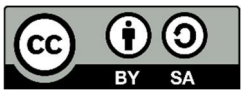
Data tracking in research: aggregation and use or sale of usage data by academic publishers

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

28 October 2021

Deutsche Forschungsgemeinschaft
Kennedyallee 40 · 53175 Bonn · Mailing address: 53170 Bonn
Phone: + 49 228 885-1 · Fax: + 49 228 885-2777 · postmaster@dfg.de · www.dfg.de

Deutsche Forschungsgemeinschaft e. V.
Kennedyallee 40 · 53175 Bonn
Mailing address: 53170 Bonn
Phone: +49 228 885-1
Fax: +49 228 885-2777
postmaster@dfg.de
www.dfg.de



October 2021

Contact:
Dr. Angela Holzer
Scientific Library Services and Information Systems (LIS)
Phone +49 (228) 885-2568
angela.holzer@dfg.de

As of: 28 October 2021

DOI: 10.5281/zenodo.5937995

Citation: Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme (2021): 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)

Data tracking in research

This briefing paper issued by the Committee on Scientific Library Services and Information Systems (AWBI) of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) on the subject of data tracking in digital research resources describes options for the digital tracking of research activities. It outlines how academic publishers are becoming data analytics specialists, indicates the consequences for research and its institutions, and identifies the types of data mining that are being used. As such, it primarily serves to present contemporary practices with a view to stimulating discussion so that positions can be adopted regarding the consequences of these practices for the academic community. It is aimed at all stakeholders in the research landscape.

1. Description of the current situation

In recent years, digital data markets of various kinds have emerged which can be categorised as either public, academic or commercial.[1] In the field of academia, in addition to very positive developments such as improved handling, legal regulations for research data and also improved use of research data, other developments need to be considered in detail and, if necessary, subjected to regulatory control. These developments are presented below. A potentially detrimental impact on the academic community arises both from the mixing of academic and commercial interests as well as from regulatory gaps and differing legal situations internationally.

For some time now, the major academic publishers have been fundamentally changing their business model with significant implications for research: aggregation and the reuse or resale of user traces have become relevant aspects of their business.[2] Some publishers now explicitly regard themselves as information analysis specialists.[3] Their business model is shifting from content provision to data analytics. This involves the tracking – i.e. recording and storage – of the usage data generated by researchers (i.e. personalised profiles, access and usage data, time spent using information sources, etc.) when they utilise information services such as when carrying out literature searches. Research tracking is carried out using an ensemble of tools ranging from tracking site visits via authentication systems to detailed real-time data on the information behaviour of individuals and institutions. The recording of such information as page

[1] Putnings, M., „Datenmarkt“, in: *Praxishandbuch Forschungsdatenmanagement*, 2021, p. 143, Praxishandbuch Forschungsdatenmanagement (degruyter.com).
[2] Aspesi, C., Allen, N. S., Crow, R., Daugherty, S., Joseph, H., McArthur, J. T., & Shockey, N., *SPARC Landscape Analysis*, 2019, March 29, https://doi.org/10.31229/osf.io/58yhb.
[3] For example, self-presentation of Elsevier: “Elsevier is a global information analytics company that helps institutions and professionals improve healthcare and academic performance to the benefit of humanity.” https://www.elsevier.com/de-de/about.

visits, accesses, downloads – including assembling granular profiles of academic behaviour – is sometimes carried out without users being sufficiently informed of the process. Data from different sources can be aggregated and combined with additional information about the individuals, including details drawn from the non-academic sphere.

There are two reasons why publishers collect this data: firstly, the aim is to tap into a new business field that enables data about knowledge, research developments and the relevant stakeholders to be used as an economic asset. Secondly, the aim is to expand the range of services offered by major academic publishers. The data can be used to improve existing services. For example, researchers can automatically receive targeted suggestions for reading and references to research results within their field based on personal profiles. It is also possible to develop new services in this way. In addition to providing and managing research findings in the form of academic literature, services are also increasingly being offered in the area of research data management and research software.

These different services could be linked to each other, making them convenient for researchers to use. For numerous activities within the research cycle – or indeed all such activities – researchers can use services of one provider, who, in addition, can offer institutions particular services (e.g. research information systems). For example, RELX – the parent company to which Elsevier belongs – is establishing the research information systems software PURE at universities around the world, explicitly indicating that it is able to provide insights into the entire research cycle.[4] [5]

This development has the potential to significantly interfere with the anonymity of researchers as fundamentally guaranteed under data protection law, thereby making research institutions jointly responsible for violating the right to informational self-determination. Data tracking also potentially encourages data misuse and academic espionage and can result in personal dis-

[4] See the description of the service at: www.elsevier.com/solutions/pure.

