

Die im Dunklen sieht man doch

Dunkle Daten und Datentracking in der Wissenschaft

In der Wissenschaft wurde schon sehr früh vom »Data Deluge«, also der gehörigen Flut an Forschungsdaten gesprochen.¹ Sie werden insbesondere in Großforschungsprojekten mittels Teleskopen, Supercomputern oder Teilchenbeschleunigern produziert. In diesem Rahmen wurde sogar ein viertes Paradigma, nämlich das der datengetriebenen Wissenschaft ausgerufen.² Im Kielwasser dieser riesigen Datenflut schwimmen jede Menge dunkle Daten.³ Dies bezeichnet Daten, die nicht ohne Weiteres (nach-)nutzbar, d.h. für Dritte oder auch nach einer gewissen Zeit nicht ohne weiteren Aufwand zugreifbar und verstehbar sind. Sie sind kein reines Phänomen der Wissenschaft,⁴ im Bereich der kommerziellen Datenverarbeitung gehen Schätzungen davon aus, dass die Klasse der dunklen Daten bis zu 80 % ausmacht.^{5,6}

Im kommerziellen Bereich werden die bewusst und als Schutzmaßnahme im Dunkeln liegenden Datenspuren der Nutzer:innen (personenbezogene oder personenbeziehbare Daten, die das Verhalten im Internet abbilden) durch Datentracking zum Zwecke der Monetarisierung an das

-
- 1 Tony Hey und Anne Trefethen: »The Data Deluge: An e-Science Perspective«, in: Fran Berman, Geoffrey Fox und Tony Hey (Hg.): *Grid Computing: Making the Global Infrastructure a Reality*, Hoboken 2003, S. 809–824.
 - 2 Tony Hey, Stewart Tansley und Kristin Tolle, K. M. (Hg.): *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Washington 2009.
 - 3 Björn Schembera und Juan M. Durán: »Dark Data as the New Challenge for Big Data Science and the Introduction of the Scientific Data Officer«, in: *Philosophy & Technology* 33 (2020), Heft 1, S. 93–115.
 - 4 Im wissenschaftlichen Bereich sind dunkle Daten zumeist unerwünscht (außer beispielsweise bei medizinischen Daten), im kommerziellen Sektor können dunkle Daten für Einzelpersonen durchaus wünschenswert sein, um die Privatsphäre zu schützen.
 - 5 Michael O'Reilly: »The Unseen Data Conundrum«, in: *Forbes*, 3.2.2022, <https://www.forbes.com/sites/forbestechcouncil/2022/02/03/the-unseen-data-conundrum/> (aufgerufen: 12.5.2023).
 - 6 Die »Data«-Unterdisziplinen der Informatik wie Data Science, Data Analytics oder Data Engineering haben allesamt zum Ziel, Licht ins Dunkel der Daten zu bringen.

profitable Licht gebracht. Auch die Bewegungen von Wissenschaftler:innen werden im Netz mittels Techniken wie Cookies, IP-Adressen-Erkennung, Einbindung von externen Trackingdiensten wie Google Analytics, Trackingpixel (oder einer Kombination) von großen Verlagen verfolgt.⁷ Dies ermöglicht das Erstellen umfassender Nutzer:innenprofile, deren Hauptgrund das gezielte Einspielen von Werbung ist. Schließlich wird dann auf Grund von Korrelationen (Nutzer:innen der Kategorie x, ausgezeichnet durch Merkmale x_{attr} , kaufen wahrscheinlich Produkt y) auch auf anderen Seiten Werbung eingeblendet.

Individuell gibt es einige technische Mittel, um das Datentracking zumindest zu erschweren. Im Browser sollte hierfür eingestellt werden, dass die besuchten Websites angewiesen werden, kein Tracking zu verwenden, sowie die Speichereinstellungen von Cookies entsprechend gesetzt werden. Ebenso sollte der Browser im privaten (Inkognito-, ...) Modus verwendet werden, was die persistente Speicherung von Cookies erschwert. Beim Besuchen von Seiten sollte das angebotene Opt-Out beachtet werden, also die Möglichkeit, gewisse Trackingmaßnahmen zu verweigern. Bei Springer können beispielsweise Performance & Analyse Cookies, Functional Cookies und Targeting Cookies abgelehnt werden. Schließlich empfiehlt sich die Nutzung von Browser-Add-Ons, welche auf jeder Website Tracker aufspüren und deaktivieren können. Beliebte Tools hierfür sind Ghostery, Disconnet, NoScript oder Privacy Badger. Natürlich schränken diese Vorgehensweisen und Tools die Nutzbarkeit von Webseiten teilweise stark ein und müssen ggf. zumindest kurz deaktiviert werden (dies kann dann in einem privaten Fenster geschehen). Sie sind auch leider kein vollumfänglicher Schutz gegen Tracking, da moderne Technologien dieses trotzdem z.B. anhand des individuellen Fingerabdrucks des Endgerätes oder sogar des Tippverhaltens ermöglichen.

