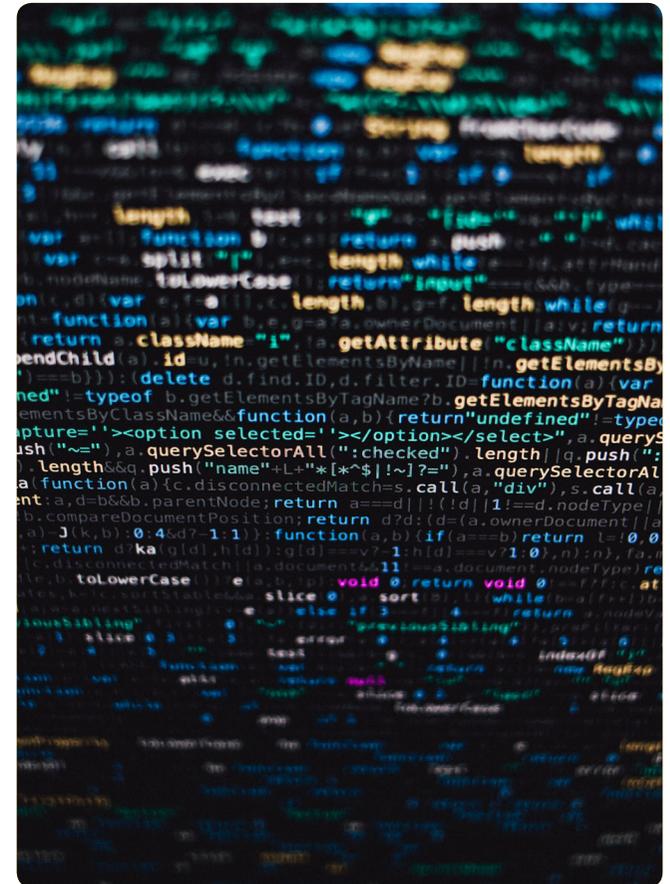




RESEARCH LIBRARIES:  
HOW YOU CAN SUPPORT  
TEXT AND DATA MINING

CONTENTS

- [Research Libraries: How You Can Support Text and Data Mining](#) 2
- [Getting Started](#) 3
- [Interacting with Publishers](#) 4
- [Researcher Privacy](#) 5



# RESEARCH LIBRARIES: HOW YOU CAN SUPPORT TEXT AND DATA MINING

Europe's incoming Directive on Copyright in the Digital Single Market contains two new exceptions allowing text and data mining (TDM). The Directive is currently being implemented by all member states and must come into force no later than June 2021.

Article 3 grants universities, educational establishments and libraries the right to mine in-copyright works to which they have lawful access, and Article 4 introduces an exception mainly aimed at commercial data analytics and AI. This second exception could be helpful in the context of Knowledge Transfer between private companies and universities.

Although a few European countries already have copyright exceptions for TDM, the majority do not. This means that most European research librarians and libraries need to start considering how they can best use this exciting new opportunity to support researchers undertaking TDM. Meanwhile, libraries who have already benefited from a TDM exception need to think how best to bridge the gap between the existing exception and the new TDM exceptions. This document contains several recommendations for libraries in both categories, who want to strengthen their TDM capacity.

## Libraries as Key TDM Partners

### *Libraries are essential to supporting TDM*

As organisations, they are often the locus of copyright law expertise at a university, have the relationship with the publisher and play an important part in supporting Open Science through overseeing the institutional repository. They also support FAIR data, open access, and open data - all of which are important precursors in order for researchers to successfully carry out automated text and data analysis.

Furthermore, librarians and libraries are a natural place where skills from across the university can come together to support researchers from different disciplines to undertake data mining. Libraries can, for example, forge relationships with colleagues across the university, in

particular with the Law and Computer Science departments, or with a centrally-managed legal office. By working with these partners, libraries can put in place TDM workflows and infrastructures, create relevant guides and run training courses.

Finally, libraries have an important role to play as advocates for the needs of researchers. The foundations of this are laid when libraries provide the technical and legal infrastructure which researchers need to undertake TDM. Libraries can also push boundaries by being robust with publishers in asserting this new right in EU law, and ensuring that any technical issues or access-blocking experienced by the institution are resolved quickly (see the section Interacting with Publishers on [p. 4](#)).

