What is in Your Cookie Box? Explaining Ingredients of Web Cookies with Knowledge Graphs

Geni Bushati [a,*], Sven Carsten Rasmusen [a], Anelia Kurteva [a], Anurag Vats [b], Petraq Nako [c] and Anna Fensel [a,d,e]

[a] *Semantic Technology Institute (STI) Innsbruck, Department of Computer Science, University of Innsbruck, Innsbruck, Austria*
E-mails: geni.bushati@sti2.at, sven.rasmusen@sti2.at, anelia.kurteva@sti2.at
[b] *Distributed and Parallel Systems Group (DPS), Department of Computer Science, University of Innsbruck, Innsbruck, Austria*
E-mail: anurag.vats@uibk.ac.at
[c] *University of Innsbruck, Innsbruck, Austria*
E-mail: petraq.nako@student.uibk.ac.at
[d] *Wageningen Data Competence Center, Wageningen University & Research, Wageningen, The Netherlands*
[e] *Consumption and Healthy Lifestyles Chair Group, Wageningen University & Research, Wageningen, The Netherlands*
E-mail: anna.fensel@wur.nl

Abstract. The General Data Protection Regulation (GDPR) has imposed strict requirements for data sharing, one of which is informed consent. A common way to request consent online is via cookies. However, commonly, users accept online cookies unaware of the meaning of the given consent and the following implications. Once consent is given, the cookie "disappears", and one forgets that consent was given in the first place. Retrieving cookies and consent logs becomes challenging, as most information is stored in the specific internet browser's logs. To make users aware of the data sharing implied by cookie consent and to support transparency and traceability within systems, we present a knowledge graph (KG) based tool for personalised cookie consent information visualisation. The KG is based on the OntoCookie ontology, which models cookies in a machine-readable format and supports data interpretability across domains. Evaluation results confirm that the users' comprehension of the data shared through cookies is vague and insufficient. Furthermore, our work has resulted in an increase of 47.5% in the users' willingness to be cautious when viewing cookie banners before giving consent. These and other evaluation results confirm that our cookie data visualisation tool helps increase users' awareness of cookies and data sharing.

Keywords: Cookies, Consent, GDPR, Knowledge Graph, Comprehension

1. Introduction

Cookies have emerged as one of the most convenient ways to request consent online and to keep track of browser-server interaction [1]. In the following years, with the rise of the number of websites and web applications, cookies

*Corresponding author. E-mail: geni.bushati@sti2.at.

have been used in such ways that they invade the users' privacy [2]. Stricter limitations have been enforced by the European Union's (EU) GDPR [3], in effect since May 2018. GDPR has highlighted the importance of consent and has set it as one of its legal basis for data processing. According to Art. 4 (11), consent must be freely given, specific, informed, unambiguous and one should be able to revoke it with the same ease it was given. However, requesting consent in an informed way does not guarantee having users who are aware of what data processing and sharing mean and the implications that follow [4]. Bechmann [5] formulates the notion of having a "blind consent" culture caused by users granting consent being unaware of the significance of the action. The requirement for informed consent prior to any data processing imposed by the GDPR, has also resulted in the so called "consent fatigue" (i.e. frequent display of cookie requests on websites) [6]. Further challenges related to consenting online, as discussed in [7] include information overload, imbalance of power when it comes to creating the consent requests and the use of dark patterns.

When it comes to cookies, once the consent is granted, the cookie is not visible to the user during their browsing activity. Retrieving cookie logs, which store specific data about the consent, data processing and the cookie's duration, can be a complex task for non-expert users. Furthermore, although the information in the cookie logs is presented in tabular format, security and privacy domain-specific terminology is used. Several solutions, in the form of browser extensions such as the Cookie Editor[1], which was utilised in this paper, help retrieve existing cookies. However, these tools focus mainly on simplifying the cookie retrieval process, leaving behind the meaning (semantics) of the stored cookies. The extensions allow users to directly export cookies instead of looking through all the browser files and settings to locate them. While simplifying this process, the cookie extensions are, in most cases, developed for specific browsers (e.g. Google Chrome [2]), which limits the audience that can use them.

The privacy concerns related to cookies stem from several factors, such as (i) the lack of consumer knowledge of what cookies are and of their functions as discussed in [8], (ii) the lack of control that users have over the data that is transmitted through cookies and (iii) the lack of feedback provided by browsers' cookie management tools [9]. These concerns highlight the need for more clarity, transparency and awareness about cookies and what happens after cookie consent is granted. This has also motivated our work, which focuses on the use case of requesting consent for data sharing through web browser cookies.

To address these challenges, by building upon the findings in [4], we present a KG-based tool for personal cookie visualisations. The main goal is to bring more transparency and awareness regarding cookies and to ease users' comprehension of cookie-based data sharing. The main research question that we focus on is: *"Can a KG-based visualisation of cookie statistics help ease one's comprehension of cookie data sharing?"*. In the context of this paper, the ease of cookie comprehension refers to the ability of users to understand what exactly a cookie is (i.e. its source, duration and type). In addition to addressing the comprehension of users, machines are also taken into consideration. Specifically, the need for information in a machine-readable format that provides machines with context, supports traceability and transparency. When it comes to consent for data sharing, multiple ontologies such as Consent and Data Management Model (CDMM) [10], Data Protection Vocabulary (DPV) [3] and GConsent [11] have been built and are widely used, as presented in [12]. However, there is a lack of semantic models for cookies. For this reason, we have built a semantic model (in Web Ontology Language (OWL)) for cookies and a KG based on it, which is the main data source for the cookie visualisation tool developed. The use of semantics, namely KGs, has a number of benefits. KGs support faster and easier knowledge discovery with the help of relationships between concepts, which can be extremely helpful in cases such as fraud detection and prediction and trace of cyber attacks (see [13–15]). Further, KGs represent information in a structured and meaningful way [16] and can be used for privacy-enabled penalisation on the web [17], intelligent decision making in the security domain as discussed in [18] and other domains such as manufacturing (e.g. [19]). In addition, KGs offer better data interpretability, transparency and traceability [20, 21] and are extendable by design, which makes them suitable for use in different ecosystems and across multiple use cases [12]. The use of semantic technology and the main trends for its application in the security and privacy domains are further discussed in [22, 23].

With this paper, we make the following contributions:

[1] https://Cookie Editor.cgagnier.ca/
[2] https://www.google.com/
[3] https://dpvcg.github.io/dpv-gdpr/#A7-3

– An ontology for cookies (referred to as OntoCookie[4]), which is openly available online.
– A cookie KG.
– A personalised cookie visualisation application.

The rest of the paper is structured as follows. Section 2 presents an overview of related work relevant to our study. Section 3 outlines our approach and the followed methodology. The implementation of our work is presented in Section 4, while its evaluation is presented in Section 5.1. The evaluation results are described in Section 5.2. Conclusions and future work discussions are presented in Section 6.

