GALE

INSIGHTS REPORT

Discover How Digital Humanities Tools Enrich Group Research: An Inside Look with Gale Digital Scholar Lab

"Working with the *Lab* has broadened my perspective on what the digital humanities are capable of and their benefit to literary studies more generally. While my previous work with digital humanities tools sought to analyze novels, working in the *Lab* has proven that historical text-based documents contain a myriad of information that has yet to be explored in full."

Ronny Litvack-Katzman
Research Assistant

MEET THE RESEARCHER

Nathalie M. Cooke, professor of English at McGill University, began her career as a math instructor. Her interests for the literature classroom have focused on literary geography and the use of ArcGIS, a cloud-based mapping and analysis solution that allowed her to explore how historical narrative functions as deep mapping. Professor Cooke was fascinated by novels that allowed reading in visual and virtual ways, and she taught graduate classes that incorporated geocoding in literature.

Attendance at the University of Victoria's Digital Humanities Summer Institute (DHSI) in British Columbia inspired her to refine her teaching to create a more immersive and engaging classroom experience for students. When Professor Cooke assumed her role as associate dean of Rare & Special Collections, Osler, Art, and Archives (ROAAr) at McGill Library, her focus shifted to research areas that shed light on McGill's collections, including two specific team-based projects: The Riddle Project and Ciphers of *The Times*.

An early adopter of *Gale Digital Scholar Lab* (the *Lab*), McGill University prioritizes developing online programs that reduce barriers to interdisciplinary teaching and research. Read on to discover how Professor Cooke used *Gale Primary Sources* and *Gale Digital Scholar Lab* to advance research grounded in McGill Library's local collections using the power of text and data mining and analysis tools.

UNEARTHING RIDDLES IN EARLY MANUSCRIPTS

McGill Library's purchase of a manuscript containing enigmatic wordplay from as early as the eighteenth century prompted the first of Professor Cooke's projects utilizing the *Lab*. In The Riddle Project, the professor and her team of digitization specialists, cataloguers, and researchers sought to identify and uncover Enigmatic Bills of Fare (EBoFs)—which are often handwritten food menus that were coded in riddles and used to spark friendly debate at the dinner table—to understand why they were created and how enigmas and wordplay were used in the West between the eighteenth and early twentieth centuries.

When the team started their research on culinary riddles, they discovered how niche the topic was, with just a couple of acquisitions in the library's special collections driving the inquiry. One of the very few sources stated definitively that the first EBoF was created in 1755 as a Christmas menu for George II. Professor Cooke's team disproved this premise, however, using Gale's archives and other research resources to find multiple examples of EBoFs published as early as 1733.

After gathering examples of EBoFs, the project team crowdsourced answers for bills of fare riddles using social media and then, using *Gale Digital Scholar Lab*'s document clustering, ngrams, parts of speech, and sentiment analysis tools, analyzed the responses they received. They endeavored to resolve questions of gender bias in the riddles, as well as location and semantic change over time.

DECRYPTING VICTORIAN AGONY COLUMNS

Professor Cooke also oversees Ciphers of *The Times*, a project dedicated to gathering and analyzing the agony columns in Gale's *The Times Digital Archive* and related period novels in the Victorian era.

Found on the front page of daily Victorian newspapers, the collection of personal advertisements known as the "agony column" was composed of largely anonymous contributions sent in by writers from across the British Empire. Over the course of the nineteenth century, these advertisements evolved into a public forum for citizens to communicate with family and friends and, in some cases, engage in clandestine and cryptic correspondences. Typically only a few dozen words in length, agony ads were read by lovers, criminals, students, and entertainment seekers alike and often included codes and ciphers.

Professor Cooke's team sought to understand the relationship between personal advertisements and the "newspaper novel" genre. The team drew on secondary criticism of agony ads, including work by Jean Palmer, who manually combed through *The Times* agony ads and recorded any that contained codes and ciphers, a project that took a decade or more. By contrast, the McGill project team used *Gale Primary Sources* and *Gale Digital Scholar Lab* to complete their searches and create a corpus in just months.

