Security Analysis of Subject Access Request Procedures How to authenticate data subjects safely when they request for their data

Coline Boniface, Imane Fouad, Nataliia Bielova, Cédric Lauradoux, Cristiana Santos

Security Analysis of Subject Access Request Procedures

How to authenticate data subjects safely when they request for their data

Coline Boniface[2], Imane Fouad[1], Nataliia Bielova[1], Cédric Lauradoux[2], and Cristiana Santos[3]

[1] Univ. Grenoble Alpes, Inria, France
{cedric.lauradoux,coline.boniface}@inria.fr
[2] Université Côte d'Azur, Inria, France
{nataliia.bielova,imane.fouad}@inria.fr
[3] School of Law, University Toulouse 1 Capitole, SIRIUS Chair
cristiana.santos@ut-capitole.fr

Abstract. With the GDPR in force in the EU since May 2018, companies and administrations need to be vigilant about the personal data they process. The new regulation defines rights for data subjects and obligations for data controllers but it is unclear how subjects and controllers interact concretely. This paper tries to answer two critical questions: is it safe for a data subject to exercise the right of access of her own data? When does a data controller have enough information to authenticate a data subject? To answer these questions, we have analyzed recommendations of Data Protection Authorities and authentication practices implemented in popular websites and third-party tracking services. We observed that some data controllers use unsafe or doubtful procedures to authenticate data subjects. The most common flaw is the use of authentication based on a copy of the subject's national identity card transmitted over an insecure channel. We define how a data controller should react to a subject's request to determine the appropriate procedures to identify the subject and her data. We provide compliance guidelines on data access response procedures.

Keywords: GDPR, data protection, privacy, right of access, identity verification, subject access request (SAR)

1 Introduction

With the GDPR in place since May 2018, the rights of the European users have been strengthened. The GDPR defines users' rights and aims at protecting their personal data. Every European Data Protection Authority (DPA) provides advices, explanations and recommendations on the use of these rights. However, the GDPR does not provide any prescriptive requirements on how to authenticate a data subject request. This lack of concrete description undermines the practical effect of the GDPR: it hampers the way to exercise the subject access right, to check the lawfulness of the processing and to enforce the derived legal rights therefrom (erasure, rectification, restriction, etc).

Every data subject would like to benefit from the rights specified in GDPR, but still wonders: *How do I exercise my access right? How do I prove my identity to the controller?* These questions are critical to build trust between the data subject and the controller. The data subject is concerned with threats like *impersonation* and *abusive identity check*. Impersonation is the case of a malicious party who attempts to abuse the subject access request (SAR) by impersonating a subject to a controller. Abusive identity check occurs when a data controller is too curious and verifies the identity of a subject by asking irrelevant and unnecessary information like an electricity bill or government issued documents.

Symmetrically, every data controller needs to know how to proceed when they receive an access request: *Is the request legitimate? What is necessary to identify the subject's data?* These concerns aggravate when controllers deal with indirectly-linked identifiers, such as IP addresses, or when they have no prior contact with data subjects, as in *Google Spain*[4]. Most of all, data controllers want to avoid data breaches, as it can

[4] Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Case C-131/12, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62012CJ0131&from=EN.

result in legal proceedings and heavy fines. Such consequence occurs in two cases: *(i)* the data controller releases data to an illegitimate subject, or *(ii)* he releases data of a subject A to a legitimate subject B.

All these questions concern the authentication procedure between the data subject and the controller. They both share a common interest in holding a strong authentication procedure to prevent impersonation and data breaches. The subject must be careful during the authentication procedure, as for providing too much personal information could compromise her right of privacy. Additionally, the controller needs to ask the appropriate information to identify the subject's data without ambiguity. There is clearly a tension during this authentication act between the controller, who tries to get as much information as possible, and the data subject who wants to provide as little as possible. Plausibly, subject access rights can probably increase the incidence of personal records being accidentally or deliberately opened to unauthorised third parties [57].

This paper studies *the tension during the authentication between the data subject and the data controller*. We first evaluate the threats to the SAR authentication procedure and then we analyze the recommendations of 28 DPAs of European Union countries. We observe that four of them can potentially lead to abusive identity check. On the positive side, six of them are recommending to enforce the data minimization principle during authentication. This principle, on one hand, protects the right to privacy of data subjects, and on the other hand prevents data controllers to massively collect personal data that is not needed for authentication, thus preventing abusive identity check (Section 2).

We have then evaluated the authentication procedure when exercising the access right of the 50 most popular websites and 30 third-party tracking services (Section 3). Several popular websites require to systematically provide a national identity card or government-issued documents to authenticate the data subject. Among third-party tracking services, 9 of them additionally to cookies demand other personal data from the data subjects, like the identity card or the full name. We explain that such demands are not justified because additional information can not prove the ownership of the cookie.

We then provide guidelines to Data Protection Authorities, website owners and third party services on how to authenticate data subjects safely while protecting their identities, and without requesting additional unnecessary information (complying with the data minimization principle). More precisely, we explain how data controllers and data subjects must interact and how digital identifiers can be redesigned to be compliant with the GDPR (Section 4). Finally, we overview related work (Section 5) and then conclude the paper.

2 Threats to SAR authentication and Recommendations of the DPAs

Chapter 3 of the GDPR [58] is dedicated to the rights of the data subject: right to access, object, rectification, erasure, restriction of processing, notification, portability and the right not to be subject to a decision based solely on automated processing. Some of these rights are not new and they already appeared in the Directive 95/46/EC, like the right of access. The right of access by the data subject is defined in Article 15, but Article 12 and Recital 64 of the GDPR are also important for our work, as these provisions regard the *modalities for the exercise of the rights of the data subject*. Three key elements can be extracted from this article. First, a data controller must answer each data subject request without undue delay. Second, the identity of the data subject making a request needs to be proven. Third, the data controller should also provide means for requests to be made electronically, where appropriate.

Access requests can be *direct*, *indirect* or *mixed*. The "normal way" for subjects to access their data is directly: the subject sends a request to the data controller and no third parties are involved. As for indirect requests, the DPA or a court can be involved only if the data controller does not respect the subject's rights, upon a complaint procedure. For certain special files, data subjects can not directly exercise their rights. The national DPA, acting as a proxy controller, has to verify the subject's identity and make the request on his behalf; then, the DPA reports its finding to the subject. Mixed access requests corresponds to situations in which some accesses are direct and other are indirect. A good example is the Schengen information system. Depending on the country, the access to this file can be direct, indirect or mixed, as defined in [27]. Our work is dedicated to direct accesses, but it also applies to the case of indirect and mixed access.

These elements are important to understand how the access right is exercised in practice, but these are still insufficient to let the subject and the data controller understand what is at stake. From now, we focus our attention on the second point: how can the data controller check the identity of the subject making a request. We first consider the issues that can occur (section 2.1), and then we examine what are the recommendations of the European Data Protection Authorities (section 2.2).

2.1 Threat model

In the last decades, researchers have made a substantial advancement in the field of authentication by means of cryptographic protocols. These protocols are often run automatically by computers and they are (almost) transparent to end users. They are the straightforward solution for a data controller to authenticate a data subject. However, this is true only if the data controller has created automatic tools for the subject to extract her data. But in practice it is often the case that the access request of the subject is handled by a human (often a data privacy officer of the data controller). All the research advancement on authentication is suddenly irrelevant because a human can not execute complex cryptographic operations.

Therefore, we question what are the consequences of weak authentication procedures. The main purpose of the authentication is to establish the identity of the data subject to the data controller. This goal is explicitly stated in the GDPR, however the GDPR does not explain the major consequence of an incorrect authentication, which we devise in this analysis. In our paper, we have considered that both the data subject and the data controller can be malicious. In our definition of the threat model, we take the perspective of the data subject making the request. In our quadrant analysis, three issues can occur: *data breach, privacy invasion and denial of access.*

(i) Data breach – A data controller discloses information of a data subject to someone else than the concerned subject. Any data controller wants to avoid this situation which can result in being fined by one of the EU DPAs. The data subject is also interested in protecting herself from such breaches and from her private data being exposed. The data can be exposed to an external adversary or to another different legitimate subject. Unauthorized disclosures are qualified as data breach, under Article 3(12) of the GDPR.

