The right to read is the right to mine: library resources for cross-disciplinary work

Sarah Bull, University of Birmingham

Neil Smyth, University of Nottingham

Launching the Corpus Statistics Group: studying large collections of electronic texts
University of Birmingham, Thursday 11 February 2016

Data-Asset-Method

Data – Asset – Method: Harnessing the Infinite Archive is an international research network led by the University of Nottingham.

Arts & Humanities Research Council

horizon
DIGITAL ECONOMY RESEARCH

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

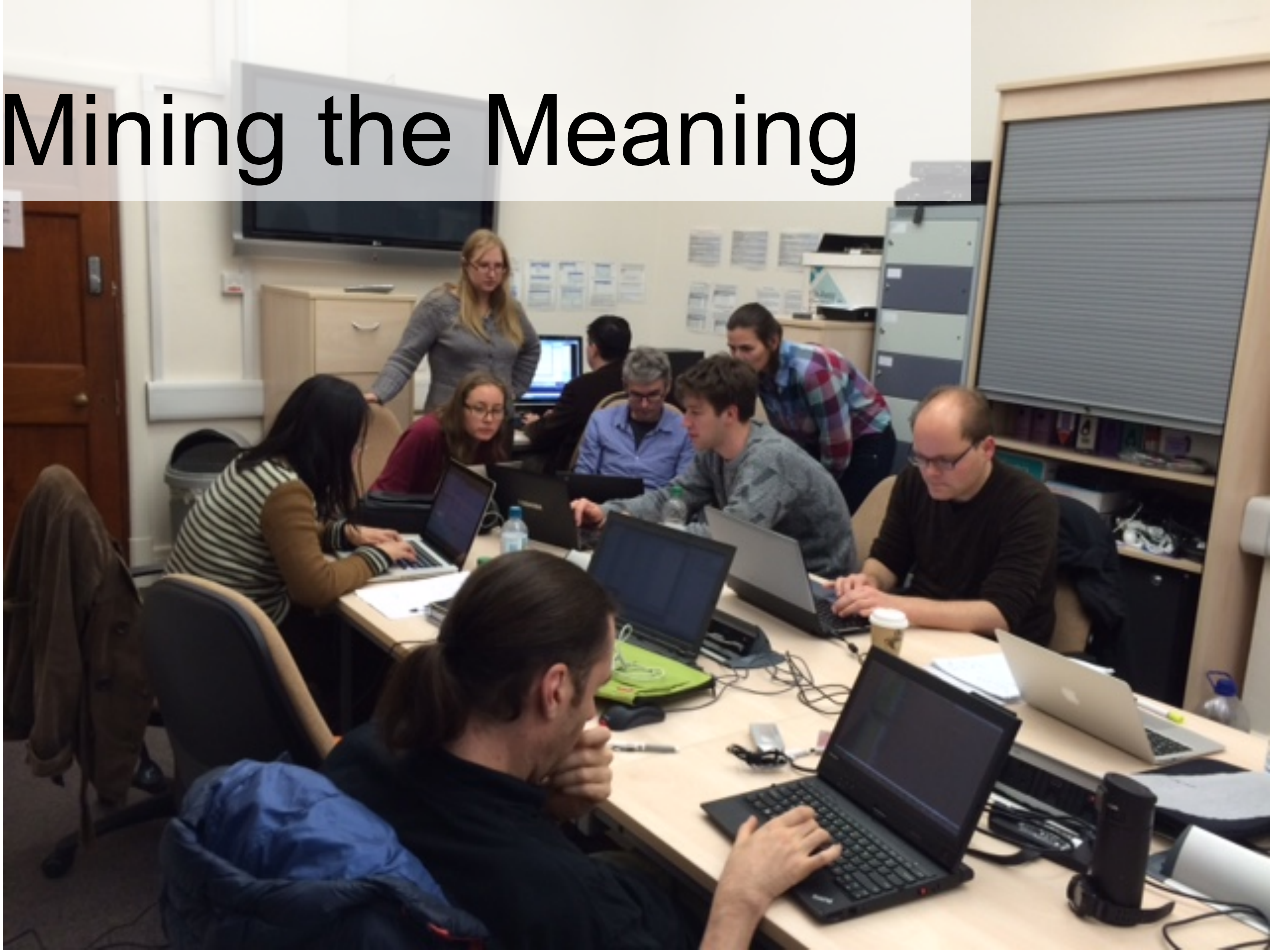
Corpus Protocols

Seth Cayley: Head of Research Solutions, Cengage Learning
Mike Gardner: Web Developer in Web Technologies, UoN
Kat Gupta: Corpus Protocols Researcher, UoN
Michaela Mahlberg: Professor , UoN
Neil Smyth: Librarian, UoN
Stella Wisdom: Digital Curator at the British Library

Arts & Humanities Research Council
horizon DIGITAL ECONOMY RESEARCH
The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

TIDAL (TImes Data Archive Lab)

