

# Gale Digital Scholar Lab updates

We are constantly working to develop and improve *Gale Digital Scholar Lab*, based on feedback from users, to best meet the needs of researchers, students, librarians and educators.

Gale strives to uphold values of openness and transparency that are core to the digital humanities, so we are happy to provide a list of the updates made to *Gale Digital Scholar Lab* below. Those at the top are most recent.

## Contents

Colour palette options for visualizations.....	2
'Introduction to Digital Humanities' course .....	2
Ngrams Start Word Lists .....	3
Parts of Speech Pie Chart visualization .....	4
Personalized Lexicons for Sentiment Analysis .....	4
Marked-Up OCR view added .....	4
Sentiment by Timeframe visualization added .....	5
Homepage development .....	6
Sentiment Analysis AFINN Lexicon update.....	6
Tool Algorithm updates.....	6
Visualization updates .....	7
Gale Research Showcase .....	7
Python Notebooks added .....	7
Learning Center update.....	8
My Research dashboard added.....	8
Projects feature added .....	9
Support for Shibboleth and IP Authentication.....	10
Highcharts Visualization update.....	11
Content Upload functionality updates .....	11
More Datasets added .....	12
Group Notebook feature added.....	13
Group Workspaces added .....	14



## Colour palette options for visualizations

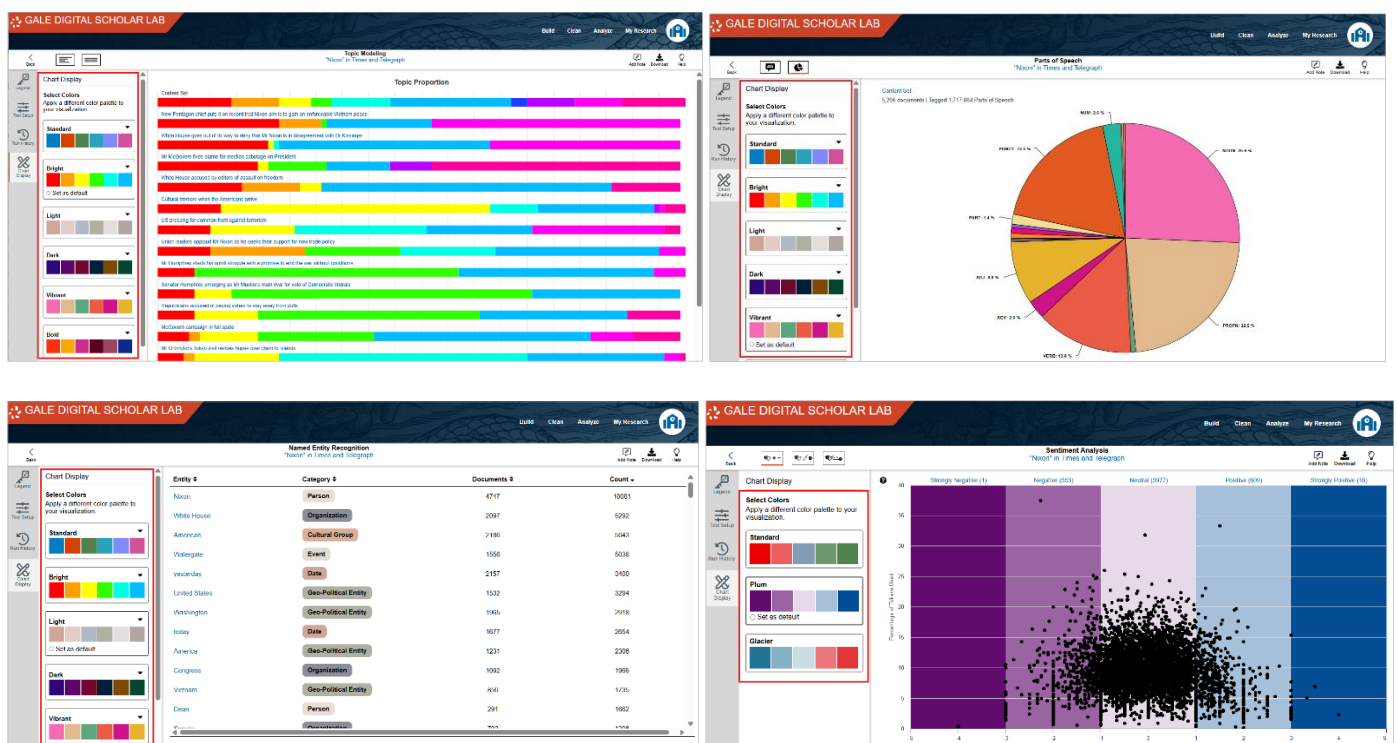
[June 2025]

This feature allows users to change the colours of their visualizations, according to preference or to better fit with their presentations or reports. Once a visualization is run, users can click on the 'Chart Display' icon and easily change the palette according to their preference.

