

# Unter Beobachtung

## Datentracking in wissenschaftlichen Infrastrukturen

**Dr. Renke Siems**

**Ringvorlesung "Bibs & Bits - Digitale Transformation in Bibliotheken"**

**16. Januar 2024**



**Baden-Württemberg**

MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST



<https://twitter.com/swibcon/status/1465326475247337478>



# Take Home Messages I

- Nicht rechtskonformer Zustand bei vielen Lizenzverträgen mit möglicher Haftungsgemeinschaft von Verlag, Bibliothek und ggf. Förderer
- Ausdehnung dieser Situation auf den gesamten Research Life Cycle
- Verletzung der Wissenschaftsautonomie und Gefährdung von Forschenden und Studierenden
- Verletzung von Zivilklauseln
- Gefahr von Wirtschaftsspionage und geopolitischen Nachteilen



# Take Home Messages II

- Digitale Souveränität macht auf europäischer Ebene Fortschritte:
  - Reform der Forschungsbewertung
  - Umbau der Publikationsinfrastrukturen
- Modell der Transformationsverträge erschöpft sich (→ CoalitionS)
- Forschungsbewertung wird von DFG, WR, KMK verändert
- NFDI wird ausgebaut (Basisdienste)
- Ein echtes Open Science Ökosystem rückt näher



# Turn Your Disparate Data into Solid Gold

---

*A (nearly) complete guide to data &  
**customer data platforms** for  
scholarly publishers*

Quelle: Hum White Paper Data Analytics Publisher



# Tracking

Ausgang: Verlage drängten auf veränderte Authentifizierung

*Publishers don't need personally identifiable information from RA21 to be able to identify library users. (Todd Carpenter)*



Die initiale Untersuchung von Cody Hanson ergab 139 Tracker auf 15 Verlagsplattformen...

<https://www.codyh.com/writing/tracking.html>

<https://twitter.com/codyh/status/1199126449925419008/photo/1>



# Tracking

Gefunden wurden

- „einfache“ Tracker (Cookies)
- Audience Tools (aggregieren Trackingdaten aus vielen Quellen und ermöglichen das Targeting)
- Fingerprinter (Identifikation trotz Privacy Mode)
- Real Time Bidding-Instrumente
- weitere Webtechnologien



# First Party Data

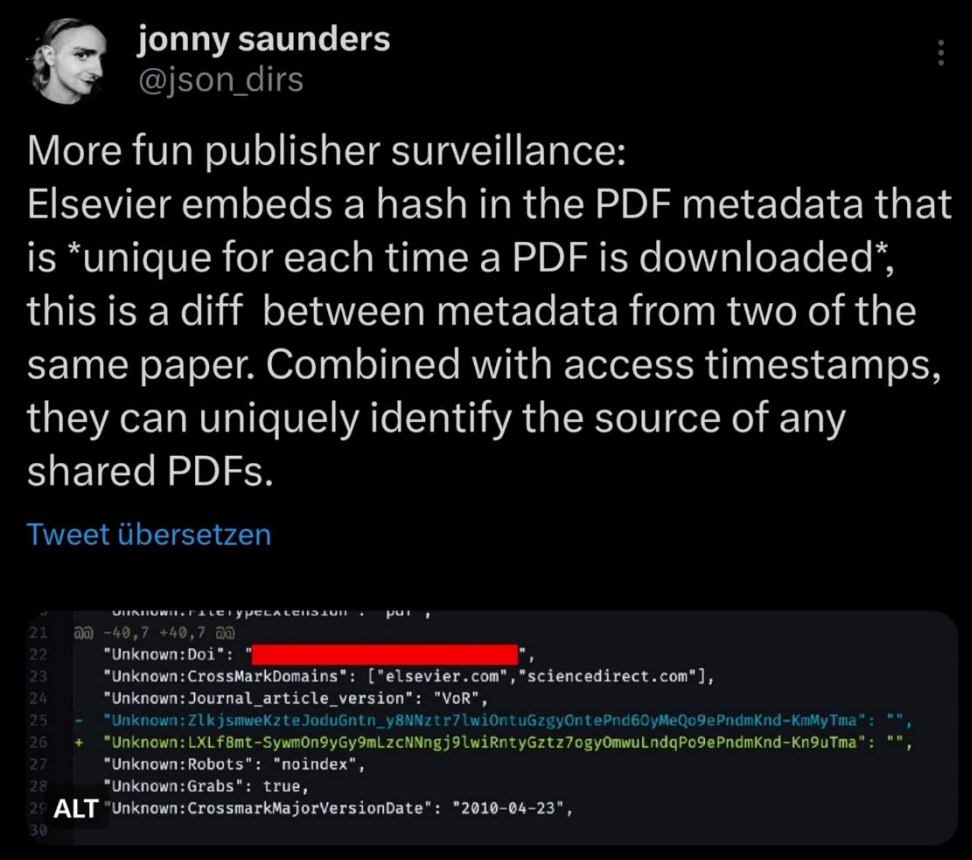


<https://twitter.com/codyh/status/1243250490403483648>





# First Party Data



jonny saunders  
@json\_dirs

More fun publisher surveillance:  
Elsevier embeds a hash in the PDF metadata that is \*unique for each time a PDF is downloaded\*, this is a diff between metadata from two of the same paper. Combined with access timestamps, they can uniquely identify the source of any shared PDFs.

