

The exception of text and data mining for the purposes of scientific research in the context of activities of libraries and repositories

Jakub Míšek

Institute of Law and Technology, Faculty of Law, MU
@jkb_misek

Conference on Grey Literature and Repositories, Prague, October 17, 2019

Supported by Czech Scientific Foundation – project GA17-22474S Adapting Exceptions and Limitations to Copyright, Neighbouring Rights and Sui Generis Database Rights to Digital Network Environment



via <http://www.nusl.cz/ntk/nusl-407828>

Contents

- TDM and the collision with IP rights
- Exceptions applicable for TDM
- Critical evaluation of new TDM exceptions (DSM Directive)
- New TDM exceptions in the context of grey literature and PSI

Text and Data Mining (TDM)

- Getting the information from a vast number of documents
- Finding new contexts, models and patterns
- The standard procedure:
 - Access
 - Mining (Copying)
 - Analysis (Reuse)

TDM and IP rights

- Data as such are not protected
- Copyright protection
 - Analysed works
 - Copyright protection of a database
 - A broad application of a reproduction right
- Sui generis database right
 - Extraction right
- Necessary condition
 - Exactness and clarity of legal regulation and publication conditions
 - A risk of a chilling effect
 - Applicable exceptions?

Exceptions in Directive 2001/29/EC

- Exceptions from the reproduction right
 - Art. 5 para. 1 – temporary reproduction (online transfers)
 - Art. 5 para. 3 letter a) – use for sole purpose for scientific research
- Exceptions from the extraction right
 - Art. 8 of directive 96/9/EC – Rights of lawful users (only insubstantial parts of DB)
 - Art. 9 letter b) of directive 96/9/ES - extraction for the purposes of scientific research
- Official work?
- Old exception – insufficient for TDM

DSM Directive (EU 2019/790)

- Exceptions apply to both copyright and sui generis database right
- Exceptions are mandatory
- Art 3 – TDM exception for the purposes of scientific research
 - Narrow application:
“reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access.”
 - Narrowing: possibility of “measures to ensure the security and integrity of the networks and databases”
 - The exception applies only to reproduction and extraction rights
- Art 4 – General TDM exception

DSM Directive (EU 2019/790)

- Art 3 – TDM exception for the purposes of scientific research
- Art 4 – General TDM exception
 - Any purpose, any person
 - Lawful access to the content is necessary
 - Problem No. 1: Reproductions and extractions may be retained only for as long as is necessary for the purposes of text and data mining
 - Future analyses? Validation? Repeatability?
 - Problem No. 2: Right holders can exclude TDM
 - Para. 3: “The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.”
- Better than before, but still insufficient

Libraries and repositories in the light of TDM exceptions

– 2 possibilities:

- 1) Library (repository operator) conducts TDM
- 2) Sources of a library (repository) are to be analysed by a third party

1) Library (repository operator) conducts TDM

- The main problem – copyright
- A number of copyrighted documents
 - Not all of them licenced for secondary use
- Database rights are not a problem – the institution is the right holder
- Libraries and repository operators are cultural heritage institutions
- For scientific research purposes Art. 3 of DSM directive is applicable
- For other purposes only Art. 4 of DSM directive is applicable

2) Sources of a library (repository) are to be analysed by a third party

- Libraries (repository operators) are subject to the PSI directive (2003/98/ES)
- The main problem is (yet again) copyright
 - Database rights can be licensed
- Publication of documents
 - Licence
 - lawful access
 - It is recommended to add a metadata record concerning applicability of Art. 4 of DSM directive
- Providing of the documents based on a FOIA request
 - Possible legal licence (in Czechia use of a work for administrative purposes – section 34 of the Copyright act)
 - Further use (re-use for TDM) legitimised by art. 3 or 4 of DSM directive

Conclusion

- The new DSM exceptions are not a substantial change for libraries and repository operators
 - TDM analysis of own documents made easier
 - It is recommended to amend metadata of the documents so, that it contains information on applicability of art. 4 of DSM directive.

Thank you for your attention!

Jakub Míšek

@jkb_misek