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

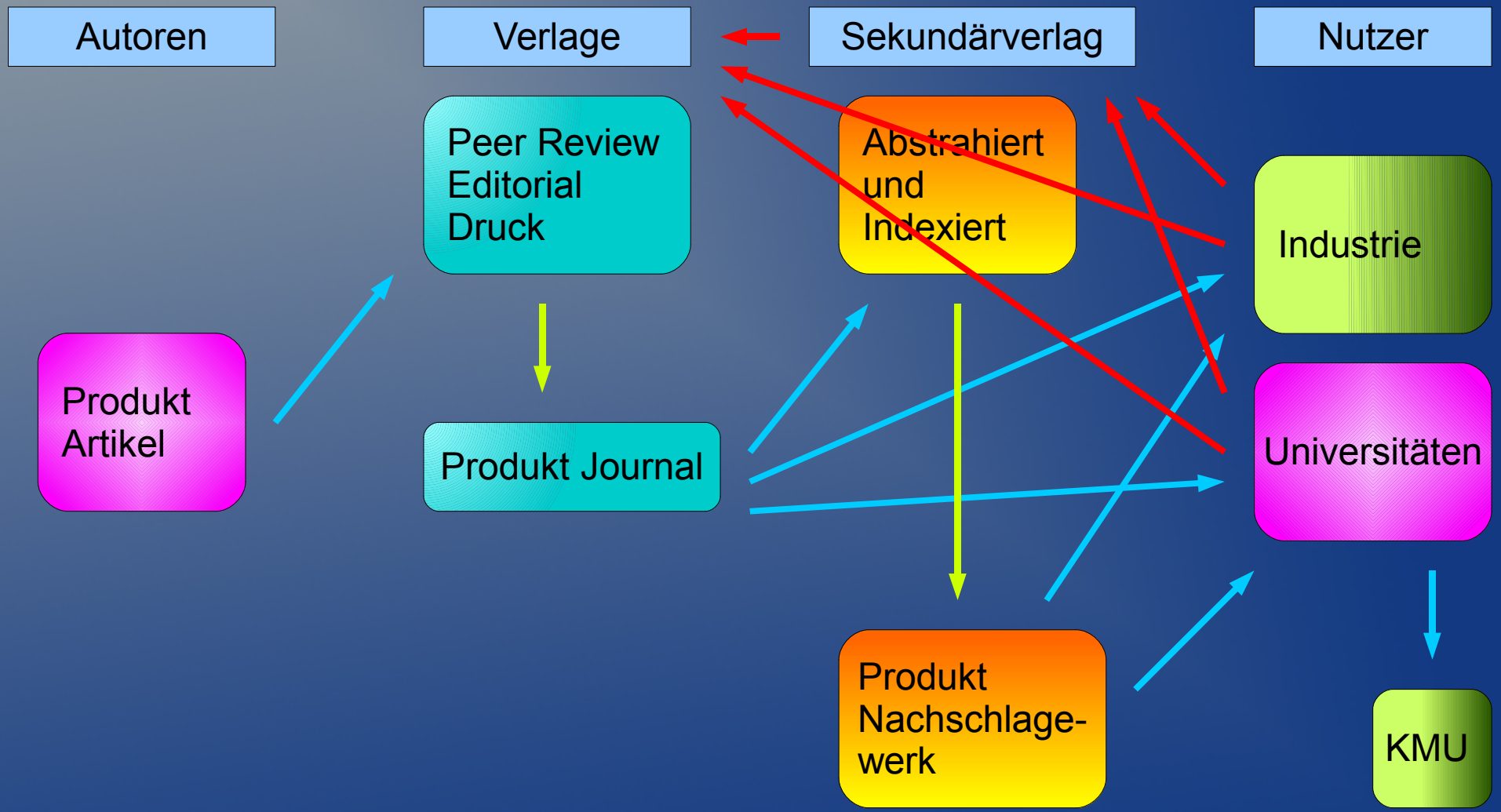
- Texte
  - Wissenschaftliche Artikel (OA oder nicht?)
  - Patente (nicht in diesem Vortrag)
- Strukturen und Verbindungen
  - Eigentlich perfekt suchbar!
  - DIE Besonderheit der chemischen Information
- Erst die Verknüpfung bringt Suchbarkeit
  - Über Sekundärliteratur (z.B. CAS, Reaxys)
- Chemische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

