Experimenting TDM
(I, Robot?)

Julien Cabay (Associate Professor ULB and ULiège, Faculty of Law, JurisLab)

julien.cabay@ulb.be

Thomas Vandamme (Ph.D. candidate ULB, Faculty of Engineering, LISA)

thomas.vandamme@ulb.be

Outline

- Introduction
- Experimenting TDM in IPSAM
- TDM: Technical Aspects
- An overview of TDM exceptions in the EU
- Belgian implementation
- Some practical and legal hurdles in IPSAM
- Call for action

Introduction

- **Text and Data Mining (TDM)?**
  - Art. 2(2) DSM Dir. : « 'text and data mining' means 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 »

- **Two exceptions**
  - Art. 3 (scientific research)
  - Art. 4 (general)

- **Much more than copyright…**
  - *« (…) It may be argued that under the misleading label Text and data mining (TDM) what has been regulated at the EU level in Arts. 3 and 4 goes far beyond a mere copyright exception. In fact, it should be reclassified as the legal regulation of AI via the allocation of property rights in its building blocks, or in other words, as a property-right approach to the regulation of AI (…) »* (Margoni/Kretschmer)

Introduction

- TDM Application
  - Search Engines
  - Consumer Statistics
  - Data-Intensive Applications :
    - Neural networks (see GPT trained on Common Crawl, 3.15B webpages)
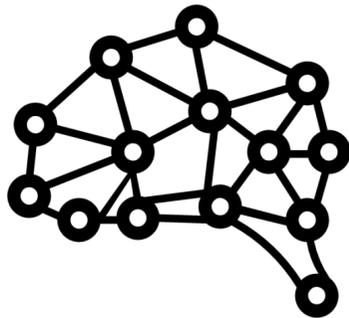
Experimenting TDM in IPSAM

- Exploratory research : Assessing **TM search engines' performances**
- Using EUIPO **Opposition Division decisions**
- Query systems with applicant TM
- Performance -> Position of Opponent TM in output

WIPO (concept filter)

BOIP

EUIPO
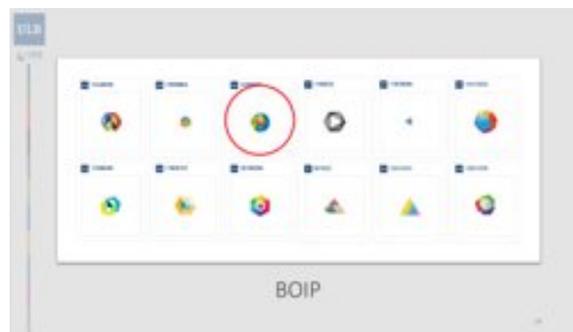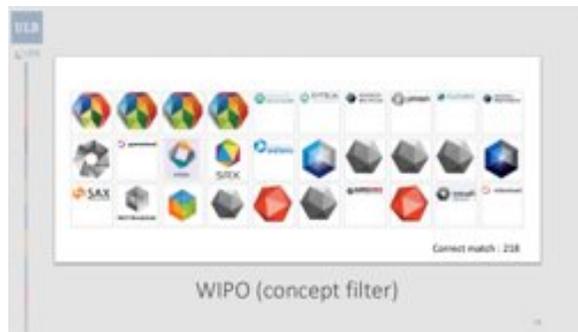
J. Cabay and T. Vandamme, *"Assessing IP Similarities Through Technology: A Trademark Exploration of Challenges and Avenues"*, *AI Tech & Policy Talks*, University of Geneva, 4 November 2021

4

Experimenting TDM in IPSAM

- Run experiment for 8k decisions (all EUIPO Opp. Div. From 23/03/2016 to 31/05/2022)

- => Necessity to automate

Forthcoming : T. Vandamme, J. Cabay, O. Debeir, "A Quantitative Evaluation of Trademark Search Engines' Performances through Large-Scale Statistical Analysis ", *ICAIL '23: Nineteenth International Conference for Artificial Intelligence and Law, Braga Portugal, June 19- 23, 2023*

TDM: Technical Aspects

- Data acquisition :
  - Methods : Applications Programming Interface (APIs, gateways made available by services), Web Scrapers (robots using human interfaces)
  - Challenges : server issues and limitations, various protections (IP blocking, Captchas)

- Data treatment :
  - Methods : Parsing (pdfs, docs, ..), cleaning and filtering, relational information
  - Challenges : Irregular data formats, broken files and data