[5] Elsevier declared on September 22, 2021: “PURE is a tool that institutional customers use to process their data. The data always belong to the customer and when the contract is terminated, the customers receive their data back (with cloud hosting) or they remain with the customer (on-premise). Elsevier does not acquire any rights to this data and does not use it for any purpose. If the customer chooses, Elsevier will not have access to a given PURE installation, even when it is hosted in the cloud. PURE and indeed all Elsevier products and services are GDPR compliant. The software offers all options for GDPR-compliant processing of personal data if it is configured correctly. As of summer 2021, both Elsevier’s hosting partner Amazon and PURE are certified according to ISO 27001. In addition, Elsevier concludes data processing agreements with customers that take local requirements into account. A strong Data Protection Agreement is in place with each university when Elsevier serves customers in this way. The data customers place in PURE is isolated from other PURE customers and from Elsevier’s use in general unless the university chooses to activate data sharing functions that they control. Like all software products, instrumental data tells Elsevier if our software is functioning correctly.”

crimination against researchers. In view of the current jurisdiction of the German Federal Supreme Court, the Schrems II ruling and the upcoming draft of an EU platforms law (Digital Markets Act)[6] , research organisations should adopt a position on these practices.

The German Federal Government's recently presented data strategy does not address this situation specifically, but it does mention the problem in principle – namely increasing monopolisation[7], abuse of market power and misuse of data: "In the use of data, not everything that is technically possible is ethically justifiable and politically desirable." [8]

All in all, researchers face the difficulty of striking a balance between being able to enjoy conveniently bundled services and maintaining control over their data. In many cases, researchers are not aware of the significance of data about their activities and the way in which it is used as an economic asset. The relationship between academia and publishers needs to be explored, with the aim of establishing a balance between these two poles of convenience and control. However, this can only be successful if it is based on clear-cut legal regulations that ensure a high degree of transparency and the participation of the academic community.

2. The transformation of the major publishers and their relationship with the academic community

Publishers began incorporating personal identification authentication solutions and user tracking in their services some time ago. By doing so, they are able to offer technical proprietary services for the entire research process and the analysis of research-related data. One example is the 2020 contract with Elsevier in the Netherlands, in which services described as Professional Services and the collection of personal data are contained.[9] Some publishers also support the SeamlessAccess or the GetFTR strategies[10] aiming to enable the major research providers to make information available in a way that is as self-contained as possible, based on straightforward, one-time authentication.[11] GetFTR and SeamlessAccess offer information

[6] For example, the Digital Markets Act explicitly addresses the objective that the data collected should serve not only the intermediaries but also promote competition and the public interest.
[7] Datenstrategie der Bundesregierung, Kabinettsfassung dated 27 January 2021, p. 21, Datenstrategie der Bundesregierung und die Ausschreibung des Bundesministeriums für Bildung und Forschung für Datentreuhandmodelle in den Bereichen Forschung und Wirtschaft vom 08.01.2021, Bekanntmachung – BMBF.
[8] Datenstrategie der Bundesregierung, Kabinettsfassung dated 27 January 2021, p. 7, Datenstrategie der Bundesregierung.
[9] Signed UKB Elsevier SD 2020-2024 agreement.pdf (vsnu.nl). We refer especially to Schedule 5, but also to section 7.6. of the contract.
[10] See www.getfulltextresearch.com and https://seamlessaccess.org
[11] Moore, S. A., "Individuation through infrastructure", in: *Journal of Documentation* 77(1) dated 28 July 2020, https://doi.org/10.1108/JD-06-2020-0090.

on what data they collect and how they address user privacy.[12][13] After initial critical feedback from librarians, adaptions have occurred.[14]

German research organisations recently concluded DEAL agreements with major academic publishers (Springer Nature[15] and Wiley[16]) to achieve open access and appropriate prices for the provision and publication of research results. When contracts are concluded with publishers, it is always important to carefully review agreements concerning data privacy and the access and authentication systems[17] that are to be used. Essentially, the most convenient access is where no authentication is required at all, i.e. open access, although here again it is possible to trace usage via publishing platforms. In addition to literature access, many institutions are also bound to a particular software, for example as supplied by a provider such as Elsevier.[18] Elsevier is also a subcontractor working on behalf of the European Commission to collect data on Open Science (Open Science Monitor).[19]

