Assessing Data Privacy on a Vendor Website

As of January 2025, Becky Yoose of LDH Consulting Services and Nick Shockey of SPARC have authored two vendor reports for SPARC's Navigating Risk in Vendor Data Privacy Practices Series - An Analysis of Elsevier's ScienceDirect and An Analysis of Springer Nature's SpringerLink. In both of these documents, the authors walk through how to use tools such as browser plugins (Ghostery, uBlock Origin and NoScript) to view the list of cookies and executable content on a web page. Additionally, Website Evidence Collector (WEC) and Blacklight are used to understand tracking behavior across multiple pages. (Yoose & Shockey, 2023, p. 19).

Mastering this analysis empowers you to gather crucial data directly from vendor websites and leverage it during negotiations. This guide equips you with the knowledge and tools to conduct your assessment effectively.

Level 1: Browser Plugins - Your Privacy Quick-Check

Think of browser plugins like handy detective tools that reveal hidden website activity. Here are a few user-friendly options:

- Ghostery: This plugin acts like an X-ray for websites, showing you all the trackers present. It helps you block trackers you don't trust.
- uBlock Origin: This powerful plugin blocks ads and trackers, making your browsing experience faster and more private.
- NoScript (Use with Caution): This plugin blocks all JavaScript by default, which can break some website functionality. Use it cautiously and allow JavaScript only on trusted sites.

How to Use Browser Plugins:

1. Installation: Search for the plugin in your browser's extension store (e.g., Chrome Web Store, Firefox Add-ons) and click "Add to [Your Browser]."
2. Activation: Once installed, the plugin's icon will appear in your browser toolbar. Click on it to see the trackers and options.

Level 2: Browser Developer Tools - Going Deeper

For a more detailed look at what's happening behind the scenes, your browser has built-in developer tools:

1. Access Developer Tools: Right-click anywhere on a webpage and select "Inspect" or "Inspect Element."
2. Navigate to Network Tab: In the developer tools panel, click on the "Network" tab.
3. Reload the Page: Refresh the webpage to see the network requests being made.
4. Analyze Network Traffic: Look for requests to unfamiliar domains or those related to tracking and data collection.

Level 3: Advanced Tools - The Big Picture

For a comprehensive analysis of website tracking, consider these advanced tools:

- Website Evidence Collector (WEC): This tool captures a snapshot of a website and analyzes it for various trackers.
- Blacklight: This tool provides a detailed report of the trackers and data collection practices used by a website.

How to Use Advanced Tools:

1. Follow the instructions on how to download, install and run WEC on the website.
2. For Blacklight, just enter a website URL on their page and let the scan run. You'll be able to review the report once complete.

Employing the Tools

Let's analyze a vendor together as an example and you can then try it on your own!

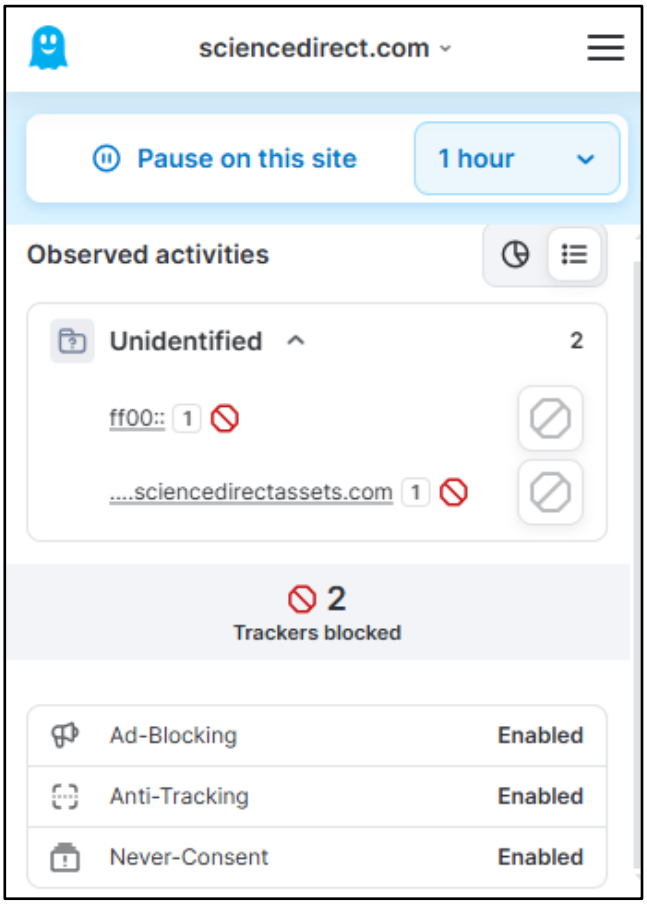
Pick your vendor and start out on their resource home page. This document will use Elsevier's ScienceDirect as an example in order to demonstrate the work Becky Yoose did in completing the vendor report - An Analysis of Elsevier's ScienceDirect - for SPARC.

