Digital Privacy

Dr. Robert Altschaffel, 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[1]

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.

[1] 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 Altschaffel, 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, FIZ 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

The Maxwell-Garfield system of academic publishing is structurally asymmetrical, putting publishers in the driver's seat and research and its libraries in the passenger seat: it is now regarded as an obstacle to the global knowledge society that is in urgent need of reform (Gulliver & Drake 2024). While already problematic in itself, this asymmetry is further exacerbated by the shift in the corporate strategy pursued by academic publishers towards that of a data analytics business. This new business model increasingly makes use of real-time data relating to the interests and behaviour of universities, institutes and individual researchers, a phenomenon which we will refer to here as "data tracking". This once again shifts market weightings to the detriment of science and the humanities (Benedikt & Schwartmann 2021; Schonfeld 2023). Not only does this data analytics business open up new markets for scientific publishers, it is also suspected of violating European data protection law and the academic freedom of individual researchers (Siems 2022).

All this is particularly sensitive for libraries: after all, it is they who provide the interface and the authentication mechanisms through which details of researchers' interests and personal data collected during the research process goes to publishers and is potentially passed on by the latter to third parties (Lamdan 2022; Siems 2024). For example, the latest SPARC report concludes that Elsevier's platform *Science Direct* is in conflict with digital privacy and is therefore not sufficiently compatible with library data protection standards (Yoose & Schockey 2023). In view of this situation, academic libraries have a particular responsibility to bear. As things currently stand, the proposal of a "Plan S" (Stern & Rooryck 2023) is no more than an indication of the direction being pursued in research policy (cOAlition S 2023; Kiley 2023; Posenato 2023), i.e. responsible publishing based on a fundamental transformation of academic publishing in Europe towards a scholarlyled system and a switch in the field of academic publishing from the shareholder-oriented principle to a stakeholder-oriented principle. The International Science Council (2021) and, more recently, the Bill and Melinda Gates Foundation (Drake 2024) have set out an even more far-reaching concept for self-regulated publishing in science and the humanities.

The recent DEAL negotiations between academic publishers and science organisations were not in the position to resolve the widespread impasse brought about by the Maxwell-Garfield system of academic publishing; however, the negotiations have enabled contractual regulation of data tracking in such a way that applicable law is better complied with and the sharing of responsibilities between publishers and libraries is handled in a legally compliant manner. As such, the negotiations also sought to contribute to transparency, firmly enshrine data tracking aspects in the DEAL contracts, and ensure co-determination on the part of the scientific community.

II. Data tracking techniques

An initial overview of the phenomenon of data tracking in research is to be found in the information paper on data tracking published by the DFG's Committee for Scientific Library Service and Information Systems. Data tracking involves "the data of researchers [...] being tracked, i.e. recorded and stored, when they use information services such as literature research" (DFG 2021). The data collected and stored by publishers is often used for other purposes, such as security assessments or further analyses, which results in the creation of personalised profiles on researchers, including access and usage data, time spent on information sources, etc.

Data is collected as a result of being transmitted through the communication network – usually when a browser is used, but also when apps or e-mail are used. A distinction has to be drawn between the flow of data to the provider of a service, such as the publisher, and the flow of data to a third party, such as booking or payment systems, or Google advertising systems. The former is referred to as *first-party tracking*, while the collection of data by external services is referred to as *third-party tracking*. The involvement of third parties in this way is not always transparent to the user and may not even be obvious to the service provider. For example, the integration of third-party services for the provision of certain fonts ("Google Fonts") is widespread: it results in the third-party service being accessed and the data transfer that this involves. In both cases, it is sometimes only possible to evade or intervene in tracking to a limited extent or not at all (Bettinger, Bursic & Chandler 2023).

Various forms of tracking are conceivable, depending on type and extent. Some of these are as follows:

- Tracking of access to websites and their content based on personalised parameters. This involves passing on information about who views a particular document and when.
- Tracking within documents, e.g. using additional software modules such as tracking pixels (also known as "beacons"). Data is regularly transmitted for this purpose, keeping the other party up to date as to whether a document is still being viewed and by whom.
- Usage analysis: While a document is being viewed, the data is transmitted in real time, e.g. which page is currently being viewed and when it is changed.

Even if much of the content of communication can only be observed to a limited extent from the outside or sometimes not at all due to encryption or coding, some properties of these data transmissions can still be observed by those being tracked (Altschaffel, Kiltz, Lucke & Dittmann 2020; Kiltz, Altschaffel & Dittmann 2023). It is possible to monitor the volume of data exchanged, the number of communication transactions and the third parties involved. This identification can also be utilised by potentially tracked users to determine the physical or logical position of third-party providers. In addition, there are databases that attribute network addresses to the address ranges of countries or internationally active companies (e.g. the Whois database; Bettinger, Bursic & Chandler 2023). What monitoring the transmitted data from the outside cannot do is provide the tracked party with information on whether and for how long this data is stored or forwarded to other providers. Answering this question would require a look at the providers' infrastructure.