> "*Gale Digital Scholar Lab* offers incredibly powerful tools that are very intuitive to use."
>
> **Nathan Drezner**
> Research Assistant

ANSWERING QUESTIONS WITH DIGITAL HUMANITIES TOOLS

Gale Digital Scholar Lab served as a springboard for both McGill projects, enabling the teams to answer foundational research questions: Can we find a corpus of material? What is the scope of the corpus? What can be discovered using specific tools from this corpus?

After sourcing a rich compendium of agony columns from *The Times* in the *Lab*, the team ran basic trials using document clustering, parts of speech, ngrams, and sentiment analysis tools to see what could be learned through text analysis. Optical character recognition (OCR) metrics were particularly helpful in filtering out unusable documents; and the team found the granular search functionality, as well as the search history, important for finding and saving data. The summary information for each content set proved similarly helpful—for teamwork, the curation, management, and analysis of team members' content sets was much easier.

SAVING TIME ANALYZING METADATA

Metadata is critical to both projects since much of the analytical work seeks to understand changes over time. For example, text mining can be used to determine whether there are discernable differences between agony columns in 1840 and 1880.

Similarly, working with OCR text and underlying metadata helped the researchers understand the relationship between specific genres of novels and newspapers of the same period. As they considered how the material transformed over time, the team looked at how various months or days of the week could affect publication content. For example, there are many coded forms of communication in the papers, but are there any patterns? Manual work can identify individual instances and anomalies but can't efficiently visualize themes and patterns at scale; for example, identifying Saturday as the day when people communicate or highlighting certain hot months or hot periods of the year, perhaps during holidays or other significant times. After making extensive use of the metadata downloaded from *Gale Primary Sources*, the topic modeling tool in the *Lab* was instrumental in uncovering these patterns, along with further analysis tools that the researchers coded themselves.

> "[The *Lab* allows us] to ask more general questions with some level of accuracy that traditional research methods cannot. The power of digital tools is primarily realized in their scope, allowing us to pose questions at a 'distance' from the individual texts themselves. By cross-analyzing large corpora, we can identify similarities and differences at the level of part-of-speech tagging, principal component analysis, text complexity, overarching topics and themes, and more."
>
> **Professor Cooke's research team**

PARSING & VISUALIZING LARGE AMOUNTS OF DATA

One of the advantages of *Gale Primary Sources* metadata is the inclusion of unique identifiers, which are a benefit of working with XML hard drives. When downloading metadata from the *Lab*, Professor Cooke's project teams used the Gale DocID as a common point to merge the downloaded metadata with the OCR text. The team also used XML coordinates to navigate directly to the pages and columns containing the agony ads.

Access to the XML drives was a turning point for the project since the team could parse the data and identify the very specific contexts of agony ad columns. With the XML coordinates in each document, the team built a parser to grab the column programmatically, using JupyterLab for manual coding of their analytical pipelines and GitHub to document their process and edits. They also used multiple Python libraries for text analysis and visualizations, including R and R Studio for supplementary data analysis and visualization.

INCREASING USAGE OF THE LIBRARY'S INVESTMENT

Professor Cooke's teams extensively used *Gale Primary Sources* archives, including *The Times Digital Archive, Seventeenth and Eighteenth Century Burney Newspapers Collection, Eighteenth Century Collections Online, British Library Newspapers*, and *The Illustrated London News Historical Archive, 1842–2003*. During the project periods in 2022, the number of documents used in content sets increased by 14%, and the number of *Gale Primary Sources* archives used across content sets in 2022 was nearly eight times the number in 2021.

COMPARISON OF USAGE BEFORE & AFTER PROJECT PERIODS

Metrics	JAN 21 – DEC 21	Jan 22 – Dec 22	Percent Change
Sessions	71	57	-20%
Searches	278	322	16%
Minutes	2309	2511	9%
Searches Per Session	3.92	5.65	44%
Minutes Per Session	32.52	44.05	36%

After working with students and researchers on The Riddle Project and Ciphers of *The Times*, Professor Cooke developed course syllabi to integrate *Gale Digital Scholar Lab* into teaching and learning at McGill University.

