



## Original software publication

## PADRES: Tool for PrivAcy, Data REgulation and Security

Fábio Pereira, Paul Crocker\*, Valderi R.Q. Leithardt

Instituto de Telecomunicações and Departamento de Informática, Universidade da Beira Interior, Covilhã, Portugal

VALORIZA, Research Center for Endogenous Resource Valorization, Polytechnic Institute of Portalegre, 7300-555 Portalegre, Portugal



## ARTICLE INFO

## Article history:

Received 27 July 2020

Received in revised form 10 August 2021

Accepted 2 November 2021

## Keywords:

GDPR

Privacy

Security

Vulnerabilities

## ABSTRACT

Since May 2018, companies have been required to comply with the General Data Protection Regulation (GDPR). The compliance process can be very expensive, for example, specialized human resources are needed who need to study the regulations and then implement any changes in company procedures, IT applications and infrastructures. With this in mind, PADRES a tool for PrivAcy, Data REgulation and Security was developed to analyse web applications and help in the compliance process. This open source software contains the main points of GDPR organized by principles in the form of a checklist and questionnaire. These questions are answered manually. Optionally a security analysis can also be performed, this is performed by integrating open source scanning tools such as NMAP, ZAP and cookie analyzers. The output of these tools is saved and a final merged report is generated with the information obtained and also a set of suggestions and recommendations.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Code metadata

Current code version

v2.0

Permanent link to code/repository used for this code version

<https://github.com/ElsevierSoftwareX/SOFTX-D-20-00011>

Legal Code License

LGPL v2.1

Code versioning system used

git

Software code languages, tools, and services used

Python, Angular, flask, Docker

Compilation requirements, operating environments &amp; dependencies

Linux, Docker

If available Link to developer documentation/manual

<https://github.com/FabioAndrePereira/PADRES/blob/master/README.md>

Support email for questions

[fabio.pereira@ubi.pt](mailto:fabio.pereira@ubi.pt)

## 1. Motivation and significance

General Data Protection Regulation (GDPR) compliance has become a priority for organizations and is still a critical challenge for businesses, especially financial and staffing resourcing [1]. Many businesses, especially smaller ones, are not prepared for the changes that have to be made and are unaware of the consequences that non compliance can bring. Studies have found that these problems happen because the actual regulations are

“vague, ambiguous and verbose”, meaning that anyone who does not have the legal and technical proficiency required can find understanding the regulations very difficult. For example, the GDPR states that companies must provide a reasonable level of protection for personal data [2], but the word “reasonable” is not well defined. Also “privacy by design” is promoted, without having a proper guide on how it can be achieved. There are two major problems that engineers and developers come across when trying to implement legal compliance [3]. The first is determining which regulations can be applied and the second is related to the ability to be able to develop the policies that enable compliance with those regulations, especially as extracting requirements from legal texts can be an error-prone job. The steep learning curve necessary to understand and comply with the GDPR also

\* Corresponding author at: Instituto de Telecomunicações and Departamento de Informática, Universidade da Beira Interior, Covilhã, Portugal.

E-mail addresses: [fabio@segal.ubi.pt](mailto:fabio@segal.ubi.pt) (Fábio Pereira), [crocker@di.ubi.pt](mailto:crocker@di.ubi.pt) (Paul Crocker), [valderi@ipportalegre.pt](mailto:valderi@ipportalegre.pt) (Valderi R.Q. Leithardt).

makes this a costly process, especially for SMEs that do not have a dedicated legal department or cannot afford legal advisory.

The PADRES (PrivAcY, Data REgulation and Security) software has been developed in order to help companies assess their GDPR compliance. The software is focused on analysing web applications. After running the software a GDPR classification is given and a final report generated that contains suggestion on what can be improved. The software contains a survey that a system administrator or developer must answer manually. The specific questions contained in the survey are constructed from a structured analysis of the GDPR. As there is no such thing as privacy without security, the tool that was developed also includes a procedure for searching for vulnerabilities. This is done by integrating a combination of open source scanning tools in order to test the platform against known risks mentioned in [4].

