

From experimenter to lab rat – Science as datafied space

Coffee Lecture Lib4RI, 05. Mai 2026



Baden-Württemberg
Ministerium für Wissenschaft,
Forschung und Kunst

Agenda

A very short history of datafication

Science Tracking

Workbenches

Data Cartels

Research Security

Frankfurter Allgemeine

ZEITUNG FÜR DEUTSCHLAND

Überwachte Wissenschaftler

Auf einmal Laborratte

Die großen Verlage haben ein neues Geschäftsfeld entdeckt: die Überwachung von Wissenschaftlern und den Verkauf ihrer Daten. Sie nehmen dadurch weitreichenden Einfluss auf die Forschung. Die Europäische Union muss dringend eingreifen.



Björn Brembs, Konrad Förstner, Peter Kraker, Gerhard Lauer, Claudia Müller-Birn, Felix Schönbrodt, Renke Siems

Datafication



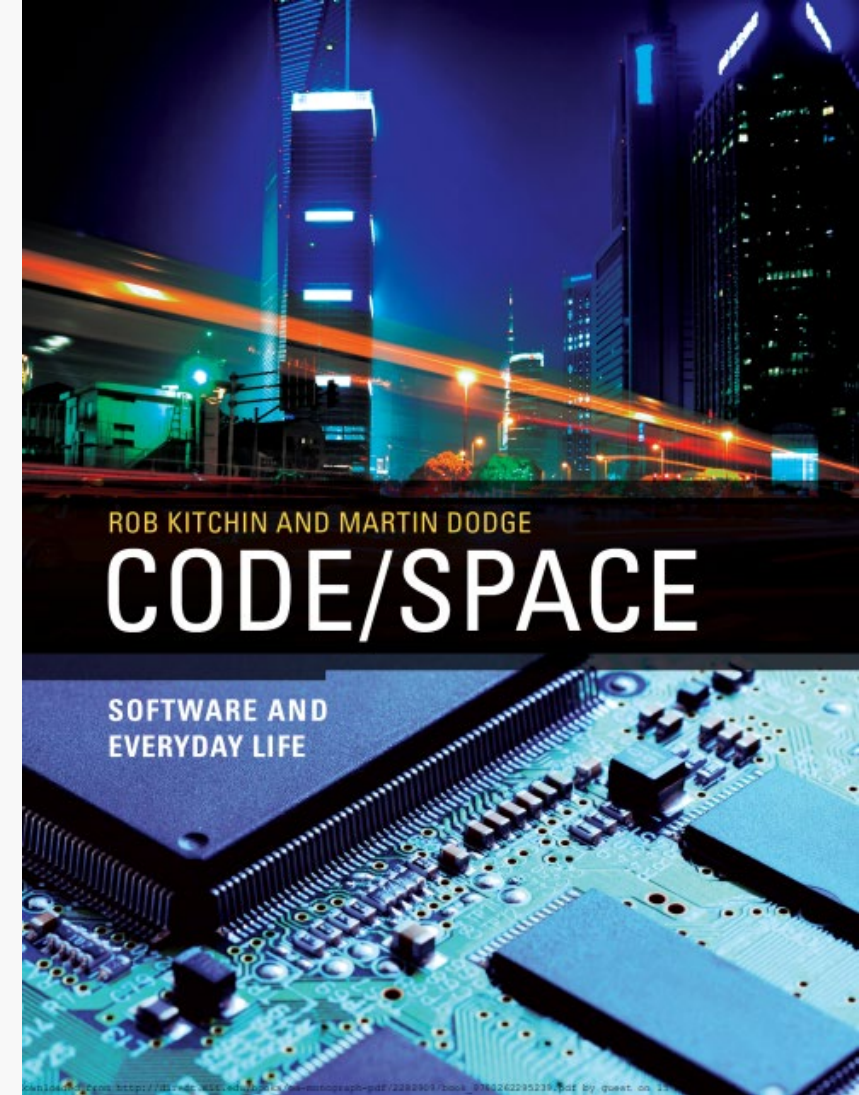
Baden-Württemberg
Ministerium für Wissenschaft,
Forschung und Kunst

