WHOIS versus GDPR

Rens Oliemans
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
r.p.oliemans@student.utwente.nl

ABSTRACT

WHOIS is a protocol for requesting information about registered domains and their registrants. The data about registrants is often personal data, including full names, phone numbers and addresses. Publicly showing this data is not in line with the EU General Data Protection Regulation of 2018 [1], and registrars showing this data about EU citizens are incompliant with EU law. This paper measures the amount of personal data publicly available via the WHOIS protocol. Since the GDPR only applies to citizens of the European Union, this paper compares the amount of personal data between 13 EU countries and 7 Non-EU countries. Before measurements can be done about the amount of personal data, however, the parsing of WHOIS data first needs to be improved. Since different Top-Level Domains – TLDs – have different formats to return this data, the parsing of WHOIS data is challenging. In this paper, we create some improvements to the state of the art of open WHOIS parsing. As for measuring personal data, a large difference in the existence of personal data can be found between different TLDs. Where 12 TLDs have no personal data at all, 3 TLDs contain personal information in more than half of the domains. Additionally, a significant difference can be found in the presence of personal data between EU countries and Non-EU countries, with Non-EU countries containing more personal information in their domains than EU countries.

Keywords

WHOIS, GDPR, personal data, parsing, rate limit

1. INTRODUCTION

Domain names are an essential part of the internet. Domain Name Servers map a domain name, like `https://duckduckgo.com`, to an IP address. This allows you to browse to that domain name instead of having to go to the IP address belonging to the server. When registering a domain name, one has to give some information to the *registrar* – a company which is responsible for controlling domain names. This usually includes personal information such as their full name, email addresses, phone numbers and physical addresses.

In many cases, this personal information is available for everyone with an internet connection, making it very easy for someone's personal information to be seen. This information about a domain name – and the person who registered it – can be accessed via the WHOIS protocol. An example of such a query can be seen in figure 1.

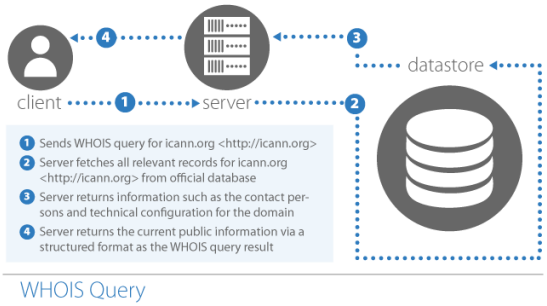

Figure 1. Overview of a WHOIS Query

The GDPR is a regulation about the collection and protection of personal data, introduced in May 2018 [1]. WHOIS, on the other hand, is a way of gathering (often personal) data about the registrants. Since the GDPR, registrars are not allowed to publicly show this data of EU citizens any more. This had lead to a couple of countries restricting the amount of personal data visible on domains within their Top-Level Domain. A Top-Level Domain, or TLD, is a domain at the highest hierarchy. For example, for the domain name `www.example.com`, the TLD is `.com`. Top-level registries handle the way WHOIS data is returned. For the Netherlands (`.nl`), the organisation doing so is the SIDN. The SIDN restricted personal data via WHOIS in March 2018 [11].

As of now, the GDPR has been in effect for over a year. Yet, no recent assessment of public personal information via the WHOIS protocol exists. Some countries have claimed they restricted the visible personal data [11, 8] but not all EU countries have done so.

All registrars which contain personal information about EU citizens have to comply with the GDPR. However, it is unclear how many registrars actually do so. This research will show the amount of personal information available on different TLDs, within an outside the European Union. The results of this research might be used for these registrars – or top-level registries such as SIDN – to see how much personal data is still – unlawfully – public via the WHOIS protocol.

In order to measure the amount of personal information available via WHOIS, a large amount of WHOIS data has to be parsed first. Parsing of WHOIS is notoriously difficult [6]. Nearly all Top-Level Domains – TLDs, such as `.nl` – have a custom structure for outputting WHOIS data. One of the most active open source tools to parse this WHOIS data is `pywhois`[1]. This library parses WHOIS data, but is not very accurate for many domains. In order

[1] `https://pypi.org/project/python-whois`

31st Twente Student Conference on IT July 5th, 2019, Enschede, The Netherlands.

to analyse WHOIS data, this library must first be improved for some TLDs.

