

The image shows the letters 'DFG' in a large, white, 3D font against a blue background. The letters are slightly out of focus, giving a sense of depth. The background is a gradient of blue with some light streaks.

# Data Tracking in Research - Experiences from the DEAL Negotiations

Dr. Michaela Bilic-Merdes, German Research Foundation (DFG)

Printemps Couperin | Paris | 19.03.2025

# Data Tracking in Research

## User Tracking on Academic Publisher Platforms

*Cody Hanson*  
*@codyh*  
*codyhanson@umn.edu*

[User Tracking on Academic Publisher Platforms](#)



## Stop Tracking Science

[Stop Tracking Science](#)

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)  
20 May 2021

Deutsche Forschungsgemeinschaft  
Kennedyallee 40 · 53175 Bonn · Postanschrift: 53175 Bonn  
Telefon: +49 228 888-11 · Telefax: +49 228 888-2177 · postmaster@dfg.de · www.dfg.de



[datentracking\\_papier\\_en.pdf \(dfg.de\)](#)

- ▶ **Information paper** of Committee on Scientific Library Services and Information Systems (AWBI), DFG
  - Presentation of **current practices of data tracking**
  - Stimulate **discussion about consequences** amongst all stakeholders in the scientific community

# Data Tracking in Research

## Digital Privacy

Dr. Robert Altschäffel, Prof. Dr. Michael Beurskens, Prof. Dr. Jana Dittmann,  
Prof. Dr. Wolfram Horstmann, Dr. Stefan Kiltz, Prof. Dr. Gerhard Lauer, Judith  
Ludwig, Dr. Bernhard Mittermaier and Katrin Stump<sup>1</sup>

### Data tracking and DEAL

– On the 2022/2023 negotiations and the consequences for academic  
libraries –

In memory of Prof. Dr. Gerald Spindler (1960-2023)

#### I. The Maxwell-Garfield system of academic publishing

After the end of the Second World War, a structure of academic publishing developed that can be specifically referred to as the *Maxwell-Garfield system*. This refers to two post-1945 developments that were initially only indirectly linked to one another. The name *Robert Maxwell* is associated with a particularly ruthless commercialisation of academic publishing (Miranda 2001) where costs are allocated solely to the public and profits solely to the publishers. With the rapid growth of universities, this business model – described by Deutsche Bank as “bizarre” (Deutsche Bank 2005; Klein 2019) – gave rise to an oligopoly of a small number of publishers, especially in the academic journal sector, while also resulting in drastic price increases for libraries and low costs for publishers. Meanwhile, the name *Eugen Garfield* is associated with an index that was initially developed to help libraries sort their acquisitions (Garfield 1955; Cronin & Atkins 2000), but which rapidly became an instrument of control, favouring the large publishers who adopted this index early on in guiding their pricing policy; at the same time, it enabled these publishers to make data available to university administrations as a basis on which to make decisions on research policy investments and structural developments.

<sup>1</sup> Members of the Ad Hoc Working Group of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) on Data Tracking in Research under the direction of Dr. Michaela Bilic-Merdes: Dr. Robert Altschäffel, Otto von Guericke University Magdeburg; Prof. Dr. Michael Beurskens, University of Passau; Prof. Dr. Jana Dittmann, Otto von Guericke University of Magdeburg; Prof. Dr. Wolfram Horstmann, IZ Karlsruhe – Leibniz Institute for Information Infrastructure; Dr. Stefan Kiltz, Otto von Guericke University of Magdeburg; Prof. Dr. Gerhard Lauer, University of Mainz; Judith Ludwig, German National Library of Science and Technology (TIB), Hanover; Dr. Bernhard Mittermaier, Forschungszentrum Jülich; Dr. Hildegard Schäffler, Bavarian State Library, Munich; Prof. Dr. Gerald Spindler, University of Göttingen and Katrin Stump, Saxon State and University Library, Dresden.

