

# Exploring the Practicalities and Processes of Developing a Collaborative Group Space in a Platform for Text Mining: Gale Digital Scholar Lab

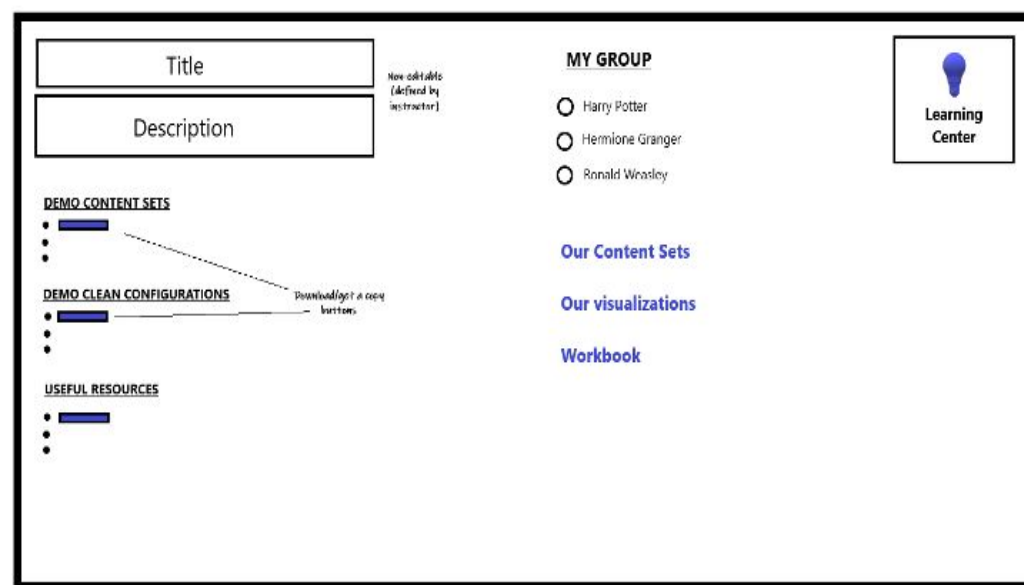
Rebecca Bowden & Sarah L. Ketchley

## OVERVIEW

This poster provides a summary of the work completed to build out two new features in *Gale Digital Scholar Lab* to support collaborative work by users of the *Lab*, and the creation of contextual notes and documentation to capture the detailed work undertaken during the process of text mining and analysis. The *Lab* was released in September 2018 and has been continually refined since then by the dedicated product and development team at Gale. This process has included a platform migration, the addition of text cleaning, document upload, a Learning Center to support new users, sample projects and pre-curated datasets. The *Lab* is fully integrated with *Gale Primary Sources (GPS)*, providing access to millions of pages of OCR text generated from underlying primary source material, based on institutional holdings. These digital archives grow in number every year; the *Lab* offers streamlined processes ('Build, Clean, Analyze') for interrogating the content sets created by users within the platform. Librarians, faculty and students engage with the *Lab* primarily for research and teaching, and the new Groups and Notebook features are intended to support both individual users as well as collaborative teams.

