CSI-COP

Citizen Scientists Investigating Cookies and App GDPR Compliance

----

Deliverable D5.2 | D19

Web-based Knowledge Resource: Repository of CSI-COP website and app investigations

Delivery date: 31 May 2023

LEAD Partner for deliverable: CU

Deliverable Authors: Shah, H., Nanneboina, V. and Winter, J.

Version: 1.2

Version control

Table of Contents

Executive Summary

CSI-COP's research and innovation (RIA) proposal promised a prototype at technology readiness level 5 (TRL5) of a repository from citizen scientists' and project team's GDPR compliance investigations in websites and apps. What the project has produced is at least **TRL7**: a free, searchable, working model of a web-based, open-access knowledge resource (**CSI-COP repository**) of digital trackers uncovered during the project's general public engagement. The repository accompanies the 'Taxonomy of Digital Cookies and Online Trackers' public report (deliverable D5.1|D18: Shah et al., 2023). This document presents an explanation on what is contained in CSI-COP's **repository** and how to use it. The repository acts as a tool that would be useful for anyone looking to see if a website they visit or an app they use has third-party trackers, or permissions to access personal data by default. The repository can be utilised in a number of ways, for example, it can serve as a training tool for junior web and app developers in their transparency efforts making it clear to users if any third-party access to personal data is extracted during engagement. We welcome feedback on the repository to continue improvements in the usability and usefulness of retrieved information to the end of the extended project: 31 August 2023.

Keywords: apps, cookies, data protection, digital trackers, GDPR, informed consent, online trackers, permissions, personal data, privacy, repository, third-party cookies, third-party requests, transparency.

Introduction

This document acts as a manual to explain what is in, and how to use CSI-COP's working model of a free, web-based, open-access searchable knowledge resource of digital trackers (CSI-COP repository). The repository (Figure 1) contains the results of website and app investigations conducted by the project's engaged citizen scientists and partners' team members. The concept for the repository in CSI-COP proposal phase aimed at a demonstrable prototype, technology readiness level 5 (TRL5). What CSI-COP have produced from a technical sub-contract by Xcel Resources Ltd.'s **Mr Venkatesh Nanneboina** is TRL7: a usable online tool to search and find what trackers were embedded in websites and apps at the time of investigations made by the general public who joined the project as citizen scientists, and the project partners.

Figure 1: Screenshot of CSI-COP Repository accessible on project website: https://csi-cop.eu/repository/

First page in the Repository

The first iteration of the repository was made available to CSI-COP project partners on Wednesday 17 May 2023. Feedback was sought by Friday 19 May 2023 allowing Coventry University's technical sub-contractor for work package 5 (WP5) to make improvements. Venkatesh Nanneboina's (VN) sub-contracted work in WP5 for the 'Open-access Knowledge Resource of Digital Trackers' involved taking on the task of producing the repository in a short time-space due to the extension of work package 4 tasks to December 2022-end. Despite the short time from January 2023 to repository launch deadline, 31 May 2023, through continual meetings between VN and Coventry University's Dr. Huma Shah (HS), CSI-COP science director, VN's design from requirements, and innovation with continual improvements in a test repository, made it possible for partners, citizen scientists and different stakeholders to explore the repository. Feedback was used to produce the 31 May 2023 version of the repository tool. Xcel Resources Ltd. will produce a technical white paper on the repository development. Next, we present the stages from citizen scientists' investigations to the repository.

Repository Concept

CSI-COP work package 5 (WP5) focused on developing an open-access knowledge resource of digital trackers. Coordinating partner, Coventry University secured a sub-contractor for the development of the repository. Following tenders, Xcel Resources Ltd. were awarded the sub-contract. Xcel Resources Ltd. onboarded software engineer Mr Venkatesh Nanneboina, a specialist in data analytics, to develop CSI-COP's repository. Mr Nanneboina attended meetings with the Coordinating partner's lead scientist, Dr. Huma Shah to understand the citizen science element, and learn what the data constituted that needed to be retrieved from the repository. Initially, in CSI-COP proposal planning phase, the idea for what information would be stored and retrieved from the repository included anonymised citizen scientists' data (Box 1).