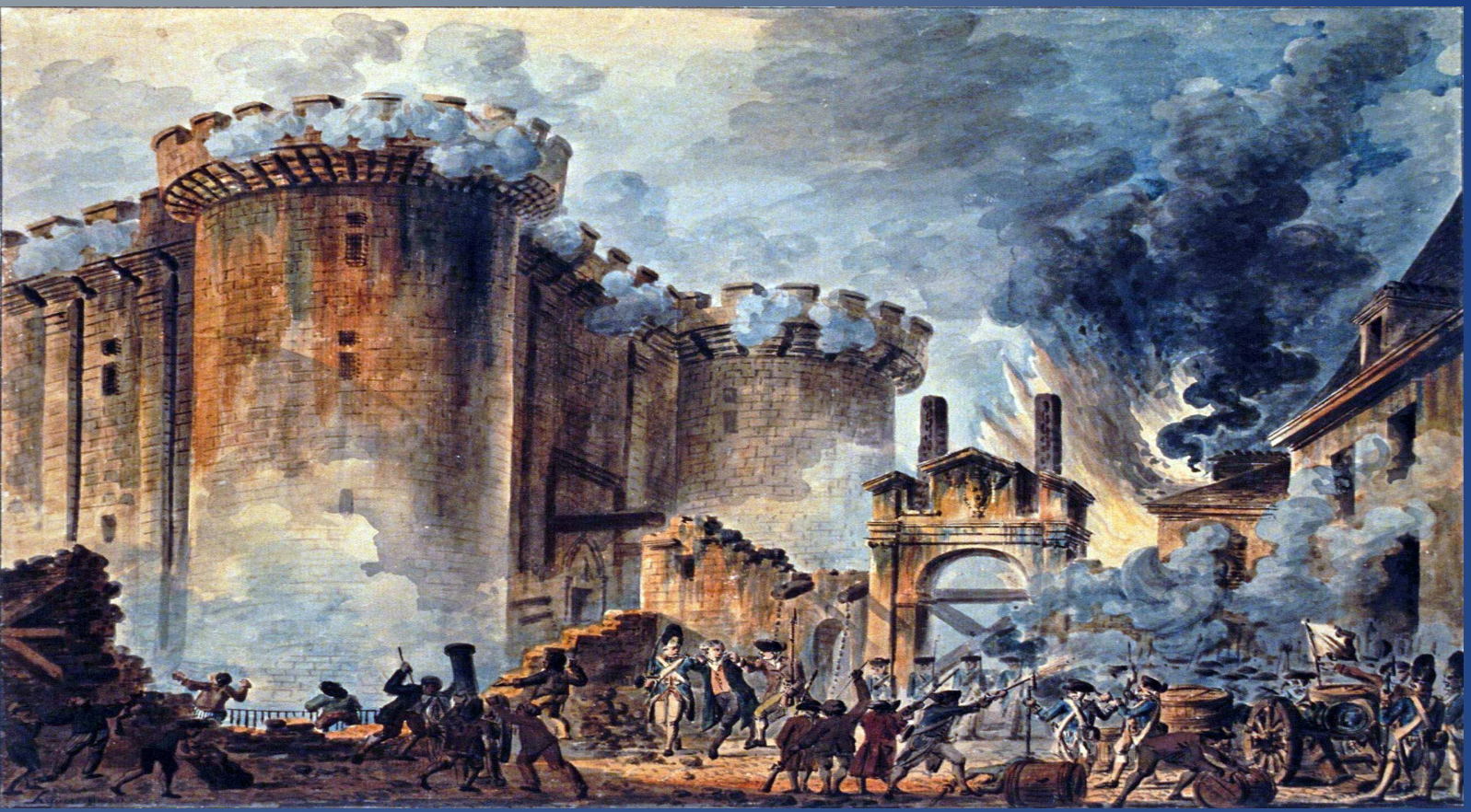
1.1. Informations- und Geldströme vor der digitalen Revolution



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?



