SPARC

NAVIGATING RISK IN VENDOR DATA PRIVACY PRACTICES

An Analysis of Elsevier's ScienceDirect

November 2023

NAVIGATING RISK IN VENDOR DATA PRIVACY PRACTICES

An Analysis of Elsevier's ScienceDirect

SPARC

November 2023

Project Statement

Becky Yoose of LDH Consulting Services prepared this report in collaboration with Nick Shockey of SPARC as part of the Navigating Risk in Vendor Data Privacy Practices Project at https://sparcopen.org/our-work/privacy-and-surveillance-community-of-practice.

The following analysis reflects publicly available information at the time of review (from Summer 2022 to Summer 2023) and focuses on the North American context. The applicability of its findings may be limited in other regions and may change over time as privacy policies and practices are revised. If you have additional information that could impact this analysis, please email Nick Shockey at nick@sparcopen.org.

This report is not an analysis of whether ScienceDirect complies with applicable privacy laws nor should it be construed as legal advice. For additional information about the limitations of the report, as well as future research opportunities, please see the "Limitations and Further Investigation" section of the report.

Acknowledgements

This report benefitted from feedback and guidance from Sarah Lamdan, Dorothea Salo, Michele Gibney, and Heather Joseph. The analysis builds on prior work by the Licensing Privacy Project, which provided the framework for the data privacy assessment. In doing this work, we are also grateful for the related work by many organizations, including the Library Freedom Project and Library Futures.

Contents

EXECUTIVE SUMMARY

As libraries transitioned from buying materials to licensing content, serious threats to privacy followed. This change shifted more control over library user data (and whether it is collected or kept at all) from the local library to third-party vendors, including personal data about what people search for and what they read. This transition has further reinforced the move by some of the largest academic publishers to move beyond content and become data analytics businesses that provide platforms of tools used throughout the research lifecycle that can collect user data at each stage. These companies have an increasing incentive to collect and monetize the rich streams of data that these platforms can generate from users. As a result, user privacy depends on the strength of privacy protections guaranteed by vendors (e.g., negotiated for in contracts), and a growing body of evidence indicates that this should be a source of concern.

User tracking that would be unthinkable in a physical library setting now happens routinely through such platforms. The potential integration of this tracking with other lines of business, including research analytics tools and data brokering services, raises pressing questions for users and institutions.

Elsevier provides an important case study in this dynamic. Elsevier is many academic libraries' largest vendor for collections, and its platforms span the knowledge production process, from discovery and idea generation to publication to evaluation. Furthermore, Elsevier's parent company, RELX, is a leading data broker. Its "risk" business, which provides services to corporations, governments, and law enforcement agencies based on expansive databases of personal data, has surpassed its Elsevier division in revenue and profitability.

For these reasons, it is important to carefully consider Elsevier's privacy practices, the risks they may pose, and proactive steps to protect users. This analysis focuses on ScienceDirect due to its position as a leading discovery platform for research as well as the Elsevier product that researchers are most likely to interact with regularly.

Based on our findings, many of ScienceDirect's data privacy practices directly conflict with library privacy standards and guidelines. The data privacy practices identified in our analysis are like the practices found in many businesses and organizations that track and harvest user data to sustain privacy-intrusive data-driven business models. The widespread data collection, user tracking and surveillance, and disclosure of user data inherent to these business models run counter to the library's commitment

to user privacy as specified in the ALA Code of Ethics, Library Bill of Rights, and the IFLA Statement on Privacy in the Library Environment. Examples of current ScienceDirect practices found in our analysis that conflict with these standards include:

- Use of web beacons, cookies, and other invasive web surveillance methods to track user behavior outside and beyond the ScienceDirect website
- Extensive collection of a broad range of personal data (e.g., behavioral and location data) from ScienceDirect combined with personal data harvested from sources beyond ScienceDirect (i.e., third parties in and outside of RELX and data brokers as stated in Elsevier's Privacy Policy and U.S. Consumer Privacy Notice)
- Collection of personal data by third parties, including search engines, social media platforms, and other personal-data aggregators and profilers such as Google, Adobe, Cloudflare, and New Relic, through extensive use of third-party trackers on the ScienceDirect site
- Disclosure of personal data to other Elsevier products and the potential for disclosure of personal data to other business units within RELX, including risk products and services sold to corporations, governments, and law enforcement agencies
- Processing and disclosure of personal data (and personal data inferred from personal data) for targeted, personalized advertising and marketing

In particular, ScienceDirect's U.S. Consumer Privacy Notice, posted and updated in 2023, raises important concerns. The notice describes the disclosure of detailed user data—including geolocation data, sensitive personal information, and inference data used to create profiles on individuals—both for wide-ranging internal use and to external third parties, including "affiliates" and "business and joint venture partners."

The collection and disclosure of data about who someone is, where they are, and what they search for and read by the same overarching company that provides sophisticated surveillance and data brokering products to corporations, governments, and law enforcement should be alarming. These practices raise the question of whether simultaneous ownership of key academic infrastructure alongside sophisticated surveillance and data brokering businesses should be permitted at all—by users, by institutions, or by policymakers and regulatory authorities.

Our analysis cannot definitively confirm whether personal data derived from academic products is currently being used in data brokering or "risk" products. Nevertheless, ScienceDirect's privacy practices highlight the need to be aware of this risk, which is not mitigated by privacy policy revisions or potential verbal assurances concerning specific data uses. Privacy policies can be changed unilaterally, and denials are not legally binding. To be meaningful, any privacy guarantee a vendor makes must be durable, verifiable, and not limited to a particular jurisdiction.

As many of the largest publishers reinvent themselves as platform businesses, users and institutions should actively evaluate and address the potential privacy risks *as this transition occurs* rather than after it is complete. In closely analyzing the privacy practices of the leading vendor in this transition, this report highlights the need for institutions to be proactive in responding to these risks and provides initial steps for doing so.

This report underscores the significant expertise and capacity required for any institution to understand even one vendor's privacy practices—and the power asymmetry this creates between vendors and libraries. Collaborative efforts, such as SPARC's Privacy & Surveillance Community of Practice, can plan a key role in supporting future action to address the real privacy risks posed by vendors' platforms. This report closes with options that institutions may consider to mitigate these risks over the short and longer term.

INTRODUCTION

The following report provides a detailed analysis of the privacy practices of Elsevier's ScienceDirect product. This analysis aims to assist libraries in better understanding the complex, rapidly evolving landscape of potential privacy risk across the policies, notices, and contract language that comprise one vendor's privacy practices and in taking action to address this risk. The resulting findings highlight the urgency for libraries to better understand these risks and, where necessary, to mitigate them. While this analysis and recommended actions are grounded in the library context, these findings also raise pressing questions for faculty, administrators, and policymakers to consider as well.

This report focuses specifically on Elsevier's ScienceDirect product as the leading academic discovery platform of the world's largest publisher and many libraries' largest vendor for collections. Elsevier is the most mature example of a wider shift by academic publishers to become data analytics businesses that provide integrated platforms of tools used throughout the research lifecycle.[1] Their products span discovery; research management, funding, and collaboration; publishing and dissemination; and, research analytics.[2] Elsevier's acquisition of Interfolio, a leading faculty information system, has further embedded them into the research lifecycle through control of the enterprise software many institutions use for hiring, promotion, and tenure.[3] While other academic publishers are pursuing similar transitions,[4,5] Elsevier appears to be the most well-developed with the most opportunity to collect and monetize user data.

Elsevier is also a subsidiary of RELX, a leading data broker and provider of sophisticated surveillance products with customers that include corporations, governments, and law enforcement agencies.[6] The

1 Kunz, Raffaela. "The Digital Threat to Science and Academic Freedom." Blog for Transregional Research. August 15, 2023. https://trafo.hypotheses.org/36984.

2 Posada, Alejandro, and George Chen. "Inequality in Knowledge Production: The Integration of Academic Infrastructure by Big Publishers." In 22nd International Conference on Electronic Publishing. OpenEdition Press, 2018. https://doi.org/10.4000/proceedings.elpub.2018.30.

3 SPARC. "Elsevier's Acquisition of Interfolio: Risks and Responses." June 29, 2022. https://infrastructure.sparcopen.org/interfolio-acquisition.

4 Lamdan, Sarah. *Data Cartels: The Companies That Control and Monopolize Our Information*. 50-71, Stanford University Press (2022).

5 Pooley, Jeff. "Surveillance Publishing." The Journal of Electronic Publishing (2022), https://doi.org/10.3998/jep.1874.

6 Lamdan, Sarah. "Librarianship at the Crossroads of ICE Surveillance," In The Library With The Lead Pipe. November 13, 2019. https://www.inthelibrarywiththeleadpipe.org/2019/ice-surveillance. See also *Data Cartels* by Sarah Lamdan.

company's "risk" products "offer an oceanic computerized view of a person's existence" and combine "10,000 different data points on hundreds of millions of people."[7] According to reporting, RELX risk products have been documented as being used in ways that raise serious concerns, including to help monitor protestors' social media feeds,[8] surveil immigrants,[9] blackmail women,[10] and help in attempting to manufacture false terrorism charges against some of those who participated in anti-racism protests in the summer of 2020.[11,12]

The simultaneous operation of academic infrastructure alongside these data brokering and surveillance products heightens the risks to users related to the collection and use of their personal data. Experts have already raised concerns about the possibility that patron data could flow from RELX research products into these surveillance products used by U.S. Immigration and Customs Enforcement (ICE) and other law enforcement.[13] Our analysis cannot definitively confirm whether personal data derived from academic products is currently being used in data brokering or other "risk" products, but the clear potential for such uses to occur underlines the need for institutions to ensure that they fully understand Elsevier's privacy practices and take informed action as necessary.

This analysis highlights the extent to which evaluating a vendor's data privacy practices has become increasingly complex due to contracts and documentation that are vague or difficult to understand as well as the opacity of data flows within and beyond a product or company. Librarians and procurement professionals tasked with evaluating a vendor's privacy practices typically do not have access to the dedicated resources needed to evaluate the full array of each vendor's data privacy practices to understand the risks they may present and develop an effective institutional response.

7 Biddle, Sam. "Lexisnexis To Provide Giant Database Of Personal Information To ICE." The Intercept. April 2, 2021. https://theintercept.com/2021/04/02/ice-database-surveillance-lexisnexis.

8 Lamdan, Sarah. "The Quiet Invasion of 'Big Information'." Wired. November 9, 2022. https://www.wired.com/story/big-information-relx-privacy-surveillance-data.

9 "LexisNexis illegally collected and sold people's personal data, lawsuit alleges." CBS News. August 16, 2022. https://www.cbsnews.com/news/lexisnexis-lawsuit-collected-sold-personal-data-immigration-advocates-allege.

10 Cushing, Tim. "Louisville Cop Used Law Enforcement Database To Seek Female Targets To Hack For Sexually Explicit Content." TechDirt. October 19, 2022. https://www.techdirt.com/2022/10/19/louisville-cop-used-law-enforcement-database-to-seek-female-targets-to-hack-for-sexually-explicit-content.

11 Cameron, Dell. "Homeland Security Admits It Tried to Manufacture Fake Terrorists for Trump." Gizmodo. November 5, 2022. https://gizmodo.com/donald-trump-homeland-security-report-antifa-portland-1849718673.

12 In response to concerns about the sale of personal data to U.S. Immigration and Customs Enforcement documented in the *Intercept* article cited in note 7 above, RELX's LexisNexis Risk Solutions has offered the following statement: "Our tool contains data primarily from public government records. The principal non-public data is authorized by Congress for such uses in the Drivers Privacy Protection Act and Gramm-Leach-Bliley Act statutes."

13 Lamdan, Sarah. "Librarianship at the Crossroads of ICE Surveillance." *In the Library with the Lead Pipe*, November 13, 2019; https://www.inthelibrarywiththeleadpipe.org/2019/ice-surveillance.

The following analysis aims to assist both users and institutions by identifying specific concerning data privacy practices through the review of ScienceDirect's contracts, public documentation, and website. The analysis also provides suggested actions institutions can take to address these issues related to both ScienceDirect and other vendor products that may present similar concerns.

By carefully analyzing one leading vendor's privacy practices, this report also contributes to a better understanding of the emergence of research platforms and their implications for privacy, academic freedom,[14] and other forms of risk that may seem unrelated to academic publishing but are a stark reality in other platform businesses.

Some of the largest providers of academic content increasingly see themselves and their peers as tech firms rather than publishers.[15,16] This shift creates a strong incentive to collect and monetize user data as a path to increased profits and higher valuations. When students or faculty access content through these platforms, the data collected may be used in ways that they would not knowingly approve of and that could negatively affect them in the future.

As publishers reinvent themselves as platform businesses, users and institutions should actively evaluate the potential privacy risks *as this transition occurs* rather than after it is complete. This is particularly true for vendors that are part of larger businesses that are involved in data brokering activities.

Vendor privacy practices are an effective indicator of potential risk related to data use and disclosure. Analyzing these can enable timely action before risks fully materialize. If action is delayed until especially concerning data uses are fully documented, users and institutions may find themselves in the position of responding to harms that have occurred rather than avoiding them in the first place.

14 German Research Foundation. "Data Tracking in Research: Aggregation and Use or Sale of Usage Data by Academic Publishers." October 28, 2021. https://www.dfg.de/download/pdf/foerderung/programme/lis/datentracking_papier_en.pdf.

15 Bert, Alison. "Bernard Marr on 'the amazing digital transformation of Elsevier from publisher to tech company'." Elsevier Connect Blog. July 10, 2018. https://www.elsevier.com/en-xm/connect/the-amazing-digital-transformation-of-elsevier-from-publisher-to-tech-company.

16 Pullman, Emma. "Company owned by Canada's richest man selling tech used to deport immigrants." The Breach. June 8, 2021. https://breachmedia.ca/company-owned-by-canadas-richest-man-selling-tech-used-to-deport-immigrants.

ANALYSIS METHODOLOGY

The structure of this analysis borrows from the data privacy criteria laid out in the Vendor Contract and Policy Rubric from the Licensing Privacy Project.[17] Created to evaluate the data privacy risks in content platform contracts and policies, the rubric measures risk in eight privacy domains listed in **Table 1**.