- For the Document Clustering, Named Entity Recognition, Parts of Speech, and Topic Modeling tools, there are seven colour palettes available: **Standard, Bright, Light, Dark, Vibrant, Bold, and Greyscale.**
- For the Ngrams and Sentiment Analysis tools, there are three colour palettes: **Standard, Plum, and Glacier.**

In Group Workspaces, any member in the group can access the palettes, even if another user created that visualization, which allows for a lot of flexibility in this collaborative setting.

This was one of the most frequently requested updates we've received, as users wanted to tailor the colours used in visualizations to their presentations/reports. By introducing this feature into the *Lab*, we meet user's desire for more customisability and control over the presentation of their research.



## 'Introduction to Digital Humanities' course

[April 2025]

Responding to demand for digital humanities curriculum and teaching support, we added a full, flexible DH course to the *Gale Digital Scholar Lab* Learning Center in April 2025.



Written by digital humanities and pedagogy experts, utilising recognised learning frameworks, and with input from practising instructors, users can be assured of the quality of the resources in this course.

Built around both *Gale Primary Sources* and *Gale Digital Scholar Lab*, the course includes six modules that increase students' understanding of digital archives, text and data mining primary sources, and how to create and publish digital humanities projects:

1. Using Digital Archives
2. Mining Text and Data
3. Building Content Sets
4. Cleaning Data
5. Analyzing Data
6. Creating and Publishing Projects

Each module includes assignable reading, discussion questions, hands-on activities, and ideas for extended activities. Instructor guides help educators navigate the materials.

Highly flexible, educators can integrate the complete course into their teaching, or select one or more modules or activities for use in their curriculum, saving time whilst ensuring a high-quality classroom experience for their learners.

[Learn More >>](#)

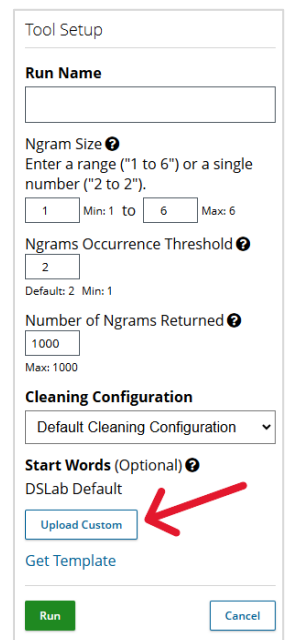
## Ngrams Start Word Lists

[February 2025]

Previously the Ngram Tool could visualize which words were most frequently found in the dataset. Users could remove words from the analysis using a Stop Words list, but not indicate if there were words they particularly wanted to focus on.

This update allows users to upload a list of words that they are interested in, and then have the tool return the frequencies for only those terms, if they appear in the content set. Thus users can create an Ngram visualization for a list of words particularly relevant to their research.

This does not replace the previous Ngram functionality, it provides an additional option for users.



Tool Setup


**Run Name**

**Ngram Size** ⓘ  
Enter a range ("1 to 6") or a single number ("2 to 2").  
1 Min: 1 to 6 Max: 6

**Ngrams Occurrence Threshold** ⓘ  
2  
Default: 2 Min: 1

**Number of Ngrams Returned** ⓘ  
1000  
Max: 1000

**Cleaning Configuration**  
Default Cleaning Configuration ▾

**Start Words (Optional)** ⓘ  
DSL Lab Default  
[Upload Custom](#)   
[Get Template](#)

[Run](#) [Cancel](#)

