The Hague
DECLARATION

The Hague Declaration on **KNOWLEDGE DISCOVERY** *in the Digital Age*

Vision

New technologies are revolutionising the way humans can learn about the world and about themselves. These technologies are not only a means of dealing with Big Data[1], they are also a key to knowledge discovery in the digital age; and their power is predicated on the increasing availability of data itself. Factors such as increasing computing power, the growth of the web, and governmental commitment to open access[2] to publicly-funded research are serving to increase the availability of facts, data and ideas.

However, current legislative frameworks in different legal jurisdictions may not be cast in a way which supports the introduction of new approaches to undertaking research, in particular content mining. Content mining is the process of deriving information from machine-readable material. It works by copying large quantities of material, extracting the data, and recombining it to identify patterns and trends.

At the same time, intellectual property laws from a time well before the advent of the web limit the power of digital content analysis techniques such as text and data mining (for text and data) or content mining (for computer analysis of content in all formats)[3]. These factors are also creating inequalities in access to knowledge discovery in the digital age. The legislation in question might be copyright law, law governing patents or database laws – all of which may restrict the ability of the user to perform detailed content analysis.

Researchers should have the freedom to analyse and pursue intellectual curiosity without fear of monitoring or repercussions. These freedoms must not be eroded in the digital environment. Likewise, ethics around the use of data and content mining continue to evolve in response to changing technology.

Computer analysis of content in all formats, that is content mining, enables access to undiscovered public knowledge and provides important insights across every aspect of our economic, social and cultural life. Content mining will also have a profound impact for understanding society and societal movements (for example, predicting political uprisings, analysing demographical changes). Use of such techniques has the potential to revolutionise the way research is performed – both academic and commercial.

[1] *For the G8 Open Data Charter, see https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex*
See also the RDA Data Harvest Report at http://europe.rd-alliance.org/sites/default/files/report/TheDataHarvestReport_%20Final.pdf
[2] *For the Berlin Declaration, see http://openaccess.mpg.de/Berlin-Declaration*
see also the Budapest Open Access Initiative 10 years on at http://www.budapestopenaccessinitiative.org/boai-10-recommendations
[3] *Where the term text and data mining / TDM is used in the Declaration, it is used to mean the mining of all forms of data irrespective of whether in text, images, sound recordings or film.*

The Hague DECLARATION
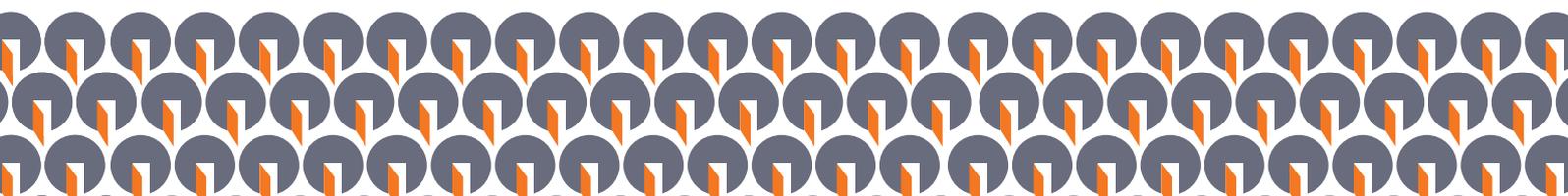
Benefits of Content Mining

The potential benefits of content mining are vast and include:
- **Addressing grand challenges such as climate change and global epidemics**
- **Improving population health, wealth and development**
- **Creating new jobs and employment**
- **Exponentially increasing the speed and progress of science through new insights and greater efficiency of research**
- **Increasing transparency of governments and their actions**
- **Fostering innovation and collaboration and boosting the impact of open science**
- **Creating tools for education and research**
- **Providing new and richer cultural insights**
- **Speeding economic and social development in all parts of the globe**

Researchers, SMEs (Small and Medium Sized Enterprises) and big technological companies have been content mining for at least 10 years, but the potential for extracting significant benefits from this work has been limited due to ongoing legal uncertainties and restrictions. However, in this age of opportunity, it is important that all members of society can benefit equally from advances in the availability of digital technology and content. This requires the creation of new principles around access to facts, data and ideas.

Principles

Therefore we, the undersigned, in recognition of the huge potential economic and societal benefits of knowledge discovery in the digital age, endorse the following principles:

1. INTELLECTUAL PROPERTY WAS NOT DESIGNED TO REGULATE THE FREE FLOW OF FACTS, DATA AND IDEAS, BUT HAS AS A KEY OBJECTIVE THE PROMOTION OF RESEARCH ACTIVITY

The free flow of information and ideas is an essential human right.[4] It is a catalyst for the production of human knowledge, which underpins welfare and prosperity. Societies around the world have chosen to protect certain limited rights in intellectual property as incentives both to innovation and the dissemination of knowledge. Intellectual property law was never intended to cover facts, ideas and pure data. The modern application of intellectual property law is increasingly becoming an obstacle to the creation and sharing of knowledge, as it can pose constraints and limitations to free sharing even when simple building blocks such as facts and data are being used.