(ii) Privacy invasion – In this situation, the data controller is perceived as malicious. He aims to exploit the authentication as a method to obtain from the data subject. This can be viewed as a sort of data breach made by the data controller himself whose goal is to access more data of the data subject. The qualification of *privacy invasion* derives from our interpretation of non-compliance to the principles of data minimization and storage limitation:

- *Minimization principle: personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed* (Article 5 (1) (c). Recital 39 specifies further that *personal data should be processed only if the purpose of the processing could not reasonably be fulfilled by other means.* The "necessity" or "proportionality" requirement that both these provisions afford, refers both to the quantity and also to the quality of personal data. It is thus clear that the controller should not process excessive data if this entails a disproportionate interference in the data subject's rights, and hence, a privacy invasion. Ultimately, if the personal data processed by a controller does not permit him to identify a user, the data controller should not be obliged to acquire additional information in order to identify the data subject for the sole purpose of complying with SAR, in accordance to Recital 57;

- *Storage Limitation Principle: personal data shall be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed*, (Article 5(1)(e)). Aligned with this principle, recital 64 further refers that a controller should not retain personal data for the sole purpose of being able to react to potential future requests. Recital 39 invites controllers to establish time limits for erasure or for a periodic review. This will ensure that the personal data are not kept longer than necessary.

(iii) Denial of access – This situation is often mentioned in the papers [43, 39, 44] testing the access right. The data controller refuses to allow a legitimate subject to access her data. The reasons can be numerous. We focus in our work on cases, where authentication is used as a mean to refuse the access to the data.

After having identified the purposes of authentication, we propose to define the threats, *i.e.* the attacks that can be carried out by an adversary. Any successful attack results in a significant privacy issue, either for the data subject making the request, or for another data subject. Instances of these threats are: impersonation, incorrect disclosure, abusive identity check and impossibility of authentication.

(i) Impersonation (data breach) – A malicious individual is able to impersonate a legitimate data subject to the data controller. The adversary forges a valid access request and goes through the identity verification enforced by the data controller. The data controller sends to the adversary the data of a legitimate data subject. Defeating impersonation is the primary objective of any authentication protocol. The result of this attack is a data breach (e.g. blaggers pretend to be someone they are not in order to wheedle out the information they are seeking obtaining information illegally which they then sell for a specified price).

(ii) Incorrect disclosure (data breach) – A data subject makes a legitimate request to a data controller to access her data. The data controller verifies successfully her identity and sends the data back to the data subject. However, some of the sent data belongs to another data subject. This is clearly an incorrect procedure made by the data controller. This error will be sooner or later exploited by an adversary who will create an account at the data controller and send a legitimate request to access the data of someone else. This is clearly a data breach. It is very easy to imagine an incorrect disclosure. Let us consider the case of a subject using IP address `XXX.WWW.YYY.ZZZ`. This address is actually shared by several subjects in Virtual Private Network (VPN). The subject asks a data controller for all the data collected and associated to `XXX.WWW.YYY.ZZZ`. If a controller sends data associated to `WWW.XXX.YYY.ZZZ`, he might commit an incorrect disclosure.

(iii) Abusive identity check (privacy invasion) – The adversary in this case is the data controller itself. The term *abusive identity check* is associated with discriminatory controls by law enforcement authorities, but we use it in a different meaning. We consider that the identity verification is abusive when the data controller asks unnecessary or irrelevant information. Let us consider a case of a subject who has registered to a service using a pseudonym. The data controller of the service has no clue on the real identity of the subject. Despite using a pseudonym, the GDPR still applies and the subject can request access to her data to the controller. The controller requires a copy of her passport to verify that the request is legitimate. We contend that this verification is abusive for two reasons:

- the information is irrelevant because getting a copy of her passport does not help the controller to check that the request is legitimate; and
- there is no reason for the data subject to reveal her real identity to the data controller through such document. The documents requested by the controller must be *proportional* or *necessary* to the controller's knowledge of the data subject. Can we state that each time a data controller asks for a copy of her passport we are dealing with abusive identity checks? No, it depends on what the data subject has already revealed to the data controller. If the data controller knows the true identity of the data subject, it is legitimate to ask for an official document. It can be the case, for instance, if the data controller is a national administration. However, as we will see in Section 3, some data controllers require extra information to authenticate data subjects (and thus perform abusive identity check), claiming they follow Article 12(6) of GDPR saying "*where the controller has reasonable doubts concerning the identity of the natural person making the request, the controller may request the provision of additional information necessary to confirm the identity of the data subject*" even though the identity check is already established.

(iv) Impossibility of authentication (denial of access) – Upon receiving an access request, the data controller can declare that he is not in a position to identify the data subject, due to difficulties to prove ownership of

the data. He cannot satisfy the condition of Article 12 of the GDPR and will not grant access to the data to prevent a data breach. Hence, the controller shall inform the data subject accordingly, if possible (Article 11 (2)), providing the reasons for his non fulfillment of a specific access request.

We excluded from our study the more generic threat of denial of service (DoS) attacks. An example of DoS attack in our settings would consist in a huge number of malicious data subjects sending access requests to the data controller at the same time of the request of the targeted subject. Being overflown by malicious requests, the data controller cannot answer the request of the targeted subject. DoS attack in this case results in a privacy issue: the data subject is deprived from her rights.

Summary: Our threat model shows that there is a *tension between privacy and security* during the authentication procedure when the data subject wants to exercise her access rights. First, a zealous data controller can ask too much information from the data subject to ensure her identity. Second, a zealous data controller can reject all the subject access requests claiming that he cannot authenticate the data subject. Finally, a negligent data controller may obtain insufficient information to prevent impersonation or incorrect disclosure.

2.2 Recommendations of the EU Data Protection Authorities

The goal of our work is to analyze the recommendations emanated by the DPAs of the members of European Union regarding authentication procedures when data subjects exercise the right to access their data. In particular, we aim to determine if DPAs provide recommendations on how a data subject can exercise her rights and if any authentication process was mentioned or suggested. In this work, we take the perspective of data subjects visiting a website of a DPA searching for recommendations to exercise their subject access requests. We acknowledge that DPA recommendations can be interpreted to be either addressed for data subjects to exercise SAR with the DPAs directly, or it can also configure a recommendation for the DPOs of the data controllers. However, since our study is data subject-centric, we analyse the issued recommendations from the perspective of the data subject who is trying to follow procedures to exercise her rights.

To achieve this goal, we adopted the following methodological steps: we have visited the webpages of French, English, Italian and Spanish speaking countries. Regarding the other countries, we have asked the assistance of colleagues who speak the language of the given country: each colleague was provided with the website of a DPA and was asked to find pages related to the exercise of access right in 30 minutes. We received answers from all members states. The results of our inspections can be found in Table 1.

From our analysis, we report that all DPAs explain what are subject rights. However not all the authorities explain how the data subject can exercise her rights. 17 authorities provide guidelines and explanations for the subject to access her data. Several of them provide also a template for the subject to make her request via email or post. It is noticeable that we have not found any authority providing guidelines or recommendations for data controllers on how to authenticate a data subject and to let her exercise her rights. It follows from the foregoing analysis that, to the best of our knowledge, the Bulgarian DPA does not provide any information on how to fill in a request. However, it has an interesting page [55] entitled "Who can copy your identity card" whose goal is to warn the subjects that the copy of an ID card is a sensitive document. This document also states that a data controller is legitimate to ask a copy of a subject's ID to authenticate her access request.

From the websites of 28 European DPAs that we have analyzed, we found that four DPAs (Belgium, Bulgaria, Ireland and Spain) require a copy of government-issued documents to make an access request by default. Such recommendations can lead to abusive identity checks. The German authority [66] suggests to use the copy of an ID card, but it strongly recommended to blur unnecessary information.