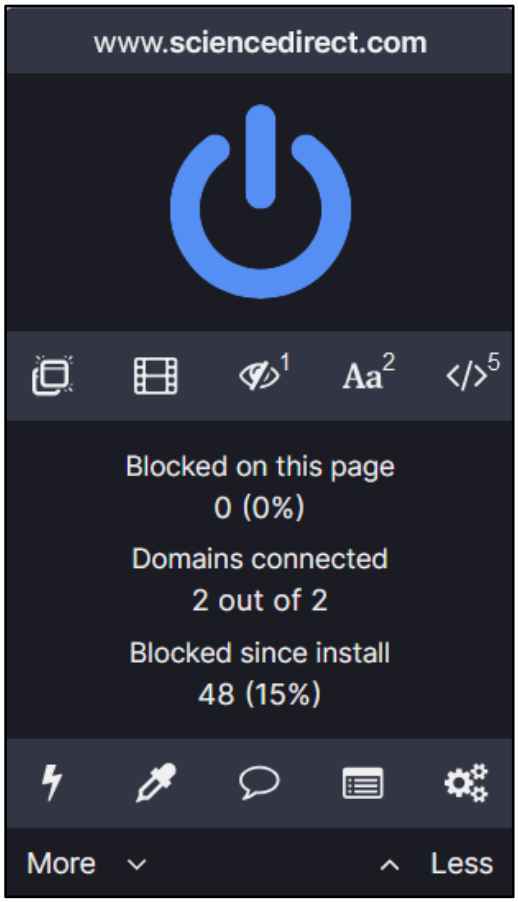
Resource home page: https://www.sciencedirect.com/

← Screenshot of the three browser plugins enabled at the top of my browser window. In this case I am using Google Chrome.

When creating this example in January 2025, we found two issues on Ghostery and NoScript. Clicking on the Ghostery icon shows me that there are two unidentified trackers in play (left image). uBlock Origin is

