

DISCOVER THE POSSIBILITIES

EXPLORE PRIMARY SOURCES
THROUGH A NEW LENS

BUILD



CLEAN



ANALYZE



INSIGHTS FROM OUR TEAM

WHERE WE CAME FROM AND WHERE WE'RE HEADING

Phil Faust, Vice President and Publisher, Academic Product, Gale a Cengage Company



Gale has been at the forefront of supporting humanities research in the digital space for many years. From the mantra “If it’s not online, it doesn’t exist” to the creation of the first text mining and analysis platform developed specifically for academic libraries by a commercial humanities publisher, Gale has watched scholarly research morph and evolve to fit the multifaceted, technology-laden interdisciplinary movement of today.

In 2014, after years of providing libraries and scholars with raw text files from our extensive archival collections, we realized that we needed to truly democratize the use of this data to a wider spectrum of users. We wanted to offer an access point without the hardware and support demands that come with physical data drives. By collaborating with scholars around the world, we created a cloud-based repository of Gale content that simplified corpus building by eliminating the need to store, mount, index, and query data drives. This broke down a significant barrier to end users: the ability to curate large content sets that will help answer key data-driven research questions that can’t be answered through close reading alone.

Gale, along with global advisors, aligned popular open-source tools, which further extend the platform’s application to both research and teaching. The result is the *Gale Digital Scholar Lab*. This new platform aligns over 600 years of multidisciplinary primary source material and optical character recognition (OCR) text with a tool set that allows students and researchers to ask new questions of the content by learning and using accessible open-source tools.

As new features and functionality are added to the platform, Gale is eager to work with current and new academic partners to better understand the research, teaching, and learning needs on campus. Most importantly, we rely on our library partners to further develop the *Lab* in ways that shed new light on the value of their roles and the resources that support scholarship across the curriculum.

THE NEW LANGUAGE OF MAKING DISCOVERIES

Chris Houghton, Head of Digital Scholarship, International, Gale, a Cengage Company

We have learned a lot in the ongoing development of the *Gale Digital Scholar Lab*. We've learned more about our customers, the nature of teaching and researching digital humanities, and how best to frame Gale data. The *Lab* has undergone significant development in its first 18 months, and I'm delighted to say that much of this development has been suggested, guided, and tested by customers—reflecting the realities of the twenty-first-century academic experience.

In addition to continual customer interaction, we have used data derived from use of the *Lab* to shape the way it evolves. An example is increasing OCR download—at the launch of the *Lab*, the limit for a content set was 10,000, and researchers could extract up to 1,000 documents in OCR form per session. There are always trade-offs with this kind of development, and for OCR download, it's technological: the need to find a happy balance between size of downloadable content and the time taken to generate the download.

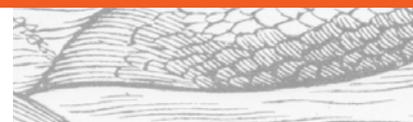
By analyzing usage, we identified that, of the 3,089 content sets created globally in the *Lab* since launch, only 166, or 5.5 percent, exceeded 5,000 documents, so this became the new download limit. We'll certainly revisit at some point, but for the moment this appears to be an ideal balance between utility and performance, based on what customers are telling us with their use.

Gale is committed to enhancing the study not only of digital humanities but of traditionally taught subjects by enabling the integration of digital research concepts. Our aim is to give everyone, regardless of technical skill, a new set of tools and a new language through which they can explore a subject and make discoveries.



Gale Digital Scholar Lab provides our community with the unique opportunity to engage with primary sources in ways they may never have considered before, without needing prior knowledge in coding or having to clean data beforehand. While the *Lab* has great potential for research, we also see it as a great investment for education, as the *Lab* enables us to provide hands-on experience with text-mining models that students can use for their studies and in their life after university.”

— Katrina McAlpine, Associate Director, Publishing and Data Services, University of Sydney, Australia





LEARNING IN THE LAB

MOVING TOWARD DIGITAL SCHOLARSHIP FOR ALL

With the development of *Gale Digital Scholar Lab*, learning takes place within the context of the *Lab* as it facilitates instruction throughout the research process. The *Lab* is a platform that empowers librarians, instructors, and students to improve their teaching, learning, and scholarly research with Learning Center materials. The Learning Center gives users the tools and support they need to understand and employ the vast amount of information and skill sets available through the *Lab*. It includes videos with live walk-throughs conducted by our digital humanities specialists, sample projects, glossaries, FAQs, and much more.

