Do funding agency data policies conflict with text mining license terms?

Jen Ferguson  
*Northeastern University*

Amanda Rust  
*Northeastern University*

Follow this and additional works at: [https://escholarship.umassmed.edu/escience_symposium](https://escholarship.umassmed.edu/escience_symposium)

Part of the Scholarly Communication Commons

This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

**Repository Citation**


Creative Commons License

This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
As text/data mining (TDM) becomes more prevalent, researchers seek to mine library resources for their projects. Many vendors are including language in their text/data mining licenses that aims to protect their investments in metadata. At the same time, researchers are increasingly being called upon to share and preserve data from their projects. We speculated that funding agency data policies might conflict with vendors’ text/data mining license terms.

**Methods**

We obtained text/data mining licenses from four vendors – here referred to as vendors A, B, C, and D. We compared language from these vendors’ licenses with guidance from several grant funding agencies. We noted potential conflicts between funding agency guidelines and vendor licensing terms.

**Results**

We identified several cases in which the language in library vendor TDM licenses is at odds with funding agency policies. These conflicts are particularly apparent in regards to vendor limits on lengths of output and data retention.

In many cases, we found the language in both funding agency policies and vendor licensing agreements imprecise, making it difficult for us to determine whether these terms are in conflict. In our judgment, most of these instances do represent at least the potential for conflict.

Librarians assisting researchers with data management plans should be aware of potential conflicts between vendor TDM licenses and funder policies on data sharing and preservation.

There is a role for cross-library collaborations to address this issue. Librarians assisting researchers with data management plans can work with colleagues who handle licensing negotiations, as well as liaisons whose researchers are doing TDM work to best support researchers wishing to use library resources for text/data mining.

There is also a unique opportunity to shape the conversation on TDM research support by working with vendors to draw attention to possible conflicts with data sharing mandates.

**Acknowledgment**

Thanks to Janet Morrow for sharing insights and excerpts from TDM licenses for this project.