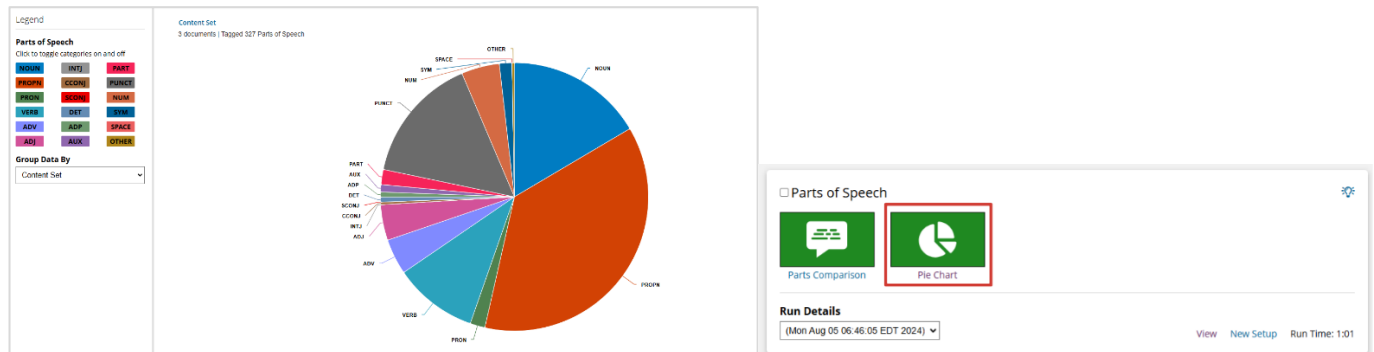


## Parts of Speech Pie Chart visualization

[February 2025]

Pie charts have been added as a second visualization option to the Parts of Speech tool, joining the existing bar graphs visualization, allowing users to see their data in different ways.

Pie charts are particularly useful for quickly conveying how large a portion certain word categories may take up (for example, what percentage of a work is nouns).



## Personalized Lexicons for Sentiment Analysis

[February 2025]

Previously, the Sentiment Analysis tool in *Gale Digital Scholar Lab* used the AFINN lexicon (simply, a list of analysed words). Whilst frequently used in the world of Digital Humanities research, the AFINN lexicon is limited because it was built around the language used in social media, making it limited in the analysis of historical documents.

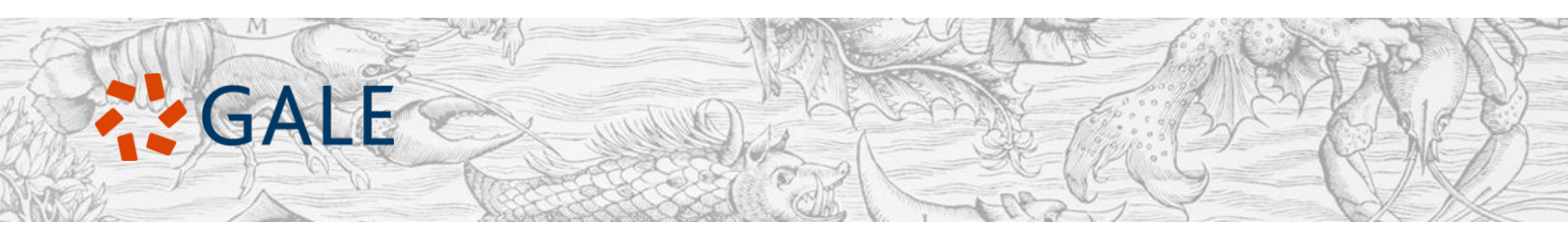
This update enables users to upload their own lexicons to the Sentiment Analysis tool. This allows users to tailor their lexicons to the materials they're researching, using their own subject expertise. An additional benefit is that it allows the creation of non-English language lexicons, and thus analysis of non-English language texts within the Sentiment Analysis tool. We have included support for accented characters to facilitate this as much as possible. (Please note it has only been tested on Latin alphabets).

This development was one of the most requested by users. We're thrilled to be able to better support the needs of users with this update.

## Marked-Up OCR view added

[October 2024]

**Parts of Speech** and **Sentiment Analysis** tools have a new view that allows users to see how the words in their documents have been identified by the tools. This is done via marking up the OCR\*. In **Parts of Speech**, the mark-up indicates how each word has been categorised (for example as a noun or adjective), and in the **Sentiment Analysis** tool, it shows the sentiment of scored words.





Please note, available Timeframes are based on the metadata available for the documents being analyzed – if none of your documents include a publication month, then the month option will not appear for that year.

---

## Homepage development

[July 2024]

Homepage redevelopments were made, giving users a cleaner and more intuitive experience, highlighting key features and research routes the user can take. Updates included:

- **Adding a Search Bar** – often the first thing a user looks for, to immediately get started!
  - **Four large buttons indicating routes to get started:** *Build, Clean, Analyze, My Research*
  - **Gale Research Showcase carousel** – a rotating display, encouraging users to explore the Showcase for more examples and inspiration!
  - **Learning Center banner**, flagging this key source of 24/7 support, including highlighting areas of particular use, and some of the questions it can answer.
- 

## Sentiment Analysis AFINN Lexicon update

[July 2024]

Sentiment Analysis tool migrated from using the 111 version of the AFINN lexicon, to the 165 version of the AFINN lexicon, which includes an additional 905 scored words/tokens.

Please note analysis runs made before the update are no longer directly replicable.

