

Copyright law and text and data mining the research community's perspective

Christoph Bruch
Helmholtz Association

80TH IFLA GENERAL CONFERENCE AND ASSEMBLY
16-22 AUGUST 2014, LYON, FRANCE



Open Science

Helmholtz Association

- More than 35,000 staff in 18 research centres of which more than 12,000 are scientists
- Helmholtz Association is Germany's largest scientific organisation
- Annual budget of approximately € 3.8 billion (30% contract funding)



Definitions

- Data** Numbers, characters or images that designate an attribute of a phenomenon.
- Information** Data become information when they are combined together in ways that have the potential to reveal patterns in the phenomenon.
- Knowledge** Information yields knowledge when it supports non-trivial, true claims about a phenomenon.
- TDM** Technique for analysing and extracting new insights and knowledge from the exponentially increasing store of digital data ('Big Data').

Hargreaves, Ian et al. (2014): Standardisation in the area of innovation and technological development, notably in the field of Text and Data Mining, Brussels, European Commission. ec.europa.eu/research/innovation-union/pdf/TDM-report_from_the_expert_group-042014.pdf

The Royal Society (VI.2012): Science as an open enterprise, Science Policy Centre report, 02/12, London, The Royal Society. royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

Lyon, 19.08.2014

Seite 3

What is it all about?

Science as an open enterprise

Summary report
June 2012

THE
ROYAL
SOCIETY



Royal Society: Intelligent (qualified) Openness

- Accessible
 - such that it can readily be found
- Assessable
 - Ability to make some judgment on reliability
- Intelligible
 - Ability to make some judgment or assessment of what is communicated
- Usable
 - Format, meta data, legal aspects

Text und Data Mining helps coping with the information deluge and ...