2. Related Work

The work in [5, 24] and [25] has shown that, for users, giving consent and being aware of what the action implies often does not imply the same thing. Bechmann [5] confirms this through a qualitative study among Danish students. The results in [5] show that there exists a non-informed consent culture among social media platform users and that although none of the participants of the study had read the privacy policies, all have given consent. Similar results were also shown in the work of Joergensen et al. [24], which further confirms that users rarely read the presented to the data-sharing terms and conditions before granting consent. Furthermore, statistics from countries within and outside the EU show that most users of social networking sites do not read the privacy policies of the sites or the third-party applications that use their data [5].

Sanchez-Rola et al. [25] present an explorative study on the topic of users' perception and reaction to cookie disclaimers. The results show that the majority of the participants view cookie disclaimers as an annoyance during their browsing time rather than a useful source for data sharing information. Although the participants of the study claimed to have privacy concerns regarding cookies and data collection practices, the evaluation results showed that the cookie disclaimers did not play a significant role in the participants' decision to continue navigating the website. Greater importance was given to factors such as the reputation of the website.

Santos et al. [26] present an in-depth analysis of how data sharing information is presented via cookie banners with the goal of receiving consent. Following the legal requirements of the ePrivacy Directive (ePD) [27] and the GDPR, around 400 cookie banners presented on the most popular English-speaking websites were manually annotated. The evaluation results showed that 89% of the cookie banners violated applicable laws. More specifically, 61% of the banners violated the purpose specificity requirement by mentioning vague purposes, including "user experience enhancement" while further, 30% of banners used positive framing, breaching the freely given and informed consent requirements.

In a similar study, Soe et al. [28] manually analysed 300 data collection consent notices from news outlets, which were built to ensure compliance with GDPR. The analysis uncovered a variety of dark patterns (i.e. deceptive design practices aimed at manipulating users' actions) [29, 30].

Ware [31], Rossi et al. [32] and Drozd et al. [33] highlight the importance of visualisation as a way to support the comprehension of the information that is being communicated to the end user. According to [31], the highest bandwidth channel of communication between humans and machines is provided by visual displays. The amount of information that can be transmitted makes data visualisation a highly appropriate method to communicate information to the users. In [32], Rossi et al. emphasise the fact that the use of visualisation is explicitly suggested by the European Union (EU) in legislations such as the GDPR (Rec. 58, Art. 12(7)) as a way to improve the comprehension of the information provided to data subjects. One can acknowledge that visual elements and visualisations in general play a crucial role in obtaining informed consent. In recent years there has been a rise in the attempts to build applications that provide more transparency regarding personal data processing through applying different visualisation approaches. Steichen et al. [34] go deeper into the topic of information visualisation by taking into consideration the role of the individual cognitive style of the users in their ability to perceive the information being communicated in a visual form. Results show that the individual cognitive style plays a significant role in tasks related to information

[4] https://github.com/STIInnsbruck/OntoCookie

visualisation in general. Findings of the presented work also provide motivation for the development of personalised information visualisation systems based on the cognitive style of the individual users.

In this context, Drozd et al. [33] present the CoRe [35] and the Consent reqUest useR intErface (CURE) user interfaces (UIs), which have the main goal of easing the process of granting consent and providing more transparency into data sharing. The evaluation of the two UIs showed that, indeed, visualisations helped raise awareness of what consent is. However, issues such as information overload due to design complexity were still present. Similar solutions for consent visualisation include the work in [36] and [37], which focuses on raising data sharing awareness with visualisations. All these studies show that there is a prominent need for consent solutions that support higher level of transparency, focus on the needs of the users and on raising awareness regarding data sharing.

These challenges are the main motivation for the work of Bless et al. [4], who not only consider the consent comprehension of users but also consider representing the data in a machine-readable format by utilising semantics. In comparison to the work in [35] and [33], the authors focus on visualising data sharing flows after consent is given. The developed visualisation is based on a KG and has helped to raise the awareness of consent data sharing significantly. Further, the later version of the KG has also shown to be beneficial when performing GDPR compliance as shown in [22]. At [38], a semantic based visualisation approach that visualises data related to the published research in the Flemish region is presented by Dimou et al. This work highlights how semantically annotating data improves the quality of the data, the diversity of the information and the knowledge integration. Brunetti et al. [39] present a formal Linked Data Visualisation Model(LDVM), which connects data and visualisation dynamically. The implementation of LDVM in this work comprises of a library of generic visualisation, which allows normal users and experts to get an overview and explore the Data Web and perform analysis on Linked Data.

The use of semantics for consent and GDPR compliance is also prominent in the work of Kirrane et al. [23], who come to the conclusion that semantics can be used to build more accurate models to detect security issues. Moreover, the meaningful interpretation of personal data that is exchanged between users and other entities on the web can be used to empower users to have better control over these interactions and therefore improve the way they manage their online privacy. The semantic approach can also bring advantages to companies through automation, which is enabled by the semantic machine-readable and machine-processable representation of data-related privacy policies. The benefits of semantics in the legal domain, especially for consent, are also discussed in [12, 16, 40, 41].

3. Selected Approach and Methodology

We approach the issue of web cookie comprehension and cookie data sharing from both human and machine perspective. However, both sides have different comprehension needs that need to be addressed. On the human side, we focus on utilising data visualisations in graphical and tabular forms. Our cookie visualisation tool provides individuals with an interface that takes as an input cookie logs and displays personalised statistics that are aimed at providing more transparency into cookie-based data sharing. Consequently, this can help raise indivdiuals' awareness regarding the implications of granting consent for cookies. Semantic technology, namely the OntoCookie [4.1] ontology and the KG build with it are used to represent the cookie data in a meaningful machine-readable and interpretable way. Our approach is motivated by the increase of cookie and consent requests online after the acceptance of the GDPR

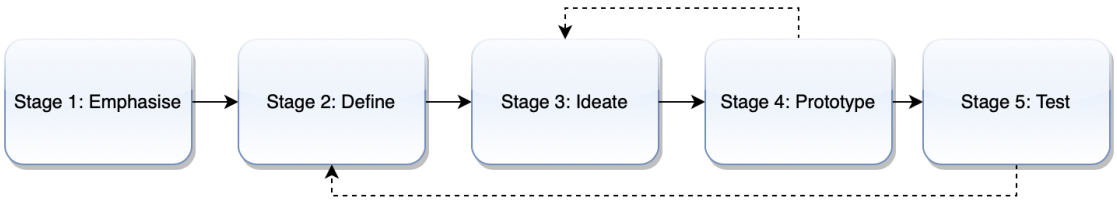


Fig. 1. Methodology overview

and tries to bridge the gap between the Semantic Web, privacy and legal domains.

The methodology (Fig. 1) followed for the development of our cookie visualisation tool is inspired by the design thinking process [42], which is a solution-based approach to solving problems by considering human needs. The development process consisted of the following stages: emphasise, define, ideate, prototype and test. The first stage was to understand the problem of cookies and consent comprehension.This included research of the privacy domain and, more specifically, of cookies and how data and consent are handled by browsers.