## Influencing Factors

The role of libraries in supporting TDM will be influenced, amongst other things, by:

- **INSTITUTIONAL POLICIES** regarding how and where to store data in line with research data management plans (which are increasingly required for academic research);
- how **KNOWLEDGEABLE** library staff are about TDM, and how supportive library leaders are in terms of providing time and encouragement for staff to undertake training;
- the role the library plays in a university's **OPEN ACCESS ACTIVITIES**; and
- its **FUNCTION** regarding the hosting of an institutional repository.



## GETTING STARTED

In order to start supporting TDM, your library could explore the following topics and activities.

### Understand the Current Situation

- Survey researchers in all domains as to how they think about data analytics, and how they would like to be supported.
- Explore the text and data mining services aimed at researchers, such as those from the [National Centre for Text and Data Mining](#) and [ContentMine](#).
- Discuss how TDM relates to funder research data management plans and the storage of research data.

### Share Knowledge & Resources

- Create a list of subscriptions, digitized content (and whether OCRred) and deposited research data in the institutional repository that could be used for TDM and communicate this to researchers.
- Collate a list of sources of tools available for data mining and share these with researchers. Available tools include, but are not limited to, [GitHub](#), [ContentMine](#), [OpenMinted](#), [Voyant](#), [TAPoRware](#), [Orange Text Mining](#), [Digital Humanities Tools](#) and [National Centre for Text and Data Mining](#). See also the [Future TDM list](#).
- Create ways for researchers to meet data scientists, legal experts and librarians.
- Run technical workshops on TDM.
- Create a website page about TDM and supporting services at the university, such as the libguides created by [Cambridge](#), [Yale](#) and the [University of Queensland](#).
- Set up a pilot TDM project with different dataset types.

### Get Leadership Buy-In

- Ask your university's leaders - Vice Chancellors, Research Vice-provosts, Senior Managers - how they would like the university to support data-driven research. It may help to link this discussion to funder requirements on data management.

- Discuss where the budget for supporting TDM will come from. e.g. Technical costs as well as purchase costs.

### Library Policies & Structures

- Brainstorm a policy on publisher relations: how to deal with publishers suspending access, protecting researchers' privacy and how to assert your rights under the new TDM exceptions. See 'Developing Institutional Policies' ([p. 4](#)).
- In line with research data management policies, ensure researchers deposit their (TDM) data sets for storage and long-term preservation. This is part of good scientific practice. It can be achieved through statutes within the university, and by highlighting funder requirements as part of the terms of the funding grant.
- Consider the best formats for data mining. e.g. XML, requirement for OCR etc.

### Network Building

- Identify pockets of legal and technical expertise across the university.
- Create ways for researchers to meet data scientists, legal experts and librarians.

### Staff Training

- There are many resources available including a LIBER webinar on [how libraries can support TDM](#), the free online course [Introduction to Text and Data Mining](#) from the FOSTER project, [Practitioner Guidelines](#) created by FutureTDM and this [introduction to TDM](#).
- Explore the TDM services aimed at researchers from the National Centre for Text and Data Mining, ContentMine etc.
- Discuss how machine learning and data mining fits into the function of institutional repositories, open science, open access and the research data management lifecycle.



## INTERACTING WITH PUBLISHERS

### Case Study: How Ghent University Started Supporting TDM

In 2015, Ghent University carried out a survey and interviewed all research and education faculty to understand what skills the university library, and the institution as a whole, should invest in and support. Among other things, the results of this project highlighted a need for better skills in data management and data science. Building on this, the university library investigated how education in data management skills could be introduced into core university curricula. Teaching of some of these skills has now been implemented in doctoral schools.

Ultimately, the library's project aims to make these part of the curricula of every Master's degree by 2018 (with the depth and detail covered varying by field). Laying the foundations for good data management practices is a key first step towards supporting greater use of data analytics, even when students are not explicitly educated about TDM.

More generally, consulting with the university community and gathering evidence to support a need for better investment in data-related skills is a strategy that can be used to drive policies of investment and support in this area. (Source: [FutureTDM Practitioner Guidelines](#), CC BY 4.0)

Also: Watch this [LIBER webinar](#) highlighting how libraries are enabling TDM.

### Developing Institutional Policies

Research by [LIBER](#) has shown that researchers are not infrequently blocked from accessing a wide range of content when undertaking TDM and that, when publishers block content, resolving the issue often takes a month or longer to resolve. Service can be suspended across the campus, impacting whole communities and not just individual researchers.