... makes possible new forms of knowledge discovery

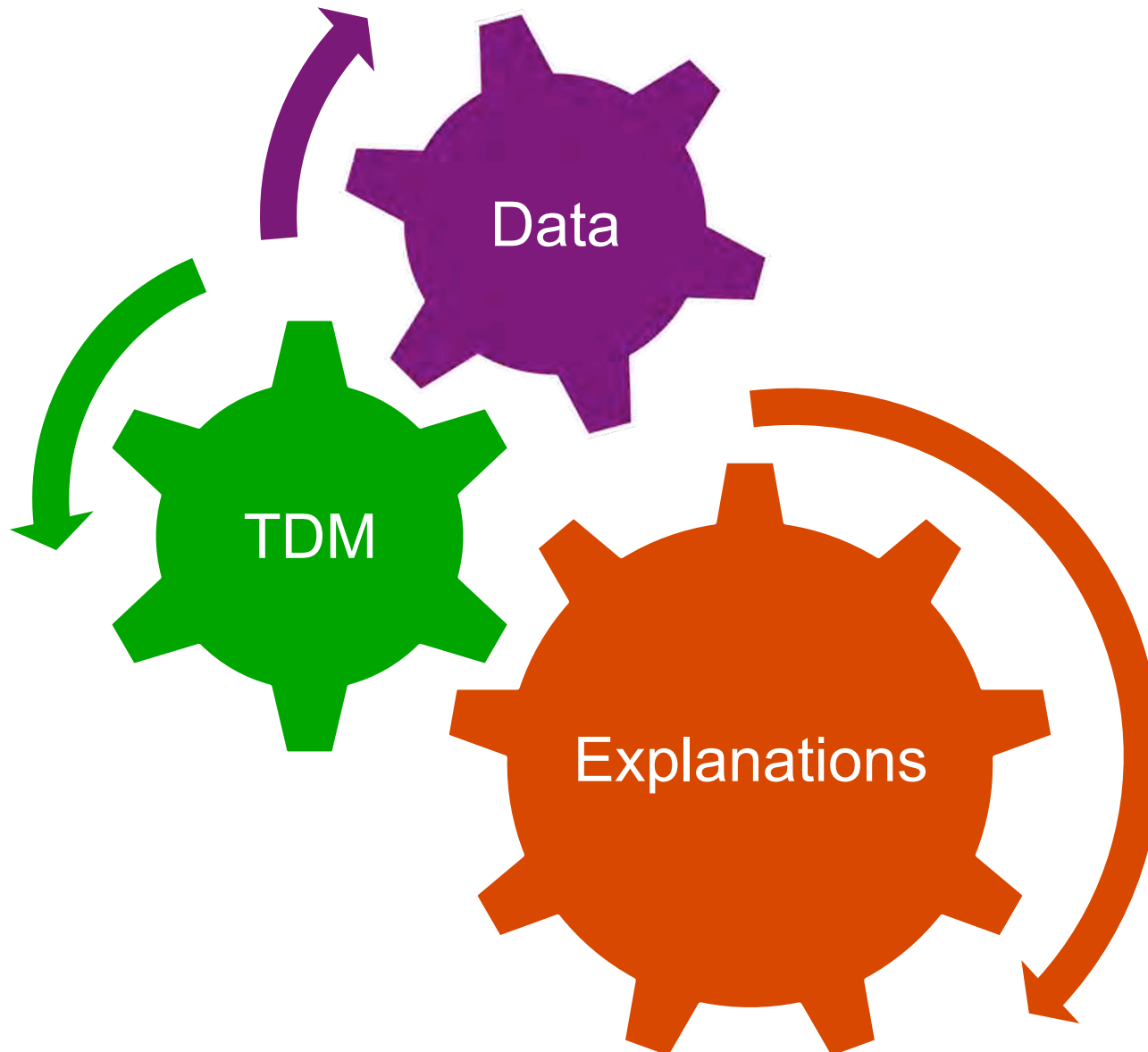


1/20/14 19:08:2014

<http://www.the-girl-who-ate-everything.com/wp-content/uploads/blogger/-eBMApY-5pjl/T9VaZdfNzml/AAAAAAAAAJAU/jhFyrtNa-U/s1600/bunless-fieta-hamburgers.JPG>



Interdependence of Access to data & publications



TDM is indispensable

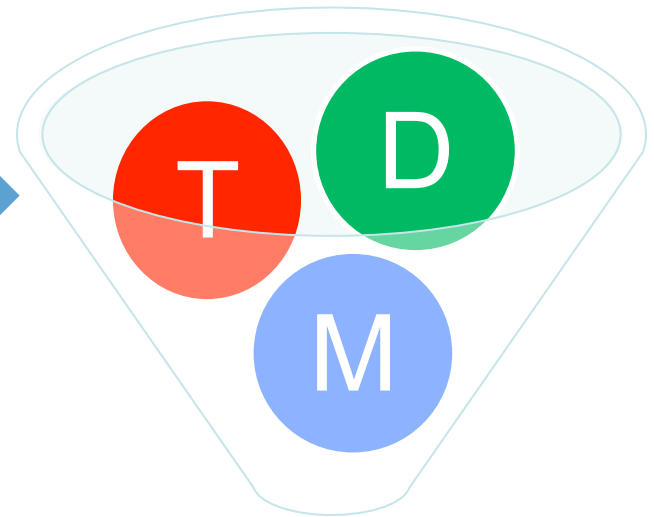
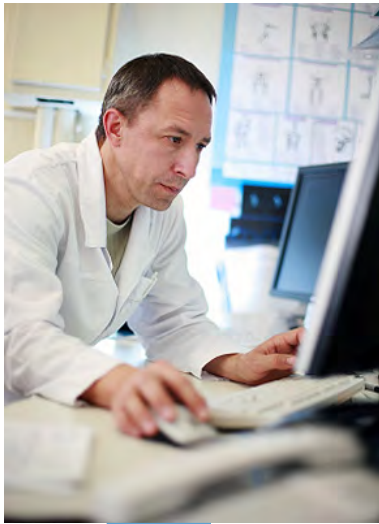
No TDM

No TDM

No TDM

Refrain from TDM would leave researchers half blind.

