

Digital Scholarship Fundamentals Workshops

Sponsored by the [University Libraries](#) and the
[Department of Library and Information Science](#)

Starting a Text Mining Project

Friday, September 30, 12:00 pm – 1:00 pm

Planning your project is a critical skill in contemporary research. Many students and faculty embark on a project without understanding fully the resources needed and the time commitment involved. This workshop will walk you through the process of a text mining project by asking the right questions: what is my research question, how can I locate and acquire the texts, what tools are relevant for cleaning and analyzing the texts, and what legal issues may limit my access and use of the texts?

Gale Digital Scholar Lab (1): Building a Dataset

Friday, October 21, 12:00 pm – 1:00 pm

Come and explore our latest acquisition devoted to text analysis, data mining, and data visualization through a series of three workshops. Using Gale's Primary Sources archive (or combining with your own text data), we will: (1) build a corpus, (2) clean the data, and (3) perform analyses. The first workshop will be building a corpus from the documents made available through our library subscription. We will learn how to search for documents, add them to a content set, upload our own .txt files, and combine them with the Gale content. No previous experience necessary.

Authors' Rights & the Publishing Industry

Monday, October 24, 12:00 pm – 1:00 pm

As a graduate student or a faculty member seeking an academic position, tenure, or promotion, you will need to establish a scholarly presence and build your curriculum vitae. A building block in this process is publishing in quality academic journals (subscription-based or open access). This workshop will assist you in selecting the right journal, ascertaining your rights as an author, and explore the problem of predatory publishing practices and how to avoid becoming a victim. Part of Open Access Week (October 24th-30th).

Gale Digital Scholar Lab (2): Cleaning a Dataset

Friday, November 4, 12:00 pm – 1:00 pm

The second workshop on using the Gale Digital Scholar Lab will focus on cleaning our data. The integral 'Clean' tool works with unstructured text data that is generated by the OCR process and shown in Lab. We will walk through the stages of the cleaning process by tackling such problems as removing ASCII characters and document sections, normalizing whitespace, and performing text corrections and modifications as necessary. How you clean your data will impact the tools used for analysis.

Gale Digital Scholar Lab (3): Analyzing a Dataset

Friday, Nov. 11, 12:00 pm – 1:00 pm

The final workshop in our Gale Digital Scholar Lab series will focus on the tools for analyzing a dataset. Document clustering, Named Entity Recognition, Ngrams, parts of speech, sentiment analysis, and topic modeling, will be covered.

Register through the Events page at libraries.catholic.edu (CUA members only) or by contacting Kevin Gunn (gunn@cua.edu). Unless otherwise indicated, the instructor for each session will be Kevin Gunn, Coordinator of Digital Scholarship. All workshops will take place on Zoom, recorded, and made available on the Catholic University Libraries' [YouTube Channel](#).



University Libraries

<https://libraries.catholic.edu/services/digital-scholarship/index.html>