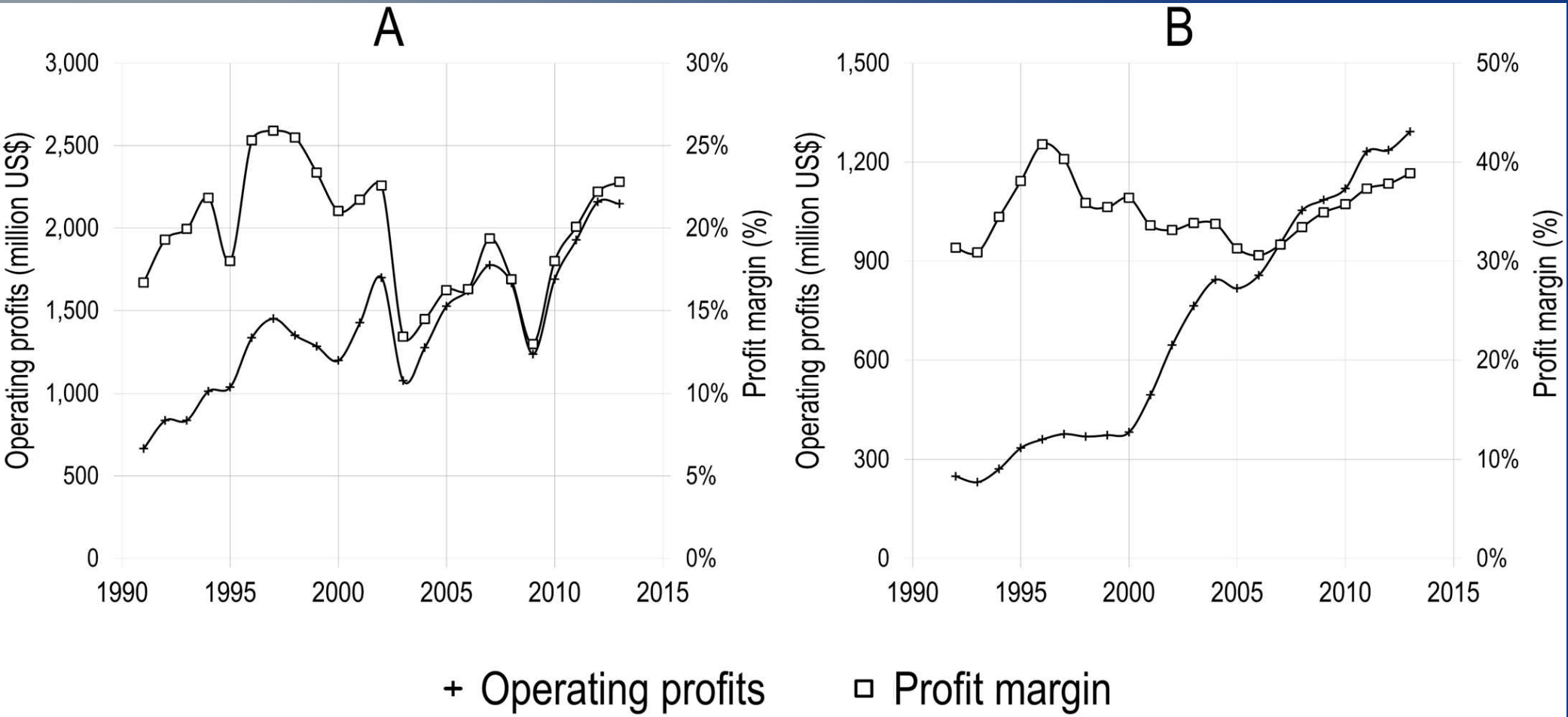
- Die Revolution ist gerade im vollen Gang!!!

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 (Divison of ACS)
  - Reed-Elsevier
  - Thomson Scientific & IP
- Konfrontiert mit der Flut an Daten!
- Kostenintensive Aufarbeitungsprozesse