Some recommendations made by the authorities are particularly interesting. The Austrian DPA [45] does not provide any specific recommendations, but gives a template form for a data subject to make a subject access request. In the part "Identity" field in the form [45], the Austrian DPA lets the data subject choose how she should be identified: (i) if the subject already had a contact with the data controller, then customer number would suffice, (ii) otherwise, the subject should attach an ID proving her identity. The Austrian

Country	Recom.	Authentication	Country	Recom.	Authentication
Austria [45]	✓	Customer ID or copy of the national identity card	Italy [84]	✗	Data minimization
Belgium [47]	✓	Copy of the national identity card	Latvia [64]	✓	Data minimization
Bulgaria [55]	✗	Copy of the national identity card	Lithuania	✗	
Croatia [61]	✗		Luxembourg [76]	✓	
Cyprus [56]	✗		Malta [81]	✓	
Czech Republic [102]	✗		Netherlands [48]	✓	Least privacy sensitive
Denmark [65]	✓		Poland [98]	✗	
Estonia [42]	✗		Portugal [54]	✗	
Finland [82]	✓		Romania [101]	✓	
France [53]	✓	Proportionality	Slovakia [80]	✓	
Germany [66]	✓	Copy of the national identity card + masking	Slovenia [69]	✓	Relevant Identifying data
Greece [68]	✗		Spain [40]	✓	Copy of the national identity card
Hungary [63]	✗		Sweden [79]	✓	
Ireland [62]	✓	Copy of the national identity card	UK [99]	✓	Any information used by the organisation to identify or distinguish you

Table 1. State of recommendations by the Data Protection Authorities of the European Union.

DPA lets the subject choose whether she wants to be contacted electronically or not, but does not provide any security conditions for the data transmission.

In France, the CNIL [53] advises to apply the proportionality principle when sharing information to authenticate to the controller. The Italian DPA [84] does not provide any specific indications on how to authenticate a data subject, but requires, like the Latvian DPA [64] that the data minimization principle is respected.

The Slovenian DPA [69] suggests to provide "*birthday or other identification data on the basis of which the manager can find in your collections your personal information you request* ". The subject needs to provide only information necessary for the controller to find her data. The recommendation of the ICO in United Kingdom [99] is very similar. The subject must provide "*Any information used by the organisation to identify or distinguish you*"

Data protection authorities are the main enforcers of the GDPR. Moreover, two additional actors at the EU level are involved in the implementation of the GDPR: the European Data Protection Supervisor (EDPS) [31] and the European Data Protection Board (EDPB) [30]. The EDPS supervises the EU institutions to help them to be exemplary. The EDPB advises the European Commission on any issue related to data protection in the EU and to rule by binding decisions on disputes regarding cross-border processing activities, ensuring therefore a uniform application of the EU rules. The EDPB is also a data controller and provides a privacy notice at `https://edpb.europa.eu/about-edpb/legal-notices/data-protection-notice_en`. Data subjects can exercise their access rights by sending an email to EDPB-DPO@edpb.europa.eu or contacting the EDPB DPO by post in a sealed envelope. It also states that:

> *Your request should contain a detailed, accurate description of the data you want access to. When there are reasonable doubts regarding your identity, you might be asked to provide a copy of a docu-*

ment, which help us to verify your identity. It can be any document such as your ID card or passport. Should you provide any other documents, personal details such as your name and your address should be in clear in order to be able to identify you, while any other data such as a photo or any personal characteristics, may be blacked out.

3 Practical Evaluation of Websites and Third Party Trackers

When a subject visits a website, data may be collected by the website owner and by third party trackers present on the website. In our work, we consider the website owner and the third party tracker present on the website as joint data controllers. Both joint controllers could distribute their responsibilities concerning SAR, and hence, data subjects could exercise their rights of access against each of the controllers. Irrespective of the contractual provision allocation tasks between joint controllers, they are both liable for non-compliance of subject access rights (Article 26 (3)).

In a practical setting, and following the cognition of Mahieu et al. [72] *"a data subject can direct a request to access to the website administrator, irrespective of the fact that the personal data is collected through the use of cookies by Facebook and the administrator has no access to data. The administrator could solve this practically by redirecting the request to Facebook. However, if Facebook would not adequately comply, the organization integrating their plugin may also be held accountable"*. The CJEU decision of Wirtschaftsakademie [28] deems both these organizations as joint controllers.

In this section we investigate authentication procedures presented in privacy policies of 50 popular websites and implemented in 30 third party services that track users on popular websites.

3.1 Evaluation of Popular Websites

In this work, we have analyzed privacy policies of the 50 top Alexa websites[5]. By doing so, we obtained information about the procedure enforced by websites for a data subject to get a copy of her data. Notice that the overall effectiveness of the GDPR rights from a European resident point of view depends on how easily and safely a resident can exercise such rights. Right of access is the most basic example of GDPR rights.

Evaluation criteria To compare procedures set up by popular websites, we propose three criteria: *Known identifiers*, *requested identifiers* and *accessibility*.

Known Identifiers – This criterion corresponds to the prior knowledge of the data controller on the subject. It is the information provided by the subject when she created her account on the website. We consider two cases. First, the data controller knows the subject through an identifier like a chosen username, an email address, a cookie or mobile identifier. Second, the data controller knows the real identity of the subject. This criterion is important to verify whether the identifiers requested to authenticate the subject are proportional to the knowledge of the data controller has about her.

Requested identifiers – During the authentication, the data controller can ask for more information on the subject to confirm her identity. For instance, he can ask for the copy of a government-issued document or a proof of residence. One important question for the requested identifier is the eligibility. According to its territorial scope, the GDPR applies to European companies handling personal data from all over the world and to companies handling data of European residents (as defined in par. 2 of Article 3 of the GDPR). European data controllers do not need to verify whether the subject is a European resident or not because they have to enforce the GDPR anyway. The case of foreign companies is different. They can decide to check the eligibility of the subject by demanding a copy of her national identity card. However, this is not sufficient

[5] Alexa measures web traffic and provides a ranking of the websites with respect to their traffic: `https://www.alexa.com/topsites`, extracted in October 2018.

because a non-EU citizen can reside in the EU and have the same GDPR rights. To check eligibility, data controllers should instead ask for a proof of residence (which may be different from citizenship), such as an electricity or a phone bill. However, this reveals more information about the data subject, such as her home address or phone number. Such collection can lead easily to an *abusive identity check* attack (see Section 2.1).

Accessibility – The data controller creates procedures to let the subject access her data. These procedures can be automatic (direct access): the subject logs into her account and can directly download her data. Another possibility is that the subject needs to send an email or a letter to the data protection officer. Data exchanged by emails are likely to be exposed to the knowledge of many people. Their use can lead to data breach.

Results of our evaluation Table 2 shows the results of our evaluation on the three criteria defined above. We have analyzed privacy policies of the top 50 Alexa websites and for simplicity we have re-grouped all the entries related to the same company in one table raw: Google (Google.com, Google.co.in, Google.co.jp, Google.com.hk, Google.com.br, Google.co.uk, Google.ru, Google.fr, Google.de, Youtube.com and Blogspot.com), Yahoo (Yahoo.com and Yahoo.co.jp) and Microsoft (Live.com, Microsoft.com, Bing.com, Microsoftonline.com, Office.com and Msn.com). After grouping, we get 27 entries in Table 2.

	Known identifiers		**Requested identifiers**			**Accessibility**	
Websites	Username	Real id.	Copy of an ID	Eligibility	Other	Direct Access	email
Google.com [5]	✓	✗	✗	✗	✗	✓	✗
Facebook.com [25]	✓	✗	✗	✗	✗	✓	✗
Baidu.com [3]	✓	✗	✓	✓	✓	✗	✓
Wikipedia.org [10]	–	–	–	–	–	–	–
Yahoo.com [21]	✓	✗	✗	✗	✗	✓	✗
Qq.com [11]	✓	✗	✓	✓	✗	✗	✓
Taobao.com, Tmall.com	–	–	–	–	–	–	–
Alipay.com [24]	✓	✗	✓	✓	✗	✗	✓
Twitter.com [2]	✓	✗	✗	✗	✗	✓	✗
Amazon [17]	✓	✗	✗	✗	✗	✓	✗
Instagram.com [13]	✓	✗	✗	✗	✗	✓	✗
Vk.com [14]	✓	✗	✓	✓	✗	✗	✓
Sohu.com	–	–	–	–	–	–	–
Reddit.com [4]	✓	✗	✗	✗	✗	✓	✗
Yandex.ru [19]	–	–	–	–	–	–	–
Weibo.com	–	–	–	–	–	–	–
Sina.com.cn	–	–	–	–	–	–	–
360.cn	–	–	–	–	–	–	–
Netflix [18]	✓	✗	✓	✓	✗	✗	✓
Pornhub.com [12]	✓	✗	✗	✗	✗	✗	✓
Linkedin.com [20]	✓	✗	✗	✗	✗	✓	✗
Mail.ru [7]	–	–	–	–	–	–	–
Twitch.tv [22]	✓	✗	✗	✗	✗	✓	✗
Ebay.com [15]	✓	✗	✓	✗	✓	✗	✓
Microsoft [6]	✓	✗	✗	✗	✗	✓	✗
Xvideos.com [23]	✓	✗	✓	✗	✗	✗	✓
Imdb.com [16]	✓	✗	✗	✗	✗	✓	✗

Table 2. Evaluation of the subject access right procedure of 50 popular websites.