Such wide-ranging services offer the opportunity to gain insights into as many phases of the research process as possible and market these to third parties: ultimately, this makes publishers or companies capable of providing research, politics, universities and society at large with the most comprehensive, data-based information about research activity. It also means that publishers are becoming indispensable for the governance of academic institutions and universities. There is already talk of an emerging “supercontinent”[20] in the supply of research information and information about research. Some data on research activity can be useful for research itself as well as for the complex governance processes that modern research involves. Good practice is when, for example, the regulations on data collection, data use and

---

[12] GetFTR: GetFTR | Why GetFTR - GetFTR (getfulltextresearch.com) in FAQ no.7: https://www.getfulltextresearch.com/why-use-getftr/

[13] SeamlessAccess: Privacy and Trust - SeamlessAccess: https://seamlessaccess.org/about/trust/

[14] Hinchcliffe, L.J.: “Why are Librarians concerned about GetFTR?”, in: *The Scholarly Kitchen* dated 10 November 2019, https://scholarlykitchen.sspnet.org/2019/12/10/why-are-librarians-concerned-about-getftr; Youngen, Ralph, Toler, Todd: "Lessons Learned: A Year with GetFTR”, in: The Scholarly Kitchen dated 16 February 2021, https://scholarlykitchen.sspnet.org/2021/02/16/guest-post-lessons-learned-a-year-with-getftr/

[15] Kieselbach, S., *Projekt DEAL – Springer Nature Publish and Read Agreement*. 2020, https://doi.org/10.17617/2.3174351.

[16] Sander, F., Herrmann, G., Hippler, H., Meijer, G., & Schimmer, R., *Projekt DEAL – John Wiley & Son Publish and Read Agreement,* 2019, https://doi.org/10.17617/2.3027595.

[17] Stellungnahme des Deutschen Bibliotheksverbands „Empfehlungen zu Methoden zur Kontrolle des Zugriffs auf wissenschaftliche Informationsressourcen“, www.bibliotheksverband.de/fileadmin/user_upload/DBV/positionen/2019_11_26_Rundgespaech_RA21_-_Stellungnahme_Empfehlungen_final.pdf.

[18] Cf. Elsevier’s list of institutions using its research information system software Pure, https://www.elsevier.com/solutions/pure/clients.

[19] See for: Microsoft Word – Open Science Monitor Methodological Note_April 2019.docx (europa.eu).

[20] Schonfeld, R.C.: “The Supercontinent of Scholarly Publishing?, in: *The Scholarly Kitchen* dated 3 May 2018, https://scholarlykitchen.sspnet.org/2018/05/03/supercontinent-scholarly-publishing.

data sharing are transparent and clear and data is also available for non-commercial purposes to stakeholders within the academic infrastructure itself (e.g. at CrossRef).

The consequences of this "data-driven organisation of research"[21] , the conditions for achieving it and the structures that provide, sell and utilise it must ultimately be reflected upon and shaped by research itself. Research organisations should advocate that data collection and use – where necessary – is not only legal but also informed by ethical values such as transparency and traceability, as well as being based on consent with full disclosure of the consequences along with other aspects of good data practice, ensuring that such data practice forms the basis for any agreement with providers.

2.1 Consequences of the transformation of publishers into data analytics businesses

There is a risk that this shift in the commercial business model towards data analytics will result in the knowledge society becoming privatised, and that ultimately it will no longer be the public sector but increasingly private companies that are privy to knowledge about research content and trends, its institutions and stakeholders. Research as a public asset is subjected to the logic of infrastructure privatisation and the consequences this entails.[22] Such a business models involves not just large publishers but also smaller-scale providers of research databases. Various studies and initiatives – including the 2012 call for "The Cost of Knowledge" as well as organisations such as Science Europe[23] and library associations – have repeatedly drawn attention to this far-reaching increase in the volume of information and data held by private-sector companies and the fact that such a concentration of knowledge about research is not only beneficial to innovation in the field of research information provision.[24]

---

[21] Herb, U.: „Zwangsehen und Bastarde", in: *Information. Wissenschaft & Praxis*, 69 (2-3), 2018, p. 87.

[22] Barlösius, E., *Infrastrukturen als soziale Ordnungsdienste. Ein Beitrag zur Gesellschaftsdiagnose*. Frankfurt/M. 2019, Chapter 6.4: „Infrastrukturierung der Forschung und infrastrukturierende Forschung".