Privacy practices in each domain are measured against three levels of data privacy based on a consensus of library data privacy standards, guidelines, and practices from the profession and major professional organizations such as ALA and IFLA (referred to as Minimum Viable Privacy or MVP).

Table 1

PRIVACY DOMAIN	DEFINITION
Data collection	What data the vendor collects, how they collect it, where they collect data, and the rationale for collecting it
User data rights	What controls users have over the vendor's ability to collect, retain, use, and share user data
Data disclosure	What data the vendor shares, which parties the vendor shares data with, the reasons the vendor shares data, and how data sharing is controlled/determined
Data processing	What data the vendor uses, for what purpose, and how data use is controlled/determined
Privacy policy	Public privacy statements on the vendor's service or website, as well as any internal vendor privacy policies the vendor provides the library
Data ownership	Who owns the data in the vendor service or product, and what rights come with data ownership in specific business scenarios
User surveillance	What tracking or logging mechanisms the vendor uses to collect user data, and the level of control users have over vendor tracking/logging behavior while using the service
Data security and accountability	How the vendor protects data in transit and storage from unauthorized access or use, how the vendor works to prevent and respond to data breaches or leaks, and what checks are in place to ensure compliance with vendor security and privacy policies and industry standards

17 Licensing Privacy. "Assessing Contracts." https://publish.illinois.edu/licensingprivacy/contracts.

Each level measures risk to patron privacy, with the caveat that a vendor meeting MVP may not automatically mean that the vendor is adequately protecting patron privacy as a whole.[18]

The ScienceDirect analysis assessed readily available materials: contracts, public documentation, and the front-end website. These materials are not in areas that require special permission to access (e.g., a support site that requires users to log in before accessing help documentation).

Two analyses were conducted on ScienceDirect's privacy policies and practices: one during the Summer of 2022 and another approximately one year later in 2023. While this was not the original intent at the outset of this project, the time involved in preparing the analysis and working toward publication highlighted the added utility of monitoring changes to policies and practices over the course of one year. This approach led to some challenges (e.g., particular website analysis tools could no longer be used on every page in the same way from the first review to the second due to website behavior changes); however, these challenges were small compared with the additional insight gained.

CONTRACTS, POLICIES, AND PUBLIC DOCUMENTATION

While the Vendor Contract and Privacy Rubric is built to assess the vendor contract and privacy policy, our analysis goes far beyond these specific documents to provide a broader picture of the data privacy practices of ScienceDirect. The documents used in the 2022 analysis include the following:

- Signed Elsevier Subscription Agreement contracts from seven academic libraries and one library consortium in the US (available in the SPARC Contract Library: https://sparcopen.org/our-work/big-deal-knowledge-base/contracts-library/)
- California Consumer Privacy Act (CCPA) Notice
- Elsevier Cookie Notice
- Elsevier Data Processing Addendum
- Elsevier Data Security Schedule
- Elsevier Privacy Policy
- Elsevier Privacy Principles
- Privacy Center (login access is required to access setting information)
- ScienceDirect Cookie Notice

18 LDH Consulting Services. "Developing the Vendor Contract and Policy Rubric," January 2022. https://publish.illinois.edu/licensingprivacy/files/2022/03/Licensing-Privacy-Vendor-Rubric-White-Paper.pdf.

- ScienceDirect and Elsevier Support Center documentation
  - » Reading history
  - » Recommendations service
  - » id.elsevier.com use of cookies
  - » Search history
- Website Terms and Conditions[19]

The 2023 analysis compared the 2022 versions of the documents with the 2023 versions. The only substantial documentation changes between the two analyses were the following:

- The supersession of the CCPA Notice by the U.S. Consumer Privacy Notice in 2023
- The addition of the U.S. Privacy Laws Addendum to Elsevier Data Processing Addendum in 2023

Links to the archived versions of the above documents for each year can be found in Appendix A.

FRONT-END WEBSITE DATA COLLECTION

The second half of the analysis focused on investigating data collection and tracking behavior occurring on the public-facing ScienceDirect website. The investigation initially examined four specific web pages at different levels of the website to capture a broader picture of the possible collection and tracking throughout the entire site:

- The ScienceDirect home page (https://www.sciencedirect.com)
- Browse Journals and Books (https://www.sciencedirect.com/browse/journals-and-books)
- The search results page for a keyword search (https://www.sciencedirect.com/search?qs=cat%20behavior)
- A page for an Open Access journal article (https://www.sciencedirect.com/science/article/pii/S016815912200020X)

19 The Privacy Shield Notice was not included in the analysis. The EU-US Privacy Shield framework was invalidated in a 2020 European court ruling. At the time of writing, Privacy Shield has been superseded by the Trans-Atlantic Data Privacy Framework.

A fifth page, https://id.elsevier.com, was tested in 2023 due to the search results page redirecting to the Elsevier login page when evaluating the search results page in 2023 with one of the tools used in the analysis.

The investigation used several mature and widely used tools available to the general public to inspect website traffic:

- **Website Evidence Collector** (WEC; https://github.com/EU-EDPS/website-evidence-collector), a tool developed by the European Data Protection Supervisor to document what information is collected and transmitted by a specific website, including web cookies, local storage, and data requests and traffic
- **Blacklight** (https://github.com/the-markup/blacklight-collector), a tool developed by the nonprofit news organization The Markup (https://themarkup.org/) to detect several standard surveillance methods, including ad trackers, canvas fingerprinting, key logging, and tracking by Google and Facebook
- **NoScript** (https://noscript.net/), a browser add-on that detects and blocks JavaScript and other executable content on websites
- **Ghostery** (https://www.ghostery.com/), a browser add-on that detects and blocks ads and trackers such as web cookies and beacons
- **uBlock Origin** (https://ublockorigin.com/), a browser add-on that detects and blocks ads, trackers, and other privacy-invasive content
- **Web Developer Tools in Firefox and Chrome**

Each URL was tested with local installations of WEC and Blacklight. The results from each test were spot-checked by the three browser plugins installed in Firefox (NoScript) and Chrome (Ghostery and uBlock Origin) and with the network monitor in the Web Developer Tools in both Firefox and Chrome.

ANALYSIS — *Experience and Assessment*

ANALYZING SCIENCEDIRECT'S CONTRACTS, POLICIES, AND PUBLIC DOCUMENTATION

Any analysis of a vendor's contracts, policies, and public documentation depends on the availability, organization, and completeness of relevant documents and the information they contain. Based on the main author's experience with vendor contracts and the Licensing Privacy Project, it is not uncommon for content platform contracts to have little to no contract language surrounding data privacy. ScienceDirect's Subscription Agreement appears to be no exception: in the contracts used for the analysis, there is typically one subsection dedicated to privacy. The following is a representative example of the language contained in this subsection:

> *7.6 Privacy. To the extent that Authorized Users provide any personal data to Elsevier during account registration or otherwise, the Subscriber acknowledges that such information will be collected, used and disclosed by Elsevier in accordance with the Elsevier privacy policy applicable to the Subscribed Products. The terms of the Elsevier Data Processing Addendum at https://www.elsevier.com/legal/processor-terms will apply.*[20]

In this case, any meaningful information about data privacy practices resides outside the signed contract.[21] For example, the contract with the above-quoted section refers to the Privacy Policy and the Data Processing Addendum (DPA). Some contracts reviewed in this analysis included references to both documents, while others only referenced the Privacy Policy.

Curiously, only one contract out of the eight contracts in this analysis included a separate schedule for data security. The schedule provided details about Elsevier's information security measures, the company's responsibilities during incident response, and the subscriber's right to audit specific policies and documentation. The information present in the schedule was largely absent from other documents.

20 Contract language from https://sparcopen.org/wp-content/uploads/2021/02/Fully-Executed-2021-2024-Agreement-Journals-East-Carolina-University-1-19106747595.pdf.

21 The Discussion section will talk more about why this approach of referencing documents and not including them in the contract in full is problematic for libraries.

Due to the lack of detail about data privacy practices in these contracts, this analysis heavily relies on the Privacy Policy and related online documents. **Figure 1** on the following page is a high-level diagram mapping the relationships between the documents used for the analysis. ScienceDirect's Privacy Policy provides general information about what data is collected, how it's used, and when it is disclosed. This Privacy Policy links to several other documents, including the U.S. Consumer Privacy Notice, while not mentioning others, such as the DPA. These documents contain specific information not provided by the Privacy Policy, such as Elsevier's general Cookie Notice's inclusion of web beacons and unique identifiers that are only referred to as "other technologies" in the Privacy Policy. ScienceDirect's Cookie Notice does not mention these other technologies, even though the tools used to investigate data collection on ScienceDirect indicate the use of web beacons and local storage.

The U.S. Consumer Privacy Notice presents the most significant instance where specific, substantive information can be found in other documents but not in the Privacy Policy. Superseding the CCPA Notice in 2023, the U.S. Consumer Privacy Notice is a document detailing the information around personal data collection, processing, disclosure, and rights as required for several US data protection and privacy laws (e.g., CCPA, Colorado Privacy Act, Virginia Consumer Data Protection Act).

While the document states that the Notice is a *supplement* to the Privacy Policy, the Notice in reality provides a clearer picture into the personal data privacy practices that is otherwise missing or vague in the main Privacy Policy. In particular, we learn in the Notice that Elsevier collected and disclosed specific types of personal information, such as:

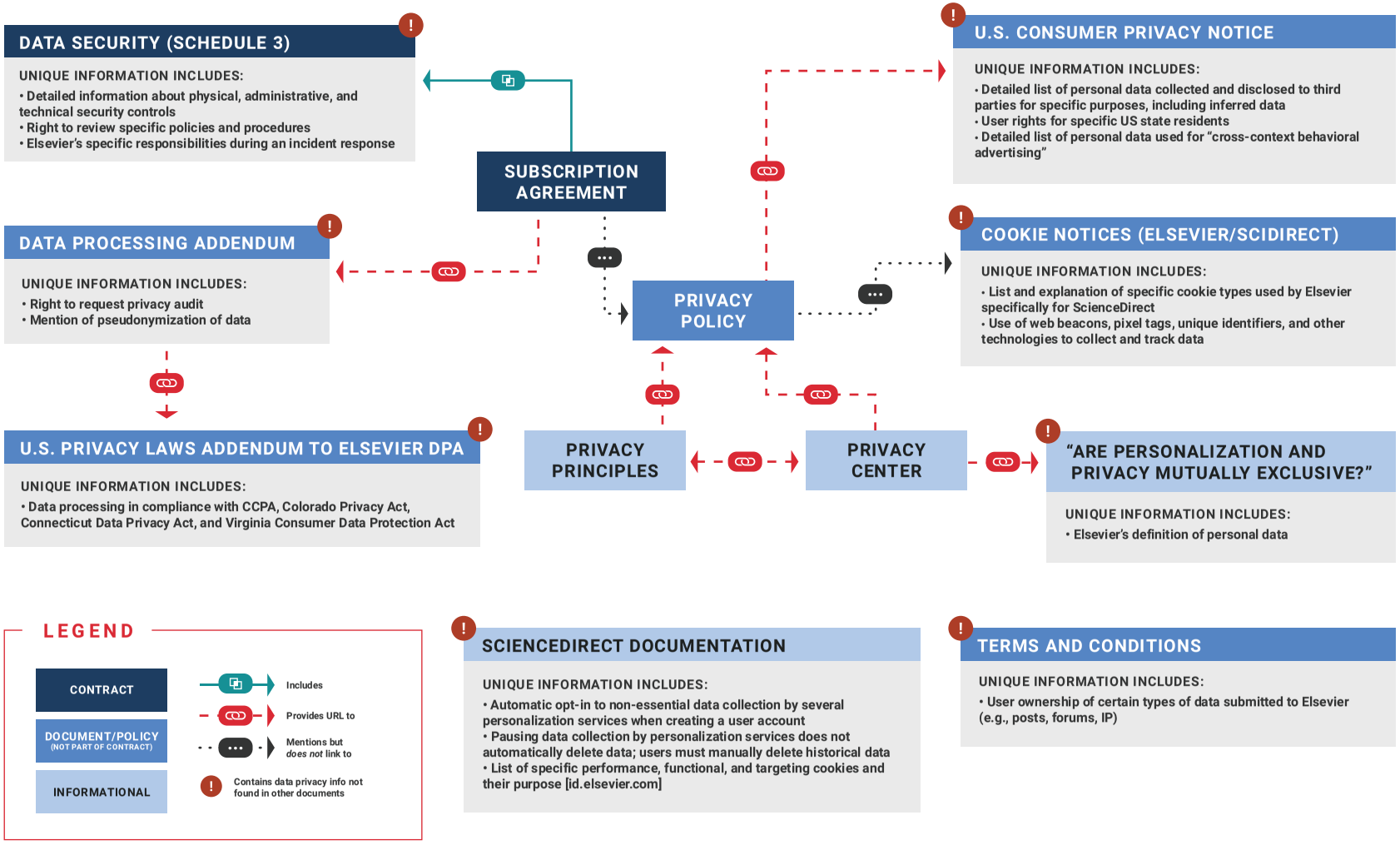
- ***Internet or other electronic network activity information****, such as browsing history, search history, online behavior, interest data, and interactions with our and other websites, applications, systems, and advertisements.*
- ***Geolocation data****, such as approximate device location.*
- ***Inferences*** *drawn from any of the personal information listed above to create a profile or summary about, for example, a consumer's preferences and characteristics*
- ***Sensitive personal information****, such as personal information that reveals a consumer's passport number, racial or ethnic origin, and account log-in and password.*[22]

While the Privacy Policy provides some information about the collection of internet and geolocation data, we learn more about the level of detailed online activity data being collected in the Notice. The Privacy Policy does not include any mention about inference data, which the Notice specifically states is used to create a user profile, documenting the user's "characteristics." The term "characteristics" is not defined in

22 Elsevier. "U.S. Consumer Privacy Notice." August 14, 2023. https://web.archive.org/web/20230814181931/https://beta.elsevier.com/legal/us-consumer-privacy-notice.

Figure 1 | **THE (MESSY) NETWORK DISTRIBUTION OF DATA PRIVACY INFORMATION: EVALUATING A SCIENCEDIRECT AGREEMENT (SUMMER 2023)**

The Subscription Agreement does not tell the entire story about the privacy of user data in Elsevier's ScienceDirect. The chart shows the wide network of resources needed to gain a more detailed account of data privacy practices. The majority of resources contain unique information not available in the main Subscription Agreement or in the Privacy Policy.



the notice; however, industry use of user profiles range from broad to highly detailed information about a particular person, usually in line with a market or audience segment.

