A CASE STUDY

Text and Data Mining for Teaching and Learning

How the University of Sydney integrated TDM and data literacy into its undergraduate courses

proquest.com

To talk to the sales department, contact us at
1-800-779-0137 or **sales@proquest.com**.

ProQuest

The University of Sydney Library has provided text and data mining (TDM) services since 2017. Recently, EdTech leader ProQuest partnered with the library to develop and to pilot its comprehensive TDM service, TDM Studio. Through this partnership, the University of Sydney (UoS) library and ProQuest worked with faculty and staff at the university to integrate TDM into undergraduate courses. They piloted TDM Studio's data visualizations feature – now widely available – with an easy-to-use, visual interface that's accessible to users without coding knowledge.

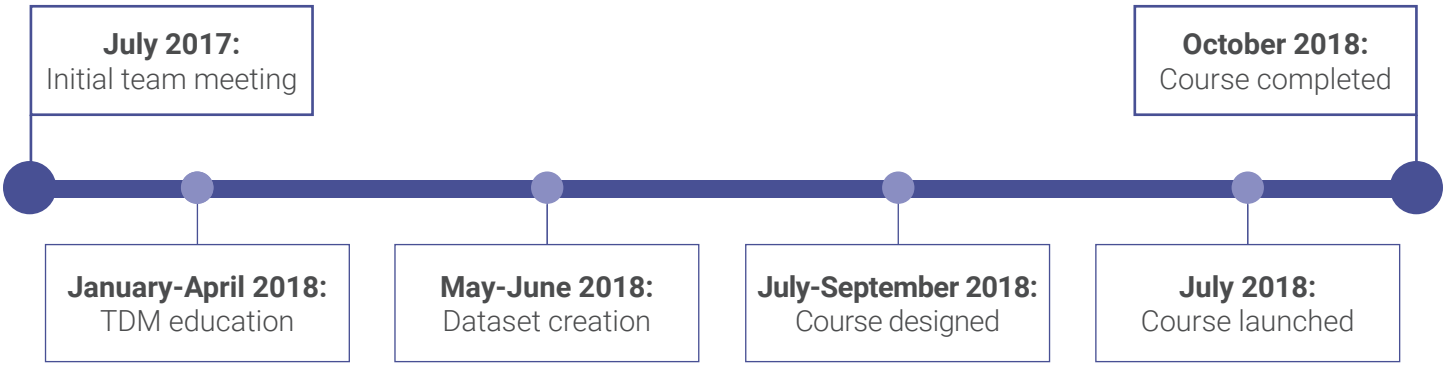
Dr. Marco Duranti, a senior history lecturer at the UoS, was a key faculty partner for this project. In 2017, Duranti won an UoS Educational Innovation grant for "developing digital literacy in human rights history." Duranti used this grant to incorporate TDM into The Human Rights Revolution, his undergraduate course covering topics like humanitarianism, transitional justice and contemporary Australian controversies. Duranti's goal was to introduce his students to TDM, an increasingly important skill in the humanities disciplines. As the first-ever faculty member at the University to integrate TDM into a contemporary history class, this was an entirely new experience for Duranti – and many others at UoS. He worked with the library and colleagues across the university to launch this program using a then-pilot version of data visualizations in TDM Studio.

This case study describes the processes, challenges and successes of this pilot and the University of Sydney's additional TDM work.

Project Team and Timeline

To successfully integrate TDM into the classroom, the UoS needed to draw from a series of broad skillsets. The team brought together a diverse group of representatives from both within and outside the university, including:

- Teaching staff and pedagogy experts from the Faculty of Arts and Social Sciences at UoS
- Data management and humanities subject experts from the UoS library
- Data scientists from the Sydney Informatics Hub
- Consultants from ProQuest



The Methodology

The library took on the role of becoming experts in TDM Studio. The library team worked with ProQuest to provide critical feedback on the pilot product to help future users. This included guidance and feedback on workflows to create datasets as well as specific suggestions for analyses and visualizations focused on geographic analysis and topic modeling. It also included creating faculty and end-user guides for working with TDM Studio. Separately, the library team was upskilling on TDM concepts and experimenting with open-source tools to grow its knowledge and skill level for TDM.

