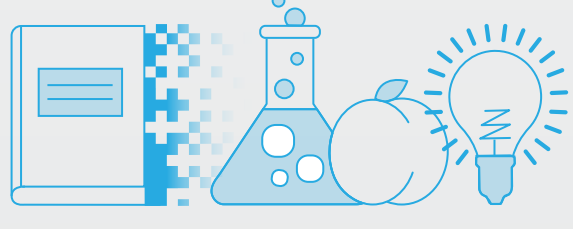# UNLOCKING TEXT & DATA MINING IN EUROPE

TDM HELPS AND ENSURES EUROPE'S COMPETITIVENESS AND FUTURE PROSPERITY

WHAT IS TDM?

SIMPLE DEFINITION OF TDM

TDM (TEXT & DATA MINING) IS A TECHNOLOGY THAT USES DATA FROM LAWFULLY OBTAINED MATERIAL, ANALYSING IT TO FIND NEW PATTERNS AND ACHIEVE NEW INSIGHTS IN LINE WITH EU DATA PROTECTION RULES.

TDM can be used by all: students, public institutions, small or large businesses, to do speculative, innovative things. It can be used by doctors to find new cures. It can be used by librarians and teachers to educate.

Researchers and businesses should be able to use information they already have legal access to. TDM will be harmed by changes to EU copyright rules that would chill new research by not allowing all users to learn from materials they have already paid to read.

TDM-led ideas and solutions are key for Europe's innovative companies, particularly SMEs: if TDM becomes too costly, this would put European innovators at a massive disadvantage compared to North America and Asia.

TDM IS CRITICAL FOR THE FUTURE

TDM USES ANOMALY DETECTION, ASSOCIATION LEARNING, CLUSTER DETECTION, CLASSIFICATION AND REGRESSION.[1] MICROSOFT'S EXCEL AND POWER BI INCLUDE TECHNOLOGY AND TOOLS THAT ENABLE RESEARCHERS AND BUSINESSES TO INNOVATE AND MAKE NEW DISCOVERIES THROUGH TDM:

TDM can help preserve Europe's vast cultural heritage data and support instant language translators.[2]
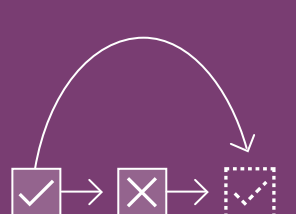
TDM is part of machine learning. It helps tackle fraud,[3] control driver-less cars[4] and reduce unnecessary readmissions to hospitals.[5]

TDM supports educators. It can be used to predict learning disabilities in school-age children.[6]

TDM has already been used to discover how existing drugs can be used to treat other conditions.

TDM can be used to anticipate Adverse Drug Reaction[7] and thus save millions of lives, and billions of euros in health-care costs.

TDM FACTS & FIGURES

WE NEED TDM BECAUSE EVERY YEAR MORE AND MORE INFORMATION IS PRODUCED AND ONLY MACHINES ARE QUICK ENOUGH TO HELP US LEARN FROM IT ALL.

The internet has indexed about **40 BILLION WEBPAGES.**[8]

There are around 28,000 active scholarly peer-reviewed journals, collectively publishing about **1.8–1.9 MILLION ARTICLES A YEAR.**[10]

YEAR 3256

JAN	FEB	MAR	APR
✓			AUG
			DEC

It would take you over **1240 YEARS** just to count the pages indexed on the internet.[9]
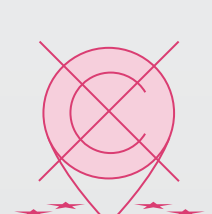
More than **70,000 PAPERS** have been published on a single protein, a tumour suppressor called p53.[11]

COPYRIGHT EXCEPTIONS ARE IMPORTANT FOR EU RESEARCHERS AND BUSINESSES ALIKE

1. **With reduced restrictions to TDM,** researchers and businesses will be able to browse scientific content from a wider pool of sources.

2. **A copyright exception** would allow them to use text and data mining methods to accelerate their research and make new discoveries by analysing large amounts of data with breakthrough technologies.[12]

3. **Universities and SMEs won't be charged additional fees to simply learn from the analysis of scientific publications.** By saving money, they will also become more competitive globally and increase their chances of making breakthrough discoveries.

WHAT NEEDS TO HAPPEN?

The EU should exempt TDM from European copyright law for **all users,** for both **commercial and non-commercial** scientific purposes. It should ensure authors aren't short-changed but that old copyright rules aren't allowed to present a barrier to modern research techniques and the pursuit of knowledge.

(Paraphrased from Hargreaves conclusion, pg 68)[13]

HELP US KEEP EUROPE COMPETITIVE IN THE PRODUCTIVE USE OF TDM

Microsoft supports a legislative environment that enables everyone to draw value from TDM and fosters creativity and innovation in Europe.

Tweet your support! Follow @MicrosoftEU | **https://blogs.microsoft.com/eupolicy/**

Microsoft

Footnotes

1. http://www.theatlantic.com/technology/archive/2012/04/everything-you-wanted-to-know-about-data-mining-but-were-afraid-to-ask/255588/
2. https://www.youtube.com/watch?v=eu9kMfe5DwQ
3. https://en.wikipedia.org/wiki/Data_analysis_techniques_for_fraud_detection#Machine_learning_and_data_mining
4. http://www.theatlantic.com/technology/archive/2014/05/all-the-world-s-music-the-trick-that-makes-googles-self-driving-cars-work/370871/
5. https://www.youtube.com/watch?v=bV8FHACTx5k
6. https://arxiv.org/ftp/arxiv/papers/1011/1011.0628.pdf
7. https://www.researchgate.net/publication/268152402_Text_and_Data_Mining_Techniques_in_Adverse_Drug_Reaction_Detection
8. http://www.worldwidewebsize.com/
9. http://www.infoplease.com/askeds/counting-billion.html
10. http://www.stm-assoc.org/2012_12_11_STM_Report_2012.pdf
11. https://www.cureforward.com/stories/gene-stories/tp53/
12. http://liberseurope.eu/wp-content/uploads/2014/11/Liber-TDM-Factsheet-v2.pdf
13. http://ec.europa.eu/research/innovation-union/pdf/TDM-report_from_the_expert_group-042014.pdf