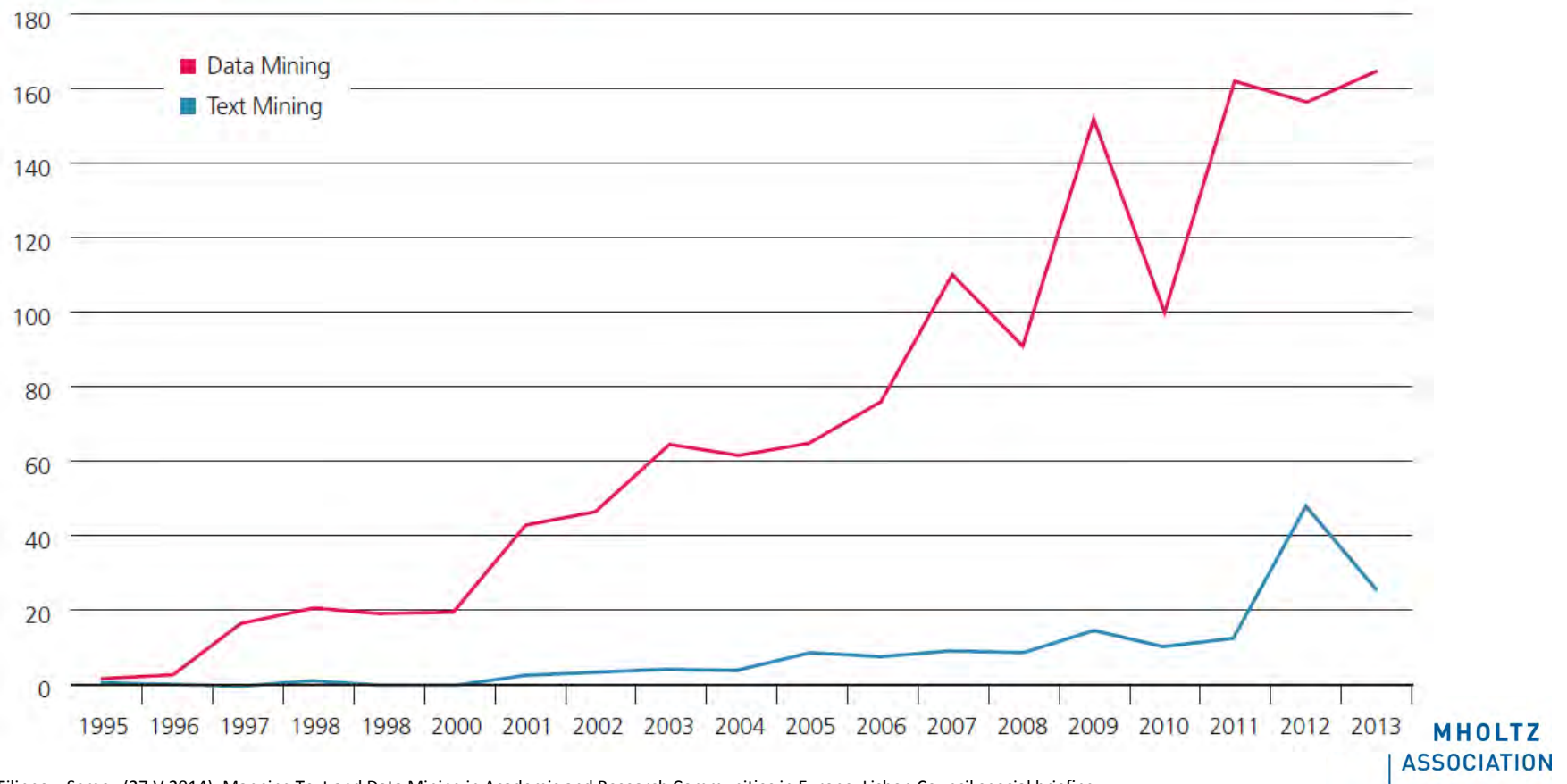
Direct and indirect use of TDM technology



TDM Service Provider

Growing importance of TDM for science

Chart 1: Number of Published Academic Articles about Text and Data Mining
In ScienceDirect database, per year



Filippov, Sergey (27.V.2014): Mapping Text and Data Mining in Academic and Research Communities in Europe, Lisbon Council special briefing: Text and data mining, 16/2014. www.lisboncouncil.net//index.php?option=com_downloads&id=1034