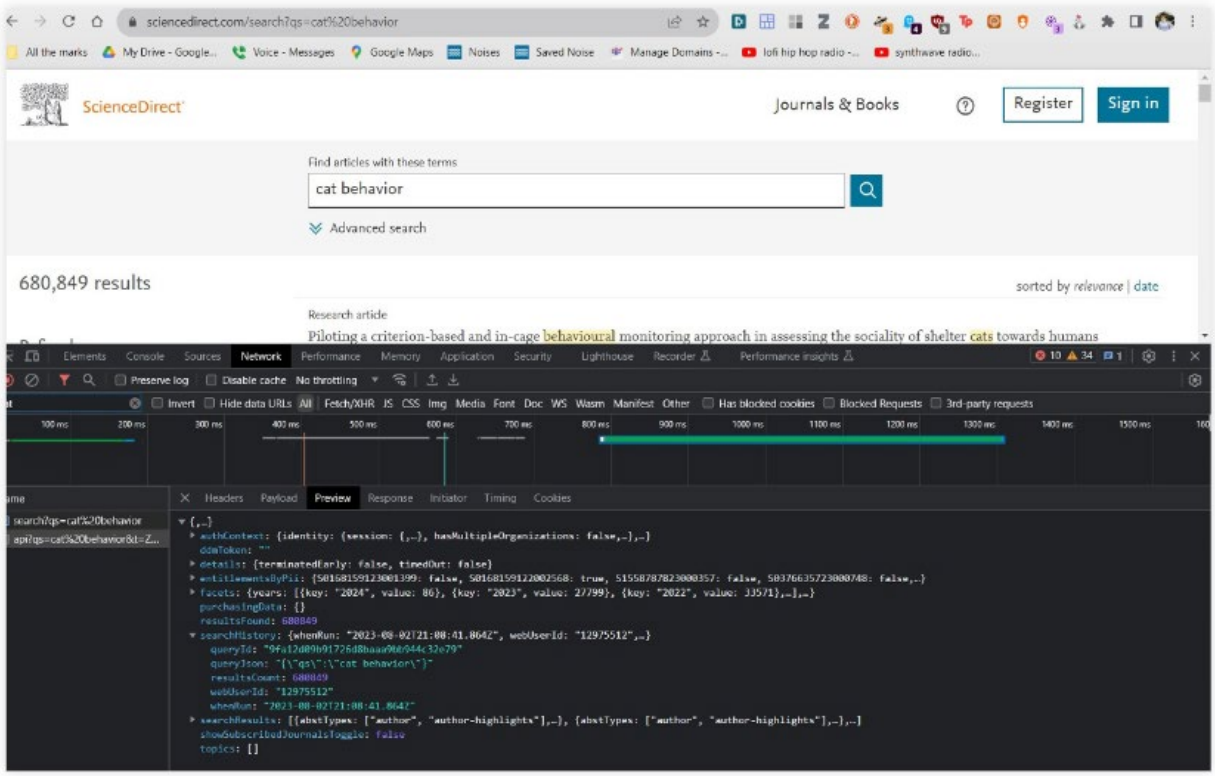
not blocking anything if I click on its icon (right image) and NoScript (bottom image) notifies me about two trackers - one of which is the same as one of Ghostery's (sciencedirectassets.com).





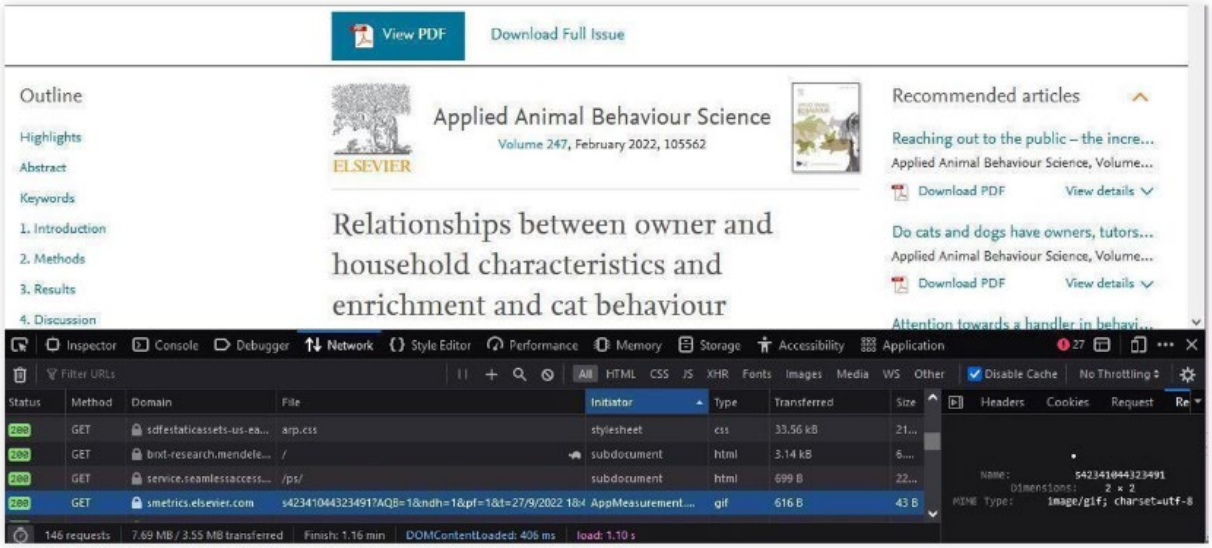
In the report An Analysis of Elsevier's ScienceDirect, Yoose and Shockey include screenshots from web testing in Appendix C of the document starting on page 45. Figures C-4, C-6, and C-8 (pp. 49, 50, and 52) capture screenshots using the built-in Chrome Web Developer Tools to inspect network logs that contain an API tracking search history, JavaScript loading a web beacon, and a third party tool called Mouseflow which can be used to track user behavior.