- Unersetzlich?

3.3. Greedy Publishers?

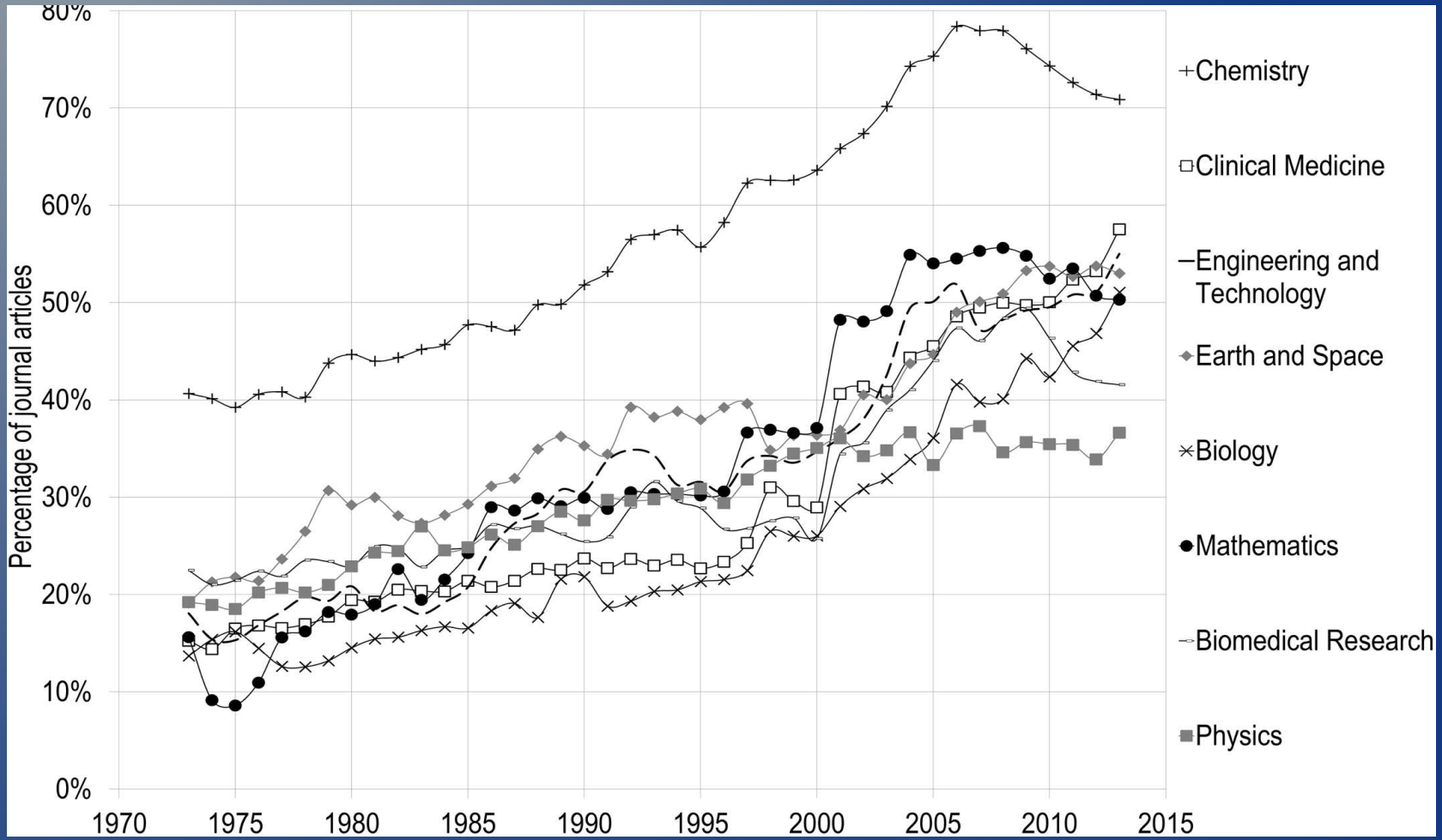
- Thomson Reuters Science & IP wurde am 11.7.2016 verkauft
- Waren 32% Gewinn zu wenig...?
- Jetzt Eigentümer: Onex Corp. und Baring Private Equity Asia
- 3,5 Mrd. US $

