Copyright and language research

Aleksei Kelli
Professor of IP Law
aleksei.kelli@ut.ee
Only one thing is impossible for God: to find any sense in any copyright law on the planet.

- Mark Twain's Notebook, 1902-1903
“intellectual property” shall include the rights relating to:
– literary, artistic and scientific works,
– performances of performing artists, phonograms, and broadcasts,
– inventions in all fields of human endeavor,
– scientific discoveries,
– industrial designs,
– trademarks, service marks, and commercial names and designations,
– protection against unfair competition,
and all other rights resulting from intellectual activity in the industrial, scientific, literary or artistic fields (art 2).
Stakeholder involvement:
Stakeholder involvement:

They always talk handsomely about the literature of the land....And in the midst of their enthusiasm they turn around and do what they can to discourage it.
- Speech in Congress, 1906

Whenever a copyright law is to be made or altered, then the idiots assemble.
- *Mark Twain's Notebook*, 1902-1903

http://www.twainquotes.com/Copyright.html
Elements of copyright system:

- Enforcement
  - Criminal
  - Civil

- Rights
  - Character & scope
  - Exceptions & limitations
  - Duration
  - Character

- Elements of copyright system
  - Right holder / author
  - Subject matter of protection
  - Criteria of protectability

- INTERNATIONAL LAW
- EU / REGIONAL LAW
- NATIONAL LAW
Trends (from Statute of Anne (1710) to InfoSoc directive):

- Duration (from 14 years to 70 pma)
- Widening of copyright protection
- Scope of protection (from literary works to software)
- Rights (from copying to broadcasting & digital rights)
Main paradigms:

- Copyright (right to copy) vs. Author’s right (droit d’auteur);
- Anglo-American (copyright) vs. Continental European (droit d’auteur);
- Positive laws (ius positum: man made law) vs. Natural law (ius naturale: inherent rights given by God, nature or reason).
The works (*œuvres de l’esprit*):

- Copyright subsists in a work, that is to say, *any expression* within the field of *literature, art* or *science* in so far as it constitutes its author’s *own intellectual creation* (European Copyright Code);

- The provisions of this Code shall protect the rights of authors in all works of the mind, whatever their kind, form of expression, merit or purpose (French IP Code).
The works (œuvre de l’esprit):

- Sweat of the brow doctrine;
- Creativity: choice & arrangement;
- Originality = independent creation + creativity;
- Independent creation;
- Any expression within the field of literature, art or science;
- Expressed in an objective form and can be perceived and reproduced in this form;
- Protection does not depend on societal value of a work.
Transformative use:

- Tension between a need to protect copyright and allow others to build upon it:
  - Legislative initiatives;
  - Private initiatives (viral copyright).
Language data & copyright:

- Press publications
- Written text
- Videos
- Oral text
- Performances
- Copyrighted databases
- Sui generis databases
Example of language data:

Two tiers of rights covering language resources

| Language resources as a database | • Copyright  
|                                | • Related rights  
|                                | • Privacy rights  
|                                | • Personal data protection |

| Material used to develop language resources | • Copyright  
|                                            | • Related rights  
|                                            | • Privacy rights  
|                                            | • Personal data protection |
Intellectual creations not protected by copyright:

- Intellectual creation is not a work / excluded works:
  - Ideas, methods, inventions,
  - Works of folklore;
  - Legislation & court decisions;
  - Facts and data;
  - Scène à faire;
  - Merger doctrine.
Intellectual creations not protected by copyright:
Moment of creation of copyright:

- No formalities (e.g., registration and other requirements);
- Copyright in a work is created with the creation of the work;
- Benefits of formalities:
  - Title search;
  - Certificate of authorship;
  - Statutory damage + attorneys fees.
Everyone has the **right to the protection of the moral and material interests** resulting from any scientific, literary or artistic production of which he is the author (The Universal Declaration of Human Rights art. 27 (2))
Moral rights:

- The right of attribution
- The right of integrity
- The right of disclosure
- The right of withdrawal or repentance
Economic rights:

• Open vs. closed catalogue of rights;
• Reproduction right;
• Distribution right;
• Right of translation;
• Right of alteration of the work;
• Right to communicate works to public;
• Making available right;
• The interplay between economic and moral rights.
Limitations and exceptions to copyright:

- Policy consideration:
  - Freedom of expression;
  - Participatory democracy;
  - Right to culture;
  - Transaction costs;

Exceptions:
- Closed list of exceptions
- Open list of exceptions
Use of LD & copyright protection:

- Quotation right (InfoSoc Art. 5 (3))
- Right holder’s consent
- Text and data mining exception (Digital Copyright Directive Art. 3, 4)

Legal bases for using data-sets:

- Temporary acts of reproduction exempted from the reproduction right (InfoSoc Art. 5 (1))
- Private use exception (InfoSoc Art. 5 (2))
- Research exception (InfoSoc Art. 5 (3))
TDM (Directive 2019/790): 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 (Art. 2);

Text and data mining can also be carried out in relation to mere facts or data that are not protected by copyright, and in such instances no authorisation is required under copyright law (Recital 9).