Table 1: CSI-COP planning phase repository data-store

1) By anonymous human citizen scientist profile:
   a. Age ranges: young adults; mature adults
   b. Male
   c. Female
   d. Inter
   e. LGBTQ
2) By Tracker feature:
   a) geographical – map/open source map,
   b) websites by type with cookie trackers: news information; weather; healthcare; travel, etc.
   c) cookie trackers by type: profiling, marketing, web analytics
   d) app trackers by type: profiling, marketing, social media analytics
   e) apps designed specifically for females
   f) apps designed specifically for males
   g) apps designed for LGBTQ
   h) apps designed specifically for children
   i) apps designed specifically for teenagers
   j) apps designed specifically for young adults
   k) apps designed specifically for mature adults

However, CSI-COP work package 4 contained a specific task presenting anonymous data on the project's citizen scientists - Task 4.3: age, socio-economic and geographical distribution – AGSEG. The public report on 'Who CSI-COP citizen scientists are?' (Rigler et al., 2023: deliverable D4.3|D17) has already been produced, submitted to the EU and is available from the project website results page: https://csi-cop.eu/project-results/who-are-citizen-scientists/

Hence the design of the repository focused on CSI-COP's citizen scientists and project partners' findings relating to **compliance of the general data protection regulation** (GDPR) with respect to **transparency** in website cookie banners /notices and privacy policies about collecting **personal data** with **informed consent**. In the next section, we explain how the repository was designed to show any third-party cookies, third-party requests in the investigated websites, and default permissions in apps uncovered by the project's citizen scientists and partners.

Design and development

Existing online, free searchable databases, or repositories, were explored before Xcel Resources Ltd. and VN joined CSI-COP as sub-contractor. Three online searchable databases were explored further to produce the design of CSI-COP's repository and to make searching it easy. Those three are:

1. 'Offshore leaks Database' or 'Panama Papers' from the International Consortium of Investigative Journalists (ICIJ), accessible from here: https://offshoreleaks.icij.org/
2. 'Legacies of British Slavery' from UCL: https://www.ucl.ac.uk/lbs/search/
3. 'Westminster Accounts', database of donations and financial interests of UK politicians from Tortoise Media and Sky News: https://www.tortoisemedia.com/westminster-accounts-explore/

To understand what the data would be in CSI-COP's repository, thereby designing its intelligent retrieval while making it easy to use for anyone, VN explored CSI-COP's two citizen scientists' databases created from WP4 by Coventry University's Jaimz Winter (JW). Citizen scientists' website investigations, and app investigations are the two WP4 databases that sit as open datasets in CSI-COP website project results page:

1. Websites investigated: https://csi-cop.eu/project-results/citizen-scientists-website-investigations/
2. Apps investigated: https://csi-cop.eu/project-results/citizen-scientists-app-investigations/

The data from the citizen scientists' investigations were used as the initial 'test data' to build the infrastructure of CSI-COP's repository. The data points extracted from the investigations to fold into the repository are:

1. Meta data on the websites investigated (see Figure 2) by CSI-COP's citizen scientists and the project's team members, include:

    a.   Website name
    b.   Website type (news, etc.)
    c.   Date website investigated (in case of changes by website owners at a later date)
    d.   URL of website
    e.   Brief information on any cookie notice, and privacy policy
    f.   Any free privacy audit tool to explore beneath the website
    g.   Any third-party cookies or third-party requests embedded in the website

Figure 2: Screenshot of website database from task T4.1

2. Meta data on the apps investigated (see Figure 3) by CSI-COP's citizen scientists and the project's team members, included:
   a. App name
   b. Operating system: Android or Apple iOS
   c. App type
   d. Date the app was investigated
   e. Language of app

f. App permissions

Figure 3: Screenshot of Apps database for from task T4.2

VN designed investigations to be presented as cards: a card for each website or app investigation. A card would display some information, such as name of website or app, and the date it was investigated (see table 1). Further information would be furnished on selecting a card. This is explained in the next section.