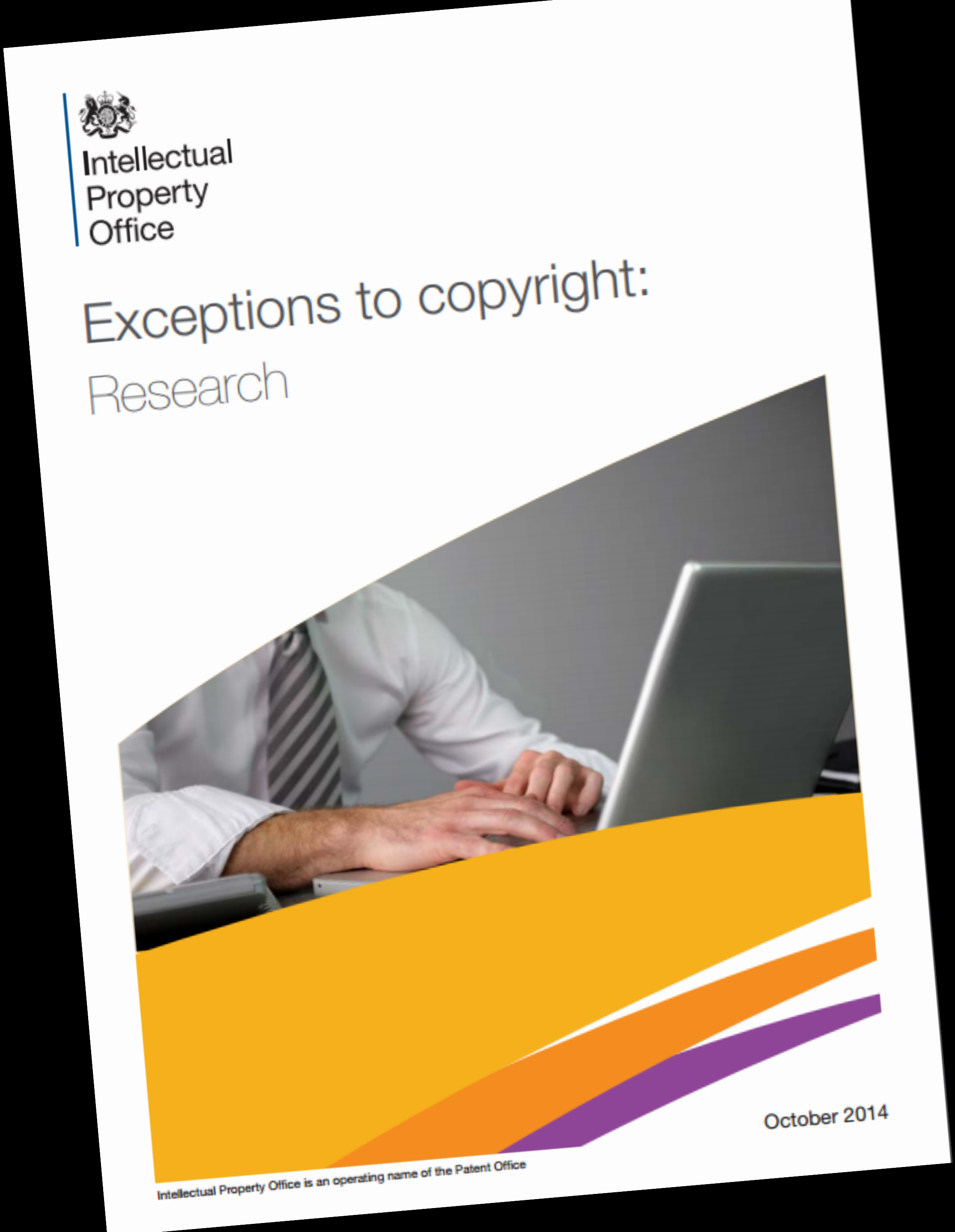
Mining the Meaning



UK Copyright legislation: 2014

“The making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that — (a) the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose, and (b) the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise)”

(The Copyright and Rights in Performances (Research, Education, Libraries and Archives) Regulations 2014.



Intellectual Property Office

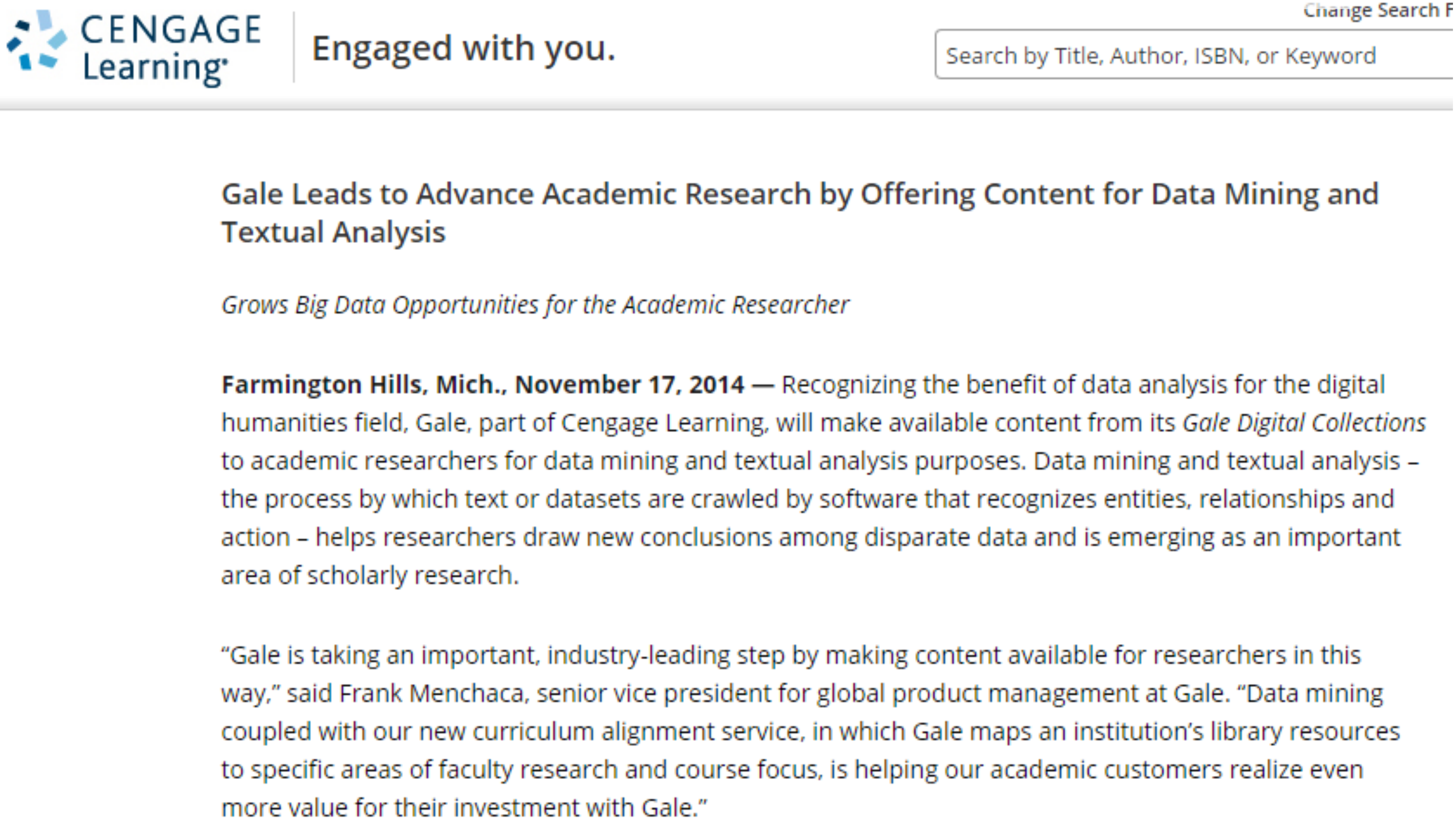
Exceptions to copyright:

Research

October 2014

Intellectual Property Office is an operating name of the Patent Office

Cengage Learning - 2014

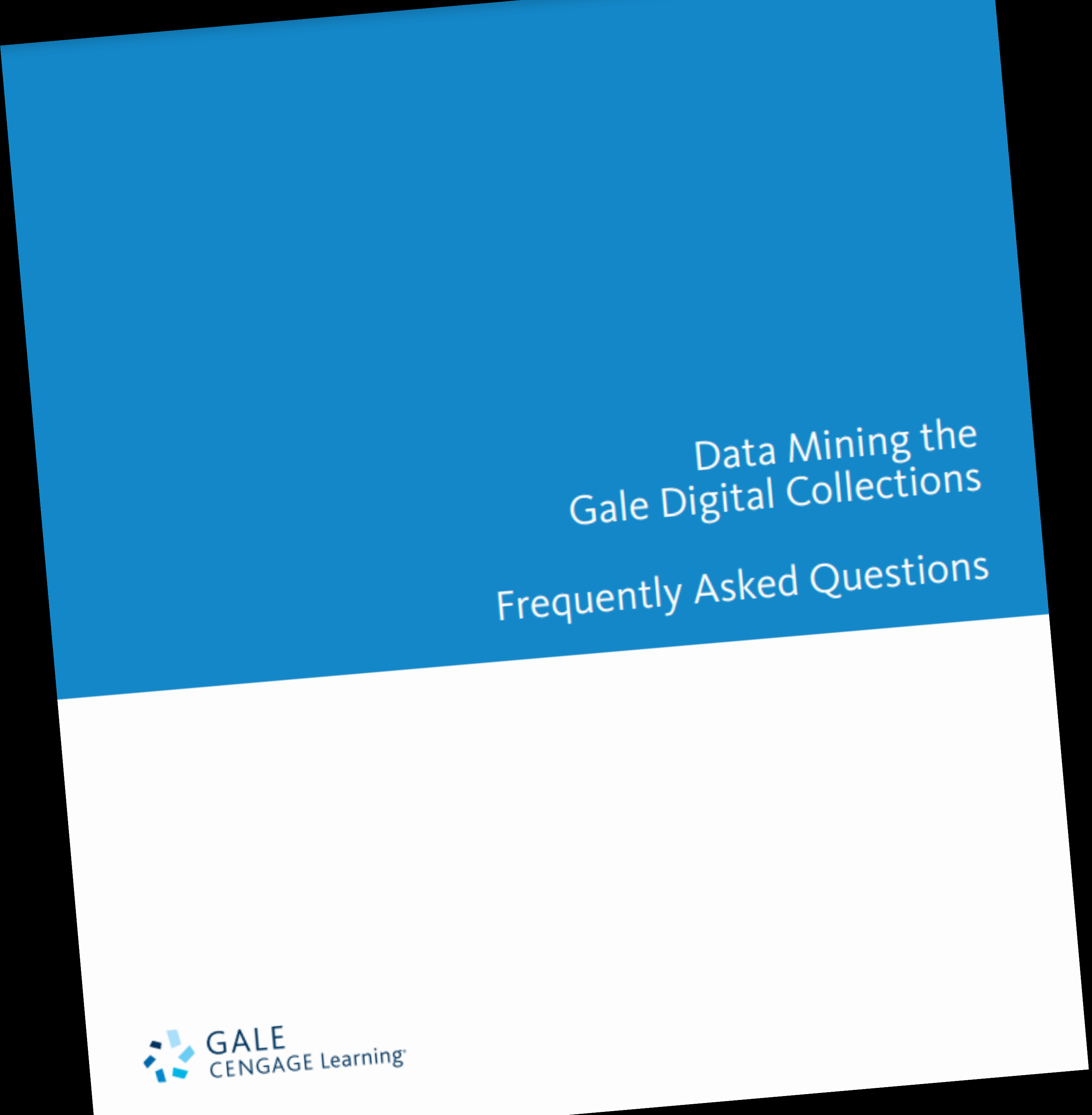


Gale Leads to Advance Academic Research by Offering Content for Data Mining and Textual Analysis

Grows Big Data Opportunities for the Academic Researcher

Farmington Hills, Mich., November 17, 2014 — Recognizing the benefit of data analysis for the digital humanities field, Gale, part of Cengage Learning, will make available content from its *Gale Digital Collections* to academic researchers for data mining and textual analysis purposes. Data mining and textual analysis – the process by which text or datasets are crawled by software that recognizes entities, relationships and action – helps researchers draw new conclusions among disparate data and is emerging as an important area of scholarly research.