- Data analysis : models, insights, applications
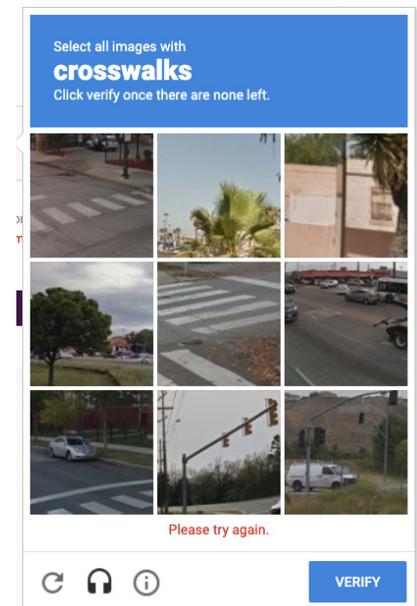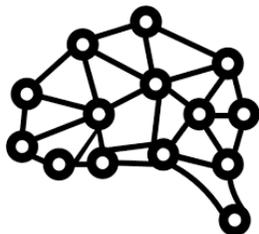
An overview of TDM exceptions in the EU

« (…) *in the Union, such organisations and institutions are confronted with* ***legal uncertainty*** *as to the extent to which they can perform text and data mining of content. In certain instances,* ***text and data mining can involve acts protected by copyright****, by the sui generis database right or by both, in particular, the reproduction of works or other subject matter, the extraction of contents from a database or both which occur for example when the data are normalised in the process of text and data mining.* ***Where no exception or limitation applies, an authorisation*** *to undertake such acts* ***is required*** *from rightholders* »

(Recital 8, DSM Dir.)

An overview of TDM exceptions in the EU

- Relevance of reproduction right?
  - CJEU, *Infopaq*, C-5/08 (2009) « ***An act occurring during a data capture process (…) is such as to come within the concept of reproduction in part (…) if the elements thus reproduced are the expression of the intellectual creation of their author*** »
  - Critics: no « **use of a work as a work** » (Ducato/Strowel)
- Irrelevance of previous exceptions?
  - Limited scope, in particular in relation to commercial uses
    - Private use exception for AI business? Art. 5(2)(b) InfoSoc Dir.: « *neither directly nor indirectly commercial (…)* »
    - Research exception for AI business? Art. 5(3)(a) InfoSoc Dir.: « *non-commercial purpose* »
    - Exception for transient copies for AI business? Art. 5(1) InfoSoc. Dir.: « *no independent economic significance* »

An overview of TDM exceptions in the EU

Art. 3 (scientific research TDM)	Art. 4 (general TDM)
(1) Member States shall provide for an exception to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, and Article 15(1) of this Directive for 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.	(1) Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining.
(2) 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.	(2) Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining.
	(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.
Juncto Art. 7(1) : Any contractual provision contrary to the exceptions provided for in Articles 3, 5 and 6 shall be unenforceable.	
Juncto Art. 7(2) : The first, third and fifth subparagraphs of Article 6(4) of Directive 2001/29/EC shall apply to Articles 3 to 6 of this Directive.	

An overview of TDM exceptions in the EU

Art. 3 (scientific research TDM)	Art. 4 (general TDM)
(1) Member States shall provide for an exception to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, and Article 15(1) of this Directive for 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.	*(1) Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining.*
(2) 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.	*(2) Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining.*
	(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.
Juncto *Art. 7(1) : Any contractual provision contrary to the exceptions provided for in Articles 3, 5 and 6 shall be unenforceable.*	
Juncto *Art. 7(2) : The first, third and fifth subparagraphs of Article 6(4) of Directive 2001/29/EC shall apply to Articles 3 to 6 of this Directive.*	

An overview of TDM exceptions in the EU

- Many issues in relation to EU provisions, especially in relation to art. 4
  - Application to works made available prior adoption/entry into force of the directive?
    - o *See* CJEU, *Soulier & Doke*, C-301/15 (2016) (no prior information on future use = hypothetical consent)
  - Applications to works wich country of origin is not EU Member States?
    - o *See* art. 5(2) Berne Convention: « The enjoyment and the exercise of these rights shall not be subject to any formality (...) »
  - Application to works not lawfully made available online?
    - o *See* in particular CJEU, *Renckhoff*, C-161/17 (2018) (unauthorized making available online of work already lawfully made available online)
- Many different regimes worldwide with broader reach

An overview of TDM exceptions in the EU (and beyond)

