CompLicy: Evaluating the GDPR Alignment of Privacy Policies - A Study on Web Platforms

Evangelia Vanezi(✉), George Zampa, Christos Mettouris, Alexandros Yeratziotis, and George A. Papadopoulos

Department of Computer Science, University of Cyprus, Nicosia, Cyprus
{evanez01,gzampa01,mettour,ayerat01,george}@cs.ucy.ac.cy

Abstract. The European Union General Data Protection Regulation (GDPR) came into effect on May 25, 2018, imposing new rights and obligations for the collection and processing of EU citizens personal data. Inevitably, privacy policies of systems handling such data are required to be adapted accordingly. Specific rights and provisions are now required to be communicated to the users, as specified in GDPR Articles 12-14. This work aims to provide insights on whether privacy policies are aligned to the GDPR in this regard, i.e., including the needed information, formulated in sets of terms, by studying the paradigm of web platforms. We present: (1) a defined set of 89 terms, in 7 groups that need to be included within a systems' privacy policy, resulting from a study of the GDPR and from an examination and analysis of real-life web platforms privacy policies; (2) the CompLicy tool, which as a first step crawls a given web platform, to infer whether a privacy policy page exists and, if it does, subsequently parses it, identifying GDPR terms and groups within, and finally, providing results for the inclusion of the necessary GDPR information within the aforementioned policy; (3) the evaluation of 148 existing web platforms, from 5 different sectors: (i) banking, (ii) e-commerce, (iii) education, (iv) travelling, and (v) social media, presenting the results.

Keywords: GDPR compliance · Privacy policies · Web platforms

1 Introduction

Privacy, constituting one of the fundamental rights of the European Union (EU) [3] was recently established in a legal framework through the EU General Data Protection Regulation (GDPR) [4], that came into effect on May 25, 2018. Privacy is becoming a critical issue as the usage of systems collecting and processing user personal data[1] is increasing. Based on recent statistics[2], it is

[1] According to GDPR, personal data are defined as information that relates to an identified or identifiable individual.

[2] https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/.

S. Cherfi et al. (Eds.): RCIS 2021, LNBIP 415, pp. 152–168, 2021.
https://doi.org/10.1007/978-3-030-75018-3_10

estimated that by 2025, 4.41 billion people will be owning social media accounts. In addition, the number of digital buyers worldwide is increasing each year. In 2020, an estimated 2.05 billion people purchased goods or services online, estimated to increase to 2.14 billions for 2021[3]. On top of that, in 2020, the usage of online educational sites has increased substantially due to the COVID-19 worldwide crisis, imposing a digital online education scheme. Large numbers of learners are impacted by this change[4]. Using such systems, users provide their personal data, which are being stored and processed in several ways. Privacy policies are included into systems, to describe the rights and obligations of the involved parties, i.e., the system and its users, in regards to personal data collection, storage, and processing. As such, all systems that are processing personal data of individuals need to include a properly formed privacy policy, providing all needed information to the users.

With the application of the GDPR, privacy policies were imposed to change and adapt their contents. Specific rights and provisions are now required to be communicated to the users, as specified in GDPR Articles 12-14. As defined, "the controller shall take appropriate measures to provide any mandatory information and communication relating to processing to the data subject", and "the controller shall provide the data subject with information necessary to ensure fair and transparent processing", including the different user rights defined within Articles 12-23 of the GDPR. Our aim is to study whether the privacy policies of software systems are following the GDPR in this regard, i.e., including and communicating the needed information to the users. The tool developed, is aiming to provide a higher level check of the inclusion of the needed GDPR information, formulated in sets of terms, within privacy policies. Extensive manual auditing or more sophisticated methodologies, e.g., exploiting Artificial Intelligence or Natural Language Processing, could be complementary, in order to examine and analyse the text in a deeper level of detail. In addition, our work does not check whether the actual system implementation (i.e., the code), complies with the content of its privacy policy, nor with the GDPR.

Towards our aim, we focus on the case study of web platforms. Web platforms were selected as they comprise a huge portion of the market, being also easily accessible in an open-manner for both users, and researchers. According to statistics[5], by 2019, 1.72 billion websites were published on the web. Our target group includes both users and software engineers. Long before the GDPR, it was shown that users do not usually read the privacy policies, but when they do, it is difficult to comprehend [8]. Even after the definition of articles mandating the clear and understandable form of privacy policies, their readability and usability is still under investigation [6,9]. Our work aims to help users understand whether a system's privacy policy follows the GDPR guidelines by providing a summary of the included and excluded terms. On the other hand, software

[3] https://www.statista.com/statistics/251666/number-of-digital-buyers-worldwide/.
[4] https://www.statista.com/chart/21224/learners-impacted-by-national-school-closures/, https://en.unesco.org/covid19/educationresponse.
[5] https://www.statista.com/chart/19058/how-many-websites-are-there/.

system engineering methodologies do not explicitly capture privacy requirements or privacy policies. Software engineers mostly use the vocabulary of data security to approach privacy challenges, which limits their perceptions of privacy mainly to third-party threats coming from outside [5]. In addition, software engineers are not familiar with legal content. Our work can also provide assistance to engineers in monitoring privacy compliance in a system design, by indicating the needed, included and missing rights and provisions from within the privacy policy text, serving as a guideline for the functionalities that are expected to be implemented into the actual system.