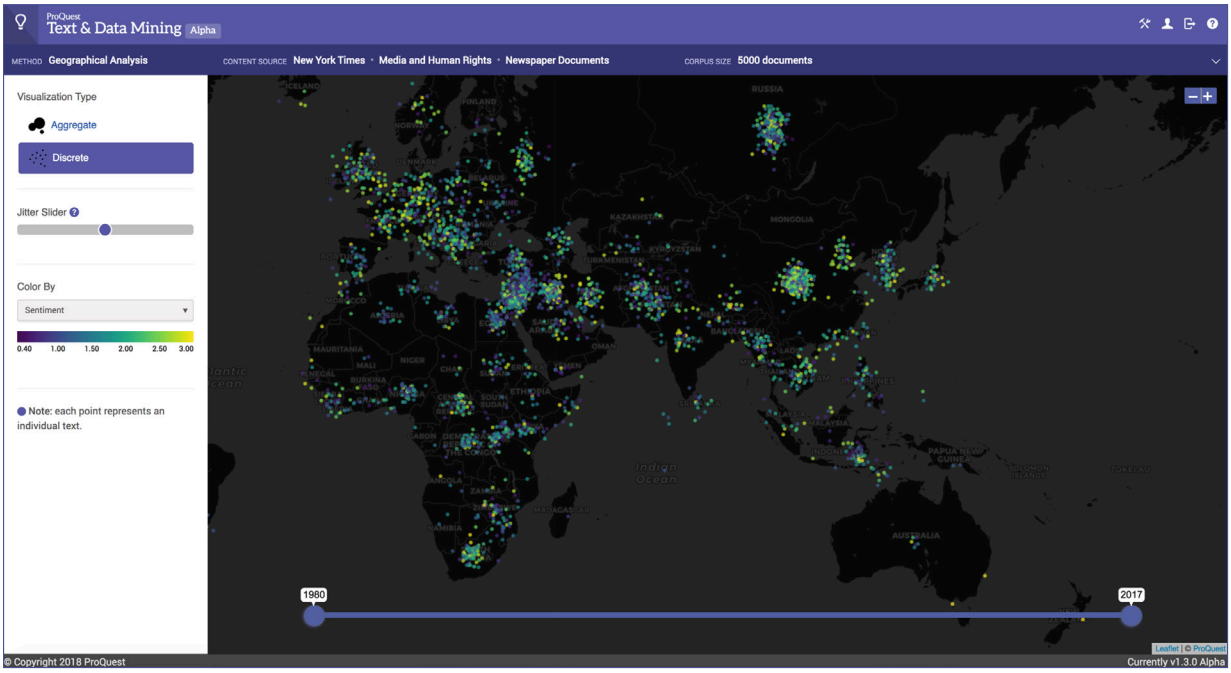
To create relevant datasets to be used in Duranti's class, the team leveraged the library's search strategy expertise and Duranti's subject knowledge to work with ProQuest to build dozens of datasets from newspaper articles. The wealth of content available from ProQuest enabled the team to create datasets on a range of topics, including datasets focusing on gay rights in the U.S. and Indigenous rights in Australia.

[TDM] helped students better understand the content they were studying or prompted them to make new discoveries in related topics.

Duranti then worked with the library to introduce text-mining methods and the data visualizations in TDM Studio. Ryan Stoker, Research Data Officer at the UoS library, provided introductory framework lectures with Duranti supplementing these with subject-specific information. The lectures were followed with hands-on classes utilizing TDM Studio focusing on two text mining methods – geographic analysis and topic modeling. When used together, these methods enabled a story to be told over both space and time. During the hands-on part of the class, students were divided into small groups and asked to build a narrative of historical events based on the TDM methods. Providing a group task rather than allowing students to individually explore the tool encouraged discussion and critical thinking throughout the class.

Prior to the in-person session, Duranti worked with educational designers, namely Bec Plumbe, to develop and integrate this unit into the course's learning management system (LMS), complete with quizzes and other online and offline learning activities. Students completed the quizzes and learning activities in their own time as a way of testing their newly acquired TDM knowledge.

Results



A prototype of TDM Studio used in Duranti's class for geographic analysis

Duranti's objective was to teach text-mining skills in conjunction with digital history methods. Were students seeing value in these emerging methods? Did they contribute to a diverse historical methodological toolkit? To evaluate these learning objectives, he asked students to answer periodic feedback quizzes.

Overall, the learnings from the pilot indicated that TDM helped:

- Understand vast sweeps of time and space with ease
- Identify possible triggers and causes for historical events
- Appreciate the complementary nature of pairing traditional historical methods with digital ones
- Discover articles via non-traditional search methods
- Gain a better understanding of large volumes of content

Students were able to learn about text and data mining and then have hands-on experience with TDM through visualizations within TDM Studio. The exposure to this type of analytics helped them better understand the content they were studying or prompted them to make new discoveries in related topics. The activities they completed also exposed some of the limitations of topic modeling, which in turn opened an opportunity to discuss critical data skills and the possible biases of algorithmic methods.

