

How can I prevent my material from being used for commercial text and data mining?

Works and other protected materials are widely used in the development of technological innovations **that utilise text and data mining techniques**, such as in the training of artificial intelligence and language models. As an author or publisher, **you can prohibit** the use of your works and other protected materials, such as books, articles or photographs, in commercial text and data mining. In legal terms, this is known as a reservation of rights.

This document provides copyright holders with examples of the reservation of rights in a machine-readable format.

If you do not wish your protected material to be used for the training of artificial intelligence, for example, you can prohibit text and data mining for purposes other than scientific research in the ways set out below. Please note that text and data mining by research organisations and cultural heritage institutions for the purposes of scientific research cannot be legally prevented by agreements or technical measures.

You must impose the prohibition explicitly and in an appropriate manner, i.e. by machine-readable means. You can include this prohibition either in the general terms of use of the material or in the metadata of the file and website. Metadata is a more effective way of enforcing a prohibition.

The right of copyright holders to reserve their rights in relation to text and data mining for commercial purposes is based on Article 4(3) of the Directive on Copyright in the Digital Single Market (DSM Directive) Article 4(3). In Finland, this right is set out in section 13b(1) of the Copyright Act.

1. Prohibition through the terms and conditions of agreement and use

You can agree on text and data mining carried out for commercial purposes in exactly the same way as for other forms of using works. This means that you can either prohibit or permit the commercial mining of text and data, for example, **in a licence agreement** to which you are a party. You can also prohibit the use of your work **in the licence** governing your material by including a prohibition on commercial text and data mining in the licence, commission, contract or other terms and conditions.

You can include a prohibition on machine-readable content in the licence terms or terms of use for your material, or in the terms of use for your website.

You can include the prohibition in the terms and conditions as follows, for example:

Example clause 1: “The copyright holder has reserved their rights to commercial text and data mining.”

Example clause 2: “Text and data mining for purposes other than scientific research is not permitted.”

You must ensure that the prohibition is clearly stated.

If you wish to ensure that permission is always sought from you for commercial text and data mining, you can state this in the terms and conditions under the clause concerning the prohibition as follows, for example:

Example clause: *"The copyright holder has reserved its rights to commercial text and data mining." "Any text and data mining carried out on the material must be agreed separately with the copyright holder."*

2. Prohibition by means of metadata

In an online environment, your digital content, such as an e-book, photograph, video, article in PDF format or other publication, may be found in a wide variety of locations. For this reason, the prohibition you have included in the terms and conditions of the original source website may not be sufficient to effectively prevent text and data mining, if your material is hosted on multiple sites.

With regard to files, a machine-readable prohibition is most effectively implemented by using metadata that automatically prevents text and data mining from being carried out on the file. In the case of material released by a publisher, the reservation of rights must be agreed with the publisher.

2.1 RESERVATION OF RIGHTS IN THE FILE

EPUB files

The World Wide Web Consortium (W3C) has developed technical guidelines for EPUB files that enable you to prevent commercial text and data mining from being carried out on your EPUB files.

You can add a restriction to your EPUB file as follows:

1. Add the 'tdm:reservation' property to the <meta> element in the EPUB package document.
2. If you wish, you can also add the optional 'tdm:policy' property to indicate that it is possible to obtain a licence for your data for commercial text and data mining.

In the example below, the copyright holder

- (1) **prohibits** commercial text and data mining of their EPUB files **using the 'tdm:reservation' property**, as well as
- (2) **specifies via the 'tdm:policy' property** that it is possible to obtain a licence for commercial text and data mining carried out on their EPUB file.

```
<package prefix="tdm: http://www.w3.org/ns/tdmrep#" ...>
  <metadata ...>
    <dc:title>Document title</dc:title>
    <meta property="tdm:reservation">1</meta>
    <meta property="tdm:policy">https://provider.com/policies/policy.json</meta>
  </metadata>
</package>
```

Other technical guidelines developed by the W3C

You can also restrict the use of your material in other ways by using metadata. The W3C has outlined various methods and provided detailed technical guidelines for them, which can be found here: [TDM Reservation Protocol \(TDMRep\) \(w3.org\)](https://www.w3.org/2011/08/tdm-reservation/).

Please note that the methods described here may have a negative impact on search engine visibility.

PDF file metadata

You can enter a restriction in the metadata of a PDF file in the same way as you would enter, for example, the file's keywords or author details.

Add the following prohibition on commercial text and data mining:

1. Select the hamburger button > Document Properties (Windows) or File > Document Properties (macOS).
2. Select the 'Description' tab and enter the author's name, the subject and the keywords.
3. **Select 'Other metadata' and add a prohibition on commercial text and data mining here.**
 - a. Word the prohibition as follows, for example: *"The copyright holder has reserved their rights to commercial text and data mining."*

You can find Adobe's English-language guidance on adding metadata to PDF files here: [PDF file properties and metadata, Adobe Acrobat](https://helpx.adobe.com/acrobat/using/pdf-file-properties-and-metadata.html).

2.2 RESERVATION OF RIGHTS ON A WEBSITE

The robots.txt robot restriction standard

The robots.txt standard is a file that can be used to provide information and set restrictions for search engines, web crawlers and other bots that scour websites.

You can place the robots.txt file in the root directory of your website's web server. Adding the robots.txt restriction standard to a website requires a certain level of technical expertise, so you may wish to ask a party, such as the technical administrator of your site, for help.

If you wish to add this file to your website, please follow these steps:

1. Place the robots.txt file in the same location where you would normally place your website's home page (/index.html). The correct location for the file is this: verkkosivustosi.fi/**robots.txt**
2. Add the desired restriction to the text of the robots.txt file
 - a. You can use any program that allows you to generate text to create the file.
 - i. Microsoft Windows -> notepad.exe, wordpad.exe or Microsoft Word
 - ii. Macintosh -> TextEdit
3. Enter the restriction you wish to apply in the text of the robots.txt file as follows:

User-agent: *

Disallow: /

The example above prevents all bots from accessing your website.

Detailed instructions in English on how to create a robots.txt file can be found here: [The Web Robots Pages \(robotstxt.org\)](https://www.robotstxt.org/). The website also provides a more comprehensive description of how the robots.txt file works.

Please note that setting up a robots.txt file has a negative impact on your search engine visibility.