---

## Tool Algorithm updates

[July 2024]

The open-source algorithms that power the following tools were updated:

- **Named Entity Recognition** and **Parts of Speech** both use spaCy, an open-source natural language processing library. Both tools were updated to using version 3.7.4 of spaCy.
- **Document Clustering** uses scikit-learn, which is a machine learning library. The Document Clustering tool was updated to using version 1.4.0 of scikit-learn.

Thanks to the updates to these two Python-based software libraries, users will notice improved performance when using the tools.



## Visualization updates

[June 2024]

A number of accessibility and styling updates were made, including:

- Improving text colours and mark-up view for Named Entity Recognition
- Enabling the user to download either the top 100 or top 25 results for Ngrams
- Adding a Tooltip to better explain functionality for Sentiment Analysis.

The ability to zoom for the Sentiment Over Time visualization was also introduced, so the user can make the visualization larger and explore it in more detail.

Lastly, the Word Cloud visualization for Ngrams was overhauled by moving to Highcharts. This brought it in line with our other visualizations, reduced white space and improved the readability of the words themselves. The results of an analysis were unchanged.

---

## Gale Research Showcase

[April 2024]

In April 2024 we launched Gale Research Showcase, a free, open repository of student-authored digital scholarship. Designed to advance early career research, it showcases high-quality, peer-reviewed essays that demonstrate best practices in digital scholarship.

Students can use Gale Research Showcase to get inspiration and guidance on using Digital Humanities techniques in their own projects – and to get published!

Publishing in Gale Research Showcase is a great opportunity for early-career researchers to experience the Peer Review process, and get free feedback, support and suggestions from Gale's DH experts on their DH project.

Gale Research Showcase is a great tool for faculty, librarians and educators to use when developing a DH curriculum, providing a hub of research examples to enhance instruction and engage and motivate students.

[View Showcase >>](#)

---

## Python Notebooks added

[March 2024]

Three Python Notebooks are now available in the Learning Center. They offer users a way to take their research further and analyse their data using annotated code blocks that can be run using an interactive coding platform like JupyterLab or Google CoLab.

The Python Notebooks offer flexible customisation options, such as the ability to process downloaded Datasets from within the Lab, or to run the user's own datasets.



The three Python Notebooks are:

- Named Entity Recognition
- Geographical Information System (GIS)
- Sentiment Analysis

#### Python Notebook 1: Named Entity Recognition

[Get a copy of Python Notebook 1: NER](#)

This Python Notebook provides instructions for cleaning and running a Named Entity Recognition (NER) analysis on the [Suffragettes Dataset](#). The framework will guide you through the steps of adding the necessary open source libraries to run the analysis, including SpaCy for analysis, the pandas DataFrame for structuring analysis output, and Counter to count occurrences of selected textual features. The analysis focuses on identifying and listing Geopolitical Entities (GPE), but the code can be adapted for element from SpaCy's NER model. The top 500 results of this Notebook's analysis can be downloaded in CSV format.

#### Python Notebook 2: Geographic Information System (GIS)

[Get a copy of Python Notebook 2: GIS](#)

This Plotting GIS Data notebook builds on the work completed on Named Entity Recognition in Notebook 1. Focusing again on the [Suffragettes Dataset](#), and in addition to SpaCy, pandas, and the Python Counter, the annotated code provides instructions on how to geocode a list of place names to generate GIS data to plot on a map. Geopy is used for geocoding, then the data is mapped using IpyLeaflet. Additional configuration options are discussed, including adding a Zoom button and saving and exporting files as static HTML, CSV, and GeoJSON.

#### Python Notebook 3: Sentiment Analysis

[Get a copy of Python Notebook 3: Sentiment Analysis](#)

The Sentiment Analysis Python Notebook offers a guided exploration of sentiment analysis using the AFINN lexicon, which is also used in Gale Digital Scholar Lab's implementation of the tool. The [Roberto Calvi dataset](#) is used to model the workflow, but the code can be run on any collection of text data. The Notebook provides instruction on how to run and adapt an AFINN analysis using python, while generating comparative output using VADER, which is another popular tool for sentiment analysis. The code also uses pandas and matplotlib for generating tables and graphs, and analysis results are downloadable and adaptable.

---

## Learning Center update

[March 2024]

A full reorganisation of the existing material makes it easier to navigate, with a particular focus on quickly getting users to the information they need. This includes a section for 'My Research', to support the new dashboard, and a 'Getting Started' section that directs users to resources that can help them start analysing straight away.

