
Articles 3 and 4 Text and Data Mining Exceptions

Maja Bogataj Slovenian Intellectual Property Institute

Benjamin White Centre for Intellectual Property Policy & Management,
Bournemouth University



What is at issue in Articles 3 & 4?

The context

- Big data is found everywhere and growing
- The importance of Big Data for Artificial Intelligence (AI)
- European based data analytics is estimated to grow to \$10.3 billion by 2021
- TDM = Text and Data Mining:
“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.” (A2.2.)

What is at issue in Articles 3 & 4?

In a nutshell

Only some EU member states have TDM exceptions.

In others there is no legal certainty for researchers and/or others involved in data analytics.

This fragments data analytics and hinders research, innovation, economic growth and competitiveness of the EU.

Articles 3 & 4, if well implemented, partially solve some of these issues.

What is at issue in Articles 3 & 4?

Two different exceptions

Article 3

- **mandatory** exception
- gives freedom to mine to **researchers in educational establishments / CHIs**
- for **research purposes**
- who have **legal access** to open web and to collections of universities, libraries, archives, other CHIs
- **protected from contracts or TPMs**

Article 4

- **mandatory** exception or limitation
- gives freedom to mine to **everyone**
- who has **legal access**
- **except when rightholders expressly prohibit**

Who can benefit from these new exceptions?

What is at issue in Articles 3 & 4?

Article 3

Anyone who has **lawful access** to content (collections and open web) via **cultural heritage institutions, universities or other research orgs for research purposes** (including commercial research) (A2.1, A2.3, A3.1, R11, R12, R14)

Article 4

Anyone who has **lawful access** to content and is not prevented by TDM and / or by contract (A2.1., A2.3., A4.1, A7, R18)

Who can make the reproduction and extractions?

Article 3

Universities and other research orgs and cultural heritage institutions (A3.1.) including the persons attached thereto (R14)

Article 4

Doesn't say - anyone who has lawful access

**What is
at issue in
Articles 3 & 4?**

What can be text and data-mined?

What is at issue in Articles 3 & 4?

Article 3

- any type of work subject to copyright
- works covered by neighbouring rights and the new publication right (A15)
- any type of work subject to *sui generis* database right
- **does not create a specific exception to the Software Directive**

Article 4

- any type of work subject to copyright
- works protected by neighbouring rights and the new publishers right (A.15)
- any type of work subject to *sui generis* database right

What copyright restricted acts are covered?

What is at issue in Articles 3 & 4?

Article 3

- reproductions
- extractions and/or reutilization

Article 4

- reproductions
- extractions and/or reutilization

Have to assume that the reproduction right includes adaption, translation etc otherwise the exception would fail.

Is commercial TDM allowed?

What is at issue in Articles 3 & 4?

Article 3

- Yes. No limits to scientific research*
- Research orgs and CHI “should be able to rely on their private partners for carrying out their text and data mining.” (R.11)

Article 4

- Yes.

**Conversely, organisations upon which commercial undertakings have a decisive influence allowing such undertakings to exercise control because of structural situations, such as through their quality of shareholder or member, which could result in preferential access to the results of the research, should not be considered research organisations for the purposes of this Directive.*

Is TDM subject to remuneration?

What is at issue in Articles 3 & 4?

Article 3

- No, no harm to rightholders (R.17)

Article 4

- Assume not. (R17: first part is about researchers, but second part talks about TDM exceptions not being subject to remuneration.)

Can the results of TDM be shared?

What is at issue in Articles 3 & 4?

Article 3

- No exception to the communication to the public right
- Nothing about sharing results

Article 4

- No exception to the communication to the public right
- Nothing about sharing results

Is EU wide / cross border TDM allowed?

What is at issue in Articles 3 & 4?

Article 3

- **Mandatory =
harmonized at
EU-level**

Article 4

- **Mandatory =
harmonized at
EU-level**

How long can derived data be stored?

What is at issue in Articles 3 & 4?

Article 3

- No limit set
- Researchers are in a best position to decide (A3.2.)

Article 4

- “As long as necessary for the purposes for text and data mining” (A4.2.)

Contractual Overrides

What is at issue in Articles 3 & 4?

Article 3

- “any contractual provision contrary to the exception... shall be unenforceable” (A.7)

Article 4

- “The exception or limitation... shall apply on condition that the uses of works and other subject matter...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” (A4.3.)

What is at issue in Articles 3 & 4?

Article 3

- measures to ensure the security and integrity of the networks and databases where their works or other subject matter are hosted.
 - shall not go beyond what is necessary to achieve that objective
 - proportionate
 - should not be excessive
 - should not undermine the effective application of the exception.
- (A3.3. R16)

Article 4

How to deal with Articles 3 & 4?

- **Concerns:**
 - **Technical protection measures**
 - **Special measures requiring secure storage**
 - **Recitals discussing depositing data with trusted intermediaries**
 - **Limitation of data retention**
 - **Problem if Robot Exclusion Standard would not be agreed upon for open web**
 - **Remuneration**
 - **Does TDM = AI?**

The Best Version of Article 3

- Demand access within maximum of 72 hours when TPMs are preventing data mining.
- Reject any specific measures regarding secure storage.
- Oppose any requirement to deposit data with a trusted intermediary.
- Discuss with the legislature to include Software under 3

Go Beyond Article 3

- Require **remote access** in national law when data mining relates to **digitised analogue** items. e.g newspapers etc.
- Advocate for the introduction of **a clear exception for sharing** the results of data mining.
- Protect all copyright exceptions from override by contracts.

The Best Version of Article 4

- **Advocate for a maximum 72 hour response time in law where access is being blocked. Compensation regime when it goes beyond this.**
- **Reject any specific limitations on being able to retain data.**
- **Reject any calls for the exception to be subject to remuneration.**
- **Require that only Robot Exclusion Standard is used to disallow data mining on websites accessible on the open web.**

Go Beyond Article 4

- Require **remote access** in national law when **data mining digitised analogue items**.
- Advocate for the introduction of a **clear exception for sharing** the results of data mining.
- Call for **clear language in contracts** when rightsholders wish to reserve rights in materials not available on the open web.

The Optimal National TDM Exceptions

For member states that want to become the center for Big Data and AI: **apply all recommendations**



The Worst Version of implementation

- A restrictive interpretation of text and data mining to exclude AI
- a very restrictive interpretation of the exception with limitations around public private partnerships under Article 3,
- inappropriate storage obligations which clash with European Commission and funder policies under Article 3
- slow / unclear resolution when researchers are blocked from data mining due to technical protection measures
- not adopting robot.txt to exclude mining under Article 4, would lead to tech start ups and other corporations doing mining outside the EU
- remuneration under Article 4



How to adapt this locally?

- **Build partnership**
 - with local researchers, scientists, libraries, companies and others who engage in data analytics exceptions and the new one
- **Get local legal support:**
 - essential where TDM exceptions already exists.
 - important to understand to what extent the national law already allows remote mining of analog materials that have been digitised as well as to share the results of mining.

Thank you!

@majamajab
maja.bogataj@ipi.si

@copywhitelaw
copyrighteu@gmail.com