[23] "Science Europe calls for a clear exclusion of data users and usage for the purposes of f from the scope of the Digital Services Act, to ensure that unintended effects on research activity are avoided. A legislative act that aims to address the selling of illegal content on large commercial platforms could have side effects on sectors of public interest unless proper exceptions are introduced." Science Europe, The Digital Services Act Should Not Have Unintended Effects on Research, 2020, www.scienceeurope.org/media/4s3bnhbr/20200908_se_response_dsa_consultation_final.pdf.

[24] For example Dobusch, L., "Kein Open-Access-Deal, dafür Spyware gegen Schattenbibliotheken", in: *netzpolitik.org*, dated 26 November 2020, https://netzpolitik.org/2020/neues-vom-grossverlag-elsevier-kein-open-access-deal-dafuer-mit-spyware-gegen-schattenbibliotheken/; die Stellungnahme des Deutschen Bibliotheksverbands „Empfehlungen zu Methoden zur Kontrolle des Zugriffs auf wissenschaftliche Informationsressourcen", www.bibliotheksverband.de/fileadmin/user_upload/DBV/positionen/2019_11_26_Rundgespaech_RA21_-_Stellungnahme_Empfehlungen_final.pdf; .

The development outlined here towards a private-sector knowledge industry[25] stands in opposition to the freedom of research as well as to the legally prescribed handling of personal data and competition law. More specifically, unregulated or undetected data tracking can

- entail a violation of academic freedom and the freedom of research and teaching;
- constitute a violation of the right to the protection of personal data;
- pose a potential threat to scientists, as the data could also become accessible to foreign governments and authoritarian regimes;
- constitute an encroachment of competition law, as new participants barely have a chance to enter the market;
- favour a reduction in the value of public research investment, since data on research activity can be collected by commercial research competitors or made available to them in return for payment in connection with industrial espionage.

The first cases of trading with data on the research interests of individual scientists illustrates just how critical this industrialisation of knowledge through tracking has already become.[26] LexisNexis, an international information solutions provider and subsidiary of the RELX Group – which includes Elsevier – has signed a deal to hand over personal data to ICE, the US Immigration and Customs Enforcement agency, for $16.8 million.[27] The situation is often further exacerbated by the fact that higher education institutions and libraries can become complicit in violating data protection law, academic freedom and competition law without their knowledge. The behavioural data profiles of German university staff can be traded and transferred in the same way, which led to the overturning of the Privacy Shield in the Schrems II ruling, i.e. the transfer of personal data to a third country outside the EU, since the same stakeholders are involved.[28] In addition, risks could potentially arise from the major publishers presenting a censored programme on the Chinese market. Tracking could also result in personalised data being generated about who uses and recommends the censored documents, without it being possible for the researchers concerned to assess who is being given access to this

[25] Burgelman, J-C.: "Scholarly publishing needs regulation", in: *Research Professional News*, dated 28 January 2021, www.researchprofessionalnews.com/rr-news-europe-views-of-europe-2021-1-scholarly-publishing-needs-regulation.

[26] Jung, J.: "UCLA School of Law Holds Contract with Companies Selling Personal Data to ICE", in: *The Daily Bruin* dated 17 July 2020, https://dailybruin.com/2020/07/17/ucla-school-of-law-holds-contracts-with-companies-selling-personal-data-to-ice.

[27] Biddle, S.: "LexisNexis to Provide Giant Database of Personal Data to ICE", in: *The Intercept* dated 2 April 2021, LexisNexis to Provide Giant Database of Personal Data to ICE (theintercept.com).

[28] Cf. Bundesland Niedersachsen, *Das SchremsII-Urteil des Europäischen Gerichtshofs und seine Bedeutung für Datentransfer in Drittländer*, 2021, https://lfd.niedersachsen.de/startseite/themen/weitere_themen_von_a_z/internationaler_datenverkehr/das_schrems_ii_urteil_des_eugh_und_seine_bedeutung_fur_datentransfers_in_drittlander/das-schrems-ii-urteil-des-europaischen-gerichtshofs-und-seine-bedeutung-fur-datentransfers-in-drittlander-194085.html.

tracking data. In response to possible amendments to the legislation, Google recently announced a change in its tracking policy, for example: in future it is to be organised more anonymously and based on which “cohorts” are identified and addressed rather than individual users. [29]

[29] Neue Spielregeln: Warum Google Cookie-Tracking abschafft (netzpolitik.org)