Our Beyond the Lab instructional materials equip users with project-based narratives that model the core critical-thinking skills necessary for ideating around research questions and phenomena and interpreting data and findings. By supporting users “beyond the lab,” we’re joining institutions as they work toward developing growth of inquiry within the digital humanities and beyond.



CHALLENGES PRESENT OPPORTUNITIES

▶ 1

Difficulties accessing and bringing together authoritative content can lead some scholars to spend years collecting, and months preparing, content sets.

▶ **Significantly decrease the time to collect quality content sets.**

▶ 2

Libraries encounter challenges in hosting content and making it available to researchers.

▶ **Uncover a vast array of cloud-hosted archive material.**

▶ 3

Exhaustive possibilities and the volume of tools can make it hard for novice researchers to engage with digital humanities data.

▶ **Offer users less intimidating, more intuitive tools.**



When we began considering the acquisition of *Gale Digital Scholar Lab*, our university was in the middle of forming a minor in digital studies. To support it, we had to give students access to online collections that could serve as a digital sandbox: something they could experiment with, research in, and manipulate digitally. The *Lab* gave us that, in addition to a suite of tools that are readily accessible in an environment with fewer barriers.”

— Hillary Richardson, Coordinator of Undergraduate Research & Information Literacy Librarian, Mississippi University for Women

Gale Digital Scholar Lab saves a lot of time. If you use it in a traditional way, it can be time consuming, but if you utilize the *Lab's* tools, it is much more convenient.”

— Wei Wang, Subject Librarian, School of Social Development and Public Policy, Fudan University

CREATING A SIMPLIFIED EXPERIENCE

The *Gale Digital Scholar Lab* guides users through three intuitive steps that simplify complex workflows and improve outcomes.

STEP 1 BUILD

- To begin, users build a corpus of documents that will provide the raw text data used to generate visualizations.
- Users can search across their library's *Gale Primary Sources* holdings and select documents to add to their custom content set.
- A side-by-side view of OCR text (raw text data) and the original scanned document image helps users quickly determine the accuracy of the OCR text.
- Users upload locally held text files, add metadata in the *Lab*, and create mixed content sets with *Gale Primary Sources* material.

STEP 2 CLEAN

- As a part of *Gale Digital Scholar Lab*, users can create multiple cleaning configurations to tailor how a content set is cleaned, depending on the analysis.
- Testing a user's cleaning configurations is an important step in determining if they're ready to start analysis. In the *Gale Digital Scholar Lab*, users can test their configurations by selecting a content set, and then the first 10 documents will be cleaned with their settings. Next, users will be able to download the original and clean version of those documents to determine if they've been cleaned to the degree needed.
- Once users are happy with their cleaning configuration, they can apply it to their analysis. Users will do this by selecting the clean configuration they want to use in Tool Setup.

STEP 3 ANALYZE

- After selecting documents and creating a custom content set, users can analyze and interrogate the data with the text mining and visualization tools.
- Six commonly used text mining methodologies are included: ngrams, named entity recognition, topic modeling, parts of speech, sentiment analysis, and clustering.
- Clear descriptions and Learning Center documentation support users new to digital humanities. Advanced researchers can customize tool configurations to generate new outputs.