Existing solutions for cookies (with and without the use of semantics) were also reviewed. During the second stage, the main research problem was defined, and system requirements were derived. The third stage focused on analysing the requirements and generating ideas for the design of the tool. The fourth stage focused on prototyping the solution. This was done in several stages as well. We started with (i) building the OntoCookie ontology, (ii) building a prototype UI for cookie import, (iii) implementing functionalities such as cookie annotation, (iv) building the cookie KG and finally (v) visualising different cookie statistics on the UI. The fifth stage consisted of the usability and design evaluation of the tool with users, analysis of the results and the comparison to existing cookie solutions.

The technologies used for the development of our cookie visualisation tool include the Flutter[5] toolkit for front-end development, the NodeJS[6] back-end environment for our middleware, Protégé[7] and GraphDB[8] for building and storing the OntoCookie KG. Cookie Editor[1] browser plug-in, available for Google Chrome, Firefox, Opera, and Microsoft Edge was used to allow users to export their cookies for each website separately. Further details about the implementation are presented in the next section.

4. Implementation

The goal of our solution is to provide an overview of a user's cookies in the form of statistics in order to ease the understanding of cookies and the data they collect. Further, we add a machine-readable format of cookies with a newly developed cookie ontology. This section presents the implementation details of our proposed KG-based tool for cookie visualisations. Section 4.1 presents an overview of the OntoCookie ontology. Section 4.2 presents the two possible action flows of using our tool, while Section 4.3 presents the implementation details of the visualisation.

4.1. OntoCookie: A Domain Ontology for Cookies

The OntoCookie ontology (Fig. 2), which this section presents, is a formal representation of the cookie domain. The ontology was built as a response to the lack of openly available semantic models for cookies. By following a top-down ontology engineering approach (see [43]), the main classes, sub-classes and the relationships between them were defined. When defining the subclasses, the "isA" constraint was followed (e.g. SessionCookie *isA* Cookie). The main source of the cookie information are the cookie log files provided by the Cookie Editor extension. The ontology was defined in Protégé[9]. Currently, the ontology comprises of 115 axioms, 20 classes, 13 object properties and 2 data properties.

The class *OntoCookie:Cookie* represents several types of cookies that are widely used, namely *OntoCookie:Session Cookie*, *OntoCookie:HostOnlyCookie*, *OntoCookie:HttpOnlyCookie*, *OntoCookie:PersistentCookie*, *OntoCookie: AuthenticationCookie*, *OntoCookie:TrackingCookie*. Definitions for each cookie type have been provided as well by reusing *dc:description* from the Dublin Core[10] vocabulary. To model metadata associated with each cookie, we have reused *Date* from schema.org[11]. In this way, one can model the *startDate*, *endDate*. For the date, the ISO 8601 date format is used. The object property *OntoCookie:hasDuration* can be used to specify the duration of a specific cookie.

[5] https://flutter.dev
[6] https://nodejs.org/en/
[7] https://protege.stanford.edu
[8] https://www.ontotext.com/products/graphdb/
[9] https://protege.stanford.edu/
[10] https://www.dublincore.org
[11] https://schema.org

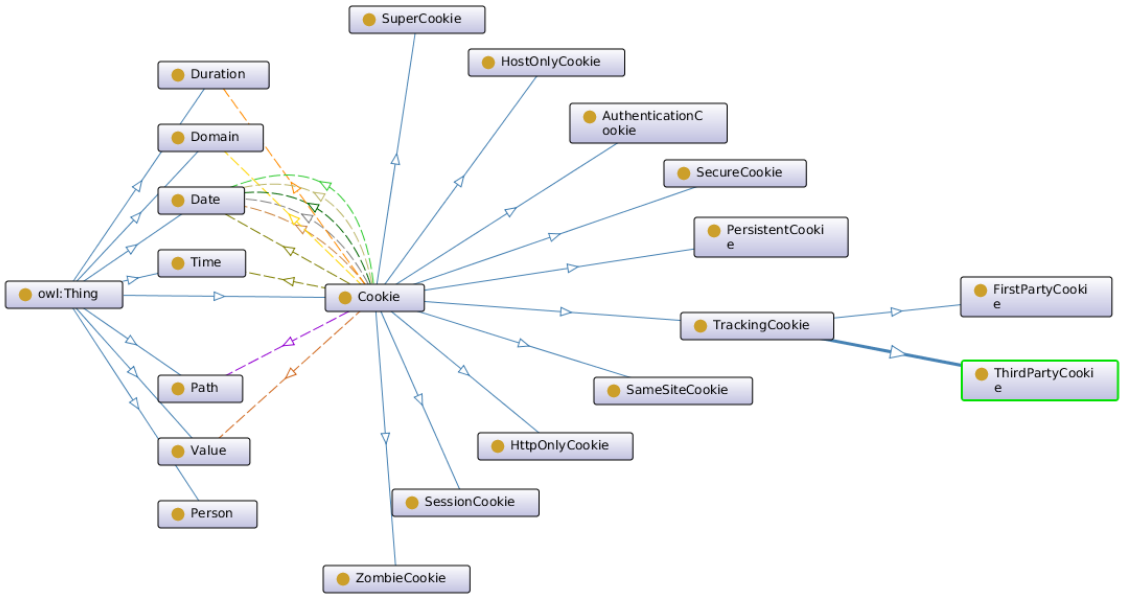


Fig. 2. The OntoCookie ontology

Cookies can also be related to a specific domain (*OntoCookie:Domain*), which signifies the domain for which the cookie is valid and can be submitted with every request for this domain or its subdomains. If a domain is not specified, then the hostname of the originating server is used as the default value. Lastly, we have defined the concept of a person (*OntoCookie:Person*) to whom the cookies belong. While using the cookie visualisation tool, each user is asked to generate a unique hash (modeled by the data property *OntoCookie:hashed_id*), which is used later for retrieving the specific cookies from the generated KG.

4.2. *The Action Flow*

In order to adhere to GDPR regulations, the users are asked for their informed consent