We have not found any websites which require or force a subject to provide her real identity at registation. The subject often provides an email address, a username or any other element of her own choice.

We have observed three behaviors in the 50 most popular websites. Some websites have created access procedures for any data subjects without checking their eligibility. Others have a special part for "*EU users only* " within their privacy policies under the section "*Additional Information for EEA users*" or "*This section (Your Rights) applies to users that are located in the European Economic Area only*". In this case, the request form is often sent by e-mail, and by regular post for only a few websites. Finally, Amazon.com and IMDb.com have no specific procedure on their websites concerning how users can access their data.

The privacy policy of Wikipedia warrants that Wikipedia is a service only dealing with public informations posted by the users as stipulated *"If you only read Wikipedia without contributing, no more personal information is collected than is typically collected in server logs by web sites in general. If you contribute to Wikipedia, assume that it will be retained forever"*. Anyone can get access to the contributions history of any Wikipedia user and to the information given in his profile. There is no dedicated procedure for a contributor to collect his data.

Terms of service of Mail.ru[6] does not mention EU regulations, GDPR or subject access rights. In addition, Mail.ru doesn't take any responsibility for not allowing to download user's data: *"2.3. All the Services Mail.Ru, including mail service, are provided "as is". Mail.Ru does not assume any responsibility for the delay, removal, non-delivery or impossibility to download any User data, including User settings..."* (translated from Russian by the author).

For six Chinese websites (Taobao.com, Tmall.com, Sohu.com, Weibo.com, Sina.com.cn and 360.cn), we examined their content with a native Chinese speaker, but we were not able to find any information related to privacy policies.

Requesting additional information– The websites QQ, Baidu, Alipay, Aliexpress, Netflix, Ebay and Xvideos ask the subject to give additional information like national identity card or government issued documents. In these seven cases, the subject needs to give her real identity to access her data. Most of the time, the motivation to request these documents is eligibility. Alibaba group uses collected information to ensure the eligibility of the request: *"verifying your identity (. . .) verifying your eligibility as an EU User of Alipay Services (including "know your customer", anti-money laundering and counter-terrorist financing verification); processing your registration as an EU User, maintaining and managing your registration"*. The website Xvideos does not provide any justification for requesting a government issued document. This website is operated by WGCZ which is located in Czech Republic. The procedure of Xvideos is clearly an abusive identity check.

Accessibility – The tech giants, such as Google and Facebook, have the best practices regarding personal data access. When the subject authenticates herself to the service (using `https`), the data controller grants her access to a copy of her personal data without much effort. For example, Google uses `TakeOut`: a tool which allows to select the subject's data for every Google service she wants to include. `TakeOut` also sends an automated confirmation in order to detect impersonation. Microsoft websites, Facebook, Instagram, Twitter and Linkedin are using the same procedure which minimizes the amount of information needed to authenticate the subject.

For seven websites, the access request was initiated using emails. The majority of these websites also asks the subject to send the copy of an ID by email. Such practice might set this information at risk.

We also discovered two websites using a privacy proxy to manage the right of the subjects. Pornhub uses managemydata.eu and Twitch uses onetrust.com. OneTrust advertises on their website (`https://www.onetrust.com/`) to have already 1500 customers.

3.2 Evaluation of Third Party Trackers

When a data subject visits a website, she is interacting and being observed not only by the owner of the website, but also by numerous third party services included in those websites. In the recent years, researchers

[6] Point 2.3 of the Terms of Service, `https://help.mail.ru/mail-help/UA` (available only in Russian).

found that more than 90% of Alexa top 500 websites [88] contain third party tracking content, while some sites include as much as 34 distinct third party content [71].

Such third party content is often tracking users: third party tracking is the practice by which third parties recognize users across different websites as they browse the web. One of the most common and basic technology to track users is via *third-party cookies*. Such cookies, installed by the third party content when the user visits a website, usually contain a unique identifier and allows third parties to track the user across different websites, recreate part of her browsing history and collect data about her.

To examine the effectiveness of the access right set up by the GDPR in case of third party tracking services, we crawled the top 100,000 websites according to Alexa ranking in October 2018 from a server located in France [41]. For each website, we visited the home page and other 10 webpages on the same website. Out of 100,000 Alexa top websites, we successfully crawled 84,094 websites with a total of 829,349 webpages. We have identified the top 30 third parties that set third-party identifying cookies in the user's browser. We have then analyzed the privacy policies of these 30 third party trackers, and interacted with them via email when privacy policy page analysis was not sufficient to draw conclusions. As a result, we extracted information on the authentication procedures implemented by the third party tracking services integrated in websites, and whether it is possible to exercise the subject access rights with them based on identifiers stored in the browser.

Evaluation criteria: To evaluate the data access procedure set up by third party tracking services,we considered two main criteria *authentication* and *simplicity*.

Authentication – Authenticating the user is one of the main requirements to allow the user to access her data. By using the online identifiers– that could be either a cookie in case of web access or a mobile ID in case of mobile, – third parties can uniquely identify the user. Notice that both identifiers stored in cookie or mobile ID are considered personal data according to the 29 Working Party Opinion 2/2010 on online behavioral advertising [26]. In some cases, the third parties require additional personal information , such as the name, email or even the ID document.

Simplicity – We evaluate simplicity by distinguishing how easy it is for the data subject to access her data collected by the third party trackers. Some third parties provide user-friendly access directly from the website, while for others the data subject need to suffer from long email exchanges making the data access very difficult for the data subject.

Results of our evaluation: Table 3 shows the results of our evaluation on the two main criteria described above. To simplify, we have grouped all the domains owned by Google (doubleclick.net, google.com, gstatic.com, youtube.com, google.fr, googlesyndication.com and 2mdn.net).

Impossible to start exercising SAR – Two companies, simpli.fi and casalemedia.com, were abusing identity check at the information extraction level. Simpli.fi refused to provide us with more information about the process unless we provide first and last name, address, phone number and email. Casalemedia.com did not explain how to exercise SAR on their website, and in order to ask a question we had to go through an online from, where we should provide additional personal data.

For four companies, teads.tv, baidu.com, innovid.com and serving-sys.com, we were not even able to start the SAR process. In their websites, teads.tv [97] and baidu.com [49] precise that data access is done upon request. We sent an email asking how we can access the third party data on December 6, 2018 and January 7, 2019 respectively but we have never received an answer as of March 18, 2019. We sent an email to innovid.com following the instruction on their website [70], but it appears that their domain isn't properly registered. Our message couldn't be delivered. The website of serving-sys.com is not accessible because of insecure connection error.

[7] Pubmatic also ask for the ID card of the witness who signs the SAR form together with the data subject.
[8] Adform declares that the provided personal data will be retained for 10 years.
[9] Quantserve provides the data subject a link that she should revisit after 30 days to fetch her data.