Daher muss das Problem des Datentrackings im Allgemeinen auf politischer und rechtlicher Ebene angegangen werden und Initiativen wie »Stop Tracking Science«⁸ verdienen volle Unterstützung. Darüber hinaus muss das gesellschaftliche System dahinter kritisch analysiert werden. Im DFG-Papier verwundert es, wenn argumentiert wird, »dass die Wissensgesellschaft durch diese Verschiebung des kommerziellen Geschäftsmodells hin zur Datenanalytik privatisiert wird [...]. Wissenschaft als öffentliches

7 Michael Freiberg: »Third-Party-Tracking bei Wiley und Springer«, in: *ABI Technik* 42 (2022), Heft 2, S. 96–104.

8 <https://stoptrackingscience.eu/> (aufgerufen: 12.5.2023).

Gut wird der Logik der Privatisierung von Infrastrukturen und ihren Folgen unterworfen.«⁹ Hier scheint es, als wäre Wissenschaft etwas von gesellschaftlichen Verhältnissen Losgelöstes, ein Tun, welches im Elfenbeinturm stattfindet. Wird dieser Argumentation gefolgt, findet sich etwas genuin Skandalöses an der Sache, dass auch Wissenschaftler:innen getrackt werden, wenn sie ihrer normalen Arbeit nachgehen. Ob eine Technik, wie hier das Tracking von Wissenschaftler:innen, »zum Heil oder Unheil gereicht, das liegt [...] am Gebrauch, den die Gesellschaft von ihr macht. Dieser Gebrauch ist keine Sache des guten oder bösen Willens, sondern hängt ab von der objektiven gesamtgesellschaftlichen Struktur.«¹⁰ Dies bedeutet insbesondere, dass Tracking von Wissenschaftler:innen zum normalen Geschäftsmodell gehören muss.¹¹ In einem profitorientierten System wie dem gegenwärtigen sollen die Bereiche erschlossen werden, die bisher noch nicht zugänglich sind. So war der Markt des Wissenschaftler:innen-trackings noch nicht in dem Maße erschlossen, wie er es bspw. für den der Nutzer:innen von Social Media war. Was den Nutzer:innen von Facebook, Instagram usw. schon lange widerfährt, wird nun auch auf Wissenschaftsverlage und -portale übertragen.

Was tun also? Wie gezeigt wurde, gibt es technische Möglichkeiten, der Tendenz zum Datentracking auf individueller Ebene entgegenzuwirken. Auf gesellschaftlich-kollektiver Ebene, die momentan eher durch eine unempörte Akzeptanz von Tracking- und Überwachungsmaßnahmen geprägt ist, sollten diese Praxen skandalisiert, analysiert und letztendlich wirkungsvolle Gegenmaßnahmen auf politischer Ebene entwickelt werden. Diese sollten nicht ausschließlich langfristig das große Ganze im Blick haben, sondern auch unmittelbare Verbesserungen möglich machen können. Dabei ist es auch an uns Wissenschaftler:innen, die gesellschaftlichen Verhältnisse in einer Form zu reflektieren, die angemessen diese Verhältnisse beschreibt, bewertet und verändert und nicht in Idealisierungen von einem inhaltsleeren Begriff der »Wissenschaftsfreiheit« verharret, die nie in der Form bestanden hat, die der Begriff gern suggerieren würde. Die Einbettung der Diskussion in eine tiefere, kritische Analyse der Produktionsver-

9 DFG: »Datentracking in der Wissenschaft: Aggregation und Verwendung bzw. Verkauf von Nutzungsdaten durch Wissenschaftsverlage«, 28.10.2021, DOI: 10.5281/zenodo.5900759 (aufgerufen: 12.5.2023).

10 Theodor W. Adorno: *Über Technik und Humanismus*, in: *Gesammelte Schriften*, hrsg. v. Rolf Tiedemann, Frankfurt am Main 1986.

11 Der Codex von Google »Don't be Evil« musste in logischer Konsequenz 2015 zu Grabe getragen werden.

hältnisse (welche die Produktion von Wissen miteinschließt) ist hierfür unabdingbar.