[Tweet übersetzen](#)

```
21   "Unknown:FileTypeExtension": "pdf",
22   "Unknown:Doi": "XXXXXXXXXXXX",
23   "Unknown:CrossMarkDomains": ["elsevier.com", "sciencedirect.com"],
24   "Unknown:Journal_article_version": "VoR",
25   "Unknown:ZlkjSmweKzteJoduGntn_y8NNztr7lwiOntuGzgyOntepNd60yMeQo9ePndmKnd-KmMyTma": "",
26   "Unknown:LXlfBmt-SywmOn9yGy9mLzcNNngj9lwiRntyGtz7ogy0mwuLndqPo9ePndmKnd-Kn9uTma": "",
27   "Unknown:Robots": "noindex",
28   "Unknown:Grabs": true,
29   ALT "Unknown:CrossmarkMajorVersionDate": "2010-04-23",
30
```

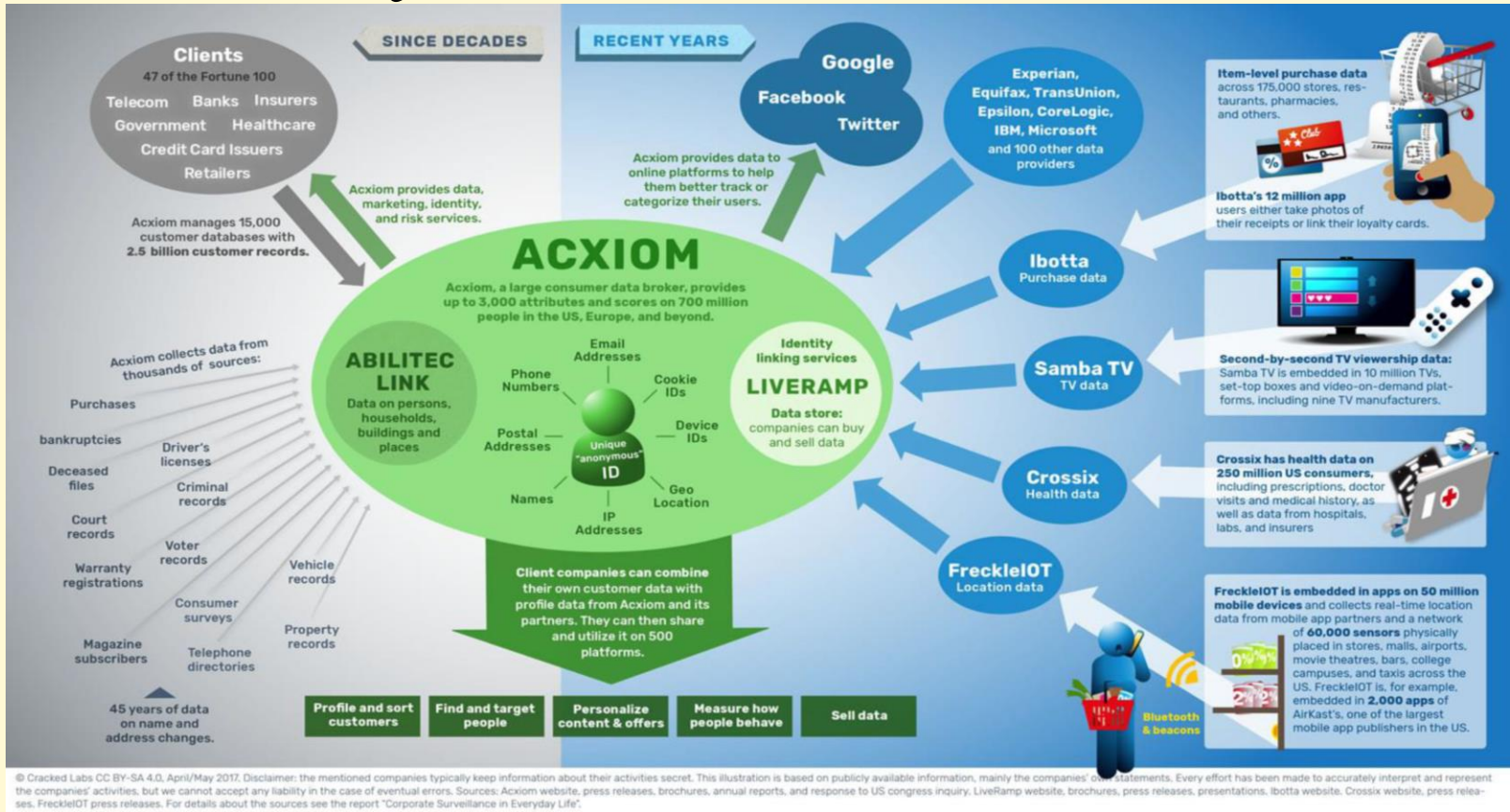
[https://twitter.com/json\\_dirs/status/1486120144141123584](https://twitter.com/json_dirs/status/1486120144141123584)