The PADRES's software fills a gap in the software engineering field for GDPR compliance software that consists of an extendable database of questions that are put to the user as well as a framework for running vulnerable scans and integrating their reports. Extendable means that the software can be easily improved on by the open source community, fine tuned towards specific country cases or by focusing on specific aspects of the legislation. The main target of analysis are web applications that consist of applications that follow a typical 3-Tier architecture, i.e a data tier accessed via an application tier or middleware that incorporates business logic accessed via a user interface/presentation tier.

PADRES was developed for the analysis of Research Infrastructures in particular the EPOS (European Plate Observatory System) infrastructure and the applications of the GNSS (Global Navigation Satellite System) community. However the software has a general scope of application and thus provides a general tool for GDPR compliance. Other layered architectures based on IOT devices such as [5] and more recent data management architectures based on Cloud and Edge Computing and 5G technologies such as [6] can also be analysed with our tool by an operator with appropriate access rights. An analysis of the cryptographic algorithms is advisable, however this is outside the scope of our tool and must be done manually consulting the relevant literature such as [7].

Given the need to be GDPR compliant several research articles and tools have been published to help and guide companies extract and identify legal requirements, analyse and achieve compliance. We briefly review here some of these.

There are existing solutions based on questionnaires and checklists such as [8]. Here several documents with guidance and checklists are made available. The checklist available in [9], splits the regulations into 4 categories, each one has the GDPR articles linked to a specific subcategory. The tool presented by ICO [10], is more intuitive than the previous ones as it is truly a checklist. The checklist is divided into 4 categories each with several questions and extra information regarding each point, thereby making it easier to answer the questions. Another GDPR assessment tool is [11]. Compared with the tools above [11] has more categories. It address topics such as the principles of processing personal data, rights of the data subject or data breaches. Also, inside each topic it clarifies exactly what is the point and its implications, complemented with a link to the corresponding GDPR article. Microsoft also has a set of tools [12]. Here a checklist to "simplify GDPR compliance efforts" for a compliance manager is given, it is possible to check the risk assessment on Microsoft cloud services and also to obtain recommendations with step-by-step guidance. All of these checklist based tools identify the most common regulation obligations, however precise techniques or mitigation's to be applied are not given. An example of a methodology that may be used to respond to questionnaires with mitigation techniques is a work described in [13]. This software library generates form-based web interfaces based on a set of inputs and rules. However

**Table 1**

List of paid Audit Software.

Software Names	URL
SolarWinds Access Rights Manager	<a href="http://www.solarwinds.com/access-rights-manager">www.solarwinds.com/access-rights-manager</a>
ManageEngine EventLog Analyzer	<a href="http://www.manageengine.com/products/eventlog">www.manageengine.com/products/eventlog</a>
LogicGate	<a href="http://www.manageengine.com/products/eventlog">www.manageengine.com/products/eventlog</a>
GDPR365	<a href="http://www.gdpr365.com">www.gdpr365.com</a>
Netwrix Auditor	<a href="http://www.netwrix.com/auditor.html">www.netwrix.com/auditor.html</a>
Really Simple Systems	<a href="http://www.reallysimplesystems.com">www.reallysimplesystems.com</a>
Vigilant Software GDPR Manager	<a href="http://www.vigilantsoftware.co.uk/topic/gdpr-manager">www.vigilantsoftware.co.uk/topic/gdpr-manager</a>
OneTrust	<a href="http://www.onetrust.com">www.onetrust.com</a>

the tool is not directly concerned with privacy issues and the developer still has to design the inputs.

Specific GDPR audit and analysis software exist in both free and open source form and via paid applications. Open source software's are mostly restricted to log file analysis, such as <https://goaccess.io/> a web server log analyzer that analyses and enables the visualization of web server statistics. In fact GDPR audit and analysis software consists mainly of paid applications. A recent list of such proprietary paid software can be found at <https://www.comparitech.com/net-admin/gdpr-compliance-software>. Here 8 software are briefly described (see Table 1).