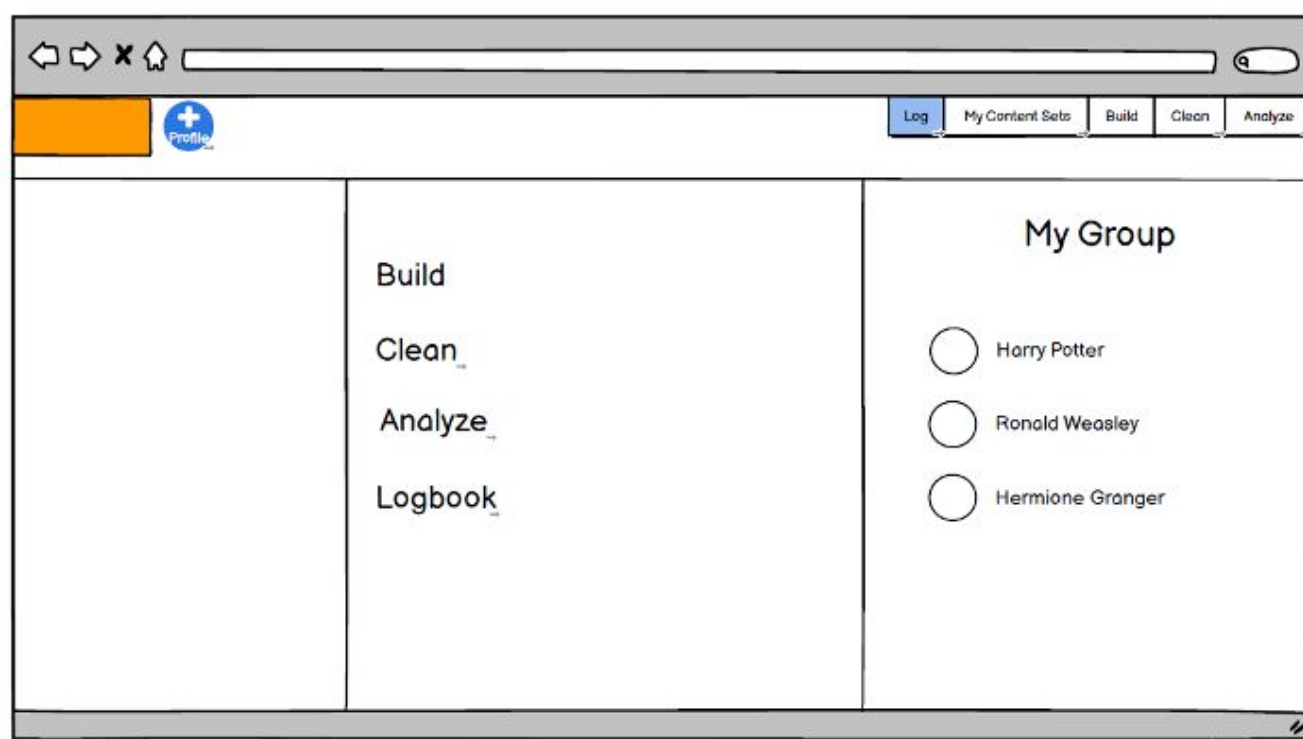
## PHASE 1: Preliminary Research

All *Gale Digital Scholar Lab* projects start with uncovering barriers to entry, and improving user experience. The Groups project was a response to one of our most requested features: collaboration within the *Lab*, especially in a classroom environment. Building on an earlier, descope project, we started by reviewing the original designs with our subject matter expert, Sarah Ketchley, to establish which aspects were still relevant. We supplemented this analysis with research into other products in the market and also conducted interviews with disciplinary researchers. We defined our key user requirements and drafted preliminary wireframes to establish the feasibility of the project. Key decisions at this stage included a move away from our previous LMS approach towards an environment to support synchronous and asynchronous working for classroom groups, as well as identifying the need for a record keeping solution to encourage good project management.



## PHASE 2: Early Design

With project feasibility confirmed, we incorporated feedback from Sarah and our software engineers into more detailed wireframes. These covered multiple aspects of the new feature, including the ability to change group; displaying a creator for content sets and analyses to emphasise authorship; and the first concepts for an accompanying 'Logbook' that would allow users to make notes on their project. At this stage we undertook conversational interviews with instructors from the UK, USA and Asia to understand how this feature would be integrated into different classroom environments and to ensure the concept was on track. These enlightening conversations enabled us to refine the designs and use cases for Groups. Significant takeaways included defining group sizes of between 2-30 students; the role of the 'Logbook' in grading and assessment, and the accompanying desire for it to be fully exportable; further emphasis on accurate authorship at all stages of classroom projects so instructors could understand a student's contribution; and the need for strong guardrails to discourage accidental or malicious deletion. At this stage we also explored a notification feature, ultimately descopeing it based on feedback from interviewees. New requirements identified during this process included avatars to differentiate users and groups, and a need to change the name 'Logbook', as the term was not consistently understood. From these conversations we were confident in the need for the feature and created a roadmap for the project, focusing on key milestones: developing the shared workspaces; developing the 'Logbook' and adding any finishing features.