> "We hope that digital humanities research will become an established and prominent avenue of research in the field. We believe that libraries specifically can benefit greatly from these forms of analysis since they can also make such data accessible and interactive for non-tech-oriented students, academics, and the general public."
>
> **Professor Cooke's research team**

EXPLORE THE DATA IN MORE DETAIL

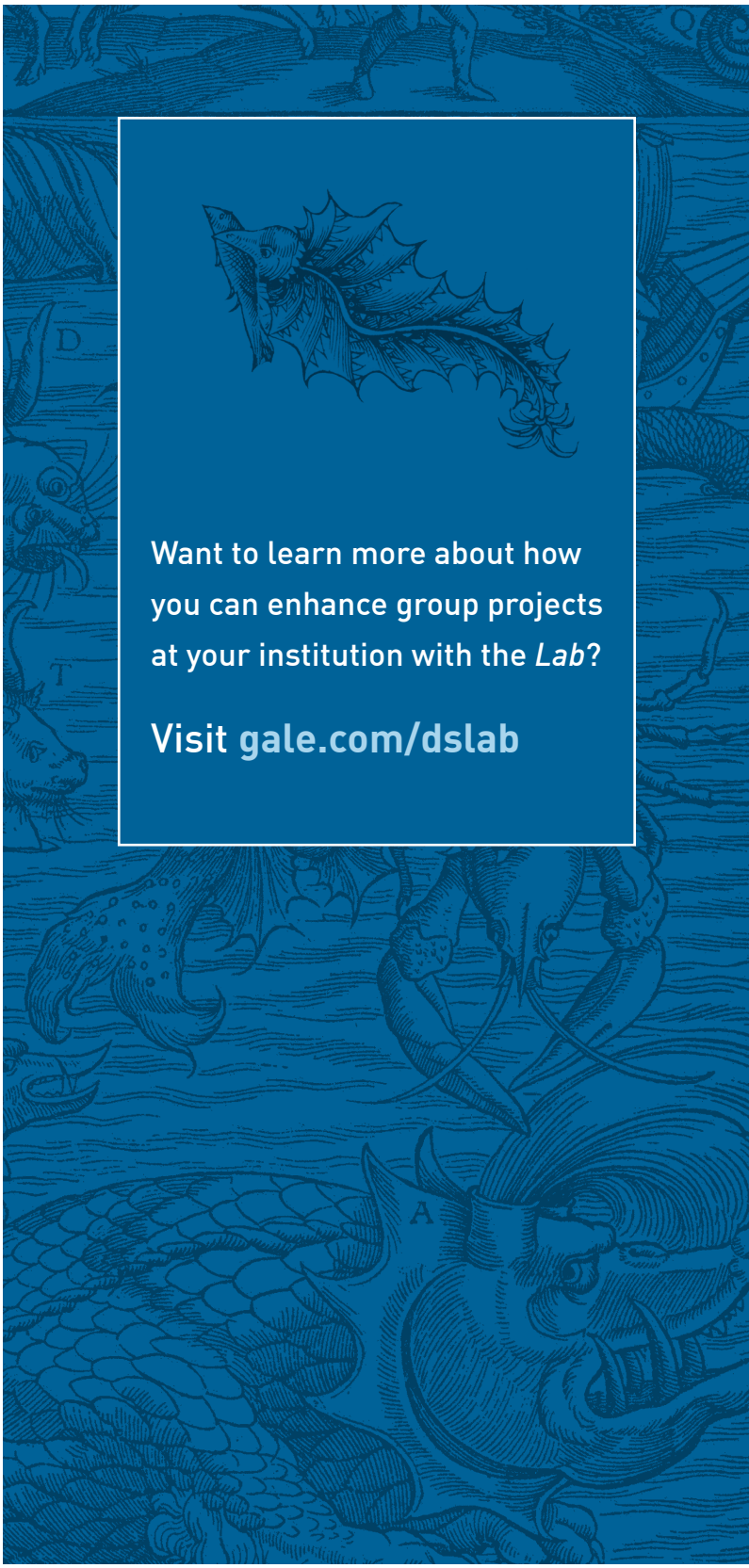
In addition to online exhibits, the project team has published articles and book chapters that discuss their research across both projects.

Want more information about their active research?

Visit riddleproject.github.io to explore The Riddle Project or libraryponders.github.io to dive into Ciphers of *The Times*.

GET IDEAS FOR TEACHING & LEARNING IN THE LAB

Gale Digital Scholar Lab is an industry-leading text and data mining research environment that removes barriers to digital scholarship for students and researchers at every level.



GALE