

# Unter Beobachtung

## User Tracking, Reading Analytics und der Wunsch nach mehr

**Dr. Renke Siems**

**20. BVB-Verbundkonferenz**

**16. September 2021**



**Baden-Württemberg**

MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST

# Agenda

- Der Ausgangspunkt: Authentifizierung
- Methoden des Datensammelns
- Neue Geschäftsmodelle
- Datenmärkte
- Bibliothekssysteme
- Was tun?



# Authentifizierung



<https://seamlessaccess.org>  
<https://ra21.org/>  
<https://www.getfulltextresearch.com/>



# Authentifizierung

- Verlage drängen auf Veränderung der Authentifizierung
- Wunschlösung: RA21 / Seamless Access / GetFTR
- Die neuen Technologien gehören zur SAML-Familie wie Shibboleth, sind aber stärker in den Plattformen integriert
- Bibliotheken würden marginalisiert werden (RA21) oder herausfallen (GetFTR)
- Bibliotheken fürchten um den Datenschutz ihrer Nutzer:
  - Empfehlungen der dbv-Erwerbungskommission



# Authentifizierung

Die Diskussion wurde weltweit geführt und dann fiel eine Bemerkung:

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

Todd Carpenter (NISO)

(Zitiert nach <https://www.codyh.com/writing/tracking.html>)



# Authentifizierung

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



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



# Authentifizierung

Aber schon zuvor war dies nicht unbemerkt geblieben:

Thursday, March 12, 2015

---

## 16 of the top 20 Research Journals Let Ad Networks Spy on Their Readers

---

A **recent query** to the "LibLicense" listserv asked:

Is there any kind of organization that has put together a website or list of database providers/publishers that indicate the extent to which they respect patron privacy?

The answer is "no", but I thought it would useful to look at the top journal publishers to

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



# Methoden des Datensammelns

- First Party Data
- Third Party Data
- Bid Streaming
- Port Scanning
- „Trojaner“
- ...



# First Party Data



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



# Third Party Data

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.

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



# 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)



# 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": "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!

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



# Port Scanning

- Port-Scanning fremder Computer und Netzwerke am Rande der Legalität nach deutschem Verständnis (§ 202c, 303b StGB)
- International benutzt für Betrugsprävention und Tracking
- Akteur u.a. der Elsevier-Mutterkonzern RELX durch Tochtergesellschaft ThreatMetrix
- ThreatMetrix auf ScienceDirect integriert, Nutzung unklar



Access through your institution

to view subscribed content from home

Download PDF Share Export

Search ScienceDirect Advanced

Outline

- Abstract
- Graphical abstract
- Keywords
- Abbreviations
- 1. Introduction
- 2. Classification of NPs



Arabian Journal of Chemistry  
Volume 12, Issue 7, November 2019, Pages 908-931



Recommended articles

- Introduction to Nanomaterials: Basic Properties...  
Nano-Sized Multifunctional Materials, 2019, pp. 1-19  
Download PDF View details
- Nanomaterials Definitions, Classifications, and ...  
Nanoengineering, 2015, pp. 3-40  
Download PDF

Review

# Nanoparticles: Properties, applications and toxicities

Inspector Console Debugger Network Style Editor Performance Memory Storage Accessibility Application What's New

online-metrix.net

Status	Method	Domain	File	Type	Transferred	S...	Res...
200	GET	h.online-metrix.net	...	html	13.25 KB	5...	2 s
204	GET	h.online-metrix.net	...	js	387 B	0 B	1.98 s
200	GET	72k6ns9a6hwsudlw2c7dlij...	...	png	438 B	8...	2.14 s
204	GET	h.online-metrix.net	clear1.png;CIS3SID=...	png	386 B	0 B	3.65 s

Headers Cookies Request Response Timings Stack Trace Security

Filter Headers

- Server: Apache
- Strict-Transport-Security: max-age=31536000
- X-Content-Type-Options: nosniff
- X-XSS-Protection: 1; mode=block

Request Headers (3.059 KB) Raw

- Accept: image/webp,\*/\*
- Accept-Encoding: gzip, deflate, br
- Accept-Language: en-US,en;q=0.5
- Connection: keep-alive
- Cookie: thx\_global\_guid=75e...
- Host: h.online-metrix.net
- Referer: https://h.online-metrix.net/ZYT...