Towards reaching our aims we present: (1) a defined set of 89 terms, in 7 groups that need to be included within a system's privacy policy, resulting from a study of the GDPR and from an examination and analysis of real-life web platforms privacy policies; (2) a tool developed (CompLicy), which as a first step crawls a given web platform, to infer whether a privacy policy page exists and, if it does, subsequently parses it, identifying GDPR terms and groups within, and finally, providing results for the inclusion of the necessary GDPR information within the aforementioned policy; (3) the evaluation of 148 existing web platforms, from 5 different sectors: (i) banking, (ii) e-commerce, (iii) education, (iv) travelling, and (v) social media, presenting the results of their alignment to the GDPR. In specifics, the evaluation examines the existence of the following GDPR provisions and rights: "*Lawfulness of Processing*", "*Right to Erasure*", "*Right of Access by the Data Subject*", "*Right to Data Portability*", "*Right to Rectification*", "*Right to Restriction of Processing*", and "*Right to Object*".

The rest of the paper is structured as follows: Sect. 2 presents an overview of the related work and background. Section 3 discusses our methodology towards creating a list of terms to be included in web platforms privacy policies, and the final list, while Sect. 4 presents the design and implementation of the crawler and parser tool (CompLicy) for automatically locating and analysing a privacy policy of a given website. Subsequently, Sect. 5 presents an evaluation of 148 websites, from 5 different sectors, and their results. Finally, Sect. 6 concludes this paper with a discussion of the conclusions and future work.

2 Background and Related Work

The GDPR [4] was enacted by the EU in an attempt to address the issue of personal data privacy. It came into effect on May 25, 2018, imposing new rights and obligations for the collection and processing of EU residents personal data, i.e., even when the system is not located within the EU. In [13] the application of GDPR articles and provisions is studied, in the design and development of web platforms. The GDPR-compliant implementation of a case study platform is demonstrated, and a set of guidelines based on the methodology followed is extracted. The work discusses all the GDPR articles judged as relevant, and additionally explains all other GDPR articles, reasoning why they were not included in the specific implementation. In [9], the authors perform an investigation on how privacy policies can be both GDPR-compliant and usable. They synthesise

GDPR requirements into a checklist and derive a list of usability design guidelines for privacy notifications. They then provide a usable and GDPR-compliant privacy policy template for the benefit of policy writers.

In [7], a large number of privacy policies was collected and analysed in regards to their versions prior and after the GDPR establishment. The authors created an automated tool for this analysis, and were focused on the change occurring for the users experience when interacting with such privacy policies. Their conclusions were that, between March and July 2018, with May being the main point, more than 45% of the examined policies had changed. Furthermore, it was shown that GDPR was mainly affecting EU policies, rather than policies of organisations outside the EU. The authors of [1] try to address the issue of privacy policies being too long and complex with poor readability and comprehensibility to users. They propose an automated privacy policy extraction system, implemented on Android smartphones. This work's main focus is addressing users' concerns and the *transparency* requirement of the GDPR. With the same aim, [10], propose a machine learning based approach to summarise long privacy policies into short and condensed notes, and [11] presents a privacy policy summarisation tool.

In [12] the authors provide automated support for checking completeness of a privacy policy in regards to the GDPR. In order to do so, metadata from privacy policies are extracted, and a set of completeness criteria based on a conceptual model is used for recognising issues. Additionally, [2] aims at automating legal evaluation of privacy policies, under the GDPR, using artificial intelligence. In their study, they present the preliminary results of the evaluation of a number of privacy policies.

Our tool aims both at providing a summarisation of the included GDPR terms to users, and in helping software engineers keep track of GDPR functionality implemented in their web systems. We focus on web platforms, creating a list of GDPR terms to be included in their policies. We then apply an automated method including a crawler to locate privacy policy pages, and a parser to analyse the policy text from within the source code. We examine web platforms' privacy policies from 5 different sectors.

3 GDPR Privacy Policy Terms List

This section presents the procedure followed for creating a *list of GDPR privacy policy terms*. As a first step, we created an initial list based on the GDPR, extracting the specific provisions and user rights terms, that should be included within a web platform privacy policy. Next, we expanded our list based on our investigation and observation of existing web platforms privacy policies.

3.1 Analysing the GDPR

The first step of our methodology included studying the 99 articles and 173 recitals of the GDPR [4], as well as the related literature and legal documentation, to conclude to a list of rights and provisions that should be clearly stated

within a web platform privacy policy. Below we discuss the 7 provisions and rights, judged as relevant to web platforms based on our work of [13], in which we designed and developed the required GDPR functionality on a web platform real case study. We refer the reader to that work, for explanations on why the rest of the articles were not judged as relevant:

1. "*Lawfulness, fairness and transparency*", is one of the provisions defined in Article 5 of the regulation, as "*Personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject*". For a processing to be lawful, the system must establish one of the *lawful bases* defined in Article 6 "*Lawfulness of Processing*". In the context of a web platform, we consider the lawful bases of *consent* and *performance of a contract*, and that the platform must first ensure one of them, before proceeding with any processing of personal data. The selected *lawful basis* should be clearly presented within the privacy policy text.
2. The "*Right of Access by the Data Subject*" defined in GDPR Article 15, states that the user has the right to be informed whether their data are being processed, and to access a copy of their personal data stored and processed in the system, i.e., the web platform.
3. The "*Right to Rectification*" defined in GDPR Article 16, states that the user "*shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data*" and "*have the right to have incomplete personal data completed*". In web platforms, a user should be able to edit their personal data at any given time.
4. The "*Right to Erasure*", defined in Article 17 of the regulation, states that the users "*have the right to obtain from the controller the erasure of personal data without undue delay*", thus in a web platform managing user accounts, one should be able to delete a part or all of their information or their account.
5. The "*Right to Restriction of Processing*", of GDPR Article 18 states that users can ask for their personal data to stop being processed for an amount of time until they decide to resume the processing. In such a case, a web platform should stop processing but not delete the respective data.
6. The "*Right to Data Portability*" defined in GDPR Article 20, states that the users "*have the right to receive their personal data, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller*". Hence, a web platform user should be able to request and obtain a package including all the respective data, in such a format as requested by the regulation.
7. Finally, the "*Right to object*", defined in Article 21 states that the user has the right to object to the processing of their personal data. Like so, one's data should not be processed in case of an objection, until it is resolved.
   All the previously described rights of the users should be clearly stated in the privacy policy text.