In some countries, copyright law[5] in particular has been interpreted to restrict the ability to apply computer reading and analysis to otherwise legally-available content. Other legislative frameworks such as patent law and database law may have a similar impact. When intellectual property law allows content to be read and analysed manually by humans but not by their machines, it has failed its original purposes.

---

[4] *http://www.un.org/en/documents/udhr/index.shtml#a19*
[5] *And in the European Union database laws also in the form of the Database Directive.*
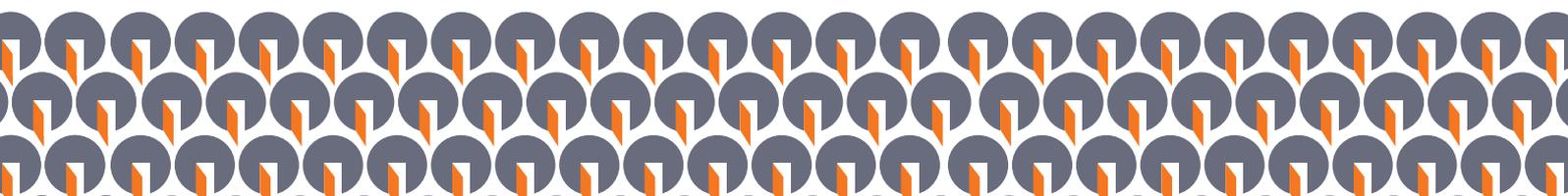
2. PEOPLE SHOULD HAVE THE FREEDOM TO ANALYSE AND PURSUE INTELLECTUAL CURIOSITY WITHOUT FEAR OF MONITORING OR REPERCUSSIONS

Providers of content should respect the intellectual privacy of individual readers and should take measures to protect readers' privacy from interference by any external body. Any exception, which for example would result in an encroachment of individual privacy, will need to be necessary and proportionate and provided for by law. The use of facts, data and ideas must not prejudice the legitimate rights of individuals to privacy and a private life.

3. LICENSES AND CONTRACT TERMS SHOULD NOT RESTRICT INDIVIDUALS FROM USING FACTS, DATA AND IDEAS

Generally, licences and contract terms that regulate and restrict how individuals may analyse and use facts, data and ideas are unacceptable and inhibit innovation and the creation of new knowledge and, therefore, should not be adopted. Similarly, it is unacceptable that technical measures in digital rights management systems should inhibit the lawful right to perform content mining.

4. ETHICS AROUND THE USE OF CONTENT MINING TECHNIQUES WILL NEED TO CONTINUE TO EVOLVE IN RESPONSE TO CHANGING TECHNOLOGY

The observation of well-established ethical norms in research and business, as well as the continued development of such standards and laws, must be supported and encouraged in order to ensure that content mining technologies are deployed for the benefit of society.

5. INNOVATION AND COMMERCIAL RESEARCH BASED ON THE USE OF FACTS, DATA, AND IDEAS SHOULD NOT BE RESTRICTED BY INTELLECTUAL PROPERTY LAW

As facts, data, and ideas are not copyrightable it does not make sense to restrict ethical commercial use of those facts, data, and ideas extracted from content which has been obtained legally. It is recognised that while patent law is designed to protect innovations and inventions, this is not meant to encompass facts and data. Restrictions on the use of facts, data and ideas can have a serious impact on innovation and on economic development globally. It can also reduce the ability to use tools and processes which can benefit citizens in the areas of health, science, employment, research, the environment and culture.

The Hague
DECLARATION
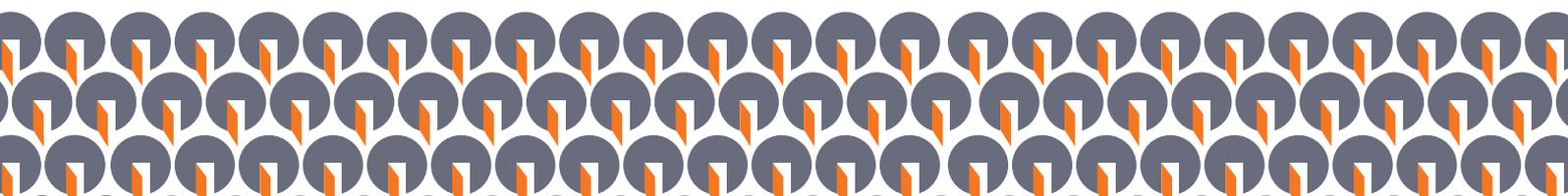
Roadmap for Action