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



# „Trojaner“

- Führende Verlage liebäugeln mit Backdoors zu universitären Netzwerken bei Benutzung ihrer Produkte
- Biometrische Identifikation durch Tippgeschwindigkeit, Mausbewegung usw., um Nutzer\*innen in anderen Bereichen des Webs wiedererkennen zu können
- Kalter Krieg 4.0: Länder sollen vom Informationszugang abgeschnitten werden
- Bibliotheken sollen Rabatte für Mitwirkung angeboten werden



# „Trojaner“

## Modern Library Design

### Library Logs/Info:

- Timestamps
- Extensive Browser Info
- Username
- Account Information
- Customer IP
- URLs Requested
- 2-Factor Device Info
- Geographic Location

- User Behavior
- Biometric Data
- Threat Correlation / Info Sharing

### Publisher Logs:

- Timestamps
- Proxy IP
- URLs Request



<https://netzpolitik.org/2020/news-from-elsevier-no-open-access-deal-but-spyware-against-shadow-libraries/>



# Neue Geschäftsfelder



[https://commons.wikimedia.org/wiki/File:ICE.Arrest\\_lg.jpg](https://commons.wikimedia.org/wiki/File:ICE.Arrest_lg.jpg)



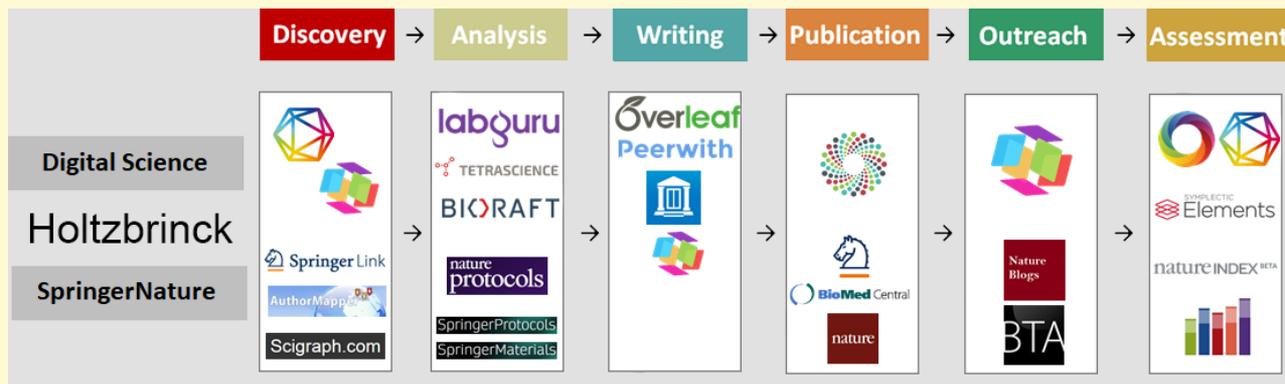
# Neue Geschäftsfelder

- 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
- Ob Nutzungsdaten wissenschaftsbezogener Produkte verwendet werden, ist mangels Transparenz unklar



# Neue Geschäftsfelder

- 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)



# Datenmärkte

- Entstehen derzeit in öffentlicher, wissenschaftlicher und kommerzieller Ausrichtung
- Wissenschaftsbezogene Stränge z.B. NFDI, EOSC, GAIA-X
- Großverlage mutieren zu Data Analytics-Firmen:
  - Publikationen nur noch ein Teil der Wertschöpfung
  - Workbenches können vertiefte Analysen ermöglichen



# Datenmärkte

- Open Access auf APC-Basis hilft nicht gegen Tracking
- Neue Form von Double-Dipping: Autor\*innen bezahlen mit Gebühren, Leser\*innen mit persönlichen Daten
- Workbenches eskalieren die Situation:
  - Verknüpfung zu Ökosystemen erzeugt mehr Daten
  - Lock-In-Effekte breiten sich in weitere Teile des Research Life Cycles aus
  - Charakter von Ransomware: der Druck, alles zu akzeptieren und zu zahlen, steigt mit jeder Akquisition



# Datenmärkte

- Vier zentrale Problemfelder:
  - Grundrechtseingriff Wissenschaftsfreiheit
  - Gefährdung von Nutzern (z.B. Wissenschaftler\*innen mit umstrittenen Themen)
  - Verletzung von Verwertungsinteressen der Hochschulen (kommerzielle Konkurrenten sind über Forschungsstand frühzeitig orientiert)
  - Wettbewerbsverzerrung zu Ungunsten der Verlage, die nicht tracken



# Bibliothekssysteme

EBSCOpost

Librarianship

News

Workflow

Technology

Library Resources

## 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

- Internationaler Trend: Learning Analytics durch die Hintertür durch Datenerfassung und -verknüpfung:
  - Mediennutzungsverhalten
  - Aufenthaltsmuster (WLAN-Tracking)
  - Nutzung von Beratungsdiensten
  - Datenintegration in die Studierendenakten



<https://twitter.com/greenarchives1/status/1400067993929912326>



# Stellungnahmen

*Solcher Datenmissbrauch und solche Wissenschaftsspionage sollten durch die Anpassung von einschlägigen Gesetzen insbesondere auf europäischer Ebene ausgeschlossen werden. Die informationelle Selbstbestimmung von Wissenschaftlerinnen und Wissenschaftlern muss im digitalen Raum oberste Priorität haben.*

Deutsche Forschungsgemeinschaft



# Stellungnahmen

- AWBI befürchtet Privatisierung der Wissensgesellschaft: nicht mehr die öffentliche Hand, sondern Unternehmen verfügen über das Steuerungswissen
- Hochschulen und Bibliotheken könnten zu Mitwirkenden bei Verletzung von Datenrecht und Forschungsfreiheit geraten
- Die geplanten „Trojaner“ könnten die Sicherheit der Hochschulnetze gefährden und Angriffe ermöglichen
- Datenschutzgerecht, risikobewusst und in der Hoheit der Wissenschaft könnte science analytics nützlich sein.