Based on the findings of studies conducted by Cody Hansen (2019), Renke Siems (2021; 2024) and Michael Freiberg (2022), it can be generally stated that both first-party tracking and third-party tracking are to be found in the services provided by academic publishers, the latter primarily in the form of integrated media delivery services (content delivery networks, CDNs), integrated social media or external analytics platforms. This area offers particular potential in terms of reducing data tracking.

Two other aspects need to be considered in order to better understand the depth of science tracking: the practice of forwarding or embedding and the so-called CNAME hiding. Embedding is where accessing a search result or link automatically redirects to another provider so that the researchers' data can be subjected to further analysis. The second method (CNAME hiding) uses a hidden technique that makes identification and tracking more difficult: a CNAME is an alternative name for an address entry on the internet – a service whose actual URL was xyz.de could also be displayed under abc.de, for example, thereby hiding the tracking from the user's perspective (this is a fictitious example, but it is a reference to a concrete example documented on 07 October 2023). All these techniques are suitable for tracking the behaviour of researchers (Kiltz, Altschaffel & Dittmann 2023). Some of the publishers' tracking techniques can be analysed using a digital fingerprint ("science tracking finger print"). Techniques installed by publishers such as CNAME hiding and additional software components such as tracking pixels can be systematically identified and subjected to critical analysis. It would then be possible to show where publishers should avoid data tracking and how it might be possible to bring about a fundamental improvement of data tracking in academic publishing.

Finally, it should be noted that the technical implementation of tracking technologies is subject to ongoing advancement, which means that research is required into new forensic technologies. One challenge in the near future is likely to be switching the system-wide address resolution system (DNS) to application-specific encrypted address queries such as DNS over HTTPS (DoH).

III. Data tracking and library practice

Academic libraries in Germany have long regarded themselves as mediators and as preservers of information and knowledge, but this task is now increasingly applicable in the context of data tracking. While academic libraries have been concerned with protecting usage data in library systems for many years, the complex issue of usage data on publishers' platforms is still frequently a new challenge: from being a tangible issue for all library employees, it has only recently been addressed at European level with regard to the democratic function of libraries (EU 2023). In practice, virtually no successful measures have been adopted against data tracking to date that libraries might adopt as a model into their practice.

One reason for this is that the major academic publishers continue to be very important to science and the humanities and therefore to the library business. These publishers control the publication and distribution of most of the media relevant to research, which in turn are licensed by the libraries. Libraries – or the library consortia acting on their behalf – do address the issue of data tracking in licence negotiations, but provisions on this issue have rarely been included in licence agreements. In most recent contracts between publishers and libraries, the publishers retain control over the data generated by the use of their media. Libraries remain dependent on publishers and their business models, and for this reason they appear to be able to take only limited measures to protect user data.

Data tracking has not been a major issue internationally to date. In the European Union in particular, data privacy law is largely harmonised thanks to the GDPR, but there is no majority position on data tracking in research, let alone a unanimous one. Individual initiatives exist that are directed against science tracking, such as SPARC, but clear guidelines and specifications as to how to handle data are only just beginning to emerge.

IV. Legal framework for data tracking

Data tracking is subject to mandatory legal requirements: regardless of the basis of the agreement, all data subjects must always be fully informed (Art. 12-14 GDPR) and rigorous standards apply to the security of data processing (Art. 33 GDPR).

Furthermore, the processing of personal data must be explicitly permitted. This is the case when the data processing is necessary to safeguard legitimate interests of the data controller or a third party (Art. 6 (1)(f) GDPR), for example to prevent copyright violations or ensure system security. However, one may only employ the least intrusive means to this end, and the data may not be used for any other purposes. Beyond this indispensable use, data tracking is permitted if the data subjects have given their express and voluntary consent (Art. 6 (1)(a) GDPR) before the processing. "Voluntariness" is lacking whenever a service can only be accessed after giving consent to the use of personal data beyond the information necessary for providing the service (Art. 7 (4) GDPR). "Explicitness" is lacking, among other things, if permission is granted by pre-checked checkboxes (Recital 32 p. 3 GDPR) – the choice to opt out is not sufficient, users must express an explicit will to opt in. Specifically for cookies, Art. 5 (3) ePrivacy Directive (implemented in Section 25 TDDDG – German Act on the Protection of Personal Data and Privacy in Telecommunications and Digital Services) adds an additional consent requirement. Finally, even stricter requirements apply if data is to be transferred to a country outside the EU (Art. 45 et seq. GDPR). It is not completely unlikely that users will, however, freely agree to the processing of their personal data– after all, personalised search results and individual suggestions can significantly improve the usability of databases.

