

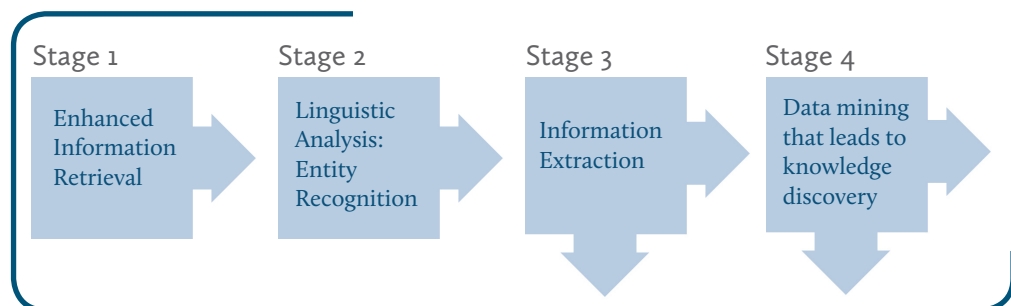
# Text and Data Mining: *The need for change in Europe.*

## 1 What is text and data mining?

*“Text and data mining (TDM) is the process of deriving information from machine-read material. It works by copying large quantities of material, extracting the data, and recombining it to identify patterns.”<sup>1</sup>*

The TDM process has four main stages. Potentially relevant documents are first identified and then turned into a machine-readable format so that structured data can be extracted. This data is subsequently mined to discover new knowledge, test hypotheses and identify new relationships.

It is essential to understand that TDM is concerned with the extraction of facts and data from vast quantities of material for research purposes, not with substituting the original material.



*The components of text and data mining.<sup>2</sup>*

## 2 Why is it important?

TDM will increase the progress of science exponentially. It has the potential to facilitate the discovery of cures for diseases such as cancer and Parkinson's. It has already been used to discover new applications for existing drugs and will act as a foundation for innovation and new industry. For libraries, it means that the researchers we support will be able to fully realise the value of our growing collections of scientific content. This will, in turn, ensure a more rigorous approach to research, including more through reviews of the literature.

## 3 Why Do Libraries Want A Copyright Exception?

Research libraries are at the centre of the data deluge. They are investing heavily in preserving born-digital content, digitising cultural heritage and in facilitating data sharing. Libraries are campaigning for freedom to apply TDM to this data because they know that doing so will maximise the return on investment of public money. A copyright exception for TDM will enable libraries and their users to contribute greatly to an innovative and competitive Europe.

This vision stands in stark contrast to the current situation. At the moment, researchers are hugely discouraged by a lack of clarity around copyright and ownership of derived works, as well as the complexity of negotiating licences. A copyright exception would allow them to use text

1. Text Mining and Data Analytics in Call for Evidence Responses. UK Government <http://webarchive.nationalarchives.gov.uk/20140603093549/http://www.ipo.gov.uk/ipreview-doc-t.pdf>

2. Value and Benefits of Text Mining. Dr Diane McDonald Joint Information Systems Committee. 2012. <http://www.jisc.ac.uk/publications/reports/2012/value-and-benefits-of-text-mining.aspx>



and data mining methods to speed the pace of their research and to make new discoveries by analysing massive amounts of data with increasingly faster and more powerful technologies.

## 4 How can you help?

If you believe that libraries and their users should be empowered to contribute to an innovative and competitive Europe, please support and share our efforts to advocate for a more flexible copyright system. We publicise our latest campaigns on the LIBER website ([www.libereurope.eu](http://www.libereurope.eu)) and on our social media channels (Twitter: [@libereurope](https://twitter.com/libereurope)).

You can also share the following Key Messages about Text and Data Mining with your national policy makers. Help them to understand why change is essential.

### Key Messages

**1. TDM will increase the progress of science exponentially.** It has the potential to facilitate the discovery of cures for diseases such as cancer and Parkinson's. It has already been used to discover how existing drugs can be used to treat other conditions. It will act as a foundation for innovation and new industry.

**2. A lack of clarity around the legality of TDM is inhibiting TDM-based research in Europe.** This in turn is impacting on the competitiveness of European research. There are far more TDM friendly copyright regimes in operation the US, Asia, Canada and the UK. A number of European based research projects have already outsourced their content mining to the US.

**3. The solutions offered by publishers are insufficient to meet the needs of researchers and are placing European researchers at a disadvantage.** One-to-one licences negotiated directly with publishers, or click-through licences such as those offered by Elsevier or through Crossref, are not scalable or interoperable and can negatively affect the dissemination of results and the availability of TDM tools. Moreover, these limited solutions do not address the legal barrier for researchers who wish to mine the open web. LIBER is calling on Elsevier to withdraw its TDM policy.<sup>3</sup>

**4. The introduction of a mandatory copyright exception to allow anyone to use computers to analyse anything to which they have legal access is essential.** It will support both science and innovation and will ultimately benefit the European economy. Such an exception should allow for the distribution of the results of TDM, as long as they do not substitute the original works.

**5. We are seeking an exception for TDM that will allow those with legal access to the content to extract facts and data using automated methods such as crawling and scraping.** We are not seeking to gain free access to copyrighted works.

**6. TDM does not put unnecessary strain on publishers' infrastructure.** Open Access publishers have confirmed this to be true. Any potential future strain can be mitigated by publishers making their crawl delay times public and by working with libraries and researchers to promote responsible content mining.

**7. TDM is concerned only with the extraction of non-copyrightable objects (facts and data). It therefore makes no sense when drafting a law to address the technical shortcomings of the current copyright framework to limit that solution to non-commercial uses.** Also, the definition of 'non-commercial' is impossible to clarify and creates a legally uncertain situation for European technology companies working in big data and data analytics markets. Limiting an exception to non-commercial work is likely to have a chilling effect on the development and availability of tools for researchers who wish to use TDM.

<sup>3</sup>. LIBER Calls on Elsevier to Withdraw TDM Policy. 2014. <http://libereurope.eu/advocacy-campaigns/liber-calls-on-elsevier-to-withdraw-tdm-policy>