Third-party domain	Authentification					Simplicity	
	Online identifier		Other data			Direct access	email
	Cookies	Mobile ID	Name and surname	email	ID card		
simpli.fi [91]	⊘	⊘	⊘	⊘	⊘	⊘	⊘
casalemedia.com [52]	⊘	⊘	⊘	⊘	⊘	⊘	⊘
teads.tv [97]	⊘	⊘	⊘	⊘	⊘	⊘	⊘
baidu.com [49]	⊘	⊘	⊘	⊘	⊘	⊘	⊘
innovid.com [70]	-	-	-	-	-	-	-
serving-sys.com	-	-	-	-	-	-	-
Google domains	⊘	⊘	⊘	⊘	⊘	⊘	⊘
facebook.com [25]	⊘	⊘	⊘	⊘	⊘	⊘	⊘
nr-data.net [77]	⊘	⊘	⊘	⊘	⊘	⊘	⊘
demdex.net	⊘	⊘	⊘	⊘	⊘	⊘	⊘
everesttech.net	⊘	⊘	⊘	⊘	⊘	⊘	⊘
yandex.ru [105]	⊘	⊘	⊘	⊘	⊘	⊘	⊘
openx.com [83]	⊘	⊘	⊘	⊘	⊘	⊘	⊘
pubmatic [85]	✓	✓	✓	✓	✓[7]	✗	✓
mathtag.com [73]	✓	✓	✓	✓	✓	✗	✓
weborama.fr [104]	✓	✓	✓	✓	✓	✗	✓
criteo.com [60]	✓	✓	✓	✓	✓	✗	✓
scorecardresearch.com [90]	✓	✓	✓	✓	✗	✗	✓
adform.com [33]	✓	✓	✓[8]	✓	✗	✗	✓
agkn.com	✓	✓	✗	✓	✗	✗	✓
smartadserver.com [92]	✓	✗	✗	✓	✗	✗	✓
adnxs.com [34]	✓	✓	✗	✗	✗	✓	✗
adsrvr.org [35]	✓	✓	✗	✗	✗	✓	✗
quantserve.com [87]	✓	✗	✗	✗	✗	✓[9]	✗
spotxchange.com [95]	✓	✓	✗	✗	✗	✓	✗

Table 3. Evaluation of the subject access right procedure of top 30 third parties: " ⊘" means that the request is denied by the third party, while "-" means it's not technically accessible.

Denial of access – Three companies answered our emails within less than one month, but their answers did not help us exercise the SAR and get the third party data. Two tech giants that set identifier cookies, Google (that covers 7 distinct third party tracking domains) and facebook.com have not given us any indication on how to access the third party data. Instead, they pointed us to their documentation and how to access the data collected directly via their services as first parties. Nr-data.net owned by New relic did not ask for the cookie identifier but only told us that the email we are using to communicate with them is not linked to any data in their dataset.

Two companies, demdex.net and everesttech.net owned by Adobe also refused to provide us with the data collected from the third party context. In our experiments, we have observed that these companies use third party cookie identifiers that allow them to identify the data subject across websites. However, when we tried to exercise SAR, these companies stated that it's not possible to confirm that any information associated with the third party cookie relates to us. On a positive side, demdex.net and everesttech.net did not ask for addition personal information, but they didn't grant us access to the third party data. According to them, their practice is in line with GDPR, they quoted:

> *This is in line with the GDPR, which recognises both that the right to obtain a copy of personal data should not adversely affect others (art.15(4)) and that rights of access do not apply where an organisation is not able effectively to identify the data subject (art.11(2)).*

Two companies, yandex.ru and openx.com refused to process our request as well. These companies claim that they act as data processors on behalf of its publisher or developer partners. Hence, the subject access requests do not apply to them and they suggest us to contact the data controllers. Notice that such interpretation is not acceptable by the recent work of Mahieu et al. [72] and the CJEU decision of Wirtschaftsakademie [28] who state that both publishers and third parties are joint data controllers (see the beginning of Section 3).

Abusive identity check – Third party domains are able to recognize the user across websites with a unique identifier, which we detected to be stored in the third party cookies. Such unique identifier is not related to the user's other personal information such as name or email. Therefore, any proof of user's name (such as the identity card) or email is not useful *to prove the ownership of the cookies*.

During our evaluation, we noticed that eight companies asked to provide not only the online identifier but other personal information as well. This practice allows third parties to link the data subject's online identifier to her personal information. Therefore, a data subject is forced to reveal even more personal data to the third party in order to practice her access right. This results in an *abusive identity check*.

Eight companies, pubmatic.com, smartadserver.com, mathtag.com, scorecardresearch.com, agkn.com, weborama.fr, adform.com and criteo.com require additional information to authenticate the user such as the full name or even the ID document. In addition to the subject's ID document, pubamtic.com asks for the name and the ID document of a witness who signs the SAR form together with the data subject. Five out of eight companies (pubmatic.com, mathtag.com, adform.com, weborama.fr and criteo.com) ask the user to fill a form, print and sign it in order to validate that she is the owner of the online identifier and of the device associated to it. Interestingly, adform.com uses this form to acknowledge the user that the company will process the additional personal data provided in the signed form (such as signature and full name) and retain it for up to 10 years! To access her data, the user has no choice except to agree and sign this form.

Direct access without requesting additional data – Four companies, adnxs.com, adsrvr.com, quantserve.com and spotxchange.com provide direct access to third party data based on the data subject's third party cookie. To verify the identity of the user and prove the ownership of the cookie, adnxs.com and adsrvr.com add a verification step where the user confirms in an online form that she is the owner of the identifier.

4 Recommendations and observations

After having analyzed the recommendations of the European DPAs in Section 2, and the practices of website owners and third party tracking services in Section 3, we have identified several major issues that data controllers face when they need to implement the software support tools for the subject access requests. Moreover, the current legal framework conveyed by the GDPR in relation to the right of access only provides for an obligation of conduct, requiring indeed certain actions to assure this right (described in the modalities of the access right – as depicted in articles 12, 15 and recitals 57, 58, 59, 60, 63, 64), but without rendering any procedural undertaking or benchmark as to an effective and specific result, which could shape the practices of the companies providing a SAR [50]. Pursuant to this normative need, in this section we give recommendations to both data controllers and data subjects concerning both problems: of authentication and validation of eligibility.

4.1 Problem of Authentication

There are two ways to authenticate the data subject by the data controller: either via the real identity of the subject (through her name surname and government-issued ID) or through the digital identity (assigned identifier, cookie, IP address, etc.).

Authentication via government-issued ID In case the data subject has never interacted directly with the data controller through electronic means (a typical case is the e-commerce discount cards), the data controller can rightfully ask the data subject to provide the proof of her identity, such as her ID card. In this case, there are two possible threats involved.

First, a security incident can occur on the data controller's side and the copy of the data subject's ID document can be leaked to attackers. Second, the data controller (or the attacker from the previous case) can *impersonate* the data subject to other data controllers by using her ID document to exercise the subject access requests on her behalf. Moreover, with the data subject's ID document it is possible to impersonate her at any point in the future (until the ID document expires). One obvious solution would be to blur some of the information on the data subject's ID document: this practice would protect some of her information from being leaked to attackers but it does not protect her from impersonation.

How to protect from impersonation?

The proofs provided by the data subject must satisfy the *non-transferability property* [74]: the documents provided by a data subject to a data controller, during a given authentication, cannot be reused in any other authentication. To protect the data subject's ID document, she should add a watermark which can not be removed from the copy of the document. This watermark must contain two elements:

- *A validity period* to prevent anyone from impersonating the data subject to the same data controller in the future.
- *The name of the data controller* to prevent anyone to use the copy with any other data controller.

Non-transferability can be implemented by signing the copy of the ID document with the date and the name of the data controller to prevent any further transfer. More complex solutions based on cryptography are also available. *Affidavits* are also an interesting alternative – they rely on a trusted third party which can be used to certify that a legal identity is bounded to an identifier. However, to protect the data subject from impersonation attacks, affidavits must also satisfy the non-transferability property.

Summary: The content that DPAs provide on their websites has a strong pedagogical role both to data subjects and data controllers. DPAs should update the information they convey publicly on their websites; specifically, they should require the non-transferability property to be applied to any usage of government-issued IDs. As a result, if data subjects follow such guidelines, and no longer share their government-issued IDs in the clear, they will avoid impersonation.

Authentication via digital identity In case the data subject has previously interacted with the data controller via electronic means, such as through an email or opening an account on the data controller's web portal, then these means of communication should be also used to authenticate the data subject. However, several security mechanisms must be put in place for a safe authentication of the data subject.

The communication through a web portal must at least use the secure channel `https` and a password. Ideally, for any online interaction, Two-Factor Authentication (2FA) is the ideal solution. 2FA requires that the user can be identified by two different factors, the most common are knowledge factors (such as password) and possession factors (such as physical or software tokens).

However, if the *data subject did not interact with the data controller via a web portal*, for example, when the data subject visited the web site where a third party (a joint controller) was tracking her, then the data subject needs to prove her identity to the controller based on her digital identifier. Examples of such identifiers are a browser cookie or an IP address.

An IP address is considered personal data according to Article 29 Working Party [8], however an IP cannot be used to uniquely identify a data subject in all cases. For example, an IP address does not allow an Internet Service Provider to distinguish data subjects who are connecting to the same wi-fi hotspot, or those using a shared computer. Hence, granting SAR within these boundary scenarios (when an IP address represents either one or many identifiable individuals) can be hard and could result in potential disclosures of other users' information.