Baden-Württemberg

MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST

# Third Party Data

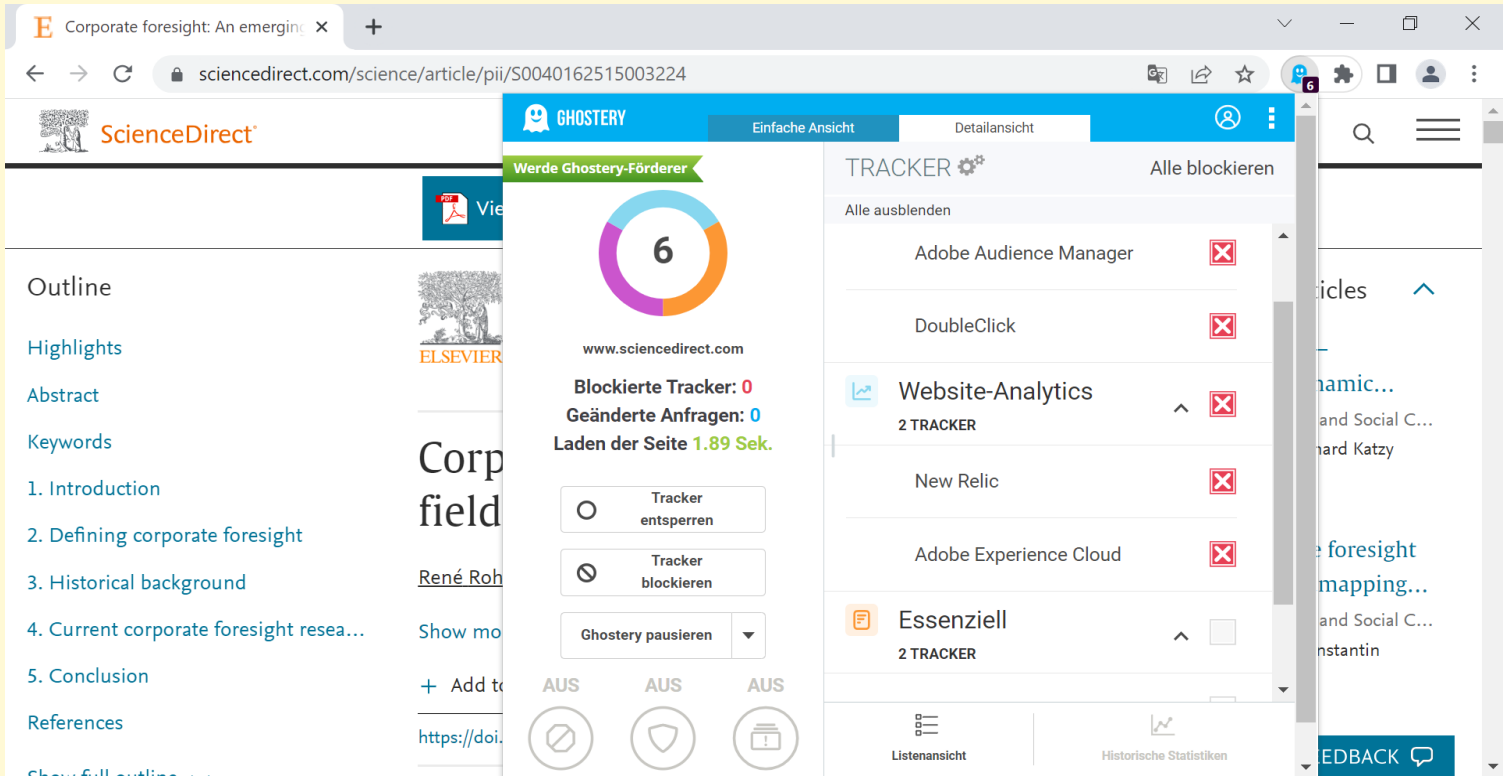


© Cracked Labs CC BY-SA 4.0, April/May 2017, Disclaimer: the mentioned companies typically keep information about their activities secret. This illustration is based on publicly available information, mainly the companies' own statements. Every effort has been made to accurately interpret and represent the companies' activities, but we cannot accept any liability in the case of eventual errors. Sources: Acxiom website, press releases, brochures, annual reports, and response to US congress inquiry, LiveRamp website, brochures, press releases, presentations, Ibotta website, Crossix website, press releases, FreckleIoT press releases. For details about the sources see the report "Corporate Surveillance in Everyday Life".

[https://crackedlabs.org/dl/CrackedLabs\\_Christl\\_CorporateSurveillance.pdf](https://crackedlabs.org/dl/CrackedLabs_Christl_CorporateSurveillance.pdf) (S. 55)



# Third Party Data



Screenshot Science Direct, 2. Juli 2023 (kein Cookie-Banner)

# Third Party Data

The screenshot displays a web browser window with the URL `nature.com/articles/s41586-020-2591-3`. The page content includes the Springer Nature logo and the article title "The causes of sea-level rise". Overlaid on the page are two browser extensions: Ghostery and Tracker.

**Ghostery Extension:**

- Header: GHOSTERY
- Buttons: Einfache Ansicht, Detailansicht
- Message: Werde Ghostery-Förderer
- Visual: A circular progress indicator showing 37 blocked trackers.
- Text: `www.nature.com`, `Blockierte Tracker: 0`, `Geänderte Anfragen: 0`, `Laden der Seite 4.07 Sek.`
- Controls: `Tracker entsperren`, `Tracker blockieren`, `Ghostery pausieren`

**Tracker Extension:**

- Header: TRACKER
- Buttons: Alle blockieren, Alle ausblenden
- Table of trackers:

Tracker Name	Status
AppNexus	Blocked (Red X)
Permutive	Blocked (Red X)
Facebook Custom Audience	Blocked (Red X)
Bidswitch	Blocked (Red X)
Media.net	Blocked (Red X)
Rubicon	Blocked (Red X)
ShareThrough	Blocked (Red X)

Bottom navigation: Listenansicht, Historische Statistiken

Screenshot Nature, 2. Juli 2023



# Bid Streaming

```
"id": "1234",
"name": "Awesome Example Site",
"domain": "examplesitedomain.com",
"mobile": 1,
"amp": 0,
"pub": {
  "id": "9876",
  "name": "Example Publisher, Inc.",
  "domain": "examplepubdomain.com"
},
"user": {
  "id": "a0af45c77890045deec100acb8443baff57c",
  "buyerid": "fcd4282456238256034abcdef220d9aa5892",
  "yob": 1990,
  "gender": "F"
},
"device": {
  "type": 4,
  "ifa": "8846d6fa10008bceaaf322908dfcb221",
  "ip": "1.2.3.4",
  "ua": "...user agent string...",
  "make": "Apple",
  "model": "iPhone",
  "hwv": "6s",
  "os": 13,
  "osv": "11.4.1",
  "mccmnc": "310-005",
  "geo": {
    "type": 1,
    "lat": 42.3601,
    "lon": 71.0581,
    "country": "USA",
    "timezone": "-0500"
  }
}
```

The website this specific person is currently viewing

Various ID codes that identify this specific person, and can tie them to existing profiles

