

# *Wem gehört die Zukunft der chemischen Information?*

ODOK, Eisenstadt 2016

Dr. Thomas Haubenreich  
Henrik Schreiber - Recherche und Beratung  
Heidelberg

# Das Buch zur Frage ...



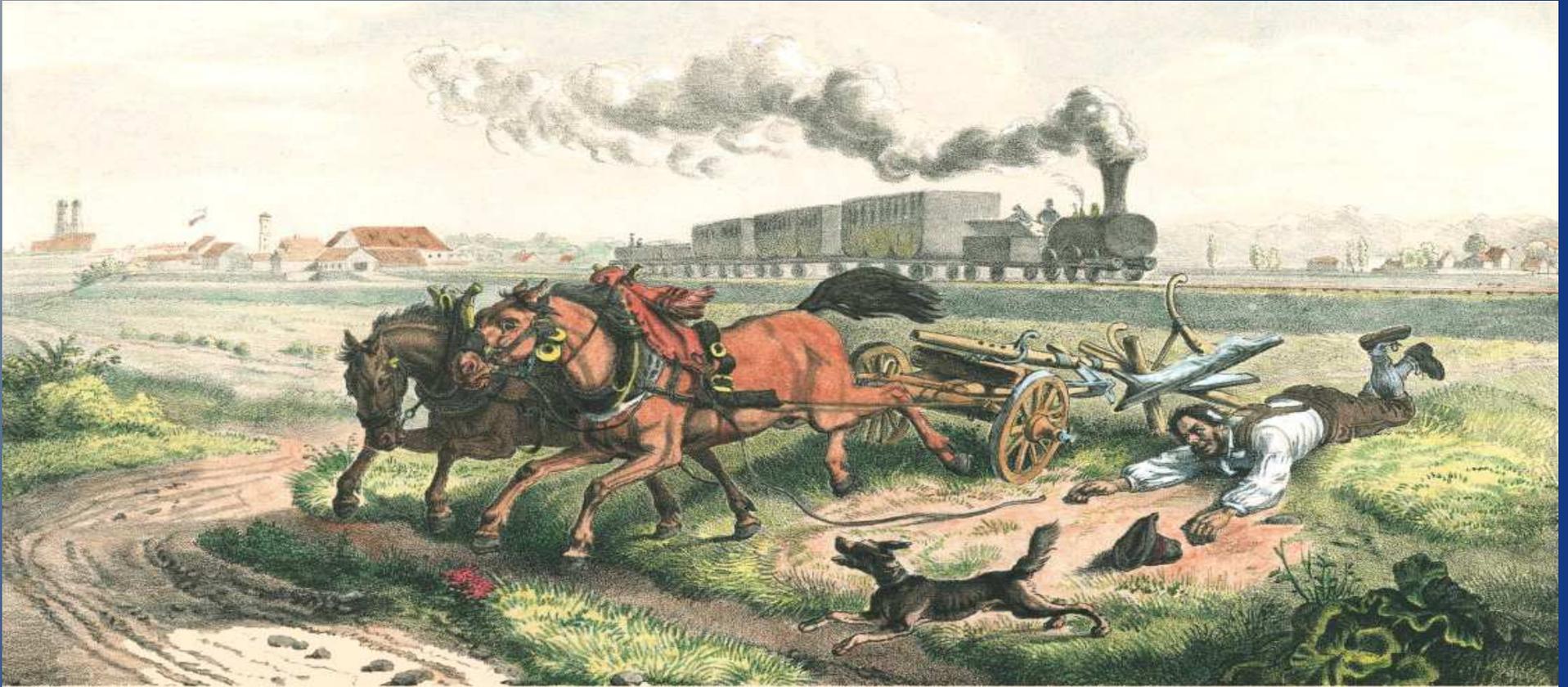
**Jaron Lanier**, Internetpionier,  
Musiker und Träger des  
Friedenspreises des deutschen  
Buchhandels 2014

- „Gewöhnliche Menschen teilen Informationen mit anderen, während ein paar elitäre Netzwerke gigantische Gewinne machen.“
- „Wem gehört die Zukunft“  
Hoffmann & Campe  
2014, Seite 39

# Chemische Information

- Texte
  - Wissenschaftliche Artikel (OA oder nicht?)
  - Patente (nicht in diesem Vortrag)
- Strukturen und Verbindungen
  - Eigentlich perfekt suchbar!
  - DIE Besonderheit der chemischen Information
- Erst die Verknüpfung bringt Suchbarkeit
  - Über Sekundärliteratur (z.B. CAS, Reaxys)
- Chemische Information veraltet nicht und ist von grossem kommerziellen Interesse!