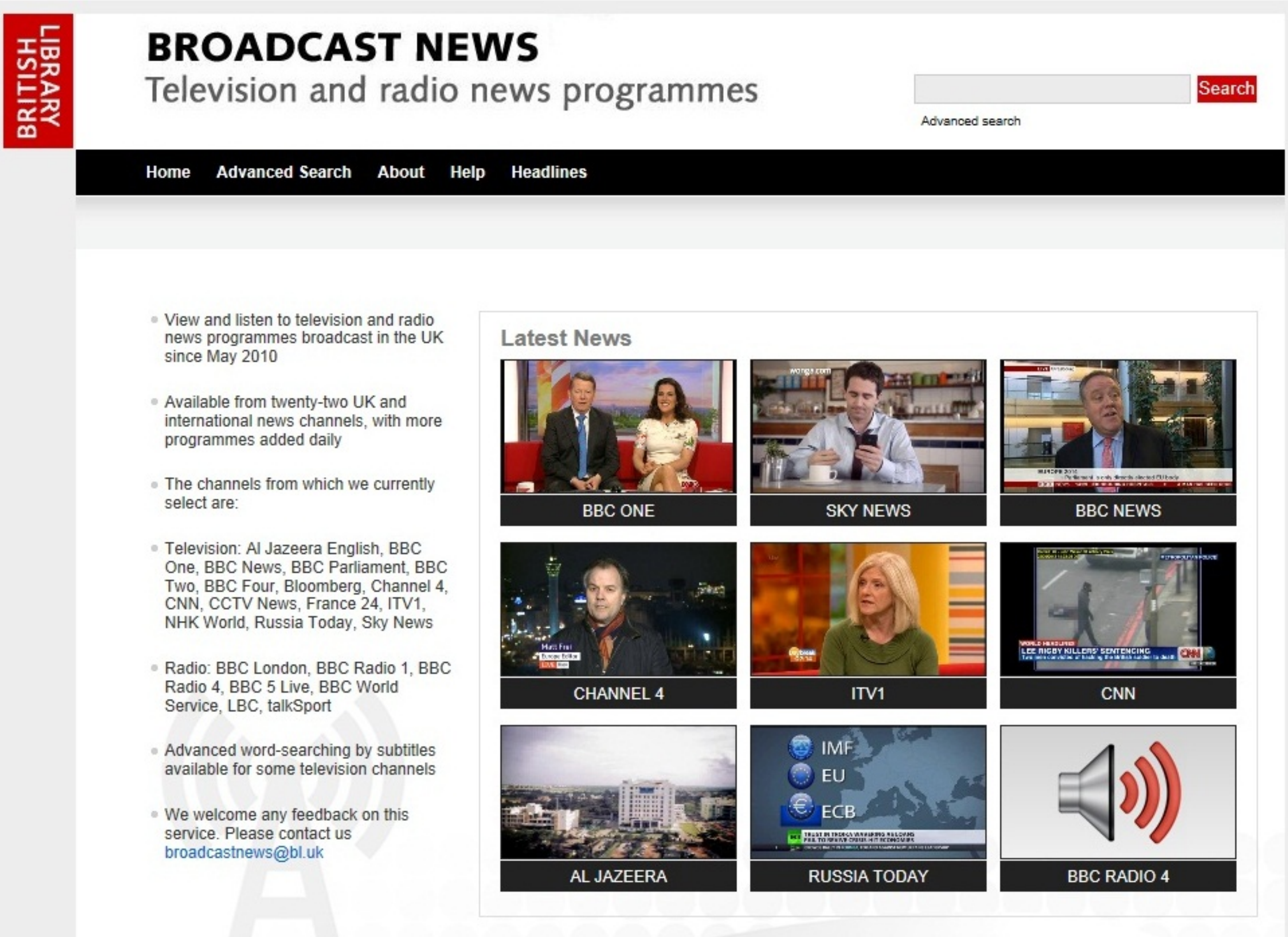
"Gale is taking an important, industry-leading step by making content available for researchers in this way," said Frank Menchaca, senior vice president for global product management at Gale. "Data mining coupled with our new curriculum alignment service, in which Gale maps an institution's library resources to specific areas of faculty research and course focus, is helping our academic customers realize even more value for their investment with Gale."