As a result of the above study, we created an initial *list of GDPR privacy policy terms*, as presented in Table 1.

Table 1. Initial list of GDPR privacy policy terms

1	`Lawfulness of Processing`	5	`Right to Restriction of Processing`
2	`Right of Access by the Data Subject`	6	`Right to Data Portability`
3	`Right to Rectification`	7	`Right to Object`
4	`Right to Erasure`		

3.2 Examining Web Platforms

As a second step towards creating our *list of GDPR terms for privacy policies*, we examined a number of web platforms. We observed the structure of their privacy policies, and how the GDPR terms of our initial list are placed and included within these policies. We then observed differences and similarities between the terms expressing the same right in different platforms but also in comparison to the GDPR official terms as collected in the first step (see Sect. 3.1). A set of such different variations for each term was collected, extending the existing list. Examples of such variations, are shown below:

Example 1 (Terms Variations).
- "*Right to Erasure*": "The Right to Request Deletion", "Right To be Forgotten"
- "*Right to restriction of processing*": "Requests the Restriction of Their Use"
- "*Right to Rectification*": "Right to Correction", "Right to Correct", "The Right to Correct and Update"

Additionally, we infused the list with some more variations by interchanging the sequence of the words, or replacing basic words with similar ones, as shown in the following example:

Example 2 (Terms Variations).
- "*Right to Object to Processing*": "Processing Objection"
- "*Right Of Access*": "Access Personal Data"

3.3 The Final Set

We resulted in a set comprised of 89 terms. We then divided the terms into groups based on the 7 initial list terms collected from the GDPR during the first step. In this way, we consider that, if any of the terms of a group is found within a policy, the whole group is considered as included. The list includes a set of exact strings to be searched within privacy policies. This implies a limitation on the tool, as, if a GDPR term variation included in a policy is not exactly the same to any of the variations of the current list, it will not be recognised. However, the list is open for future editing and expansion. Improving and enhancing the list is envisioned as a continuous iterative process. After each iteration of evaluation of a set of platforms, a manual auditing is planned and terms that were not included, but shown to be needed, added to the list. In case the GDPR will be

updated, or new regulations need to be added, the list is flexible enough to be adapted to accommodate them. At the moment the list includes only English language GDPR terms. The final *list of GDPR privacy policy terms* in groups, is presented in Table 2 below.

Table 2. List of GDPR privacy policy terms

`Lawfulness of Processing (5 terms)`	`Right to Restriction of Processing`
Lawfulness of Processing	Restriction of Processing
Consent	Restrict Your data
Contract	Right to Restrict
Right to Withdraw Consent	Right to demand processing restrictions
Withdraw consent	Right to restriction of processing
	Request the Restriction of Their Use
	Request the Restriction of Data Use
	Request the Restriction of Personal Data Use
	Right of data subjects to be informed about the restriction
	Right to propose other restriction
`Right of Access by the Data Subject`	`Right to Data Portability`
Right Of Access	Right to Data Portability
Right To Access	Right of Portability
Access Personal Data	Right to Portability
Access Your Personal Data	Right to Transmit Those Data
Access your data	Right to Transmit Personal Data
Access Your Personal Information	Right to Transmit My Data
Access Personal Information	The Right to Transmit Those Data
Right to Lodge a Complaint	Right to Transmit Data
Right to complaint	Transmit Your Data
Right to File a Complaint	Transmit Your Personal Data
Right to Obtain a Copy	Right to Transmit Personal Data
Request a Copy of your Information	Request the transfer of your personal data
Request a Copy of your personal Information	Request the transfer your personal data
Request a Copy of your data	Request the transfer personal data
Request a Copy of your personal data	Request the transfer data
Request access to a copy of your personal data	Right to Receive the Personal Data
Right to Information	Right to Receive your Personal Data
Right to request and receive information	Right to Receive Personal Data

(*continued*)

Table 2. (*continued*)

Request access	Right to Receive a Subset of the Personal Data
	Right to receive a copy of your personal information
`Right to Rectification`	`Right to Object`
Right To Rectification	Right to Object
Right to have incomplete personal data	Right to Object at any time to Processing of Personal data
Right to complete incomplete personal data	Right to Object at any time to Processing
Right to Request Proper Rectification	Right to Object to Processing
Right to Request Rectification	Processing Objection
Rectify Your Data	Object to processing
The Right to Correct and Update	`Right to Erasure`
The Right to Correct	Right to Erasure
The Right to Update	Right of Erasure
Update or Correct your Information	Right to Request Deletion
Update your Information	Right To be Forgotten
Correct your Information	Erase your Information
Right to request the correction	Request erasure
Right to request correction	Erase the Personal data
Right to request update	Erase your Personal data
Right to Correction	To Erase Your Data
Right to Correct	Erase any personal data
Rectify	Erase personal data

4 The CompLicy Tool

In order to provide insights on whether privacy policies are aligned to the GDPR, by studying web platforms, we designed and developed the CompLicy[6] tool that uses the results of the previous steps of this work as the foundation to automatically locate and evaluate such platforms' privacy policies. First, the tool crawls a given web platform, starting from the given homepage URL, to infer whether a privacy policy page exists and, if it does, subsequently parses it, identifying GDPR terms and groups within based on our *list of GDPR privacy policy terms*, and finally, provides results for the compliance of the aforementioned policy with the GDPR, by displaying which terms out of the list were located within the policy, and which groups are thus represented.

