

The right to read is the right to mine!

Christoph Bruch
Helmholtz-Gemeinschaft

17. Helmholtz Open Access Webinar zu Forschungsdaten
10. Dezember 2013, 15:00

Was Sie erwartet

- Text und Data Mining
- TDM an Helmholtz-Zentren
- Konflikte um die rechtlichen Rahmenbedingungen von TDM
- Was tun?
- Fazit



- Dieser Vortrag gibt ausschließlich die Meinung des Autoren wieder.



Text und Data Mining

Definition des Begriffes Daten/Forschungsdaten

- In diesem Vortrag werden mit dem Begriff Daten oder Forschungsdaten alle Informationen bezeichnet, die von WissenschaftlerInnen im Rahmen ihrer Arbeit genutzt werden bzw. interessant sind.
- Diese Verständnis schließt Texte ein.



Text und Data Mining hilft bei der Bewältigung der Informationsflut und ...

- [Bild: Zwischen Büchern schlafender Wissenschaftler]



... ermöglicht neue Formen der Erkenntnisgewinnung

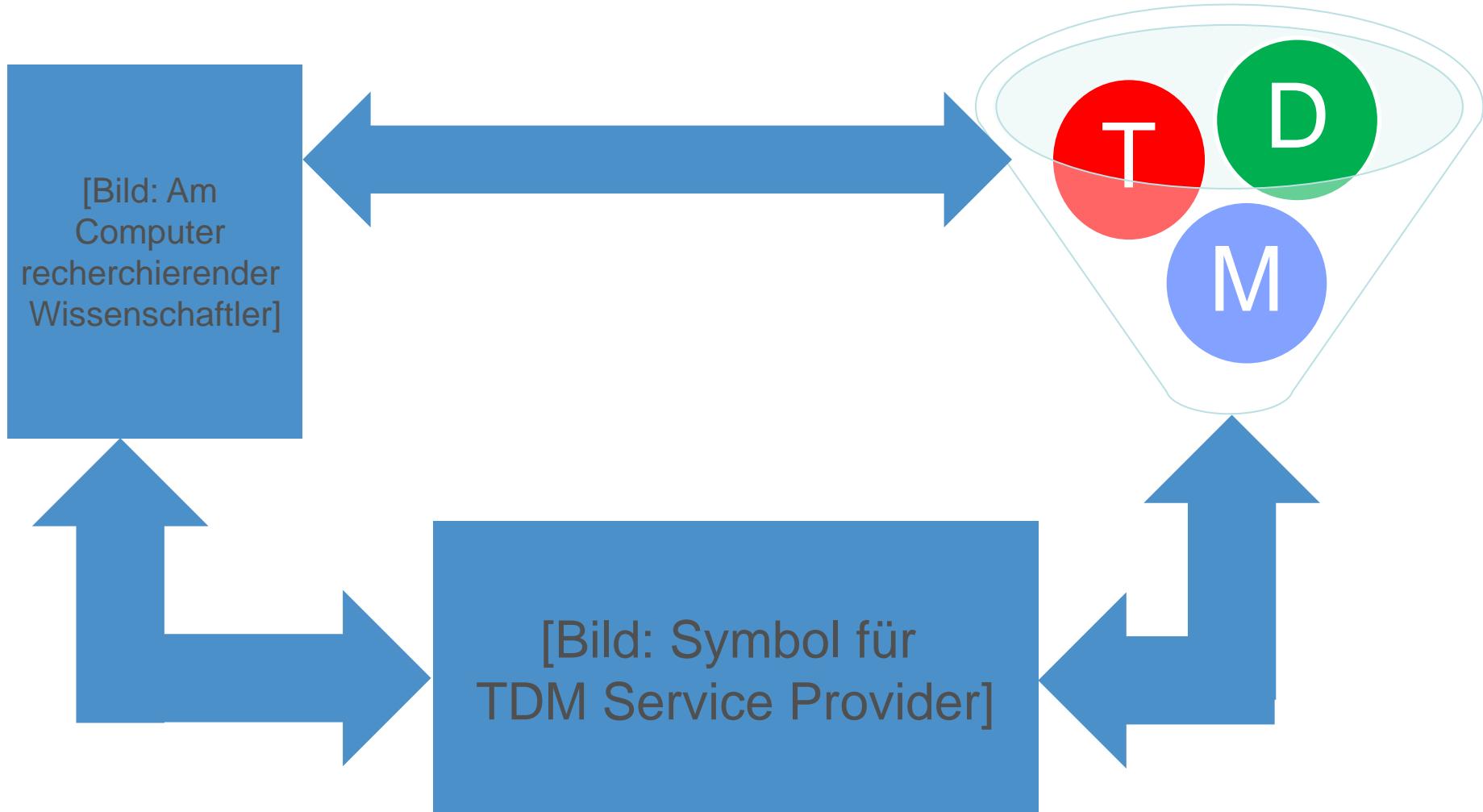
- [Bild: Hamburger]