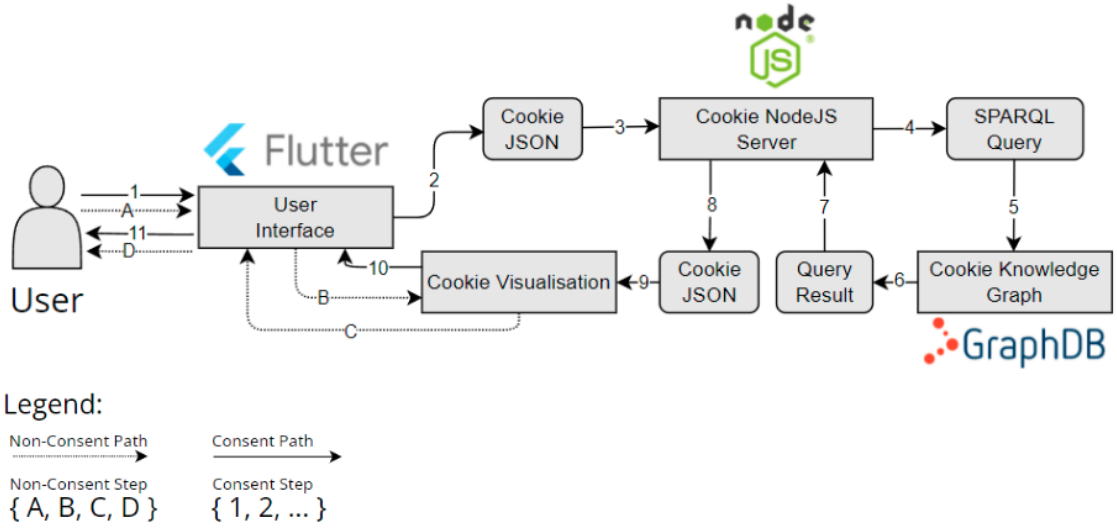


Fig. 3. The action flow of the application

(i.e. users are explicitly asked whether they want their cookie data saved in the KG via a consent dialog). If consent is given, the action flow consists of 11 steps (numbered from 1 to 11 in Fig. 3). In step **1**, the users are provided with details about the application. In step **2**, the users can import the collected cookie data in JSON format into a designated text field for this purpose. Next, the users have the option to visualise their data. However, before continuing to the next step they are asked for consent via a consent dialog.

In the case of consent, the action flow continues to step **3**, where the JSON file with cookie data is directed to the NodeJS middleware. A SPARQL `INSERT` query is constructed by the middleware in step **4**, where it is executed in the KG during step **5**.

Upon completion of the query, the middleware executes a `SELECT` query in steps **6** and **7**, where it fetches all the uploaded cookie data by the current user. In step **8**, the queried data is stored in a JSON format. In step **9**, the data is processed and consequently visualised in step **10**. In step **11**, the users can view their cookies data in a human-readable format.

In case the users do not consent to have the cookies stored in the KG, the data will not be annotated to the KG. Users can import the cookies collected into the designated text field created for this purpose. After deciding not to consent (i.e. users have decided not to save the cookie data consumed in the KG), the data will be locally processed and accordingly visualised. In this case, cookie data imported through the Cookie Editor extension, will be deleted

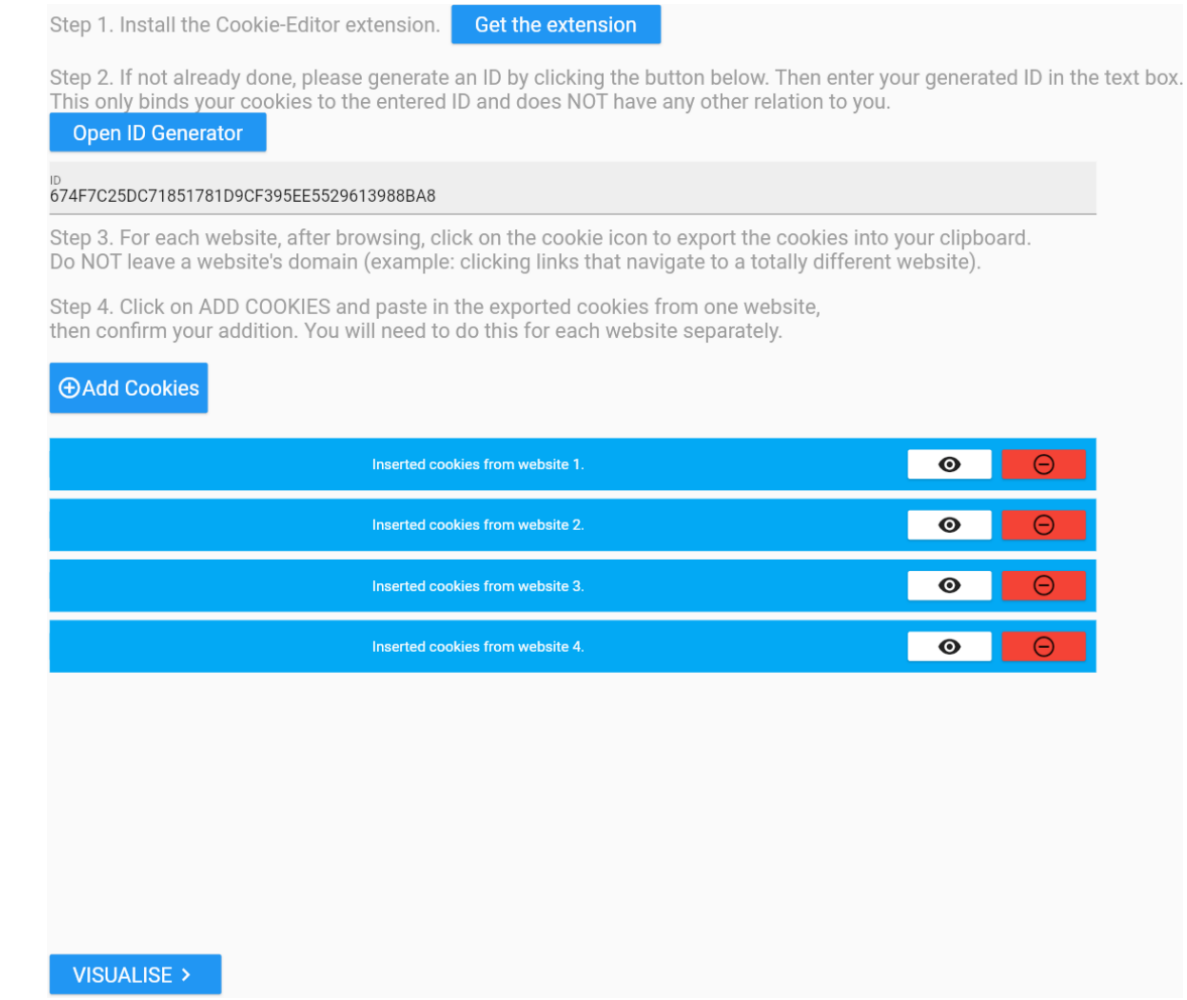


Fig. 4. Main input page of the cookie visualisation tool

once the application window is closed. The non-consent action flow follows steps **A**, **B**, **C** and **D** in succession (Fig. 3).

4.3. UI and the Connection to the Back-End

The UI is organised in two parts. The first part is a general guide of six steps on how to use the visualisation tool (Fig. 4). Step **1** contains a link to the extension we use to export cookies in JSON format. Step **2** asks the user to enter their randomly generated ID, which is created at the start of our evaluation process. Step **3** and **4** explain how the user should use the Cookie Editor browser extension to import their cookies into the visualisation tool. For each website, a separate import has to be done, as the browser extension loses the cookies of a website once the user navigates to a different website domain (e.g., navigating from "Wikipedia.org" to "Euronews.com"). Once the users click on the *Visualise* button, they are asked if they would like to consent to store their cookies on the KG for 10 days explained in sections 4.2.

Consequently, the second part of the UI displays all the cookie data, except the stored value, retrieved from the browser extension (Fig. 5). Here, the information is divided into four segments. Segment **1** lists all cookies with their domain, name, type, and duration. Segment **2** contains a bar chart grouping the amount of cookies based on their duration. A pie chart containing the distribution of the cookies among all visited websites is visualised in segment **3**. Charts were created with the help of the *charts_flutter*[12] library. Segment **4** contains a button that give the users the opportunity to withdraw consent and erase data from the KG if agreed to share it previously. Segment **5** illustrates which websites stand out for storing cookies (i.e. longest cookie, shortest cookie, the total amount of cookies and the average duration of all cookies combined).