One might assume that those legal requirements would only concern the publishers. But according to ECJ judgements starting in 2018 (ECJ judgement dated 05 June 2018, Case C-210/16 – Facebook Fan Pages Recital 36 f.; ECJ judgement dated 10 July 2018, Case C-25/17 – Jehovah's Witnesses Recital 73; ECJ, judgement dated 29 July 2019, Ref. C-40/17 – Fashion ID Recital 80; ECJ judgement dated 05 December 2023, Case C-683/21 – Nacionalinis visuomenes sveikatos centras Recital 43 f.; ECJ judgement dated 07 March 2024, Case C-604/22 – IAB Europe Recital 77) academic institutions (which conclude the contracts and are therefore acting through their libraries) can be held jointly responsible (Art. 26 GDPR) if they (1) enable the data processing by the publishers and (2) benefit themselves (even indirectly) from the data processing. Libraries provide researchers with access to platforms and databases for their institutions, and thereby grant publishers the means to personal data of their users. Without the aid of the institutions, user access to the databases and data-gathering would be far more complicated or even impossible. The academic institutions in turn receive statistics from the publishers, some of which require user identification, i.e. data tracking – and thereby profit from the act of data processing. Although German law exempts public entities (i.e. purely state-owned institutions) from administrative fines and criminal sanctions, private institutions (such as foundations) may be fined under Art. 83 GDPR. Data subjects can also assert claims for damages (Art. 82 GDPR) and injunctive relief (Art. 79 GDPR). Both public and private institutions are obliged to protect their employees from harm (Section 241 (2) BGB – German Civil Code, and Section 45 (2) BeamtStG – German Civil Servants' Status Act), i.e. institutions have a protective duty towards researchers also covering personal data. This applies in particular where access to necessary resources is subject to limitations of their rights, e.g. by requiring exclusive access via electronic databases that process personal data not necessary for their operation.

In the run-up to the DEAL contract negotiations, the DFG set up a working group on data tracking in research. Its purpose was to make use of ongoing contract negotiations to initiate discussion on the extent of data tracking and appropriate assignment of responsibility. The working group also sought to clarify how decisions can be placed in the hands of users and what specific requirements to impose on data tracking. The group was further requested to ensure that liability lies with the publishers to the farthest extent possible. As consent can only be granted personally by the individual users, the group focused on preventing limitations on the legal requirements in the DEAL-contracts.

V. Legal opinions and draft joint controllership arrangement

To clarify the legal framework, the DFG commissioned two legal opinions from Härting, a law firm specialising in IT law (https://haerting.de). The first opinion discussed the general sharing of responsibility and the legal limits of data tracking, while the supplementary opinion focussed on ways of limiting liability, making specific recommendations on the DEAL negotiations which at that time were yet to commence.

The reviews clarified that, in principle, both the institutions and the publishers are liable to the researchers for data privacy violations and can therefore be sued independently of each other. Compensation can only be sought internally through recourse by the institution against the publisher.

Therefore, the reviews suggested drawing up detailed provisions on joint responsibility in the form of a "joint controllership arrangement" (JCA, Art. 26 (1)(2) GDPR), the substance of which was also to be made available to all library users of the licensed platforms (Art. 26 (2)(2) GDPR). The option of determining publishers as mere data processors (Art. 28 GDPR) was rejected as the institutions lack any significant control.

Another approach discussed in the reviews was certification of the publishing platforms by independent third parties (e.g. in accordance with Art. 42(f) GDPR). Although such certification would not result in a general exclusion of liability, it would ensure that the institutions could not be accused of negligence in damage cases; this would enable a defence against private claims, albeit an uncertain one.

In line with the recommendations of these legal opinions, the DFG working group proposed data tracking in research to be prohibited as far as possible.

To facilitate negotiations, the working group formulated a sample joint controllership arrangement (JCA) based on the legal opinions. This agreement would have to be signed not only by MPDL Services gGmbH and the respective publisher but also by each individual organisation as a co-responsible party technically involved in the data processing (Art. 26 (1)(1) GDPR). The proposed JCA tried to construct the cases of joint responsibility as narrowly as possible by specifically defining those instances in which the publishers transmit data to the institutions. In addition, the working group formulated requirements which also applied to third-party companies engaged by publishers as data processors (e.g. for hosting, embedded libraries, etc.) – which included e.g. information obligations as well as security and confidentiality under the GDPR. The proposed JCA would have further required Publishers to carry out an annual data protection impact assessment (Art. 35 GDPR), independently of legal obligations, and make the results available to MPDLS and the institutions.

VI. The DEAL negotiations and their outcomes in relation to data tracking

Clear-cut, specific recommendations for action are required so as to ensure that data tracking is something that can be tangibly grasped by all stakeholders. In addition, librarians and researchers in particular should be given the opportunity to engage with the issue of data tracking on a low-threshold basis. The model contract clauses mentioned above were the working group's initial recommendations but putting these into practice proved difficult in the DEAL use case. Ultimately it was not possible to negotiate either a joint controllership arrangement or an impact assessment.