Data Mining the
Gale Digital Collections

Frequently Asked Questions

GALE
CENGAGE Learning

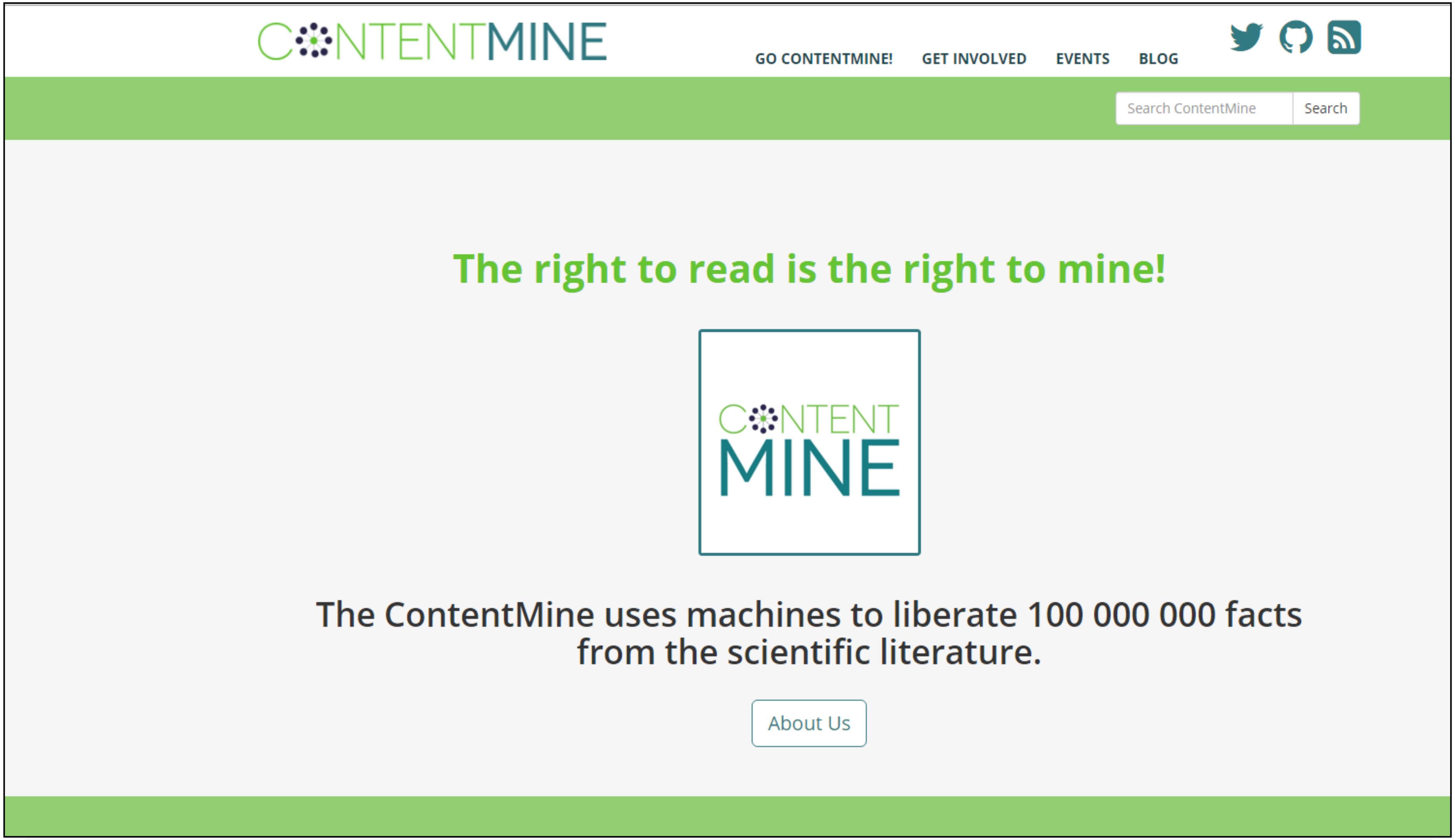


BROADCAST NEWS

Television and radio news programmes

- View and listen to television and radio news programmes broadcast in the UK since May 2010
- Available from twenty-two UK and international news channels, with more programmes added daily
- The channels from which we currently select are:
- Television: Al Jazeera English, BBC One, BBC News, BBC Parliament, BBC Two, BBC Four, Bloomberg, Channel 4, CNN, CCTV News, France 24, ITV1, NHK World, Russia Today, Sky News
- Radio: BBC London, BBC Radio 1, BBC Radio 4, BBC 5 Live, BBC World Service, LBC, talkSport
- Advanced word-searching by subtitles available for some television channels
- We welcome any feedback on this service. Please contact us broadcastnews@bl.uk

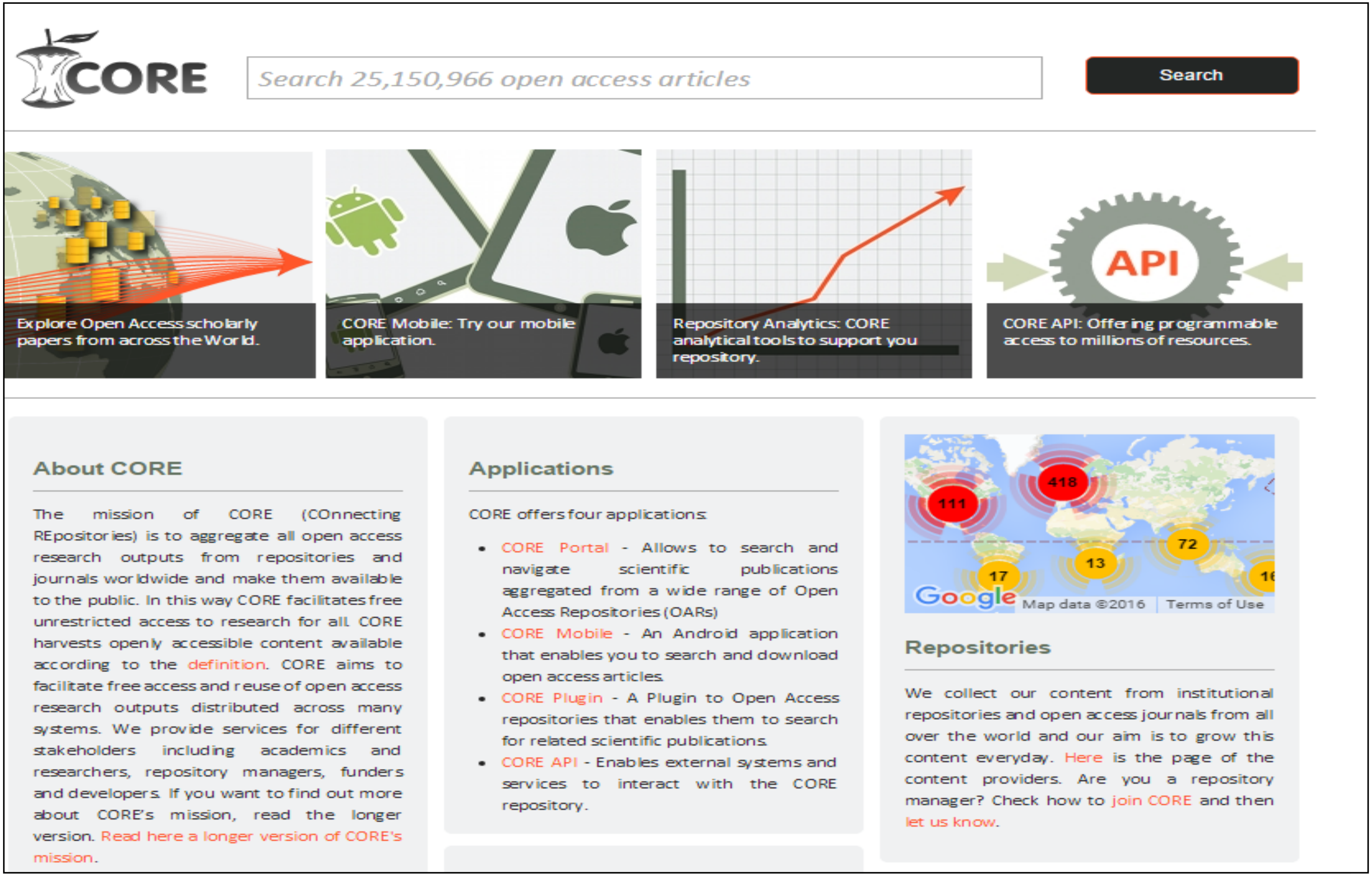
Latest News



The right to read is the right to mine!

