Innovation often requires the analysis of thousands – even millions – of documents. That’s an impossible task for any human, and it’s exactly where text and data mining (TDM) comes in. With TDM Studio, the end-to-end TDM solution from ProQuest, researchers have a new path to innovation. From an initial idea to the final output, TDM Studio leverages the power of a library’s content to help researchers at all levels make new connections and uncover career-defining outcomes.

TDM Studio offers researchers two pathways to perform text and data mining:

Accustomed to using coding for text analysis? The Workbench provides programmatic access to millions of ProQuest content using R or Python.

Want to visualize your data quickly without coding? Visualizations offer a new option to interrogate ProQuest content and easily discover insights across thousands of documents.

“We’ve been searching for an accessible path to current and historical content for our faculty and students to text and data mine. We expect this offering will foster community engagement with library collections and give users the means for collaborative research discoveries that are in line with Dartmouth’s mission.”

- Kenneth Peterson, Associate Librarian for Access & Collection Strategies, Dartmouth College
Two paths to new discoveries:

**Workbenches in TDM Studio: For Researchers with Coding Experience**

Designed for researchers who use the coding languages of R or Python.

**TDM Studio Dashboard**

Manage datasets and access the development environment.
- Create datasets of up to 2 million documents each by searching publication titles or ProQuest database names
- Manage up to 10 datasets simultaneously for a total of 20 million documents
- Access the development environment, a Jupyter Notebook, and restart your environment if necessary

**Creating a New Dataset: Finding Content**

Select from your ProQuest subscriptions – all rights-cleared for TDM.
- Access a wealth of content across disciplines, from technology to literature, all available with a few clicks
- Curate a dataset from current content or historical archives
- Benefit from the consistent data structure provided across content types – including current and historical newspapers, dissertations, scholarly journals and primary sources

**Creating a New Dataset: Refining Content**

Filter and refine your dataset based on your specific research targets and keywords.
- Narrow content and target the dataset specifically for your research topic saves valuable research time by eliminating the need to process unrelated documents
- Provides ability to specify topics by entering keywords and supports full Boolean search syntax
- Target specific date ranges, source types, and document types

**Jupyter Notebook**

A familiar integrated development environment.
- Cloud-based development environment that can be accessed from anywhere – on or off campus
- Preconfigured with standard data science libraries in R and Python
- Organize and access datasets, and manage running jobs
- Upload your content and easily incorporate rights-cleared datasets – including open-access content and social media content
**Visualizations Dashboard**

Create, access and manage your own projects.

- Launch several types of visualizations to explore supported data analysis methodologies
- Create new projects to answer research questions and explore new documents
- Manage and access up to five projects of 10,000 documents each

**Tested in Classrooms Across Disciplines**

Simple-yet-powerful search syntax can quickly be refined to reflect your area of interest. Based on your library's subscriptions:

- Analyze current content from sought-after newspaper titles
- Use ProQuest Dissertation & Theses to understand trends in research
- Quickly narrow content by keywords, publication title, date range, or document type

**Powerful Visualizations in Minutes**

Conduct research or teach data analysis with or without coding skills. These data visualizations provide the capability to enhance data literacy in the classroom, as part of a lab or workshop, or for individual research.

**Geographic Analysis**

- Map display shows dispersion of locations mentioned in the selected content
- Interactive display allows the exploration of patterns over time – or click on a cluster to explore documents associated with a specific locale

**Topic Modeling**

- Uncover keywords based on their prevalence in the selected documents, keywords are grouped into topics and track trends over time
- Gain a better sense of a topic, and find areas that may be overlooked. Topic modeling can help identify research topics, track trends, or conduct literature reviews
A text and data mining solution for research at all levels and across disciplines

Sought-After Content
TDM Studio offers access to rights-cleared content from ProQuest subscriptions and purchases. This content is applicable across disciplines and can be applied to many different research questions.

Flexibility
By combining an intuitive dataset creation capability, Jupyter Notebook access, and data visualizations, TDM Studio is able to support data literacy as well as research across disciplines for researchers with or without coding skills.

Efficiency
Significantly decrease the time to analysis, quickly create datasets, and either visualize the content relationships or access content in a consistent data schema across ProQuest sources.

Research, Teaching & Learning
Enhance research, teaching and learning outcomes across campus with a solution designed for varied skill levels in research methodologies and approaches.

Explore TDM Studio today.
Visit www.proquest.com/go/tdm-studio