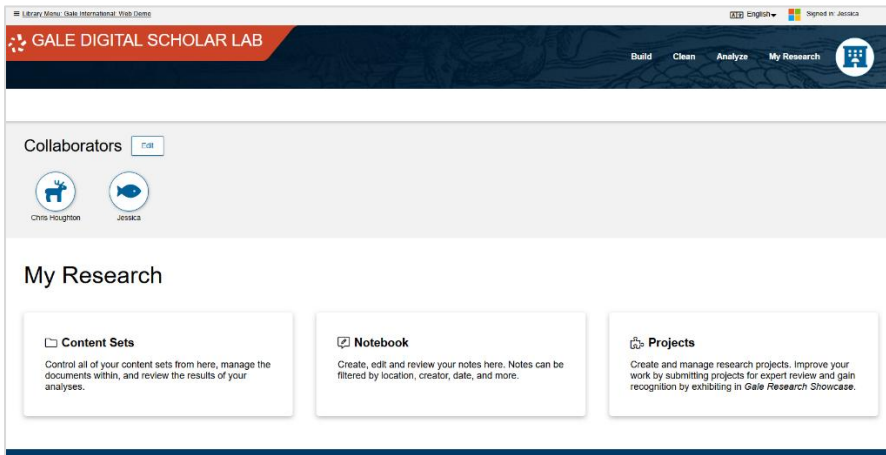
---

## My Research dashboard added

[March 2024]

As part of the introduction of the Projects feature, Content Sets and Notebook have been moved to the new 'My Research' dashboard, where they sit alongside Projects. This groups together all the user's personal materials into one hub, making it easy for researchers to organise and locate the materials that are central to their research, while moving seamlessly between the different aspects of creating a research project.





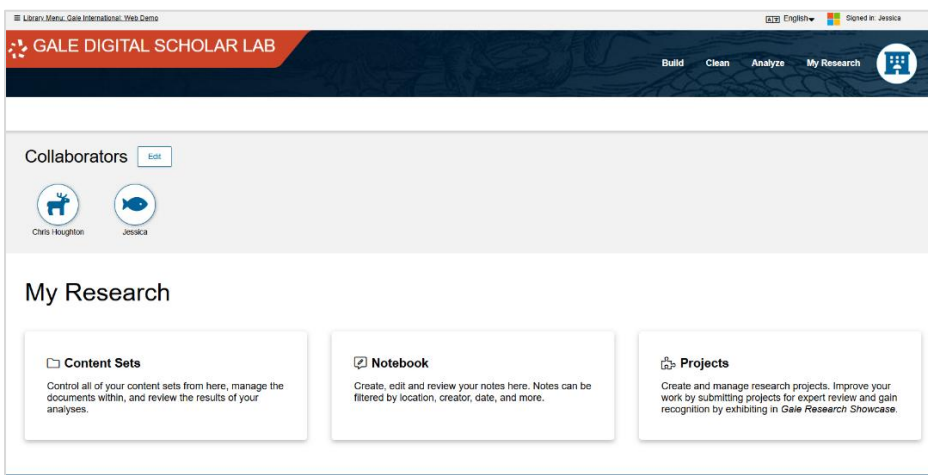
## Projects feature added

[March 2024]

The new Projects feature allows users to take copy materials from their Notebook, and elsewhere in the *Lab*, and thoughtfully compile them together, along with newly written text, into a formal report. This can then be submitted to Gale Research Showcase (see update above), or downloaded for use elsewhere, such as submitting an assignment or creating a research paper.

The Projects feature is found within the 'My Research' dashboard (see update above). Working in tandem with existing *Lab* features such as the Notebook, Projects allows the researcher to undertake their entire research workflow within the platform.

Both Personal and Group Workspaces support the Projects feature, meaning groups of researchers can work on a project simultaneously within a Group Workspace. Multiple Projects can also be created within the same workspace.



**My New Project**  
This is an abstract.

- Introduction**  
This is a test project. First we start with an introduction, where you can explain what your project is about. This section can be as long as you want it to be. Let's get started!
- Starting A Project**  
You can make a project in lots of different ways. If you've used the Notebook to gather and record your thought process while you make content sets and analyse them, you can add those notes to a project. Do this from the Notebook using the 'Add to Project' button, and then your notes will appear in the Notebook. You can then rearrange and edit them however you'd like, as you put together your formal report, without losing the original notes themselves. You can also add additional information, as well as use the section and header features to organise your project.
- Adding Visualizations**
- How to Capture Visualizations**  
You can capture a visualization in the 'Add Note' dialogue as you go. Once it's saved to your notebook, you can send it to your project.