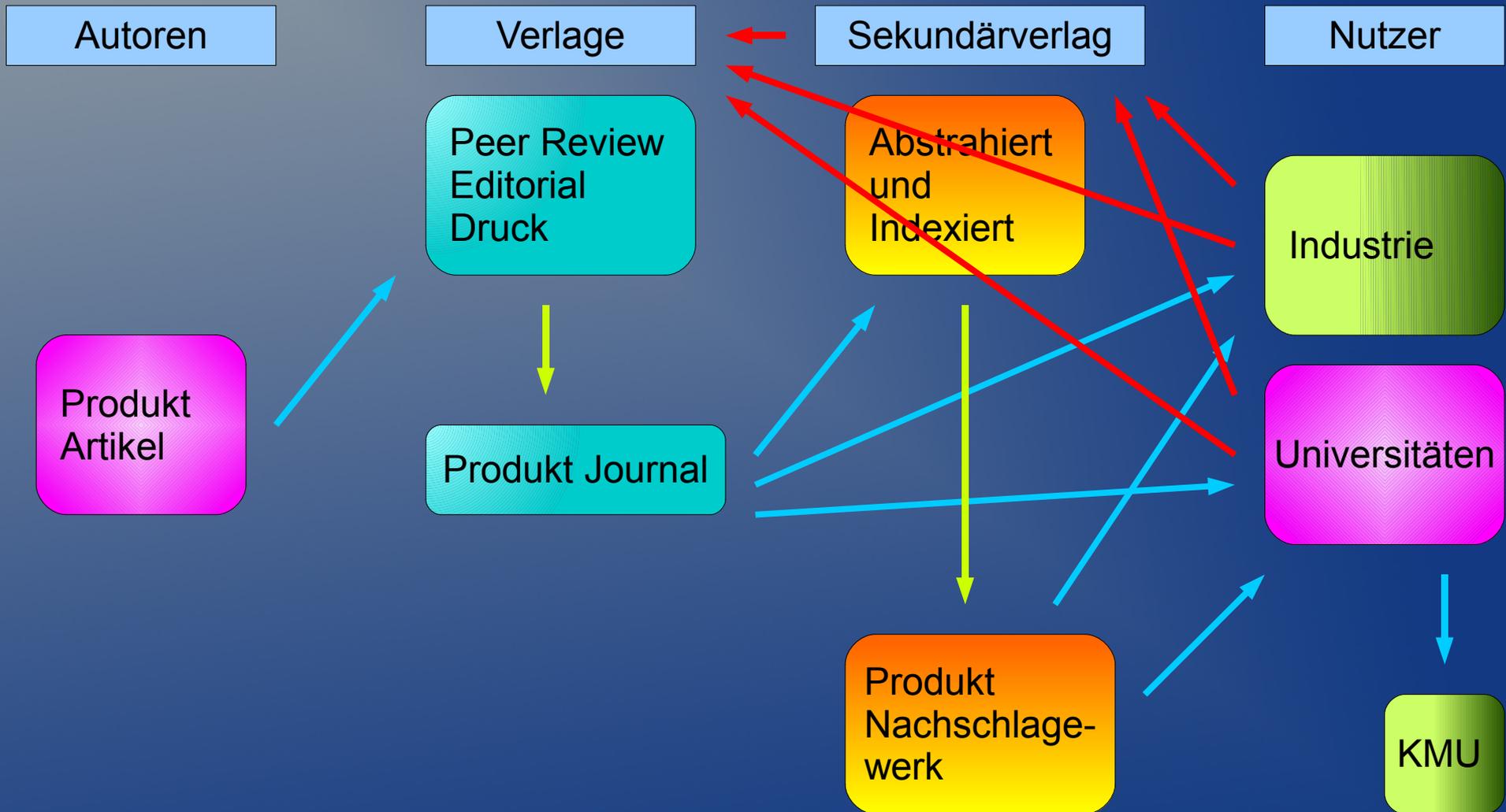
# Fortschritt oder Bedrohung?



# 1. Vor der digitalen Revolution

- Eine klar strukturierte Welt der wissenschaftlichen Literatur:
- **Autoren** – liefern die Artikel
- **Verlage** – kommerzielle und besonders Fachgesellschaften kuratieren und liefern das Produkt Journal
- **Sekundärliteratur** – z.B. Beilstein, Zentralblatt, CAS abstrahieren die verfügbaren Journale
- **Abonnenten** – Universitäten *und* Industrie bezahlen durch die Abonnements

# 1.1. Informations- und Geldströme vor der digitalen Revolution



## 2. Die digitale Revolution

- Eine kleine Auswahl an „buzzwords“
  - PDF
  - Open Access
  - Single Article
  - Online Zeitschriften im Paket
  - Konsortien
  - Piracy Server (SciHub...)
  - Social Media (Research Gate, Facebook, Linked In, Xing etc.)
  - Google!!!

## 2.1. Nach der Revolution?



- Die Revolution ist gerade im vollen Gang!!!