FIGURE C-4.



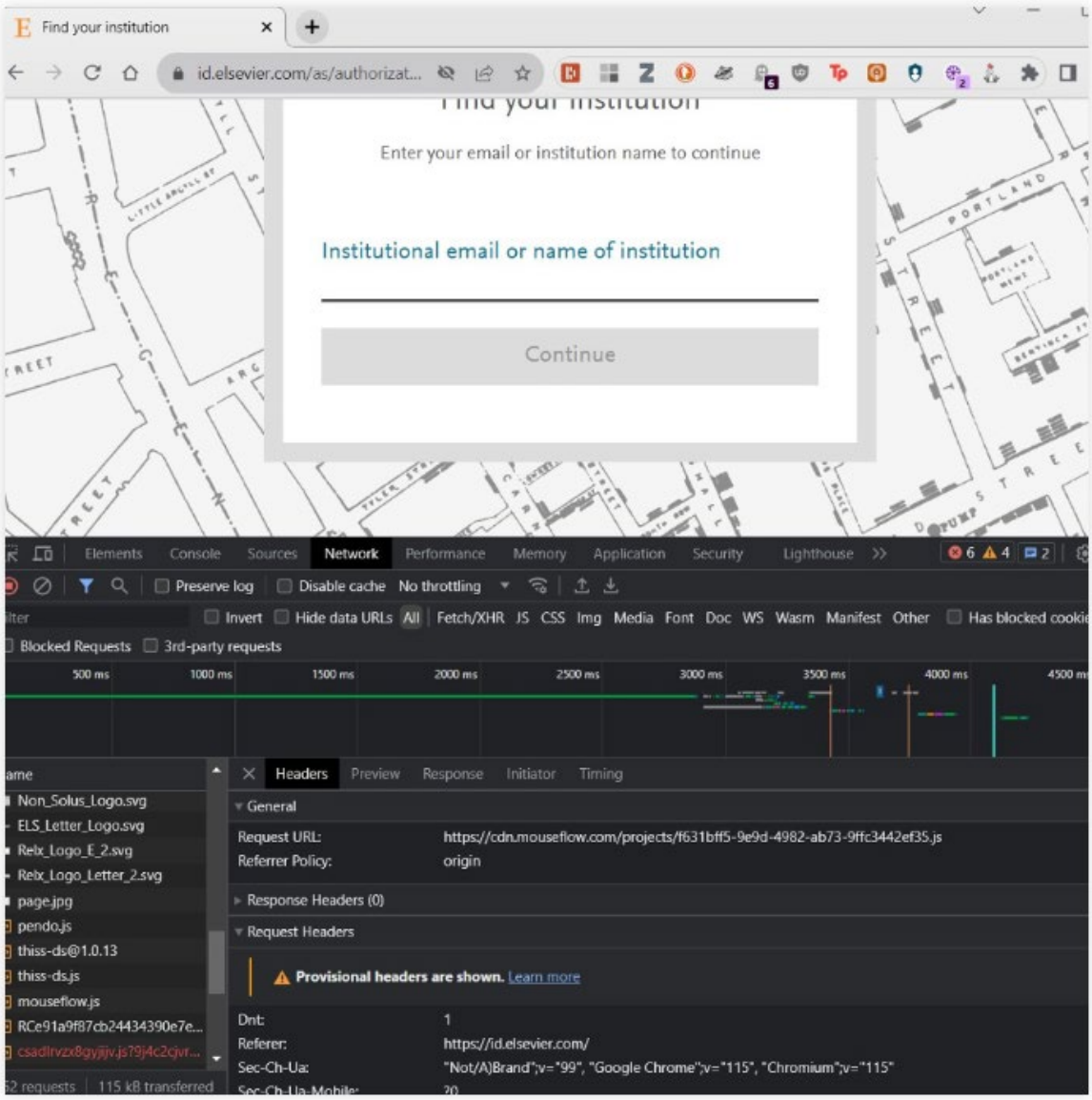
A screenshot of the search results page in 2023 using Chrome's web developer tools that shows a call to an API script that contains information about search history under a searchHistory parameter.

