



Foto: Trabold/IDS

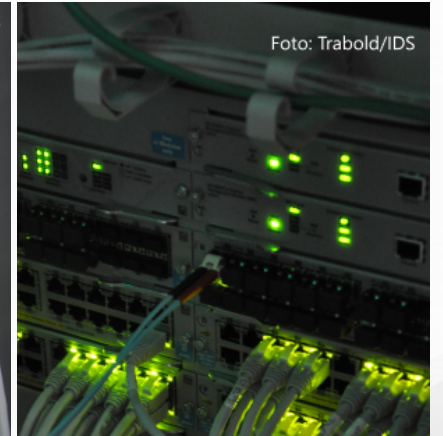


Foto: Trabold/IDS

Paweł Kamocki

## NEW COPYRIGHT EXCEPTION FOR TEXT AND DATA MINING FOR SCIENTIFIC RESEARCH PURPOSES -- A PYRRHIC VICTORY FOR OPEN SCIENCE?



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

## 1. Context

What led to the adoption of the new copyright exception for scientific TDM?

## 2. Content

...of the exception.

## 3. Comment

Is it good for Open Science?

...or rather: how to make the best of it for Open Science?

---

## CONTEXT:

# COPYRIGHT LANDSCAPE IN THE EU

- **copyright axiom:** idea/fact - expression dichotomy
    - pure information (fact, idea) is not protected, only its expression is
  - Directive on copyright in the Information Society (2001)
    - broadly defined **exclusive rights**
      - **reproduction**
      - **communication to the public**
    - 21 narrowly defined, (mostly) optional **exceptions**
      - non-commercial research, citation, parody, uses made by libraries...
  - low threshold of **originality** ('author's own intellectual creation')
  - CJEU's tendency for **extensive interpretation** of originality and exclusive rights
-

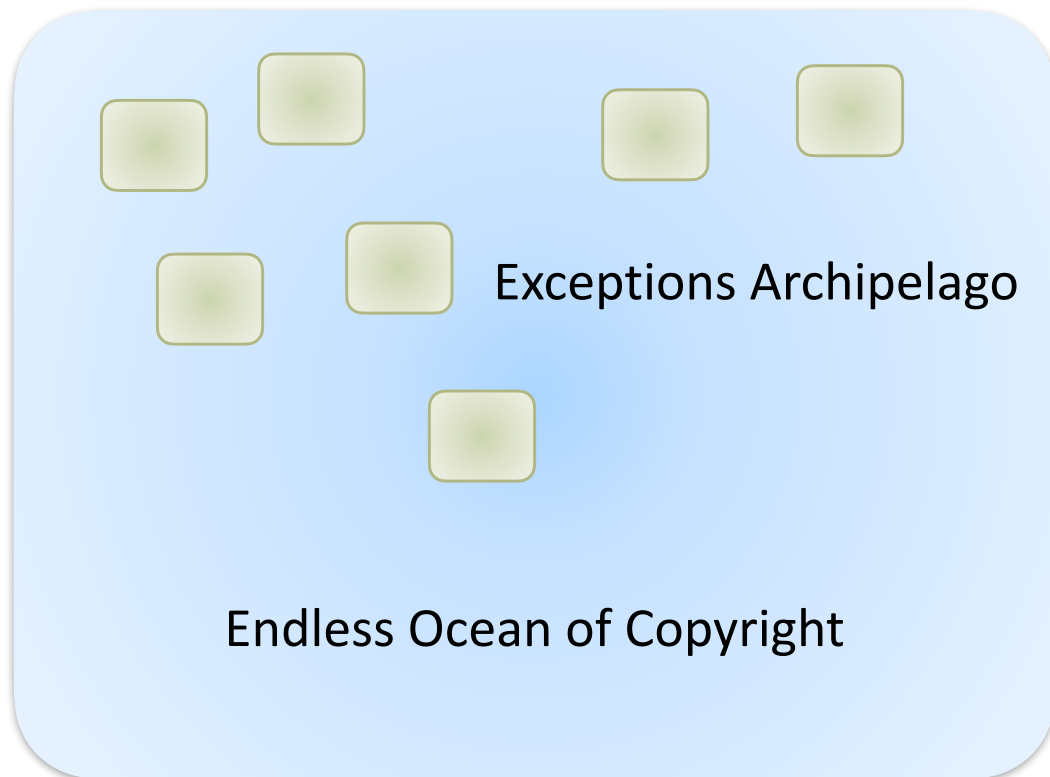
## CONTEXT:

# COPYRIGHT LANDSCAPE IN THE EU VS. THE US

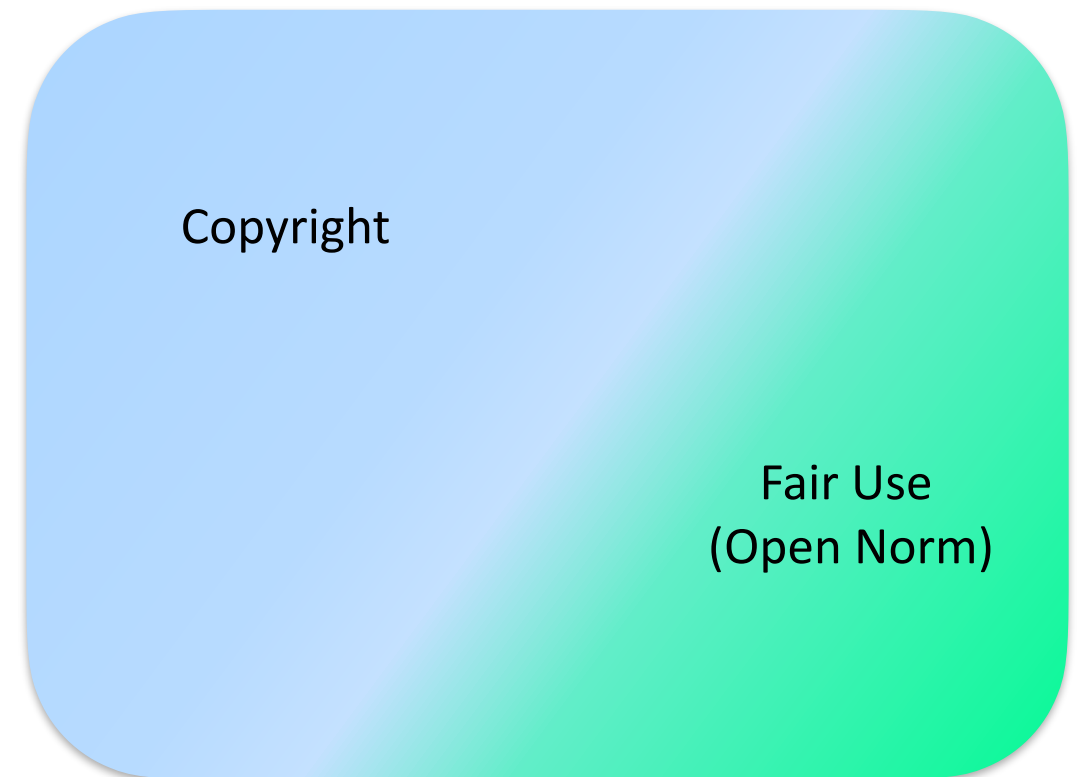
IDS

LEIBNIZ-INSTITUT FÜR  
DEUTSCHE SPRACHE

## European Union



## United States



## CONTEXT: THE TDM PROBLEM

- **TDM** entails acts of **reproduction** (copying)
  - research exceptions insufficient (narrow transpositions, non-commercial research only)
  - how to do TDM legally?
    - authorisation (license) from the author [*Licenses for Europe*]
    - dedicated copyright exception
      - ‘national’ TDM exceptions in the UK (2014), France (2016), Germany (2017) - limited to non-commercial research
      - re-definition of exclusive rights, introduction of an open norm...
  - the research community generally called for a dedicated exception
-

## CONTEXT: THE DSM DIRECTIVE

- **Directive on Copyright in the Digital Single Market**
    - September 2016 — first draft published by the EC
    - until April 2019 — turbulent legislative process
    - 7 June 2019 — publication (entry into force)
    - 7 June 2021 — deadline for transposition
  - **Art. 3** — exception for TDM for scientific research purposes
  - Art. 4 — general TDM exception (rightholders can ‘opt-out’, no storage of copies, no sharing)
-

## CONTENT:

### DEFINITION OF TDM AND BENEFICIARIES

- **TDM:** *any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations*
  - **beneficiaries** of the exception:
    - **research institutions** (universities, research institutions)
      - carry out research on a non-profit basis or re-invest profits in research
      - public-interest mission recognised by the State (e.g. through funding)
      - public-private partnerships expressly included
    - **cultural heritage institutions** (libraries, museums, archives)
-

## CONTENT: PERMITTED ACTS

- **reproduction** of protected material
    - sole purpose: TDM for scientific research (no ‘close reading’)
    - condition: lawful access (license/subscription, open Internet)
      - *the right to read is the right to mine*
  - **copies** (corpora):
    - can be re-used for research purposes (incl. for verification of results)
    - should be stored ‘with an **appropriate level of security**’
  - **result** (output, e.g. trained model) is copyright-free [?]
  - no communication to the public allowed (...apart from quotations in research papers?)
-



## CONTENT: TECHNICALITIES

- the exception **overrides contracts**
  - rightholders can apply **measures necessary to ensure the security and integrity** of their networks or databases
    - BUT the exception can **override technological protection measures** (the procedure is cumbersome and rarely used)
  - Member States should encourage adoption of **best practices** regarding secure storage of copies and technological protection measures on content
  - important: **can be 'mixed' with the exception for non-commercial research** (e.g. in Germany)
-

# CONTENT: NATIONAL TRANSPOSITIONS

Aspect	Germany (60d UrhG)	France (L122-5-3, II CPI)
Citizen scientists	yes, for non-commercial purposes only	no
Sharing of copies	<ul style="list-style-type: none"><li>• limited circle of persons for joint research (must stop upon completion of the project)</li><li>• individual persons for verification of results</li><li>• only for non-commercial institutions</li></ul>	no

---

## COMMENT: MANY FRICTION POINTS

- **‘the exception that proves the (new) rule’**
    - now it’s clear that every TDM use not covered by the exception needs authorisation (license)
  - **no sharing** (at best: very limited)
  - **lawful access** requirement
    - increase of subscription costs for research institutions?
    - BUT: preserves scientific publishers
  - requirement for **secure storage** —> chilling effect?
  - unclear relation with **technological protection measures** —> digital arms race?
-

## COMMENT: IMPACT ON OPEN SCIENCE

- **disincentivisation of data licensing** and open sharing
    - **FAIR principle R1.1:** *Data are released with a clear and accessible data usage license*
    - license = authorisation from rightsholders
    - why seek authorisation if there is a (limited) exception?
    - why license my data if research institutions can use them without a license?
  - solutions?
    - Open Access / Open Science **mandates** (research institutions, funders)
    - role of **research infrastructures** to mitigate problems with secure storage / sharing
-

kamocki@ids-mannheim.de

THANK YOU FOR YOUR ATTENTION!