,
©1999

Click Here for Copyright and Usage Information

ProQuest Start here.

ProQuest Start here.



- My Library
  - Prosody DB
    - Cultural Prosody Reference (in a periodical, book,...
    - Elocution Manual
    - Grammar-Book (study of language for use in sch...
    - Grammars & Standards (basic textbooks for scho...
    - Introduction to Language (for use in schools)
    - Introduction to Language (general)
    - Introduction to Literature (general)
    - Introduction to Poetry (for use in schools)
    - Introduction to Poetry (general)
    - Introduction to Single Author (gen'l / for use in s...
    - Lectures
    - Letters
    - Periodical Articles
    - Style Manual
    - Technical Manual (for poets)
    - Tracts
    - Treatises
    - Versification Manual (for use in schools)
    - Versification Manual, Popular (beginner / amateu...
  - Trash

ete w/ Google!

lots of restrictions!

TRUST

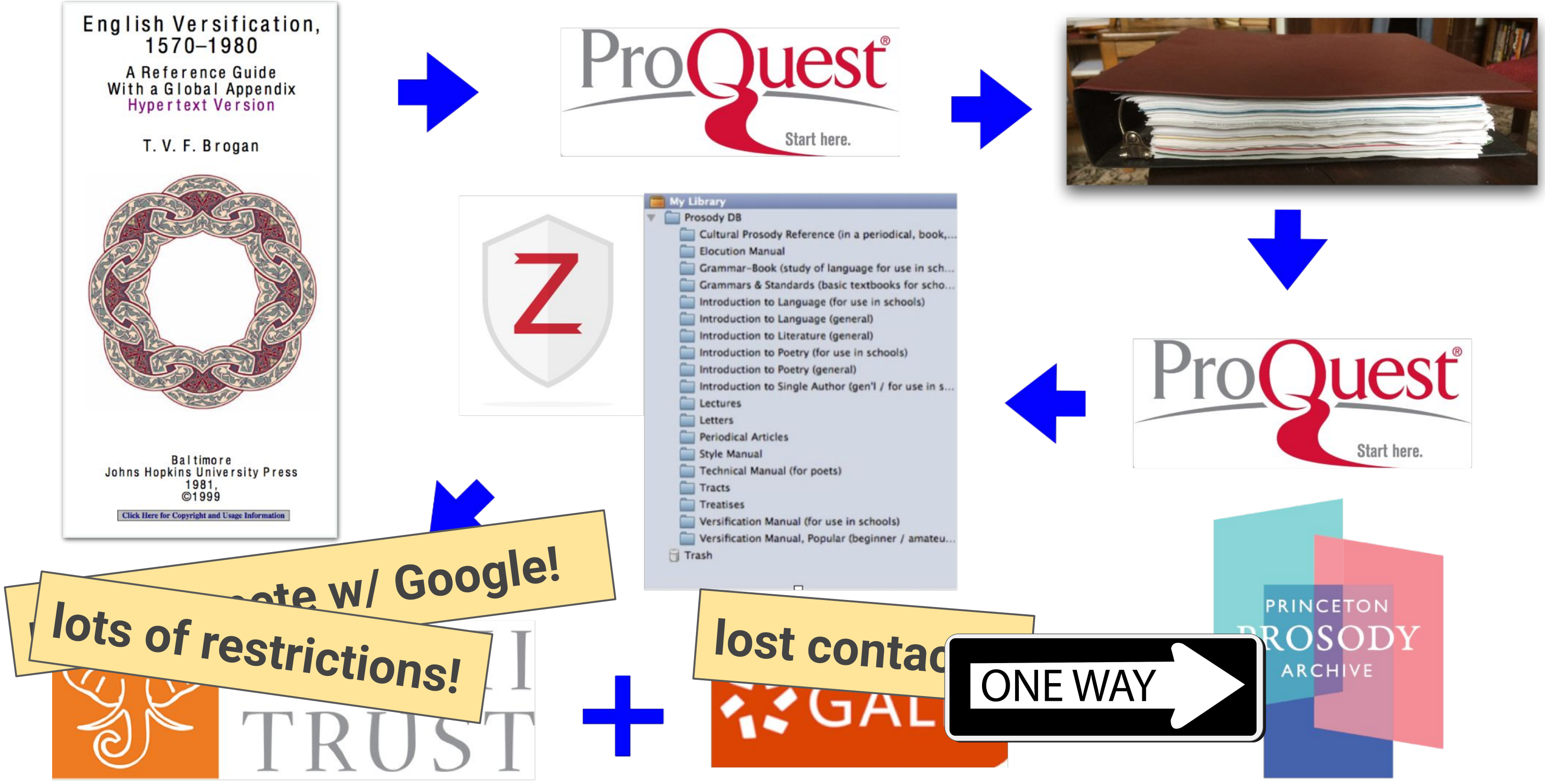
+

lost contacts

GALE

PRINCETON PROSODY ARCHIVE

“Dead Ends” & Roadblocks



English Versification, 1570–1980

A Reference Guide With a Global Appendix

Hypertext Version

T. V. F. Brogan

Baltimore
Johns Hopkins University Press
1981,
©1999

Click Here for Copyright and Usage Information

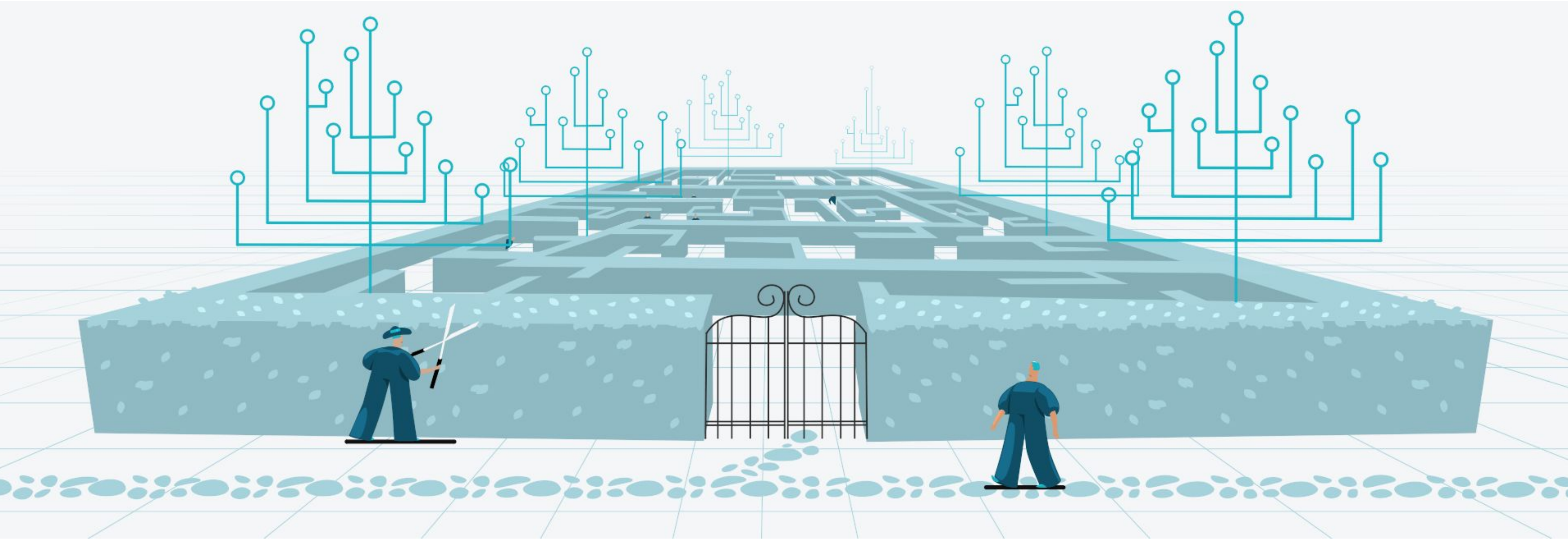
My Library

- Prosody DB
  - Cultural Prosody Reference (in a periodical, book,...
  - Elocution Manual
  - Grammar-Book (study of language for use in sch...
  - Grammars & Standards (basic textbooks for scho...
  - Introduction to Language (for use in schools)
  - Introduction to Language (general)
  - Introduction to Literature (general)
  - Introduction to Poetry (for use in schools)
  - Introduction to Poetry (general)
  - Introduction to Single Author (gen'l / for use in s...
  - Lectures
  - Letters
  - Periodical Articles
  - Style Manual
  - Technical Manual (for poets)
  - Tracts
  - Treatises
  - Versification Manual (for use in schools)
  - Versification Manual, Popular (beginner / amateu...
- Trash

The Path through the Hedge Maze

- Leverage the power of your institution