Recht und Zugang, issue 1-2024 DOI: 10.5771/2699-1284-2024-1-23

<https://doi.org/10.5771/2699-1284-2024-1-23>, am 04.07.2024, 11:15:07

Open Access: <https://www.nomos-elibrary.de/jku>

[Data tracking and DEAL: On the 2022/2023 negotiations and the consequences for academic libraries](#)

SPARC\*

NAVIGATING  
RISK IN VENDOR  
DATA PRIVACY  
PRACTICES

An Analysis of Elsevier's  
ScienceDirect

November 2023  
© 2023 SPARC. Licensed to a Private Company Member of a Commercial Library



SPARC\*

NAVIGATING  
RISK IN VENDOR  
DATA PRIVACY  
PRACTICES

An Analysis of Springer  
Nature's SpringerLink

October 2024  
© 2024 SPARC. Licensed to a Private Company Member of a Commercial Library



► **Data tracking issues in the context of the DEAL negotiations** with the academic publishers Elsevier, Springer Nature and Wiley

- Inclusion of data protection clauses in license agreements
- Consultation and assessment of the results by the DFG Working Group Data Tracking and DEAL

# Shift in Data Privacy Responsibility

## Analog Library World



Pixabay

- ▶ Privacy-protecting environment
- ▶ High library privacy standards
- ▶ Responsibility and control by librarians

## Digital Information Platforms



DALL-E

- ▶ Privacy-intrusive collection of personal usage data
- ▶ Commercial privacy „standard“ / „commercial surveillance“
- ▶ Control by users / out of control for third-parties

**Potential for data misuse**

**Take up responsibility!**


# Transformation to Data Analytics Business

## Changing business model of scientific publishers: from **content provider** to **data analytics businesses**

- ▶ Information platforms, databases and research tools, providing **services & insights** into the **entire research cycle**
- ▶ **Comprehensive data collections** are created in the hands of few large corporations
- ▶ **Research Information Systems**
- ▶ **AI Analytics & Research Intelligence Solutions**



[Research - Business School Library - LibGuides at Edinburgh Napier University](#)

 <https://www.elsevier.com> :

**Elsevier | An Information Analytics Business**

Elsevier is a global information analytics company that helps institutions and professionals progress science, advance healthcare and improve performance.

[Publish with Elsevier](#) · [Elsevier Connect](#) · [Elsevier for Reviewers](#) · [Books & Journals](#)