ACCEPTING THE VISION AND PRINCIPLES

1. The vision embodied in this Declaration is that intellectual property was not designed to regulate the free flow of facts, data and ideas, but has as a key objective the promotion of research activity.
2. Where copyright frameworks do not currently support such a vision, legislators should immediately work to support the introduction of changes which would allow users to undertake content mining on materials to which they have lawful access.
3. Where Exceptions or Limitations are introduced into copyright law to allow content mining, these should be mandatory and may not be overridden by contracts.
4. It is unacceptable that technical measures in digital rights management systems should inhibit the legal right to perform content mining.
5. There should be no need to sign separate licences to undertake content mining activity, since the right to read is the right to mine where those performing mining activities already have lawful access to relevant content.

POLICY AND INCENTIVES

6. Policy makers should aim to provide legal clarity by ensuring that content mining is not an infringement of copyright or related rights. We believe that the right to read includes the right to mine, but only where individuals have lawful access to that content.
7. Where research funders or other bodies require, and where authors wish, research outputs to be made available under specific licences, these should typically be CC-BY for publications and CC0 for research data.
8. Every university, research organisation, research funder and commercial business should ensure that their policies advocate content mining as a research methodology which has the potential to transform the way research is performed. The growth of open access and open data has been, and will continue to be, a key enabler of content mining.
9. Researchers should recognise the right of the authors of publications to be acknowledged as such, and to be respected and acknowledged as data producers, wherever possible.
10. Policies on content mining should respect the legitimate rights of authors and publishers, and be driven by the needs of researchers and businesses in the digital age.
11. To encourage the uptake of content mining activity, universities, research organisations, research funders and businesses should consider introducing incentives to reward those who use these new techniques – for example, content mining activities should be noted and commended in appraisal/evaluation processes.

The Hague
DECLARATION
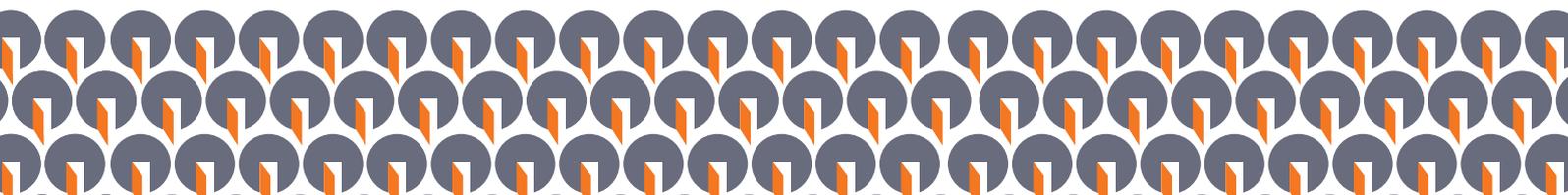
INFRASTRUCTURE AND TOOLS

12. Research organisations, universities and businesses should ensure that they maintain and develop repository infrastructures to provide storage for, and access to, publications which can legally be made available for content mining. Independent researchers should strive to make use of such facilities where they are available.

13. Research organisations, universities and businesses should provide access to suitable infrastructures to enable research data to be made available, where it is lawful and ethically possible to do so, for content mining. Independent researchers should strive to make use of such facilities where they are available.

14. Developments including technical infrastructures, standards, ethical norms and funding requirements should make research results available as open outputs.

15. Open standards such as XML and JSON for data transport, ORCID for author IDs and CC licenses for open licensing needs should be used wherever possible.

16. If material is made available under a CC BY licence, content creators should make the following available for download: the XML or other high quality file format, high resolution images, supporting data behind the images; the unformatted accepted article from the author.

ADVOCACY

17. Bodies such as universities, research organisations, library associations, the medical community, businesses, and members of the content mining community should advocate the benefits of content mining.

18. Research libraries are well placed to take on this advocacy role as part of their activities in research support.

19. Libraries should provide training for researchers on content mining literacy, including legal advice.

DEVELOPMENT AND MONITORING OF THE HAGUE DECLARATION & ROADMAP

20. LIBER (Association of European Research Libraries) has been instrumental in working with stakeholders to develop the Hague Declaration and Roadmap. LIBER will continue to monitor and oversee progress towards obtaining signatures for the Declaration, and will advocate the implementation of the Roadmap.

Learn more & sign the Declaration
www.thehaguedeclaration.com