The Notice also provides more information about the processing and disclosure of personal information that again is either missing or briefly described in the Privacy Policy. All personal data categories listed in the Notice have been disclosed to "affiliates; customers; *service providers, agents, and representatives; business and joint venture partners*; and other parties where required by law or to protect our rights" (emphasis added).[23] These service providers and business and joint venture partners could potentially include companies that engage in data brokering.[24,25]

In addition, the Notice lists the types of personal information shared for the "purposes of cross-context behavioral advertising or targeted advertising, and have shared for such purposes, in the preceding 12 months, with advertising networks, internet service providers, data analytics providers, operating systems and platforms, and social networks," which includes online activity data, geolocation, and inference data.[26]

Again, this level of detail about the disclosure of personal data for behavioral advertising is absent from the Privacy Policy. Placing important details that adversely affect patron privacy—such as the extent of how personal data is used for user profiling and disclosed to data brokers—in a supplemental policy described in the Privacy Policy as "information about our processing of personal information as required by applicable U.S. laws"[27] can leave some libraries to overlook the notice, particularly if they are in a state where there are no data privacy regulations in place.

Elsevier is only one example among a growing number of companies that place detailed data collection, processing, and disclosure information in supplemental notices that are separate from the main privacy policy as part of compliance to specific US state laws.[28] While there is some consolation that there is a link in the main Privacy Policy to the U.S. Consumer Privacy Notice in the case of Elsevier, many reviewing

23 Ibid.

24 ALA. "Privacy Policies." Accessed October 15, 2023. https://libraryprivacyguides.org/privacy-policies/understanding-commonly-used-phrases-and-terms.

25 Other RELX risk-related activities appear to use similar language. For example, reporting has described RELX as a "data partner" for Palantir. See https://theintercept.com/2021/04/02/ice-database-surveillance-lexisnexis.

26 Elsevier. "U.S. Consumer Privacy Notice." August 14, 2023. https://web.archive.org/web/20230814181931/https://beta.elsevier.com/legal/us-consumer-privacy-notice.

27 Elsevier. "Privacy Policy." July 31, 2023. https://web.archive.org/web/20230731225516/https://www.elsevier.com/legal/privacy-policy.

28 Keegan, Jon, and Jesse Woo. "How to Quickly Get to the Important Truth Inside Any Privacy Policy." *The Markup*, August 3, 2023. https://themarkup.org/the-breakdown/2023/08/03/how-to-quickly-get-to-the-important-truth-inside-any-privacy-policy.

a vendor's privacy practices may not be aware that they should be paying attention to these types of notices which again provide much more detail about a vendor's privacy practices.

Even with the information provided by the various supplementals to the Privacy Policy, the Policy does not capture the entire data privacy picture. Instead, pertinent information is also located in other documents that might easily be overlooked. One example is the Terms and Conditions, which contains information about ownership of specific types of user data (i.e., intellectual property, content posted by users in blogs and discussion posts). Searching ScienceDirect support documentation reveals that several types of non-essential and privacy-invasive data collection (in particular, reading and search histories) are turned on by default at the point of account creation or sign-in. The documentation also states that pausing data collection for these features does not delete historical data, so users must manually delete the collected data.

Additional public documentation aimed at audiences other than library workers also provides information about Elsevier's data privacy practices and processing. Company articles such as "Are personalization and privacy mutually exclusive?" offer some information about Elsevier's data privacy approach.[29] These documents also characterize the company's processing and sharing of personal data across different products and services in the Elsevier Research Products Group (which includes ScienceDirect, SSRN, and Mendeley), as providing "data-derived added value." However, while knowing how Elsevier markets its company's approach to data privacy may be helpful for internal conversations where libraries might encounter marketing copy being repeated by faculty or administrators, these articles do not contain the level of information provided by other documents, and so were not included in this analysis.

INVESTIGATING DATA COLLECTION AND TRACKING ON THE SCIENCEDIRECT WEBSITE

The Privacy Policy refers to specific types of user tracking on ScienceDirect and other Elsevier websites as "cookies and other technologies." The Cookie Notice and ScienceDirect documentation provide more information about the types of cookies used by ScienceDirect and other Elsevier sites. However, information about the use of web beacons and other tracking methods is scant in the documentation. The extent of data collection and use remains largely hidden from library workers who do not have the time or resources to go beyond the contract and the Privacy Policy.

29 Jan Aalbersberg, Ijsbrand. "Are Personalization and Privacy Mutually Exclusive?" Elsevier, September 21, 2017. https://web.archive.org/web/20220709005621/https:/www.elsevier.com/connect/archive/are-personalization-and-privacy-mutually-exclusive.

Our analysis takes a multifaceted approach to test front-end website data collection and tracking. Popular browser plugins such as Ghostery, uBlock Origin, and NoScript grant quick access to the list of cookies and executable content (e.g., JavaScript) found on a web page. Built-in browser web developer tools provide a more detailed view of real-time network traffic for a specific page, including what data trackers and scripts are requesting and collecting. More advanced tools such as Website Evidence Collector (WEC) and Blacklight offer a better overview of the types of tracking present on multiple pages across the ScienceDirect website and the information being collected, stored, and transmitted. Both WEC and Blacklight output several text files after using each tool. These files contain lists of specific trackers on the site and overall inspection report files. WEC and the web version of Blacklight also produce a summary report of the findings in a more friendly, human-readable HTML file.

The overall number of cookies and trackers differs between the four pages in the website analysis (see the tracker table and chart in Appendix D). The home page level, for instance, has fewer trackers than the journal article page. This discrepancy in the number of trackers between pages could be explained in a few ways. One explanation is the likelihood of a page being visited by users (i.e., users landing on an article page through an external search engine result such as Google Scholar). Another possible explanation is that the search result and article pages offer the opportunity to collect behavioral data tied to specific subject interests that could then be used to refine ScienceDirect's search engine and recommendation service. Yet another explanation points to increased user tracking on article and search results pages allowing for more detailed user profiles, tying specific interests or search/reading behaviors around specific topics to an individual user. Elsevier's Privacy Policy and Cookie Notice provide general information about user profiling via sharing data between "related Elsevier services." However, the extent to which this happens is still unknown. That said, clues were found in the analysis as to how this profiling takes place, even when users block cookies.

The WEC reports provided more information about how ScienceDirect uses cookies and the "other technologies" mentioned in the Privacy Policy and Cookie Notice. Some of the more notable findings from the WEC summary reports include:

- Cookies and other trackers that collect personal data are typically persistent, sometimes long after the user session ends. Persistent cookies (or cookies not deleted after the user session ends) that are set to automatically expire have long life spans. Six-month to two-year cookie expiration dates are common in our 2022 and 2023 analyses. Some personally-identifying persistent data—either as cookies or local storage—had up to a 50-year expiration date when the websites were tested in 2022 (see Figure C-1 in Appendix C).

- ScienceDirect heavily uses non-cookie tracking methods, such as persistent HTML5 local storage and web beacons:
  - » Websites often use local storage to store data locally in the user browser profile. Data in local storage can persist after a user shuts down their browser. Unlike cookies, data stored in local storage do not have automatic expiration dates.
  - » Web beacons take many forms: a single (1x1) pixel image, an empty file, or a transparent image embedded into the page. Beacons allow for unintrusive (often undetectable without tracker-blocking tools) tracking of user behavior.

  Both methods can bypass common mechanisms used to reduce or prevent user surveillance. Data in local storage can persist even when a privacy-conscious user sets their browser to clear all cookies. Depending on the browser, users wanting to clear data stored in persistent storage will need to adjust additional settings (e.g., in Firefox, setting the browser to delete cookies, site data, and cached web content at browser close). Browser add-ons such as uBlock Origin and Ghostery block some web beacons. Still, many of these add-ons depend on filter lists that may not contain the domain or rule to block a specific web beacon. Using add-ons with current filter lists became more important when ScienceDirect pages in 2023 showed a marked increase in the use of web beacons across the board (see Appendix D for a breakdown of tracking methods on each page per year plus a comparison of total trackers from each year).

- Several cookies and web beacons that appear to come from the sciencedirect.com or elsevier.com domain names redirect to third-party domains. This is standard practice in other markets for placing third-party trackers for advertising and analytics, concealing from users the extent to which user data is collected and disclosed on third-party platforms. One example is smetrics.elsevier.com which resolves to elsevier.com.ssl.d1.sc.omtrdc.net. The omtrdc.net domain is tied to Adobe Experience Cloud, a cloud-based integrated suite of marketing, web analytics, and data-broker products (see Figure C-2 in Appendix C).
  - » Notably, the cookie and beacon host lists from the WEC reports (located in Appendix D) do not include nav.sciencedirect.com or content.id.elsevier.com, which redirects to a ThreatMetrix address (the significance of ThreatMetrix is described further below). Nevertheless, the absence of these domain names does not necessarily mean that ScienceDirect is no longer using ThreatMetrix after users publicly pointed out its use on the website in 2020 (see the "ThreatMetrix and ScienceDirect" section for additional explanation).
- Even when browsing the site without logging into a personal or institutional account, ScienceDirect collects personal data, including IP address and information about the user's browser and device. The cookies and beacons can contain pseudonymous identifiers tied to the collected personal data (see Figures C-1, C-3, C-4, and C-5 in Appendix C).

- Blocking JavaScript to prevent data collection or user tracking breaks the website's basic functionality, such as searching the site and downloading Open Access articles. Privacy-conscious patrons, therefore, lose access to the desired information (see Figure C-6 in Appendix C).

Blacklight reported an additional tracking method on the ScienceDirect website—canvas fingerprinting—designed to identify specific users without using cookies (see Figure C-7 in Appendix C). Canvas fingerprinting uses HTML5 to draw shapes and text on a page, hidden from the end user's view. Differences or variations in the drawing indicate specifics about the user's computer, including hardware and software present. These fingerprints are typically used to protect against fraud and abuse but can also be used to track behavior for user profiling, personalization, and advertising purposes.[30]

Even though the 2022 website analysis did not include the login page (https://id.elsevier.com), the automatic redirection to the login page when testing the search results page during the 2023 analysis raised additional tracking concerns. Using the redirected login URL presented in the WEC report as the test URL, both WEC and Blacklight reported a beacon belonging to Mouseflow (https://mouseflow.com/), a company that specializes in tracking mouse movements and keystrokes (see Figure C-8 in Appendix C). WEC and Blacklight reported a specific script that may be recording use of the login page, including possible keylogging and mouse movements. These recordings can potentially reveal behavior that the user otherwise might not be aware is being captured by the vendor, such as hovering over a link without clicking it or typing text in a form without submitting the form. The same Mouseflow script was not detected when a WEC report was run on the domain name of the login page minus the redirection parameters found in the search results redirect report.

30 Mattu, Surya, and Aaron Sankin. "How We Built a Real-Time Privacy Inspector—The Markup." Accessed November 11, 2022. https://themarkup.org/blacklight/2020/09/22/how-we-built-a-real-time-privacy-inspector.

THREATMETRIX AND SCIENCEDIRECT

In 2020, technologist and researcher Wolfie Christl posted screenshots on Twitter documenting the use of ThreatMetrix on an article page in ScienceDirect.[31] Elsevier states that their use of ThreatMetrix primarily protects the website against fraud and abuse.[32] However, the presence of a product from RELX's risk management portfolio on their journals database raises serious questions about the possibility that data collected by ThreatMetrix through academic products could potentially feed into other risk management services or RELX's broader data-brokering business. The 2022 analysis did not include domain names associated with ThreatMetrix—nav.sciencedirect.com, content.id.elsevier.com, or the h.online-metrix.net—in their first- and third-party host lists. The 2023 analysis did find one domain name that is associated with ThreatMetrix (h-elsevier-fame.online-metrix.net; see Figure C-9 in Appendix C) on the login page.

Regardless, the absence of these domains from the home, browsing, search results, and article pages does not necessarily mean that Elsevier does not feed ThreatMetrix data collected from ScienceDirect users by other means. Even though ScienceDirect's Cookie Notice changed on June 17th, 2022, striking ThreatMetrix by name from the Strictly Necessary Cookies section, users who wish to log into ScienceDirect encounter ThreatMetrix on the login page at id.elsevier.com.[33] The id.elsevier.com domain is used when authenticating user accounts for other Elsevier products, such as Mendeley and Scopus.[34] Given Elsevier's aggregation of personal data across related products when a user creates a personal account, it is feasible that data collected by ThreatMetrix at the time of authentication could be combined with the data in the personal account through matching user login activity between the two data sets.

31 Wolfie Christl [@WolfieChristl]. "Does RELX, the Scientific Publisher, Use Personal/Behavioral Data on Academic Scholars Who Access Publications via Elsevier for Its Data Brokerage and Risk Management Services? I Don't Know. In Any Case, ScienceDirect (RELX/Elsevier) Embeds ThreatMetrix (RELX/LexisNexis Risk). https://T.Co/BqoO38tojQ." Tweet. Twitter, August 18, 2020. https://twitter.com/WolfieChristl/status/1295655040741445632.

32 Elsevier Support Center. "How Does Id.Elsevier.Com Use Cookies?" https://web.archive.org/web/20221017174021/https:/service.elsevier.com/app/answers/detail/a_id/28037.

33 Two versions of ScienceDirect "Use of Cookies" page from June 17, 2022: Version with mention of ThreatMetrix - https://web.archive.org/web/20220617003826/https://www.sciencedirect.com/legal/use-of-cookies and version without mention of ThreatMetrix https://web.archive.org/web/20220617234129/https://www.sciencedirect.com/legal/use-of-cookies; Elsevier Support Center. "How Does Id.Elsevier.Com Use Cookies?" https://web.archive.org/web/20221017174021/https:/service.elsevier.com/app/answers/detail/a_id/28037.

34 Mendeley Support Center. "How Do I Enable Cookies to Allow Sign in for Mendeley Web Importer?" Accessed November 2, 2022. https://web.archive.org/web/20221102212747/https://service.elsevier.com/app/answers/detail/a_id/19574/supporthub/mendeley/kw/cookies/p/16075/; https://www.scopus.com/signin.uri?origin=&zone=TopNavBar redirects to a https://id.elsevier.com login page.