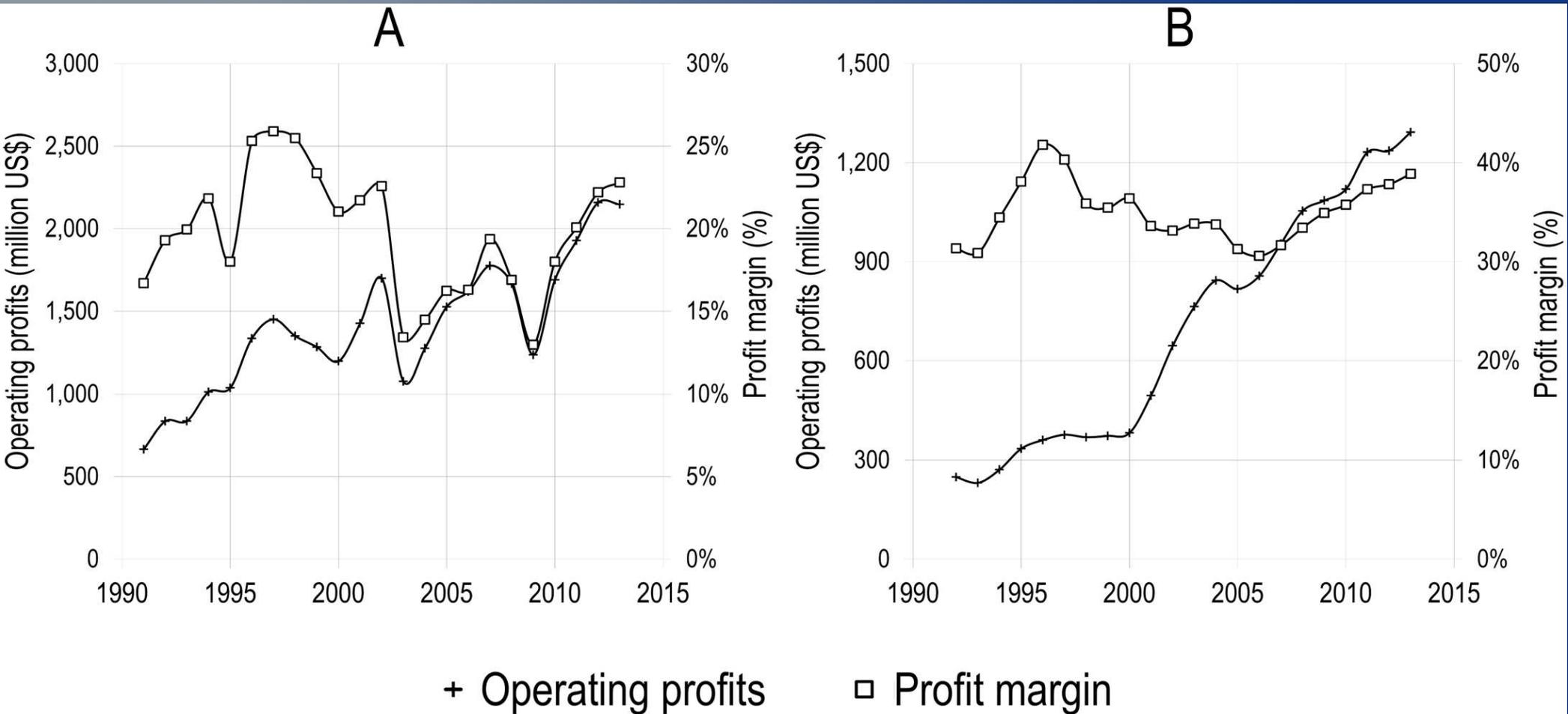
# 3.1. Verlage - Primärliteratur

- Die wichtigsten Primär-Verlage im Bereich Chemie:
  - Reed-Elsevier
  - Wiley-Blackwell
  - Springer
  - Taylor & Francis
  - ACS
  - RSC
- Über 70% der veröffentlichten Artikel



<https://svpow.com/2016/05/21/i-did-my-research-yes-i-think-academic-publishers-are-greedy-with-notes-on-publishers-rhetoric-and-creationism/>

# 3.1. Gewinnentwicklung Elsevier



Larivière V, Haustein S, Mongeon P (2015)

The Oligopoly of Academic Publishers in the Digital Era. PLoS ONE 10(6): e0127502. doi:10.1371/journal.pone.0127502

Operating profits (million USD) and profit margin of Reed-Elsevier as a whole (A) and of its Scientific, Technical & Medical division (B),

1991–2013. Compilation by the authors based on the annual reports of Reed-Elsevier.

(<http://www.reedelsevier.com/investorcentre/pages/home.aspx>)

## 3.2. Verlage - Sekundärliteratur

- Die wichtigsten Sekundär-Verlage im Bereich Chemie:
  - CAS (Division of ACS)
  - Reed-Elsevier
  - Thomson Scientific & IP
- Konfrontiert mit der Flut an Daten!
- Kostenintensive Aufarbeitungsprozesse
- Unersetzlich?

## 3.3. Greedy Publishers?

- Thomson Reuters Science & IP wurde am 11.7.2016 verkauft
- Waren 32% Gewinn zu wenig...?
- Jetzt Eigentümer: Onex Corp. und Baring Private Equity Asia
- 3,5 Mrd. US \$