FIGURE C-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-8.



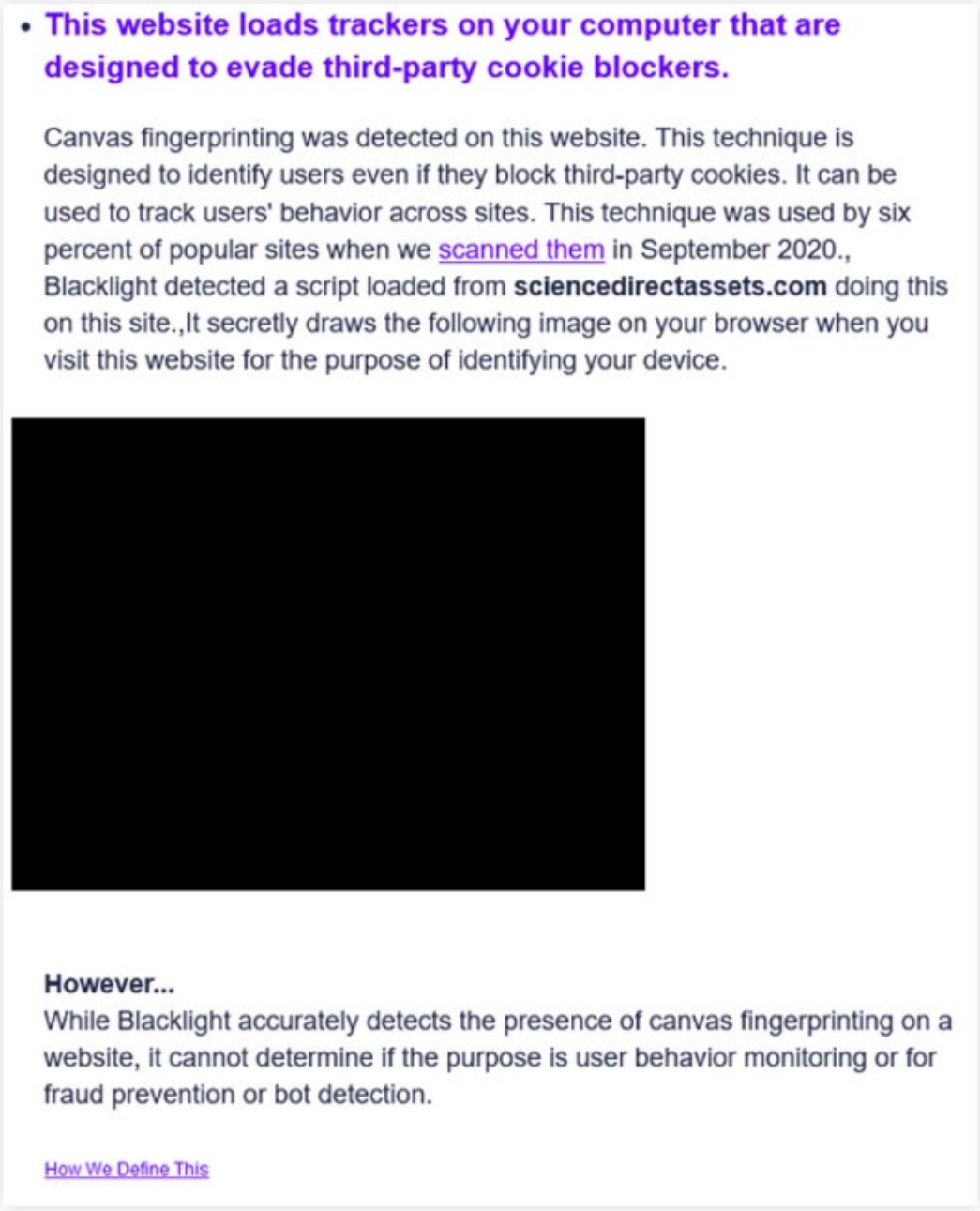
Screenshot of the network log in Chrome's web developer tools showing the header for the Mouseflow script on the id.elsevier.com page.