Datafication

Digital transformation creates spaces shaped by software

This affects everyday life - such as travel -, creates prediction markets, and shapes information infrastructures

„Datafication is not the same as digitization, which takes analog content—books, films, photographs—and converts it into digital information, a sequence of ones and zeros that computers can read. Datafication is a far broader activity: taking all aspects of life and turning them into data. Google’s augmented-reality glasses datafy the gaze. Twitter datafies stray thoughts. LinkedIn datafies professional networks. Once we datafy things, we can transform their purpose and turn the information into new forms of value.“ (Cukier / Mayer-Schoenberger)



Datafication

Military use case as an extreme example of datafication:

- The rise of platform capitalism in warfare: Corporations such as Alphabet, Microsoft, and Amazon are helping to build the “Kill Cloud”
- Palantir generates AI-driven target selection systems by merging data sources (satellite imagery, biometric data, ICT data, the web)
- Individuals are selected based on scoring values, not factual relationships: “We kill people based on metadata.” (Michael Hayden)

„Palantir is here to disrupt and make all the institutions we partner with the very best in the world. And when it’s necessary to scare enemies, and on occasion kill them. And we hope you are in favor of that. We hope you’re enjoying being a partner, and we’re really happy and very very focused on what we’re doing.“ (Alex Karp)



Datafication

- Today, we all find the infrastructure of the Kill Cloud in our offices
- Data-driven border technologies are moving inland and into cities
- Relevant providers like Palantir also rely on mission-critical partners from the academic world: LexisNexis (RELX) and Thomson Reuters
- Both companies diversified early on as data brokers to become less dependent on the digital transformation of academia



Datafication

- The Big 5 publishers, in particular, have equipped their platforms with data analytics and ad tech tools
- This creates customizable snapshots of the academic landscape across the entire research cycle
- The data collected does not remain with the service providers but flows through the participating data brokers
- It is impossible to track the data's whereabouts



Data tracking in research: aggregation and use or sale of usage data by academic publishers

A briefing paper of the Committee on Scientific Library Services and Information Systems of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)

28 October 2021

Science Tracking



Baden-Württemberg
Ministerium für Wissenschaft,
Forschung und Kunst

Science Tracking

“The reason I was interested in third-party assets being loaded on these sites is that any JavaScript loaded on these pages has access to the entire DOM, or document object model, meaning it can read the address and contents of the page. It also has access to every user action that happens on that page, and can itself load additional scripts from additional sources. So when, for example, a publisher puts JavaScript from Google on its pages, Google can record any information from the page about the article being sought, or search terms from a library user in the publisher platform. Fourteen of the fifteen publisher platforms included Google code on the article page.”

Cody Hanson



Science Tracking

Cody Hanson found

- Simple trackers („cookies“)
- Audience tools (aggregate tracking data from multiple sources and enable micro-targeting)
- Fingerprinting tools (identification even in private browsing mode)
- Real-time bidding tools
- Other web technologies



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 @ -40,7 +40,7 @
22 "Unknown:Doi": "[REDACTED]",
23 "Unknown:CrossMarkDomains": ["elsevier.com","sciencedirect.com"],
24 "Unknown:Journal_article_version": "VoR",
25 - "Unknown:ZlkjSmweKzteJoduGntn_y8NNztr7lwiOntuGzgyOntepNd60yMeQo9ePndmKnd-KmMyTma": "",
26 + "Unknown:LXLf8mt-SywmOn9yGy9mLzcNNngj9lwiRntyGztz7ogyOmwuLndqPo9ePndmKnd-Kn9uTma": "",
27 "Unknown:Robots": "noindex",
28 "Unknown:Grabs": true,
29 ALT "Unknown:CrossmarkMajorVersionDate": "2010-04-23",
30
```

Science Tracking

raw.githubusercontent.com_the-markup_xandr-audience-segments_main_data_marketplace_public_segments_pricing_05212021.csv - HP Sure Click Secure View

Datei Start Einfügen Seitenlayout Formeln Daten Überprüfen Ansicht Was möchten Sie tun? Anmelden Freigeben

Einfügen Zwischenablage Schriftart Ausrichtung Zahl Formatierung Formatvorlagen Zellen Bearbeiten

A545784 LiveRamp Data Store,8082,23584537,LexisNexis > Stirista Fusion > Professional License Type > Pharmacists Engineers Scientists Lawyers and Architects

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
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														

raw.githubusercontent.com_the-m

Bereit 100%

Science Tracking

- The information behavior of researchers is not private; rather, it is tracked on an individual basis
- The data does not remain with the publishers; instead, it is also aggregated and exploited by data brokers
- Comprehensive online profiles are created: what a researcher looks up is linked to
 - shopping
 - social media
 - her jogging route
 - whether she has children etc.