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

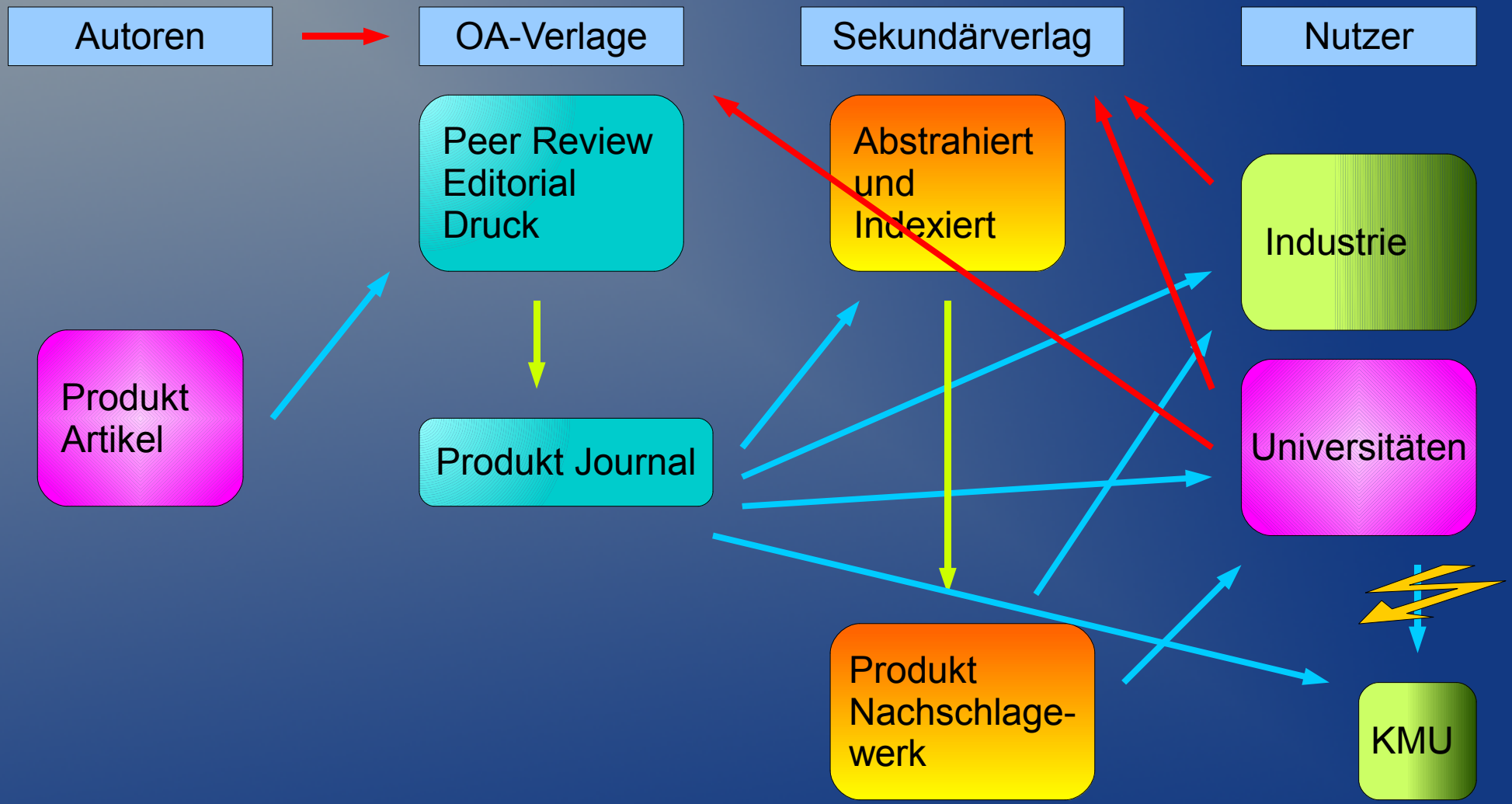
4. Open Access

- 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

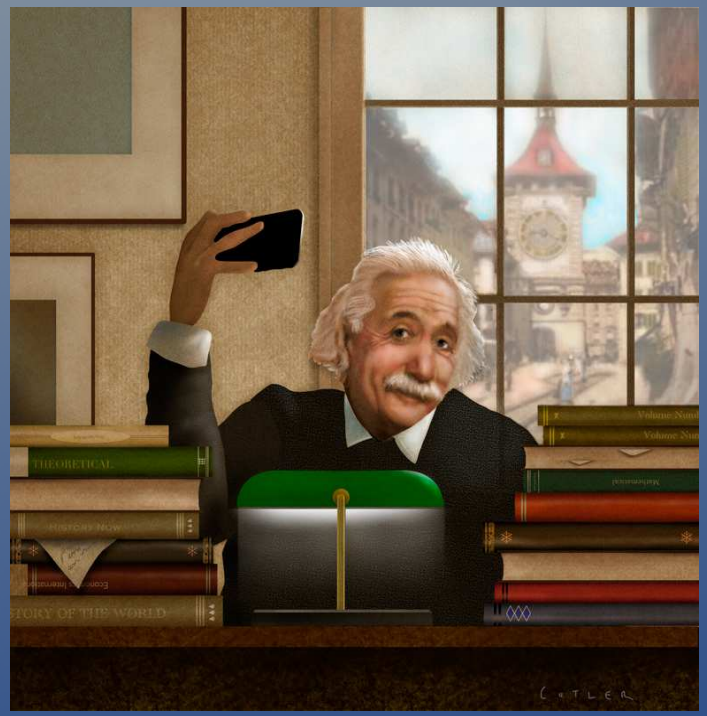
- Umkehr der Geldströme
  - Autor bezahlt – Industrie nutzt?
  - Der Autor ist der Geldgeber – Qualität?
- Kein Vertrag zwischen Nutzer und OA Verlag
  - Beständigkeit und Qualität?
- Predatory Journals
  - Bealls List - https://scholarlyoa.com/publishers/
- Cleanup bei DOAJ
  - https://doajournals.wordpress.com/2016/05/09/doaj-to-remove-approximately-3300-journ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

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

„Science in the Age of Selfies“
PNAS | August 23, 2016 | vol. 113 | no. 34 www.pnas.org/cgi/doi/10.1073/pnas.1609793113

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!
- Sirenenserver: Google? IBM? CAS? Elsevier?

- Diskussion nach dem Vortrag erwünscht!