## PHASE 3: UX Development

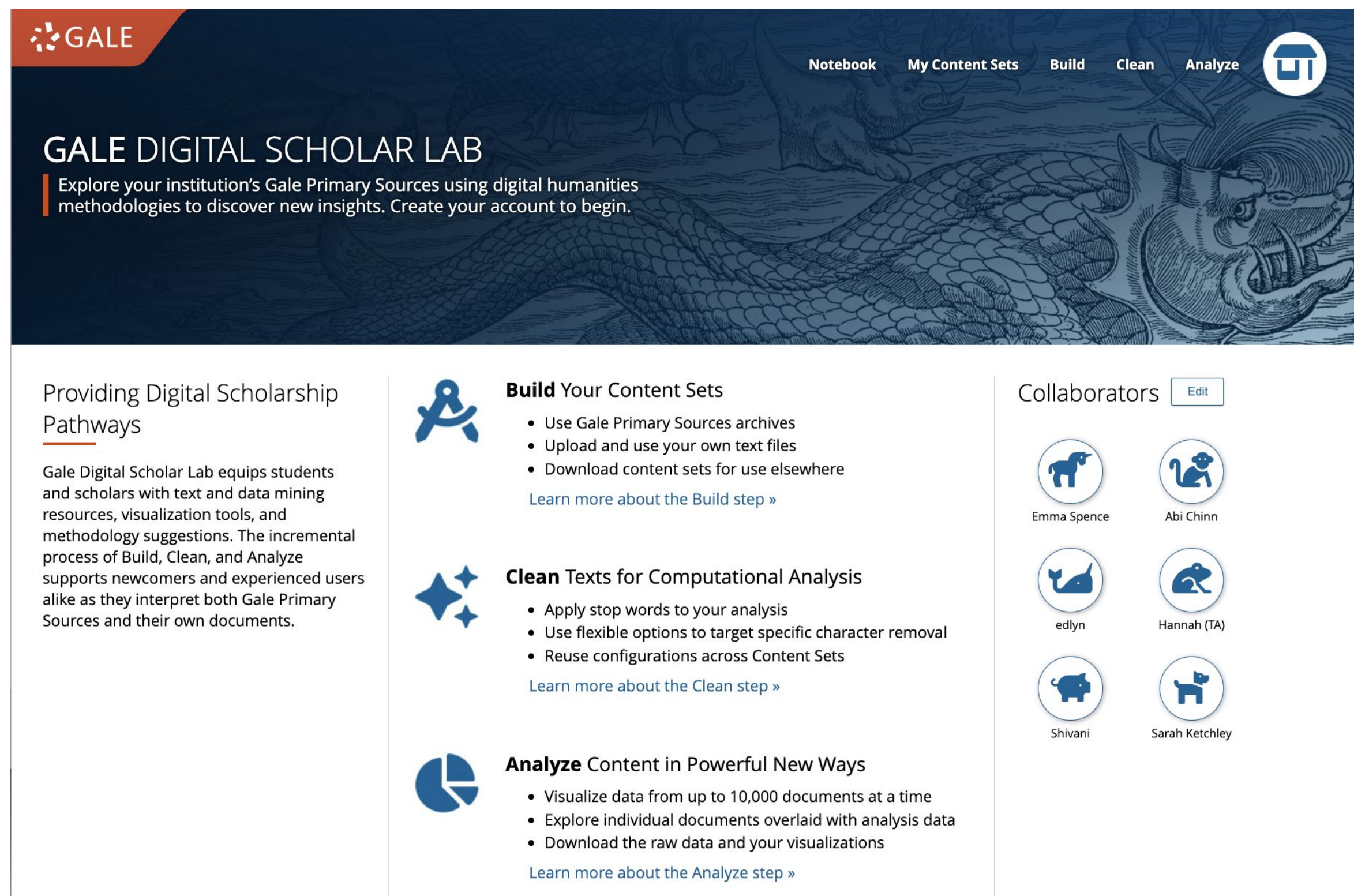
Having established our requirements, we created high-fidelity prototypes. These were fully interactive and included multiple options to test button locations, aesthetic design and workflows, allowing us to refine the details of the feature. We focused on three particular areas: how to show the current workspace; the 'Add Note' modal; and a design for the newly named 'Workbook' (previously 'Logbook'). Working closely with our UX designer, Cara Suriyamongkol, we ran observational testing exercises with Gale's *Digital Humanities User Engagement Group*. This group consisted of 7 instructors, librarians and PhD students who worked closely with us over the course of a year to provide continued feedback on product design and development. Our usability tests indicated that there was confusion around the purpose of the notes and with connecting them to the 'Workbook'. This led us to iterate on name for a third time, with users guiding us towards 'Notebook'. Our observations also prompted a change to the location of the group avatar, making it more intuitive for users to find, and to add multiple download formats for the Notebook feature.