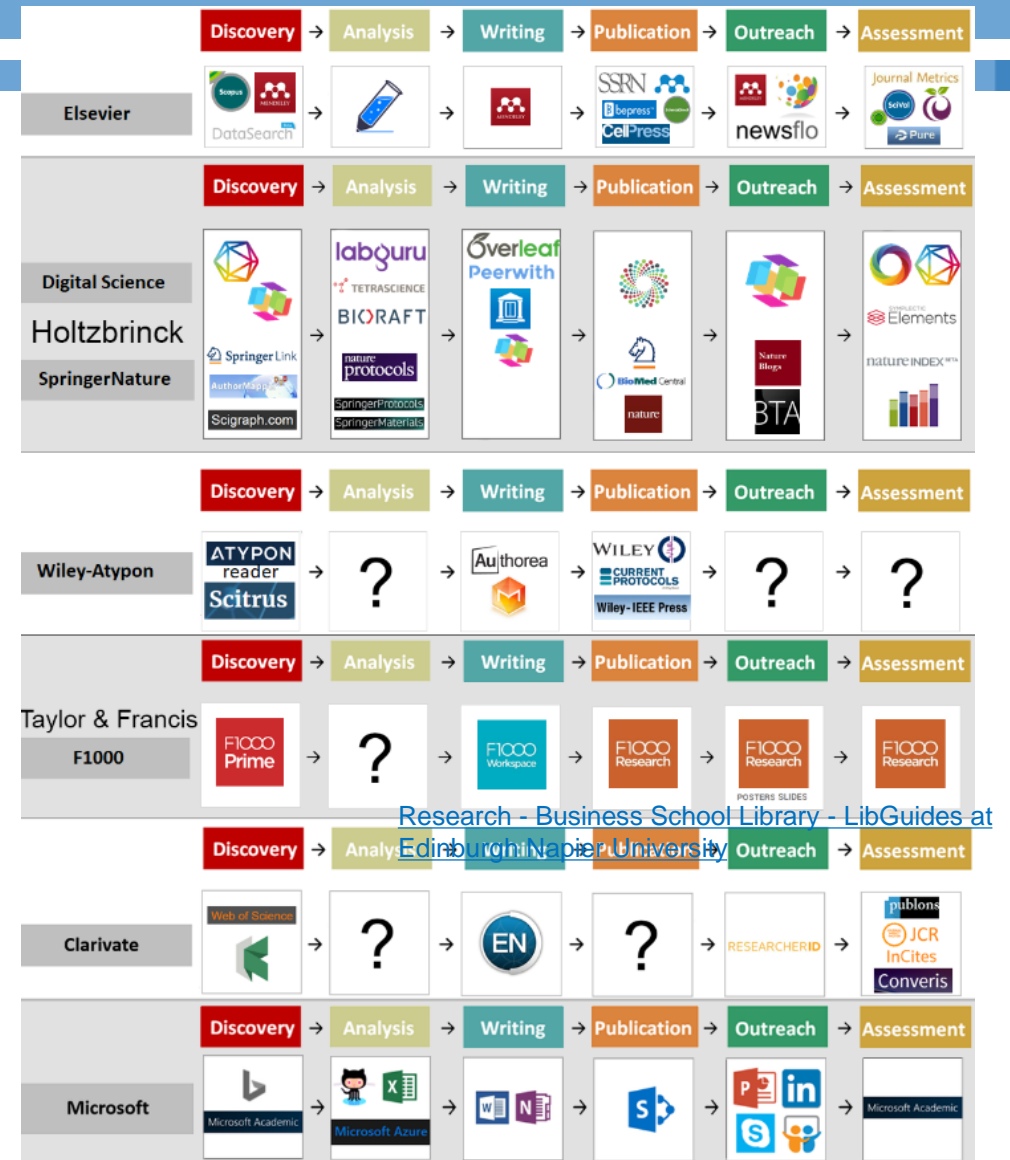


# Transformation to Data Analytics Business

## Question of Data Sovereignty

- ▶ „Data analytic businesses“ are creating data on knowledge, research developments and stakeholders
- ▶ Lock-up into proprietary systems for all activities in research workflows
- ▶ Decision-making and control instruments in commercial hands

Control over data?



Research - Business School Library - LibGuides at Edinburgh Napier University

# What is Data Tracking?

## Personal Data Traces and Usage Data

- ▶ Capturing and **storing data of researchers** when using information services, e.g. during literature research
- ▶ **Personalized profiles**, access and usage data, length of stay, downloads, mouse movements etc.
- ▶ Partly under **insufficient information** of the users and **without active consent**
- ▶ Data can be **aggregated and enriched** with additional data from the non-scientific environment to form profiles

**Academic information platforms collect, share and make use** of the recorded data and **personal information in often non-transparent and unregulated ways**



# How is Data Collected? Tracking by Publishers and Third-Parties



## ► First-Party (by publishers)

- **First-party cookies**: direct user traces
- **Tracking-pixel technologies** allow for extensive collection and **usage & interest analysis** in real time

## ► Third-Party (by commercial service providers, data brokers, social media platforms etc.)

- **Third-party cookies**: Javascript code has access to the Document Object Model, can read out text users are engaging with, what they browse to next, what search words are entered etc.
- **Browser-fingerprinting technologies** collect e.g. **behavioral and location data**

## ► Bidstream data, beacons, port scanning, pdf tracking and other tracking technologies

- Background collection of data about **location, devices, data used**; **geofencing** and **IP-targeting**
- Linking **real-time data to an identifier**; **aggregating data**
- Collection of **biometric data** (e.g. typing speed, type of mouse movements) to personalize users despite the use of proxy servers and VPN tunnels



# Why is it Problematic?

- ▶ Use of vast range of **invasive web tracking methods** to track user behaviour inside, outside and beyond academic information platform
- ▶ **Extensive collection** of broad range of personal (usage) data
- ▶ **Outside of publisher's influence or control** - data processing, sharing, trading, selling by third parties and tracking service providers is often **non-transparent and unregulated**
  - Search engines
  - „Security“ tools providers, other service providers
  - Social media platforms
  - Personal-data aggregators and profilers such as Google, Adobe, Cloudflare, New Relic, etc.
  - Data brokers, Ad Tech, Real-time bidding platforms, etc.
- ▶ Data can be **aggregated and enriched** with additional data from the non-scientific environment to **form profiles**