Distinctive characteristics of this specific person

Distinctive information about this specific person's device

This specific person's IP address

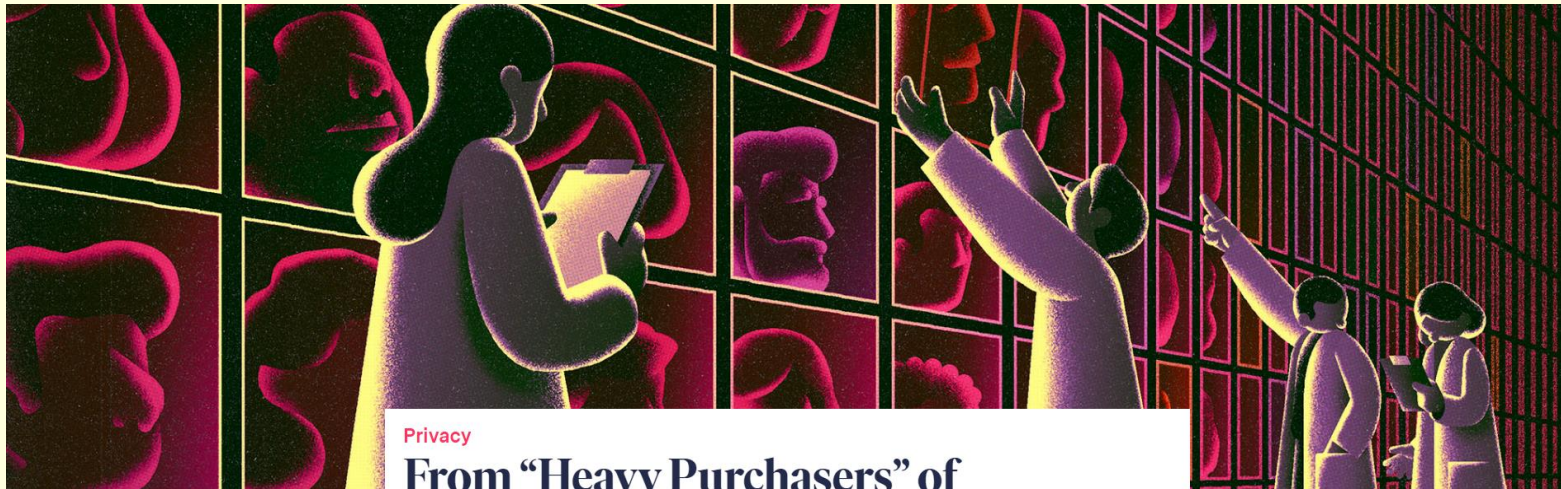
Distinctive information about this specific person's device

This young woman's GPS coordinates!

<https://www.slideshare.net/JohnnyRyan/briefing-on-adtech-rtb-and-the-gdpr-at-dmexco-brave-event>



# Bid Streaming



June 8, 2023 06:00 ET

Privacy

## From “Heavy Purchasers” of Pregnancy Tests to the Depression-Prone: We Found 650,000 Ways Advertisers Label You

A spreadsheet on ad platform Xandr’s website revealed a massive collection of “audience segments” used to target consumers based on

Adrián Astorgano

<https://themarkup.org/privacy/2023/06/08/from-heavy-purchasers-of-pregnancy-tests-to-the-depression-prone-we-found-650000-ways-advertisers-label-you>



# Bid Streaming

The screenshot shows an Excel spreadsheet with the following data rows (ID | Description):

ID	Description
545777	LiveRamp Data Store,8082,23584509,LexisNexis > Stirista Fusion > Number of Household Members > 6
545778	LiveRamp Data Store,8082,23584518,LexisNexis > Stirista Fusion > Number of Household Members > 3
545779	LiveRamp Data Store,8082,23584529,LexisNexis > Stirista Fusion > Financial > Total AVM Real Estate Value Of All Properties > \$900k-\$1M
545780	LiveRamp Data Store,8082,23584530,LexisNexis > Stirista Fusion > Financial > Original Mortgage Amount > \$200k-\$225k
545781	LiveRamp Data Store,8082,23584531,LexisNexis > Stirista Fusion > Demographic > Individual Income > \$60k-\$80k
545782	LiveRamp Data Store,8082,23584533,LexisNexis > Stirista Fusion > Number of Business Registrations at Current Residence > 4
545783	LiveRamp Data Store,8082,23584535,LexisNexis > Stirista Fusion > Financial > Number of Personal Properties Owned > 3
545784	LiveRamp Data Store,8082,23584537,LexisNexis > Stirista Fusion > Professional License Type > Pharmacists Engineers Scientists Lawyers and Architects
545785	LiveRamp Data Store,8082,23584559,LexisNexis > Stirista Fusion > Financial > Current AVM Property Value > \$125k-\$150k
545786	LiveRamp Data Store,8082,23584566,LexisNexis > Stirista Fusion > Financial > Number of Personal Properties Owned > 0
545787	LiveRamp Data Store,8082,23584568,LexisNexis > Stirista Fusion > Education > Attended Private College
545788	LiveRamp Data Store,8082,23584569,LexisNexis > Stirista Fusion > Demographic > Gender > Female
545789	LiveRamp Data Store,8082,23584571,LexisNexis > Stirista Fusion > Financial > Total Tax Assessed Real Estate Value Of All Properties > \$1M-\$2M
545790	LiveRamp Data Store,8082,23584583,LexisNexis > Stirista Fusion > Home Value Compared to Block Median > Home Value is 50-60% Below Block Median
545791	LiveRamp Data Store,8082,23584584,LexisNexis > Stirista Fusion > Demographic > Individual Income > \$50k-\$60k
545792	LiveRamp Data Store,8082,23584595,LexisNexis > Stirista Fusion > Home Value Compared to Block Median > Home Value is 30-35% Above Block Median
545793	LiveRamp Data Store,8082,23584596,LexisNexis > Stirista Fusion > Home Value Compared to Block Median > Home Value is 5-10% Below Block Median
545794	LiveRamp Data Store,8082,23584598,LexisNexis > Stirista Fusion > Demographic > Age > 33
545795	LiveRamp Data Store,8082,23584599,LexisNexis > Stirista Fusion > Demographic > Age > 114
545796	LiveRamp Data Store,8082,23584601,LexisNexis > Stirista Fusion > Home Value Compared to Block Median > Home Value is 15-20% Above Block Median
545797	LiveRamp Data Store,8082,23584602,LexisNexis > Stirista Fusion > Last 5 Year Residence Changes Economic Trajectory > Moved To Another Single Family Home
545798	LiveRamp Data Store,8082,23584603,LexisNexis > Stirista Fusion > Financial > Original Mortgage Amount > \$225k-\$250k
545799	LiveRamp Data Store,8082,23584625,LexisNexis > Stirista Fusion > Financial > Current AVM Property Value > \$400k-\$450k