## PHASE 4: Agile Implementation

From here, our prototypes were finalized and development got underway. As our software engineers began work we realised that some of our planned approaches would not work in practice. Privacy was a consideration that needed particular attention: the importance of assigning authorship was emphasized during user testing, yet storing user's full names and emails raised concerns around personal data. The solution was to keep users' personal information anonymous to the platform and instead allow them to actively select an avatar and username that would be associated with their anonymized ID and then displayed within Groups to assign authorship. We also encountered difficulty with our plan to add images to notes in the Notebook - this functionality was discovered to be a security risk but discussion with Sarah further emphasized images as an essential component. Ultimately, we altered the design, allowing users to capture images of visualisations from directly within the *Lab*. These examples represent two of many aspects of Groups that relied heavily on cross-team collaboration and iteration to reach the design that you can see in the platform today.

## FUTURE ROADMAP

*Gale Digital Scholar Lab* is continuously developing, with an exciting roadmap planned for the next year. We're continuously iterating on the Groups feature as we receive feedback from our users, including adding tooltips to further support new users, making changes to the Notebook to enhance collaborative workflows (for example, adding pins to notes), and addressing problems around using uploaded documents in the Groups workspace. Additionally, we are exploring Jupyter Notebooks to extend the analysis possibilities of work done in the *Lab*, with the goal of providing sample code to engage students and support instructors who want to stretch their students in the classroom. We're also working on a significant project to showcase projects done in the *Lab* and celebrate the success of the students and researchers undertaking it.