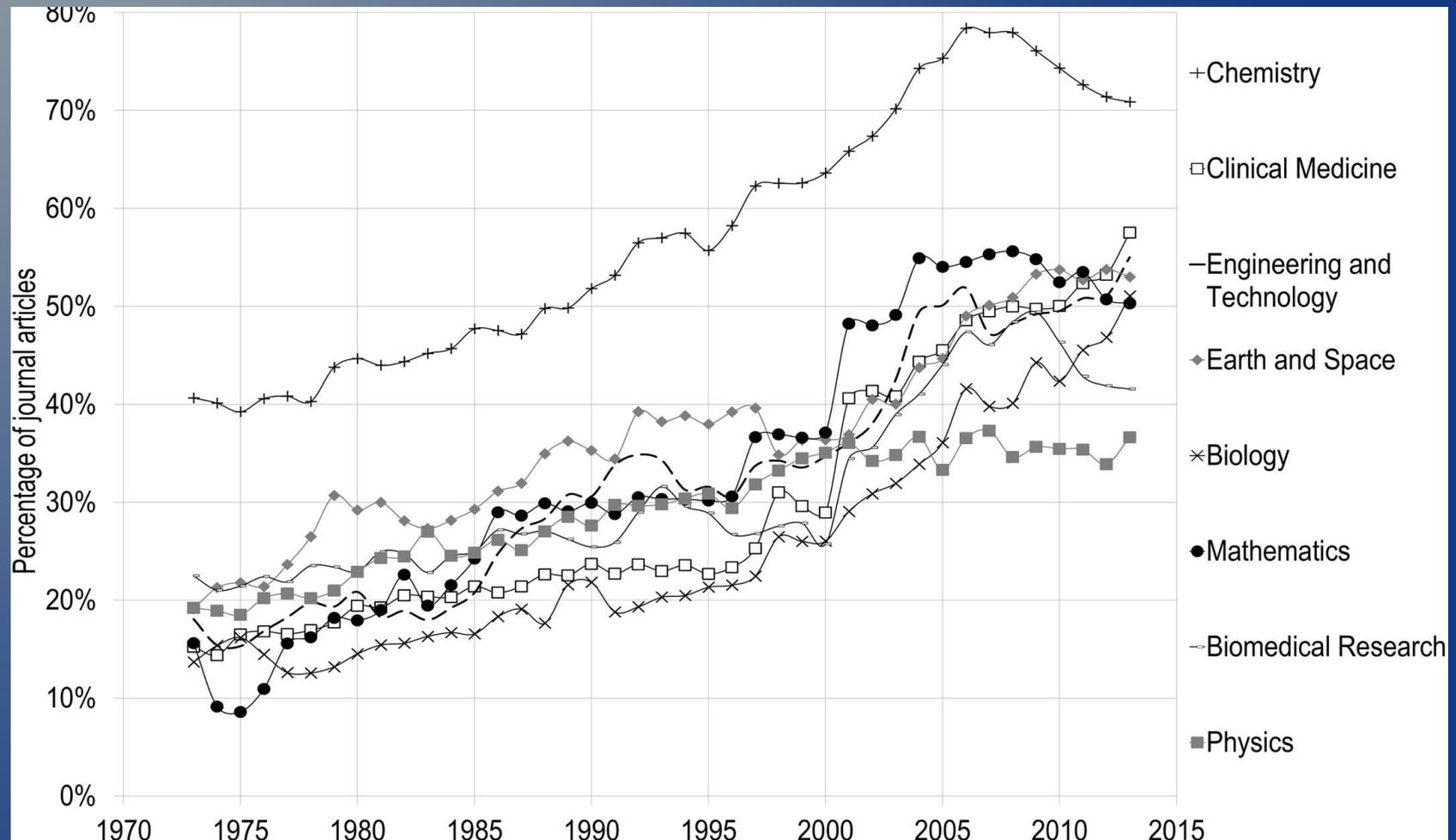
<http://www.reuters.com/article/us-thomsonreuters-ipbusiness-onex-corp-idUSKCN0ZR13T>

## 3.4. Sind die Verlage Sirenenserver?



- „Die Sirenenserver lenken einen Grossteil der Produktivität gewöhnlicher Menschen in Richtung einer informellen Tausch- und Reputationswirtschaft, während sie das gewonnene altmodische Vermögen auf sich selbst konzentrieren.“
  - „Wem gehört die Zukunft“, Seite 90

# 3.5. Trotzdem: Publikationen bei den grossen 5



Larivière V, Haustein S, Mongeon P (2015)

The Oligopoly of Academic Publishers in the Digital Era. PLoS ONE 10(6): e0127502. doi:10.1371/journal.pone.0127502