OVERALL DATA PRIVACY ASSESSMENT

Our analysis indicates that, except for Data Security and Accountability, there are no other privacy domains where most of ScienceDirect's data practices appear to exceed Minimum Viable Privacy (MVP). The following table breaks down each privacy domain and the significant findings for each privacy level.

PRIVACY DOMAIN:

Data Collection

EXCEEDS MVP	—
MEETS MVP	• The Privacy Policy lists the type of data collected and the business reasons why the data is collected. • There is no indication that ScienceDirect collects physical or behavioral biometric information[35] when not logged into the site (outside of the login page).
DOES NOT MEET MVP	• Users must enable JavaScript for core site functionality, such as searching and downloading the full text of Open Access articles. In some instances, the HTML full text of Open Access articles does not display if JavaScript is disabled. Enabling JavaScript allows the site to use web beacons and other data collection methods to collect personal data. • Collection of behavioral biometrics and other personal data by ThreatMetrix might occur during the account login process.[36] • Users signing up for an account (or signing into their account) will automatically have Recommendations Service, Reading History, and Search History turned on by default. Users must go into their account settings to "pause" these services and manually delete historical data collected via these features. • As noted in the Privacy Policy and U.S. Consumer Privacy Notice, the categories of data collected (e.g., usage and location data) and the sources where data is harvested (third parties within and outside of RELX) are extensive. In addition, there is the creation of inferred data that can contain potentially sensitive data about an individual's characteristics based on collected personal data. These practices do not abide by data minimization standards.

35 Behavioral biometric data can include typing or swiping speed, mouse movement, shape and pressure of finger on screen, and keyboard behavior (shortcut or special key use).

36 ThreatMetrix customers have the option to add the LexisNexis® Behavioral Biometrics to their ThreatMetrix instance—https://risk.lexisnexis.com/global/en/products/behavioral-biometrics. It is unknown if this feature is enabled for the authentication pages that redirect to id.elsevier.com.

PRIVACY DOMAIN:

User Data Rights

EXCEEDS MVP	—
MEETS MVP	• Depending on applicable regulations, ScienceDirect users have rights to access, correct, restrict processing, delete, and export personal information. • Users can opt-out of certain marketing and invitation communications. • Users can opt-out of specific non-essential data collection in their account settings.
DOES NOT MEET MVP	• It is unclear in the Privacy Policy and the Support Hub if the data rights afforded by data protection and privacy regulations (e.g., GDPR, CCPA/CPRA) are available to ScienceDirect users outside of those states or regions where those regulations apply. If not, ScienceDirect users outside of those specific jurisdictions only have the ability to access to some of their data that appears on account-related pages. • There seems to be no mechanism for all users, regardless of jurisdiction, to opt-out of non-essential data processing.

PRIVACY DOMAIN:

Data Disclosure

EXCEEDS MVP	—
MEETS MVP	• If the Data Processing Addendum (DPA) U.S. Privacy Laws Addendum is included in the Agreement and user data falls under CCPA, the vendor cannot disclose data outside of the business purposes specified in the Agreement unless it is explicitly permitted by the CCPA.
DOES NOT MEET MVP	• There seems to be no mechanism for all users to opt-out of all non-essential data disclosure. • There seems to be no mechanism for users to opt-out of data disclosure/sharing across different Elsevier products when users create a personal account. • The legal reasons for data disclosure are broad and vague and rely on "good faith belief" to determine the appropriateness of disclosure. It is unclear from the Privacy Policy if the internal policy explicitly states that personal data will be disclosed only with a valid court-issued order. • While there is mention of pseudonymization of personal data in the DPA, there is no explicit statement of de-identification of personal data before disclosure to third parties. • Personal data is disclosed to third parties for behavioral advertising and targeted marketing purposes. • Personal data is disclosed to other Elsevier services. While it is technically possible, it is unclear from the documentation and website analysis if personal data is shared with specific data brokering business units in RELX (e.g., LexisNexis Risk Solutions Group). • Personally-identifiable data is disclosed to third parties such as Google, Adobe, Cloudflare, and New Relic through cookies and web beacons. While Facebook is listed on the ScienceDirect Cookie Policy page, WEC or Blacklight did not find a cookie from Facebook on the ScienceDirect site.

PRIVACY DOMAIN:

Data Processing

EXCEEDS MVP	—
MEETS MVP	• If the DPA is included in the Agreement, personal data will only be processed based on the Agreement unless the processing is expressly permitted by the CCPA and other regulations. • If the DPA is included and user data falls under the scope of CCPA, personal data cannot be used, retained, or disclosed for any other purpose outside of its original purpose (secondary use) unless expressly permitted by the CCPA and other regulations.
DOES NOT MEET MVP	• It is unclear whether data processing limitations in the DPA extend to Agreements that do not list the DPA as part of the Agreement. • Data retention practices appear vague, with no explicit statement of secure deletion of personal data after no longer needed for operational purposes. • While the Privacy Policy states the use of aggregation for some cases of data reporting and the DPA states the use of pseudonymization for personal data, the rigor and extent of de-identification of personal data are unknown. • Personal data, including behavioral data, is processed for targeted, personalized advertising and marketing.

PRIVACY DOMAIN:

Privacy Policy

EXCEEDS MVP	—
MEETS MVP	• The publicly available Privacy Policy generally explains the collection, use, and disclosure of personal data provided by the subscriber, user, and third parties.
DOES NOT MEET MVP	• The language in the Privacy Policy is typically not included in the signed Agreement, leaving little recourse for libraries or users with personal accounts to object before the policy changes take place. • The supplemental U.S. Consumer Privacy Notice contains specific data collection, use, and disclosure practices that contradict library or organizational privacy policies that adhere to data protection standards and best practices in the library profession.

PRIVACY DOMAIN:

Data Ownership

EXCEEDS MVP	—
MEETS MVP	• The Terms and Conditions indicate that users retain ownership of certain types of personal data, such as intellectual property and user-generated content (i.e., public comments and discussion posts). • If the DPA is included in the Agreement, personal data is deleted at the end of the business relationship.
DOES NOT MEET MVP	• It is unclear whether the personal data deleted at the end of the business relationship includes aggregate and de-identified personal data. • There are no explicit personal data ownership statements outside the Terms and Conditions. • There is no explicit statement about subscribers or users exercising data ownership rights in the case of a merger or acquisition (M&A). The Privacy Policy states that Elsevier will disclose personal data as part of an M&A, which implies personal data is a company-owned asset.

PRIVACY DOMAIN:

User Surveillance

EXCEEDS MVP	—
MEETS MVP	• Users are not required to create a separate personal account to access core site functionality. • The Privacy Center allows users to opt-in/out of certain types of personal data collection.
DOES NOT MEET MVP	• Users must allow JavaScript to access core site functionality, potentially opening the door to web beacons and local storage to track user behavior on ScienceDirect. • ScienceDirect uses several web tracking methods to evade or work around user privacy protections, such as canvas fingerprinting. • ScienceDirect uses third-party beacons, cookies, and other web tracking methods that could be used to track user behavior outside of the ScienceDirect website. • ScienceDirect widely uses third-party services such as Google DoubleClick and Adobe Target and Audience Manager cookies and web beacons to collect personal data, including user behavior on ScienceDirect. Blocking third-party cookies does not entirely block these third-party services from collecting user data (e.g., web beacons from smetrics.elsevier.com collecting personal data for Adobe marketing and advertising products). • The Privacy Policy and Cookie Notice state that user behavior is tracked across Elsevier products if signed into their account. Based on the information gathered from documentation and website analysis, user tracking occurs regardless of whether the user is signed into ScienceDirect. • It is unclear whether the security and fraud detection mentioned in the Cookie Notice includes behavioral data. It is additionally unclear if "unauthorized use," as mentioned in the agreement, involves the analysis of collected behavioral data.

PRIVACY DOMAIN:

Data Security and Accountability

EXCEEDS MVP	• If the Agreement includes the Elsevier Data Security Schedule, there are specific details about the main requirements and responsibilities as laid out in the Information Security Program, including technical, physical, and administrative information security policies and practices. The Data Security Schedule also states Elsevier's role and responsibilities in its incident response plan. Finally, the Schedule explicitly grants the subscriber the right to audit specific data protection documentation.
MEETS MVP	• If the Agreement includes the DPA, subscribers have the right to a yearly audit of Elsevier's compliance with the DPA. However, the only mention of the audit being conducted by an independent third party is in the Colorado, Connecticut, and Virginia section of the U.S. Privacy Laws Addendum for the DPA.
DOES NOT MEET MVP	• If the Agreement does not include the DPA or the Elsevier Data Security Schedule, the subscriber's right to audit data security/privacy documentation or practices is unclear from existing documentation. • Existing documentation outside the DPA and the Data Security Schedule does not provide detailed information about data security and accountability. While the practices are in place to meet contractual requirements as laid out in the DPA and Data Security Schedule, Subscribers without these documents as part of the Agreement are limited to what they can negotiate.

LIMITATIONS AND FURTHER INVESTIGATION

Our analysis encompassed a variety of public documentation, contracts, and the front-end (user-facing) website, providing a broad overview of ScienceDirect's data privacy practices. Nevertheless, there are limitations to our analysis based on the scope of the analysis and the nature of the resources used. Reviewing compliance with specific data protection and privacy regulations was beyond the scope of our analysis, partly because a legal review heavily depends on which libraries and users fall under which jurisdictions or scopes of specific regulations. Sections of the rubric overlap with rights and requirements governed by regulations, such as data user rights. Still, our analysis is not a legal review of whether ScienceDirect complies with specific laws.

The analysis provides a reasonably high-level summary of what data is collected, processed, disclosed, and retained based on publicly available information. Contract language variability prevented a complete picture of data flows and practices, particularly with the inconsistent inclusion of data protection schedules and signed DPAs among the contracts. The lack of sign-in credentials also limited the ability of the analysis to determine any additional data collection that might happen when a user is logged into the site. Some of these limitations can be addressed in future analyses, but lack of access to the back-end systems will most likely leave gaps in any external analysis.

KEY FINDINGS

Based on our assessment with the Licensing Privacy Rubric and website testing, **ScienceDirect is an example of a vendor product with data privacy practices that directly conflict with library privacy standards and guidelines.** Some of these problematic data privacy practices are clearly stated by Elsevier; others are not. The Privacy Policy and U.S. Consumer Privacy Notice provide an overview of how personal data—including behavioral data—is collected and used for targeted marketing and advertising purposes and disclosed to affiliates and partners within Elsevier, RELX, and others outside the parent company. We also know from the Privacy Policy that there is data sharing between Elsevier services when a user creates a personal account, opening a user profile containing personal data (including behavioral data) across those services.[37] The specifics in the U.S. Consumer Privacy Notice of how personal data and the user profile are shaped by data from other data brokers and the extent of sharing back to these brokers and third parties should give users and institutions significant concerns when evaluating ScienceDirect's data privacy practices.

The magnitude of user tracking, data collection, and data disclosure to third parties is not apparent until one goes past the main contract and Privacy Policy—where users are less likely to find them. For example, the U.S. Consumer Privacy Notice gives information about specific personal data collected and disclosed to specific third parties. The Cookie Notice for ScienceDirect and Elsevier describes the advertising and analytic products used to collect, track, and process personal data. The id.elsevier.com documentation also points to the continued use of ThreatMetrix on ScienceDirect login pages even though the ScienceDirect Cookie Notice no longer lists it as one of the security cookies in use on the website. Even when the user is not signed into a personal account, personal and behavioral-tracking data (e.g., session ID, IP address, digital fingerprint) can be found in the cookies, web beacons, and local storage that may persist across browsing sessions if users do not thoroughly clear that data after every browsing session.

The analysis also came across several public documents from Elsevier describing the purported benefits of personalization through personal data. These arguments are tied to Elsevier's Privacy Principles:

37 From Elsevier's Privacy Policy, under the "Your personal account" section—"We will share your usage activity, preferences and other information amongst these and related Elsevier services, like bepress and PlumX, to help you improve your productivity. For example, ScienceDirect may recommend better content for you based on your Mendeley library content and the abstracts you read on Scopus, and Mendeley may present you a better tailored content feed based on your ScienceDirect and Scopus activity."

value, transparency, choice, anonymization, and accountability.[38] To the first principle—value—some users will find the personalization services that come with creating a personal account to be of great value in their work. Others might not, and some mechanisms are in place where users can opt out of some data collection and sharing. However, personal account holders who do not want the whole personalization experience might not realize that ScienceDirect begins collecting their search and reading histories as soon as they create an account.

Overall, users have limited options for controlling the collection and disclosure of their data. For users who opt into personalization features, the ability to control the disclosure and processing of personal data by ScienceDirect and third parties is limited. In certain types of data disclosure and processing, opting out appears unlikely if a user is not granted the same rights as a user in the EU or California. While there is some mention of pseudonymization of personal data in the DPA, there is no indication in the public documentation that the data is consistently and rigorously de-identified when disclosed to third parties, including other Elsevier and RELX services and products. This extends to user tracking across the Elsevier Research Products group, creating a user profile constructed from detailed personal data (including behavioral, geolocation history, and digital fingerprints) and inferred data that can infer characteristics of a person without directly collecting that data.

Even though several data protection regulations prohibit reusing personal data for purposes other than the originally intended use, not all users are covered under these regulations, including many US-based ScienceDirect users. Like personal data, it is unclear whether or how other RELX services and products would exactly utilize these user profiles.

The substantial changes in documentation and website functionality over the course of one year highlight the power asymmetries between vendors and institutions. Few, if any, institutions have the resources necessary to dedicate sufficient staff time to regularly monitor an interconnected web of privacy policies and practices for changes (which are not included in contracts themselves), analyze the effect of and risks posed by any changes, and constantly revise institutional policy accordingly. Again, an example of such a substantial change is the superseding of the CCPA Notice with the U.S. Consumer Privacy Notice which provides some of the clearest descriptions of problematic data privacy practices to date.