```
"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",
  "buyeruid": "fcd4282456238256034abcdef220d9aa5892",
  "yob": 1990,
  "gender": "F"
},
"device": {
  "type": 4,
  "ifa": "8846d6fa10008bceaaf322908dfcb221",
  "ip": "1.2.3.4",
  "ua": "...user agent string...",
  "make": "Apple",
  "model": "iPhone",
  "hvw": "6s",
  "os": 13,
  "osv": "11.4.1",
  "mccmnc": "310-005",
  "geo": {
    "type": 1,
    "lat": 42.3601,
    "lon": 71.0581,
    "country": "USA",
    "timezone": "EST"
  }
}
```

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

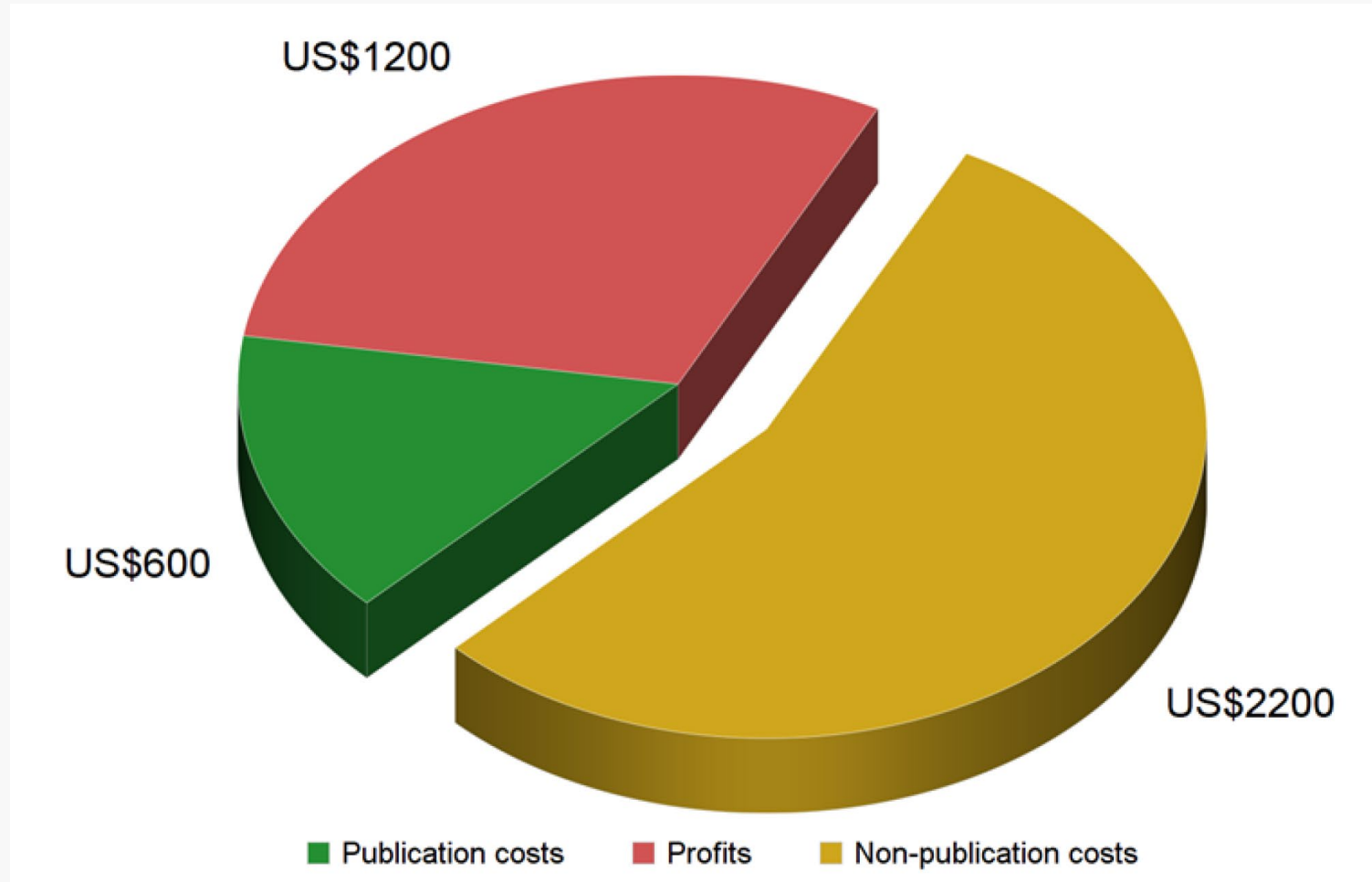
This young woman's GPS coordinates!

Workbenches



Baden-Württemberg
Ministerium für Wissenschaft,
Forschung und Kunst

Workbenches



Workbenches

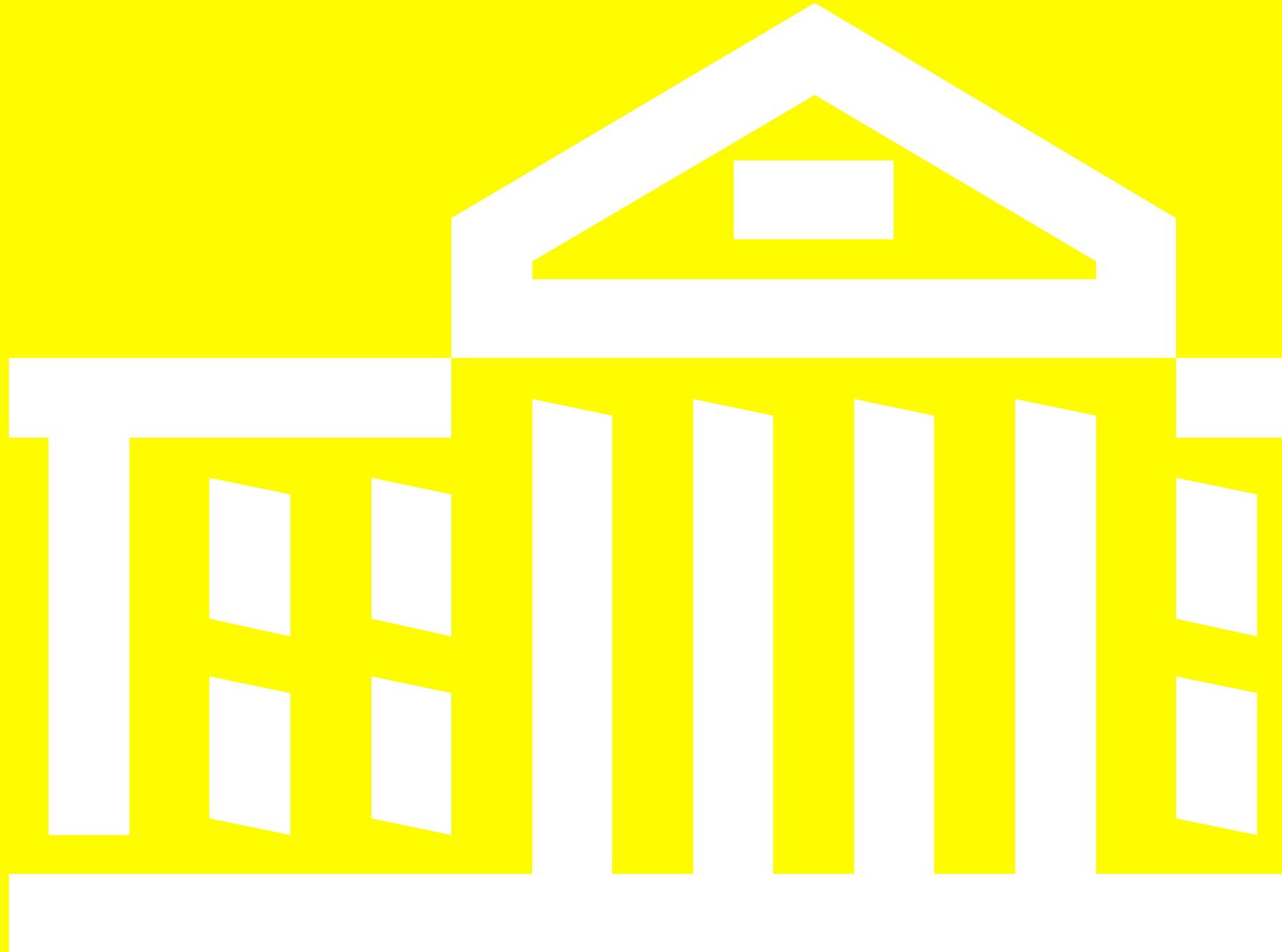


Workbenches

The Supercontinent of Scholarly Publishing?

By **ROGER C. SCHONFELD** | MAY 3, 2018 | **14 COMMENTS** | TIME TO READ: 8 MINS

Data Cartels



Baden-Württemberg
Ministerium für Wissenschaft,
Forschung und Kunst

Data Cartels

- transform public goods into club goods, as evident in the provision of information through paywalls and APCs
- are the “GAFAM of curated information” and create a hierarchy in access to knowledge
- generate advantages from economies of scale: their size enables an ever-growing array of data products, which in turn benefit the big players the most
