

Emory University Libraries
Inventory for TDM and AI Research Projects
Fall 2024

This document is designed to help guide subject librarians in discussions with researchers needing to use content from Emory-licensed resources, or other free and open sources for purposes of text and data mining and/or AI-powered research.

- What is the current research project? What is its stated purpose?
 - Will it incorporate AI tools?¹
- Do you have a third-party tool in mind to work with any raw data? Ex: SPSS, STATA, etc.?
- Are there Emory or non-Emory co-authors? Will you need access to the raw data source? Much will depend on licensing terms.
- What is the ideal output?
 - What format do you need? Do the files contain both text and images? Ex: csv file, pdfs, etc.
 - Will the data need significant cleanup prior to analysis?
- Does the output have requirements for data citing/sharing?
- Have you already identified a data set? Can that data be found in multiple sources?
 - Do you know if we already have a license in place for the needed data? Or if it is freely available?
 - If the content is not available via the library, do you have a budget for acquisition?
- What are the range of dates needed?
- Do you have a secure place for your computational analysis?
 - Most of our licenses, particularly for machine learning, require significant security measures and restrict training of third-party AI models.²
- Have you considered your storage needs?
 - The Library does not provide a storage option at this time, but there are options available to Emory researchers by way of AWS.
- Are you interested in available library tools that allow you to conduct text & data mining without coding?

¹ . As part of the Emory AWS service, there are a number of AI tools Amazon provides. We recommend getting in touch directly with the Emory AWS service to learn more about these tools; you can also find a comprehensive list of AWS AI Tools at: <https://aws.amazon.com/ai/services/>

² Emory Amazon Web Services (AWS) is Emory University's preferred and recommended cloud service for faculty-led computational needs. The service provides access to Amazon's cloud computing services, including computing, storage, database, etc. You can learn more about this service on OIT's website: <https://it.emory.edu/aws-overview/index.html>.

- Some of our vendors have their own services, such as ProQuest (TDM Studio) and Gale (Digital Scholar Lab) for text and data mining.
- Do you need advice on a tool or language for analyzing the data?
-
- What is your ideal timeframe for this request?

Questionnaire adapted from Erin Loree and Kelly Lavoice at Vanderbilt University Libraries.