Crawler. As a preliminary step for implementing the crawler, we manually examined the source code of a set of web platforms. We additionally examined

[6] "*CompLicy*" is a portmanteau, i.e., a made-up word, coined from the combination of the words "*Compli*ance" and "Po*licy*".

their front-ends. In both source codes and front-end interfaces, we observed: (1) how the navigation was done from each platform homepage towards the privacy policy page, (2) the key-words included in the respective privacy policy URLs, and (3) the different variations of names for the specific page in the menu or other interface parts of the web platforms. Based on these observations, we created a list of respective key phrases, that would help the crawler in locating and recognising a privacy policy page. In the case of examining the URLs, the key-phrases extracted were part of them. For example, in the case of a web platform of which the privacy policy page URL is *http://<an-example-page>/data-processing/*, the key-phrase extracted would be "*data-processing*". The final key-phrases set is the following: {*data-processing, data-privacy, privacy, policy, privacy-policy, legal, privacy-policy-link, privacy_check, guidelines*}.

4.1 System Design

Architecture. Figure 1 presents the architecture of the tool. The *User*, through the *graphical user interface (GUI)*, gives as input the URL of the web platform (*platform link)* to be examined, which is subsequently passed on to the *Crawler*. The web *Crawler* (or spider) is aimed to realise a systematic navigation into the website directed by the URL. In order to do so, it also receives as input the text file with the key-phrases, i.e., the *list of terms for crawling*. The *Crawler* then searches within the given website, to locate the privacy policy page, and if found, subsequently passing the *privacy policy page source code* to the *Parser* which in turn receives also as input the *list of GDPR privacy policy terms* in the format of a text file, searches and recognises the list terms within the privacy policy text and returns the *results* to the GUI. The GUI also inputs the list of GDPR terms as categorised in the groups, printing both the list and the results, and presenting them to the user. Figure 2 presents a more detailed version of the system architecture.

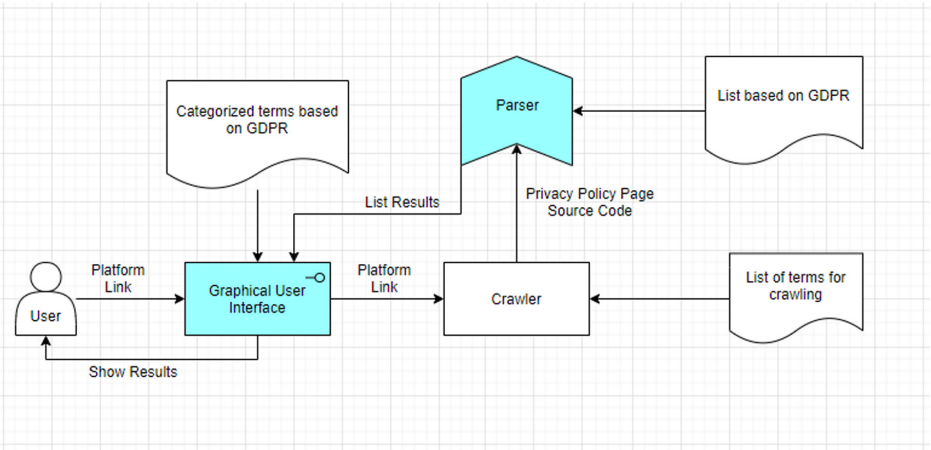

Fig. 1. Tool architecture

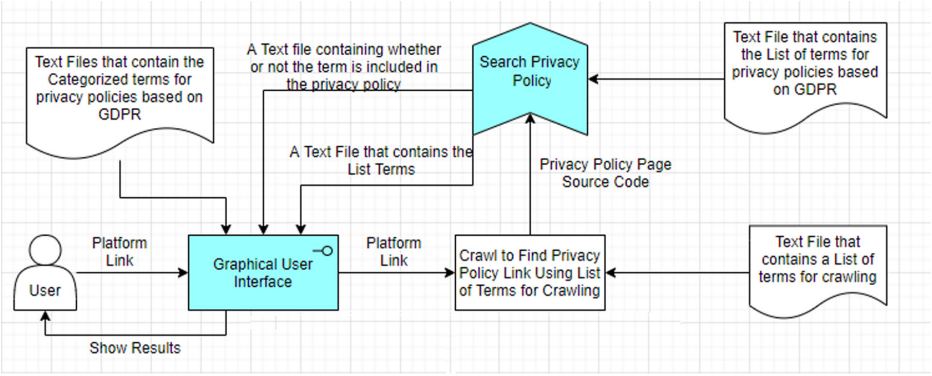

Fig. 2. Tool detailed architecture

4.2 System Implementation

Tools. The system was implemented using the Python programming language, in the PyCharm integrated development environment (IDE). The Qt Designer and the PyQt5 library were used for designing and implementing the GUI.