3. Types of data mining

There are potentially three different types of data mining, i.e. methods by which data could be collected and stored by publishers (third-party data through microtargeting, bidstream data and port scanning, and "spyware"), which are described below. There are also different tools: the portfolio currently in use in research includes page visit trackers, audience tools for aggregating different data sources into profiles, fingerprinters that identify even those users who seek to prevent identification through browser settings, and tools for real-time auctioning of user data. The tracking tools are mostly produced by third-party providers under contract to the major internet companies, but also by specialised companies such as BlueKai, the Big Data platform belonging to Oracle, which is itself the subject of class action lawsuits for misuse of personalised data.[30] Since it is already institutionally linked to other data aggregators operated by internet services, the data can be condensed into profiles and combined with further data from other areas of life.[31] The publishers do not disclose how deep their tracking goes, so at the moment we can only refer to various tests[32] showing that anyone who accesses an article in the journal *Nature*, for example, is tracked by more than 70 instruments.[33] Finally, the tools used can be flawed, resulting in even more detrimental consequences for individual researchers.[34] The three main types of data mining mentioned above are briefly outlined below. All in all, it can be assumed that research tracking instruments will be constantly refined and expanded in their application since they afford suppliers and corporations with considerable competitive advantages.

3.1 Third-party data through microtargeting

Microtargeting is the addressing of very specific target groups. Both first-hand and second-hand data are used by publishers. First-hand data consists of direct user traces, while second-hand data is purchased data, which in turn is condensed into precise data profiles by third

---

[30] Lomas, N.: "Oracle and Salesforce Hit with GDPR Class Action Lawsuits Over Cookie Tracking Consent", in: *TechCrunch* dated 14 August 2020, https://techcrunch.com/2020/08/14/oracle-and-salesforce-hit-with-gdpr-class-action-lawsuits-over-cookie-tracking-consent/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+Techcrunch+%28TechCrunch%29&guccounter=1.

[31] Vogel, C.: „Kennen Sie Google CASA?", in: *Medinfo. Informationen aus Medizin, Bibliothek und Fachpresse*, www.medinfo-agmb.de/archives/2020/07/08/6880.

[32] "Digital Library Federation, Endangering Data. Interview with Sarah Lamdan" see for www.diglib.org/endangering-data-interview-with-sarah-lamdan or Lamdan, S.: "Social Media Privacy: A Rallying Cry to Librarians", in: *The Library Quarterly* 85 (3), 2015, p. 261-277 https://academicworks.cuny.edu/cl_pubs/52; Wolfie Christl's studies on RELX and ThreatMetrix, https://twitter.com/wolfiechristl/status/1295655040741445632 and https://crackedlabs.org.

[33] Brembs, B.: https://twitter.com/brembs/status/1301897878387003398.

[34] See for example Lamdan, S.: "Librarianship at the Crossroad of ICE Surveillance", in: *In the library with the lead pipe* dated 13 November 2019 and Swauger, S., https://twitter.com/SheaSwauger/status/1205587676172144641.

parties, mainly the large internet companies. Publishers have established a wide variety of these *third-party asset sources* on their platforms, be it the widespread trackers used by Google or Facebook, those used by providers such as BlueKai and Krux Digital, browser fingerprinting tools such as Double Click or data-aggregating audience tools by Adobe, Neustar, Oracle, AddThis and others. The third-party Javascript code can access the Document Object Model of the website in question, so it is able to read out which text the user engages with, which text they browse to next and which search words they enter on the platform. Since many providers include the same third parties to some extent or these exchange data with each other in some cases, the information behaviour of university members can be collected across platforms and, in the case of Google, Facebook or Twitter, can be linked to knowledge already available about their other online activities.[35] In the case of providers such as Acxiom/Liveramp, online and offline activity can be synchronised too, since data is also available regarding purchases, driving licences, TV consumption, electoral rolls, criminal records and the like.[36]

3.2 Bidstream data and port scanning

The integration of third parties in websites is often criticised and in some cases is no longer supported by major internet companies and institutions, so alternatives such as the harvesting of bidstream data (real-time bidding data) are currently used as well, i.e. background collection of data about location, devices and data used. User data is auctioned on a real-time basis, including a variety of individual items of information such as localisation data, IP number, device details and much more; this is then transmitted and linked to an identifier so as to be able to reliably identify individuals even without a cookie.[37] Simply searching for open ports on other people's computers and/or networks in order to infiltrate malware or surveillance software, for example, borders on the illegal under German law since it can be considered a preliminary stage of certain sanctioned offences (§§ 202c, 303b German Criminal Code – StGB). This method is still widely used nonetheless, partly for the purpose of fraud prevention and partly as a tracking tool.