Table 2: Test repository card display of app, and website investigations

App investigations: card display in test repository	Website investigations: card display in test repository

Search Requirements

Once the infrastructure was built around the citizen scientists' website and app investigations, then VN was able to add the CSI-COP team member investigations into the test repository. The requirements for the repository related to information about investigations to be retrieved on search. The search would be related to what trackers were found in websites, and what permissions were found in apps. This was dependant on 'clean data'. It was essential that the information to be retrieved, such as images of websites, etc., taken during investigations, had to match the website or app. This is where we found that

more time was necessary to ensure the labelling of images matched precisely with the investigation. At launch of repository on 31 May 2023, this work was still ongoing for some investigations. However, and especially to make a version of the repository ready for its demonstration to a variety of stakeholders attending CSI-COP's main project results' event in Brussels: 23-26 May 2023, another test repository site was created by VN. Partners' feedback had produced some further developments, and search improvements in a second iteration of the repository. This second iteration, in VN's repository developmental phase, remains viewable on a test website until 9 June 2023 at this web link: https://csicoprepo.azurewebsites.net/cookies/# The purpose of leaving the test site open for a further two weeks is for citizen science projects aiming to build a repository from their investigations to gain insight into CSI-COP repository's innovation. Examples of information extracted from test search in CSI-COP repository's second iteration are shown in Tables 3-5.

Table 3: Test repository website cards

Test site website home page	Test site website card displays

Table 4: Test site app cards

Test site app search bar	Test site app card displays

Table 5: Test site information on selecting card

Test site: app image on selecting ard	Test site: website tracker information on selecting card

![CSI-COP logo — Citizen Scientists Investigating Cookies & App GDPR Compliance]

Feedback on second iteration

After demonstrating the repository across two days during the project's main event in Brussels 23-24 May, the test site was made accessible to the project's citizen scientists and online and in-person attending stakeholders in Brussels. Anonymous feedback was collected from a page on the test site between 24 and 28 May 2023 to further iterate and improve the repository before official launch on 31 May 2023. The deadline of end-of-day 28 May 2023 allowed VN to make further improvements based on the feedback and to increase usefulness of CSI-COP's repository by 31 May 2023.

Appendix 1 presents anonymous qualitative feedback, which included recommendations to correct typos. Suggestions for "worthy additions", for example, "Have the ability to rerun an analysis on today's date to see if anything changed" are not possible within CSI-COP person-month effort, time-frame, and sub-contract agreement. It would be a useful exercise post-project with new funding to find if the websites and apps investigated improved their GDPR compliance with respect to transparency and informed consent on collecting personal data.

Overall, the feedback on the second iteration was positive with more people finding the second iteration of the repository usable, the information provided having clarity and being useful (Chart 1).

Chart 1: Anonymous feedback on second iteration of CSI-COP repository
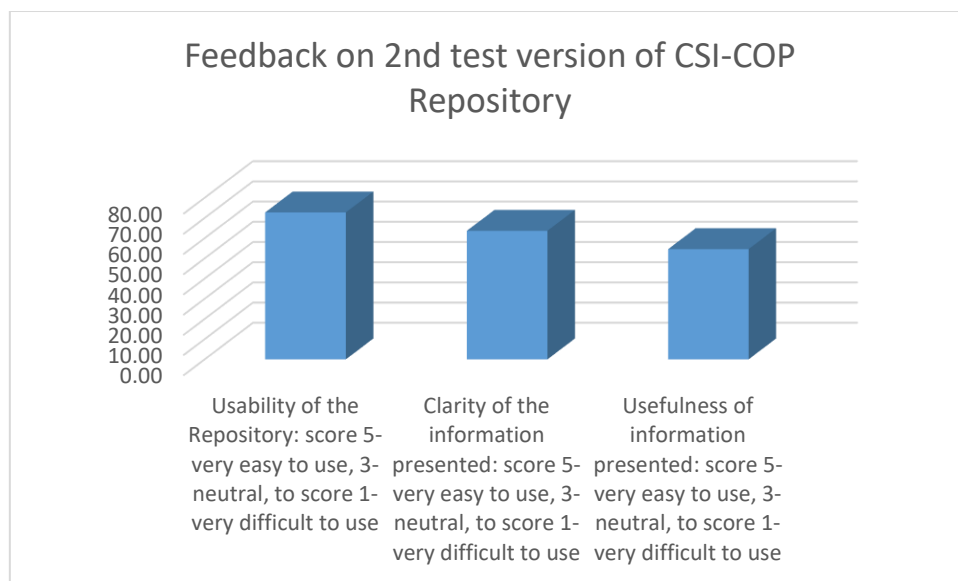
One other particular investigation that CSI-COP included in the repository's second iteration was EU funded project websites. Over one hundred investigations of EU funded project websites were conducted by Coventry University's junior research assistant, Giacomo Masone. The work was evaluated for accuracy by Coventry University's senior research assistant, Jaimz Winter. VN then added the EU funded project website investigations into the Repository. Together with all the investigations, CSI-COP will recommend standardisation of cookie notices and privacy policies in the final project policy brief.