If the data subject uses Privacy Enhancing Technologies (PETs), such as VPNs, anonymous networks like TOR[10], or cleans the browser cookies regularly, then it becomes nearly impossible to identify the data subject and hence prevents her from being able to exercise her subject access rights. Let us imagine a data subject who is visiting the website and uses TOR. Let us assume that the only digital identifier of the data subject visible to the data controller is the IP address observed by the data controller. However, because of the TOR network, this IP address does not belong to the data subject: it is a TOR exit point. Therefore, the data controller cannot identify the data subject by this IP address.

However, if the data subject browses the websites and is tracked by third party content present on the websites, and does not use any PETs, then third party trackers can use pseudoidentifiers (for example, stored in third party cookies) to track and recognize the data subjects. Interestingly, the IAB Europe GDPR Implementation Working Group raises the concern that pseudonymous data that is not linked to the individual's name and address cannot confirm that the data belongs to the requestor [29] and raise the subsequent question: *Should digital marketing companies that only collect pseudonymous data respond to data subject right requests?* Our answer to this question is definitely "yes", but their concern is valid: data subjects need to demonstrate and prove that the pseudoidentifiers (third party cookies, in our examples) indeed belongs to the data subject. In the following, we propose a procedure that would allow the data controllers to use pseudoidentifiers that are linked to the data subject's identity elements, like email address, yet the email address is not observable by any third party.

Without loss of generality, we consider the case of third party tracking via cookies. Cryptographic techniques can be used to bind the cookies with some identity elements, such as email, that can be checked later by the third parties. We provide a proof of concept algorithm on a cookie generation technique which is compatible with the GDPR.

In order for the data subject to be able to prove a cookie ownership, *the cookies must be generated on the client side* (in the web browser) rather than set by the server, as it is done in today's web standards. We assume that the subject has an email address denoted `address`, a public key K_{pub} and the corresponding private (RSA or ECC) key K_{priv}. The third party is associated with an identifier, such as third party's name or domain, denoted `tp_id`. The email `address`, K_{pub}, K_{priv} and third party identifier `tp_id` can be embedded in a web browser to make their usage transparent to the data subject. For a third party with identifier `tp_id`, we propose to compute a cookie value using the digest of a cryptographic hash function H (SHA256 or SHA3):

$$\texttt{cookie} = H(\texttt{tp\_id}, \texttt{address}, K_{pub}, N),$$

where N is a number (128-bit for instance).

When the data subject requests an access to her data, she provides her cookie `cookie`, K_{pub}, her email address `address` and the value N used to create the cookie. The third party tracker can recompute the cookie on his own and checks if it matches with the cookie sent by the subject. The third party can send an email to the subject at `address`. This email is encrypted with the data subject public key K_{pub} using software like pretty easy privacy (`https://www.pep.security/`). The data subject can now decrypt the message using her private key K_{priv}. Upon reception of an acknowledgement of the data subject, the third party is sure that the cookie indeed belongs to the data subject, and can now send her the data directly. An attacker that observes the communication between the data subject and the third party cannot predict or forge by himself the `cookie` of a legitimate user. The third party cannot attempt to recover by himself the value `address` and K_{pub} if the subject has not provided N. After getting her data, the data subject can renew `cookie` by changing the value N.

Currently, cookies are either set by servers (of publishers or third parties) or are programmatically set up in the browser by the JavaScript code running on a visiting webpage. Our protocol would require *to generate all the cookies at the browser side* and we believe it is possible to make it work even in a case when cookies are installed by a server: it's enough to run a client-side code that substitutes the cookies with the freshly-generated cookies that follow our algorithm. We believe it is better to have subject centric

[10] TOR is an anonymity network, directs Internet traffic through a worldwide overlay network, and therefore the IP address of the user's device is not visible to the server that receives requests from the user, `www.torproject.org`

approach to create digital identifiers. There are other possibilities than our scheme like the initiative of W3C on Verifiable Claims and Distributed Identifiers [94].

Summary: As of today, we are not aware of any GDPR compliant implementation of the pseudoidentifiers that would allow data subjects to be authenticated to exercise their rights and at the same time be protected from impersonation attacks. In this section, we have proposed a scheme that allows to generate pseudoidentifiers and protect the data subjects. To protect all the components of such scheme, it has to be implemented in the trusted environment of the data subject, which is her web browser.

4.2 Problem of Validating Eligibility

Data controllers also need to validate the SAR eligibility. If a data controller is European, he should review the Subject Access Request protocol and ensure that whenever enough information is already obtained to authenticate the data subject, no additional information should be requested. This approach would prevent *abusive identity check* attack. It is harder to verify eligibility of data subjects for non-European data controllers: they need to determine whether a request is legitimate or not by identifying whether the requestor is a resident in the European Union. Therefore, eligibility checks are legitimate in this case.

We draw the attention of the data subjects that they need to be aware that eligibility checks by non-European data controllers are required and do not constitute an *abusive identity check* attack. However, it is true that it is also complicated for data subjects themselves to establish whether a certain data controller is European or not.

Additionally, eligibility checks can be done via IP address of the requestor. In this case, Privacy Enhancing Technologies (PETs) play a dual role in the validation of eligibility. On one hand, as we have described before, if the European data subject uses PETs, such as TOR network, then she will likely maintain her anonymity at the cost of not being able to exercise the rights provided by the GDPR. On the other hand, a non-European data subject can use PETs, such as VPN, to pretend to be a EU resident to the data controller. If the data controller only relies on the IP address as a proof of eligibility, then he will allow a non EU resident to exercise her rights as well.

5 Related work

In 1969, Miller in [75] already considered that the right to access can be abused through impersonation. He pointed the risk of sharing personal data and violating people's privacy by unthoughtfully accommodating access requests.

The most notorious case of the right to access was given by Max Schrems [93]. In 2011, he contacted Facebook to exercise his right to access. He received a 1200-page document and discovered many anomalies showing Facebook was not compliant with the European laws and created `http://europe-v-facebook.org`.

The AFCDP (*Association Française des Correspondants aux Données Personnelles*) is a French association of the french data privacy officers. They publish every year a report [36, 37, 38, 39] on the right to access. They benchmark between 150 and 200 companies, administrations and organizations to test how they answer to data access requests. Their work is very close to ours. They primarily focus on measuring how many data controllers respond in time. Their reports also included anomalies and observations of misbehavior concerning the access right. We extend their work to evaluate more precisely how requests are treated by the data controllers. Our evaluation criteria could be re-used in the future by AFCDP during their benchmarks.

Asghari *et al.* [43] presented a benchmark of 32 data controllers in the Netherlands at HotPETS 2017. They acknowledge in their paper the fact that all the organizations they contacted authenticate the subject making the request. However, they did not analyze the authentication process nor if secure channels were used. They also mentioned in their work an upcoming benchmark of larger scale. Our work could help to obtain more precise results.

Ausloos *et al.* [44] also conducted a benchmark of the right to access on 60 organizations. Their tests asserted some organizations requested additional information to authenticate the users and especially copy of ID card or driving license. They observe that many obstacles exist for a subject who wants to exercise

her rights. They also point out the frequency with which an access request leads to an endless sequence of e-mails. Moreover, this sequence never resulted in the transfer of all the data legally allowed to be obtained. They have not taken into account security considerations as it is done in our paper.

In [78] the author points out that *"data protection law should apply to information that is used to single out people, even if no name can be tied to the information" Seeing data used to single out a person as personal data fits the rationale for data protection law: protecting fairness and fundamental rights. Data that are used to single out a person should be considered personal data"*. Although this might be enough to prevent impersonation, it could be dangerous to provide government issued documents to data controllers (unless they were required for the registration). When a data subject makes a request, she should obtain what she discloses to the data controller or what is related to her pseudonym. It is disproportional to provide governmental issued documents when the data subject has not used her real name to register on a website. Some authors [78] refer to "the visibility paradox" when dealing with the issue of disclosing additional information in order to obtain the data already disclosed.

The work of Urban *et al* [103] is very close to ours: the authors have studied the economy of web tracking by making subject access requests to third party websites. They have observed procedures of third parties to authenticate data subjects. Our observations and conclusions in Section 3.2 are very similar to those of Urban *et al*: the authors needed to sign several affidavits to access their third party data.