C-4, C-6, and C-8 Screenshots from An Analysis of Elsevier's ScienceDirect (CC-BY-4.0).

Figure C-7 (p. 51) from the report highlights the results from Blacklight which indicate that the webpage loaded scripts for canvas fingerprinting from sciencedirectassets.com (see Ghostery and NoScript screenshots above!) which can be used to track user behavior on and across websites. Blacklight does

provide the caveat that they cannot tell if ScienceDirect is using the script for tracking user behavior monitoring, fraud prevention, or bot detection.

FIGURE C-7.

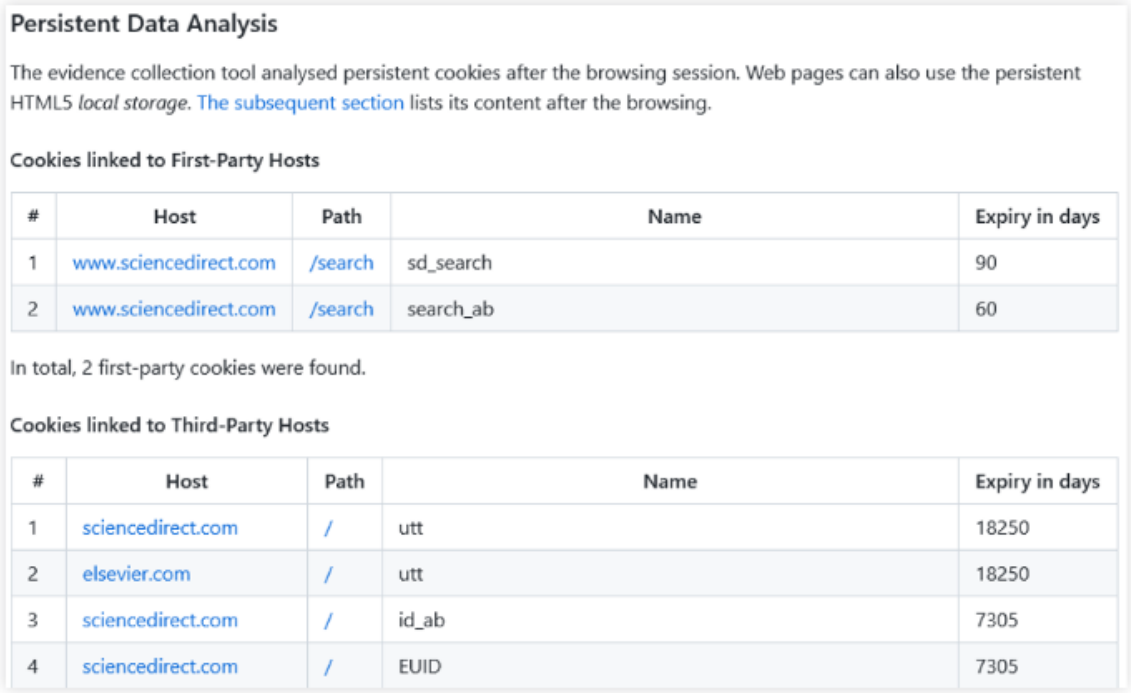


Screenshot of Blacklight's report on ScienceDirect's website. The screenshot shows that the site uses canvas fingerprinting, along with a redacted fingerprint drawn by the site to identify the user device.

C-7 Screenshot from An Analysis of Elsevier's ScienceDirect (CC-BY-4.0).

Finally, Website Evidence Collector (WEC) is the most advanced and comprehensive of the tool arsenal used in SPARC's vendor reports. More information on how to install and run it is available from their website and GitHub repository page. In the ScienceDirect report, screenshots from the WEC report run are included again in Appendix C as Figures C-1, C-3, and C-5 (pp. 45, 46, and 48).