The ContentMine uses machines to liberate 100 000 000 facts from the scientific literature.

http://contentmine.org/



Explore Open Access scholarly papers from across the World.

CORE Mobile: Try our mobile application.

Repository Analytics: CORE analytical tools to support you repository.

CORE API: Offering programmable access to millions of resources.

About CORE

The mission of CORE (COnnecting REpositories) is to aggregate all open access research outputs from repositories and journals worldwide and make them available to the public. In this way CORE facilitates free unrestricted access to research for all. CORE harvests openly accessible content available according to the definition. CORE aims to facilitate free access and reuse of open access research outputs distributed across many systems. We provide services for different stakeholders including academics and researchers, repository managers, funders and developers. If you want to find out more about CORE's mission, read the longer version. Read here a longer version of CORE's mission.

Applications

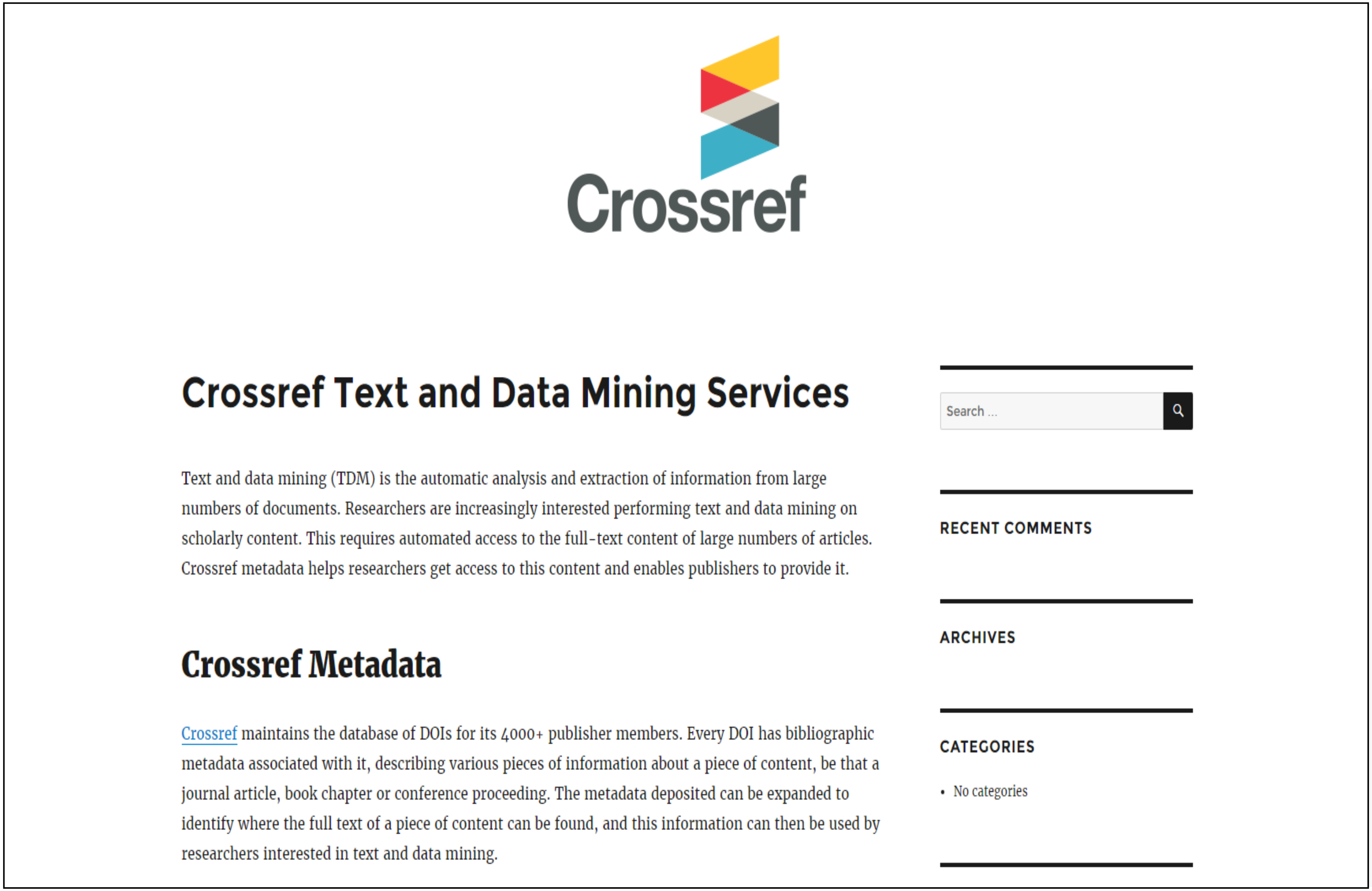
CORE offers four applications:

- CORE Portal - Allows to search and navigate scientific publications aggregated from a wide range of Open Access Repositories (OARs)
- CORE Mobile - An Android application that enables you to search and download open access articles.
- CORE Plugin - A Plugin to Open Access repositories that enables them to search for related scientific publications.
- CORE API - Enables external systems and services to interact with the CORE repository.

Repositories

We collect our content from institutional repositories and open access journals from all over the world and our aim is to grow this content everyday. Here is the page of the content providers. Are you a repository manager? Check how to join CORE and then let us know.

https://core.ac.uk/



Crossref Text and Data Mining Services

Text and data mining (TDM) is the automatic analysis and extraction of information from large numbers of documents. Researchers are increasingly interested performing text and data mining on scholarly content. This requires automated access to the full-text content of large numbers of articles. Crossref metadata helps researchers get access to this content and enables publishers to provide it.

Crossref Metadata

Crossref maintains the database of DOIs for its 4000+ publisher members. Every DOI has bibliographic metadata associated with it, describing various pieces of information about a piece of content, be that a journal article, book chapter or conference proceeding. The metadata deposited can be expanded to identify where the full text of a piece of content can be found, and this information can then be used by researchers interested in text and data mining.

