

# Text and Data Mining (TDM), Machine Learning & Copyright

WIPO Conversation on Intellectual Property (IP) and Frontier Technologies

Panel 4

22-23 September 2021

Carlo Scollo Lavizzari

Lenz Caemmerer Attorneys, Basel, Switzerland

# Why is TDM important? – opportunities and current challenges

- **New tool to use copyright-protected and other content**
- **TDM applicable and so far of distinct value for:**
  - **biomedical research (e.g. PubGene)**
  - **security (Echelon, Europol)**
  - **sentiment analysis (marketing, public relations)**
  - **machine learning: training, calibration, de-biasing, testing, verification**
  - **searching (Google, Bing, DuckduckGo)**
- **Current challenges:**
  - **fast evolving demand for published content (requests in non-commercial space growing)**
  - **missing building blocks and difficulty of defining customer in TDM projects (case by case defining and building a bespoke minable haystack needed)**
  - **creation and transfer of economic value to “tech” third-parties for a ridiculously low return**

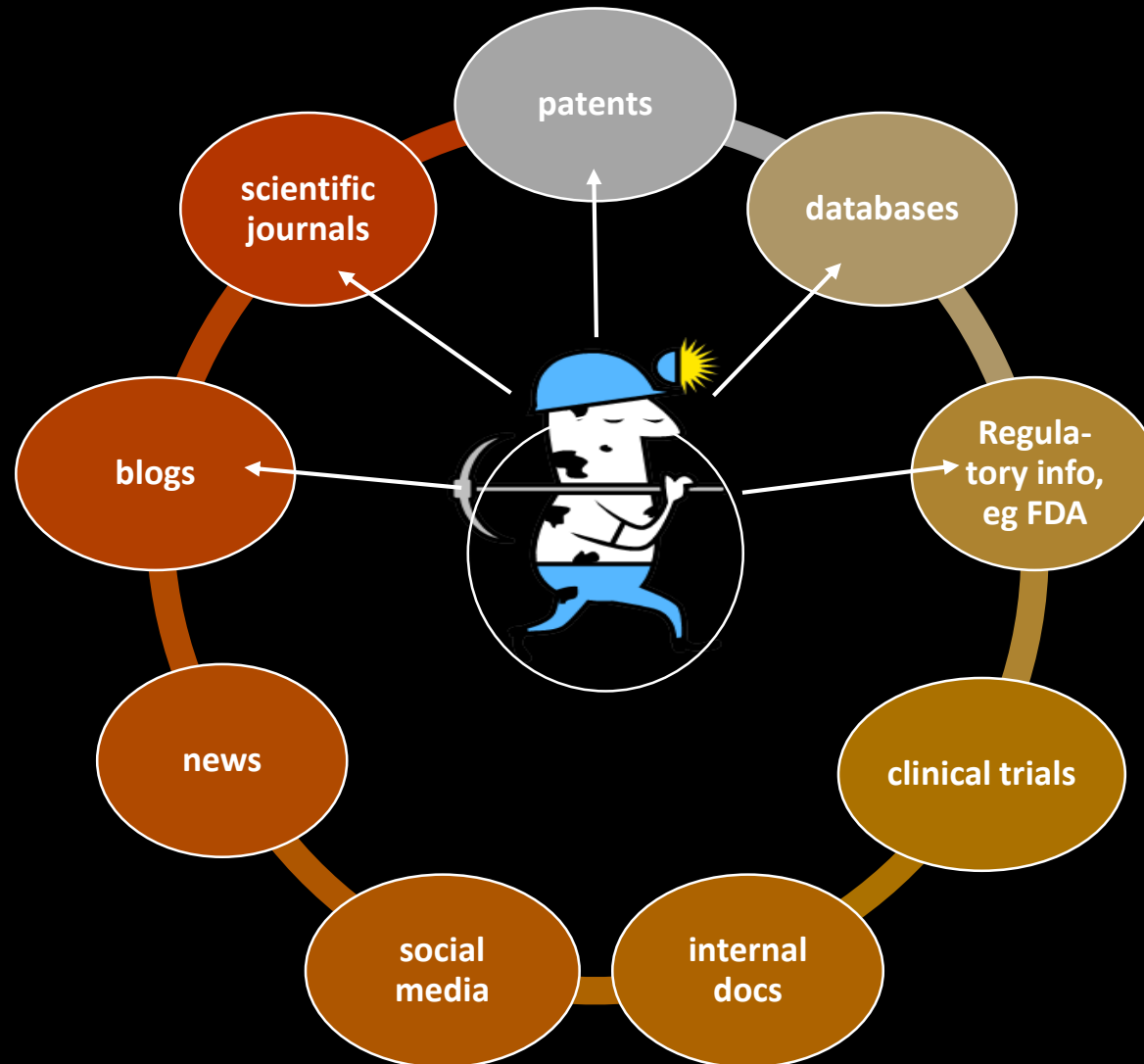
Where is the knowledge we have lost in information?

*T. S. ELIOT, CHORUSES FROM THE ROCK, 1934*

The answer:

*“enable successful TDM by letting a viable ecosystem evolve: a collaborative system that allows fair returns on investments in necessary TDM infrastructure”*

# The TDM wheel of fortune ...



## TDM stakeholders:

- **Rightsholders:** enrich published content, add value for customers, develop new products
- **Researchers:** identify new hypotheses, discover new patterns, facts, knowledge
- **Corporate R&D** (e.g. pharma): as above, accelerate drug discovery and development, maximise value of information spend
- **Commercial “tool-makers”:** software and other tech companies that develop new products based on cross-publisher mining & entity extraction – patents, drug discovery, product information

# What can text miners learn from coffee machines? – Quality of inputs matters

**Before mining**



**After mining**



## What is the right environment for text and data mining?

- *TDM is the new normal – not possible to exempt wholesale  
See: Berne Convention & TRIPS 3-step test*
- *the future of reading is mining and the future of mining is content worth mining*
- *EU leads the way on content and mining technology – tools & collaboration needed, value transfer to US tech industry would harm European ecosystem*

# Parameters for a sensible TDM exception: EU DSM Article 3 – non-commercial sphere

Beneficiary: public institution

Purpose:  
Non-commercial scientific research

Restricted Acts:  
Reproduction & obligation to store safely & delete

Source:  
Acquired lawful access

Safeguards:  
Platform stability,  
No prejudice to cross-industry standards and best practices



# Parameters for a sensible TDM exception: EU DSM Article 4 – disclaimer

Beneficiary: any user

Purpose:

Any purpose for which  
TDM used

Restricted Acts:

Reproduction & retain  
only until needed

Source:

«Open licenses» or  
online unrestricted  
lawfully accessed

Safeguards:

No RH disclaimer/waiver: see W3C  
Working Group on Disclaimer,  
Platform stability,  
No prejudice to cross-industry  
standards and best practices

Copying for TDM is an infringing activity which requires a license or an exception

### **USA & Japan**

- Commercial uses in USA: licensing
- infringing or fair use, based on fact dependent review when no license is in place, e.g. word count (US «is» vs US «are»)
- Japan allows computerised technical analysis, yet preserves viability of commercial licensing markets

### **EU, UK, Switzerland**

- Commercial uses in EU & UK: licensing
- UK narrow exception for mining of subscribed content for researchers, non-commercial, allows technical safeguards
- EU non-commercial / commercial, if no disclaimer, allows technical safeguards
- obligation to safely store and delete content used in the mining process in EU & UK

*"If your plan is for one year  
plant rice. If your plan is for ten years  
plant trees. If your plan is for one hundred years  
educate children."*

*Confucius*

*Thank you!*

*Carlo*