<https://github.com/the-markup/xandr-audience-segments>



# Zwischenstand

- Das Informationsverhalten von Hochschulangehörigen ist nicht privat, sondern wird personalisiert beobachtet
- Die Daten bleiben nicht bei den Verlagen, sondern werden auch von Data Brokern aggregiert und verwertet
- Es entstehen durchgängige Onlinebiographien: was eine Wissenschaftlerin recherchiert, verbindet sich mit
  - Social Media
  - Shopping
  - ihre Joggingstrecke
  - ob sie Kinder hat – etc.







## Programm der 15. Göttinger Urheber- und Datenschutzrechtstagung

**“Webtracking durch Wissenschaftsverlage – eine Spurensuche: Was wird genutzt, was darf man und was kann/muss/sollte man dagegen tun?”**

*Prof. Dr.-Ing. Jana Dittmann & Dr.-Ing. Robert Altschaffel  
Institut für Technische und Betriebliche Informationssysteme (ITI), Universität  
Magdeburg*

*Prof. Dr. Michael Beurskens, LL.M. (Chicago)  
Lehrstuhl für Bürgerliches Recht, Deutsches, Europäisches und Internationales  
Wirtschaftsrecht, Universität Passau*

# Aussagen Urheberrechtstagung

- Plattformen der Verlage sind genauso trackerverseucht wie deren mobile Apps, Mails und Newsletter
- Die meisten Cookies / Tracker sind nicht gerechtfertigt
- Nicht rechtskonforme Praxis trotz Hinweisen
- Datenschutz im DEAL-Vertrag wahrscheinlich nicht ausreichend
- Mitverantwortung Einrichtungen nach Art. 26 DSGVO möglich
- Verletzung Fürsorgepflicht gegenüber Beschäftigten



# »The Next Battleground«

Science

Current Issue

First release papers

Archive

About ▾

Submit manuscript

HOME > SCIENCE > VOL. 368, NO. 6491 > IN PURSUIT OF OPEN SCIENCE, OPEN ACCESS IS NOT ENOUGH

🔒 | POLICY FORUM | SCIENTIFIC PUBLISHING

## In pursuit of open science, open access is not enough

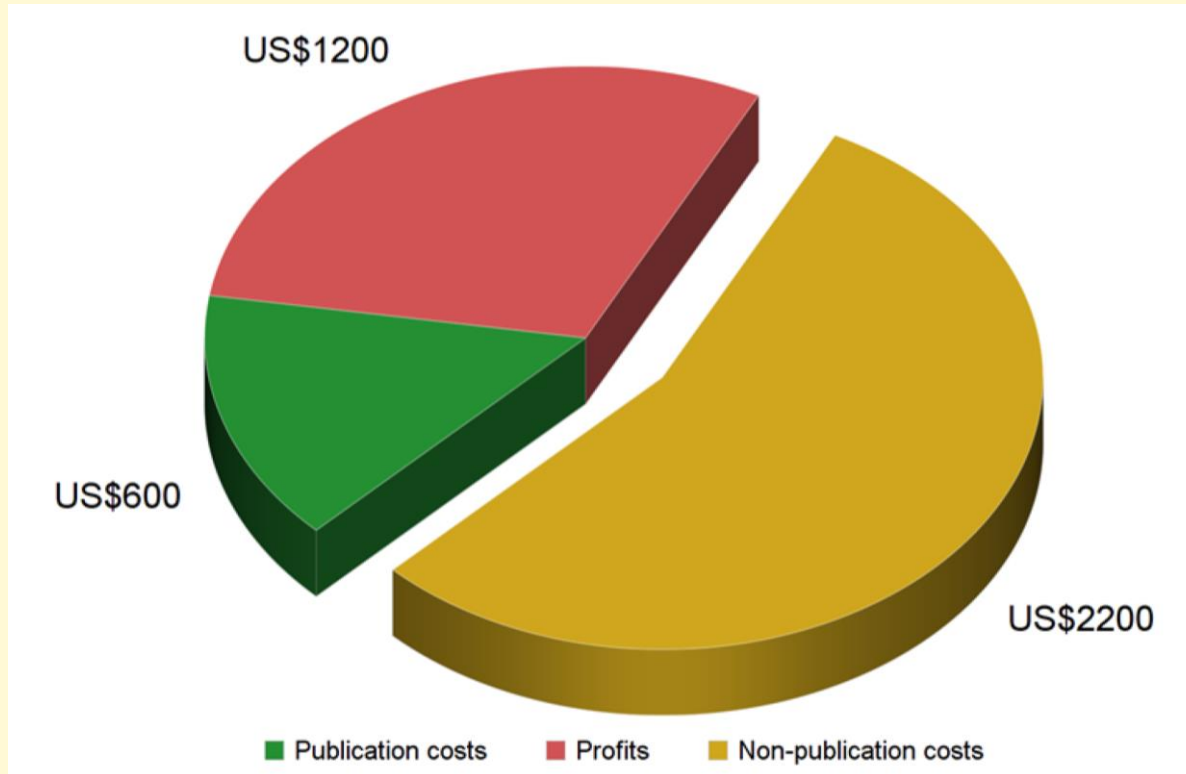
Preventing monopolies in knowledge infrastructure is the next battleground for publishers and research institutions