### 1.1. PADRES development methodology

The most important aspect of the PADRES software is the extraction of pertinent questions for our GDPR questionnaire. To be able to this we first reviewed existing and relevant methodologies.

Christmann et al. [14] state that a major problem for small companies is to access expertise related with privacy and legal concepts. They propose a structured method for identifying IT security and legal requirements for cloud services depending on the functional and non-functional requirements.

Boella et al. in [15] reviews and analyzes different approaches to representing legal knowledge for legal requirements engineering. Existing mechanisms for extracting legal requirements are compared and then presented in a way that "industry experts" can use to make judgements concerning these requirements.

The solution presented by Gjermundrod et al. [16], gives a concrete solution for addressing the GDPR data processing requirements. Their solution consists of a privacy by design framework, based on 3 modules. In their solution they give technical information on how to collect the data, how to provide data tractability and also on how to share the data with other entities. More recently Tsohou et al. [17], discuss the requirements for a GDPR compliance platform. This paper reports on the DEFEND EU project <https://www.defendproject.eu> and its aims for constructing an overall platform for supporting GDPR (for instance functionalities that data controllers request for supporting GDPR such as monitoring GDPR requests and other supporting tools)

The recent paper by [18] reports on a formal method of analysis of the GDPR that aims to extract knowledge from the regulations. Here the GDPR is extracted to a concept lattice, featuring 144372 records that can be used for many purposes, including helping to preparing GDPR questionnaires and is available online.

It is also important to refer to the following "The nightmare letter: A subject access request under GDPR". [19]. This letter describes a worst case scenario in terms of GDPR information requests that a company can come across. It aims to investigate how a company would react to the scenario of a "Data Subject Access Request" made by someone with a wide knowledge in law and technologies that support data management.

**Table 2**  
GDPR principles [2].

1	Lawfulness, Fairness and Transparency
2	Purpose Limitation
3	Data Minimization
4	Accuracy
5	Storage Limitation
6	Integrity and Confidentiality
7	Accountability

After reviewing the previously mentioned articles it was concluded that the most appropriate framework for the regulation was to split it into GDPR principles, as seen in Table 2. Then for each principle a list with the most important points that need to be followed and investigated is drawn up, more details can be found in [20].

The points for each principle were then extracted manually. Each point has to be simple and concise. For example for the first principle the following points “Does the consent inform the Individual about the processing objectives ?” or Does your application provide any information regarding the Individual’s rights ? were defined. The points and principles are available in the database and can be accessed either through the application or through an endpoint of the API.

Finally to develop the suggestions that accompany the questions the same sources of information were also used, but this time converting them not into rules but into possible approaches to be implemented. Suggestions were written for the rules that can be more difficult to understand and implement. For example consider the rule *Do you have any mechanisms to pseudonymize data?* then if the user answers no, the report will contain the suggestion “pseudonymize of data is a data management technique, where the data controller swaps the individual’s direct identifiers, such as email or phone number, with a pseudonym..”.

## 2. PADRES software description

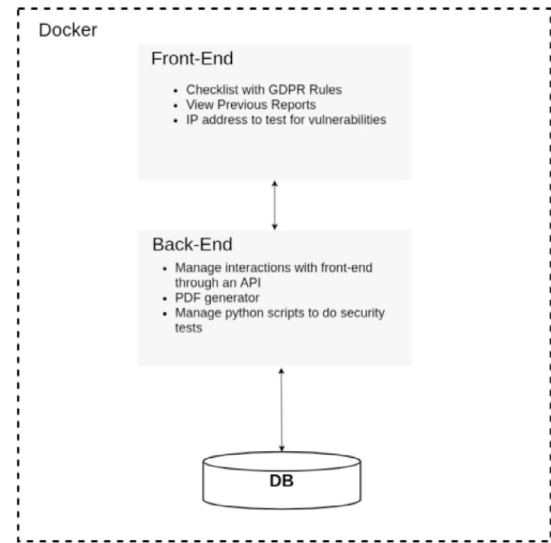
PADRES uses a client–server model. The front end web application was built using the Angular framework. The back-end uses a REST architecture and is built using the python Web framework. The back-end connects to a database that comes pre-filled with the GDPR questions that need to be answered and associated suggestions. Each question is associated with its own group, allowing a clear way to understand the connections between the questions and the GDPR and also making easier in the future to add new ones. Also provided is a structure to store the metadata of the final report and the results of executing software tools in BLOB format.