Given this, and with demand for TDM rising, libraries must establish policies on how to react if access to paid-for eResources is suspended. Often suspension of service happens when the publisher's servers are modelled around human access, not computer access, and TDM therefore is seen as an atypical behaviour.

*Policies relating to how to deal with suspension to subscribed-to content should reflect the following:*

- Under the Copyright in the Digital Single Market Directive, universities and research organisations have the right to undertake TDM in regards to content to which they have legal access;
- Individual researchers that can be identified by the library (or the publisher) have certain rights under data protection law to protect their privacy. As data controllers, universities must comply with the law;
- Privacy and academic freedom should be protected;
- Given the above, the extent to which the university feels it is lawful and appropriate to share information with publishers, or respond to publisher requests regarding legitimate lawful activities performed by researchers should be formed into a policy.

### Legal Considerations

#### *Contracts Cannot Remove the Right to Undertake TDM*

Europe's incoming TDM exceptions grant anyone in a university, cultural heritage institution or research organisation the right to undertake TDM. This right cannot be removed by contract, irrespective of what jurisdiction the contract is signed in, if your institution is physically based within the European Union.

**If the contract is silent on TDM or includes provisions banning TDM, you can still undertake text and data mining.** Essentially this means that as long as you comply with the new law, there is absolutely no need to engage with or seek approval from publishers for TDM as long as you comply with the exceptions in copyright law. Researchers are free to perform TDM on any in-copyright material to which the library has legal access, without seeking permission from publishers.

To summarise, libraries should:

- **REMEMBER** that this new right to perform TDM cannot be removed by licences. There is no requirement to inform or seek permission from publishers in order to perform TDM.
- **DECIDE** whether to highlight contractual provisions in licences that conflict with TDM<sup>1</sup>, or simply ignore them as contracts cannot remove the right in law to perform TDM.
- **PAY ATTENTION** to when Articles 3 and 4 from the DSM Directive are transposed into national law.
- **KNOW** the legal procedure set out in your country's copyright law for getting access to resources, if access is blocked and not quickly resolved by the publisher (see the section Reacting to Content Blocking on [p. 5](#)).



<sup>1</sup> There is no need to remove any conflicts in the licence with the right to data mine because the law makes it very clear that text and data mining cannot be removed by contracts. If for whatever reason this is still not enough the following type of clause can be inserted " Nothing in this contract shall undermine limitations and exceptions in copyright law."



## RESEARCHER PRIVACY

In a world where user behaviour is constantly tracked online, and where some publishers ask to be informed as to what research is being undertaken by researchers employing TDM, we believe libraries have a particularly important role to play regarding researcher privacy.

LIBER believes that concepts of privacy (including but not limited to data protection law) and academic freedom mean that any libraries should consider carefully whether or not to respond to any request from publishers about specific TDM activities being undertaken. This is consistent with the situation in the analogue world, when libraries bought printed journals and books, and the purchase of those items was the limit of publishers' knowledge about user behaviour.

To summarise, libraries should:

- **RESPECT** the privacy and academic freedom of researchers by not automatically complying with publisher requests for further information about TDM activities;
- **CONSULT** with the individual(s) doing TDM for their opinion, in cases where publishers wish to understand more about the research being undertaken in order to help the research;
- **MAKE CLEAR** to researchers that there is no obligation to discuss with the university or the publisher their TDM activities, as long as it complies with the law and institutional policies;
- **DECIDE** on best practices for dealing with publishers where TDM is taking place generally, and the library is concerned that Technical Protection Measures (TPMs) used by publishers may lead to a campus-wide suspension of access to an e-resource.

### Reacting to Content Blocking

Despite the new exceptions, LIBER sees a risk that publishers may continue to use TPMs or other means to frustrate the right to data mine. Once the new law is implemented, we therefore recommend that libraries immediately take the following actions when publishers fail to resolve content blocking within 72 hours.

- 1. INFORM THE GOVERNMENT.** National processes will be different but in general you should contact the government department which oversees, in line with the law, the blocking of access by technical protection measures. Say that you wish to invoke the 'circumvention of technical protection measures' process in copyright law.
- 2. TELL THE PUBLISHERS ASSOCIATION** and any national consortia licensing organisation that a publisher is preventing your institution from exercising the right in law to data mine.
- 3. ALERT LIBER** by filling in the following [form](#).