## Support for Shibboleth and IP Authentication [October 2023]

When accessing resources through an institution, users are first required to **authenticate**, to confirm they're at the institution. With *Gale Digital Scholar Lab*, they then also need to **login** to their personal account, to see their own Content Sets, cleaning configurations and analyses.

Users can **authenticate** in different ways – automatically via IP address/ EZProxy, or via Shibboleth, which requires them to input their university credentials.

This update introduces a new authentication page for the *Lab* that allows users to select how they want to authenticate – allowing them to bypass automatic IP address/ EZProxy authentication for this product only. This will enable institutions to use IP address/ EZProxy authentication for *other* Gale products, while allowing *Lab* users to access the university credentials login option.

Please note, this new page will display only for institutions that have *both* Shibboleth and IP authentication enabled.

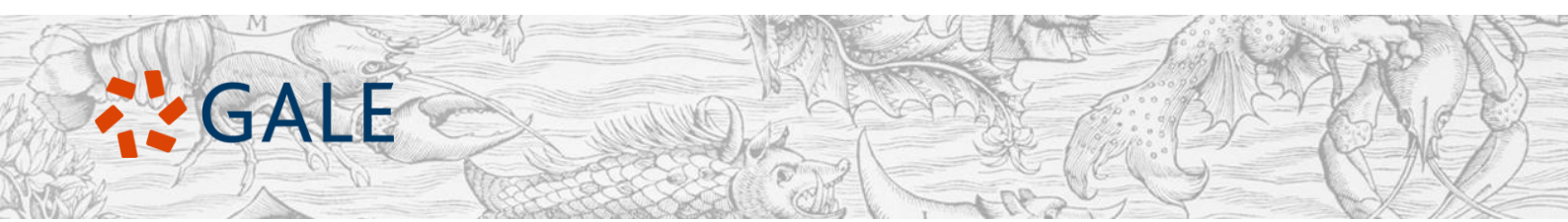
Further information in [this FAQ](#).

**GALE DIGITAL SCHOLAR LAB**

Choose your sign in method

- Institution Credentials** (Recommended)  
Sign in with your university or institution credentials. You will be brought to your institution's login page to authenticate.
- On-Campus**  
Only available if you're on campus. You will be go to the DS Lab to authenticate through your institution, Microsoft or Google.

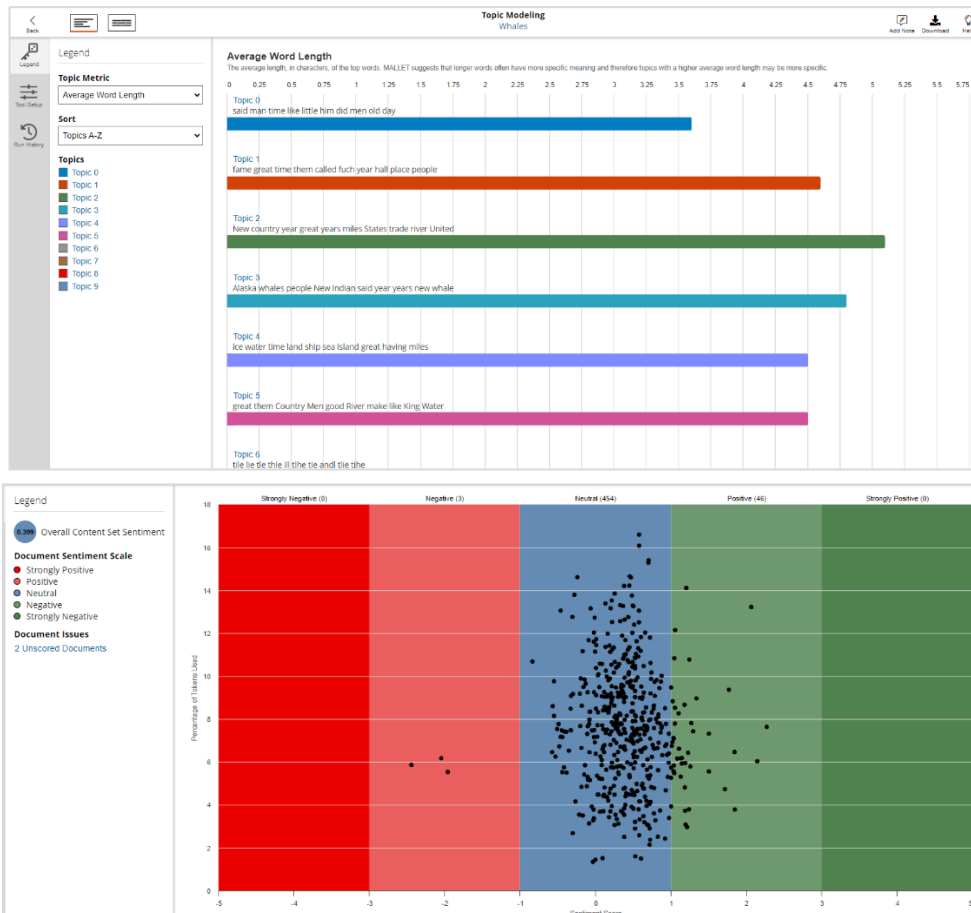
**GALE** A Cengage Company | ABOUT | HELP | CONTACT US | TERMS OF USE | PRIVACY | ACCESSIBILITY | END SESSION | Gale, here for everyone.