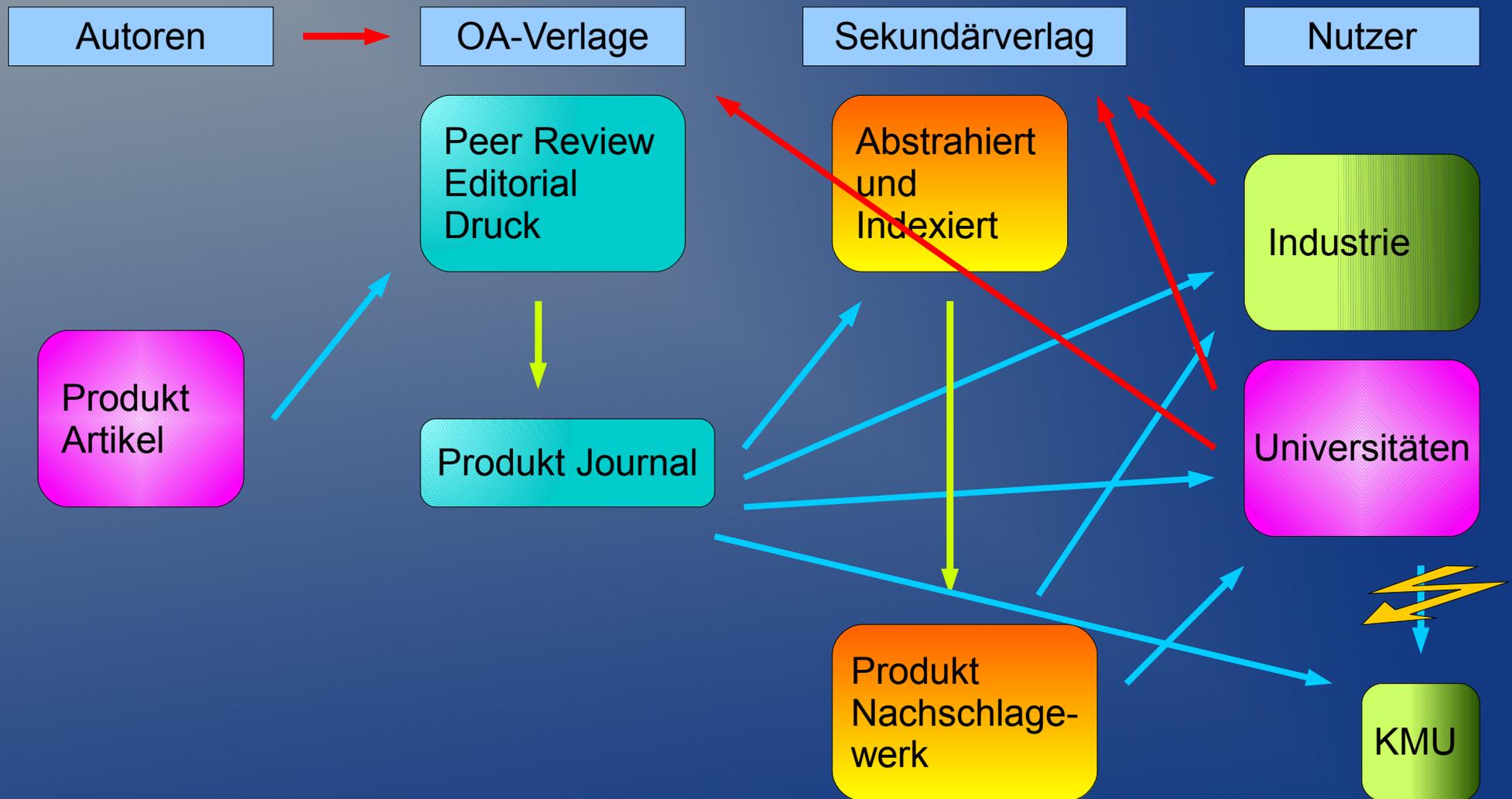
# 4. Open Access

- Initiative Europa
  - Erklärtes Ziel: OA für alle Publikationen bis 2020
  - Science M.Enserink DOI:  
10.1126/science.aaf9911
- Max Planck Gesellschaft
  - Die MPG ist besonders aktiv und fördert OA
  - <https://openaccess.mpg.de/>

## 4.1. Chemistry Preprint Server

- WASHINGTON, Aug. 10, 2016 — The American Chemical Society (ACS) today announced its intention to form ChemRxiv, a chemistry preprint server for the global chemistry community, proposed as a collaborative undertaking that will facilitate the open dissemination of important scientific findings. The Society is presently in the process of inviting interested stakeholders to participate in helping to shape the service ahead of its anticipated launch.

# 4.2. Informations- und Geldströme bei Open Access



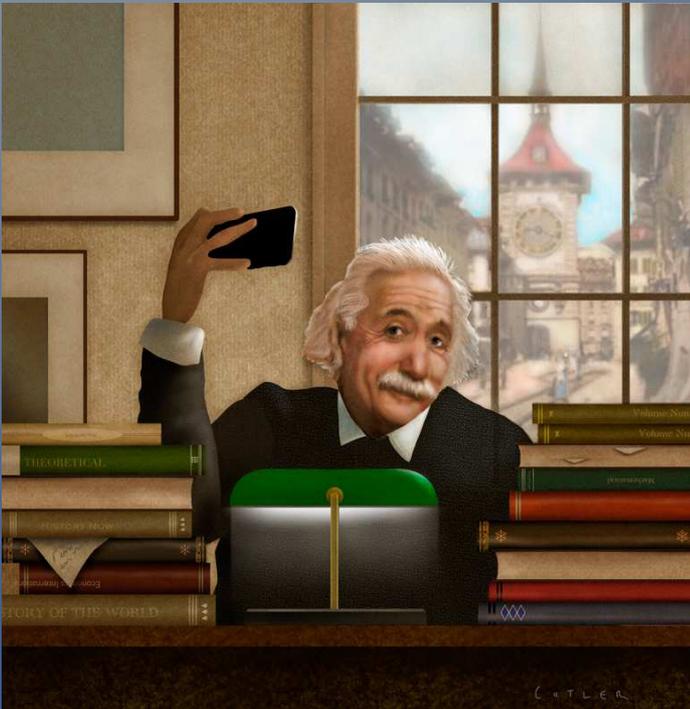
# 4.3. OA – Risiken und Nebenwirkungen

- Umkehr der Geldströme
  - Autor bezahlt – Industrie nutzt?
  - Der Autor ist der Geldgeber – Qualität?
- Kein Vertrag zwischen Nutzer und OA Verlag
  - Beständigkeit und Qualität?
- Predatory Journals
  - Bealls List - <https://scholarlyoa.com/publishers/>
- Cleanup bei DOAJ
  - <https://doajournals.wordpress.com/2016/05/09/doaj-to-remove-approximately-3300-journals/>

# 5. Informationsflut

- Patente
- Journale
- Open Access
  
- Ich verzichte in diesem Vortrag auf viele exponentiell nach oben wachsende Kurven :-) !!!