For several reasons, the DEAL negotiations with Elsevier from autumn 2022 onwards and with Springer Nature and Wiley from spring 2023 onwards were initially predestined to systematically highlight the issue of data tracking/data privacy as a subject of negotiation for the first time:

The DEAL publishers are the largest academic publishers in the world and regard themselves first and foremost as "leaders in information and analytics" (https://www.elsevier.com
/about, and also https://epdos.nl), not primarily as publishing companies. It is no coincidence that this shift in the publishers' self-image places the emphasis on data analytics, thereby implicitly highlighting the role of data tracking.

- Being a co-partner of MPDLS, the DFG has the possibility to establish its positions on data tracking in research directly in a contractual context when engaging in DEAL negotiations.
- Any agreement reached with the major publishers at national level on important aspects of protection against data tracking could have been applied to contracts between individual consortia and small to medium-sized publishers: in terms of a joint controllership arrangement, the plan was to make this available to

libraries and consortia as a model agreement for reuse in further negotiations following successful implementation in the DEAL context.

In each of the DEAL negotiations with the three major publishers, a separate working group was formed on the issue of data tracking and data privacy. These working groups were made up of representatives of the respective publishing company and two people representing DEAL/MPDLS. On the side of sciences, the people involved in the DEAL negotiations were identical in all cases, thereby guaranteeing a standardised approach on the part of the research community.

In the case of all three publishers, the discussions began by undertaking an analysis of the current situation with the aim of determining actual data tracking practices pursued . At this stage, the DFG working group in particular contributed in-depth computer forensics expertise, in part also through direct participation in discussions with the publishers. In their response to the unanimous opinion put forward by the lawyers in the DFG working group and the law firm involved, the publishers were likewise unanimous – to the surprise of DEAL – in their rejection of joint responsibility within the meaning of Art. 26 (1)(2) GDPR: according to the publishers, the processing steps can be clearly differentiated from one another, and each step can be clearly assigned to one sphere. Despite several rounds of negotiations, it was not possible to resolve the disagreement on this issue specifically. In particular, the publishers do not regard the ECJ case law described above as being applicable to this specific constellation. The issue can only be resolved definitively based on complaints lodged with the respective data protection supervisory authorities or through legal action on the part of researchers concerned.

The DEAL data tracking negotiations have drawn the attention of publishers and science organisations to the problem, revealing to both sides the shortcomings in the handling of data tracking under data privacy law. Some of the problematic data tracking techniques have now been scaled back by publishers; at the same time, more legally compliant policy regulations have been established, data policy governance has been introduced, and a cookie management platform has been announced, which has already been implemented in some cases. The result is that the contracts do not assume joint responsibility on the part of academic publishers and academic institutions within the meaning of Art. 26 (1)(1) GDPR. Only internally did the publishers indemnify the MPDLS and the institutions participating in the contracts against claims for damages and other consequences in the event that a different legal opinion was subsequently expressed by the data protection authorities and/or the courts. Nonetheless, the most important goal was achieved from the point of view of DEAL and the academic institutions: the latter do not have to bear the financial consequences of any data privacy breaches. Nevertheless, the institutions are of course still responsible to the authorities and to those affected, potentially also in the public domain and in court.

A temporary or possibly even a definitive ban on the relevant processing operations by a supervisory authority (Art. 58 (2)(f) GDPR) – which would effectively mean blocking database access – continues to be suspended over the contracts as a sword of Damocles that cannot be eliminated. In such a case, only potential financial consequences would be compensated by the publishers. Apart from this aspect, the three contracts differ significantly from each other: on the one hand there is Springer Nature, where the data privacy regulations are satisfactory, and on the other hand there is Elsevier and Wiley, where unfortunately not all desired outcomes were achieved.

For the first time in publishing contracts of this nature, the DEAL contracts contain regulations relating to the collection, processing and utilisation of user data, thereby improving transparency as to what is tracked. There is no prohibition of tracking per se, however. Tracking is permitted (in this respect the provisions are identical in wording to the existing legal situation under Art. 6 (1) GDPR) if

- users themselves consent, or
- it is necessary for the fulfilment of a contractual or legal obligation, or
- if there is a legitimate interest in accordance with Art. 6 (1)(f) GDPR.