Bild: Rind



Direkte und indirekte Nutzung von TDM Technologien



TDM Service Provider



www.exchange3d.com/images/uploads/aff197/office_building_2.jpg

<http://www.automarque.biz/images/researcher.jpg>

Berlin, 06.12.2013

Page 9

TDM ist unverzichtbar für Wissenschaft, Gesellschaft und Wirtschaft

- Ein stetig steigender Anteil von Forschungsdaten liegt in digitaler Form vor.
- Diese Daten können mit Hilfe von TDM-Technologien auf neue Weise analysiert werden.
- Von besonderem Interesse für WissenschaftlerInnen ist die kombinierte Auswertung von Forschungsdaten und wissenschaftlichen Publikationen.
- Der Nutzen der öffentlichen Forschungsförderung wird maximiert, wenn WissenschaftlerInnen ermöglicht wird
 - State of the art TDM Technologien zu nutzen
 - die erzielten Ergebnisse öffentlich zugänglich und nachnutzbar zu machen



Ohne TDM wären WissenschaftlerInnen halbblind.

No TDM

[Bild: Gruppe von WissenschaftlerInnen mit
„No-TDM-Balken“ vor den Augen]

No TDM

No TDM



TDM an Helmholtz-Zentren

Helmholtz-Gemeinschaft: Large Scale Data Management and Analysis

■ Partner aus der Helmholtz-Gemeinschaft

- Deutsches Elektronen-Synchrotron (DESY)
- Forschungszentrum Jülich GmbH
- GSI Helmholtzzentrum für Schwerionenforschung
- Karlsruher Institut für Technologie
- Deutsches Klimarechenzentrum

■ Mission

- Large Scale Data Management and Analysis" (LSDMA) extends the data services for research of the Helmholtz Association of research centres in Germany with community specific Data Life Cycle Laboratories (DLCL).

<http://www.helmholtz-lsdma.de/>



Helmholtz Zentrum München

Ausgründung: Clueda

- „Stümpflen und sein Kollegen beschäftigten sich am Helmholtz Zentrum München unter anderem mit dem Problem der Publikationsflut: Allein im Bereich Diabetes existierten damals etwa 400.000 Publikationen. Wollte eine Person alle diese Publikationen tatsächlich lesen, so wäre sie etwa 200 Jahre beschäftigt. Aus den von den Wissenschaftlern entwickelten Ansätzen, mittels Computer die Informationen zu sortieren und Zusammenhänge zu erkennen, entstand die Idee für die Entwicklung von weiteren Produkten und der Gründung einer Firma.“
- „Die Software analysiert nicht nur die grammatischen Strukturen der Sätze, sondern auch die sich daraus ergebenden Zusammenhänge und speichert sie in einer Art assoziativem Wissensnetzwerk.“