## Highcharts Visualization update

[August 2023]

Highcharts underpins visualizations in multiple Gale products, including the analyses in *Gale Digital Scholar Lab*. This update brings the most up-to-date version of Highcharts to the platform, providing new features to enhance the visualizations and fixing a variety of bugs. Alongside the update, we adjusted the colours of the visualizations to be fully accessible against both white and black backgrounds.



## Content Upload functionality updates

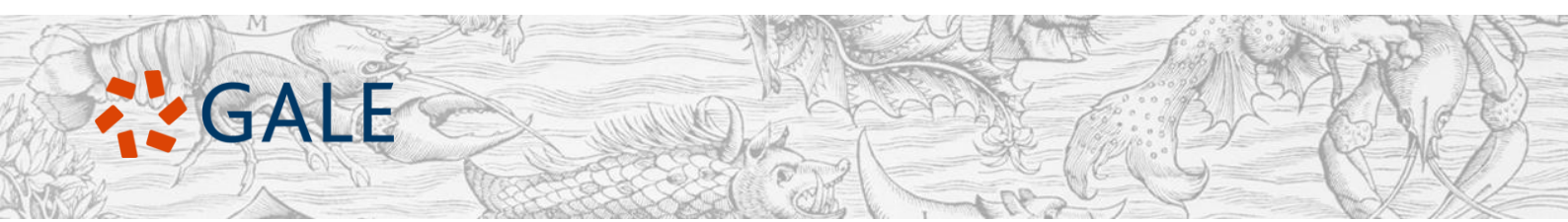
[May 2023]

### New Upload Format

Users can now upload documents via .csv file, through which the text and metadata for multiple documents can be uploaded in one go. This functionality works alongside the existing .txt file upload option.

### Workflow changes

When a user uploads a document, they are now rerouted to the 'Manage All Uploads' page, where they can add metadata and fully interact with their new documents. This creates more transparency and makes it easier to manage uploaded documents.



**Build**  
Build Content Sets by finding documents in your institution's Gale Primary Sources holdings or by uploading your own documents.

**Search**  
Search by keyword

**Upload**  
Drag and Drop .txt or .csv Files here or [Browse](#)

[Manage All Uploads](#)

**Manage All Uploads**  
Add your own documents, with metadata, for use in the Digital Scholar Lab.

Date Uploaded	Title	Owner	Published	Author	Publisher	Publication Title	Document Type	Subject	Content Set
<input type="checkbox"/>	8/6/2024	SA Test							SA Word Clo
<input type="checkbox"/>	7/17/2023	Hausarbeit_Despot*Or kett* 1650-1750Witchcraft in Popular CultureWitchcraft Cluster 1Witchcraft - Harry Potter							
<input type="checkbox"/>	7/17/2023	Alisigh* faint* NOT sigh* 1680-1760sigh* faint* NOT sigh* 1761-1780sigh* faint* NOT sigh*						Historic England Project	

## New Metadata columns

Two new metadata columns have been added to the 'Manage All Uploads' page:

- Content Set (indicates all the Content Sets a document is included in)
- Date Uploaded

Users can hide columns as needed. Each column also has its own filter.

## More Datasets added

[May 2023]

Datasets are small collections of carefully chosen documents that focus on a specific topic and allow users to quickly jump into using the tools within *Gale Digital Scholar Lab*, right from the get-go. They can be easily added to the user's My Content Sets dashboard using the 'Get a copy' button.