Grogan *et al.* [67] have analyzed how Internet users react to their right to access their data. They created a survey and distribute it to collect answers from US and Irish citizens. They observe that citizens are rather confused about their right to access and its application.

6 Conclusion

The right to access is the first and basic user right set up by the GDPR. In this paper, we have analyzed security aspects of the authentication procedures set up for subject access requests recommended by the DPAs and implemented by the website owners and third party tracking services.

While reviewing the recommendations of all the European DPAs, and the practice of the most popular websites and third party trackers, we have discovered several issues: abusive identity checks, potential data breach or denial of access. These issues are the results of incorrect procedures or a lack of means. Data controllers need to enforce the proportionality principle when they authenticate the requests to avoid abusive identity checks. The eligibility controls encountered during this work are a reminder that the relation between a data subject living in the European area and non-European data controllers is complex. Finally, webpages and third party trackers need to change their practice for the generation of identifiers to be compliant with the GDPR and avoid denial of access.

We hope that the materials provided in this paper can help to shape the design of better guidelines regarding the exercise of the users' rights and future benchmarking campaigns for the right to access.

Acknowledgments

This work is supported by the French National Research Agency in the framework of the *Investissements d'Avenir* program (ANR-15-IDEX-02) and project PrivaWEB (ANR-18-CE39-0008-01), and as well ANSWER project PIA FSN2 (P159564-2661789\DOS0060094).

Bibliography

[1] 29 working party opinion 2/2010 on online behavioural advertising, adopted on 22 june 2010, (wp 171), p. 9. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp171_en.pdf.

[2] 4.1 accès ou rectification de vos données à caractère personnel. https://twitter.com/fr/privacy, accessed on 28 September 2018.

[3] Access, rectification, opposition and cancellation rights. https://www.baidu.eu/privacy-policy, accessed on 28 September 2018.

[4] Accessing your reddit data. https://www.reddithelp.com/en/categories/using-reddit/your-reddit-account/accessing-your-reddit-data, accessed on 28 September 2018.

[5] Googletakeout. https://takeout.google.com/?utm_source=pp&hl=en, accessed on 28 September 2018.

[6] I want to make a request regarding personal data microsoft has about me related to my microsoft account. https://www.microsoft.com/en-us/concern/privacy, accessed on 28 September 2018.

[7] Mail.ru terms of service. https://help.mail.ru/mail-help/UA, accessed on 1 October 2018.

[8] Opinion n° 4/200 on the concept of personal data - wp 136, p.17. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf.

[9] Privacy policy. https://alidropship.com/privacy-policy/, accessed on 28 September 2018.

[10] Privacy policy/faq. https://foundation.wikimedia.org/wiki/Privacy_policy/FAQ#anonymize, accessed on 28 September 2018.

[11] Qqi ds rights request form. https://dl.url.cn/myapp/bhqq/iQQ/QQi_DS_RIGHTS_REQUEST_FORM.pdf, accessed on 28 September 2018.

[12] Request a copy of my personal data. https://fr.pornhubpremium.com/terms, accessed on 28 September 2018.

[13] Vi. how can you exercise your rights provided under the gdpr? data download. https://www.instagram.com/about/legal/terms/api/, accessed on 28 September 2018.

[14] Vk.com privacy policy. https://vk.com/privacy/eu for logged-in users, accessed on 1 October 2018.

[15] Ways you can access, control, and correct your personal information. https://www.ebay.com/help/policies/member-behaviour-policies/user-privacy-notice-privacy-policy?id=4260#section6, accessed on 28 September 2018.

[16] What choices and access do i have. https://www.imdb.com/privacy?ref_=helpms_helpftr_privacy, accessed on 28 September 2018.

[17] What information can i access. https://www.amazon.co.uk/gp/help/customer/display.html?nodeId=502584, accessed on 28 September 2018.

[18] What personal information netflix holds about you and how to request a copy. https://help.netflix.com/en/node/100624?ba=SwiftypeResultClick&q=request%20a%20copy%20of%20my%20data, accessed on 28 September 2018.

[19] Yandex.ru privacy policy. https://yandex.com/legal/privacy/, accessed on 1 October 2018.

[20] Your choices and obligations. https://www.linkedin.com/legal/privacy-policy, accessed on 28 September 2018.

[21] Your control and privacy rights. https://policies.oath.com/ie/en/oath/privacy/index.html, accessed on 28 September 2018.

[22] Your privacy choices. https://www.twitch.tv/p/legal/privacy-choices/, accessed on 28 September 2018.

[23] Your rights. https://info.xvideos.com/legal/privacy/, accessed on 28 September 2018.

[24] Your rights with respect to your personal information. https://render.alipay.com/p/f/agreementpages/alipayeuprivacypolicy.html, accessed on 28 September 2018.

[25] Yourfacebookinformation. https://www.facebook.com/full_data_use_policy, accessed on 28 September 2018.

[26] Opinion 2/2010 on online behavioural advertising. Technical Report 171, 2010.
[27] The Schengen Information System A Guide For Exercising The Right of Access, 2015. https://edps.europa.eu/sites/edp/files/publication/16-11-07_sis_ii_guide_of_access_en.pdf.
[28] Case C-210/16 Wirtschaftsakademie Schleswig-Holstein. 2018. ECLI:EU:C:2018:388, http://curia.europa.eu/juris/document/document.jsf?docid=202543&doclang=EN.
[29] Data subject requests, working paper 04/2018, 2018. available online at https://www.iabeurope.eu/wp-content/uploads/2018/04/20180406-IABEU-GIG-Working-Paper04_Data-Subject-Requests.pdf.
[30] European Data Protection Board, 2018. urlhttps://edpb.europa.eu.
[31] European Data Protection Supervisor, 2018. urlhttps://edps.europa.eu.
[32] Addthis - privacy policy. https://www.addthis.com/privacy/privacy-policy/.
[33] Adform - privacy policy. https://site.adform.com/privacy-center/website-privacy/website-privacy-policy/.
[34] Adnxs – appnexus data subject rights. https://www.appnexus.com/data-subject-rights-policy.
[35] Adsrvr. https://www.adsrvr.org/.
[36] AFCDP. Données personnelles - Index AFCDP du Droit d'accès. Technical report, 2013. In french.
[37] AFCDP. Données personnelles - Index AFCDP du Droit d'accès. Technical report, 2014. In french.
[38] AFCDP. Données personnelles - Index AFCDP du Droit d'accès. Technical report, 2015. In french.
[39] AFCDP. Données personnelles - Index AFCDP du Droit d'accès. Technical report, 2017. In french.
[40] Agencia de Protección de Datos. Ejerce tus derechos. https://www.aepd.es/media/formularios/formulario-derecho-de-acceso.pdf, accessed on 28 September 2018.
[41] Alexa. https://www.alexa.com/.
[42] Andmekaitse Inspektsioon . Andmekaitse Inspektsioon. http://www.aki.ee/, accessed on 28 September 2018.
[43] Hadi Asghari, Rene L.P. Mahieu, Prateek Mittal, and Rachel Greenstadt. The Right of Access as a tool for Privacy Governance. In *Proceedings of Hot Topics in Privacy Enhancing Technologies (HotPETs 2017)*, 2017.
[44] Jef Ausloos and Pierre Dewitte. Shattering one-way mirrors – data subject access rights in practice. *International Data Privacy Law*, 8(1):4–28, 2018.
[45] Ihre rechte als betroffener, 2018. https://www.dsb.gv.at/rechte-der-betroffenen, accessed on 28 September 2018.
[46] Antrag gemäß art. 15 DSGVO auf auskunft, 2018. https://www.dsb.gv.at/at.gv.bka.liferay-app/documents/22758/844171/Antrag+an+den+Verantwortlichen+Recht+auf+Auskunft+Art+15.pdf/00315f65-1ea8-438b-8f1f-766d20002702, accessed on 28 September 2018.
[47] Autorité de protection des données. Lettre Type Droit Acces Direct . https://www.autoriteprotectiondonnees.be/node/3995, accessed on 28 September 2018.
[48] Autoriteit Persoonsgegevens. Recht op inzage. https://autoriteitpersoonsgegevens.nl/nl/zelf-doen/privacyrechten/recht-op-inzage#, accessed on 28 September 2018.
[49] Baidu - privacy policy. http://usa.baidu.com/privacy/.
[50] Emre Bayamlıoğlu. Transparency of automated decisions in the gdpr: An attempt for systemisation. 2018. Available at https://ssrn.com/abstract=3097653.
[51] Frederik Zuiderveen Borgesius. Singling Out People Without Knowing Their Names – Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation. 2016. https://ssrn.com/abstract=2733115.
[52] Casalemedia - privacy policy. http://casalemedia.com/.
[53] CNIL Commission Nationale de l'Informatique et des Libertés. Guide sécurité des données personnelles. https://www.cnil.fr/fr/le-droit-dacces-connaitre-les-donnees-quun-organisme-detient-sur-vous, accessed on 28 September 2018.
[54] Comissão Nacional de Protecção de Dados . Comissão Nacional de Protecção de Dados . https://www.cnpd.pt, accessed on 28 September 2018.