User Interface. Figure 3 presents the tool GUI. Users can provide a URL at the text box located at the top left side of the screen, and press the "*start*" button to initiate the crawling and parsing procedure, which when completed, the results will be presented in two tables as shown. The top table presents the complete *list of GDPR privacy policy terms*, accompanied with a "yes" or "no" answer, depending on whether the term was included within the policy or not, and the total number of terms located. The bottom table presents the 7 groups, accompanied by a mapping to the top list terms, and a 1-point scoring system for each group if at least one of its terms was included in the policy text. Finally, the total score of the privacy policy is presented in the format of points score and percentage.

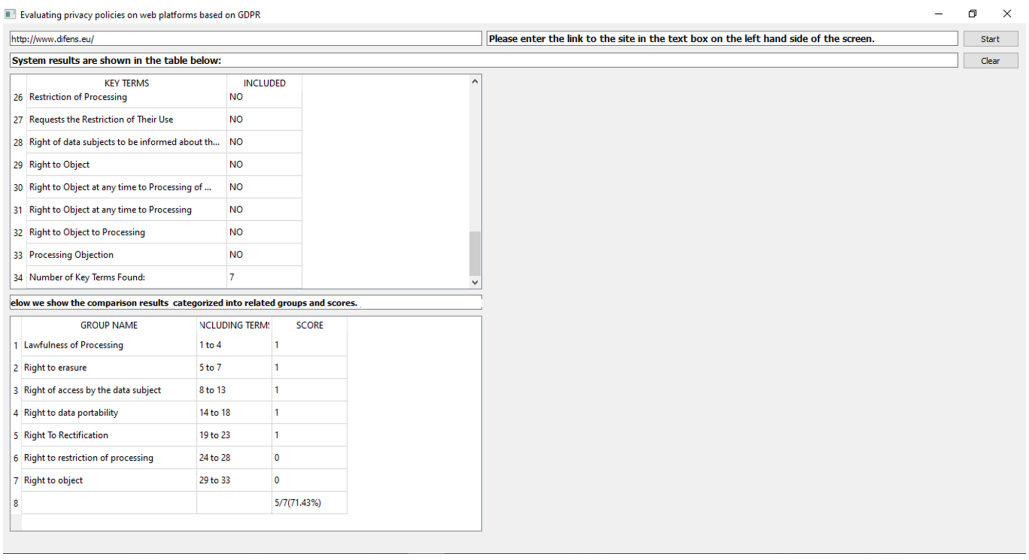

Fig. 3. Graphical user interface

5 Evaluating Compliance

For evaluating the compliance of web platforms, we first examined a preliminary set of websites, which we then manually audited to verify correctness of results. From the manual audit, we gathered a set of term variations for the rights and provisions of our list, that were not already included, but should have been. We then proceeded with adding the terms in the list, thus enhancing it. Next, we examined a second preliminary set of websites observing again by manual auditing the improvement in the results. Subsequently, we proceeded with the evaluation of the actual set of web platforms.

5.1 Dataset

A total of 148 websites were considered in the dataset, i.e., given as input to the tool, from five different sectors: (i) banking; (ii) e-commerce; (iii) education; (iv) travelling; and (v) social media. As described in the introduction (Sect. 1), the increasing usage of web platforms of these sectors, as well as the importance of the data processed and the actions carried out within, mandate for privacy protection. We have collected the URLs of the biggest and most popular websites of each respective sector. The procedure for selecting the sector-based websites was founded on three main stages. These included the following:

1. Stage 1 – Online review per sector. Desk research was conducted on the five different sectors.
2. Stage 2 – Define sector shortlist. Results from Stage 1 aided in identifying the more popular websites with larger user bases within each sector. An important requirement for the website to be eligible for selection was the existence of an English version of it. With the aim to approach the number of 20 websites being evaluated for each sector, the sector shortlist included a higher number of websites (n$\geq$ 20 websites per sector).
3. Stage 3 – Test sector shortlists. Once the shortlists were defined, each website on a sector-shortlist was evaluated in order of its popularity (as defined in Stage 2) with the CompLicy tool. In cases where a sector website could not be evaluated, e.g. for reasons such as not having an English version, privacy policies in PDF format, etc., the next website on the sector shortlist was evaluated.

As discussed, not all of these websites policies' could be successfully evaluated with the tool. The two main reasons for a website policy not being successfully evaluated were: 1) the privacy policy was in PDF; 2) the privacy policy was not written in English. Focusing on the successfully evaluated policies of the selected websites, in total these were 80 (mean GDPR compliance = 67.67, SD = 22.92, range = 14–100). Sector wise, the successful website policy evaluation rates were: (i) banking (57%); (ii) e-commerce (54%); (iii) education (59%); (iv) travel (52%); and (v) social media (50%). Figure 4 presents the number of successful (represented as "Successfully evaluated policy") and unsuccessful (represented as "Not evaluated policy") evaluated website policies for each sector, i) banking (n = 21); (ii) e-commerce (n = 39); (iii) education (n = 27); (iv) travel (n = 21); and (v) social media (n = 40).

Figure 5 presents the mean GDPR compliance score for the privacy policies of the successfully evaluated websites with the tool. Sector wise, the following scores were recorded: (i) banking (mean GDPR compliance = 75); (ii) e-commerce (mean GDPR compliance = 70.75); (iii) education (mean GDPR compliance = 52.68); (iv) travel (mean GDPR compliance = 72.73); and (v) social media (mean GDPR compliance = 69.29). It is not surprising that privacy policies of websites in the banking sector scored the highest in terms of GDPR compliance with 75%, yet one might expect an even higher compliance rate. Standing out however is the compliance score of websites in the education sector with 52.68%. This can be regarded as poor and does require further investigation, considering the large adoption of e-learning in 2020 and respectively large numbers of students using educational web environments.