THE LAB AT WORK

BUILD



CLEAN



ANALYZE



REFINE ATTRIBUTES

ADVANCED SEARCH



EVALUATE RESULTS

RESULTS PAGE



EXAMINE OCR

DOCUMENT EXPLORER



**EXPORT OCR CONTENT
FOR EXTERNAL USE**

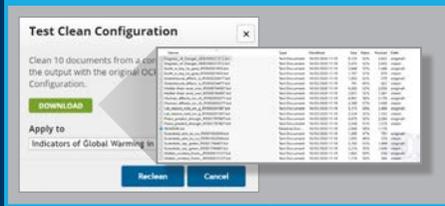
CONFIGURE CLEANING ATTRIBUTES

CLEANING TOOL



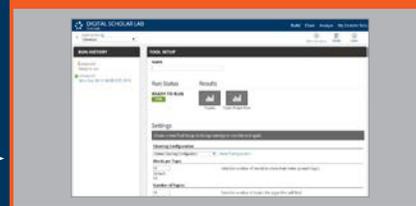
CLEAN OCR DATA

CLEANING TOOL



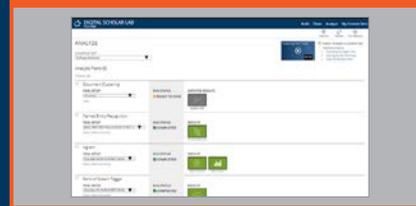
DETERMINE ANALYSIS APPROACH

TOOL CONFIGURATION



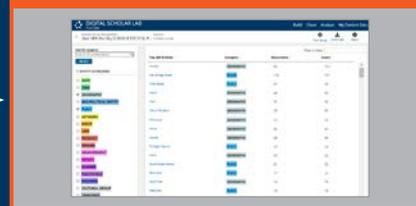
ANALYZE DATA

TOOL LIBRARY



EXAMINE VISUALIZATIONS

VISUALIZATION RESULTS



**EXPORT
ANALYTICAL DATA
& VISUALIZATION**

IDEATE

Discover phenomena  Construct research questions



THOUGHTS FROM Lindsey Gervais, Ph.D., Digital Learning Manager, Gale, a Cengage Company

“Gale has a reputation for empowering learning with high-quality depth and breadth of content and tools that enhance learning opportunities for librarians, instructors, and students. This pairing, between content and effective learning tools, inspires my work in digital pedagogy, providing users with the instructional tools they need to create positive outcomes within their institutions. With the best practices in teaching and learning instilled within the *Gale Digital Scholar Lab's* workflow and Learning Center, we equip users with the tools and scaffolds they need to understand and make use of the knowledge and skill sets required. The skill sets cultivated from the *Lab* are not only essential for success as a scholar, but as digital citizens in our evolving and innovating world.”

THOUGHTS FROM Maggie Waligora, Product Manager, Gale, a Cengage Company

Throughout the development of *Gale Digital Scholar Lab*, we actively partnered with librarians and practitioners in the field of digital humanities to guide our journey. These partnerships informed our alpha, beta, and production versions of the *Lab* and continue to drive the long-term strategy of the platform. We collaborated with Gale's user experience team to test design and functional concepts with over 80 users. These steps helped us make key decisions around all elements of the *Lab*. The conversations led to more than 100 features and enhancements, which can be directly linked to user feedback we have gathered over the past 18 months. As we continue to enhance the overall experience of the *Gale Digital Scholar Lab*, we plan to strengthen our partnerships with current and new users to ensure we are helping them accomplish their core research goals."



Reflect on research questions  Develop inferences

INTERPRET





HEAR FROM YOUR PEERS

HOW THE UNIVERSITY OF ADELAIDE IS DRIVING DEPARTMENTAL CHANGE AND TEACHING DIGITAL HUMANITIES COURSES

Whether on campus or in their online presence, a clear desire to grow, evolve, and improve shines through from the University of Adelaide. This ambition is reflected in results, with Adelaide seeing consistently impressive growth in their global university ranking over the past six years. For Professor Jennifer Clark, head of the department of humanities, “this means a number of new programs, it means different approaches to things, different styles of teaching . . . different ways of exploring the humanities.” As Helen Attar, liaison librarian for the faculty of arts, explains, “*Gale Digital Scholar Lab* will have a significant place in that future.”

Dr. Aaron Humphrey, lecturer in media and digital humanities, explains, “Adelaide is one of the few places in Australia where you can pursue a major or a minor in digital humanities. We are looking to increase that in addition to introducing students to the tools and techniques used in the digital humanities.” For Dr. Humphrey, the benefit of digital humanities is that it “makes it very easy for students to do research that’s never been done before.” Professor Clark sees this benefit as crucial to the aims of the ambitions of the department and the university: “The most interesting history is that which asks questions which haven’t been asked before or explores new material that hasn’t been looked at.”