# Datenflut - ironisch



- “an academic career, in which a person is forced to produce scientific writings in great amounts, creates a danger of intellectual superficiality” - Albert Einstein

# 6. Wie findet man das gesuchte?

- Traditionelle Sekundärliteratur
  - CAS
  - Elsevier
  - Thomson Reuters (wie lange noch?)
- Internet (for free?)
  - Google
  - PubChem / NIH
  - Surechem/ChEMBL
- IBM SIIP Cloud Solution (?)

# 7. Wem gehört die Information?

- Commercial vs. Non-for-Profit
  - Merke: There is nothing like a free lunch!
- Sirensenserver: Google? IBM? CAS? Elsevier?
- Diskussion nach dem Vortrag erwünscht!

# 8. Gamechangers?

- IBM Watson
  - Zusammenarbeit mit NIH
- Google CPC Indexing für NPL
  - Immer für Überraschungen gut
- Text- und Datamining (2 Beispiele)
  - Minesoft
  - Infoapps
- **Es gilt: Neue Tools kritisch evaluieren!**

# 8.1. IBM Watson

- Sicher die grösste Investition in Augmented Intelligence (2 Mrd. US\$, 2000 Mitarbeiter)
- Vortrag Dr. Stephen Boyer
- IBM BAO strategic IP insight platform (SIIP)
  - <http://www-935.ibm.com/services/us/gbs/bao/siip/>
- Kooperation mit dem NIH und EMBL
  - Verknüpfung von Substanzinformation mit Quellen (Patente, Open Access Papers ( ? ) )
  - <http://www-935.ibm.com/services/us/gbs/bao/siip/nih/>

# 8.1. IBM - SureChEMBL

- IBM arbeitet auch mit SureChEMBL zusammen
  - Beta Version beobachten!
  - Verknüpfungen zu IBM Patent Server

The screenshot displays the SureChEMBL website interface. The main content area shows search results for the chemical structure 5-[4-(2H-1,3-benzodioxol-5-yl)-4-oxobut-2-yn-1-yl]-5-(3-chlorophenyl)-2,2-dimethyl-1,3-dioxolan-4-one. The InChI key is DJGUAWSNAUSHFX-UHFFFAOYSA-N. A table indicates the number of occurrences across different document sections: Title (5), Abstract (134), Claims (50), Description (1514), Images (344), and CWL's (2) (45). A bar chart visualizes these counts, with a note 'Coming soon: Analysis of structure occurrence'. Below the chart, there are sections for UniChem Cross References, Related structures, and Patent hits, with a list of patent sources including IBM Patent System, PubChem-Thomson Pharma, PubChem, and Nikkaji.

5-[4-(2H-1,3-benzodioxol-5-yl)-4-oxobut-2-yn-1-yl]-5-(3-chlorophenyl)-2,2-dimethyl-1,3-dioxolan-4-one

InChI key: DJGUAWSNAUSHFX-UHFFFAOYSA-N

Section	Occurrences
Title	5
Abstract	134
Claims	50
Description	1514
Images	344
CWL's (2)	45

Coming soon: Analysis of structure occurrence  
A breakdown of document locations and data sources for structures across all SureChEMBL patents

Mol Weight	Lipinski Ro5	
412.820	Yes	
Log P	Donor Cnt	Ring Cnt
4.70	0	4
Accept Cnt	Rotable Bond Cnt	
5	5	

UniChem Cross References: Related structures Patent hits

UniChem cross references (Click to expand)

IBM Patent System (1) PubChem-Thomson Pharma (1) PubChem (1) Nikkaji (1)

# 8.2. Google CPC Indexing für NPL

- Google investiert in automatische Indexierung

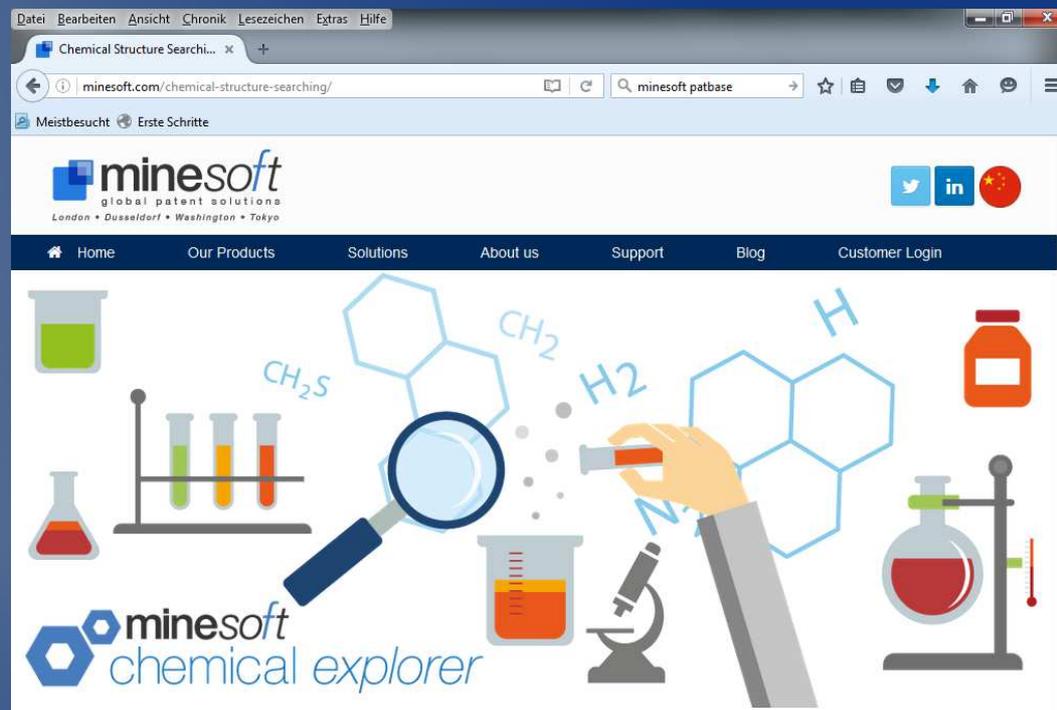
- Aber kein primärer Fokus des Giganten
- Zugänglich in Google Patents mit NPL Suche