Relying on the company's word—the privacy policy—without having stringent data privacy protections explicitly spelled out in the signed contract is not a viable strategy for protecting library user privacy. The lack of language around data privacy in most of the publicly available signed ScienceDirect Subscription Agreements compounds the concerns above. Out of the analyzed contracts, only one

38 Elsevier. "Elsevier Privacy Principles," August 22, 2022. https://web.archive.org/web/20220822132224/https://www.elsevier.com/about/policies/privacy-principles.

included a signed Data Processing Addendum, while one other included a data security schedule. Any reference to specific data privacy practices or policies in the contracts took the form of a mention of the existence of the Privacy Policy. *In US contract law, courts have not consistently recognized privacy policies as contracts.*[39] Specific regulations, such as the California Online Privacy Protection Act and GDPR, require companies to publish privacy policies publicly, but there is no legal guarantee that these policies would be considered enforceable contracts.[40]

In addition, any changes in the Privacy Policy do not automatically trigger a renegotiation of the signed contract, leaving libraries with limited options when a change in policy or documented practices further conflicts with patron privacy. Public privacy policies have many functions, but they primarily serve to minimize compliance liability and are not designed for or to be readable by the average user.[41]

39 Citron, Danielle Keats, and Daniel J. Solove. "Privacy Harms." SSRN Scholarly Paper. Rochester, NY, February 9, 2021. https://doi.org/10.2139/ssrn.3782222.

40 Tarr, Madelyn. "Accountability Is the Best (Privacy) Policy: Improving Remedies for Data Breach Victims Through Recognition of Privacy Policies as Enforceable Agreements." *Georgetown Law Technology Review* 3, no. 1 (2018): 40. https://georgetownlawtechreview.org/wp-content/uploads/2019/01/3.1-Tarr-pp-162-201.pdf.

41 Cranor, Lorrie Faith. "Necessary but Not Sufficient: Standardized Mechanisms For Privacy Notice and Choice." J. on Telecomm. & High Tech. 10 (2012): 36. https://www.law.nyu.edu/sites/default/files/upload_documents/Cranor%20-%20Necessary%20but%20Not%20Sufficient.pdf; Waldman, Ari Ezra. "Privacy, Notice, and Design." SSRN Scholarly Paper. Rochester, NY, March 16, 2016. https://doi.org/10.2139/ssrn.2780305.

SUGGESTED ACTIONS

Users and institutions should carefully consider these new and growing privacy risks in the context of the transition by vendors like Elsevier into data analytics businesses that provide increasingly integrated research platforms. Types of data collection and use that may be less concerning by a business solely involved in publishing (e.g. search history, logs of content accessed, sensitive personal information, geolocation data) can become significantly more concerning when users consider how this same data may be used for future analytics products assessing productivity or integrated into databases of personal information that may be disclosed to others (including data brokers, corporations, governments, and law enforcement agencies). These possibilities underline the urgency of a proactive, multi-pronged approach to protecting the privacy and best interests of users, both collectively and by individual institutions.

Libraries should expect that ScienceDirect will track users in much the same way as e-commerce, news, and social media sites. The data privacy practices described in the analysis are ones that users encounter through everyday interactions with businesses and organizations that track and harvest user data to sustain profitable data-driven business models. The widespread data collection, user tracking and surveillance, and disclosure of user data inherent to these business models run counter to the library's commitment to patron privacy as described in the ALA Library Bill of Rights (LBoR).[42] The library's responsibility to protect the patron's right to privacy is expanded upon in the interpretation of the LBoR:

> *The right to privacy includes the right to open inquiry without having the subject of one's interest examined or scrutinized by others, in person or online… Lack of privacy and confidentiality has a chilling effect on users' selection, access to, and use of library resources. All users have a right to be free from any unreasonable intrusion into or surveillance of their lawful library use.*[43]

Libraries provide some level of privacy for patrons who use physical library resources. On a local level, libraries can decide what data to collect, how to use the data, and how others can use the data if they

42 ALA. "Library Bill of Rights." https://www.ala.org/advocacy/intfreedom/librarybill.

43 ALA. "Privacy: An Interpretation of the Library Bill of Rights." https://www.ala.org/advocacy/intfreedom/librarybill/interpretations/privacy.

even have access. However, **the shift from in-house physical resources to online content platforms managed by third-party vendors has largely shifted the control over what data is collected and tracked from libraries to vendors.**[44] There is little to nothing to stop vendors who collect and track patron data from feeding that data—either in its raw form or in aggregate—into their data brokering business. The contribution of library patron data into the data broker economy through the use of library products and services also runs counter to the user's rights to privacy and confidentiality in the library:

> *The library profession has a long-standing ethic of facilitating, not monitoring, access to information… Libraries should not monitor, track, or profile an individual's library use beyond operational needs.*[45]

While libraries continue to describe themselves as protectors of patron privacy, the marketplace shifted toward business models based on commercial surveillance.[46] Libraries must be proactive in changing their dealings with content platform providers to reflect this reality and better protect users. There are a variety of steps that libraries can consider to take action in both the shorter and longer term. Some of the immediate actions library might consider include:

- **Reviewing existing contracts for inclusion of data protection and security documents**
  - » Check if the Data Processing Addendum (DPA) and Elsevier Data Security Schedule are included in the signed contract. These documents provide specific data security and some privacy protections for patron data that are enforceable if added to the signed contract. If they are not included, work with the vendor to include these documents in the signed contract.
  - » Check with other departments at the institution for other Elsevier contracts and review for any conflicts or inconsistencies that might arise from pursuing additions or modifications to the library contracts, such as the inclusion of a DPA or Data Security Schedule (e.g., a contract for Interfolio from Academic Affairs or the Office of Research).

44 For more information about the differences around privacy affordances between physical and electronic resources, see Dorothea Salo's "Physical-Equivalent Privacy," https://minds.wisconsin.edu/handle/1793/81297.

45 ALA. "Privacy: An Interpretation of the Library Bill of Rights." https://www.ala.org/advocacy/intfreedom/librarybill/interpretations/privacy.

46 This shift has drawn concern from the U.S. Federal Trade Commission, which has begun a rulemaking process related to "commercial surveillance and data security." See https://www.ftc.gov/legal-library/browse/federal-register-notices/commercial-surveillance-data-security-rulemaking. As part of this process, SPARC submitted comments that outline concerns related to commercial surveillance in academic markets. See https://downloads.regulations.gov/FTC-2022-0053-1082/attachment_1.pdf.

- **Reviewing and revising policies and settings around public computing security and privacy**
  - » Install privacy-protecting browsers (e.g., Firefox, Tor) and browser add-ons (e.g., uBlock Origin, Privacy Badger) on public computers to limit tracking through cookies or beacons.
  - » Configure public computers to automatically reset the system after each user session (i.e., reimaging). This will remove data stored on the computer from the user session, including persistent data in local storage.

  Library workers can use the following resources to assist them in reviewing and securing public computing:

  - » Electronic Frontier Foundation's (EFF) Cover Your Tracks (https://www.eff.org/pages/cover-your-tracks)
  - » ALA's Library Privacy Guidelines and Checklists (https://www.ala.org/advocacy/privacy/guidelineschecklists), specifically *Assistive Technologies and Public Access Computers and Networks*
  - » Choose Privacy Every Day's "Guidelines for Private Online Searching & Browsing" (https://chooseprivacyeveryday.org/guidelines-for-private-online-searching-browsing/)
  - » Security-in-a-Box (https://securityinabox.org/en/)
  - » Digital Privacy & Technology Guide's Privacy Resources for Library Tech Management (https://guides.masslibsystem.org/digital-privacy-and-technology/resources-for-lib-tech-management)

- **Educating campus users about ScienceDirect data privacy practices**
  - » Use the forthcoming talking points that are currently in production by the Resource Library Working Group of SPARC's Privacy & Surveillance Community of Practice for library workers to use in instructional or reference sessions about how personal data is collected, used, and shared by ScienceDirect, Elsevier, RELX, and third parties based on the Privacy Policy and other public documentation.[47]
  - » Provide information and training around privacy-protecting tools for campus users who want to protect their privacy while using ScienceDirect.
  - » Update library privacy notices or other public policies about vendor privacy practices to include the level of tracking and disclosure of user data by ScienceDirect.

47 SPARC. "Privacy and Surveillance Community of Practice." https://sparcopen.org/our-work/privacy-and-surveillance-community-of-practice.

» Work with campus privacy advocates, concerned campus community members, IT, instructional technologists, and people involved in vendor selection and negotiation in addressing the campus' continued relationship with Elsevier (and other vendors with similar data privacy practices) and how the library will protect user privacy on the users' behalf.

Library workers can use the following resources to assist them in their patron communication and education work:

» ALA's Privacy Field Guides (https://libraryprivacyguides.org/), specifically *How to Talk About Privacy, Privacy Policies, and Vendors and Privacy*

» Digital Privacy & Technology Guide's Privacy Resources for the Public (https://guides.masslibsystem.org/digital-privacy-and-technology/resources-for-the-public)

» Library Freedom Project Resources (https://libraryfreedom.org/resources/)

» Digital Library Federation (DLF) Privacy and Ethics in Technology Working Group

— Digital Privacy Instruction Curriculum (https://osf.io/sebhf/)

— Advocacy Action Plan (https://osf.io/2smrf/)

» Digital Shred's Privacy Literacy Toolkit (https://sites.psu.edu/digitalshred/)

It is critical to note that the above steps are only short-term workarounds that limit some of the more well-documented surveillance harms of profitable data-intensive business portfolios. Though entrenched vendor data practices do not change overnight, longer-term strategies with widespread community support could create productive pressure for changes in ScienceDirect's data privacy practices. Toward this end, there are additional steps that individual libraries can consider taking over the longer term. These include:

- **Negotiating stronger privacy terms in vendor contracts.** The Navigating Risk in Vendor Data Privacy Practices Project will provide tools and resources for libraries to negotiate, such as model contract language for contract negotiations. SPARC's Privacy & Surveillance Community of Practice[48] can also support libraries in sharing information and experiences in trying to secure stronger privacy contract terms.
- **Removing confidentiality clauses that prevent sharing of privacy terms.** Libraries must be free to discuss the privacy terms they successfully negotiate (and those they do not) in order to effectively learn from and leverage others' experiences. By refusing to sign NDAs and

48 SPARC. "Privacy and Surveillance Community of Practice." https://sparcopen.org/our-work/privacy-and-surveillance-community-of-practice.

sharing terms, libraries can reduce information asymmetries with vendors when negotiating. SPARC's brief resource on pushing back against NDAs[49] may be useful to libraries pursuing this strategy.

- **Recalibrating relationships with vendors whose privacy practices conflict with library expectations and could pose risks to users.** If a vendor is unwilling to commit to sufficient contractual privacy guarantees, libraries may consider adjusting their spend with that provider over time. As more libraries unbundle from "Big Deal" subscription packages,[50] there is growing experience with how to successfully navigate the transition to a dramatically lower spend with a given vendor while maintaining access to content through alternative means.[51] Those institutions interested in considering an unbundling project can find resources through SPARC, such as the Big Deal Cancellation Tracker[52] and Negotiations Community of Practice.[53]

Through a combination of the above actions, individual academic libraries can better shore up privacy protections while raising awareness about vendor data privacy practices with patrons. This can reinforce and extend existing privacy advocacy work on campus. However, individual libraries can only change the landscape so much. Institutions and consortia can share their experience pushing back on vendor surveillance practices, leverage individual advances, and collaborate to build and support privacy-preserving scholarly infrastructure.

There could be a market opportunity for a vendor that can provide such an alternative. It is not uncommon for a vendor to become a more prominent player in the library marketplace after securing a large client for the long term—the ILS marketplace in the late 1980s to the early 2000s is one example.[54] Building up privacy-friendly alternatives in the marketplace, whether for profit or non-profit, will take time.

49 SPARC. "Pushing Back Against Confidentiality Clauses & Non-Disclosure Agreements." https://sparcopen.org/our-work/big-deal-knowledge-base/confidentiality-clauses-and-ndas.

50 SPARC. "Big Deal Cancellation Tracking." https://sparcopen.org/our-work/big-deal-cancellation-tracking/; Aiwuyor, Jessica. "More Research Libraries Decline 'Big Deal' Subscription Contracts with Publishers." Association of Research Libraries (blog), May 14, 2020. https://www.arl.org/news/more-research-libraries-decline-big-deal-subscription-contracts-with-publishers.

51 SPARC. "Recommendations for Providing Alternative Access After a Big Deal Cancellation." https://sparcopen.org/our-work/negotiation-resources/alternative-access.

52 SPARC. "Big Deal Cancellation Tracking." https://sparcopen.org/our-work/big-deal-cancellation-tracking.

53 SPARC. "Negotiation Community of Practice." https://sparcopen.org/our-work/negotiations-community-of-practice.

54 A specific example would be Innovative Interfaces, Inc. (III) and their contracts with large library consortia such as OhioLINK starting back in 1991 — https://librarytechnology.org/document/4943.

While challenging, it is possible to provide some of the same personalization services in a privacy-friendly alternative that is as stable and mature as the ones found in existing products like ScienceDirect. The recent advances in privacy-enhancing technologies (PETs) provide viable solutions for privacy-forward companies wanting to build more responsible and ethical personalization or recommendation features into their products that do not compromise their users' privacy.[55]

Libraries still have the power to shift the marketplace to once again reflect librarianship's commitment to patron privacy. Any large-scale approach to create alternatives will take time and resources. More importantly, large-scale approaches require a long-term commitment from a critical mass of individuals and organizations. Past actions to build protection mechanisms and meaningful alternatives were primarily successful thanks to the countless hours of privacy work, relationship-building, and advocacy by both individual library workers and organizations. Deliberate, dedicated, sustained action on a local level is a necessary foundation for cross-institutional collaborations that can shape markets and provide valuable resources, partnerships, and commitment needed for these actions to succeed.

In collaboration with the many other organizations and individuals that are actively leading library privacy work, SPARC is committed to supporting libraries in addressing vendor privacy concerns. This work is an essential component of SPARC's broader commitment to ensure that privacy is foundational to equitable systems for openly sharing knowledge.

55 Readers wanting to learn more about the latest PET developments should review the recent proceedings from the PETs Symposium — https://www.petsymposium.org/popets.

APPENDIX A
— *Documents Used for Rubric Analysis*

Most of the links to the documents point to two archived versions of the page in the Wayback Machine: one from the Summer 2022 analysis period and the other from the Summer 2023 analysis period.