CLAUDIO ASPESI AND AMY BRAND [Authors Info & Affiliations](#)

SCIENCE • 8 May 2020 • Vol 368, Issue 6491 • pp. 574-577 • DOI: 10.1126/science.aba3763



# Workbenches



Alexander Grossmann, Björn Brembs: Current market rates for scholarly publishing services; DOI: [10.12688/f1000research.27468.1](https://doi.org/10.12688/f1000research.27468.1)

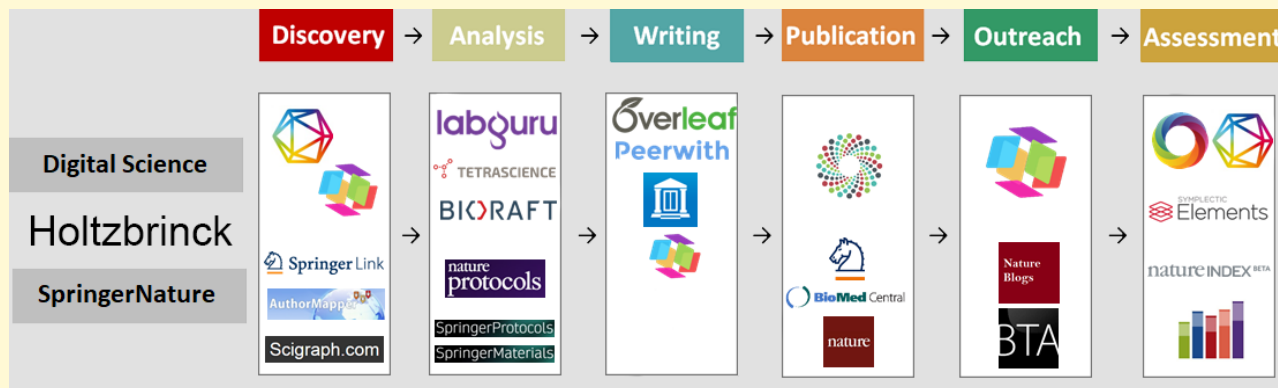


Baden-Württemberg

MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST

# Workbenches

- Die vielen Zukäufe bilden zunehmend datengetriebene Workbenches über alle Forschungsbereiche hinweg
- Produkte werden damit beworben, Einblick in den gesamten Forschungszyklus bieten zu können



Grafik: Björn Brembs auf Basis von 101innovations; <https://stoptrackingscience.eu/background-information/> (Auszug)

# Bibliothekssysteme

EBSCOpost


Librarianship

News

Workflow

Technology

Library Resources

 Search blog

## The Evolving Landscape of Library Data Analytics

Where is the evolution of library data going? And how is your library adapting?

<https://www.ebsco.com/blogs/ebscopost/evolving-landscape-library-data-analytics>



# Bibliothekssysteme



The screenshot shows the OCLC website interface. At the top left is the OCLC logo. To the right is a search bar with the text "OCLC-Webseiten suchen" and a magnifying glass icon. Below the search bar is a navigation menu with items: "Mitgliedschaft", "Produkte", "Support", "Research", "Events", and "Über". On the right side of the menu are "Kontakt" and "Log in für Services". The main content area features the text "University of Gloucestershire" in green, followed by the headline "Erfahren Sie, welche Bibliotheksressourcen den größten Einfluss haben".

Die Integration des Datenfeeds in das Studierendenaktensystem der Universität ermöglicht weitere Einblicke. ... „Über die Benutzer-ID können wir jetzt auf das Studienprogramm und den Fachbereich des Studierenden zugreifen“ erklärt Lee. „Auf diese Weise können wir beispielsweise erkennen, wie viele Studierende eines bestimmten Programms auf ScienceDirect zugegriffen haben. Unsere Daten aus Studierendenakten sind ziemlich umfangreich, daher können wir sogar die Nutzung von Bibliotheksressourcen in unterschiedlichen demografischen Gruppen vergleichen.“

„Was wir für die University of Gloucestershire getan haben, können wir auch anderen OCLC-Mitgliedern bieten. Wir haben Standardprozesse und Datenstrukturen eingerichtet, um WMS- und EZproxy-Daten sofort zu verarbeiten.“