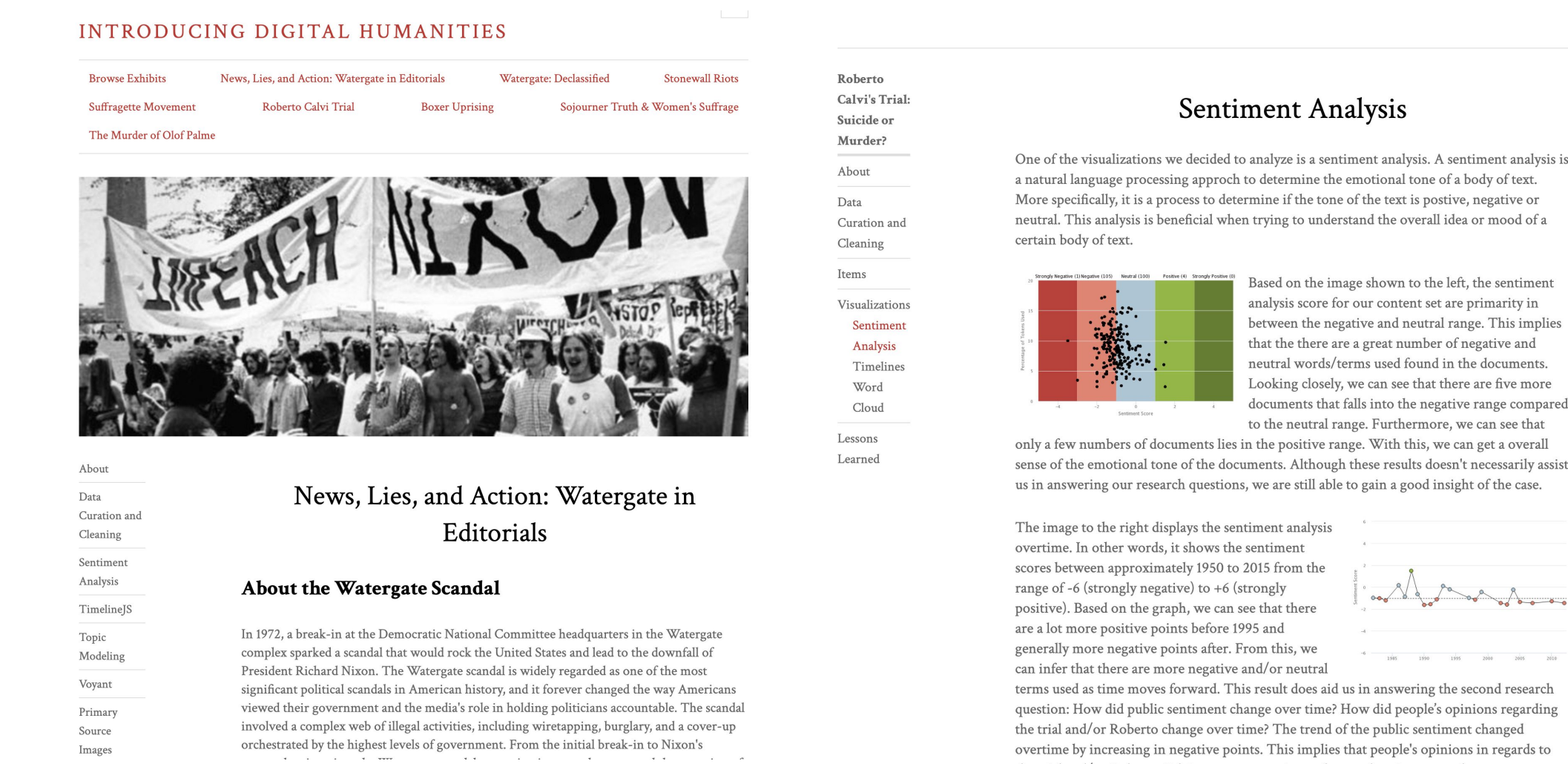
## SUSTAINABLE PROJECT-BUILDING

An integral component of developing digital projects that have a defined start and end point, and then a planned-for archival phase, is the creation of complete project documentation. This includes administrative recordkeeping, such as a Project Charter detailing how collaborators will work together, how they will communicate, and so on, alongside project-level considerations including scope and roadmap documents, sustainability and risk management plans. Details of stakeholders, funding and agreements with library repositories may also be included. While the scope and content of classroom projects was necessarily constrained by the time available in the quarter, the expectation was that groups would integrate the process of creating documentation within their workflow. To this end, students used the inbuilt 'Notebook' in *Gale Digital Scholar Lab* to capture process-level notes including details about search terms used when refining datasets, content set clean configuration choices, and tool setup options. Students embedded relevant visualizations and explained their interpretation of results.

These records served a dual purpose: they also provided me with insights into student comprehension, and into equity of labor within the team environment. The notes highlighted gaps in understanding that were then addressed in subsequent class sessions.

The Notebook contents are fully exportable in various formats; teams leveraged this functionality to funnel the notes they kept while working in the *Lab* into their digital exhibits, which were built on *Omeka*, an open source content management system ideal for displaying digital material with full metadata and embedded text