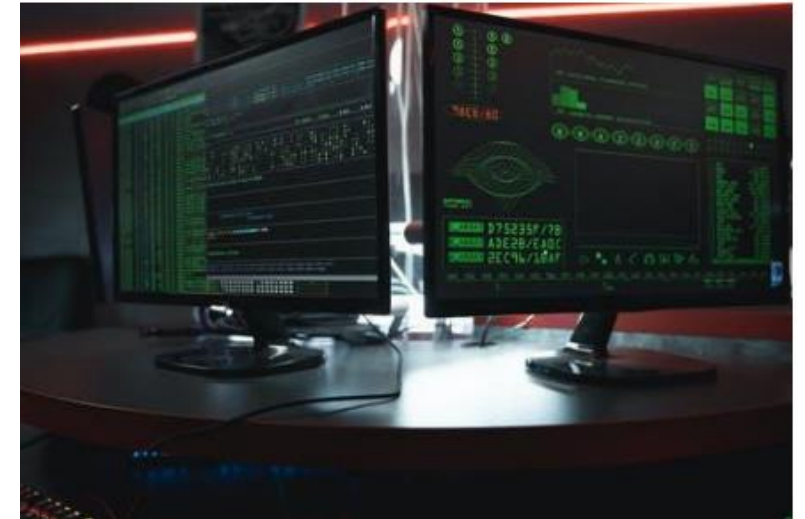
**many have  
critical privacy evaluations**

**may form individual and  
identifiable researcher profiles**

# Possible Consequences of Data Tracking and Data Aggregation

- ▶ Freedom of science and freedom of research and teaching at risk
- ▶ Right of informational self-determination / interference with anonymity of researchers guaranteed by data protection law
- ▶ May promote data misuse and scientific espionage
- ▶ Personal discrimination, exclusion, other social implications

***Ethical reflection needed, exploration of technical and legal frameworks & change of current practice***



ALLIANZ DER  
WISSENSCHAFTSORGANISATIONEN

DFG

- ▶ Not only technical and legal issue, but one of **research ethics** and **research policy**
- ▶ Privacy-intrusive data collection and user tracking **conflicts with library privacy standards**
- ▶ Importance of **transparency** and **trust** in **ethically integer** scientific infrastructure platforms

# How Can We Change Current Malpractice?

Use **national consortial negotiations** to address data tracking issues with the large academic publishers



## Aim

- ▶ **Transparency** on personal usage data collection, processing, storage etc.
- ▶ Proper legal handling of **cookie management** and **consent management**, with **fully informed active consent** of users
- ▶ **Reduction of Data Tracking / Default: NO TRACKING**
  - **No use** of invasive tracking technologies
  - **No tracking** by third parties
  - Data tracking **only if necessary** for **performance** of agreement, **compliance** with legal obligation or **legitimate interested** (with the necessity of the interest explained)
- ▶ **Data Minimization** – keep collected data and the collectors to a minimum

**DEAL project** - Rector's Conference (HRK)

- ▶ Initiated to negotiate **nationwide transformative “Publish and Read” agreements** with major publishers of academic journals that enable the open dissemination of research results from Germany
- ▶ Moving from subscription paywalls to **open access publishing services**
- ▶ Negotiations on behalf of **all German academic institutions**, including universities, universities of applied sciences, research institutions and state and regional libraries



Springer Nature 2020-2023

**DEAL Contract Negotiations 2024-2028**

in 2022/2023 with

Elsevier

Springer Nature

Wiley



Wiley 2019-2023

# Data Privacy Negotiations in DEAL

## ▶ DFG Working Group „Data Tracking and DEAL“

- AWBI, DEAL-negotiators, researchers, librarians, technical and legal experts
- Anchoring data privacy aspects in (DEAL) licence agreements
- Ensuring co-determination of science and transparency in dealing with personal data
- Raising awareness for science tracking - librarians, academic institutions, researchers
- Establishing best practice and strategies on how to deal with data tracking
- Legal opinion = Joint responsibility of libraries and publishing company in case of damage

## ▶ Working Groups: DEAL & Publisher on “Data Protection and Data Tracking”

- Separate working group for each publisher
- Discussion, consultation and preparation of standard contract wording



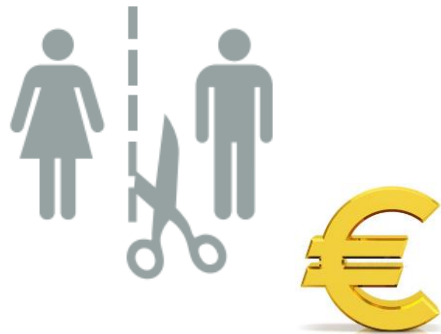
# (No) Joint Responsibility of Libraries and Publisher

Is there a **joint responsibility of libraries and publishing company** in case of damage?