- **Signed Elsevier Subscription Agreement contracts from academic libraries in the US**
  - 2022 version—https://web.archive.org/web/20220517200458/https://sparcopen.org/our-work/big-deal-knowledge-base/contracts-library/
  - 2023 version—https://web.archive.org/web/20230609104304/https://sparcopen.org/our-work/big-deal-knowledge-base/contracts-library/
- **CCPA Notice** (superseded by Elsevier U.S. Consumer Privacy Notice in 2023)
  - 2022 version—https://web.archive.org/web/20220402191824/https://www.elsevier.com/legal/california-privacy-notice
- **Elsevier U.S. Consumer Privacy Notice** (preceded by CCPA Notice in 2022)
  - 2023 version—https://web.archive.org/web/20230814181931/https://beta.elsevier.com/legal/us-consumer-privacy-notice
- **Elsevier Cookie Notice**
  - 2022 version—https://web.archive.org/web/20220802081210/https://www.elsevier.com/legal/cookienotice
  - 2023 version—https://web.archive.org/web/20230801143708/https://www.elsevier.com/legal/cookienotice
- **Elsevier Data Processing Addendum**
  - 2022 version—https://web.archive.org/web/20220202112633/https://www.elsevier.com/legal/data-processing-terms/processing-terms
  - 2023 version—https://web.archive.org/web/20230122163253/https://www.elsevier.com/legal/data-processing-terms/processing-terms
- **Elsevier Data Security (Schedule 3)**
  - https://web.archive.org/web/20220601033406mp_/https://sparcopen.org/wp-content/uploads/2021/03/ElsevierColoradoAllianceContract2021-2023signed.pdf

- **Elsevier Privacy Policy**
  - 2022 version—https://web.archive.org/web/20220731043736/https://www.elsevier.com/legal/privacy-policy
  - 2023 version—https://web.archive.org/web/20230731225516/https://www.elsevier.com/legal/privacy-policy
- **Elsevier Privacy Principles**
  - 2022 version—https://web.archive.org/web/20220822132224/https://www.elsevier.com/about/policies/privacy-principles
  - 2023 version—https://web.archive.org/web/20230504133532/https://www.elsevier.com/about/policies/privacy-principles
- **Privacy Center** (login access is required to access setting information)
  - 2022 version—https://web.archive.org/web/20220722182107/https://privacy.elsevier.com/
  - 2023 version—https://web.archive.org/web/20230725094224/https://privacy.elsevier.com/
- **ScienceDirect Cookie Notice**
  - 2022 version—https://web.archive.org/web/20220729093603/https://www.sciencedirect.com/legal/use-of-cookies
  - 2023 version—https://web.archive.org/web/20230801070830/https://www.sciencedirect.com/legal/use-of-cookies
- **Support Center documentation**
  - id.elsevier.com use of cookies
    - https://web.archive.org/web/20221017174021/https://service.elsevier.com/app/answers/detail/a_id/28037/
    - Page last updated in 2020
  - Reading history
    - https://web.archive.org/web/20221017173639/https://service.elsevier.com/app/answers/detail/a_id/28358/
    - Page last updated in 2021
  - Recommendations service
    - https://web.archive.org/web/20220624003340/https://service.elsevier.com/app/answers/detail/a_id/18726/c/10547/supporthub/sciencedirect/
    - Page last updated in 2021

  - » Search history
    - — https://web.archive.org/web/20221017174024/https://service.elsevier.com/app/answers/detail/a_id/34024/
    - — Page last updated in early 2022
- **U.S. Privacy Laws Addendum to Elsevier Data Processing Addendum**
  - » 2023 version—https://web.archive.org/web/20230821161812/https://www.elsevier.com/legal/data-processing-terms/us-privacy-addendum
- **Website Terms and Conditions**
  - » 2022 version—https://web.archive.org/web/20220802081313/https://www.elsevier.com/legal/elsevier-website-terms-and-conditions
  - » 2023 version—https://web.archive.org/web/20230801010009/https://www.elsevier.com/legal/elsevier-website-terms-and-conditions

APPENDIX B
— *Glossary*

N.B.—Several term definitions come from the Licensing Privacy Glossary (https://publish.illinois.edu/licensingprivacy/glossary/), published under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Aggregation
A method of de-identification that reduces the granularity of personal data through grouping data into categories or ranges (such as using age ranges to report on user birthdate or age data). Aggregation carries some risk of **re-identifying** an individual if there are outliers in the data set, if the data set size is small overall, or if categories or ranges are granular enough to identify an individual in the data set.

Behavioral tracking
The practice of surveilling users' actions over a period of time.

Canvas fingerprinting
A specific browser fingerprinting method that uses the HTML5 canvas element to track unique users on a website. This method draws shapes and text on a page hidden from the end user. Differences or variations in the drawing indicate specifics about the user's browser and computer, including hardware and software. These fingerprints are typically used to protect against fraud/abuse but can also be used to track behavior for user profiling, personalization, and advertising purposes.

Consent
The act of giving or denying permission for the use or disclosure of an individual's data. Explicit consent requires an affirmative action from the individual, while implicit consent implies consent, such as continued use of a service or website.

Cookies
A small data file sent from the website and stored on the user's computer or the user's browser. Web cookies can be used to manage user authorization and session management, as well as track users through web analytic software and other tracking software. Some cookies last until a web session ends ("session" cookies), while other cookies ("persistent" cookies) can remain on the user's computer after the end of a session. A website can have web cookies from the site itself ("first-party" cookies) as well as cookies from external sites ("third-party" cookies).

Data breach (or data leak)
The unauthorized access of personal data by an individual, organization, or system process. A data breach is an intentional act of gaining unauthorized access, such as an attacker gaining access to personal data for the purpose of identity theft, while a data leak is unintentional, such as an employee losing a laptop or mobile device containing personal data.

Data brokers
Entities that sell personal data collected from private and public data sets.

De-identification
The process of transforming personal data to remove identifiable aspects of the data. This includes a variety of methods, including aggregation, stripping or truncating personal data, or removal or pseudonymization of personal data. De-identified data carries a risk of **re-identification**, meaning that an individual person can be identified from the dataset by reattaching parts of the dataset back to the individual.

Digital fingerprinting
A set of data about a person's device, browser, and other hardware and software that, when combined, can identify an individual. Often referred to as **browser fingerprinting**.

Digital surveillance
The act of monitoring and capturing a person's activities through various technologies, including web analytics, cookies, trackers, and other data observation and capture techniques.

Domain name system (DNS)
The service that associates numerical Internet Protocol (IP) addresses with domain names, which are more friendly for everyday human use. A domain name can point to another domain name through a **Canonical Name (CNAME) record**. The domain name mapped to the other domain name is called an **alias**.

Incident response
The process of responding to and managing a data breach or leak. This can include:

- Identification and detection of a breach or leak
- Containing and eliminating the cause of the breach or leak
- Communications with affected parties

Local storage

HTML5 local storage allows for offline storing of data in the client web browser. Compared to cookies, local storage has a larger storage capacity to store potentially personally identifiable data, persists after the end of the user session and browser shutdowns, and does not have an automatic expiration date.

Opt-in/out

Opt-in is a choice made by a user involving active affirmation, such as checking a box or toggle to turn on specific data sharing or collection settings.

Opt-out is a choice made by a user through inaction unless the user makes an action to choose otherwise. An example is a product collecting user data until the user unchecks the box that controls the data collection setting.

Personal data

Data relating to an identifiable individual. This includes single points of data that can identify a person (direct identifier); data that, when combined, can identify a person (indirect identifiers); and data about a person's behaviors (behavioral data).

Direct identifiers can include:

- Name
- Email or physical address
- Government or organization-issued identification numbers
- Account username and password
- Biometric information
- IP address
- Device information (operating system, browser, device unique ID, etc.)

Indirect identifiers can include:

- Age or date of birth
- Race/ethnicity
- Gender identity
- Education level
- Major or minor field of study
- Disability status
- Veteran status
- Geographical information, such as regions or zip code

Behavioral data can include:

- Search history
- Electronic content access histories
- Circulation histories
- Website activity
- Geolocation history

Web analytics

The collection and analysis of website data. Web analytic applications track and capture data from a variety of sources, including user data. Depending on the application, user data can range from search term results, page hits, and landing/exit pages to user demographic data, behavioral data, and even data of users visiting sites outside of the original website.

Web beacon

A web tracking method that is often undetectable by the end user. Web beacons commonly take the form of a small image file (often a single transparent pixel) loaded with a web page or email but can also be embedded in visible graphics (e.g., buttons, banners) and HTML elements. Beacons can collect personal data such as behavioral data, digital fingerprints, and IP addresses.

APPENDIX C
— *Screenshots of Web Test Reports*

FIGURE C-1.

Persistent Data Analysis

The evidence collection tool analysed persistent cookies after the browsing session. Web pages can also use the persistent HTML5 *local storage*. The subsequent section lists its content after the browsing.

Cookies linked to First-Party Hosts

#	Host	Path	Name	Expiry in days
1	www.sciencedirect.com	/search	sd_search	90
2	www.sciencedirect.com	/search	search_ab	60

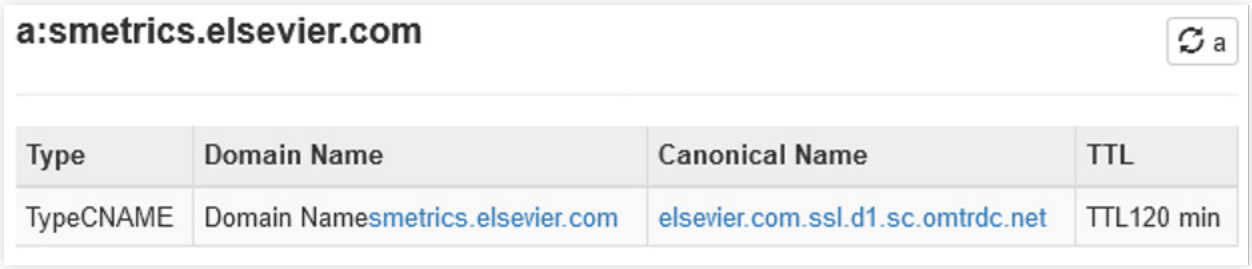
In total, 2 first-party cookies were found.

Cookies linked to Third-Party Hosts

#	Host	Path	Name	Expiry in days
1	sciencedirect.com	/	utt	18250
2	elsevier.com	/	utt	18250
3	sciencedirect.com	/	id_ab	7305
4	sciencedirect.com	/	EUID	7305

Screenshot of the first part of the Persistent Data Analysis from the 2022 WEC report of the search results page. The Cookies linked to Third-Party Hosts table shows cookies from sciencedirect.com and elsevier.com containing potential session id and other unique identifiers with 20- and 50-year expiration dates. The 2023 Blacklight report of the same page reported no change to the id_ab and EUID cookies while the utt cookie was not found.

FIGURE C-2.

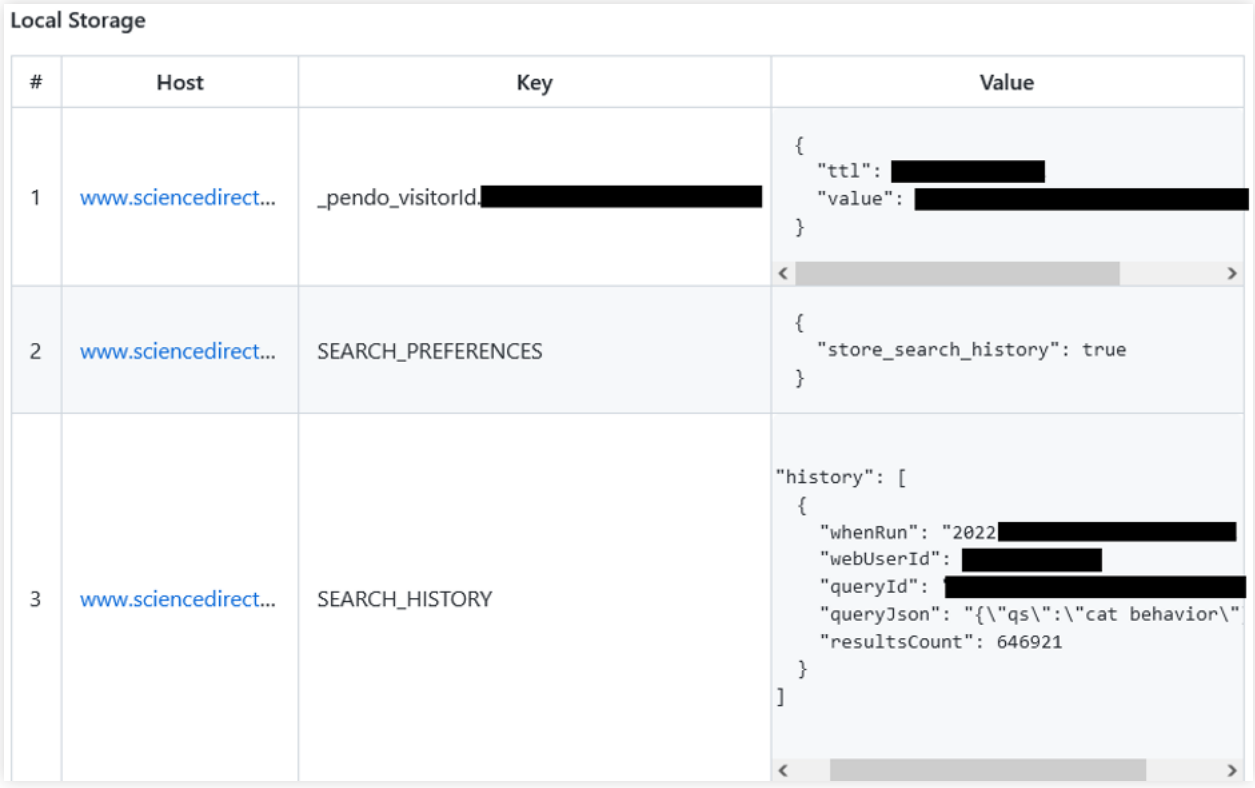


a:smetrics.elsevier.com

Type	Domain Name	Canonical Name	TTL
TypeCNAME	Domain Namesmetrics.elsevier.com	elsevier.com.ssl.d1.sc.omtrdc.net	TTL120 min

Screenshot of the MXToolbox DNS report on smetrics.elsevier.com resolving to elsevier.com.ssl.d1.sc.omtrdc.net.