The *Gale Digital Scholar Lab* is central to the university’s goal of providing new avenues for discovery. As a teacher of digital humanities, Dr. Humphrey is enthusiastic about the *Lab*. “The tools we use to explore are immense. In the classroom, that’s the benefit of the *Gale Digital Scholar Lab* platform—to get students in there and quickly build a dataset, talk about it, look at the results, and then refine that.”

“The most interesting history is that which asks questions which haven’t been asked before or explores new material that hasn’t been looked at.”

— Jennifer Clark, Professor, University of Adelaide



The University of North Carolina at Greensboro is pleased to be leading a collaborative multistate project funded by IMLS [the Institute of Museum and Library Services] entitled the Carolina Digital Library Network (CDLN). In this project, we are evaluating several shared digital library infrastructures, starting with the remarkable new *Gale Digital Scholar Lab*. The *Lab* offers an amazing array of textual pattern analysis tools for scholarly explorations into historical documents. Our project will feature a series of competitions over the next two years by researchers at each of our collaborating universities; cash incentives will be awarded for the winning research projects. Our faculty members are intrigued with the possibilities of digital humanities toolkits such as the *Lab*, not only for their own established research agendas but also because of the next-generation scholarship training that the project will offer to their graduate students. What we like most about the *Lab* is that it integrates advanced textual pattern analysis tools into a simple-to-use interface that can be learned quickly by scholars from virtually any discipline.”

— Martin Halbert, Professor at the University of North Carolina – Greensboro

Ready incorporation into his workflow is key. Dr. Humphrey adds, “We have 10 weeks of teaching time. I do not have time to teach students command line coding. To get them to do topic modeling in Python, that’s going to take more time than I have in a class. I like how the *Gale Digital Scholar Lab* makes this kind of analysis very easy to do, allowing us to get robust and compelling results quickly enough to be able to dissect them in a tutorial session.”

Professor Clark sees strategic benefits. “We want to provide, through digital humanities, opportunities to bridge those two worlds of the sciences and the arts and humanities,” she explains. There is also an advantage for the department to attract students. “We can say to them, ‘You can still do the arts and humanities, but you can also get these [technical] skills.’” Attar concurs. “The biggest growth in jobs is in data science. Anybody who has skills in data management, data manipulation, data visualization, or data wrangling is going to be well sought after,” she says. Dr. Humphrey agrees with the benefits to students: “Some of the most vital work in digital technologies is figuring out where disciplines cross over and how we can create students who are technically literate and who have an understanding of the human heart . . . The possibilities are almost unimaginably full of potential.”



ADDITIONAL INSIGHTS FROM OUR TEAM

THOUGHTS FROM Wendy Kurtz, Ph.D., Gale Digital Humanities Specialist and UCLA Lecturer



As a researcher and instructor in the digital humanities, I am a strong believer in the *Gale Digital Scholar Lab* because it exposes the complex nature of the workflow for a text-mining project and facilitates a user through the various steps. By aligning dataset building with the analytical tools to analyze and visualize texts, the processes highlighted in the *Lab* encourage users to learn through the project-building process.

As one of the digital humanities specialists at Gale, I help promote and facilitate use of the *Lab* by librarians, faculty, and students. Working with dozens of academic institutions since its launch in fall 2018 has reinforced the fact that the functionality of the platform offers immense value for a broad spectrum of users—from novice to seasoned—and institutional personas.”