8. Gamechangers?

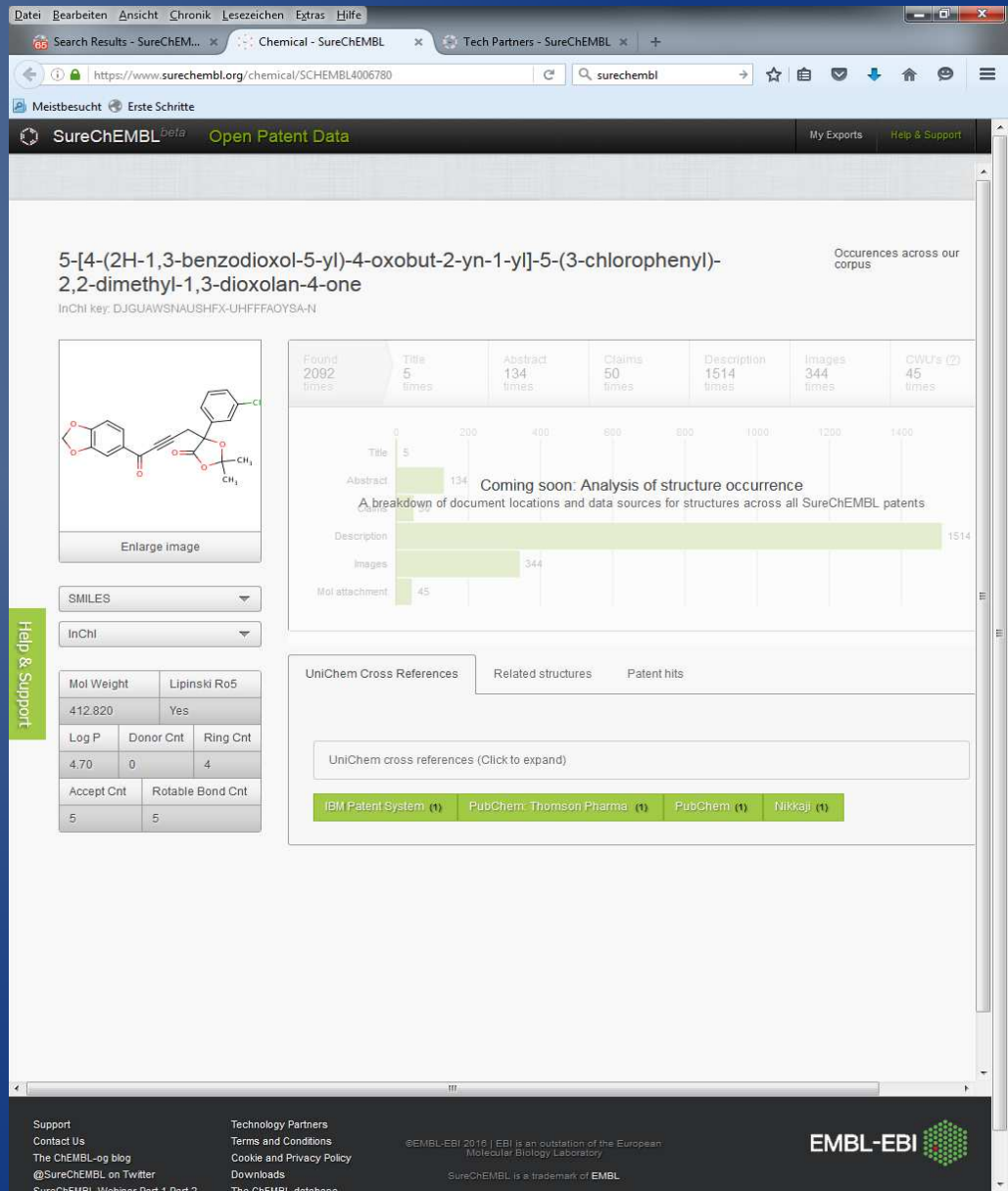
- IBM Watson
  - Zusammenarbeit mit NIH
- Google CPC Indexing für NPL
  - Immer für Überraschungen gut
- Text- und Datamining (2 Beispiele)
  - Minesoft
  - Infoapps
- Es gilt: Neue Tools kritisch evaluieren!

8.1. IBM Watson

- Sicher die grösste Investion 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/