	EU	China	USA	Japan
TDM	Art. 3 (scientific research) and 4 (general) Dir. 2019/790	Closed list, no TDM but 'open' clause (Art. 24(13) refering to 'other circumstances' possible 'fair use/fair dealing'	17 USC 107 (*fair use*, after *Authors Guild v. Google Inc.*, 804 F.3d 202 (2d Cir. 2015)	Art. 30-4 Japanese Copyright Act (exception 'not for enjoying the idea or emotions expressed in a work')
Reach	**Limited**	« Chinese courts have realised the disadvantage of limiting fair use to enumerated circumstances and brought in a **more flexible approach for finding fair use** by learning from their US counterparts » (Wang, He)	« **Uses involving robotic readers are fast-tracked for fair use** » (Grimmelmann) ; but doubts as to « expressive machine learning » (Sobel)	« **Japan as a paradise for machine learning** » (Ueno)

Belgian implementation

- Art. 2(2) (definitions)
  - I.13, 10° (text and data mining)
  - I.13, 9° (research organisation)
  - No use of wording 'cultural heritage institution' (and no definition), directly implemented in substantial provisions (*see Doc. Parl.*, *Doc. parl.*, Chambre, 2021-2022, No 2608/1, p. 66)

-

DSM Dir. / CDE	3 (scientific research TDM)	4 (general TDM)	7(1) (contradictory contractual provisions unenforceable)	7(2) (technological protection measures)
General © ; Database ©	XI.191/1, §1er, 7°	XI.190, 20°	XI.193	XI.291, § 2 *
Software ©	-	XI.299, § 5	XI.301	XI.291, § 2 *
Database *sui generis* right	XI.310, § 3, 1°	XI.310, § 3, 2°	XI. 314	XI.291, § 2 *✢
Neighbouring right, including new publisher's right	XI.217/1, 6°	XI.217, 19°	XI.219	XI.291, § 2 *

Belgian implementation

- Explicit choices:
  - **No remuneration** for (scientific research) TDM exception
    - // recit. 17
    - *See Doc. Parl.*, *Doc. parl.*, Chambre, 2021-2022, No 2608/1, p. 57 (limited prejudice)
  - General TDM opt-out in the case of content made publicly available online **only if machine-readable means**
    - Art. XI.190, 20° CDE: « *En ce qui concerne les contenus mis à la disposition du public en ligne, la réservation n'est considérée appropriée que si elle est effectuée au moyen de procédés lisibles par machine.* »
    - *Contra a*rt. 3(3) (« such as ») ; // recit. 18
    - *See Doc. Parl.*, *Doc. parl.*, Chambre, 2021-2022, No 2608/1, p. 49 (*contra* opinion Conseil d'Etat *; « cette manière de transposer favorise la sécurité juridique tant des titulaires de droits que des utilisateurs »*)
  - General TDM exception also '**imperative**', if contractual reservation not deemed 'approriate'
    - Comp. art. 7(1), that only applies to scientific research TDM
    - *See Doc. Parl.*, *Doc. parl.*, Chambre, 2021-2022, No 2608/1, p. 57 : « *L'exception de l'article 4 concernant l'extraction de textes et de données permet aux titulaires de droits de faire une "réserve appropriée". Ainsi, on ne peut déroger à cette exception que si une telle réserve appropriée est faite. S'il n'y a pas de réserve appropriée, on ne peut s'écarter de cette exception, quelle que soit la stipulation contraire. L'interprétation de la réserve appropriée dépend de la jurisprudence. Toutefois, on ne peut exclure a priori que cela puisse se faire par convention.* »

Belgian implementation

- Choices or mistakes?
  - ✣ Technological protection measures reservation for database *sui generis* right in XI.291, § 2
    - ○ XI.291 refers to rightholders in the field of copyright ; should have been implemented in XI.316, § 2 (// in relation to the producer's database *sui generis* right)
  - * XI.291, § 2 (voluntary measures by rightholders to ensure benefit of exceptions, notwithstanding technological protection measures), does not apply to works made available to the public online on agreed terms (*see* XI.291, § 3)
    - ○ = all works freely available online? (maybe yes, *see* Dusollier)
    - ○ If yes, contradiction with
      - Art. 7(2) DSM Dir. *juncto* art. 6(4), suparagraph 4 InfoSoc Dir. (not refered to)
      - Recit. 14 in relation to scientific research TDM : *« Lawful access should also cover access to content that is freely available online »*
      - Systematic interpreation Art. 3 and 4 DSM Dir. and Art. 6(4), suparagraph 4 InfoSoc Dir. (scientific research TDM on works freely accessible online should not be subject to contractual arrangement)