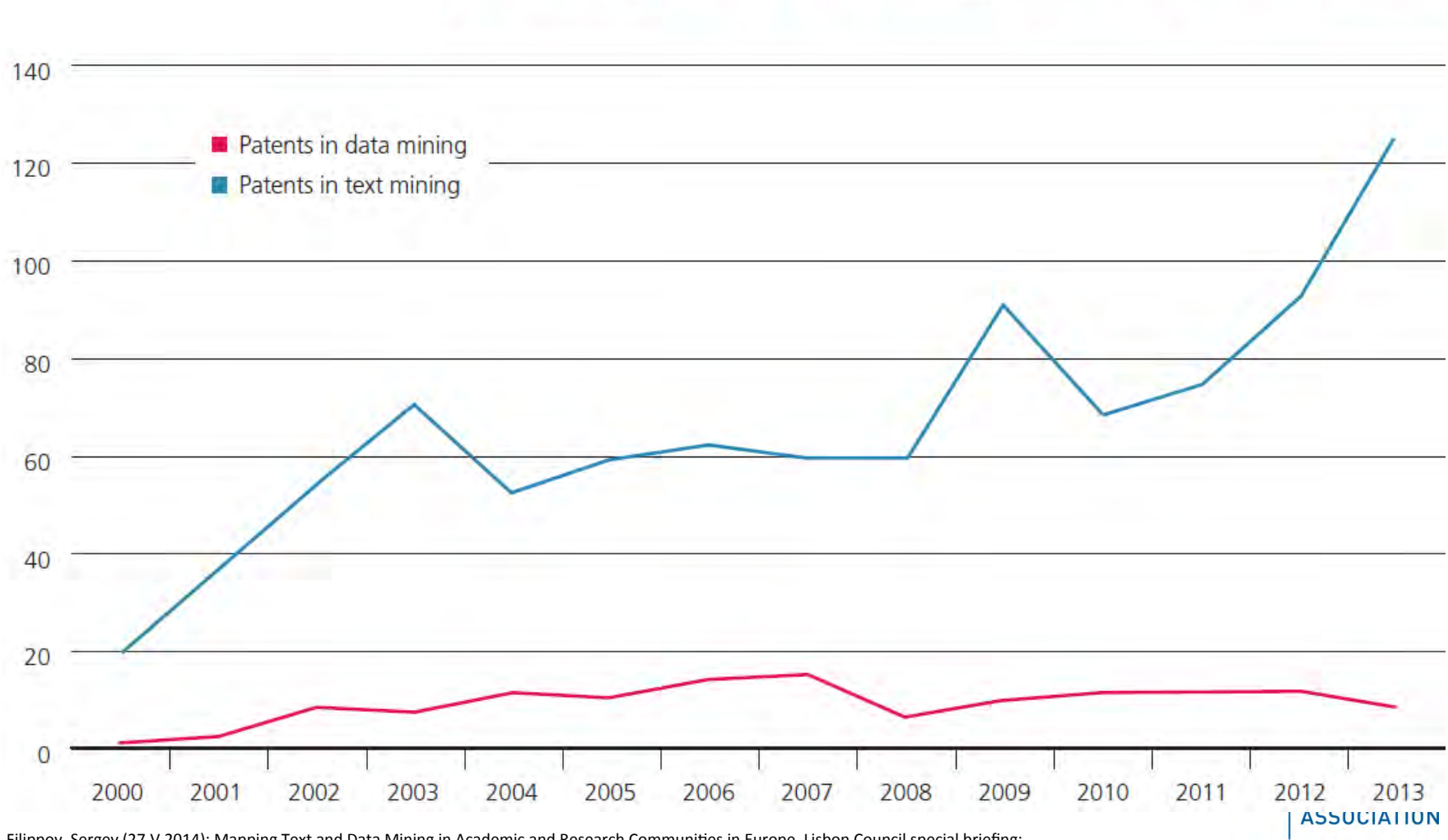
Lyon, 19.08.2014
Seite 13

Open Science

MHOLTZ
ASSOCIATION

Growing importance of TDM for science

Chart 2: Patents in Data Mining and Text Mining Granted in 2000-2013



Filippov, Sergey (27.V.2014): Mapping Text and Data Mining in Academic and Research Communities in Europe, Lisbon Council special briefing: Text and data mining, 16/2014. www.lisboncouncil.net//index.php?option=com_downloads&id=1034

Four hurdles against free TDM



Competitive advantage

- **USA**
 - Fair use probably covers TDM for non-commercial research
- **Canada**
 - Fair dealing probably covers TDM for non-commercial research
- **Israel**
 - Situation unclear but tendency to realign with US interpretation of fair use
- **Japan**
 - Statutory exception for TDM – no limitation to non-commercial research
- **United Kingdom**
 - Statutory exception for TDM for non-commercial research

Hargreaves, Ian et al. (2014): Standardisation in the area of innovation and technological development, notably in the field of Text and Data Mining, Brussels, European Commission. ec.europa.eu/research/innovation-union/pdf/TDM-report_from_the_expert_group-042014.pdf

EU Directives: exceptions

- **Copyright** Directive (2001/29/EC) &
- **Database** Directive (1996/09/EC)