To build the back-end, we used NodeJS and also Express for the routing. This has made the creation of our Application Programming Interface (API) easy to use. As mentioned in previous sections, we have created a KG in order to save the information on cookies and their relations with each other. Our KG is contained in GraphDB, a Resource Description Framework (RDF) database for KGs. The back-end is connected to

Fig. 5. Overview of the cookie visualisation statistics

[12]https://pub.dev/packages/charts_flutter

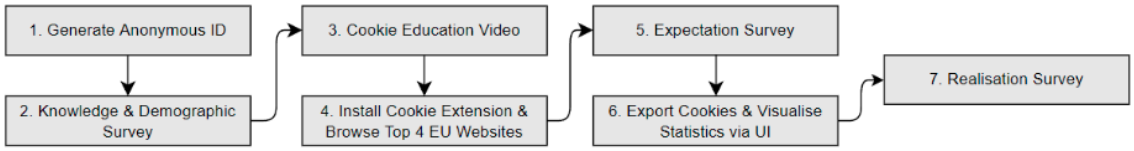


Fig. 6. Evaluation flow

GraphDB[13] using the *sparql-client* library. In this way, we perform `SELECT`, `INSERT`, and `DELETE` queries against a SPARQL endpoint via HTTP. All of the source code can be found at https://github.com/STIInnsbruck/OntoCookie including the OntoCookie ontology and a link to try out the cookie visualisation tool.

5. Evaluation

This section presents details about the evaluation of the presented cookie visualisation tool, namely the evaluation set up (Section 5.1) and the evaluation results (Section 5.2).

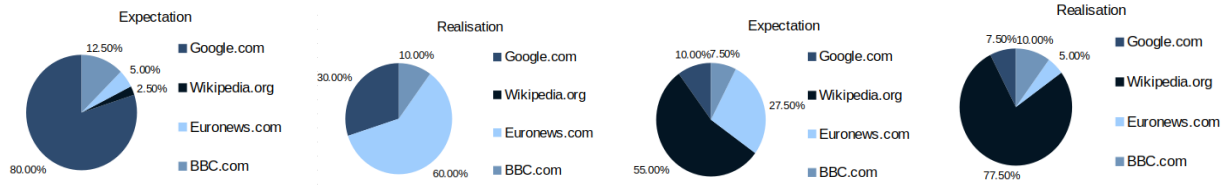
5.1. Evaluation Set Up

To evaluate our solution, its usability and design, three questionnaires (i.e. demographics, expectation and realisation) using Google Forms were created. The evaluation was done in seven stages (Fig. 6). First, the participants were asked to generate a unique ID using the SHA1[14] online hash generator and then to complete a demographics survey. Next, in stage 3, they were presented with an introductory video[15] that contains general information on cookies (what cookies are, different cookie types etc.). The goal of this was to familiarise the end-user with the topic. Following the video, in stage 4, the participants were asked to install the Cookie Editor extension and to browse four websites for the time span of two minutes. The extension is available for the Google Chrome, Firefox and Microsoft Edge browsers and provides an export button that allows the users to export their cookie data into their clipboard. For the work with the cookies collection, we have selected four highly used websites ("Google.com"[16], "Wikipedia.org"[17], "BBC.com"[18], and "Euronews.com"[19]) that do not require users to register to access information. In this way, the cookies which we collect do not have sensitive information such as usernames and passwords. During stage 5, the participants were presented with a pre-use (i.e. before using the tool) expectation survey, which contains questions to evaluate their general knowledge of cookies and the expectation of what data cookies can collect. Having completed that, in stage 6, the participants were asked to export their cookies with the Cookie Editor and to import them into the cookie visualisation tool and visualise the data. To measure in a quantified manner whether participants' comprehension of cookies has changed after using the presented tool, all participants were asked to fill in a post-use realisation survey. The analysis of the results are presented in the next sections.

5.2. Evaluation Results

For the evaluation, 40 participants (25 male and 15 female) took part in the survey. The age of the participants varied between 18-35 years old where 92.5% were within the range of 18-30 years old and 7.5% were within the range of 30-35 years old. The participants were selected from different backgrounds (computer science students, non-computer science students, researchers, computer-science experts, non-computer science experts) and were

[13] https://www.ontotext.com/products/graphdb/
[14] https://passwordsgenerator.net/sha1-hash-generator/
[15] https://www.youtube.com/watch?v=KKZlEaAWAao
[16] https://www.google.com/
[17] https://www.wikipedia.org/
[18] https://www.bbc.com
[19] https://www.euronews.com



(a) Expectation for the website with the highest amount of cookies

(b) Realisation for the website with the highest amount of cookies

(c) Expectation for the website with the least amount of cookies

(d) Realisation for the website with the least amount of cookies

Fig. 7. Survey results for the amount of cookies collected by each website

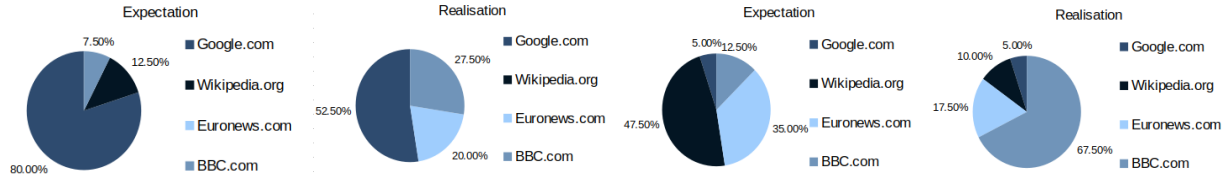
based in different countries in Europe (namely, Austria, the Netherlands, Bulgaria and Albania). They were recruited via university network and personal connections. Out of the 40 participants, 20% acknowledged that their highest level of education completed was a high school degree. For 57.5%, the highest level of education obtained was a bachelor's degree, and for 22.5%, the highest level of education obtained was a master's degree. 30% of the participants declared a very high-level internet surfing competency, 47.5% a high level of competency, 20% declared an average internet surfing level competency, and 2.5% declared a low level of internet surfing competency. 65% of our participants spend more than 4 hours per day on on the internet, 25% spend 3-4 hours per day while 10% spend 1-2 hours per day on the internet.

5.2.1. Expectation vs. Realisation

In order to measure the level of comprehension of the users in regards to the cookies collected during the browsing time of the four websites, we at first asked them about their expectation (i.e. how the users expected the results to be before using the application) and compared them with the personalised data (i.e. the factual results) which were visualised by the application. For this purpose, questions related to the amount, duration and source of the cookies collected were asked.