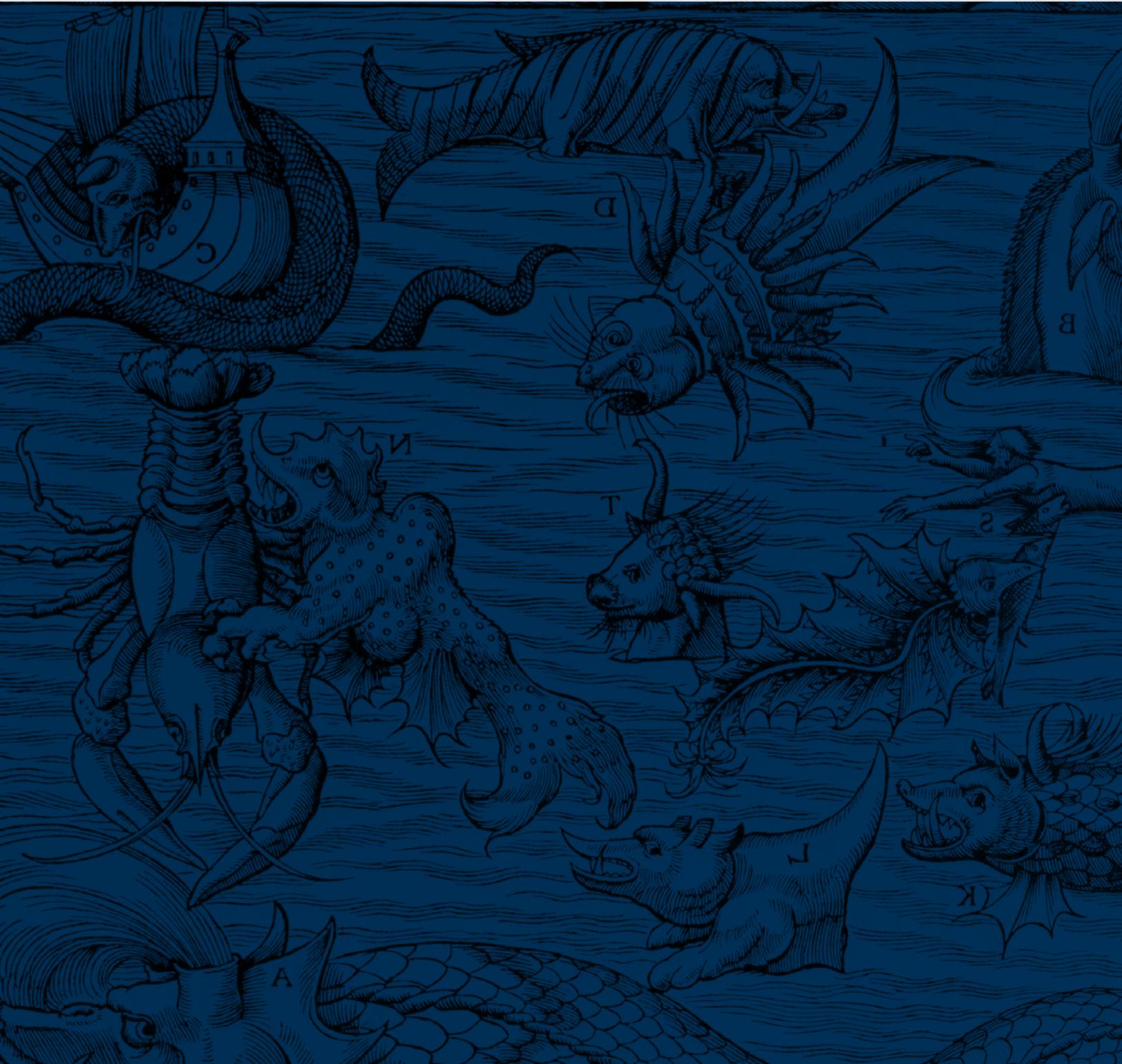
THOUGHTS FROM
Sarah Ketchley, Ph.D.,
Gale Digital Humanities Specialist
and University of Washington Lecturer

As a faculty member at the University of Washington, I regularly teach students skill- and research-based classes in the digital humanities. In choosing tools to use for data curation and text analysis in the classroom, I look for those that are straightforward to set up and manage and can be used effectively by all students—no matter what their experience level. I have used the *Gale Digital Scholar Lab* as the primary digital tool in three classes I have taught: two in-person introductory sessions and one fully online class. In each case, the cloud-based *Lab* provided an accessible entry point into the process of working with primary source data, building data sets, cleaning OCR text, and performing text analysis based on research questions the students developed. Students reported that they found the platform intuitive to navigate, with supplementary video and text tutorials to support classroom learning. Students were able to export their visualizations, raw analysis data, and relevant primary source images to work with and embed in external tools, including OpenRefine, Lexos, Voyant, Omeka, and StoryMapJS.”





GALE DIGITAL SCHOLAR LAB



**LEARN
MORE**



800.877.GALE



gale.com/GaleDigitalScholarLab