The project also helped demonstrate the limitations of text mining, such as availability of content and the parameters of the algorithms, and how to address those limitations. From a data literacy standpoint, this was a great benefit – it allowed students to explore general benefits and issues around data visualizations and representations.

As an educator implementing TDM in his course for the first time, Duranti noted that "ProQuest TDM Studio is an exciting and accessible platform for introducing scholars to the possibilities of text mining. The tool transformed how my students and I approached our exploration of human rights controversies in historical perspective. It illuminated not only the transformational impact of data science on our field, but also the value of combining computational analyses with traditional humanities approaches."

Challenges

The team encountered several challenges throughout the project, and this paper highlights two that may be helpful to other libraries and faculty staff considering integrating TDM into student courses.

- Learning curve: Initially, most members of the project team did not have extensive experience with TDM. They quickly familiarized themselves with TDM concepts and TDM Studio independently, and then used weekly meetings with ProQuest to ask questions, provide feedback, and inspire the creation of a user guide.
- During this project, the team gained a greater understanding of text mining processes, what types of research text mining can help with and how to create a dataset. The library team also created reusable text mining training content for an audience new to text mining. The ultimate payoff for the library was becoming further integrated into text mining activities both in teaching and research.
- Recruitment and audience: For a project like this one, it's important to find teaching staff who have an interest in exploring a digital research method in their course or are looking at a way to make their course more interactive. From a student perspective, TDM Studio's visualizations feature works well for novice researchers, particularly undergraduate students. It's an easy and low-stress way to teach the concepts of TDM without students needing to understand coding or technical languages.

Additional Projects and the Future of TDM at the University of Sydney

Librarians at the University of Sydney

Duranti's project was the first TDM Studio pilot at the UoS, but not the last. The library also worked with Jonathon Hutchinson, online media lecturer, to pilot TDM Studio in his Online Media course. In addition, the library continued working with Duranti in subsequent history courses including Genocide in Historical Perspective, The History of Human Rights, and Twentieth-Century Europe. In addition to these larger-scale pilots, TDM Studio Visualizations is used for smaller-scale library workshops to introduce graduate students to TDM. At UoS, these courses and workshops focused on humanities research, but TDM and TDM Studio can be used in nearly any discipline.

The UoS library, through its involvement in this project, has continued to grow its TDM services. Throughout 2020, the library trialed a set of text and data mining services that leveraged traditional library knowledge – such as copyright, licensing and information literacy pedagogy – packaged with new skills, such as text analysis techniques and concepts, to offer TDM services for both research and education. The library sees TDM skills as a key development area for library professionals as the way information is produced, accessed and used evolves.

We just give the students access to TDM Studio and let them run with it. Each year, less explanation and less setup is needed.

— Marco Duranti

Data skills are only becoming more important as research develops. TDM Studio enables students, teaching staff and researchers at all levels to engage with text and data mining and increase their level of comfort with using digital research methods. As Marco Duranti, who kicked off this project back in 2017, said: "At the outset, we give the students access to TDM Studio and let them run with it. Each year, less initial explanation and setup is needed. We taught this again in a recent history course – online, due to the COVID-19 pandemic – and we were still able to run with it. The interesting thing is that students are learning data literacies from lots of different everyday sources, so they were able to pick up TDM Studio almost on the fly."

Key team members

Marco Duranti
Senior Lecturer, Modern European and International History

Bec Plumbe
Educational Designer

Ryan Stoker
Research Data Officer

Jennifer Stanton
Manager, Digital Collections

Gene Melzack
Digital Curation Officer

Chao Sun
Research Engineer Group Lead

Jonathon Hutchinson
Lecturer, Online Media

Mathew Toll
PhD Candidate, Sociology and Social Policy

Kim Wilson
Academic Liaison Librarian

Andrew Inman
Manager, Academic Services

Katrina McAlpine
Associate Director, Publishing and Data Services

About TDM Studio

TDM Studio is a text and data mining solution for research across disciplines and enables researchers with or without knowledge of coding.

ProQuest's workflow solution for text and data mining is designed for research, teaching and learning. TDM Studio provides access to sought-after content including current and historical newspapers, primary sources, scholarly journals, and dissertations and theses. It empowers researchers, students and faculty to analyze documents by uncovering connections and patterns that lead to career-defining discoveries.

Learn more at www.proquest.com/go/tdm-studio or contact your ProQuest representative today.

proquest.com

To talk to the sales department, contact us at **1-800-779-0137** or **sales@proquest.com**.

10532-LTR/SB-AR-CG/12-20