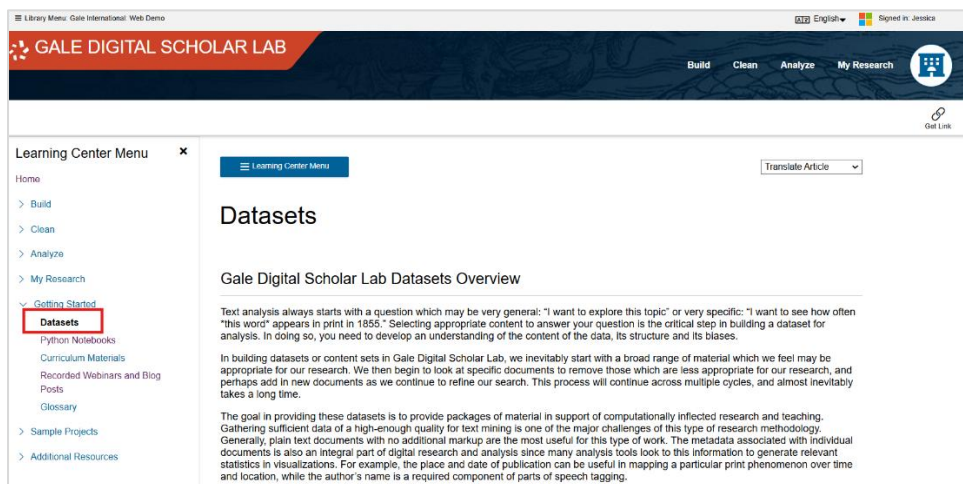
The four new datasets are called:

- Indigenous Leaders 1
- The Campaign Against the Contagious Diseases Act
- The Triangle Shirtwaist Fire
- The Battle of Cable Street



They were created in direct response to user feedback, which indicated that additional example materials were in demand. Instructors requested more example datasets for teaching, and newer researchers suggested additional datasets would help them to better conceptualise how they might apply the *Lab* to their own research.

These new datasets were curated by our Senior Digital Humanities Specialist, with topics suggested by the Gale Digital Humanities User Engagement Group. All datasets are available in both the Personal and Group workspaces, enabling them to be used in both solo and collaborative research. Datasets differ from Sample Projects. Sample Projects are full projects that guide the user from content curation to analysis; datasets are smaller collections that have not been processed or cleaned, and do not include example analyses.



The screenshot shows the Gale Digital Scholar Lab interface. At the top, there is a navigation bar with 'Build', 'Clean', 'Analyze', and 'My Research' buttons. Below this is a 'Learning Center Menu' on the left, which includes 'Home', 'Build', 'Clean', 'Analyze', 'My Research', 'Getting Started', 'Datasets' (highlighted with a red box), 'Python Notebooks', 'Curriculum Materials', 'Recorded Webinars and Blog Posts', 'Glossary', 'Sample Projects', and 'Additional Resources'. The main content area is titled 'Datasets' and contains a 'Gale Digital Scholar Lab Datasets Overview' section. The text explains that text analysis always starts with a question, either very general or very specific, and that selecting appropriate content is the critical step in building a dataset for analysis. It also mentions that in building datasets or content sets in Gale Digital Scholar Lab, they start with a broad range of material and then begin to look at specific documents to remove those which are less appropriate for their research.

Please note, four further datasets have been released since this update:

- *Indigenous Leaders Part 2*
- *Sea Serpents*
- *Sarah and Angelina Grimke*
- *Female Stunt Reporters*

## Group Notebook feature added

[December 2022]

Released at the same time as Group Workspaces, the Group Notebook is a collaborative note-taking tool which allows students and researchers to communicate, ideate, and iterate on their projects together, helping to develop good research methodologies, as well as project management and report writing skills.

Significantly it allows students to create collaborative notes on content sets, clean configurations, and visualizations – *from directly within those features*, thus not disrupting their workflow. Users can also add images and links to their notes. The feature also records information such as the author of each note and date it was added, allowing group member contributions to be tracked.

The Group Notebook can also be used by educators to provide students with support materials, resources, and feedback to guide project-based learning.



Later updates included:

- A Sort feature, making it easier to navigate the Notebook.
- Additional Export options, allowing users to download notes as a CSV, Word or PDF file.
- Pan-Zoom functionality for images in the Notebook, allowing users to interact with notes more effectively.

[Download Factsheet >>](#)

---

## Group Workspaces added

[December 2022]

Released at the same time as Group Notebooks, the Group Workspace feature was introduced to allow users to view other group members' content, build content sets collaboratively, and share visualizations between group members.

This functionality helps students and researchers develop project management and collaboration skills. It also allows educators to oversee their class's work by joining each group, meaning they can offer support when needed and keep projects on track.

Collaboration is a huge part of digital humanities work, and this update was in direct response to extensive user requests.

Later updates included:

- The ability to select avatars for each collaborator, and for the group itself.
- Replicating the Merge, Move and Duplicate Content Set functions in the Group Workspace, aligning it with the functionality of the Personal Workspace.

Collaborators must all be at the same institution, due to rights limitations.

[Download Factsheet >>](#)