Some practical and legal hurdles in IPSAM

- Data acquisition online => ???
  - Robot.txt
    - o = TDM reservation through machine-readable mean?
    - o = Technological protection measure?
    - o Disregard (scientific research TDM), VPN to avoid 'IP blocking'
      - ▪ Lawfull access?
      - ▪ = Circumvention?
  - Captcha
    - o = TDM reservation through machine-readable mean?
    - o = measures to ensure the security and integrity of the networks and database?
    - o = Technological protection measure?
    - o Use of Mechanical Turk (cheap outsourced labor)
      - ▪ = Lawfull access?
      - ▪ = Circumvention?
  - General/scientific research TDM overlap, yet former shall not affect the latter (art. 4(3) DSM Dir.) => solution in contractual arrangement?
    - o = compatible with scientific research TDM?
    - o = compatible with Academia?
    - o = compatible with Open Science?
- Data treatment => OK
- Data analysis => OK

?

Applicable Law?
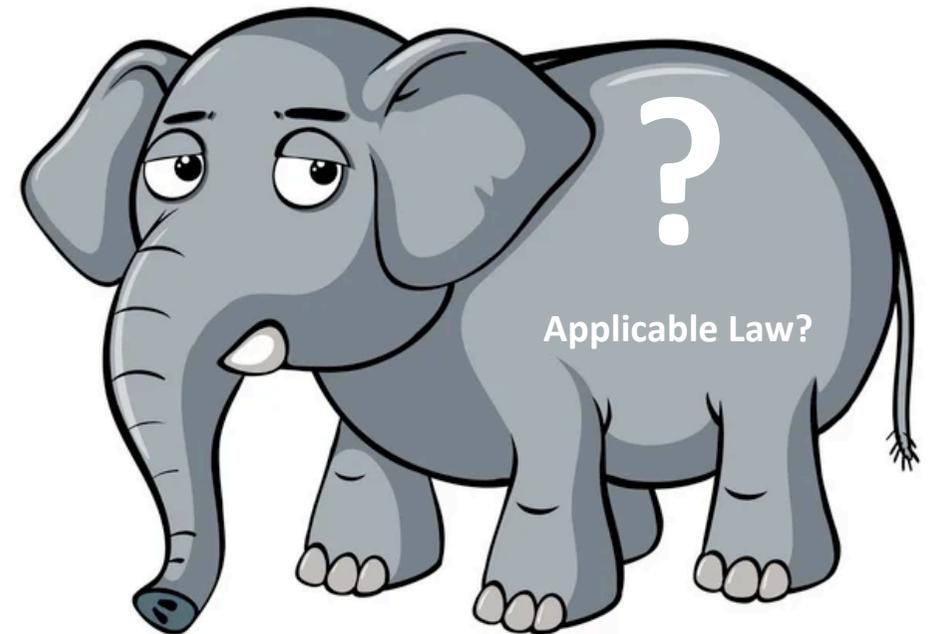
Call for action

- Need for more legal certainty & technical standardisation

- Belgium
  - Making use of Art. 3(4) DSM Dir. : *« Member States shall encourage rightholders, research organisations and cultural heritage institutions to define commonly agreed best practices concerning the application of the obligation and of the measures referred to in paragraphs 2 and 3 respectively »*

- EU
  - Amending the law with the introduction of **country of origin principle** (e.a. art. 5(3) DSM Dir.)