MDC for Molecular Medicine Group: Computational Biology and Data Mining

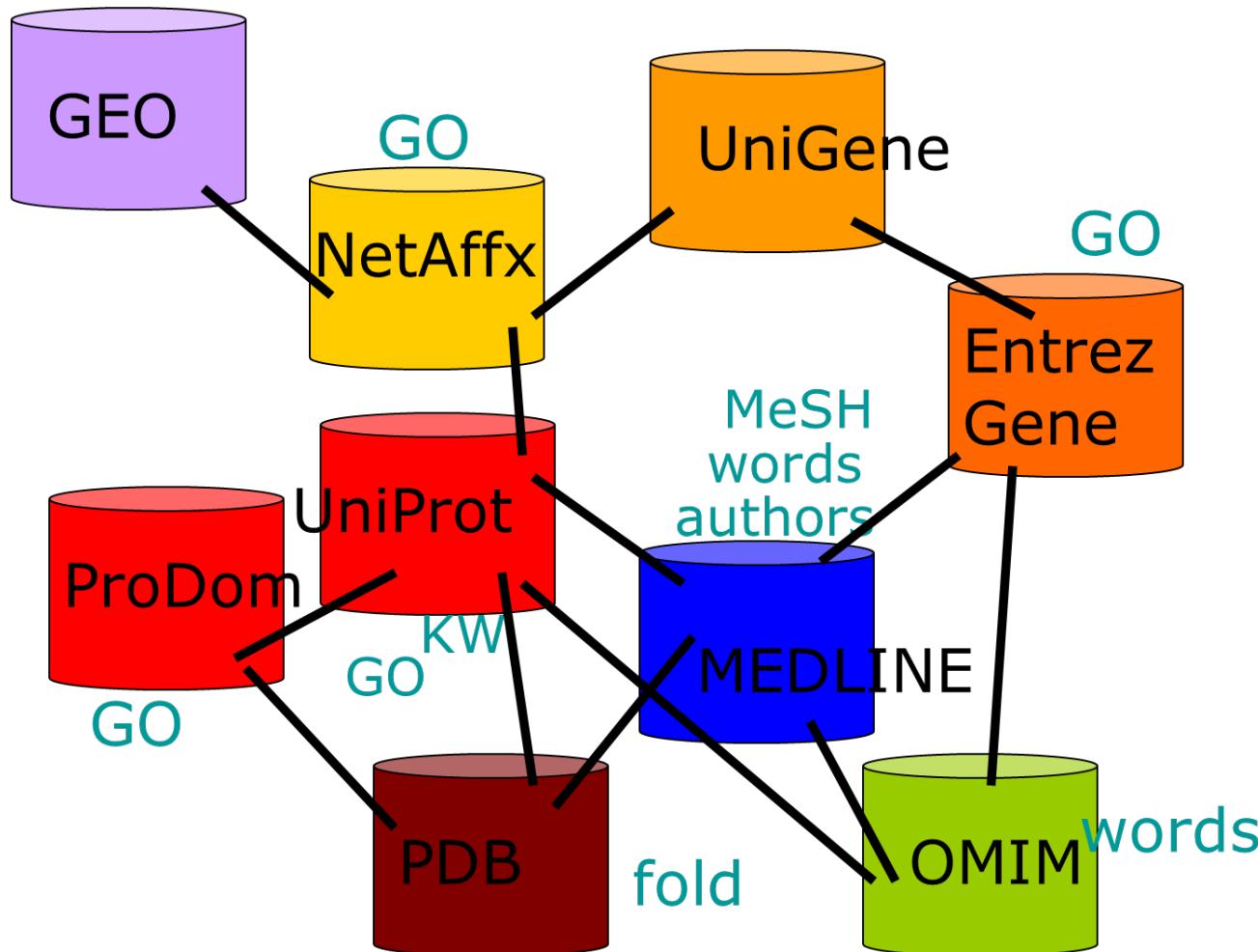


Illustration von Miguel Andrade, Group Leader Computational Biology and Data Mining

Berlin, 06.12.2013

Seite 15

Konflikte um die rechtlichen Rahmenbedingungen von TDM

[Bild: Ölfeld mit Fördertürmen]

refined to be used in a meaningful way to provide for value creation.

ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=2118

Das sieht die Content-Industrie wie die EU Kommission und beansprucht deshalb Kontrolle

[Bild: Gruppe von Personen,
die die Content-Industrie
symbolisieren sollen]



Drei Hürden beeinträchtigen freies TDM



WissenschaftlerInnen sagen: “Wir kopieren nicht!”

- “Vorübergehende Vervielfältigungshandlungen”, die bei der Anwendung von TDM-Technologien anfallen, mögen aus juristischer Sicht als Kopien gewertet werden, sie sind es jedoch nach verbreitetem Verständnis der WissenschaftlerInnen nicht.
- WissenschaftlerInnen bestehen darauf: “Beim Text und Data Mining produzieren wir keine Kopien!!”
- Die geltende Rechtslage ist zu kompliziert. WissenschaftlerInnen sind in dieser Situation damit überfordert, die Legalität ihres Handelns einzuschätzen.