Currently the three Open source security assessment tools, Wapiti, ZAP and NMAP are installed. These are executed by the user through the graphical interface. The reports from each tool are merged together using these tools API’s with the report from the GPDR questionnaire.

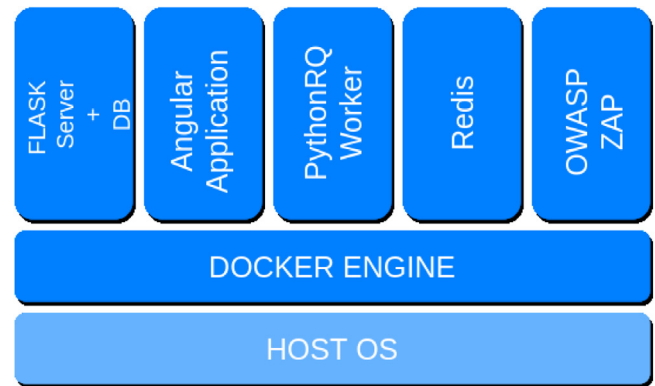
### 2.1. PADRES software architecture

Fig. 1 gives an overview of the PADRES architecture. This architecture makes it possible to have different clients interacting with the application, accessing the data provided by the database on the front end, through the REST API. Also, it is possible to add more rules to each principle and this allows PADRES to be easily extendable.

The security assessment tools may take a long time to execute. In order to execute these tools asynchronously a PythonRQ worker and a Redis database are used. This way scan requests



**Fig. 1.** PADRES architecture.



**Fig. 2.** Docker diagram of the collaborating containers.

are added to a queue that, when concluded, will show up as a notification on the front end application.

The entire solution is packaged using Docker. Fig. 2 illustrates the Docker layers and information about the technologies used in each container. Docker Compose was used for defining and running the multi-container application.

### 2.2. PADRES software functionalities

The major functionalities of the software are the GDPR questions that have been created, the integration of security assessment tools and the creation of a final report with comments and suggestions.

The GDPR questions are presented in the front end of the application. When submitted by the user, these are evaluated and depending on the answers a set of suggestions, for the questions that had a negative answer, are created. At the end of this process a partial report is generated

The second functionality, the security scans, is activated optionally by the user. The user selects which security tools are to be used and the target IP(s). The security scans are launched and when they finish the report from each one is merged with the partial report from the functionality above, thus producing one final report in PDF format. This is accessed on the front end, specifically on the history navigation tab, where it is also possible to see the previous reports.

Fig. 3. Example of GDPR questions.

### Summary of Alerts

Risk Level	Number of Alerts
<a href="#">High</a>	2
<a href="#">Medium</a>	5
<a href="#">Low</a>	16
<a href="#">Informational</a>	5

Fig. 4. ZAP scan result for <https://gnssproducts.epos.ubi.pt/>.

### 3. Illustrative examples

In this section we present results from applying PADRES on two applications. The first was the EPOS GNSS Products Portal, <https://gnssproducts.epos.ubi.pt/>, a web application dedicated to collecting, analysing and disseminating GNSS products such as time series and velocity maps. The GNSS Products Portal collects personal data due to its authentication and statistical inventory mechanisms. The GDPR thus applies directly on data collected from users logging in and downloading data and products and indirectly as personnel data is contained in some GNSS metadata. The PADRES tool was run by the EPOS GNSS IT staff responsible for maintaining the portal infrastructure and who understand the data flows used by this application. The front end application shows the questions to be answered as can be seen on Fig. 3

The PADRES tools user selects the security tools to be used and these are run in the background. The results are only available in the final report. One of the outputs is shown in Fig. 4 that illustrates the vulnerabilities found by the ZAP scanner.