- The monetization of usage data and knowledge work as a whole amounts to the exploitation of very personal and personally identifiable information

DATA CARTELS

The Companies that Control and
Monopolize Our Information

SARAH LAMDAN

Data Cartels

- They are extremely opaque: one only sees more and more datagoing in and new data products coming out
- They reinforce hierarchies and insulate against external oversight through the use of AI
- They are becoming embedded in more and more decision-making processes in the public sector (key performance indicators, rankings, “security”)
- They are the “informational landlords who decide who can swim in the ocean of knowledge”



Research Security



Baden-Württemberg
Ministerium für Wissenschaft,
Forschung und Kunst

Research Security



Wolfie Christl
@WolfieChristl



This unredacted 214p contract between LexisNexis Risk Solutions and the US migrant persecution agency ICE is quite something:
s3.documentcloud.org/documents/2385...

...it suggests, once again, that LexisNexis aka Elsevier aka RELX Group is a kind of outsourced domestic US intelligence agency.

„Regardless of which company will offer this as a product - in a situation where scientific publications, funding decisions, and web resources of scientific institutions are being withdrawn and blocked en masse in order to check them against lists of banned words - tracking data of researchers collected over many years can also become a threat and be used to generate a loyalty score based on previous behavior. Who worked on the wrong topics? What are their networks? What are their other activities, movements, lifestyle, family relationships?“

Research Security

„It is important to bear in mind that not only academic institutions but also research-based companies and providers of knowledge-intensive services may rely on digital science and data platforms, and thus the issues outlined above apply in principle to companies as well. If the large consortium in the DEAL project has failed, after protracted negotiations, to negotiate contracts that are fully satisfactory in all cases with regard to data protection and “science tracking,” individual companies will certainly also face difficult negotiations. It therefore seems necessary to devote further attention to the issue of tracking in science within the framework of research and innovation policy, while taking into account the implications for the economy.“

Tracking in der Wissenschaft



**Wir
gestalten
Zukunft**

VDI Research

Thank you!



Ministerium für Wissenschaft, Forschung
und Kunst Baden-Württemberg (MWK)

Dr. Renke Siems