One important aspect of the negotiations was the issue of the transfer of data to countries outside the EU, especially the USA, as this not only violates applicable data and personal data privacy legislation but also constitutes an expansion of data tracking, At the beginning of the talks, and as a result of the "Schrems II" judgement (ECJ, judgement date 16 July 2020 – Ref. C-311/18), the sharing of personal data in the USA was only permitted to a very limited extent. But the publishers Elsevier and Wiley insisted on the possibility of such data transfer outside Europe, citing the (IT) structures maintained by their global corporations. By contrast, Springer Nature placed less emphasis on this aspect from the outset, since it processes data inside the EU. A new situation arose in the course of the negotiations, however, when an adequacy decision was issued by the European Commission on the EU-US data protection framework (EU-US Data Privacy Framework dated 10 July 2023, OJ L 231 dated 20 September 2023, p. 118 ff., https://eur-lex.europa.eu/eli/dec_impl/2023/1795/oj?locale=de, status 06 December 2023): since this decision, the transfer of data to the USA is no longer prohibited in principle (European Commission 2024). The contracts with Elsevier and Wiley stipulate that the transfer of data outside the EU is only permitted if the requirements of Art. 44 et seq. GDPR are met. In addition to the adequacy decisions (Art. 45 GDPR), this also includes mere contractual guarantees on the part of the controller (Art. 46 GDPR) – here, the Wiley contract explicitly emphasises the standard EU contractual clauses (Art. 46 (3 GDPR, EU SSCs), which are also a permissible option for Elsevier due to the broad scope of the reference. As a result, the level of data privacy is now barely in line with the European legal framework since data from German academic institutions is now also processed in countries such as India, China, the USA and the Philippines. Here the detailed provisions differ between publishers. According to its privacy policy, for example, Elsevier has data processed in "Australia, China, France, Germany, India, Ireland, the Netherlands, the Philippines, Singapore, the United

Kingdom, and the United States" (https://www.elsevier.com/de-de/legal/privac y-policy, status 06 December 2023), while the Wiley contract is subject to the Wiley-VCH data privacy provisions: "If data transfers take place abroad, they are based within the EU or the EEA or in a country that has an adequate level of data protection in accordance with a decision of the EU Commission. In the case of data transfers to Wiley-VCH Group companies based in other countries, Wiley-VCH will ensure by way of guarantees that the Wiley-VCH Group company importing the data has been committed to an adequate level of data protection." (https://www.wiley-vch.de/de/ueber-wiley/impressum, status 05 October 2023). The Wiley corporation also states the following: "We may transfer, store or process personal information in any country where Wiley or its service providers operate. This includes the United States, the United Kingdom, Germany, Singapore, Brazil, India, Sri Lanka, Australia, the Philippines, and China. These countries might have privacy laws that do not provide the same level of protection as those of your country of residence. We take all necessary steps to ensure that personal information is treated securely and in accordance with all applicable data protection laws." (https://www.wil ey.com/en-us/privacy, status 05 October 2023). In this respect, the data privacy situation with Wiley is similar to that of Elsevier. In the case of both publishers, however, it is not immediately clear whether personal data from Germany is indeed transferred to these countries, and if so, subject to what conditions; Elsevier's Open Access Agreement team is based in Chennai, India, for example. Since the contract was concluded based on there being no joint responsibility within the meaning of Art. 26 (1) GDPR, it does not provide for any rights of control or information rights on the part of the institutions or MPDLS GmbH. Researchers do have the option of requesting detailed information on their collected data, the data processing and the level of data privacy (Art. 15 GDPR), however. For this reason, requests for information or complaints lodged by researchers are a possible way of clearing up what is otherwise a legal grey area.

In the case of Springer Nature, the collection, processing and utilisation of user data is subject to the proviso that the tracking is absolutely necessary for the fulfilment of the contract. This restriction is very far-reaching, which is why data tracking is largely prevented here. Wiley and Elsevier are not subject to this restriction, so these two publishers reserve the right to track data. According to the wording of the contract, "the existence of a legitimate interest pursuant to Art. 6 (1)(f) GDPR" can apply to a many constellations; however, these must be weighed against the interests, fundamental rights and freedoms of the data subjects (as also provided for in Art. 6 (1)(f) GDPR). It will now be necessary to observe and possibly also assess how Wiley and Elsevier utilise these possibilities and apply this weighing-up of interests (see Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, Article 29 Data Protection

Working Party 2014, in particular p. 76 f.). In the case of Wiley, it is also stipulated that the aforementioned conditions likewise apply to tracking outside the DEAL contract (which of course is required by law anyway).

With regard to the use of cookies, the contract with Springer Nature stipulates that the publisher must generally limit the storage period to one month or less and may not use technical methods to restore erased cookies. The latter also applies to Elsevier. By contrast, Wiley only commits to "reasonable endeavours" in this regard. Wiley and Elsevier undertake to set up a cookie management platform where users can find information to help them decide on the extent and type of data tracking to which they consent.

Springer Nature undertakes to store IP addresses in anonymised form only (deletion of the last two octets in the case of IPv4). There is no restriction of this nature in the case of Elsevier and Wiley, though workshops are to be organised to enable further discussion in the first year of the contract term.