The second case study was the site that provides services for the Portuguese GeoSciences community <https://intranet.c4g-pt.eu/>. This intranet gives its users the possibility to register and login and interact with the platform. Again, PADRES was executed by the IT staff responsible for maintaining the platform. It resulted in a non compliance of 15 points out of 31, one of the non-compliant points is shown in Fig. 5. From that is possible to say the possibly that, even though not using pseudonymization is not

a major issue, with the suggestion the user is more aware of the pseudonymization purpose and if necessary can implement it. The next point in the image regarding encryption was also answered as not in compliance.

Possibly the user answered this point incorrectly since portal access is only over HTTPS. The following item from the report was from the cookie scanner. This reported that only two cookies are being saved. One related with the PHP session and the second, the XSRF-TOKEN, to protect against CSRF attacks. The other security tools were also ran (NMAP, ZAP), the results from ZAP are omitted here but indicated some vulnerability alerts which were duly analysed by the staff.

### 4. Impact

The two case studies were used to not only validate the software concept but as an opportunity to fine tune the current questionnaire. However, the overall impact of the software is currently limited due to the size and scope of the questions and the need for more end user input and case studies. Extending the tool by adding more questions is easily done by adding more question directly to the database. This is a very important and great way of constructing a more comprehensive and complete set of compliance questions and can be a community based effort. Currently the authors will consider any pull request made using GitHub. The fact that the tool is easily deployed as it is built on docker and is open source are also important factors. The software can also be extended by adding further security assessment tools. In this case there is the need to study the tool to explore the best way to integrate such a tool either via a tools API (if it has one) or through its command line interface and output formats.

### 5. Conclusions

The PADRES tool is open source and designed to be extensible by adding new GDPR related questions and suggestions and by adding more open source tools for vulnerability and cookie analysis. This way the tool can be easily extended to become a simple but important tool that helps companies, IT infrastructure managers and developers analyse their applications in light of the GDPR.



- Do you have any mechanisms to pseudonymize data?
- Suggestions to be in compliance
- Is a data management technique, where the data controller swaps the individual's direct identifiers, such as email or phone number, with a pseudonym. Then the data processor can process the data without exposing the sensitive data. Then when data goes back to the data controller, he can rebuild the original data through re-identification techniques
- Does your application use encryption?
- No suggestions available

Fig. 5. C4G GDPR points not in compliance.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by national funds through the Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) by the project UIDB/05064/2020 (VALORIZA – Research Centre for Endogenous Resource Valorization), by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/EEA/50008/2020 and also by the EPOS-IP European Union Horizon 2020 research and innovation program under grant agreement No 676564.