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

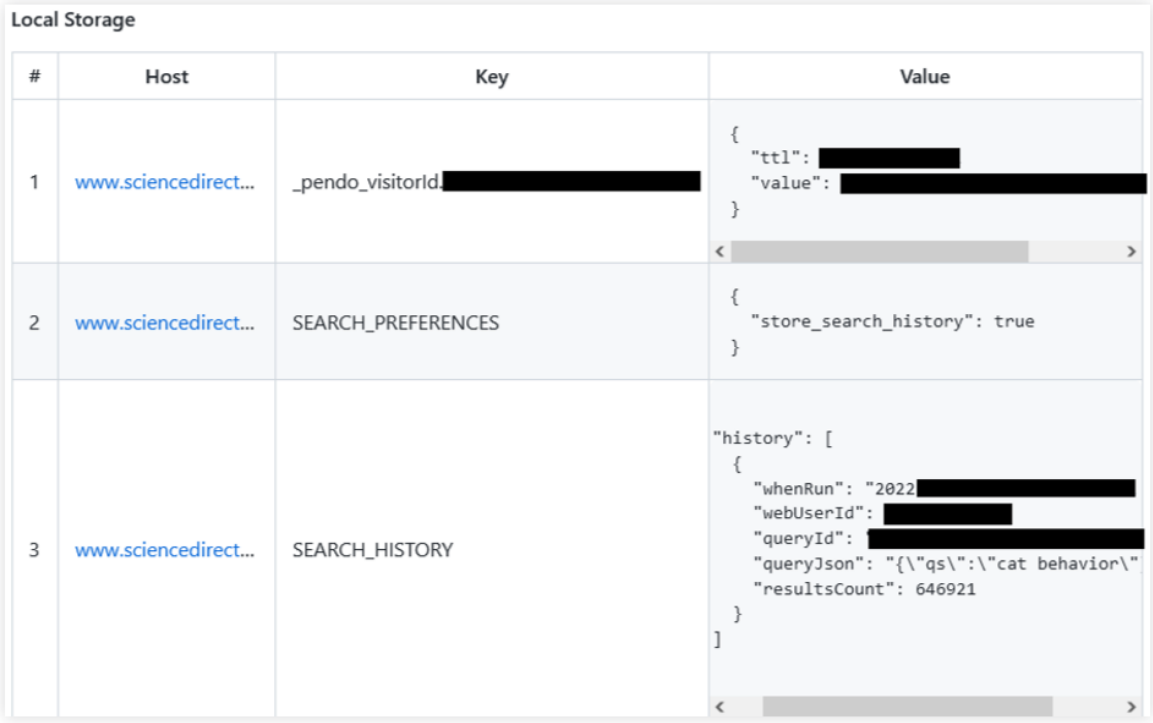
In total, 2 first-party cookies were found.

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-3.



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-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).

C-1, C-3, and C-5 Screenshots from An Analysis of Elsevier's ScienceDirect (CC-BY-4.0).

Additional information on the WEC report findings, plus the other tools used, can be found in the report section titled "Investigating Data Collection and Tracking on the ScienceDirect Website" on pages 18-21.

Overall, the assessment of data privacy via use of these tools can assist in identifying concerns and areas to limit invasive data collection practices via contractual language. For example, the ScienceDirect report used the tools and reported that the website employs various tracking technologies, including persistent

cookies, HTML5 local storage, and web beacons, some with expiration dates up to 50 years. These trackers collect personal data, such as IP addresses and device information, and redirect to third-party domains for advertising and analytics. Even without logging in, users are tracked, and blocking JavaScript to prevent tracking disrupts website functionality. The use of canvas fingerprinting and potential tracking on the login page raise further privacy concerns, highlighting the need for transparency and user awareness regarding data collection practices on ScienceDirect, (pp. 18-21). This type of information and analysis can assist in vendor negotiations by pointing out inherent risks inherent to the library's users and raising concerns both with the vendor and among your campus community.

Go ahead and try out the tools yourself and see what you discover! If you're not sure where to start, try checking out SpringerNature's SpringerLink and compare your findings with the SPARC report, An Analysis of Springer Nature's SpringerLink.