- Devise Memoranda of Understandings (MOUs)
- Cultivate relationships with librarians and vendors
- Prepare for change (in contacts & data)
- Create a culture of documentation
- Team continuity is key
- Be persistent advocating for your research needs

Futures Beyond Walled Gardens



Futures Beyond Walled Gardens

- Move toward Open Source options like Internet Archive, which just made ECCO available
- Generalizable code like the PPA that cuts across walled gardens
  - Partnership/Fellowship with HathiTrust so that others can create “PPAs” with different content?
- Cross-institutional leverage
- Cultural Data Collectives movement

THE CENTER FOR DIGITAL HUMANITIES @PRINCETON

References

- Giancarlo F. Frosio, “Google Books Rejected: Taking the Orphans to the Digital Public Library of Alexandria,” *Santa Clara Computer and High Technology Law Journal* 28, no. 1 (Nov. 2011): 81-141.
- Roopika Risam, *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy* (Northwestern UP, 2019).
- Patricia McCracken and Emma Raub, “Licensing Challenges Associated With Text and Data Mining: How Do We Get Our Patrons What They Need?”, *Journal of Librarianship and Scholarly Communication* 11 no. 1 (2023): 1-14, https://doi.org/10.31274/jlsc.15530.
- Patricia Aufderheide, Brandon Butler, and Kimberly Anastacio, “The Chilling Effects of Obstacles to Accessing, Using, and Sharing In-Copyright Data for Quantitative Research,” *Information & Culture* 59, no. 1 (2024): 44-65, https://doi.org/10.7560/IC59103.
- “About the Collection,” HathiTrust, https://www.hathitrust.org/the-collection/.
- Ted Underwood, Patrick Kimutis, and Jessica Witte, “NovelTM Datasets for English-Language Fiction, 1700-2009,” *Journal of Cultural Analytics* 5, no. 2 (2020): 1-30, https://doi.org/10.22148/001c.13147.
- John A. Walsh et al., “‘The Library is Open!’: Open Data and an Open API for the HathiTrust Digital Library,” *CHR2023: Computational Humanities Research Conference* (December 6–8, 2023, Paris, France), https://ceur-ws.org/Vol-3558/paper7875.pdf.