RECENT COMMENTS

ARCHIVES

CATEGORIES

- No categories

http://tdmsupport.crossref.org/

Copyright Exceptions - Constraints

- Only applicable in the UK
- Researcher must have lawful access
- Must be for non-commercial purposes
- Work within publisher systems for TDM (technical protection measures)
- Collaborative research projects
- Limited quotation under ‘fair dealing’
- Sufficient acknowledgement

Technological Protection Measures

“It is important to be aware that media, such as DVDs and e-books, are often protected by Technological Protection Measures (TPMs) (also known as copy protection measures or DRM) which prevent unauthorised access or copying.

TPMs can play an important role in enabling copyright owners to offer content to consumers in different ways, as well as preventing piracy. EU and UK law protects the right of copyright owners to use TPMs to protect their works, and circumvention of such technology is illegal”

https://www.gov.uk/exceptions-to-copyright

Publisher licences

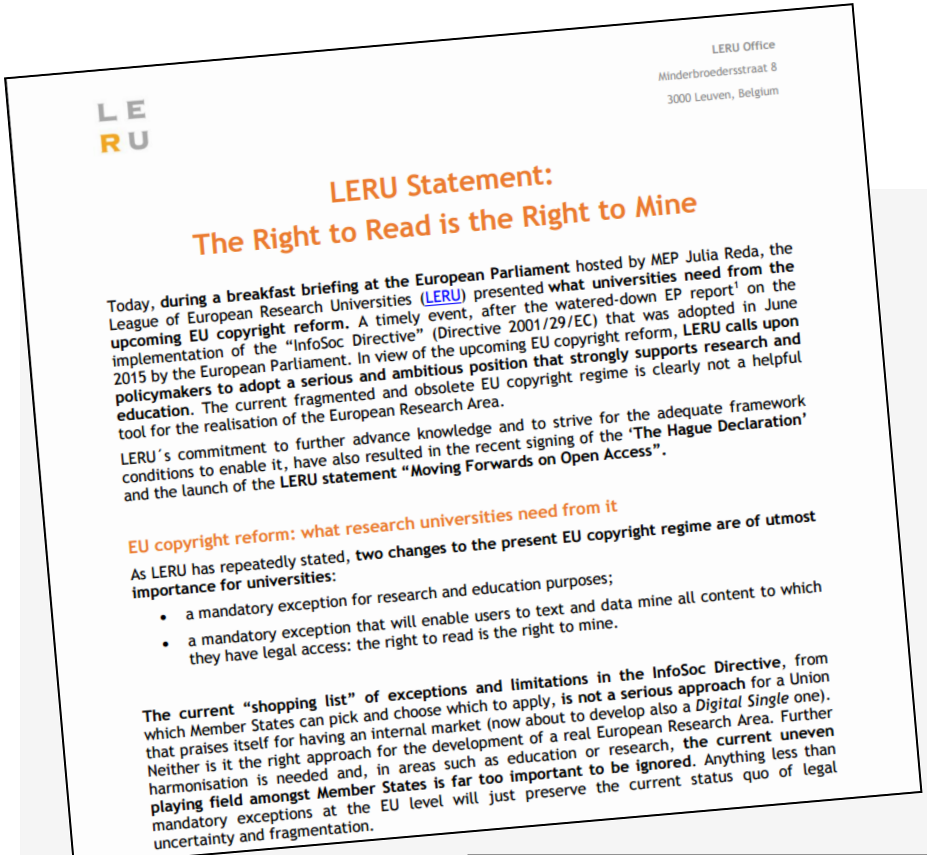
- Library signs a web licence for end user access on behalf of University
- For TDM library also has role in ensuring security of licensed content
- Use of the term “snippets” and restrictions on content quotation in research output

Community clarification ongoing based on real world projects

Infrastructure considerations

- Publisher readiness – capacity, service
- Institutional capacity – network, security
- Hard drive or API – no ‘one size fits all’

Ensuring security of content balanced with enabling innovative research methods



LERU Office
Minderbroedersstraat 8
3000 Leuven, Belgium

LE
RU

LERU Statement: The Right to Read is the Right to Mine

Today, **during a breakfast briefing at the European Parliament** hosted by MEP Julia Reda, the League of European Research Universities (LERU) presented **what universities need from the upcoming EU copyright reform.** A timely event, after the watered-down EP report[1] on the implementation of the "InfoSoc Directive" (Directive 2001/29/EC) that was adopted in June 2015 by the European Parliament. In view of the upcoming EU copyright reform, **LERU calls upon policymakers to adopt a serious and ambitious position that strongly supports research and education**. The current fragmented and obsolete EU copyright regime is clearly not a helpful tool for the realisation of the European Research Area.