More specifically, to the question: *"Which of the websites do you think has the highest amount of cookies?"*, 80% of our participants expected it to be "Google.com", 2.5% expected it to be "Wikipedia.org", 5% expected "Euronews.com" and 12.5% expected "BBC.com" to have the highest of cookies. In contrast to the users' expectations, the results showed that for 60% of the participants, the highest amount of cookies consumed were generated from "Euronews.com", for only 30% the highest amount of cookies collected were from "Google.com" and for 10%, the highest amount of cookies collected were from "BBC.com" (Fig. 7).

To the question: *"Which of the websites do you think has the least amount of cookies?"*, 55% of the participants expected it to be the website "Wikipedia.org", 27.5% expected it to be "Euronews.com", 10% expected it to be "Google.com" and 7.5% expected to be "BBC.com". Results showed that in 77.5% of the cases, "Wikipedia.org" had the least amount of cookies, 10% of the cases "BBC.com" had the least amount of cookies collected, followed by "Google.com" and "Euronews.com" with 7.5% and 5% respectively (Fig. 7).



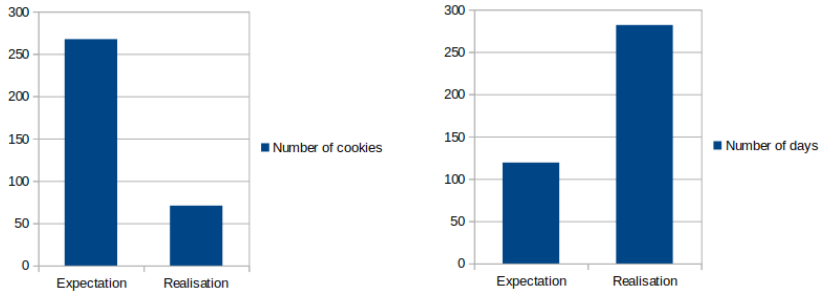
(a) Expectation for the website with the longest lasting cookie

(b) Realisation for the website with the longest lasting cookie

(c) Expectation for the website with the shortest lasting cookie

(d) Realisation for the website with the shortest lasting cookie

Fig. 8. Survey results for the duration of cookies collected by each website



(a) Expectation vs. realisation of the average number of cookies collected

(b) Expectation vs. realisation of the average number of days the cookies collected last

Fig. 9. Comparison of expectation vs. realisation averages

When asked: *"Which of the websites do you think has the longest lasting cookie?"*, 80% of the participants expected it to be "Google.com", 12.5% expected it to be "Wikipedia.org" and 7.5% expected that the longest lasting cookie originated from "BBC.com". Meanwhile, the results obtained show that in 52.5% of the cases, "Google.com" had the longest lasting cookies, "BBC.com" had the longest lasting cookie in 27.5% of the cases and on 20% of the cases the longest lasting cookie belonged to "Euronews.com" (Fig. 8).

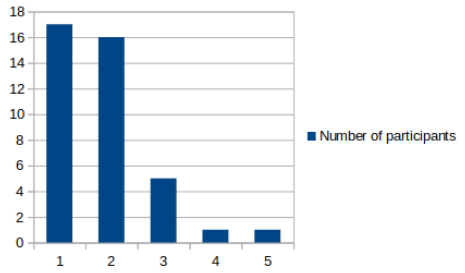
To the question: *"Which website do you think has the shortest lasting cookie?"*, 47.5% of the participants expected "Wikipedia.org" to have the shortest lasting cookie, 35% answered "Euronews.com", 12.5% answered "BBC.com" and 5% expected the shortest lasting cookie to belong to "Google.com". On the contrary, the realisation results showed that in 67.5% of the cases, "BBC.com" had the shortest lasting cookie, in 17.5% of the cases "Euronews.com" had the shortest lasting cookie, in 10% of the cases "Wikipedia.org" had the shortest lasting cookie while on 5% of the cases the shortest lasting cookie belonged to "Google.com" (Fig. 8).

The claim that the users' knowledge on cookie data is vague and insufficient was further strengthened by the significant differences detected between expectation and realisation, with respect to the total amount of cookies collected and their duration. More precisely, on average, the participants expected the total number of cookies collected during the two minutes of website browsing to be 267.4. Results from the realisation survey showed that, on average, a total amount of 70.8 number of cookies were collected during their surfing time, approximately 73% less than the users' expectation. Regarding the duration of cookies collected, when asked: *"How many days on average do you think cookies last?"*, the response mean was 119.2 days. Results obtained from the realisation survey show that on average, the cookies collected during the session lasted 281.8 days, approximately 137% more than the expectation (Fig. 9).

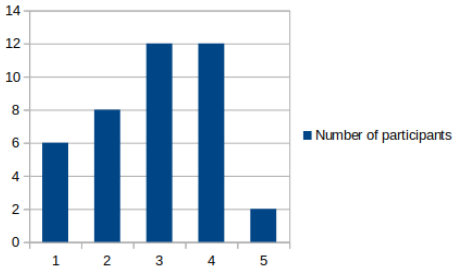
The question: *"How carefully do you read the cookie notification banner before proceeding to give consent or not?"* was numerically encoded on a scale from 1 ("Not carefully at all") to 5 ("Very carefully") and was asked to the participants before using the application. 82.5% said that they do not read the banner carefully at all or not carefully, while 17.5% were neutral, read the banner carefully, or very carefully. After the participants used the application, we asked the question: *"How carefully will you be reading the cookie notification banner before agreeing to give consent or not?"*. Participants responded that they were willing to be more careful when reading the cookie notification banner before agreeing to cookies, showing a significant increase in awareness related to the process of web cookie agreement. Specifically, 65% of the participants were neutral, willing to be careful or willing to read very carefully the cookie notification banner, and 35% of the participants confessed that they would continue not to be careful or not careful at all when agreeing to the cookie notification banner (Fig. 10).

5.2.2. *Further Survey Results*

Furthermore, participants answered a set of questions related to their general feeling about cookie data privacy after using the application and also how they will approach internet cookies in the future.



(a) Scaled answers to the question: *"How carefully do you read the cookie notification banner before proceeding to give consent or not?"*



(b) Scaled answers to the question: *"How carefully will you be reading the cookie notification banner before agreeing to give consent or not?"*

Fig. 10. Comparison of results regarding the carefulness with which the participants read and will read the cookie notification banner before and after using the application

Participants of the survey exhibited the feeling that cookies were intrusive to their online privacy. Specifically, to the question: *"Do you feel as if the website knows more than you expect"*, 62.5% of the participants answered "Yes" while 37.5% answered "No". The possible answer to the question: *"How do you feel about your privacy when browsing the internet in regards to the safety of your data, after being presented with information on the cookies you consumed?"* was numerically encoded on a scale from 1 ("Not safe at all") to 5 ("Very safe"). 57.5% of the participants replied either 1 or 2, meaning that they did not feel safe about their data privacy, 37.5% were neutral, while 5% felt safe in regards to their data privacy on the internet. To the question: *"Do you feel it is fine for websites to collect the given amount of cookies?"*, 82.5% of the participants answered "No" and that they "Wished for fewer cookies to be collected", 15% answered "Yes" and that "Things may continue unchanged", and 2.5% answered "No" and that they "Wished for more cookies to be collected".