TDM (Directive 2019/790): There can also be instances of text and data mining that do not involve acts of reproduction or where the reproductions made fall under the mandatory exception for temporary acts of reproduction provided for in Article 5(1) of Directive 2001/29/EC, which should continue to apply to text and data mining techniques that do not involve the making of copies beyond the scope of that exception (Recital 9).
TDM beneficiaries (Art. 2):

**Research organisation**: a university, including its libraries, a research institute or any other entity, the primary goal of which is to **conduct scientific research or to carry out educational activities** involving also the conduct of scientific research:

(a) on a **not-for-profit basis** or by reinvesting all the profits in its scientific research; or

(b) pursuant to a public interest mission recognised by a Member State; in such a way that the access to the results generated by such scientific research **cannot be enjoyed on a preferential basis** by an undertaking that exercises a decisive influence upon such organisation;

**Cultural heritage institution**: a publicly accessible library or museum, an archive or a film or audio heritage institution.
TDM and public-private partnership:

While research organisations and cultural heritage institutions should continue to be the beneficiaries of TDM exception, they should also be able to rely on their private partners for carrying out text and data mining, including by using their technological tools (Recital 11).
TDM conditions:

- Right to make reproductions and extractions for TDM;

- Subject matter covered:
  - Works;
  - Objects of related rights:
    - Performances;
    - Phonograms;
    - Broadcasts;
  - *Sui generis* databases;
  - Press publications
TDM (Directive 2019/790):

**Lawful access**: This exception or limitation should only apply where the work or other subject matter is accessed lawfully by the beneficiary, including when it has been made available to the public online (Recital 18);

Any **contractual provision contrary** to the exceptions shall be **unenforceable** (Art. 7).
TDM and technological measure: Rightholders are allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective (Art. 3 (3)); Member States shall take appropriate measures to ensure that rightholders make available to the beneficiary of an exception the means of benefiting from that exception (InfoSoc Art. 6 (4))
TDM datasets:

• Copies of works or other subject matter made in compliance with paragraph 1 shall be stored with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results (Art. 3 (2));

• Uses for the purpose of scientific research, other than text and data mining, such as scientific peer review and joint research, should remain covered by research exception in InfoSoc.
Text and data mining (Digital Copyright Directive Art. 3)

Any contractual provision contrary to the exception is unenforceable.

Limitation for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access

Definition: 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 (Digital Copyright Directive Art. 2)

Research organisations: a) on a not-for-profit basis or by reinvesting all the profits in its scientific research; b) pursuant to a public interest mission recognised by a Member State

Cultural heritage institutions: a publicly accessible library or museum, an archive or a film or audio heritage institution

The entitled party (Digital Copyright Directive Art. 2)

Right to make reproductions and extractions

Right to store copies

Limitation of technological protection measures to guarantee TDM (Art. 6 (4) of InfoSoc)

Exception to

Reproduction right (Art. 2 of InfoSoc)

Reproduction right of copyrighted database (Art. 5 of Database Directive)

Extraction and/or re-utilization right related to sui generis database (Art. 7 of Database Directive)

Reproduction right related to software (Art. 4 of Software Directive)

Press publisher’s right (Art. 15 of Digital Copyright Directive)

with an appropriate level of security

for the purposes of scientific research (e.g., verification)
Liability of depositories:

- **No general obligation to monitor**
- **Hosting service provider is not liable if (Art. 14-15 of Directive on electronic commerce 2000/31)**
  - Information society service is provided that consists of the storage of information provided by a recipient of the service
  - The provider does not have actual knowledge of illegal activity or information and, as regards claims for damages
- **The provider, upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information**
Liability of depositories:

Information society service provider

Purpose is to store and give the public access to a large amount of copyright-protected works or other protected subject matter uploaded by its users

Organises and promotes content for profit-making purposes

Online content-sharing service provider (Directive 2019/790)

The following are not covered

Not-for-profit online encyclopedias

Not-for-profit educational and scientific repositories

Open source software-developing and-sharing platforms

Online marketplaces

...
Tänan