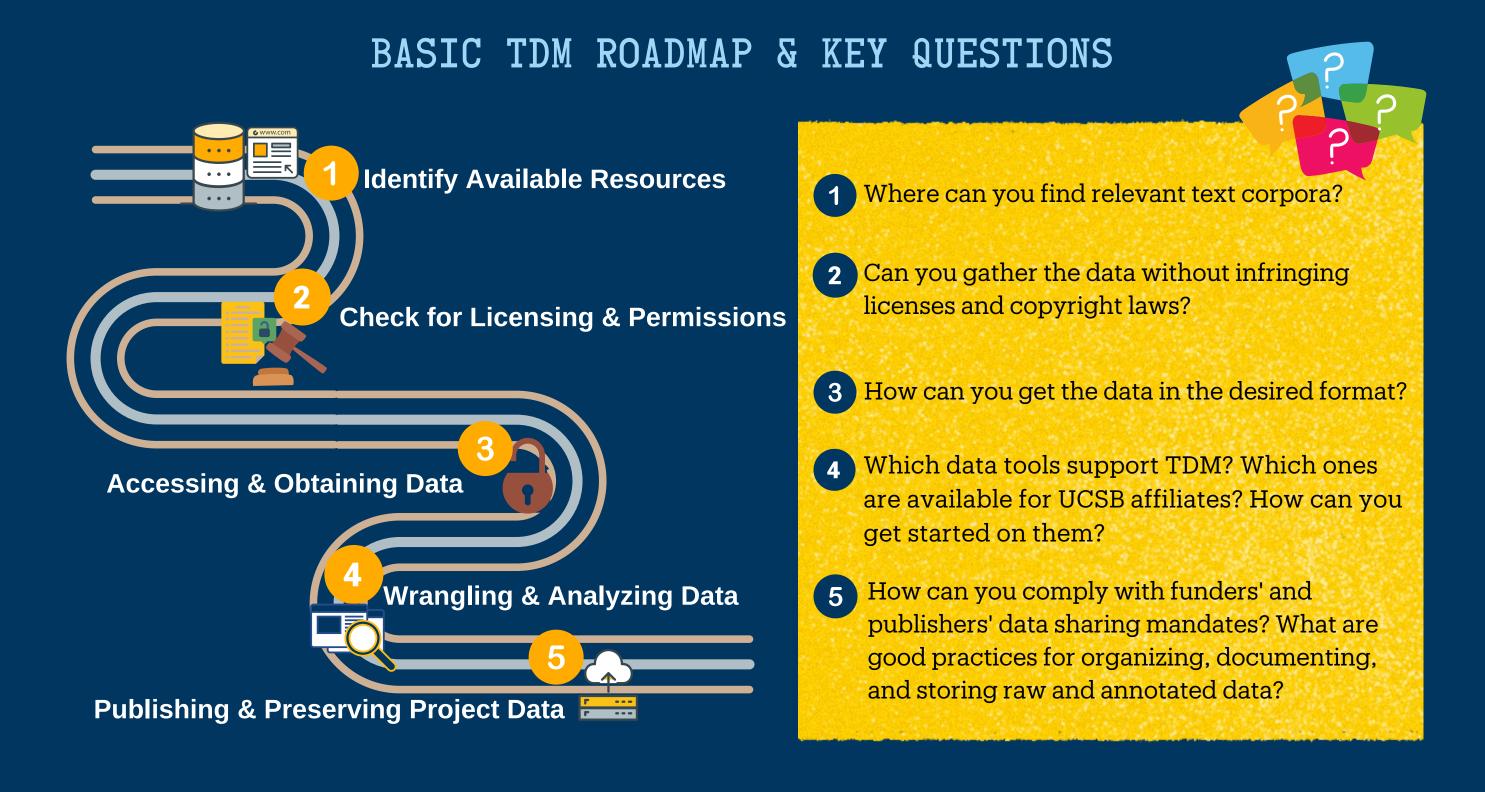# Minding Text Data Mining

Text Data Mining (TDM) is a research process for deriving high-quality information based on insights and patterns from text corpora. In a TDM project, you may encounter obstacles around copyright and license agreements or technical challenges obtaining, analyzing, and preserving your project data.

BASIC TDM ROADMAP & KEY QUESTIONS

1. **Identify Available Resources**
2. **Check for Licensing & Permissions**
3. **Accessing & Obtaining Data**
4. **Wrangling & Analyzing Data**
5. **Publishing & Preserving Project Data**

1. Where can you find relevant text corpora?

2. Can you gather the data without infringing licenses and copyright laws?

3. How can you get the data in the desired format?

4. Which data tools support TDM? Which ones are available for UCSB affiliates? How can you get started on them?

5. How can you comply with funders' and publishers' data sharing mandates? What are good practices for organizing, documenting, and storing raw and annotated data?

THE UCSB LIBRARY CAN HELP YOU NAVIGATE THESE STEPS MORE EASILY. CHECK OUT THE TDM LIBGUIDE BELOW FOR MORE INFORMATION, OR CONTACT OUR TDM TEAM: TDM@LIBRARY.UCSB.EDU

MORE INFO

bit.ly/tdm-help

UC SANTA BARBARA
Library

www.library.ucsb.edu