FIGURE C-3.

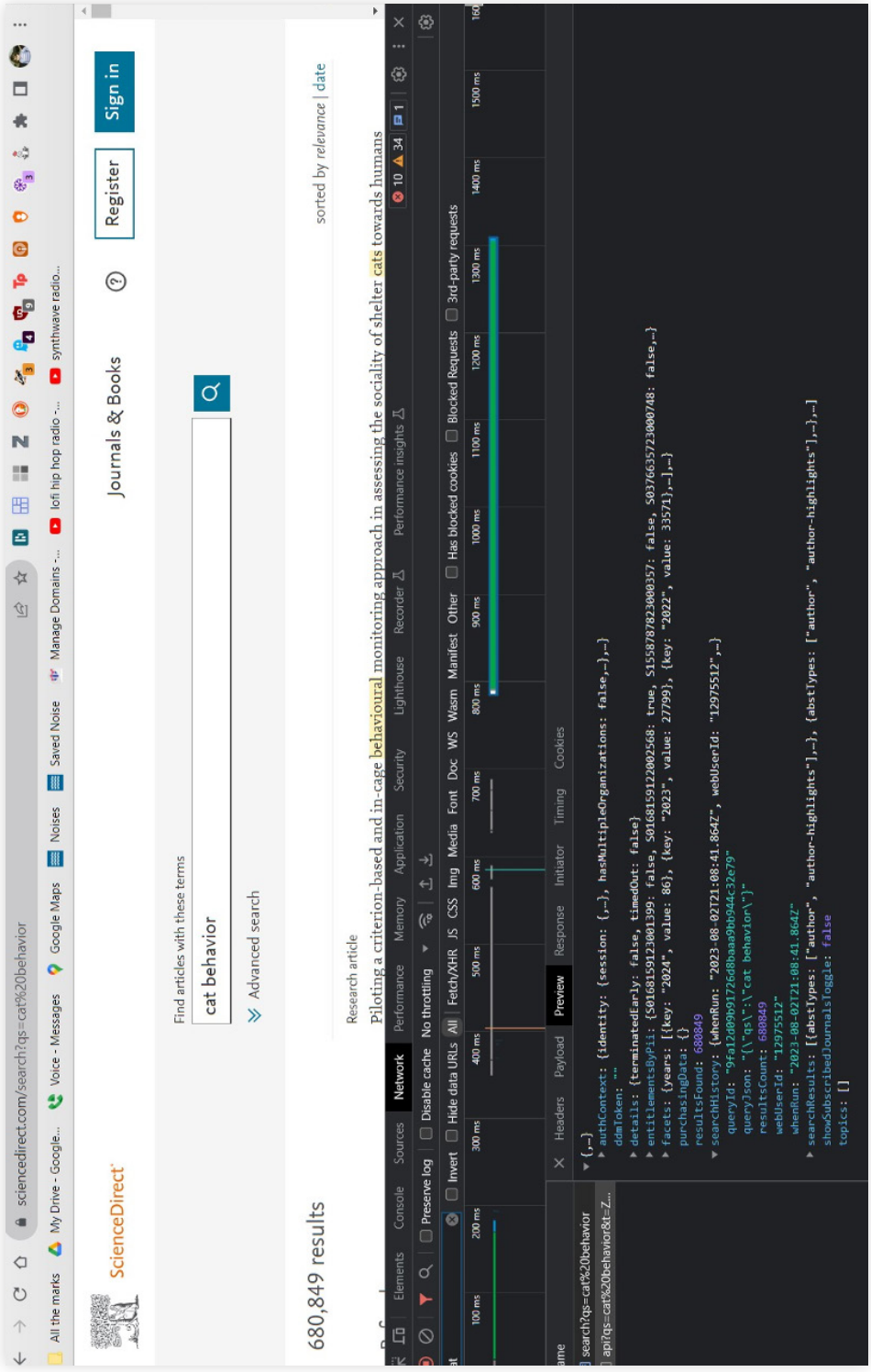


Local Storage

#	Host	Key	Value
1	www.sciencedirect...	_pendo_visitorId.[redacted]	{ "ttl": [redacted] "value": [redacted] }
2	www.sciencedirect...	SEARCH_PREFERENCES	{ "store_search_history": true }
3	www.sciencedirect...	SEARCH_HISTORY	"history": [{ "whenRun": "2022[redacted] "webUserId": [redacted] "queryId": [redacted] "queryJson": "{\"qs\":\"cat behavior\" "resultsCount": 646921 }]

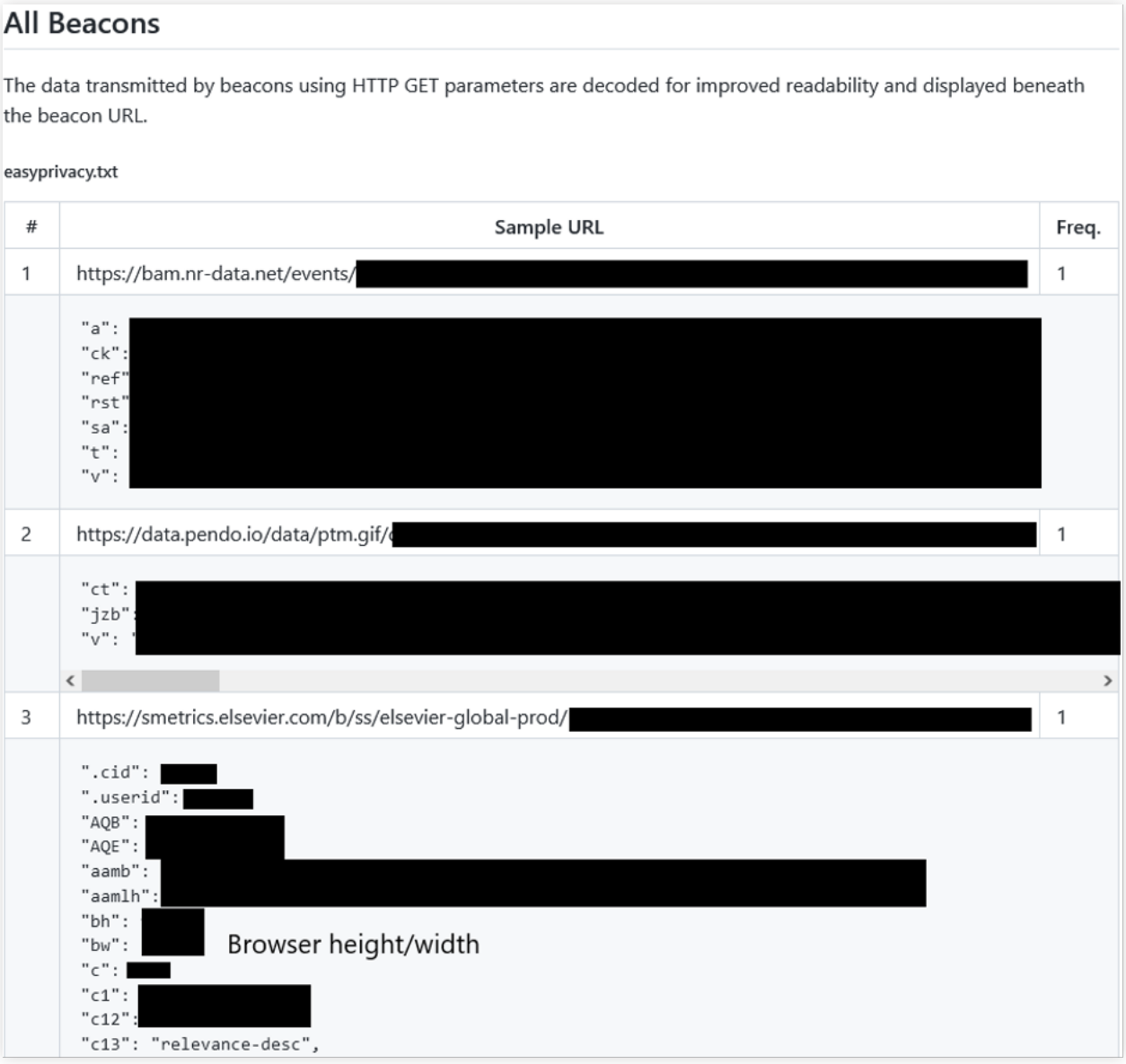
Screenshot of the first part of the Local Storage list from the 2022 WEC report of the search results page. The bottom two items on the list show that the search results page is collecting information on search history, including query information.

FIGURE C-4.



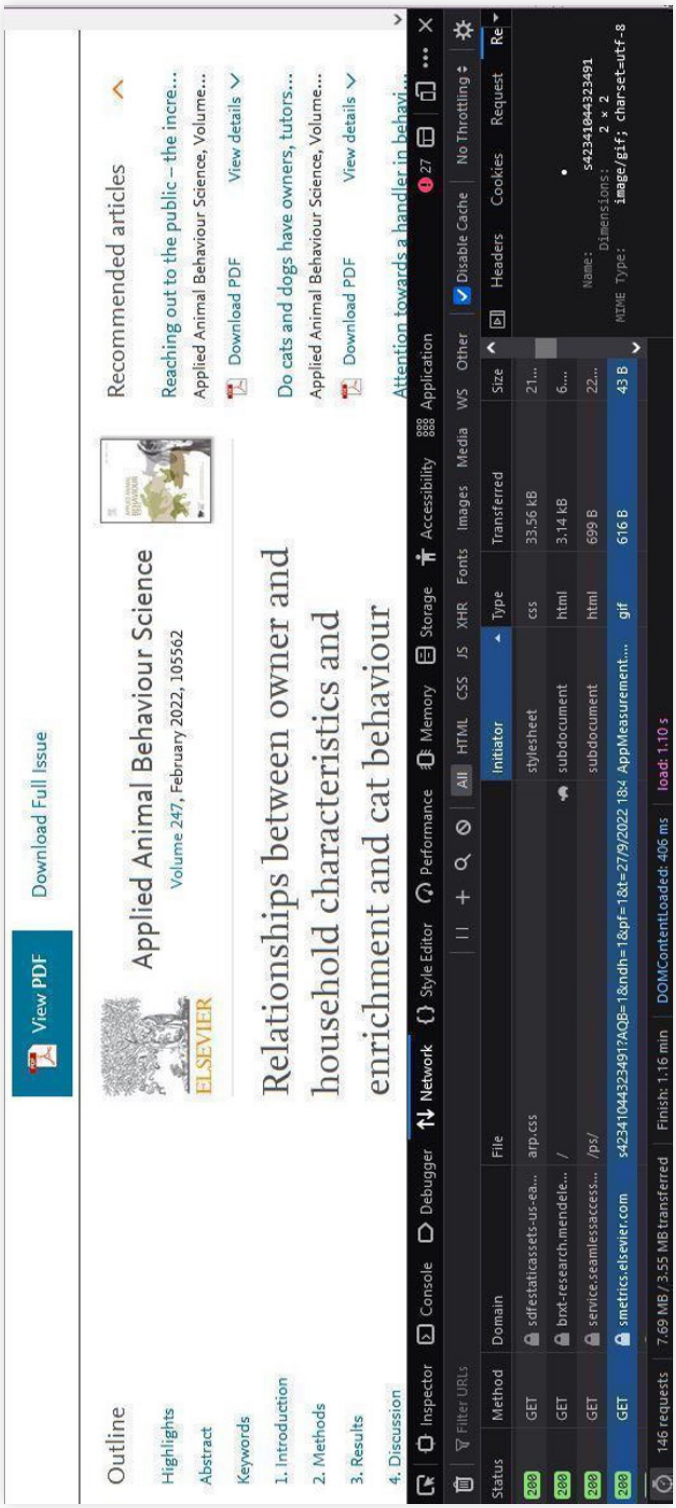
A screenshot of the search results page in 2023 using Chrome's web developer tools that shows a call to an API script that contains information about search history under a searchHistory parameter.

FIGURE C-5.



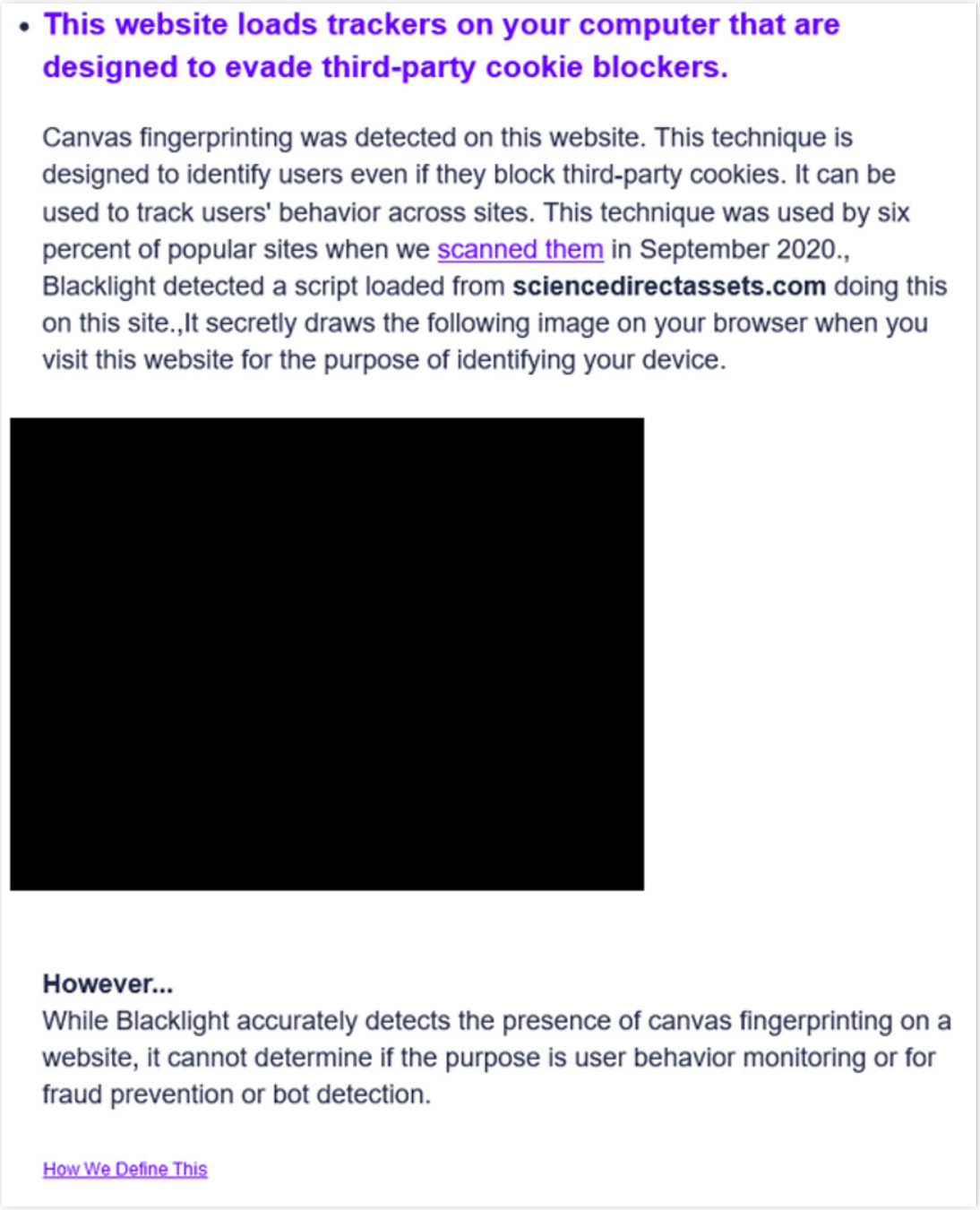
Screenshot of the first part of the Web Beacon list from the 2022 WEC report of the search results page. The third entry—smetrics.elsevier.com—is transmitting digital fingerprint data through the "bh" and "bw" tags (browser height and width, respectively).