All in all, the data privacy regulations hold much that is positive, but also – in the case of Elsevier and Wiley – many negative aspects: On the one hand, it is the first time that significantly more specific provisions were agreed upon in the area of data privacy in national contracts. On the other hand, it is regrettable to note that with regard to data tracking in particular, not much more has been achieved than a reiteration of the legal requirements that apply anyway (i.e. irrespective of any contractual clause).

VII. Assessment

Data tracking in research is a problem to which academic institutions have only just started to come up with a response (Zweck, Holtmannspötter & Freund 2024); for publishers, on the other hand, the possibilities of data tracking – especially in connection with the use of artificial intelligence – open up whole new business areas that have more to do with data analysis and less to do with conventional publishing activities. The publishers' current stance is not driven purely by economic interests, but very likely also by the desire to avoid precedents and the impact this would have on their contracts with other (European) academic institutions.

Taking stock at this interim stage, it can be said that for the first time, the issue of data tracking has been successfully incorporated in negotiations with academic publishers such as the DEAL negotiations. The DEAL contracts are now all publicly accessible (Elsevier B.V. & MPDL Services gGmbH 2023; Wiley-VCH GmbH & MPDL Services gGmbH 2023 and Springer Nature Customer Service Centre GmbH & MPDL Services gGmbH 2023): even though they have now been approved by the various academic institutions, their benefits to science continue to be the subject of controversial debate nonetheless. Essentially, the agreed provisions do no more than spell out the validity of the GDPR, in this sense they fall far short of expectations – and not just those of the DFG working group. On the issue of liability, it was not possible to agree on a joint controllership arrangement. Only an exemption clause that protects institutions such as libraries from being exposed to possible financial

claims should a lawsuit be filed for violation of data privacy law and personal rights could be achieved in the negotiations. Another unsatisfactory aspect is the inconsistent and often inadequate level of data protection in relation to user data processing outside Europe. Here, the publishers were able to fall back on the EU's relaxed regulations: as a result, it is only possible to get a rudimentary handle on how tracking data is further processed outside Europe. One point to have emerged clearly, however, is that tracking techniques using cookies are becoming less important, while browser data and search history are becoming more the focus in terms of data tracking, as are links to ORCID numbers.

What can be learned from the experience of the three DEAL negotiations and developments towards increasingly detailed data tracking techniques is that the issue of data tracking must be incorporated in primary negotiations with publishers early on and not dealt with in separate negotiations, as has been the case here. If this is not done, the economic framework conditions are established during primary negotiations, thereby restricting the scope for regulating data tracking. It would definitely be desirable to achieve better legal certainty for the academic institutions, ideally either based on a joint controllership arrangement or by means of independent third-party publisher certification relating to data privacy and data security. As matter of urgency, coordination of the academic institutions is to be prepared at European level with the aim of improving the regulation of data tracking.

Awareness of the issue of data tracking must be raised in libraries and academic institutions as well as at the level of university administrations and science organisations. For libraries in particular, the issue of data tracking should also be incorporated in traditional licence agreement negotiations, preferably based on a model clause. An assessment should be undertaken of the technical methods that might be used to restrict data tracking. Given the oligopoly structures among publishing companies, it remains to be seen whether or not the latter are interested in engaging in partnership-based dialogue with the research community and are willing to adopt principles of transparency, data economy and green IT. Researchers and scientific institutions are called upon to take on responsibility for their data and digital privacy.

VIII. Note

Robert Altschaffel, Jana Dittmann and Stefan Kiltz contributed to the data-related aspects of this publication, Michael Beurskens and Judith Ludwig contributed to the legal aspects, Wolfram Horstmann, Judith Ludwig, Bernhard Mittermaier and Katrin Stump contributed to the library-related aspects and Gerhard Lauer contributed to those aspects relating to science policy and the historical development of publishing. Michaela Bilic-Merdes led the Ad Hoc Working Group on Data Tracking in Research and was in charge of the content structure of this publication. If you have any questions regarding the legal opinion or the JCA, please contact the DFG.

IX. Literature

Altschaffel, Robert; Kiltz, Stefan; Lucke, Thorsten; Dittmann, Jana (2020): *Introduction to Being a Privacy Detective: Investigating and Comparing Potential Privacy Violations in Mobile Apps Using Forensic Methods*, in: *Proceedings of the Fourteenth International Conference on Emerging Security Information, Systems and Technologies (Securware)*, Valencia, Spain, 21–25/09/2020, p. 60–68.

Artikel-29-Datenschutzgruppe (2014): *Stellungnahme 06/2014 zum Begriff des berechtigten Interesses des für die Verarbeitung Verantwortlichen gemäß Artikel 7 der Richtlinie 95/46/EG*. Accessible at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_de.pdf, last accessed on 15 February 2024.