- Int'l
  - Reaching an **int'l consensus** to adress TDM (arguably outside of current int'l copyright law, *see* Senftleben)
  - Setting **technical standards** (TDM.txt // robot.txt?)
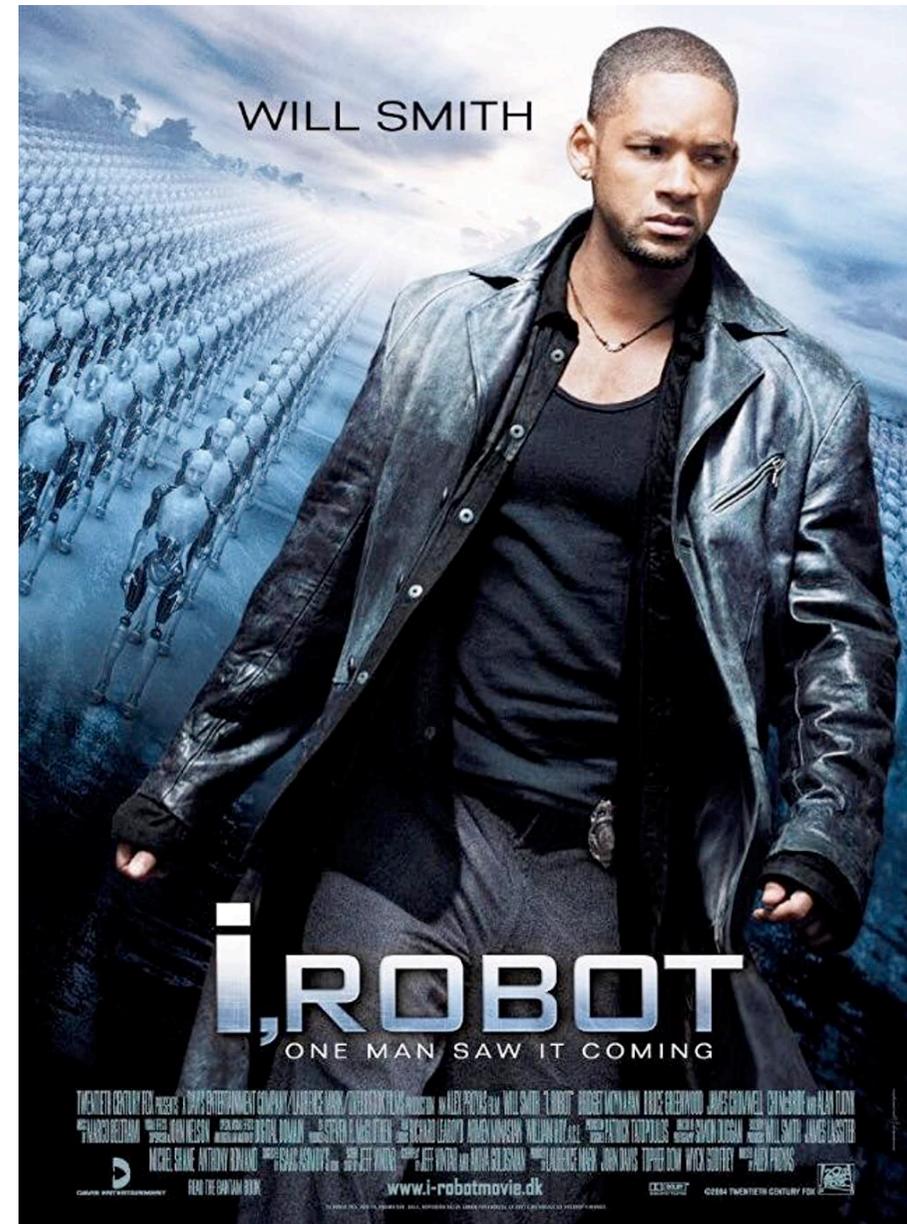
Cited sources

- Ducato, R., Strowel, A., "Ensuring text and data mining: remaining issues with the EU copyright exceptions and possible ways out", *European Intellectual Property Review,* 2021, 43(5), 322-337

- Dusollier, S., *Droit d'auteur et protection des œuvres dans l'univers numérique : droits et exceptions à la lumière des dispositifs de verrouillage des œuvres*, Bruxelles, Larcier, 2007

- Grimmelmann, J., "Copyright for Literate Robots", *101 Iowa Law Review 657* (2016)

- Margoni, T., Kretschmer, M., "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology" (July 14, 2021). Available at SSRN: https://ssrn.com/abstract=3886695 or http://dx.doi.org/10.2139/ssrn.3886695

- Sobel, B. L. W., "Artificial Intelligence's Fair Use Crisis", *41 Columbia Journal of Law & the Arts 45* (2017)

- Senftleben, M., "Compliance of National TDM Rules with International Copyright Law: An Overrated Nonissue?", *IIC - International Review of Intellectual Property and Competition Law*, 53 (10), 2022, 1477-1505

- Ueno, T., "The Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication", *GRUR International*, 70(2), 2021, 145–152

- Vandamme, T., Cabay, J., Debeir, O., "A Quantitative Evaluation of Trademark Search Engines' Performances through Large-Scale Statistical Analysis ", *ICAIL '23: Nineteenth International Conference for Artificial Intelligence and Law, Braga Portugal, June 19- 23, 2023*

- Wang, J., He, T., "To share is fair: The changing face of China's fair use doctrine in the sharing economy and beyond", *Computer Law & Security Review*, 35(1), 2019, 15-28

Many thanks for your attention, comments and questions !

julien.cabay@ulb.be
thomas.vandamme@ulb.be

Find out more about IPSAM :
https://droit-prive.ulb.be/ipsam-adressing-intellectual-property-relevant-similarities-in-images-through-algorithmic-decision-systems/