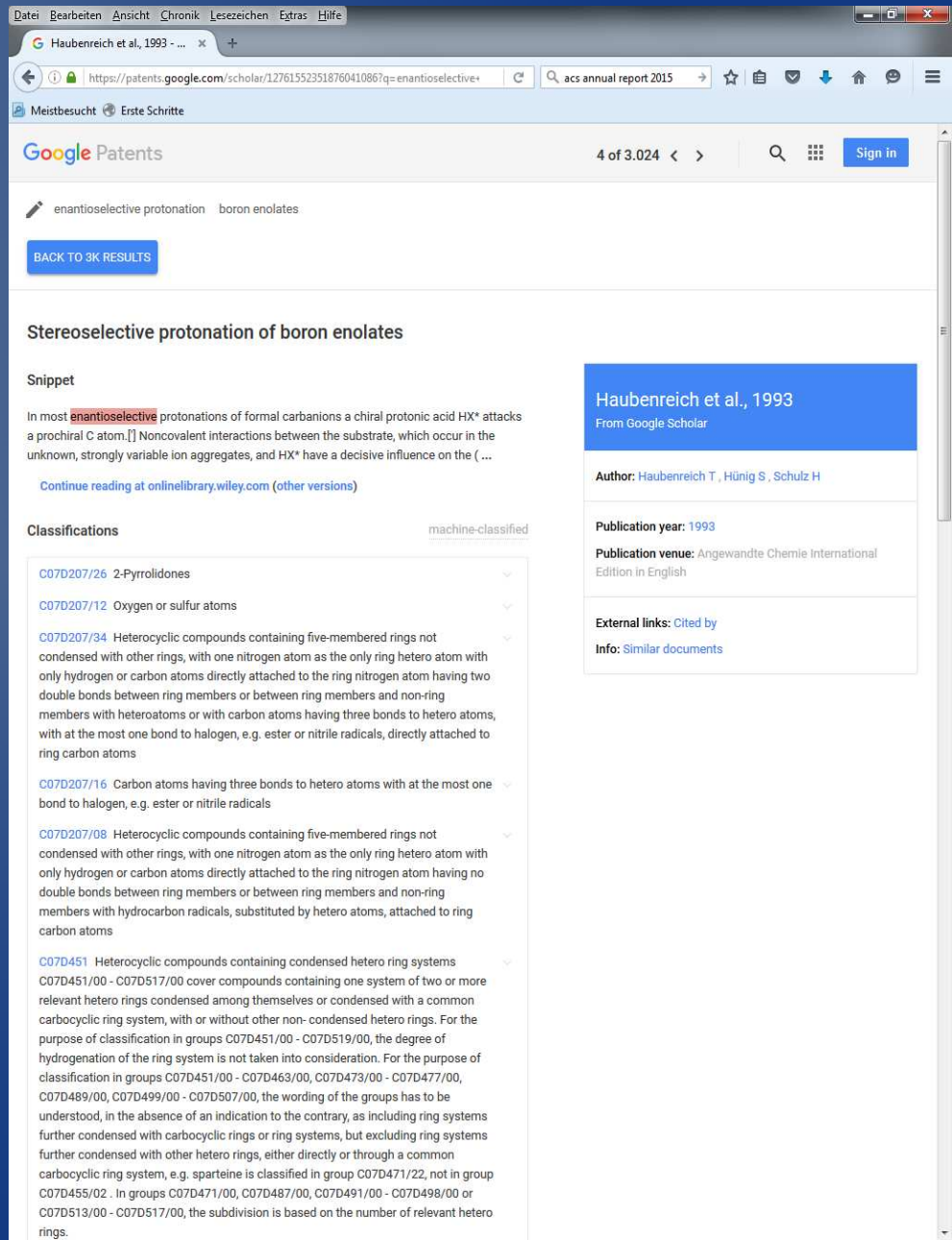
8.1. IBM - SureChEMBL

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



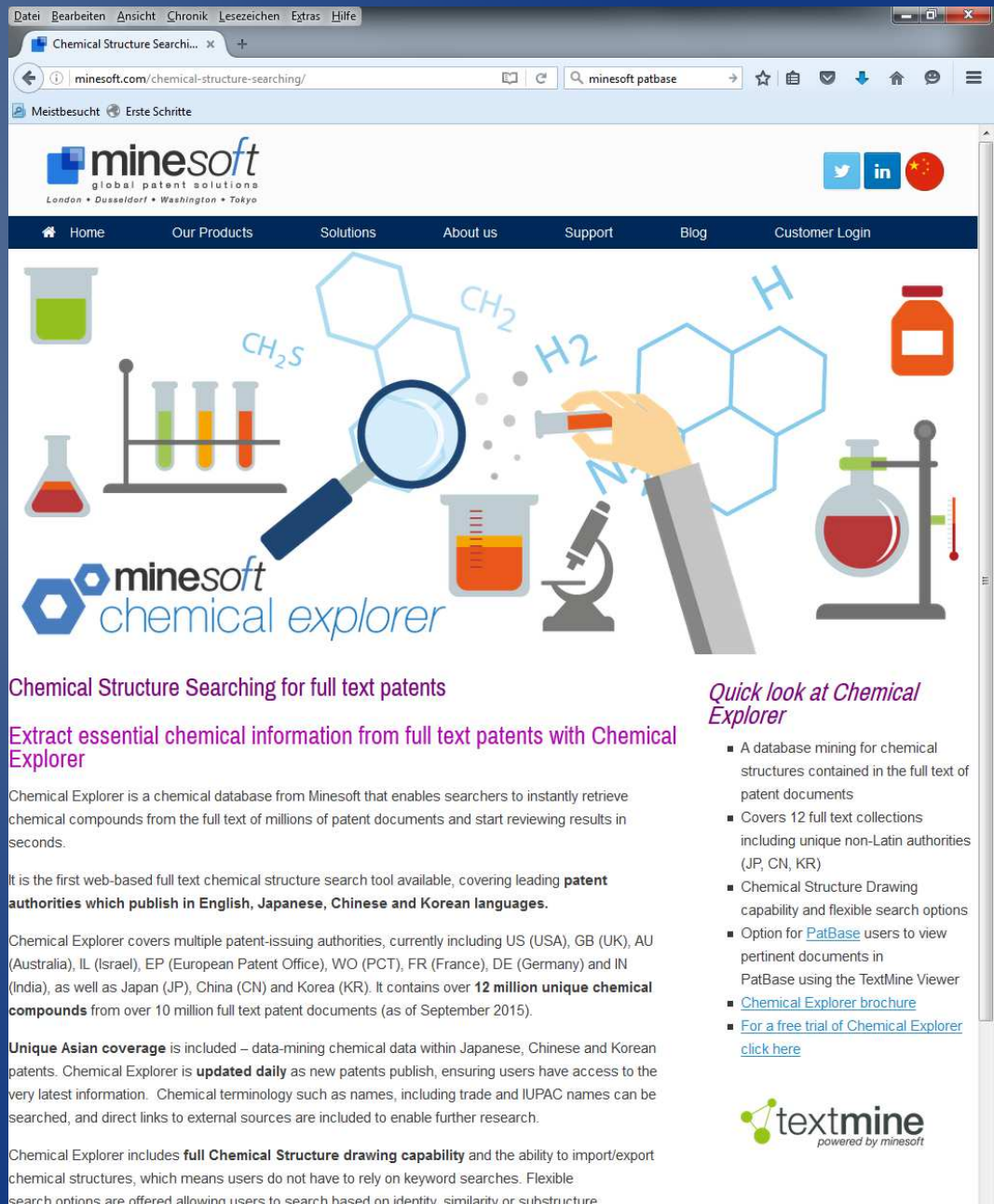
8.2. Google CPC Indexing für NPL

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



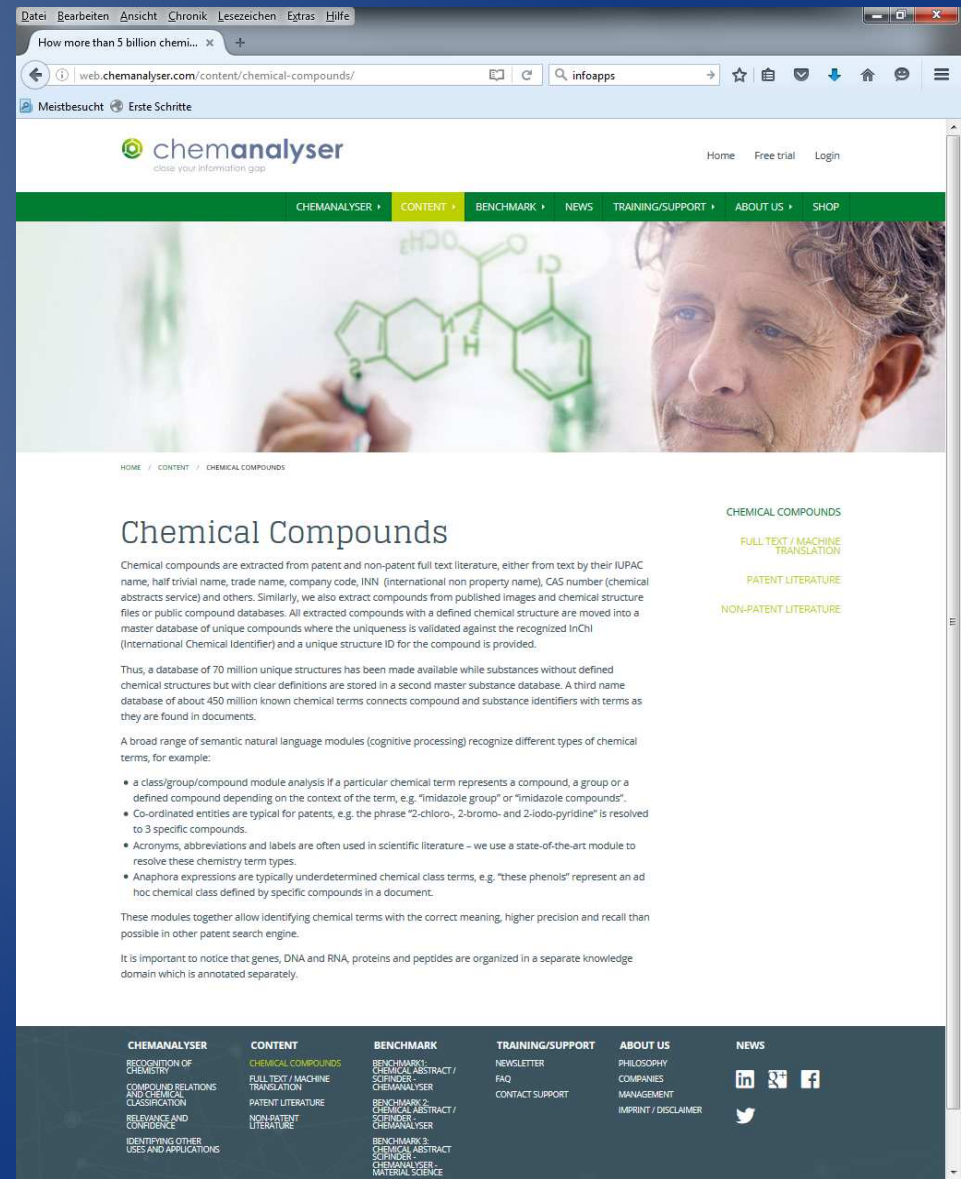
8.3. Minesoft Chemical Explorer

- Minesoft bietet chemische Informationen
  - Nur in Patenten (Volltexte frei!)



8.4. Infoapps chemanalyser

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



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.