## References

- [1] Breitbarth P. The impact of GDPR one year on. *Netw Secur* 2019;2019(7):11–3. [http://dx.doi.org/10.1016/S1353-4858\(19\)30084-4](http://dx.doi.org/10.1016/S1353-4858(19)30084-4).
- [2] Ayala-Rivera V, Pasquale L. The grace period has ended: An approach to operationalize GDPR requirements. In: 2018 IEEE 26th international requirements engineering conference. 2018, p. 136–46. <http://dx.doi.org/10.1109/RE.2018.00023>.
- [3] Otto PN, Anton AI. Addressing legal requirements in requirements engineering. In: 15th IEEE international requirements engineering conference. 2007, p. 5–14. <http://dx.doi.org/10.1109/RE.2007.65>.
- [4] OWASP Top 10 Application Security Risks – 2017 [online]. Available at: <https://www.encurtador.com.br/amqAV>. [Accessed 14 July 2021].
- [5] Tewari A, Gupta B. Security, privacy and trust of different layers in Internet-of-Things (IoTs) framework. *Future Gener Comput Syst* 2020;108:909–20. <http://dx.doi.org/10.1016/j.future.2018.04.027>.
- [6] Stergiou CL, Psannis KE, Gupta BB. IoT-based big data secure management in the fog over a 6G wireless network. *IEEE Internet Things J* 2021;8(7):5164–71. <http://dx.doi.org/10.1109/JIOT.2020.3033131>.
- [7] Saraiva D, Leithardt V, de Paula D, Sales MA, González G, Crocker P. PRISEC: Comparison of symmetric key algorithms for IoT devices. 2019;19(19):4312. <http://dx.doi.org/10.3390/s19194312>.
- [8] Preparing your organisation for the general data protection regulation - your readiness checklist [online]. Available at <https://www.dataprotection.ie/sites/default/files/uploads/2019-04/A-Guide-to-help-SMEs-Prepare-for-the-GDPR.pdf>. [Accessed 14 July 2021].
- [9] GDPR checklist for data controllers [online]. Available at <https://gdpr.eu/checklist/>. [Accessed 13 July 2021].
- [10] Controllers checklist [online]. Available at <https://ico.org.uk/for-organisations/data-protection-self-assessment/controllers-checklist/>. [Accessed 4 August 2021].
- [11] EU GDPR Readiness Assessment Tool [online]. Available at: <https://advisera.com/eugdpracademy/eu-gdpr-readiness-assessment-tool/>. [Accessed 14 July 2021].
- [12] How Microsoft tools and partners support GDPR compliance [online]. Available <https://www.microsoft.com/security/blog/2017/12/19/how-microsoft-tools-and-partners-support-gdpr-compliance/>. [Accessed 12 July 2021].
- [13] Galizia A, Zereik G, Roverelli L, Danovaro E, Clematis A, D'Agostino D. Jsn-GUI—A module for the dynamic generation of form-based web interfaces. *SoftwareX* 2019;9:28–34. <http://dx.doi.org/10.1016/j.softx.2018.11.007>.
- [14] Christmann C, Falkner J, Horch A, Kett H. Identification of IT security and legal requirements regarding cloud services. In: IEEE CLOUD 2015.
- [15] Boella G, Humphreys L, Muthuri R, Rossi P, van der Torre L. A critical analysis of legal requirements engineering from the perspective of legal practice. In: 2014 IEEE 7th international workshop on requirements engineering and law. 2014, p. 14–21. <http://dx.doi.org/10.1109/RELAW.2014.6893476>.
- [16] Gjermundrød H, Dionysiou I, Costa K. Privacytracker: A privacy-by-design GDPR-compliant framework with verifiable data traceability controls. In: Current trends in web engineering - ICWE 2016. Lecture notes in computer science, vol. 9881 LNCS, Springer Verlag; 2016, p. 3–15. [http://dx.doi.org/10.1007/978-3-319-46963-8\\_1](http://dx.doi.org/10.1007/978-3-319-46963-8_1).
- [17] Tsohou A, Magkos E, Mouratidis H, Chrysoloras G, Piras L, Pavlidis M, Debussche J, Rotoloni M, Crespo BG-N. Privacy, security, legal and technology acceptance elicited and consolidated requirements for a GDPR compliance platform. *Inf Comput Secur* 2020. <http://dx.doi.org/10.1108/ICS-01-2020-0002>.
- [18] Tamburri DA. Design principles for the General Data Protection Regulation (GDPR): A formal concept analysis and its evaluation. *Inf Syst* 2020;91:101469. <http://dx.doi.org/10.1016/j.is.2019.101469>.
- [19] Karbaliotis C. The nightmare letter: A subject access request under GDPR [online]. Available at: [encurtador.com.br/cqx3D3](https://www.encurtador.com.br/cqx3D3) [Accessed 15 July 2021].
- [20] de Sousa Pereira FA. EPOS security & GDPR compliance. Portugal: University of Beira Interior; 2020, doi:10.4006/10811.