Future Improvements

CSI-COP repository launched on 31 May 2023 has taken into consideration the invaluable feedback from CSI-COP partners, citizen scientists and anonymous stakeholders received by 28 May 2023. Improvements were made in the time available from 28 May 2023 to the repository deliverable due date, 31 May 2023. This was to ensure issues in search were ironed out. We now invite readers of this document, citizen science project teams, website and app developers, GDPR researchers and others to explore CSI-COP's repository of website and app investigations from the project's website at this link: https://csi-cop.eu/repository/

Users might find that some data issues remain. CSI-COP sub-contractor Xcel Resources Ltd.'s Venkatesh Nanneboina, and the Coventry University team as task lead for the repository, are working on it constantly. The aim is, following further data-cleaning, by the end of the project in August 2023, CSI-COP repository's data cards will show more information extracted from investigations. Nonetheless, CSI-COP's repository tool is beyond a TRL5 prototype. Venkatesh Nanneboina of Xcel Resources Ltd. development of the repository has innovated it to at least TRL7: a usable tool for a wide range of purposes, including as a training tool for web and app developers, and an educational open, web-based dataset for data protection and technology ethics undergraduate and postgraduate courses.

References

Rigler, D., Hinsenkamp, M. Shah, H. and Winter, J. (2023). *Age, Gender, Socio-economic and Geographical Distribution Report* (anonymised information on CSI-COP citizen scientists) . CSI-COP EU Horizon2020 SwafS 15-2019 funded project (GA 873169). Accessible from CSI-COP project results page here: https://csi-cop.eu/project-results/who-are-citizen-scientists/

Shah, H., Gialelis, Y., Konstantinopoulos, I., Lantavou, K., Karakonstanti, A., Rigler, D., Stepankova, O., Ignat, A. and Celentano, U. (2023). *Taxonomy of Digital Cookies and Online Trackers*. CSI-COP EU Horizon2020 SwafS 15-2019 funded project (GA 873169). Accessible from CSI-COP project results page here: https://csi-cop.eu/project-results/taxonomy-of-cookies-and-online-trackers/

Further project results can be found on CSI-COP website project results page here: https://csi-cop.eu/projectresults/

CSI-COP
Citizen Scientists
Investigating Cookies
& App GDPR Compliance

Appendix 1: Anonymous feedback

Anonymous feedback received via a web page on the second iteration of the test repository between 24-28 May 2023.

Anonymous feedback on the repository received between 24 - 28 May 2023					
Date of feedback	Language of Repository investigations for feedback	Usability of the Repository: score 5- very easy to use, 3- neutral, to score 1- very difficult to use	Clarity of the information presented: score 5- very easy to use, 3- neutral, to score 1- very difficult to use	Usefulness of information presented: score 5- very easy to use, 3- neutral, to score 1- very difficult to use	Overall Feedback
24.05.23	English	5	5	4	None provided
24.05.23	Hungarian	5	5	5	None provided
24.05.23	English	5	5	5	Very good in that it makes visible the hidden cookies and trackers to a wider range of potential users.
24.05.23	Hungarian	5	4	4	None provided
24.05.23	Hungarian	5	5	5	None provided
25.05.23	English	4	3	3	**Minor suggestions:**
					- When you change from App to Website (and vice versa) clear the results shown.
					- When you open a card if there is no image available for that entry then do not have the clickable image tab.
					- Under info include the URL of the website so it can be easily copied and pasted.
					- Entry Shein is English and Hebrew: the first letters of these words should be capital.
					Worthy additions:
					- Have an ability to export the data produced to a spreadsheet.
					- Have the ability for citizen scientists to add new entries.
					- Have the ability to rerun an analysis on today's date to see if anything changed
25.05.23	English	4	4	4	Mispelling of "government" under categories drop down Governament VS Government
26.06.23	Romanian	5	5	5	None provided
26.05.23	Greek	3	4	2	Good idea, but not clear visualization
27.05.23	Hungarian	5	5	5	I really like it but a few mistakes were left in the beta version (mainly typos, in some cases the main info such as title of the webpage missing and some image files)
					One improvement I could suggest is to make the images in the detailed info page bigger and correct the typos.
28.05.23	English	5	5	5	None provided