## analysis/evaluation output: OUTCOMES

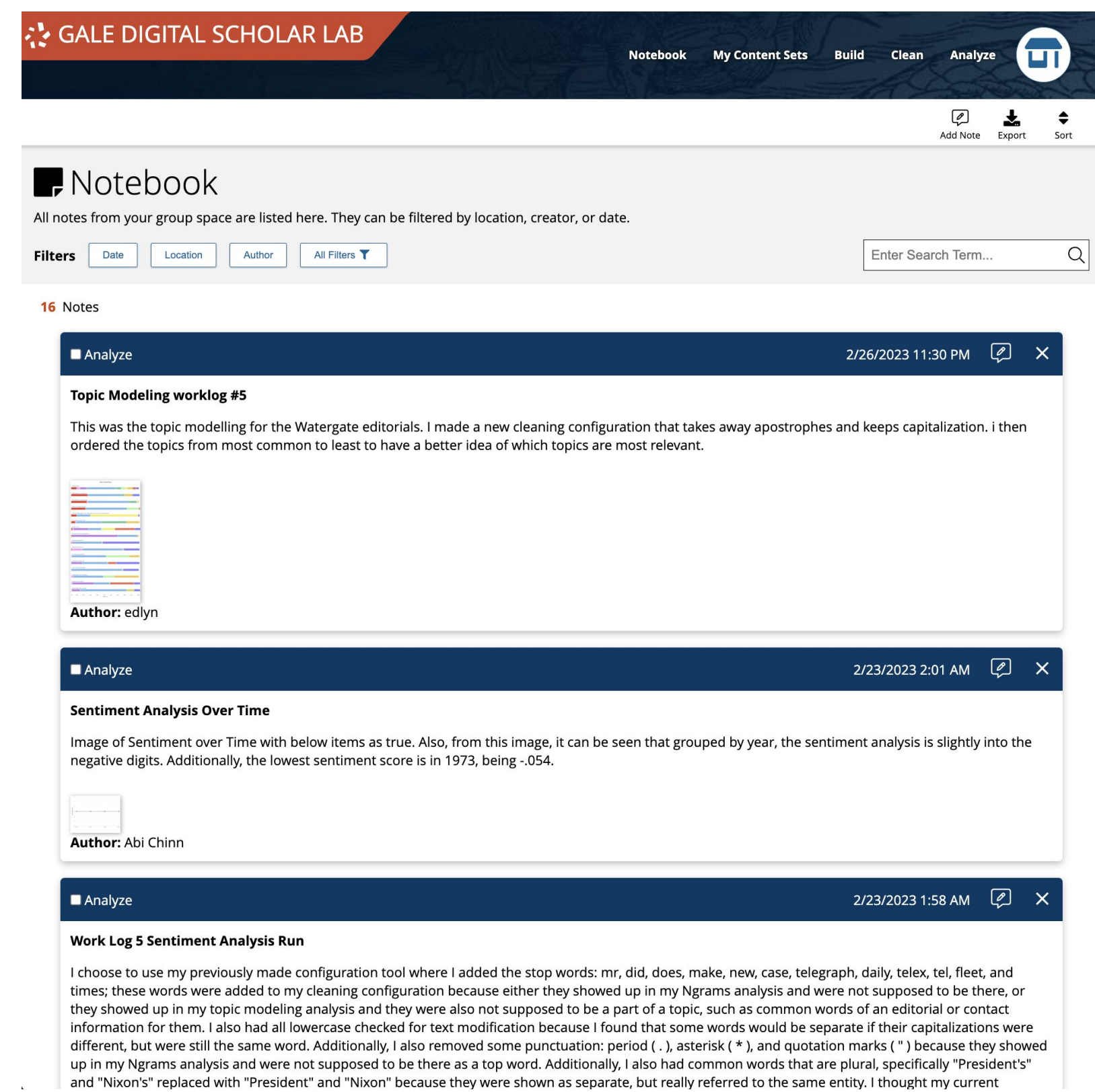


*Working collaboratively (with) a group in the DSL was good because we got to see each other's configurations and how each person used an analysis tool. I liked that I was able to compare my results with other group member's configurations. If our results were different, then I was able to further investigate why the differences occurred and either fix my tool or better understand each other's cleaning process.* (Student feedback, INFO 498)

## PEDAGOGICAL APPROACHES TO GROUP WORK IN THE UNDERGRADUATE DIGITAL HUMANITIES CLASSROOM

The average DH project has many collaborators, since it is understood that no one person can be expert in all aspects of successful project-building. Bringing this ethos to the undergraduate classroom necessarily involves developing strategies for facilitating student group work with the end goal of creating successful team norms for collaborative practices and delivering a completed digital project at the end of the class session. The pedagogical nature of the experience differentiates classroom project-building from the regular operations of an ongoing DH project, so for educators, it becomes important to streamline the practicalities of group work to create space to focus on theory, methodology and the application of digital tools. Student work described in this poster was undertaken during Winter Quarter 2023 by a cohort of Information School undergraduates at the University of Washington, taught by Sarah Ketchley.

Project-based learning and group work serve as the basis for classroom experience, both of which rely on developing successful team dynamics. Some of the practical tasks undertaken in class in support of this include creating a working agreement, project scoping, and establishing parameters and expectations. The latter is largely driven by project rubrics, which provide student teams with the scaffolding necessary to successfully complete a DH project.



The Winter Quarter INFO 498 class was an invaluable testing ground for the new *Lab* features. The cohort of 32 students raised questions, suggested enhancements and areas of improvement, and affirmed features that worked. They worked in teams of four to develop research questions based on one of eight datasets in the *Lab*. The Groups space enabled them to collaboratively curate further material, to develop cleaning configurations specific to their project, and then conduct either quantitative or qualitative analysis to develop theories and find answers. The Notebook was integral to transparency and clarity of working practices; often questions that one student raised in this document were answered by another team member. For the final digital exhibit, teams exported a variety of outputs from the *Lab* and brought the material into their *Omeka* digital exhibits, including:

- Images of primary source documents,
- Metadata spreadsheets,
- Tabular analysis data,
- Visualization images,
- Downloaded OCR text from team datasets,
- Notebook records (in both PDF and Word formats), filtered to display Individual student contributions.