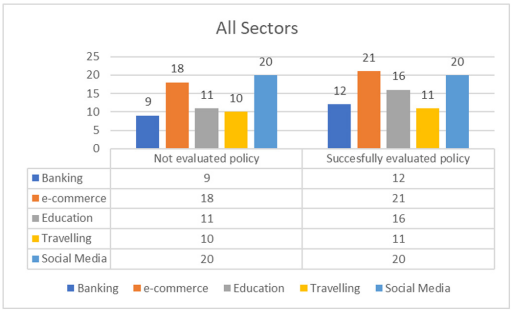

Fig. 4. Summarising successful and unsuccessful evaluation attempts according to sector websites'

Fig. 5. Summarising average score of GDPR compliance according to sector websites'

5.2 Results

For all the results following, we will be using the following abbreviations for the 7 groups: Lawfulness of Processing (LP), Right to erasure (RE), Right of access by the data subject (RA), Right to data portability (RDP), Right To Rectification (RR), Right to restriction of processing (RR2), Right to object (RO). Data and results are available at: http://www.cs.ucy.ac.cy/seit/resources/RCIS21.zip

Banking Sector. A total of 21 websites were given as input to the tool. From these, 12 websites policies' could be successfully evaluated. The GDPR compliance score for each of the 12 websites is presented in Fig. 6. In addition to the GDPR compliance score, also evident is the number of groups that each specific website's privacy policy includes from the 7 groups of the list of GDPR privacy policy terms. Only one website (i.e., no.3), had all 7 groups included, thus resulting in a GDPR compliance score of 100%. The majority (i.e., 7 websites), had 6 groups included. Figure 7 presents the 7 groups of GDPR provisions and rights included within the list, and for the 12 websites it is evident how many in total included each respective provision and right. RO and LP were the most included, appearing in 11 websites, whereas RR and RDP were the least included, appearing in 7 websites.

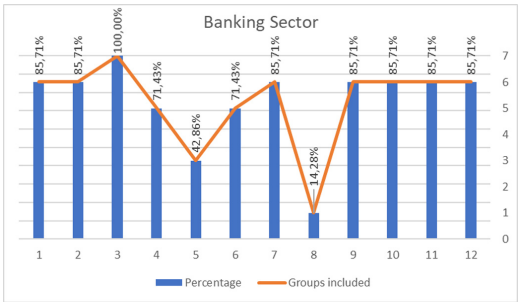

Fig. 6. Summarising average score of GDPR compliance and respective groups included in the privacy policy for successfully evaluated banking sector websites'

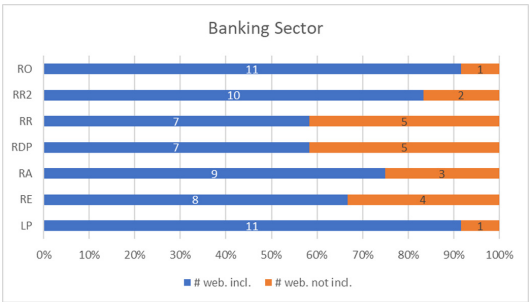

Fig. 7. Summarising the total inclusion of each group within the privacy policies of the successfully evaluated banking sector websites'

e-Commerce Sector. A total of 39 websites were given as input to the tool. From these, 21 websites policies' could be successfully evaluated. The GDPR compliance score for each of the 21 websites is presented in Fig. 8. In addition to the GDPR compliance score, also evident is the number of groups that each specific website's privacy policy includes from the 7 groups of the list of GDPR privacy policy terms. Three websites (i.e., no.14, no.17, no.21), had all 7 groups included, thus resulting in a GDPR compliance score of 100%. The majority (i.e., 6 websites), had 5 groups included. Figure 9 presents the 7 groups of GDPR provisions and rights included within the list, and for the 21 websites it is evident how many in total included each respective provision and right. Similarly to the banking sector websites, RO and LP were the most included, appearing in 19 and 21 websites respectively, whereas RR was the least included, appearing in only 6 websites.

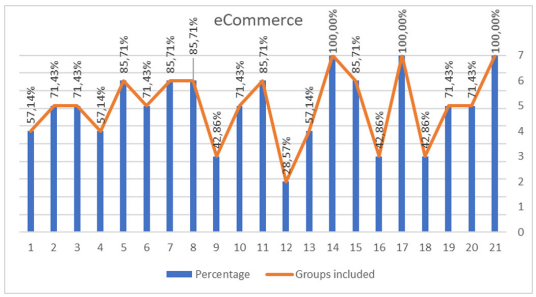

Fig. 8. Summarising average score of GDPR compliance and respective groups included in the privacy policy for successfully evaluated e-commerce sector websites'

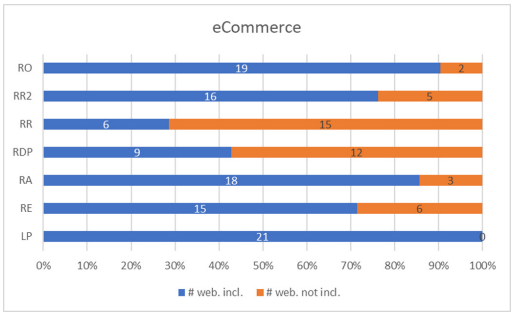

Fig. 9. Summarising the total inclusion of each group within the privacy policies of the successfully evaluated e-commerce sector websites'