Results showed that participants would embrace an overview tool to manage their cookies. Precisely, the question *"Would you feel more confident surfing the internet if you were given an overview tool to manage your cookies?"*. 72.5% answered "Yes", 25% were neutral and 2.5% answered "No". Further, we asked the participants: *"Would you feel more knowledgeable about cookies and your browsing privacy if you were given an overview tool to manage your cookies?"*. Results show that 95% of the participants would feel more knowledgeable about cookies and browsing privacy if they were given an overview tool to manage them. 2.5% were neutral while 2.5% responded "No".

6. Conclusions

In this paper, we presented a KG based tool for cookie information visualisation, which focuses on easing users' comprehension of cookies and on raising awareness of cookie data sharing. Our evaluation confirms that users (even proficient web surfers) lack detailed knowledge about cookies and the consequences of granting consent for them. For example, the duration of the cookies being stored, the amount of cookies collected during the browsing time and practices of different websites with regards to the cookies they use, commonly do not match the users' expectations. The results also showed that the cookie visualisation tool presented helped improve users' comprehension of cookies and has raised awareness regarding data sharing on the web. More specifically, after being presented with the application, an increase of 47.5% of the the users willingness to be more cautious when reading the cookie consent banner before giving consent was noticed. The outcome of the evaluation also confirms that users are ready to embrace an overview tool that helps them manage their cookies. 72.5% of the participants agreed that they would feel more confident about their privacy on the web if they were given such overview tool, and 95% of the users admitted that they would feel more knowledgeable about cookies if an overview tool to manage cookies was in their disposal . In addition, we believe that this work helps breach the gap between the Semantic Web and the security and privacy domains.

Complex issues related to the privacy domain in general [44], specifically privacy issues related to the collection of cookie data, are far from being resolved. Currently, our cookie visualisation tool is dependent on the Cookie Editor [1] extension and the information captured by it. Our future goal is to remove this dependency by extending the functionalities of our cookie visualisation tool (i.e. implement a cookie capture functionality). Another possible future direction is to extend the use case of our application such that not only it serves as a tool to communicate information, but it also allows users to act on it by offering them the possibility to manage cookies. On the semantic side, we have presented a novel ontology for cookies that can be extended for different domains and use cases. We believe that its reuse and extension will inspire further collaboration between semantic and privacy experts. The uses of the KG for detecting security breaches and data sharing patterns (within the cookies) can be explored as well.

Acknowledgements

This research is supported by the CampaNeo project funded by FFG (grant 873839) as well as the smashHit EU project funded under Horizon 2020 (grant 871477). We would like to thank Harshvardhan J. Pandit for sharing insights on cookies.

References

[1] R. Tirtea, C. Castelluccia and D. Ikonomou, Bittersweet cookies. Some security and privacy considerations, *European Union Agency for Network and Information Security-ENISA* (2011).

[2] S. Jegatheesan, Cookies Invading Our Privacy for Marketing Advertising and Security Issues, *ArXiv* **abs/1305.2306** (2013).

[3] European Parliament, Council of the European Union, Regulation (EU) 2016/679 of the European Parliamentand of the Council of 27 April 2016 on the protectionof natural persons with regard to the processing of per-sonal data and on the free movement of such data, andrepealing Directive 95/46/EC (General Data ProtectionRegulation), *Official Journal of the European Union, L119* (May 2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[4] C. Bless, L. Dötlinger, M. Kaltschmid, M. Reiter, A. Kurteva, A.J. Roa-Valverde and A. Fensel, Raising Awareness of Data Sharing Consent Through Knowledge Graph Visualisation, *Studies on the Semantic Web* (2021). doi:10.3233/ssw210034.

[5] A. Bechmann, Non-Informed Consent Cultures: Privacy Policies and App Contracts on Facebook, *Journal of Media Business Studies* **11** (2015), 21–38. doi:10.1080/16522354.2014.11073574.

[6] C. Utz, M. Degeling, S. Fahl, F. Schaub and T. Holz, (Un)Informed Consent: Studying GDPR Consent Notices in the Field, in: *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, CCS '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 973–990–. ISBN 9781450367479. doi:10.1145/3319535.3354212.

[7] S. Human, H.J. Pandit, V.P. Morel, C. Santos, M. Degeling, A. Rossi, W. Botes, V. Jesus and I. Kamara, Data Protection and Consenting Communication Mechanisms: Current Open Proposals and Challenges, International Workshop on Privacy Engineering – IWPE'22, Co-located with 7th Ieee European Symposium on Security and Privacy, 6 June 2022, Genoa, Italy.

[8] A.D. Miyazaki, Online privacy and the disclosure of cookie use: Effects on consumer trust and anticipated patronage, *Journal of Public Policy & Marketing* **27**(1) (2008), 19–33. doi:10.1509/jppm.27.1.19.

[9] V. Ha, K. Inkpen, F. Al Shaar and L. Hdeib, An Examination of User Perception and Misconception of Internet Cookies, in: *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '06, Association for Computing Machinery, New York, NY, USA, 2006, pp. 833–838–. ISBN 1595932984. doi:10.1145/1125451.1125615.

[10] K. Fatema, E. Hadziselimovic, H.J. Pandit, C. Debruyne, D. Lewis and D. O'Sullivan, Compliance through Informed Consent: Semantic Based Consent Permission and Data Management Model, in: *PrivOn@ISWC*, 2017. http://ceur-ws.org/Vol-1951/PrivOn2017_paper_5.pdf.

[11] H.J. Pandit, C. Debruyne, D. O'Sullivan and D. Lewis, GConsent-a consent ontology based on the GDPR, in: *European Semantic Web Conference*, Springer, 2019, pp. 270–282. doi:10.1007/978-3-030-21348-0_18.

[12] A. Kurteva, T.R. Chhetri, H.J. Pandit and A. Fensel, Consent through the lens of semantics: State of the art survey and best practices, *Semantic Web* (2021), 1–27. doi:10.3233/SW-210438.

[13] A. Piplai, S. Mittal, A. Joshi, T. Finin, J. Holt and R. Zak, Creating cybersecurity knowledge graphs from malware after action reports, *IEEE Access* **8** (2020), 211691–211703. doi:10.1109/ACCESS.2020.3039234.

[14] Y. Jia, Y. Qi, H. Shang, R. Jiang and A. Li, A practical approach to constructing a knowledge graph for cybersecurity, *Engineering* **4**(1) (2018), 53–60. doi:10.1016/j.eng.2018.01.004.

[15] Y. Qi, R. Jiang, Y. Jia and A. Li, Attack analysis framework for cyber-attack and defense test platform, *Electronics* **9**(9) (2020), 1413. doi:10.3390/electronics9091413.

[16] D. Fensel, *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, 2nd edn, Springer-Verlag, Berlin, Heidelberg, 2003. ISBN 3540003029. doi:10.1007/978-3-662-09083-1.