Benedikt, Kristin; Schwartmann, Rolf (2021): *Im Drachenblut des Digitalen. Facebook und Google machen es vor: wie große Verlage in Zukunft die Wissenschaft steuern*, in: *Frankfurter Allgemeine Zeitung*, 04.08.2021, p. N 4.

Bettinger, Eliza; Bursic, Meryl; Chandler, Adam (2023): *Disrupting the Digital Status Quo: Why and How to Staff for Privacy in Academic Libraries.* Accessible at: https://publish.illinois.edu/licensingprivacy/files/2023/06/Whitepaper-on-Privacy-Staf fing-Licensing-Privacy.pdf, last accessed on 15 February 2024.

Bundesbeauftragte für Datenschutz und Informationsfreiheit (2023). *Relevante Auswirkungen des Schrems II Urteils.* Accessible at: https://www.bfdi.bund.de/DE/Fachthemen/Inhalte/Europa-Internationales/Auswirkungen-Schrems-II-Urteil.html, last accessed on 15 February 2024.

cOAlition S: More than 2000 journals share price and service data through Plan S's Journal Comparison Service, in: cOAlition S. Accessible at: https://www.coalition-s.org/more-than-2000-journals-share-price-and-service-data-through-journal-comparison-service, last accessed on 15 February 2024.

Cronin, Blaise; Atkins, Helen B. (Eds.) (2000): *The Web on Knowledge. A Festschrift in Honor of Eugene Garfield*, Medford 2000.

Deutsche Bank AG (2005): *Reed Elsevier: Moving the Supertanker*, Company Focus. Global Equity Research Report, 11.01.2005, p. 34 and 37.

Deutsche Forschungsgemeinschaft (2021): *Datentracking in der Wissenschaft. Aggregation und Verwendung bzw. Verkauf von Nutzungsdaten durch Wissenschaftsverlage.*

Accessible at: https://www.dfg.de/download/pdf/foerderung/programme/lis/datentracking_papier_de.pdf, last accessed on 15 February 2024.

Drake, Tom (2024): *Advancing Equity and Innovation in Research Publishing. Time for a New Era in the Open Access Movement?* Blog Post Center for Global Development. Accessible at: https://www.cgdev.org/blog/advancing-equity-and-innovation-research-publishing-time-new-era-open-access-movement, last accessed on 28 March 2024.

Elsevier B.V.; MPDL Services gGmbH (2023): *Projekt DEAL – Elsevier Publish and Read Agreement.* Accessible at: https://dx.doi.org/10.17617/2.3523659.

European Commission (2024): *Adequacy Decisions. How the EU Determines if a Non-EU Country Has an Adequate Level of Data Protection.* Accessible at: https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-p rotection/adequacy-decisions_de?prefLang=de, last accessed on 15 February 2024.

EU (2023): *Recommendation CM/Rec(2023)3 of the Committee of Ministers to Member States on library legislation and policy in Europe.* 1462nd Meeting of the Ministers' Deputies (5 April 2023). Accessible at: https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680aaced6, last accessed on 14 February 2024.

Freiberg, Michael (2022): *Third-Party-Tracking bei Wiley und Springer. Analyse und Ausblick*, in: ABI-Technik 42, 2, p. 96–104. Accessible at: https://doi.org/10.1515/abitech-2022-0017.

Garfield, Eugene (1955): *Citation Indices for Science. A New Dimension in Documentation through Association of Ideas*, in: *Science*, 122, p. 108–111. Accessible at: https://www.science.org/doi/10.1126/science.122.3159.108.

Gulliver, Sophie; Drake, Tom (2024): *The $1 Trillion Paradox. Why Reforming Research Publishing Should be a Global Priority.* CGD Note 362 (March). Accessible at: https://www.cgdev.org/sites/default/files/why-reforming-research-publishing-should-be-global-priority.pdf, last accessed on 28 March 2024.

Hanson, Cody (2019): *User Tracking on Academic Publisher Platforms.* Accessible at: https://www.codyh.com/writing/tracking.html, last accessed on 15 February 2024.

International Science Council (ISC) *The Future of Scientific Publishing.* Accessible at: https://council.science/actionplan/why-scientific-publishing-matters, last accessed on 15 February 2024.

Kiley, Robert (2023): *Supporting Open Access for 20 Years.* Accessible at: https://www.coalition-s.org/blog/supporting-open-access-for-20-years-five-issues-that-have-slowed-the-transition-to-full-and-immediate-oa, last accessed on 15 February 2024.

Kiltz, Stefan; Altschaffel, Robert; Dittmann, Jana (2023): *Science-Tracker Fingerprinting with Uncertainty: Selected Common Characteristics of Publishers from Network to Application Trackers on the Example of Web, App and Email*, in: *Proceedings of the Seventeenth International Conference on Emerging Security Information, Systems and Technologies (Securware)*, Porto, Portugal, 25–29/09/2023, p. 88–97.