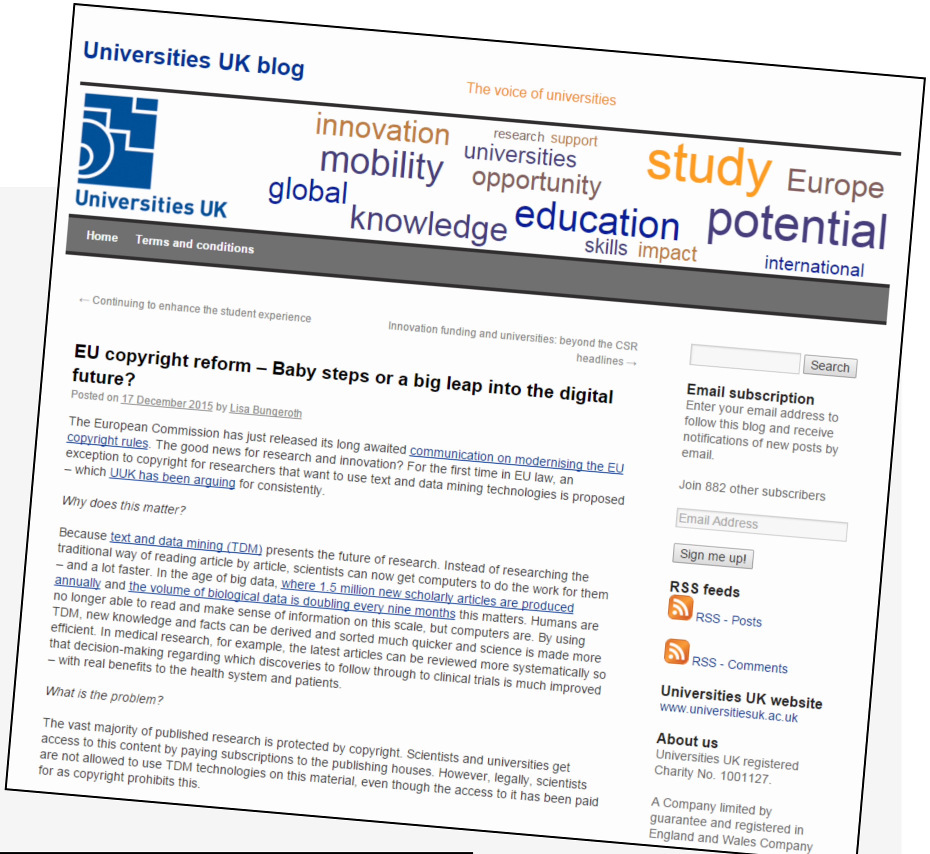
LERU´s commitment to further advance knowledge and to strive for the adequate framework conditions to enable it, have also resulted in the recent signing of the **'The Hague Declaration'** and the launch of the **LERU statement "Moving Forwards on Open Access"**.

EU copyright reform: what research universities need from it

As LERU has repeatedly stated, **two changes to the present EU copyright regime are of utmost importance for universities**:

- a mandatory exception for research and education purposes;
- a mandatory exception that will enable users to text and data mine all content to which they have legal access: the right to read is the right to mine.

The current "shopping list" of exceptions and limitations in the InfoSoc Directive, from which Member States can pick and choose which to apply, **is not a serious approach** for a Union that praises itself for having an internal market (now about to develop also a *Digital Single* one). Neither is it the right approach for the development of a real European Research Area. Further harmonisation is needed and, in areas such as education or research, **the current uneven playing field amongst Member States is far too important to be ignored**. Anything less than mandatory exceptions at the EU level will just preserve the current status quo of legal uncertainty and fragmentation.



← Continuing to enhance the student experience

Innovation funding and universities: beyond the CSR headlines →

EU copyright reform – Baby steps or a big leap into the digital future?

Posted on 17 December 2015 by Lisa Bungeroth

The European Commission has just released its long awaited communication on modernising the EU copyright rules. The good news for research and innovation? For the first time in EU law, an exception to copyright for researchers that want to use text and data mining technologies is proposed – which UUK has been arguing for consistently.

Why does this matter?

Because text and data mining (TDM) presents the future of research. Instead of researching the traditional way of reading article by article, scientists can now get computers to do the work for them – and a lot faster. In the age of big data, where 1.5 million new scholarly articles are produced annually and the volume of biological data is doubling every nine months this matters. Humans are no longer able to read and make sense of information on this scale, but computers are. By using TDM, new knowledge and facts can be derived and sorted much quicker and science is made more efficient. In medical research, for example, the latest articles can be reviewed more systematically so that decision-making regarding which discoveries to follow through to clinical trials is much improved – with real benefits to the health system and patients.

What is the problem?

The vast majority of published research is protected by copyright. Scientists and universities get access to this content by paying subscriptions to the publishing houses. However, legally, scientists are not allowed to use TDM technologies on this material, even though the access to it has been paid for as copyright prohibits this.

Email subscription
Enter your email address to follow this blog and receive notifications of new posts by email.

Join 882 other subscribers

Email Address

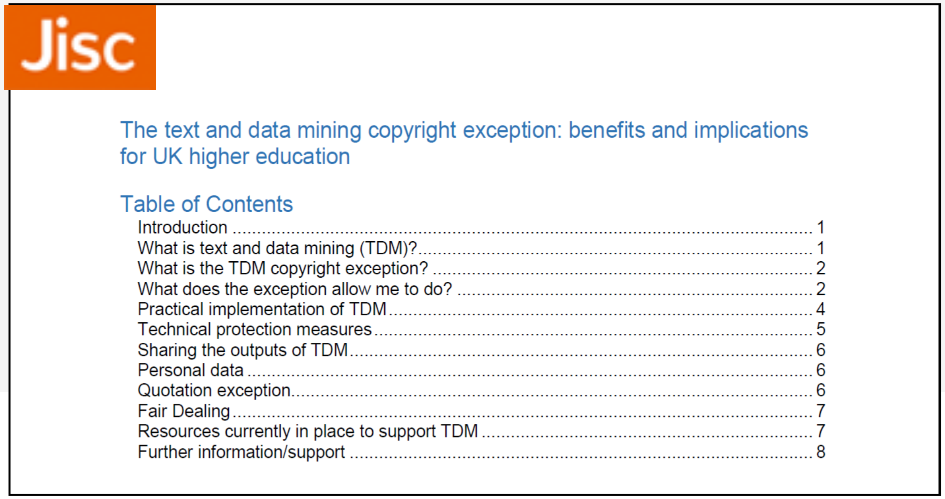
RSS feeds
RSS - Posts

RSS - Comments

Universities UK website
www.universitiesuk.ac.uk

About us
Universities UK registered Charity No. 1001127.

A Company limited by guarantee and registered in England and Wales Company



Jisc

The text and data mining copyright exception: benefits and implications for UK higher education

Table of Contents

Further advice:

Library Services

copyright@contacts.bham.ac.uk
copyright@nottingham.ac.uk

The right to read is the right to mine: library resources for cross-disciplinary work

Sarah Bull, University of Birmingham

Neil Smyth, University of Nottingham

Launching the Corpus Statistics Group: studying large collections of electronic texts
University of Birmingham, Thursday 11 February 2016