Educational Sector. A total of 27 websites were given as input to the tool. From these, 16 websites policies' could be successfully evaluated. The GDPR compliance score for each of the 16 websites is presented in Fig. 10. In addition to the GDPR compliance score, also evident is the number of groups that each specific website's privacy policy includes from the 7 groups of the list of GDPR privacy policy terms. Only two websites (i.e., no.3, no.15), had all 7 groups included, thus resulting in a GDPR compliance score of 100%. The majority (i.e., 6 websites), had only 2 groups included, which is concerning, considering the large number of students sharing personal information on such websites. Figure 11 presents the 7 groups of GDPR provisions and rights included within the list, and for the 16 websites it is evident how many in total included each respective provision and right. RO and LP were again the most included, appearing in 14 and 16 websites respectively, whereas RR was also once again the least included, appearing in only 2 websites.

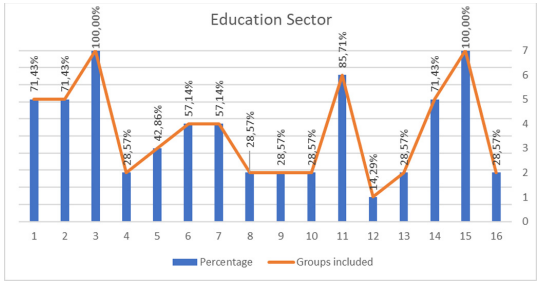

Fig. 10. Summarising average score of GDPR compliance and respective groups included in the privacy policy for successfully evaluated education sector websites'

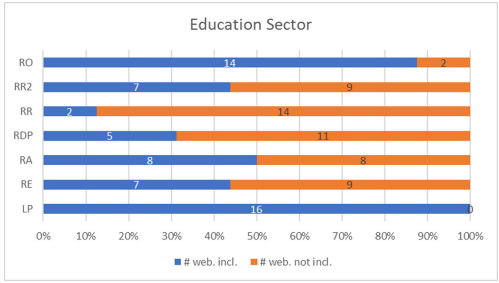

Fig. 11. Summarising the total inclusion of each group within the privacy policies of the successfully evaluated education sector websites'

Travelling Sector. A total of 21 websites were given as input to the tool, focused on booking flights, other travel activities, or accommodation. From these, 11 websites policies' could be successfully evaluated. The GDPR compliance score for each of the 11 websites is presented in Fig. 12. In addition to the GDPR compliance score, also evident is the number of groups that each specific website's privacy policy includes from the 7 groups of the list of GDPR privacy policy terms. No website had all 7 groups included, thus achieving a GDPR compliance score of 100% was not possible for a travel sector website. The majority however (i.e., 5 websites), had 6 groups included. Figure 13 presents the 7 groups of GDPR provisions and rights included within the list, and for the 11 websites it is evident how many in total included each respective provision and right. RO, LP and RR2 were the most included, appearing in 11 websites, whereas RR was also once again the least included, appearing in only 4 websites.

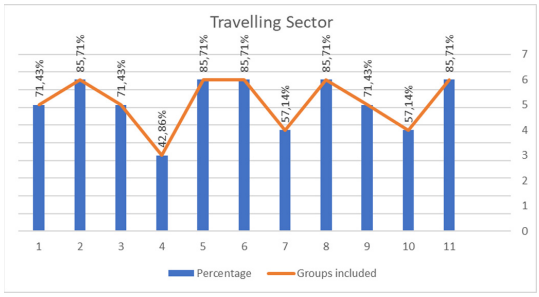

Fig. 12. Summarising average score of GDPR compliance and respective groups included in the privacy policy for successfully evaluated travelling sector websites'

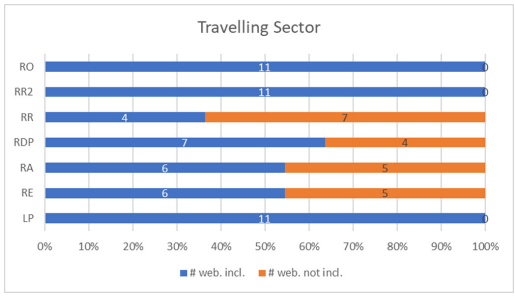

Fig. 13. Summarising the total inclusion of each group within the privacy policies of the successfully evaluated travelling sector websites'

Social Media Sector. A total of 40 websites were given as input to the tool. From these, 20 websites policies' could be successfully evaluated. The GDPR compliance score for each of the 20 websites is presented in Fig. 14. In addition to the GDPR compliance score, also evident is the number of groups that each specific website's privacy policy includes from the 7 groups of the list of GDPR privacy policy terms. Two websites had all 7 groups included, thus achieving a GDPR compliance score of 100%. The majority (i.e., 7 websites), had 4 groups included. Figure 15 presents the 7 groups of GDPR provisions and rights included within the list, and for the 20 websites it is evident how many in total included each respective provision and right. LP was the most included, appearing in 20 websites, whereas RDP was the least included, appearing in 5 websites.