FIGURE C-6.



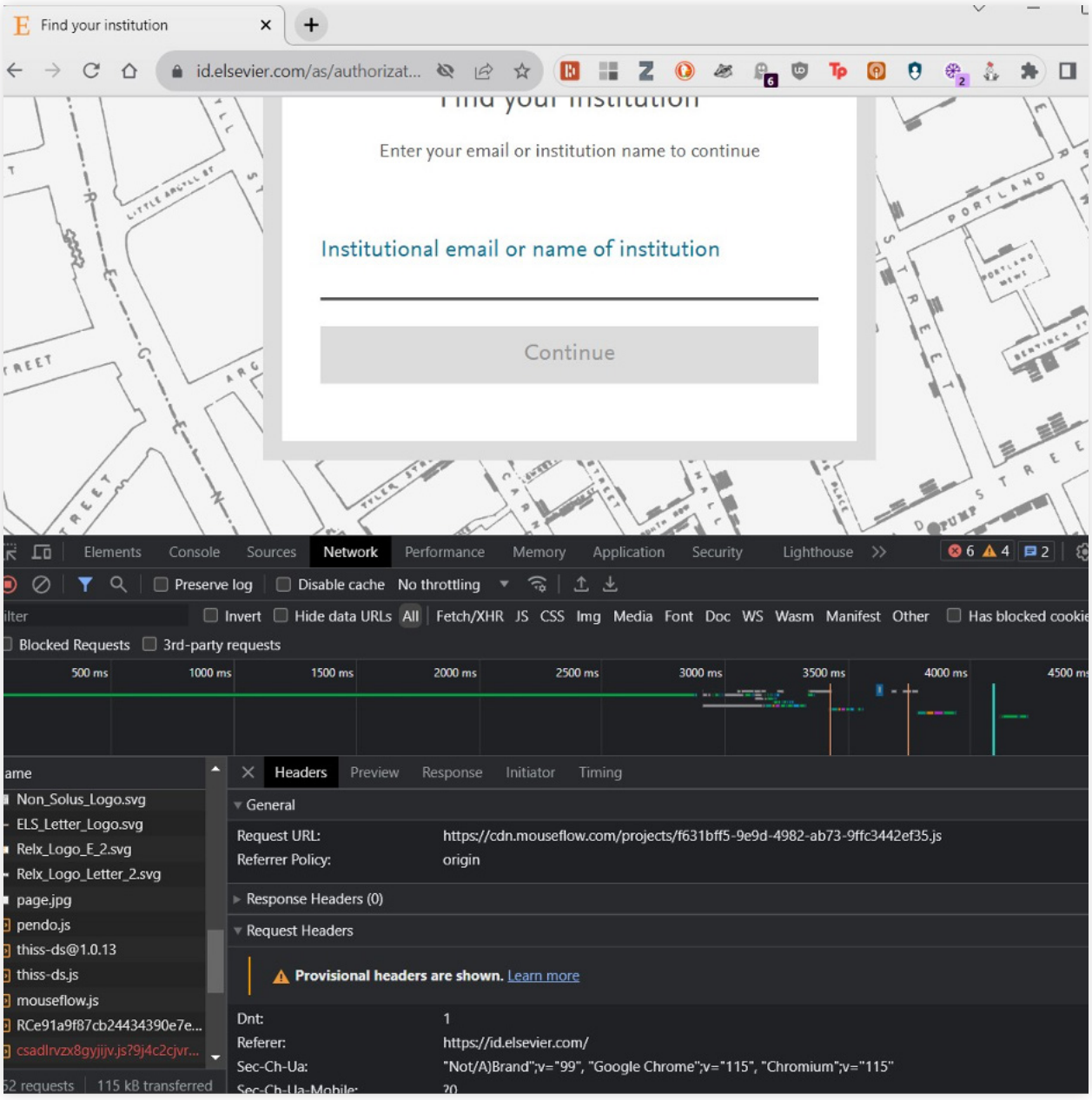
Screenshot of the network log in Chrome's web developer tools listing the different types of files being called by a ScienceDirect article page. The log entry for a JavaScript file being called by the site to load a web beacon is highlighted.

FIGURE C-7.



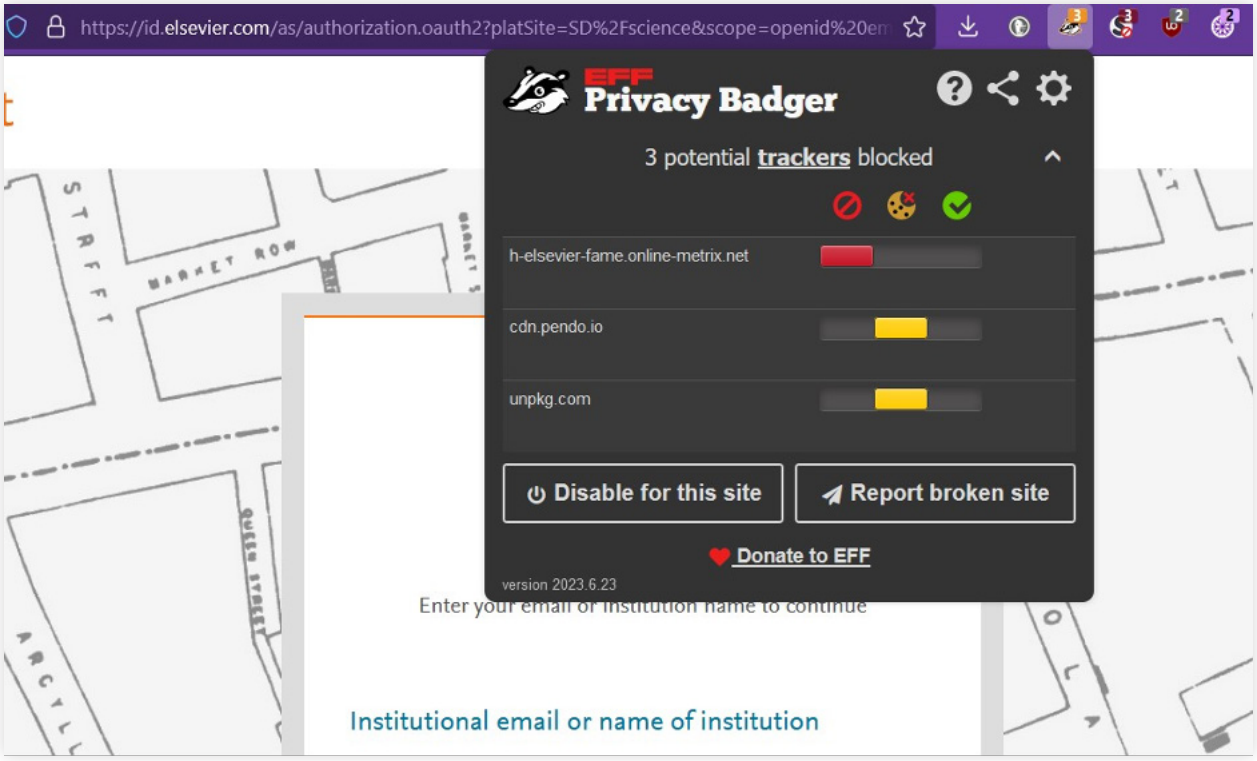
Screenshot of Blacklight's report on ScienceDirect's website. The screenshot shows that the site uses canvas fingerprinting, along with a redacted fingerprint drawn by the site to identify the user device.

FIGURE C-8.



Screenshot of the network log in Chrome's web developer tools showing the header for the Mouseflow script on the id.elsevier.com page.

FIGURE C-9.



Screenshot of EFF's Privacy Badger browser extension listing h-elsevier.fame.online-metrix.net as a tracker on the id.elsevier.com page.

APPENDIX D
— *Web Tracker List and Counts*

Cookie and Web Beacon Host List

The following is a list of the unique domains serving cookies and web beacons found in the 2022 and 2023 website test WEC and Blacklight reports.

1. acw.elsevier.com
2. acw.sciencedirect.com
3. ajax.googleapis.com
4. api.plu.mx
5. ars.els-cdn.com
6. assets.adobedtm.com
7. bam.nr-data.net
8. bam-cell.nr-data.net
9. brxt-research.mendeley.com
10. cdn.mouseflow.com
11. cdn.pendo.io
12. cdn.plu.mx
13. cdnjs.cloudflare.com
14. cm.everesttech.net
15. content.id.elsevier.com
16. data.pendo.io
17. dpm.demdex.net
18. elsevier.demdex.net
19. elsevierlimited.tt.omtrdc.net
20. id.elsevier.com
21. js-agent.newrelic.com
22. pagead2.googlesyndication.com
23. pendo-static-5661679399600128.storage.googleapis.com
24. prod-id-assets.elsevier-ae.com
25. relx-elsevier-erms--c.documentforce.com
26. scholar.google.com
27. sdfestaticassets-us-east-1.sciencedirectassets.com

28. securepubads.g.doubleclick.net
29. service.seamlessaccess.org
30. smetrics.elsevier.com
31. static.cloudflareinsights.com
32. static.mendeley.com
33. unpkg.com
34. www.googletagservices.com
35. www.sciencedirect.com

Number of Trackers by Page

2022 WEB RESULTS

PAGE	1ST AND 3RD PARTY COOKIES	LOCAL STORAGE	WEB BEACONS	TOTAL # TRACKERS
Home page	20	4	14	**38**
Browse page	26	4	16	**46**
Search results page	27	6	18	**51**
Article page	29	7	20	**56**

Number of Trackers by Page

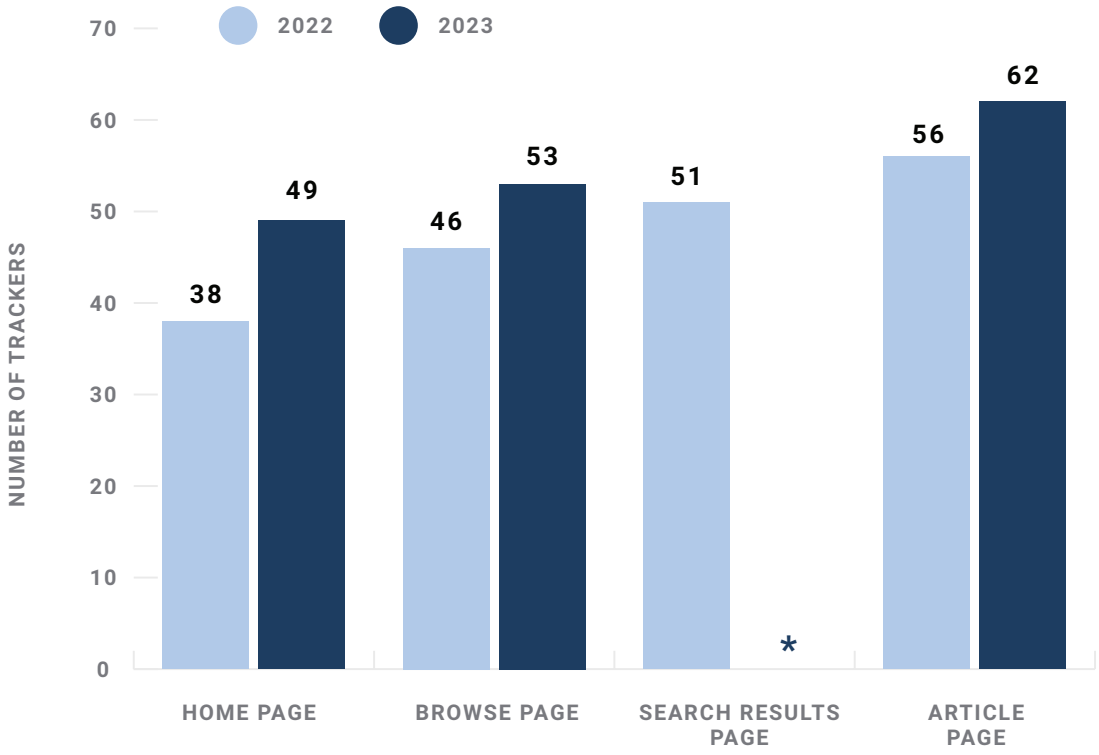
2023 WEB RESULTS

PAGE	1ST AND 3RD PARTY COOKIES	LOCAL STORAGE	WEB BEACONS	TOTAL # TRACKERS
Home page	22	4	23	**49**
Browse page	23	4	26	**53**
Search results page	*	*	*	*
Article page	27	7	28	**62**

**WEC was unable to capture the number of trackers for the search results page numbers in 2023 due to the search results page automatically redirecting to the login page during the test.*

Year-by-Year Comparison of Total Trackers by Page

(WEC RESULTS)



**2023 results not available*

November 2023