Klein, Samuel (2019): *Turning the Supertanker. Deutsche Bank on Elsevier's excess*, in: *Knowledge Futures Group*, 05.08.2019. Accessible at: https://notes.knowledgefutures.org/pub/supertanker/release/3, last accessed on 15 February 2024.

Lamdan, Sarah: *Data Cartels. The Companies that Control and Monopolize Our Information.* Stanford 2022.

Miranda, Robert (2001): Robert Maxwell. *Forty-Four Years as Publisher*, in: Einar H. Frederiksson (Ed.): *A Century of Science Publishing. A Collection of Essays.* Amsterdam, p. 77–90.

Posenato Garcia, Leila (Ed.): *We so Loved Open Access.* São Paolo 2023. Accessible at: https://doi.org/10.21452/abec.2023.isbn.978-65-993452-6-5.

Reuter, Peter; Söllner, Konstanze (2022): *Datentracking in den Wissenschaften – ein Sonderfall?*, in: ABI Technik 42,2, p. 94–95. Accessible at: https://doi.org/10.1515/abitech-2022-0016.

Roßnagel, Alexander (2022): *Third-Party-Tracking – ein Problem aus Sicht des Datenschutzes?*, in: ABI Technik 42,2, p. 105–107. Accessible at: https://doi.org/10.1515/ abitech-2022-0018.

Schonfeld, Roger C. (2023): *The Supercontinent of Scholarly Publishing?*, in: *The Scholarly Kitchen* (3 May 2018). Accessible at: https://scholarlykitchen.sspnet.org/2018/05/03/supercontinent-scholarly-publishing, last accessed on 15 February 2024.

Siems, Renke (2021): *When Your Journal Reads You – User Tracking on Science Publisher Platforms*, in: *Elephant in the Lab.* Accessible at: https://doi.org/10.5281/zenodo.4683778.

Siems, Renke (2022): *Das Lesen der Anderen. Die Auswirkungen von User Tracking auf Bibliotheken*, in: o-bib. *Das offene Bibliotheksjournal* 9.1, p. 1–25. Accessible at: https://doi.org/10.5282/o-bib/5797.

Siems, Renke (2024): *Subprime Impact Crisis. Bibliotheken, Politik und digitale Souveränität*, in: *BIBLIOTHEK – Forschung und Praxis 48.2*, p. 1–11. Accessible at: https://doi.org/10.1515/bfp-2024-0008.

Springer Nature Customer Service Center GmbH; MPDL Services gGmbH (2023): *Projekt DEAL – Springer Nature Publish and Read Agreement, 2024–2028.* Accessible at: https://dx.doi.org/10.17617/2.3551270.

Stern, Bodo; Rooryck, Johan (2023): *Introducing the "Towards Responsible Publishing" proposal from cOAlition S, in: cOAlition S.* Accessible at: https://www.coalition-s.org/blog/introducing-the-towards-responsible-publishing-proposal-from-coalition-s, last accessed on 15 February 2024.

Wiley-VCH GmbH, MPDL Services gGmbH (2023): *Projekt DEAL – Wiley Publish and Access Agreement, 2024–2028.* Accessible at: https://dx.doi.org/ 10.17617/2.3551268.

Yoose, Becky; Shockey, Nick (2023): *Navigating Risk in Vendor Data Privacy Practices. An Analysis of Elsevier's Science Direct*, in: SPARC. Accessible at: https://sparcopen.org/our-work/privacy-and-surveillance-community-of-practice/navigating-risk-in-vendor-data-privacy-practices bzw. https://doi.org/10.5281/zenodo.10078610.

Zweck, Axel; Holtmannspötter, Dirk; Freund, Sebastian (2024): *Tracking in der Wissenschaft.* VDI Research-Paper 17, VDI Technologiezentrum GmbH Düsseldorf. Accessible at: https://www.vditz.de/service/publikationen/details/tracking-in-der-wissenschaft, last accessed on 27.05.2024.

Summary: This article discusses the problem of data tracking in academic libraries in the context of the DEAL negotiations with the academic publishers Elsevier, Springer, and Wiley. The authors call for libraries and academic institutions to be made more aware of the issue of data tracking and for the inclusion of data protection clauses in license agreements with publishers. The article assesses the results of the recently concluded negotiations concerning data tracking and comes to a mixed conclusion. On the positive side, both sides have acknowledged the data protection shortcomings in the handling of data tracking and, although libraries and higher education institutions are liable to researchers for data protection violations and can therefore be sued independently of each other, compensation is provided internally through recourse by the institutions to the publishers. On the negative side, it should be noted that the agreed regulations only spell out the validity of the GDPR and thus fall far short of the expectations of academic institutions and others. It was not possible to agree on a joint controller agreement on the issue of liability, nor was it possible to prevent the further processing of data outside of Europe in general.