# Was tun?

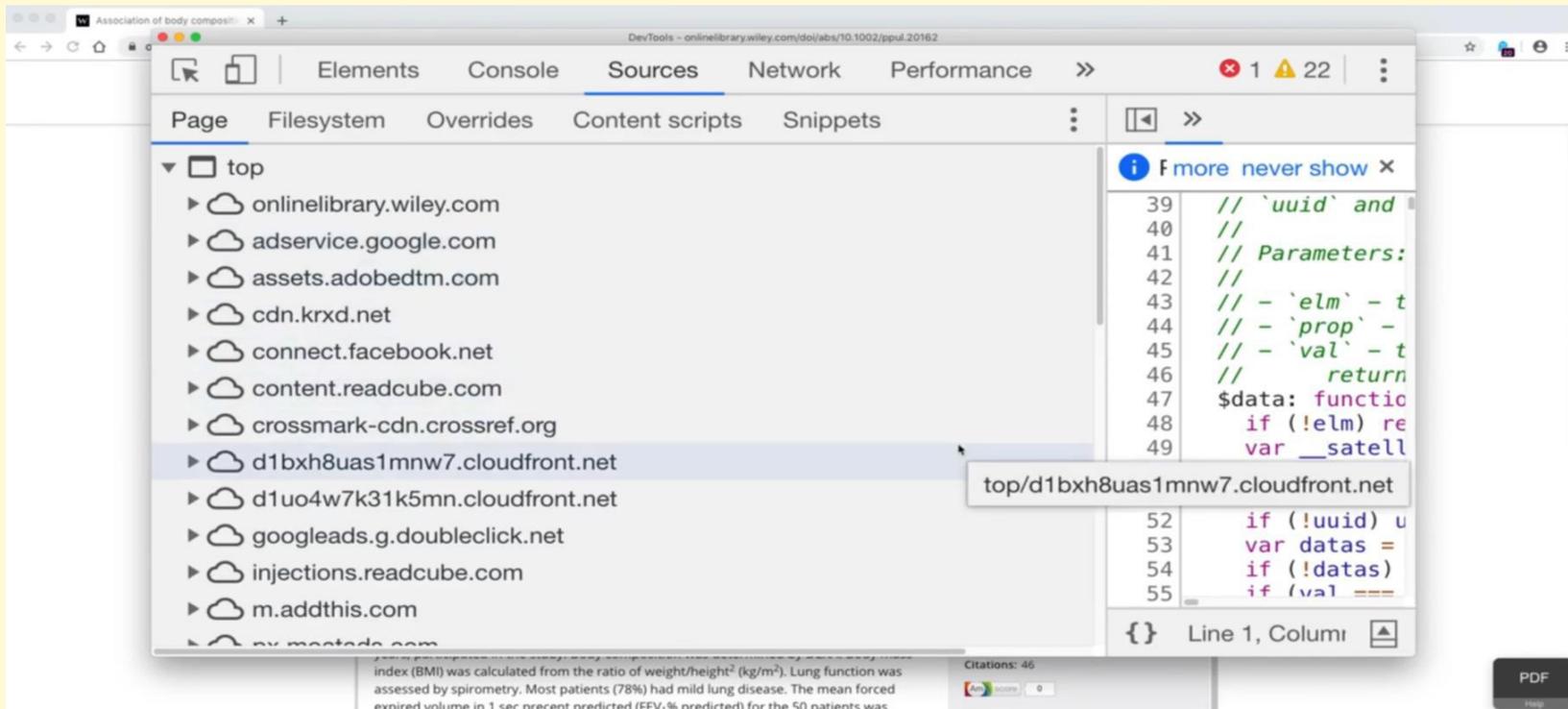
- Aufklärung und Prüfung: technische und rechtliche Abschätzung tut not auch für das eigene Handeln
- Awareness: (berufsethische) Diskussion über die Rolle von Bibliotheken und Bibliothekar\*innen in diesem Kontext
- Fortlaufende Aufmerksamkeit bei „tollen neuen Services“ wie Google CASA

*Wir müssen mehr lernen, in Daten zu denken, um eine „Lebendfalle für Forschende“ (Petra Gehring) zu verhindern*

<https://www.forschung-und-lehre.de/das-schicksal-von-open-science-steht-auf-dem-spiel-3902/>



# 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/>