The screenshot shows a web browser window displaying a Google Patents search result. The search query is "enantioselective protonation boron enolates". The search results show 4 of 3,024 results. The top result is titled "Stereoselective protonation of boron enolates" by Haubenreich et al., 1993. The snippet reads: "In most enantioselective protonations of formal carbanions a chiral protonic acid HX\* attacks a prochiral C atom.[] Noncovalent interactions between the substrate, which occur in the unknown, strongly variable ion aggregates, and HX\* have a decisive influence on the (...". The classification section lists several CPC codes and their descriptions, such as C07D207/26 (2-Pyrrolidones), C07D207/12 (Oxygen or sulfur atoms), C07D207/34 (Heterocyclic compounds containing five-membered rings not condensed with other rings, with one nitrogen atom as the only ring hetero atom with only hydrogen or carbon atoms directly attached to the ring nitrogen atom having two double bonds between ring members or between ring members and non-ring members with heteroatoms or with carbon atoms having three bonds to hetero atoms, with at the most one bond to halogen, e.g. ester or nitrile radicals, directly attached to ring carbon atoms), C07D207/16 (Carbon atoms having three bonds to hetero atoms with at the most one bond to halogen, e.g. ester or nitrile radicals), C07D207/08 (Heterocyclic compounds containing five-membered rings not condensed with other rings, with one nitrogen atom as the only ring hetero atom with only hydrogen or carbon atoms directly attached to the ring nitrogen atom having no double bonds between ring members or between ring members and non-ring members with hydrocarbon radicals, substituted by hetero atoms, attached to ring carbon atoms), and C07D451 (Heterocyclic compounds containing condensed hetero ring systems). The right sidebar shows the author information: Haubenreich et al., 1993, From Google Scholar, Author: Haubenreich T, Hüinig S, Schulz H, Publication year: 1993, Publication venue: Angewandte Chemie International Edition in English, External links: Cited by, Info: Similar documents.

# 8.3. Minesoft Chemical Explorer

- Minesoft bietet chemische Informationen

- Nur in Patenten (Volltexte frei!)



**Chemical Structure Searching for full text patents**

**Extract essential chemical information from full text patents with Chemical Explorer**

Chemical Explorer is a chemical database from Minesoft that enables searchers to instantly retrieve chemical compounds from the full text of millions of patent documents and start reviewing results in seconds.

It is the first web-based full text chemical structure search tool available, covering leading **patent authorities which publish in English, Japanese, Chinese and Korean languages.**

Chemical Explorer covers multiple patent-issuing authorities, currently including US (USA), GB (UK), AU (Australia), IL (Israel), EP (European Patent Office), WO (PCT), FR (France), DE (Germany) and IN (India), as well as Japan (JP), China (CN) and Korea (KR). It contains over **12 million unique chemical compounds** from over 10 million full text patent documents (as of September 2015).

**Unique Asian coverage** is included – data-mining chemical data within Japanese, Chinese and Korean patents. Chemical Explorer is **updated daily** as new patents publish, ensuring users have access to the very latest information. Chemical terminology such as names, including trade and IUPAC names can be searched, and direct links to external sources are included to enable further research.

Chemical Explorer includes **full Chemical Structure drawing capability** and the ability to import/export chemical structures, which means users do not have to rely on keyword searches. Flexible search options are offered allowing users to search based on identity, similarity or substructure