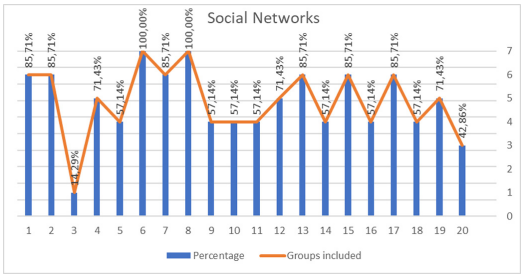

Fig. 14. Summarising average score of GDPR compliance and respective groups included in the privacy policy for successfully evaluated social media sector websites'

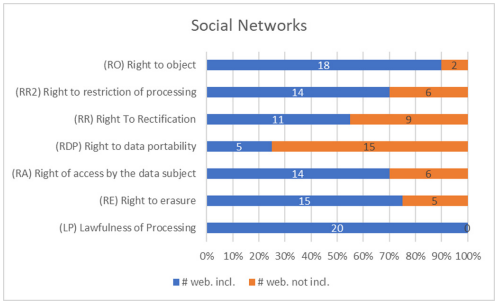

Fig. 15. Summarising the total inclusion of each group within the privacy policies of the successfully evaluated social media sector websites'

Websites in the banking sector scored the highest (75%), while websites in the education sector scored the lowest (52,68%). In all five sectors, "Lawfulness of Processing" was amongst the most included, appearing in all websites except in the educational and banking sectors appearing in 14 out of the 16, and 11 out of the 12 total websites evaluated respectively. On the contrary, the "Right

to Rectification" was the least included in all sectors, with the exception of the social media sector, where the "Right to Data Portability" was the least included.

6 Conclusions

Recognising the importance of *privacy*, and the mandatory nature of the GDPR, in this work we aim to give insight on whether software systems privacy policies have been aligned to comply with the GDPR. In this context we focus on web platforms, and present a list of GDPR terms that need to be included within privacy policies, and the CompLicy tool which locates a given web platform privacy policy and subsequently parses it, identifying GDPR terms and groups within, providing results for its compliance. Based on their increasing usage and popularity, and their importance, we focus on web platforms of the following sectors: (i) banking; (ii) e-commerce; (iii) education; (iv) travel; and (v) social media. We evaluate a set of 148 such platforms and present the results obtained.

We observe that websites in the banking sector scored the highest, while websites in the education sector the lowest. A more thorough investigation can be conducted focusing on the education sector, especially with the increased usage due to COVID-19. Future work should also investigate the reason for the least included provisions for each sector, as shown by the results ("Right to Rectification" in 4 sectors, "Right to Data Portability" in social media). We also observe that there is still a respectful percentage of policies not being fully aligned with the GDPR. Incorporating a privacy design step explicitly within the software engineering methodologies could assist developers. We are also planning the enhancement of our list of terms increasing the accuracy of our tool. As future work, we envision to enhance the tool with PDF scanning ability to capture more websites privacy policies. New languages can also be added by translating the existing terms in English.

References

1. Chang, C., Li, H., Zhang, Y., Du, S., Cao, H., Zhu, H.: Automated and personalized privacy policy extraction under GDPR consideration. In: Biagioni, E.S., Zheng, Y., Cheng, S. (eds.) WASA 2019. LNCS, vol. 11604, pp. 43–54. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23597-0_4
2. Contissa, G., et al.: CLAUDETTE meets GDPR: Automating the evaluation of privacy policies using artificial intelligence. SSRN 3208596 (2018)
3. European Parliament and Council of the European Union: Charter of fundamental rights of the European union. Official Journal of the European Union (2012)
4. European Parliament and Council of the European Union: General data protection regulation. Official Journal of the European Union (2015)
5. Hadar, I., et al.: Privacy by designers: software developers' privacy mindset. Empirical Softw. Eng. **23**(1), 259–289 (2018)
6. Krumay, B., Klar, J.: Readability of privacy policies. In: Singhal, A., Vaidya, J. (eds.) DBSec 2020. LNCS, vol. 12122, pp. 388–399. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-49669-2_22

7. Linden, T., Khandelwal, R., Harkous, H., Fawaz, K.: The privacy policy landscape after the GDPR. Priv. Enhanc. Technol. **2020**(1), 47–64 (2020)
8. McDonald, A.M., Reeder, R.W., Kelley, P.G., Cranor, L.F.: A comparative study of online privacy policies and formats. In: Goldberg, I., Atallah, M.J. (eds.) PETS 2009. LNCS, vol. 5672, pp. 37–55. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-03168-7_3
9. Renaud, K., Shepherd, L.A.: How to make privacy policies both GDPR-compliant and usable. In: International Conference on Cyber Situational Awareness, Data Analytics and Assessment, pp. 1–8. IEEE (2018)
10. Tesfay, W.B., Hofmann, P., Nakamura, T., Kiyomoto, S., Serna, J.: I read but don't agree: Privacy policy benchmarking using machine learning and the EU GDPR. In: The Web Conference, pp. 163–166 (2018)
11. Tesfay, W.B., Hofmann, P., Nakamura, T., Kiyomoto, S., Serna, J.: PrivacyGuide: towards an implementation of the EU GDPR on internet privacy policy evaluation. In: International Workshop on Security and Privacy Analytics. pp. 15–21 (2018)
12. Torre, D., Abualhaija, S., Sabetzadeh, M., Briand, L., Baetens, K., Goes, P., Forastier, S.: An AI-assisted approach for checking the completeness of privacy policies against GDPR. In: International Requirements Engineering Conference, pp. 136–146. IEEE (2020)
13. Vanezi, E., et al.: GDPR Compliance in the Design of the INFORM e-learning platform: a case study. In: International Conference on Research Challenges in Information Science, pp. 1–12. IEEE (2019)