Lizenzpflicht für TDM: Probleme

- Transaktionskosten
 - Aufwand der Identifikation und Kontaktaufnahme zu Rechteinhabern
 - Aufwand der Verhandlung von Nutzungsbedingungen
- Vertraulichkeit
- Wertschöpfungskette
- Lizenzgebühren
- Wettbewerbsnachteile gegenüber internationalen Konkurrenten

[Bild: Fast nackter Mann als Symbol für die Problematik der Verletzung der Vertraulich bei einem Zwang, Lizenzen für TDM abschließen zu müssen.]



Competitive advantage of USA and Japan

■ USA

- fair use which could enable text mining usage in non-commercial research

■ Japan

- The Japan Copyright Act (2011) is the only copyright law explicitly allowing text mining with a limitation - Article 47:
 - “For the purpose of information analysis (‘information analysis’ means to extract information, concerned with languages, sounds, images or other elements constituting such information, from many works or other much information, and to make a comparison, a classification or other statistical analysis of such information; the same shall apply hereinafter in this Article) by using a computer, it shall be permissible to make recording on a memory, or to make adaptation (including a recording of a derivative work created by such adaptation), of a work, to the extent deemed necessary”



TDM Mining wird so zur verbotenen Stadt

[Bild: Bild der verbotenen Stadt in Peking]



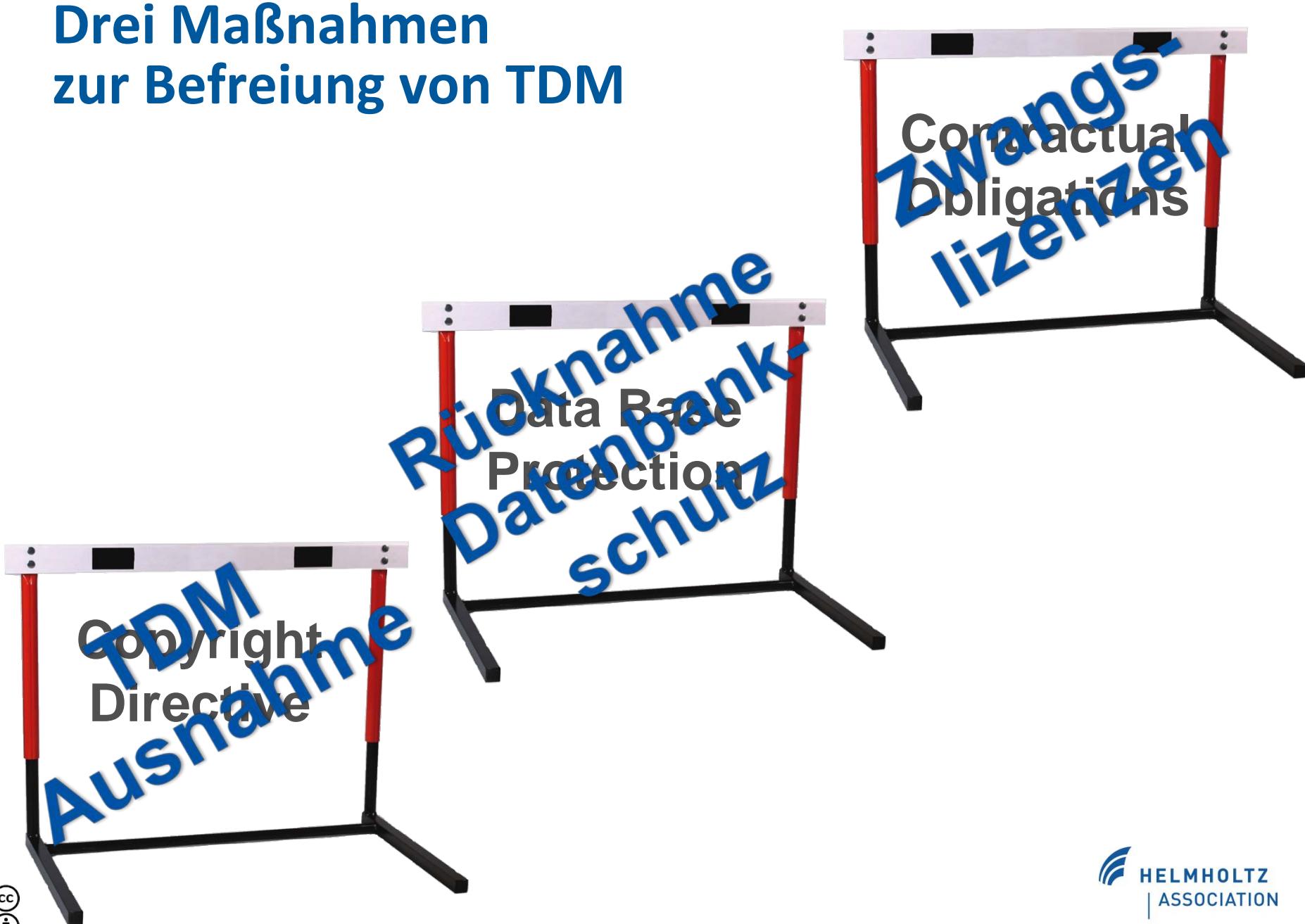
<http://hdwallpaperbackgrounds.info/wp-content/uploads/2012/04/Forbidden-City-China-Pictures.jpg>