**Quick look at Chemical Explorer**

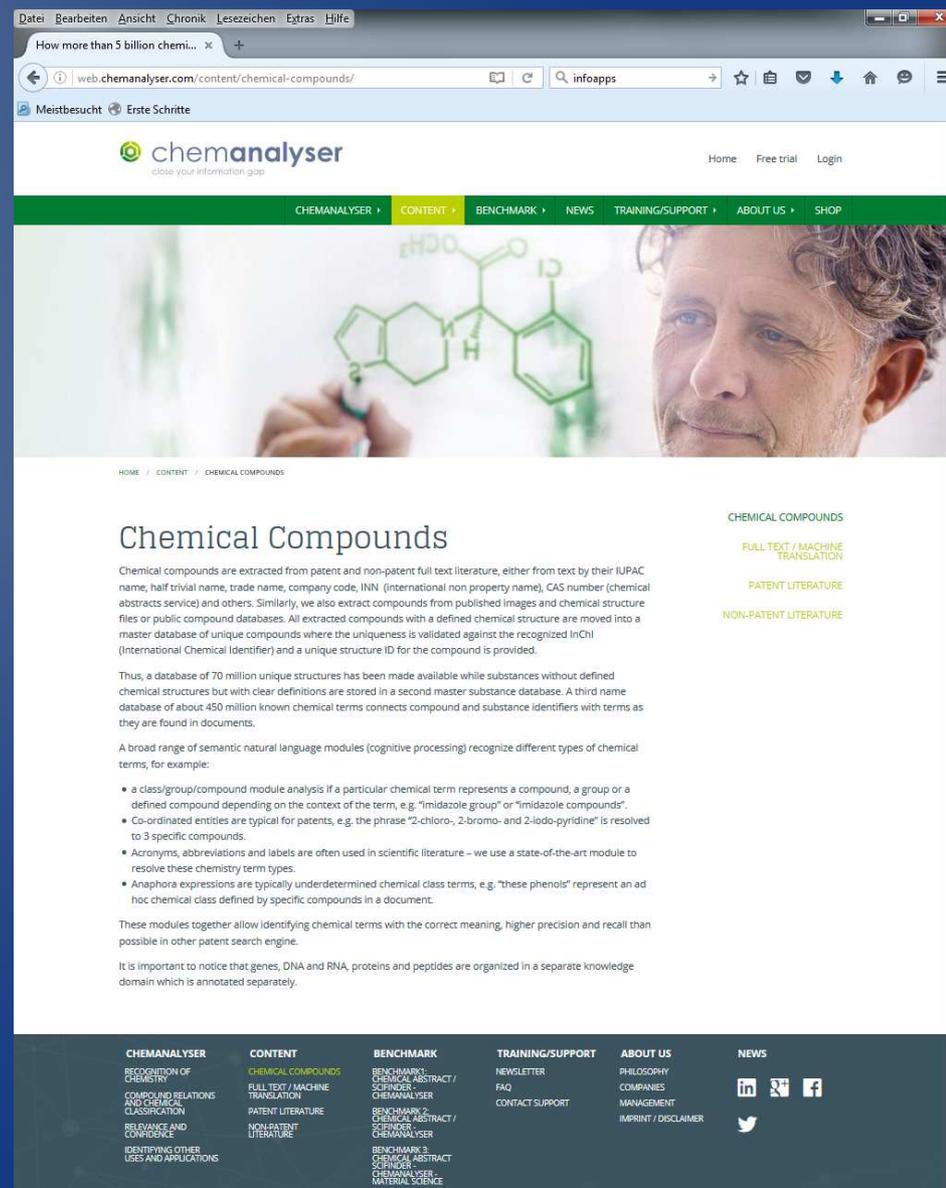
- A database mining for chemical structures contained in the full text of patent documents
- Covers 12 full text collections including unique non-Latin authorities (JP, CN, KR)
- Chemical Structure Drawing capability and flexible search options
- Option for [PatBase](#) users to view pertinent documents in PatBase using the TextMine Viewer
- [Chemical Explorer brochure](#)
- [For a free trial of Chemical Explorer click here](#)

 **textmine**  
powered by minesoft

# 8.4. Infoapps chemanalyser

## chemanalyser

- in Patenten (Volltexte frei!) und NPL (?)
- Ging am Montag online
- YAT???



The screenshot shows a web browser window displaying the chemanalyser website. The address bar shows the URL [web.chemanalyser.com/content/chemical-compounds/](http://web.chemanalyser.com/content/chemical-compounds/). The website header includes the logo "chemanalyser" with the tagline "close your information gap" and navigation links for Home, Free trial, and Login. A green navigation bar contains links for CHEMANALYSER, CONTENT (active), BENCHMARK, NEWS, TRAINING/SUPPORT, ABOUT US, and SHOP. The main content area features a large image of a scientist in a lab coat looking at a chemical structure. Below the image, the page title is "Chemical Compounds". The text describes the extraction of chemical compounds from patent and non-patent full text literature, mentioning the use of IUPAC names, JUPAC names, trade names, company codes, INN, CAS numbers, and chemical structures. It also mentions a database of 70 million unique structures and a second master substance database. A list of chemical terms is provided as an example of semantic natural language modules. The footer contains a grid of links for CHEMANALYSER, CONTENT, BENCHMARK, TRAINING/SUPPORT, ABOUT US, and NEWS.

## 8.5. Text- und Datamining – Risiken und Nebenwirkungen

- Man muss Zugriff auf den Volltext haben!
  - Wem gehören die Daten / Server! Was darf man damit tun?
- Aufwändige Lern- und Trainingsprozesse
  - Wer kann den Aufwand finanzieren?
- Sprachbarrieren
- Keine Validierung der Qualität der Indexierung?
- Vollständigkeit

# 9. Zusammenfassung

- Vorhersagen sind schwierig, besonders wenn es um die Zukunft geht!
- Eine weitere Aufspaltung der chemischen Information ist mehr als wahrscheinlich!
- Die Welt der chemischen Information wird erheblich komplexer – Informationskompetenz!
- Es gibt in einer hochkomplexen Welt keine einfachen Lösungen – ein gesundes Misstrauen ist angebracht!

# 10. Dank an alle :-)

**Besonders für Ihre Aufmerksamkeit!**

Henrik Schreiber, Recherche und Beratung  
Albrecht Storz, Recherche und Beratung

Thorsten Rohde, ACSI  
Prof. Damon Ridley  
Jochen Lennhof, Minesoft  
Sascha Kammhuber, InfoApps  
Tony Trippe, Patinformatics  
u.v.a.