Exception for education and research



Optional only



For non-commercial education/research only



Exceptions can be overridden by contractual agreements



Exceptions can be thwarted by technical protection measures

The content industry claims control

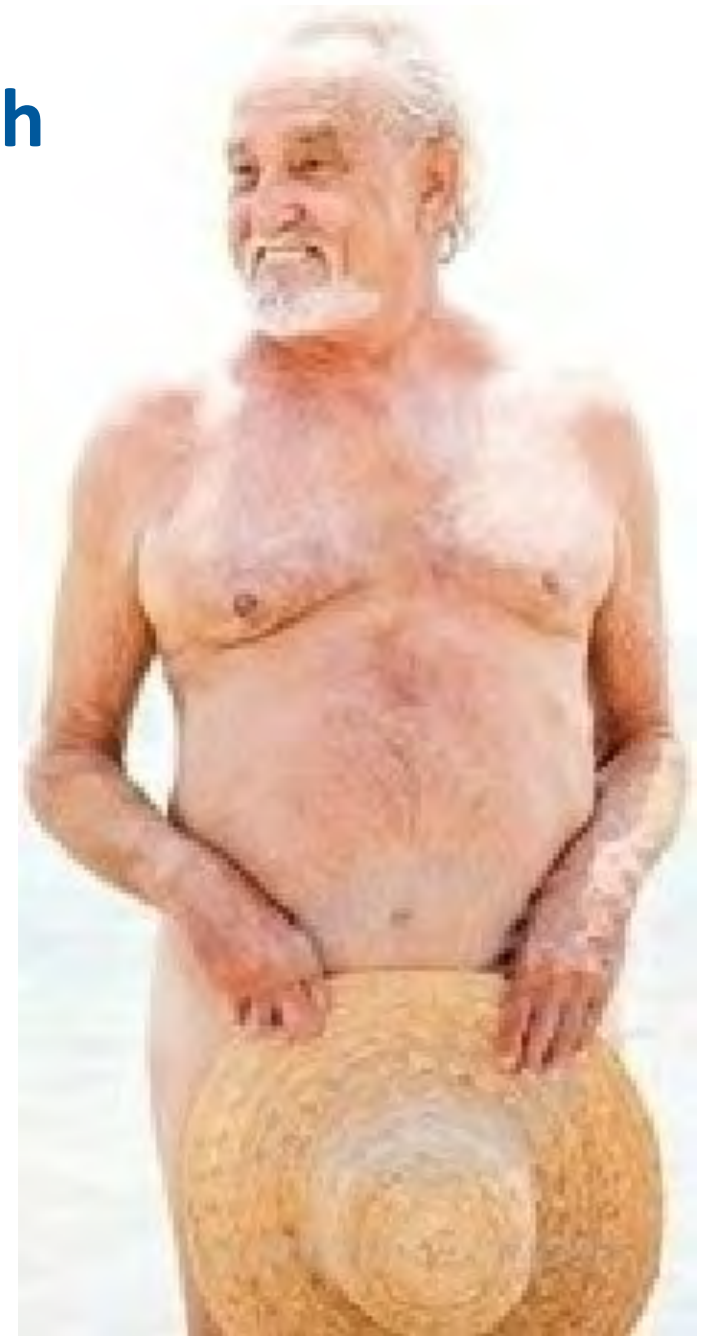


Researchers: We do not copy!

- Transient reproductions made in the cause of TDM may be reproductions from the legal point of view but they are not reproductions according to the researchers common understanding of the term.
- Researchers insist “We do not produce any copies in the cause of mining text and data!”
- The current legal situation is too complicated for the researchers to judge what they may do within legal boundaries.

Licenses are the wrong approach

- Transaction costs
 - Effort to identify and contact rights holders
 - Effort to negotiate terms of use
- Confidentiality issues
- Value creation chain
- License fees
- Competitive disadvantages compared to researchers in some other countries
- Control stays with the content industry



Need to raise awareness within the research community



- Many of the researcher using TDM technology are not aware of the associated legal implications.
- Likewise the management of research performing organisations lacks information on the extent to which TDM technology is use or of the associated legal implications.
- Science Europe does bring the legal issues and contractual hurdles to TDM to the attention of its member organisations.

Agreement – goal – disagreement

- **Agreement**

- There is unanimous agreement on the importance of TDM.

- **Goal**

- Assure that publicly financed content is freely (gratis and libre) available for reuse in order to maximise societal benefits including economic growth.

- **Disagreement**

- There is disagreement on the legal regime and the policies that should be employed to maximise the societal benefits of TDM.

What needs to be done



Possible remedies against TDM hurdles



What needs to be done: Policy measures

- Advocacy for real open access publishing = CC-BY
- Advocacy for adequate licences for research data – e.g. CC-0 (zero)
- No rights transfer for research data to publishers
- Funding for build-up and operation of publicly controlled infrastructures to make research publications and research data freely available independent of business interests of the content industry
- Policy to assure research data get fully archived in trusted repositories

Thank you very much for your attention

christoph.bruch@oa.helmholtz.de

CC-BY

- Where not indicated otherwise, the text of this work is licensed under the Creative Commons – Attribution 4.0 Unported License.
- To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0>
- Helmholtz Logo copyright Helmholtz Association