

Paweł Kamocki, Erik Ketzan, Julia Wildgans & Andreas Witt

New exceptions for Text and Data Mining and their possible impact on the CLARIN infrastructure

- Copyright - exclusive rights to copy and share
- TDM requires (temporary) copying
- Copyright Directive 2001 allows for exceptions :
 - Temporary copies (web browsing)
 - Non-commercial research
 - Quotations etc.
- Principle : exceptions are overridden by contracts and technological protection measures (DRM)

National exceptions for TDM I : the UK

- Introduced in 2014
- Beneficiaries: everybody
- Allowed acts: reproduction for ‘computational analysis’ (no sharing)
- Allowed purposes: non-commercial research
- Requirements: **lawful access**, attribution
- Non-overridable by contracts
- Overridable by technological protection measures

National exceptions for TDM II : Germany

- UrhWissG (Urheberrechts-Wissensgesellschafts-Gesetz)
- Introduced in 2017/2018
- Beneficiaries: everyone
- Allowed acts: reproductions for automatic analysis, sharing within research team
- Allowed purposes: non-commercial research
- Limitation: after the end of the project, the corpus must be deleted or transferred to a specialised archive
- Requirement: flat-rate remuneration
- Non-overridable by contracts, can possibly override TPM ?

National exceptions in some other Member States

- **Austria:**

- No research exceptions (only private copies for research purposes)

- **Poland:**

- Very broad exception for non-commercial research (only for public institutions) seems to cover TDM activities

- **France:**

- Proposed, not yet enacted — dead letter for now

TDM exception(s) in the new Copyright Directive

- Directive on Copyright in the Digital Single Market
 - Proposed in 2016
 - Approved by the European Parliament on 12.09.2018
 - Expected to be adopted in early 2019
 - Twelve months for implementation

TDM exception(s) in the new Copyright Directive I

- Art. 3 – mandatory exception
 - Beneficiaries: public research organisations, educational establishments, cultural heritage institutions
 - Allowed acts: reproductions for text and data mining (no sharing)
 - Allowed purposes: research (also commercial)
 - Requirements: lawful access
 - Non-overridable by contracts
 - TPM only possible in limited circumstances

TDM exception(s) in the new Copyright Directive II

- Art. 3a – optional exception
 - Beneficiaries: everyone
 - Allowed acts: reproductions for text and data mining (no sharing)
 - Allowed purposes: any
 - Requirements: lawful access, use has not been expressly reserved by rightholders

TDM exception(s) in the new Copyright Directive III

- ‘Trilogue’ meeting — European Commission, Parliament, Council — are currently having closed-door meetings to agree on a final text.
- Meetings ongoing through Christmas 2018
- Upload filters and ‘link tax’ (snippets of news articles must be licensed) are most controversial aspects.
- Text and data mining Article 3 likely (?) to stay the same as current drafts...

Closing remarks – impact on the CLARIN community?

- Lower transaction costs for research activities — certain types of language research will certainly benefit
- These exceptions do not — and should not — replace proper licensing in many cases
- These exceptions are designed for **individual researchers and small teams** — research infrastructures were not taken into account
 - Exceptions do not allow easy sharing
 - Exceptions will impact researchers and individual centers, rather than CLARIN

Paweł Kamocki, Erik Ketzan, Julia Wildgans & Andreas Witt

New exceptions for Text and Data Mining and their possible impact on the CLARIN infrastructure