One example that has attracted public interest is ThreatMetrix, part of LexisNexis Risk Solutions/RELX, which claims to be able to identify 4.5 billion devices. ThreatMetrix is implemented on ScienceDirect, for example, the platform through which researchers consult the content of

---

[35] Hanson, C.: *User Tracking on Academic Publisher Platforms*, 2019, www.codyh.com/writing/tracking.html.
[36] Cf. the graphic in Christl, W.: *Corporate Surveillance in Everyday Life*, p. 55, https://crackedlabs.org/dl/CrackedLabs_Christl_CorporateSurveillance.pdf.
[37] Cf. Ryan, J.: *Briefing on adtech, RTB, and the GDPR at dmexco Brave Event*, Slide 45, www.slideshare.net/JohnnyRyan/briefing-on-adtech-rtb-and-the-gdpr-at-dmexco-brave-event.

journals published by Elsevier. The RELX Group's connection to various government agencies in the US is already the subject of public petitions in the US.[38] As long as publishers do not reveal their tracking practices, it remains a matter of speculation whether data collected using such trackers is also used in connection with other products of the Risk Solutions division[39], e.g. in the area of analyses for companies and authorities.[40]

3.3 "Spyware"

If offered to libraries in connection with discounts on other services, "spyware" can serve the purpose of scaling research tracking. "Spyware" is additional software to be installed in the libraries which collects biometric data such as typing speed or type of mouse movement in order to be able to personalise users despite the use of proxy servers and VPN tunnels.[41] Organisations might argue that such software allows for the prosecution of users of shadow libraries.[42][43] However, such "spyware" undermines the security of university networks and potentially exposes universities to all kinds of attacks. Its use can therefore not be recommended.[44]

---

[38] American Civil Liberties Union: *ACLU Calls On Tech Companies to End Their Alliance with ICE and CBP*, 2020, www.aclu.org/news/immigrants-rights/aclu-calls-on-tech-companies-to-end-their-alliance-with-ice-and-cbp.

[39] Risk & Business Analytics – RELX

[40] Cf. the documentation by Wolfie Christl at: https://twitter.com/wolfiechristl/status/1286341387718397952

[41] Cf. Mehta, G.: "Proposal to Install Spyware in Universities Libraries to Protect Copyrights Shocks Academics", in: *Coda* dated 13 November 2020, www.codastory.com/authoritarian-tech/spyware-in-libraries.

[42] The German equivalent – *Schattenbibliotheken* – has become an established term and is used here for example: Ball, R.: *Wissenschaftskommunikation im Wandel*, Springer, 2020, p. 127.

[43] In an earlier version of this paper, PSI was referenced and that reference has been deleted in this version due to a declaration by PSI on June 9, 2021:

"1. *PSI does not work with SNSI in any capacity.*
2. *PSI does not track users of shadow libraries.*
3. *PSI has no involvement with or knowledge of any spyware whatsoever.*
4. *PSI has not had any involvement in the prosecution of "users" of shadow libraries and do not believe "users" have been prosecuted by anyone.*"

[44] In an earlier version of this paper, the Scholarly Networks Security Initiative (SNSI) was mentioned in this paragraph. This mention has been deleted due to a declaration by SNSI on September 8, 2021:

"*SNSI strongly encourages its customers to maintain strong security over access to the data of researchers and students and the content provided to them by its members but it does not:*

- *Operate or advocate the use of spyware (e.g. to collect biometric data such as typing speed or type of mouse movement in order to be able to personalise users despite their use of proxy servers and VPN tunnels);*
- *Provide or advocate for incentives to libraries to install and operate their own spyware.*"

4. Conclusion

Potentially, research tracking of this kind can fundamentally contradict academic freedom and informational self-determination. It can endanger scientists and hinder the freedom of competition in the field of information provision. For this reason, scholars and academic institutions must become aware of the problem and clarify the legal, technical and ethical framework conditions of their information supply – not least so as to avoid involuntarily violating applicable law, but also to ensure that academics are appropriately informed and protected.

AWBI's aim in issuing this briefing paper is to encourage a broad debate within the academic community – at the level of academic decision-makers, among academics, and within information infrastructure institutions – so as to reflect on the practice of tracking, its legality, the measures required for compliance with data protection and the consequences of the aggregation of usage data, thereby enabling such measures to be adopted.

The collection of data on research and research activity can be useful as long as it follows clear-cut, transparent guidelines, minimises risks to individual researchers and ensures that academic organisations are able to use such data if not have control over it.