**Lee Baylis**, Senior Analytics Innovator

<https://www.oclc.org/de/member-stories/gloucestershire.html>



# Neue Geschäftsfelder

☰ GEORGETOWN LAW  
Center on Privacy & Technology

POLICE  
ICE

## AMERICAN DRAGNET

DATA-DRIVEN DEPORTATION IN THE 21ST CENTURY

May 10, 2022

<https://americandragnet.org/>





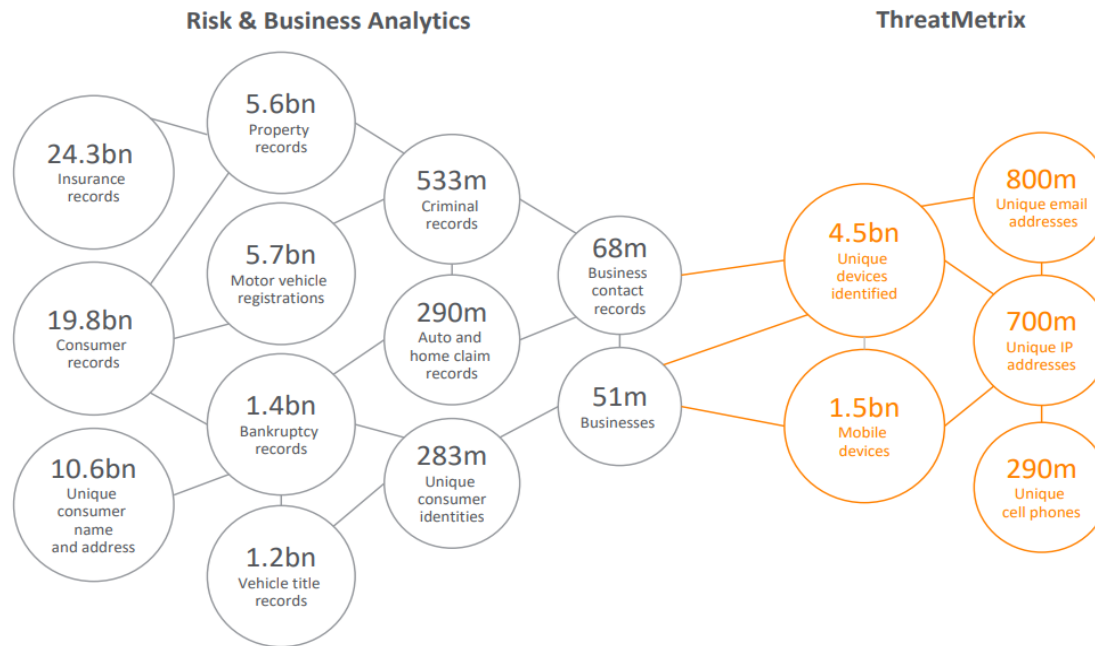
# Neue Geschäftsfelder

- Informationsanbieter wie Thomson Reuters und RELX sind ins Big Data Policing eingestiegen
- Aggregierte Daten-Produkte wie CLEAR und LexisNexis Risk Solutions werden an die Sicherheitsindustrie lizenziert
- Anbieter wie Clarivate agieren in ähnlich tracking-intensiven Bereichen wie fraud prevention
- Belege für die Verwendung von Nutzungsdaten wissenschaftsbezogener Produkte mehren sich



# Neue Geschäftsfelder

Combining physical and digital identity data assets positions us for strong future growth



RELX Group

40

<https://www.relx.com/~media/Files/R/RELX-Group/documents/presentations/risk-teach-in-8Nov18.pdf>



# Neue Geschäftsfelder

The screenshot shows the ScienceDirect website interface. At the top, there is a navigation bar with the ScienceDirect logo, 'Journals & Books' link, search icon, and 'Register' and 'Sign in' buttons. Below this is a banner for 'Access through your institution'. The main content area features the 'Arabian Journal of Chemistry' (Volume 12, Issue 7, November 2019, Pages 908-931) with a review article titled 'Nanoparticles: Properties, applications and toxicities'. A sidebar on the left contains an 'Outline' with links to 'Abstract', 'Graphical abstract', 'Keywords', 'Abbreviations', and a numbered list of sections. A 'Recommended articles' section on the right lists 'Introduction to Nanomaterials: Basic Properties...' and 'Nanomaterials Definitions, Classifications, and ...'. At the bottom, the browser's developer tools are open to the 'Network' tab, showing a list of requests to 'online-metrix.net'. The selected request is a GET request for 'clear1.png;ClS3SID=...' with a status of 204. The 'Request Headers' panel is expanded, showing headers such as 'Accept: image/webp,\*/\*', 'Accept-Encoding: gzip, deflate, br', 'Accept-Language: en-US,en;q=0.5', 'Connection: keep-alive', 'Cookie: thx\_global\_guid=75...', 'Host: h.online-metrix.net', and 'Referer: https://honline-metrix.net/Z/T/...'.

<https://twitter.com/WolfieChristl/status/1295655040741445632>



# Digitale Souveränität

„Die digitale Selbstbefähigung von Wissenschaftseinrichtungen muss gestärkt werden. Der Zugang zu digitalen Diensten, Infrastrukturen und Daten ist essentiell für die Qualität der Forschung und die Überprüfbarkeit wissenschaftlicher Ergebnisse. Der Schutz vor Cyberangriffen wird immer wichtiger“, so der WR-Vorsitzende, Wolfgang Wick. „Zugleich kann die Wissenschaft selbst mit ihrer Kompetenz und Kreativität den digitalen Fortschritt in Deutschland vorantreiben und mit Innovationen helfen, Alternativen zu den Angeboten globaler Konzerne zu schaffen“, betont Wick.

[https://www.wissenschaftsrat.de/SharedDocs/Pressemitteilungen/DE/PM\\_2023/PM\\_2323.html](https://www.wissenschaftsrat.de/SharedDocs/Pressemitteilungen/DE/PM_2023/PM_2323.html)



# Digitale Souveränität

- NFDI positives Beispiel wissenschaftsgetriebener digitaler Infrastrukturen
- Dateninfrastrukturen beginnen, Keimzelle eines öffentlichen Forschungsökosystems werden
- Zentral strategische Assets: Elektronische Laborbücher, Forschungsinformationssysteme, Bibliothekssysteme

## Future of Libraries is Open

FOLIO is a collaboration of libraries, developers and vendors building an open source library services platform. It is a modern platform that inspires innovation and delivers both new and traditional resource management functionality.

<https://folio.org/>



# Digitale Souveränität

Die größte Baustelle zur Zeit: Forschungsbewertung

**CoARA**

About Agreement ▾ Coalition ▾ News Resources Contact **Sign**

## Coalition for Advancing Research Assessment

Our vision is that the assessment of research, researchers and research organisations recognises the diverse outputs, practices and activities that maximise the quality and impact of research. This requires basing assessment primarily on qualitative judgement, for which peer review is central, supported by responsible use of quantitative indicators.