[55] Commission for Personal Data Protection . Who can copy your identity card. https://www.cpdp.bg/index.php?p=element&aid=423, accessed on 28 September 2018.
[56] Commissioner for Personal Data Protection . Commissioner for Personal Data Protection . http://www.dataprotection.gov.cy/, accessed on 28 September 2018.
[57] Andrew Cormack. Is the Subject Access Right Now Too Great a Threat to Privacy? *European Data Protection Law Review*, 2(1):15–27, 2016.
[58] Council of European Union. Council regulation (EU) no 2016/679, 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679.
[59] Council of European Union. Council regulation (EU) no 2016/679, 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679.
[60] access right criteo. https://www.criteo.com/privacy/.
[61] Croatian Personal Data Protection Agency. Croatian Personal Data Protection Agency . https://azop.hr/, accessed on 28 September 2018.
[62] Data Protection Commissioner. A guide to your rights. https://www.dataprotection.ie/docs/A-guide-to-your-rights-Plain-English-Version/r/858.htm, accessed on 28 September 2018.
[63] Data Protection Commissioner of Hungary. Annual report of the Hungarian National Authority for Data Protection and Freedom of Information (NAIH) 2017. http://www.naih.hu/annual-reports.html, accessed on 28 September 2018.
[64] Data State Inspectorate. Datu subjekta tiesibas. http://www.dvi.gov.lv/lv/wp-content/uploads/DVI_broshura_datusubjekt_ties.pdf, accessed on 28 September 2018.
[65] Datatilsynet . Guidance on the registrants' rights. https://www.datatilsynet.dk/media/6893/registreredes-rettigheder.pdf, accessed on 28 September 2018.
[66] Die Bundesbeauftragte für den Datenschutz und die Informationsfreiheit . Auskunftsrecht. https://www.bfdi.bund.de/DE/Datenschutz/Ueberblick/MeineRechte/Artikel/Auskunftsrecht.html, accessed on 28 September 2018.
[67] Samuel Grogan and Aleecia M. McDonald. Access Denied! Contrasting Data Access in the United States and Ireland. *PoPETs*, 2016(3):191–211, 2016.
[68] Hellenic Data Protection Authority. Law 2472/1997 & Citizen's rights. http://www.dpa.gr/portal/page?_pageid=33,43290&_dad=portal&_schema=PORTAL, accessed on 28 September 2018.
[69] Information Commissioner. Request for acquaintance with your own personal data. https://www.ip-rs.si/fileadmin/user_upload/doc/obrazci/ZVOP/Zahteva_za_seznanitev_z_lastnimi_osebnimi_podatki__Obrazec_SLOP_.doc, accessed on 28 September 2018.
[70] access right innovid. https://www.innovid.com/privacy-policy/.
[71] Adam Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, 2016.
[72] Rene Mahieu, Joris van Hoboken, and Hadi Asghari. Responsibility for data protection in a networked world – on the question of the controller, "effective and complete protection" and its application to data access rights in europe. 2019. Available at https://ssrn.com/abstract=3256743.
[73] Mathtag - privacy policy. http://www.mediamath.com/privacy-policy/#Section-11.
[74] Alfred Menezes, Paul C. van Oorschot, and Scott A. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1996.
[75] Arthur R. Miller. Personal privacy in the computer age: The challenge of a new technology in an information-oriented society. *Michigan Law Review*, 67(6):1089–1246, 1969.
[76] National Commission for Data Protection. The right of access. https://cnpd.public.lu/en/particuliers/vos-droits/droit-acces.html, accessed on 28 September 2018.
[77] New relic - privacy policy. https://www.simpli.fi/site-privacy-policy/.
[78] Paul Norris, Cliveand de Hert, Xavier L'Hoiry, and Antonella Galetta. *The Unaccountable State of Surveillance Exercising Access Rights in Europe*. Springer International Publishing, 2017.
[79] Office for Personal Data Protection of the Slovak Republic. Dina rättigheter enligt personuppgiftslagen . https://www.datainspektionen.se/globalassets/dokument/gammalt/dina-rattigheter-enligt-personuppgiftslagen.pdf, accessed on 28 September 2018.

[80] Office for Personal Data Protection of the Slovak Republic. How to submit a petition initiating the procedure of personal data protection . https://dataprotection.gov.sk/uoou/en/content/how-submit-petition-initiating-procedure-personal-data-protection, accessed on 28 September 2018.
[81] Office of the Data Protection Commissioner . What is the Right of Access? . https://idpc.org.mt/en/Pages/faq.aspx#3, accessed on 28 September 2018.
[82] Office of the Data Protection Ombudsman . When you want to inspect your data. https://tietosuoja.fi/en/when-you-want-to-inspect-your-data, accessed on 28 September 2018.
[83] Openx - privacy policy. https://www.openx.com/legal/privacy-policy/.
[84] Garante per la protezione dei dati personali. Guida all'applicazione del regolamento europeo in materia di protezione dei dati personali - diritti degli interessati, 2018. https://www.garanteprivacy.it/regolamentoue/diritti-degli-interessati, accessed on 28 September 2018.
[85] Data subject rights notice, pubmatic. https://pubmatic.com/legal/eea-data-subject-rights-notice/.
[86] Pubmatic - cookie policy. https://pubmatic.com/legal/platform-cookie-policy/.
[87] Quantserve - privacy policy. https://www.quantcast.com/privacy/.
[88] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. Detecting and defending against third-party tracking on the web. In *Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2012*, pages 155–168, 2012.
[89] access right rubiconproject. https://rubiconproject.com/terms-conditions/subject-access-request-policy/.
[90] access right scorecardresearch. https://www.scorecardresearch.com/privacy.aspx.
[91] Simpli - privacy policy. https://www.simpli.fi/site-privacy-policy/.
[92] Smartadserver - privacy policy. https://smartadserver.com/end-user-privacy-policy//.
[93] Olivia Solon. How much data did facebook have on one man ? 1.200 pages of data in 57 categories. *Wired*, 2012. https://www.wired.co.uk/article/privacy-versus-facebook.
[94] Manu Sporny and Dave Longley. Verifiable Claims Data Model and Representations. Technical report, W3C, 2017. https://www.w3.org/TR/verifiable-claims-data-model/.
[95] Spotxchange - privacy policy. https://www.spotx.tv/privacy-policy/.
[96] Spotxchange portal. https://www.spotx.tv/privacy-policy/gdpr/.
[97] Teads - privacy policy. https://www.teads.tv/privacy-policy/.
[98] The Bureau of the Inspector General for the Protection of Personal Data - GIODO. Rights of data subject. https://giodo.gov.pl/en/293, accessed on 28 September 2018.
[99] The Information Commissioner's Office . Your right of access. https://ico.org.uk/your-data-matters/your-right-of-access/, accessed on 28 September 2018.
[100] The Information Commissioner's Office. Your right to get copies of your data. https://ico.org.uk/your-data-matters/your-right-of-access/, accessed on 28 September 2018.
[101] The National Supervisory Authority for Personal Data Processing. Derptul de Acces. http://www.dataprotection.ro/servlet/ViewDocument?id=386, accessed on 28 September 2018.
[102] The Office for Personal Data Protection . The Office for Personal Data Protection . http://www.uoou.cz/, accessed on 28 September 2018.
[103] Tobias Urban, Dennis Tatang, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. The Unwanted Sharing Economy: An Analysis of Cookie Syncing and User Transparency under GDPR. *CoRR*, abs/1811.08660, 2018.
[104] Weborama - privacy policy. https://weborama.com/weborama-privacy-commitment/.
[105] Yandex.ru - privacy policy. https://yandex.com/legal/privacy/.