Was tun?

[Bild: Zweit Fäuste, die ihre Fesseln zerreißen]



Drei Maßnahmen zur Befreiung von TDM



Was tun: Legislative Maßnahmen

- Unterstützung für den Vorstoß von Großbritannien, das europäische Urheberrecht um eine Fair Use-Schranke, die auch TDM umfassen soll, zu ergänzen.
- Gegebenenfalls eine eigenständige Schranke für TDM die auch eine kommerzielle Nutzung einschließt.
- Einführung einer Wissenschaftsschranke
- Urhebervertragsrechtliche Regelung: Zwangslizenz, um einen Zugang zu wissenschaftlichen Publikationen und die Nutzung von TDM-Technologien sicher zu stellen
- Rücknahme des Datenbankschutzesrechtes

Was tun: Policies

- **Ablehnung der Content-Industrie favorisierten „Lösung“ Lizenzierung**
 - Die Entscheidung über die Freigabe publizierte Inhalte für TDM bliebe bei den Rechteinhabern
 - Die Problematik der Verletzung der Vertraulichkeit (Forschungsfragen, Analysemethoden)
- **Publikationen**
 - Werbung für und Unterstützung des Open Access: Publizierens (CC-BY)
- **Forschungsdaten**
 - Schaffung von Anreizen für die Ermöglichung der Nachnutzung
 - Werbung für die Nutzung der CC-0 (zero) für Forschungsdaten
 - Bereitstellung finanzieller Mittel für den Aufbau & Betrieb öffentlich kontrollierter Infrastrukturen für wissenschaftliche Publikationen und Forschungsdaten
 - Keine exklusive Nutzungsrechtübertragungen an Verlage
 - Verankerung von Arbeitsabläufen, die sicher stellen, dass Forschungsdaten „nachnutzungskompatibel“ archiviert werden.



Fazit

Einigkeit – Ziel – Uneinigkeit

■ Einigkeit

- Einigkeit herrscht hinsichtlich der Einschätzung der großen Relevanz von TDM Technologien.

■ Ziel

- In der Politik und beim Wissenschaftsmanagement zeichnet sich eine Mehrheit für dafür ab, öffentlich finanzierte Inhalte (fast)kostenfrei zugänglich und nachnutzbar zu machen, um so den gesellschaftlichen und wirtschaftlichen Nutzen der eingesetzten Steuergelder zu maximieren.

■ Uneinigkeit

- Uneinigkeit herrscht hinsichtlich in der Frage, welche rechtlichen Regelungen am besten dem Erreichen des oben genannten Ziels dienen.

Was Sie tun können

- Das Bewusstsein für die Bedeutung von TDM und die damit verbundenen Konflikte steigern
- Bei WissenschaftlerInnen
 - Viele WissenschaftlerInnen, die indirekt TDM-Techniken nutzen, sind sich dessen nicht bewusst.
 - Auch die WissenschaftlerInnen, die bewusst TDM-Technologien anwenden, sind oft nicht mit den rechtlichen Aspekten von TDM vertraut.
- Beim Management von Wissenschaftsorganisationen
 - Dort liegen in der Regel weder Informationen zum Umfang der Nutzung von TDM Technologien vor, noch ist man mit der Rechtslage vertraut.
- Open Access Publizieren

Herzlichen Dank für Ihre Aufmerksamkeit

Christoph.bruch@oa.helmholtz.de



Open Access



- Where not indicated otherwise, the text of this work is licensed under the Creative Commons – Attribution 3.0 Unported License.
- To view a copy of this license, visit
<http://creativecommons.org/licenses/by/3.0>
- Helmholtz Logo copyright Helmholtz Association