[17] B. Heitmann and C. Hayes, An architecture and methodologies for federated, privacy-enabled personalisation on the Web of Data, *Semantic Web* (2011).
[18] K. Zhang and J. Liu, Review on the Application of Knowledge Graph in Cyber Security Assessment, *IOP Conference Series: Materials Science and Engineering* **768** (2020), 052103. doi:10.1088/1757-899X/768/5/052103.
[19] T.R. Chhetri, A. Kurteva, J.G. Adigun and A. Fensel, Knowledge Graph Based Hard Drive Failure Prediction, *Sensors* **22**(3) (2022). doi:10.3390/s22030985. https://www.mdpi.com/1424-8220/22/3/985.
[20] S. de Lusignan, S. Shinneman, I. Yonova, J. van Vlymen, A. Elliot, F. Bolton, G. Smith and S. O'Brien, An Ontology to Improve Transparency in Case Definition and Increase Case Finding of Infectious Intestinal Disease: Database Study in English General Practice, *JMIR Medical Informatics* **5** (2017), e34. doi:10.2196/medinform.7641.
[21] N. Freire and S.d. Valk, Automated interpretability of linked data ontologies: : an evaluation within the cultural heritage domain, in: *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 3072–3079. doi:10.1109/BigData47090.2019.9005491.
[22] T.R. Chhetri, A. Kurteva, R.J. DeLong, R. Hilscher, K. Korte and A. Fensel, Data Protection by Design Tool for Automated GDPR Compliance Verification Based on Semantically Modeled Informed Consent, *Sensors* **22**(7) (2022). doi:10.3390/s22072763. https://www.mdpi.com/1424-8220/22/7/2763.
[23] S. Kirrane, S. Villata and M. d'Aquin, Privacy, security and policies: A review of problems and solutions with semantic web technologies, *Semantic Web* **9**(2) (2018), 153–161. doi:10.3233/SW-180289.
[24] R. Joergensen and I. Review, The unbearable lightness of user consent, *Internet Policy Review* **Volume 3** (2014). doi:10.14763/2014.4.330.
[25] I. Sanchez-Rola, M. Dell'Amico, P. Kotzias, D. Balzarotti, L. Bilge, P.-A. Vervier and I. Santos, Can I Opt Out Yet?: GDPR and the Global Illusion of Cookie Control, 2019, pp. 340–351. doi:10.1145/3321705.3329806.
[26] C. Santos, A. Rossi, L. Sanchez Chamorro, K. Bongard-Blanchy and R. Abu-Salma, Cookie Banners, What's the Purpose? Analyzing Cookie Banner Text Through a Legal Lens, in: *Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society*, WPES '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 187–194–. ISBN 9781450385275. doi:10.1145/3463676.3485611.
[27] European Parliament, Council of the European Union, Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications), 2002. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32002L0058.
[28] T.H. Soe, O.E. Nordberg, F. Guribye and M. Slavkovik, *Circumvention by Design - Dark Patterns in Cookie Consent for Online News Outlets*, in: *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, Association for Computing Machinery, New York, NY, USA, 2020. ISBN 9781450375795. https://doi.org/10.1145/3419249.3420132.
[29] A. Mathur, M. Kshirsagar and J. Mayer, What makes a dark pattern... dark? Design attributes, normative considerations, and measurement methods, in: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–18. doi:10.1145/3411764.3445610.
[30] C.M. Gray, Y. Kou, B. Battles, J. Hoggatt and A.L. Toombs, *The Dark (Patterns) Side of UX Design*, in: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, USA, 2018, pp. 1–14–. ISBN 9781450356206. https://doi.org/10.1145/3173574.3174108.
[31] C. Ware, *Information visualization: perception for design*, Morgan Kaufmann, 2019. ISBN 0123814642. doi:10.1016/C2009-0-62432-6.
[32] A. Rossi and M. Palmirani, A Visualization Approach for Adaptive Consent in the European Data Protection Framework, in: *2017 Conference for E-Democracy and Open Government (CeDEM)*, 2017, pp. 159–170. doi:10.1109/CeDEM.2017.23.
[33] O. Drozd and S. Kirrane, Privacy CURE: Consent Comprehension Made Easy, 2020. ISBN 978-3-030-58200-5. doi:10.1007/978-3-030-58201-2_9.
[34] B. Steichen and B. Fu, Towards Adaptive Information Visualization - A Study of Information Visualization Aids and the Role of User Cognitive Style, *Frontiers in Artificial Intelligence* **2** (2019). doi:10.3389/frai.2019.00022.
[35] O. Drozd and S. Kirrane, I Agree: Customize Your Personal Data Processing with the CoRe User Interface, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2019), 17–32. ISBN 9783030278120. doi:10.1007/978-3-030-27813-7_2.
[36] J. Angulo, S. Fischer-Hübner, T. Pulls and E. Wästlund, Usable Transparency with the Data Track, *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (2015), 1803–1808. doi:10.1145/2702613.2732701.
[37] P. Raschke, A. Küpper, O. Drozd and S. Kirrane, Designing a GDPR-Compliant and Usable Privacy Dashboard, *Privacy and Identity Management. The Smart Revolution. Privacy and Identity 2017. IFIP Advances in Information and Communication Technology* (2017). doi:10.1007/978-3-319-92925-5_14.
[38] A. Dimou, L. De Vocht, G. Van Grootel, L. Van Campe, J. Latour, E. Mannens and R. Van de Walle, Visualizing the Information of a Linked Open Data Enabled Research Information System, *Procedia Computer Science* **33** (2014), 245–252, 12th International Conference on Current Research Information Systems, CRIS 2014. doi:https://doi.org/10.1016/j.procs.2014.06.039. https://www.sciencedirect.com/science/article/pii/S1877050914008291.
[39] J.M. Brunetti, S. Auer, R. García, J. Klímek and M. Nečaský, Formal Linked Data Visualization Model, in: *Proceedings of International Conference on Information Integration and Web-Based Applications amp; Services*, IIWAS '13, Association for Computing Machinery, New York, NY, USA, 2013, pp. 309–318–. ISBN 9781450321136. doi:10.1145/2539150.2539162.
[40] J. Cardoso and A. Sheth, *The Semantic Web and Its Applications*, 2006, pp. 3–33. ISBN 978-0-387-30239-3. doi:10.1007/978-0-387-34685-4_1.

[41] F. Zhang, N.J. Yuan, D. Lian, X. Xie and W.-Y. Ma, Collaborative Knowledge Base Embedding for Recommender Systems, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 353–362–. ISBN 9781450342322. doi:10.1145/2939672.2939673.

[42] T. Brown et al., Design thinking, *Harvard business review* **86**(6) (2008), 84.

[43] N. Noy and D. Mcguinness, Ontology Development 101: A Guide to Creating Your First Ontology, *Knowledge Systems Laboratory* **32** (2001). doi:10.1.1.136.5085.

[44] A. Oltramari, D. Piraviperumal, F. Schaub, S. Wilson, S. Cherivirala, T. Norton, N. Russell, P. Story, J. Reidenberg and N. Sadeh, PrivOnto: A semantic framework for the analysis of privacy policies, *Semantic Web* **9** (2017), 1–19. doi:10.3233/SW-170283.