9

<https://coara.eu/>



# Digitale Souveränität

## Die allergrößte Baustelle zur Zeit: Diamond OA

 Council of the European Union

Brussels, 23 May 2023  
(OR. en)

9616/23

RECH 190  
EDUC 169  
PI 77  
DIGIT 96

**OUTCOME OF PROCEEDINGS**

---

From: General Secretariat of the Council  
On: 23 May 2023  
To: Delegations

---

No. prev. doc.: 8827/23

---

Subject: High-quality, transparent, open, trustworthy and equitable scholarly publishing  
- Council conclusions (approved on 23 May 2023)

---

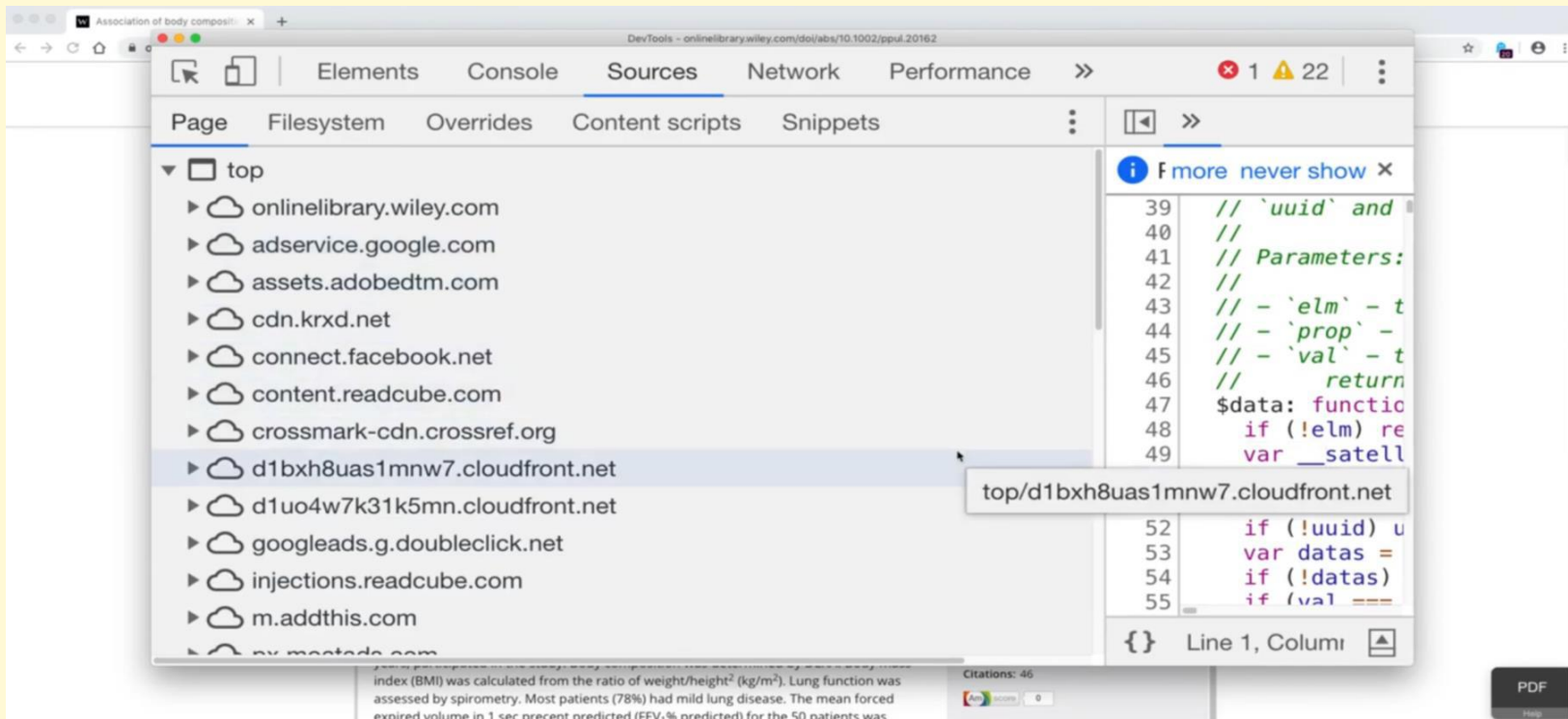
### The hazards of scholarly publishing

Scientific articles and other forms of scholarly publishing continue to be the primary means of disseminating research results and scientific findings. However, far from every article is available to other researchers or other interested readers. The costs of paywalls to access and publish articles are becoming unsustainable and the publication channels for researchers are frequently in the hands of private companies, which often control the intellectual property of the articles. The peer review of articles is essential to assure the quality control of the articles, but this process comes with challenges such as increased numbers of submissions and reviewers' fatigue. There is also an issue with predatory and questionable publishing practices.

<https://data.consilium.europa.eu/doc/document/ST-9616-2023-INIT/en/pdf>



# Vielen Dank für Ihr Interesse!



Screenshot <https://www.youtube.com/watch?v=uAzt-iJEkU&feature=youtu.be>





# Zusätzliche Nachweise

## **Initiale Untersuchungen:**

Eric Hellman: 16 of the top 20 Research Journals Let Ad Networks Spy on Their Readers;

<https://go-to-hellman.blogspot.com/2015/03/16-of-top-20-research-journals-let-ad.html>

Clifford Lynch: The rise of reading analytics and the emerging calculus of reader privacy in the digital world; <https://firstmonday.org/ojs/index.php/fm/article/view/7414>

Cody Hanson: User Tracking on Academic Publisher Platforms;

<https://www.codyh.com/writing/tracking.html>

## **Deutsche Forschungsgemeinschaft:**

Erkenntnisgeleitete Forschung stärken, von Wissensspeichern profitieren;

[https://www.dfg.de/service/presse/pressemitteilungen/2021/pressemitteilung\\_nr\\_13/index.html](https://www.dfg.de/service/presse/pressemitteilungen/2021/pressemitteilung_nr_13/index.html)

Datentracking in der Wissenschaft. Aggregation und Verwendung bzw. Verkauf von Nutzungsdaten durch Wissenschaftsverlage. Ein Informationspapier des Ausschusses für Wissenschaftliche Bibliotheken und Informationssysteme der Deutschen Forschungsgemeinschaft;

[https://www.dfg.de/download/pdf/foerderung/programme/lis/datentracking\\_papier\\_de.pdf](https://www.dfg.de/download/pdf/foerderung/programme/lis/datentracking_papier_de.pdf)

## **Weitere Literatur und Informationen unter:**

Stop Tracking Science; <https://stoptrackingscience.eu/>