**Controller-to-Controller Relationship**

vs.

**Joint-Controllership**



**Publishers**

legal opinions differ



**DEAL / legal opinion DFG**

- ▶ **Legal certainty** at national & European level missing
  - Court decision or involvement of **data protection agencies** necessary
  - In Germany: **legal complaint** by Prof. Brembs against Nomos, Springer and Wiley [Gesellschaft für Freiheitsrechte e.V.](#)
- ▶ **Compromise & workaround**: Responsibilities regulated in agreement - **indemnification clause**



# Important Data Protection Aspects in the DEAL Contracts

	Elsevier	Wiley	Springer Nature
<b>Joint Responsibility</b>	Rejected by all publishers, workaround instead		
<b>(International) Data Transfer</b>	Permitted according to Art. 44 ff GDPR		Permitted only after approval by DEAL
<b>Use of Tracking Tools</b>	Permitted for legitimate interests		Only if necessary for contract fulfillment
<b>Cookies Recovery</b>	Shall refrain from	Shall use reasonable efforts	Shall refrain from
<b>Cookies Storage Duration</b>	No limitation		Max.1 month
<b>Data Minimization / Anonymisation</b>	Full IP-address: no longer than necessary (?); Joint Publisher/DEAL Workshop		Full IP-address: max. 48 h

Adapted from Bernhard Mittermaier, DFG Working Group Member

# Lessons Learned

## **Novelty: a data privacy clause in an national transformative Open Access agreement**

- ▶ Potential for improvement!!!
- ▶ Other **European countries** / consortia are taking up the topic in their negotiations, taking DEAL contracts as example (e.g. new SURF-Elsevier contract – **new „standard“**)
- ▶ Make data privacy claims part of the **main negotiations** and subject to **commercial terms** of the agreement / an essential part of the **legal and ethical framework**
- ▶ Ongoing **DIALOG with scientific publishers & information providers** needed
  - **Workshops** with the DEAL publishers on data privacy topics
  - Initially contractually planned only for the topic of IP-addresses, but goes well beyond
- ▶ **International dialog and awareness raising** with academic publishers / STM Association & all academic information providers

# Impulse & Appeal to Academic Publishers

- ▶ **TRANSPARENCY** in all steps of personal (usage) data life cycle
- ▶ Special **high demands** for **privacy standards for researcher** and
- ▶ „Personal freedom of usage“ / **anonymity of a researcher** on an
- ▶ Academic publishers need to be **trustworthy, reliable and ETHICAL**
- ▶ Voluntary ethical „**Code of Conduct**“ to go **BEYOND COMPLIANCE** and researcher
- ▶ Safe and **privacy-protected academic search and read environment**
- ▶ **Personal DATA MINIMIZATION** in all aspects

**Shift from extensive tracking to a privacy-friendly and trustful information platform**

## ▶ Ideas & Suggestions

- Default: **No Tracking**
- Real choice for users: opt-in to „tailored services“ e.g. for search history, specific social media, ...
- No tracking by third parties
- „Positive list“ of tracking service providers/ third parties
- Ethically integer „Data Privacy Board“
- Critical assessment of all tracking technologies service providers/third parties with respect to necessity, user privacy, compliance and high ethical privacy standards
- ...

# Recommendations & Future Perspectives

- ▶ Constant **monitoring** of the data tracking, its extent and tracking technologies applied
  - *Allianz Task Force Data Tracking – handout planned for end of 2025*
- ▶ Apply **technical precautions** on individual and organisational level
- ▶ Raise **awareness** on topic of data tracking in research
- ▶ Expand **Open Access** and create **non-commercial infrastructure** being **privacy-protective**
- ▶ **International approach** on data tracking and topics of data sovereignty



**„Tracking“ / science analytics can be useful**  
**if used data privacy-consciously, risk aware and**  
**under the sovereignty of science**



# Thank you for your attention!

## More Information

- ▶ on DFG : [DFG - Deutsche Forschungsgemeinschaft](#)
- ▶ on AWBI information paper : [Data tracking in research](#)
- ▶ on Data tracking and DEAL paper : <https://zenodo.org/records/13837664>
- ▶ on DEAL: [About DEAL - DEAL Konsortium \(deal-konsortium.de\)](#)