To summarise, there are no easy ways to accurately parse WHOIS data on a large scale as of now. Additionally, it is unknown how much personal information is available via WHOIS, while showing this data is incompliant with the GDPR. This creates some interesting research questions:

1 *In what way can the current public way of parsing* WHOIS *data be improved?*

2 *What personal data is still publicly accessible on* WHOIS*?*

3 *Is there a difference between the amount of personal data available via* EU TLD*s and Non-*EU TLD*s?*

4 *Is there a difference between the amount of personal data available within* EU TLD*s?*

To answer these questions, WHOIS data must first be parsed. However, in order to sufficiently say something about the parsed data, one needs to be sure that the data is accurately parsed. In WHOIS, two different lookup models exist: *Thick* and *thin* [4]. The models refer to the way labelling and displaying of WHOIS output is done. In thick models, the format of WHOIS is the same for all domains within that TLD. In thin models, however, the domains are not required to have a certain format. This leads to the registrars within a TLD themselves determining how the format should be [4], making parsing thin lookup models quite difficult due to the different formats. Therefore, this research will only parse domains within a TLD that has a thick lookup model.

First, the parsing of `pywhois` will be improved. This question is answered in section 6.2.

Secondly, the amount of personal data will be analysed. This question will be answered by filtering personal information from WHOIS data, as seen in section 6.3. Finally, the last two questions will be answered via statistical measures in section 6.4.

2. RELATED WORK

This research consists of two partly separated sections. One is the improving of WHOIS parsing. The other is analysing the parsed results and drawing conclusions about possible personal information available from WHOIS data.

Some current solutions for parsing WHOIS already exist. The most popular open-source parser is `pywhois`[2]. Additionally, there are also other open parsers, such as `whoisrb`[3], but these have lower accuracy than `pywhois`, or are more difficult to extend and improve. Many paid WHOIS APIs also exist, such as WHOISXML[4] or WHOAPI[5]. For this research, only free clients are looked at. The current solutions are not without flaws. Mainly, `pywhois` lacks some information about certain domains, mostly "unconventional" or less popular domains, such as `.nl` domains.

The second part is analysing WHOIS data and finding personal information. Since the GDPR is a recent development, few papers exist looking at WHOIS after the GDPR. An important paper is 'Balancing Privacy and Security in a Multistakeholder Environment. ICANN, WHOIS and GDPR' [5], which looks at privacy and security of WHOIS data. However, that paper mainly focuses on ICANN and its temporary specification for tiered levels of access [3], which also concerns WHOIS data, but is outside the scope of this research.

[2]`https://pypi.org/project/python-whois/`
[3]`https://github.com/weppos/whois`
[4]`https://www.whoisxmlapi.com/`
[5]`https://whoapi.com/page/api`

In conclusion, some work exists for parsing WHOIS data, but these solutions lack accuracy, are proprietary, or suffer from other issues. As for personal information, barely any prior research exists, likely due to the fact that the GDPR is rather new.

3. MEASURING DATA

This section contains the methodology of the research: how the results were measured and computed. The first part concerns how Research Question 1 will be answered, so how the current state of the art can be improved. This section will also cover the ways the data was gathered. Determining personal data is the last part of this section, related to the last three Research Questions.

The first Research Question reads *In what way can the current public way of parsing* WHOIS *data be improved?*.
There are many ways to obtain WHOIS data, as seen in section 2. For this study, the `pywhois` library was used, as it was the best viable option available which was open to use. However, the library was still not accurate enough, especially when parsing Top-Level Domains that were not common worldwide (such as `.nl`, `.se`, etc.). In order to properly parse WHOIS data and analyse it for personal data, the library had to be fixed. Since there was limited time, the improvements made are the ones which would have the largest impact given the time put in. Support for some top-level domains had to be added completely, but most fixes of `pywhois` were fixes that were small to implement, yet gained a large benefit to the working of the library. Some examples of fixes are shown in section 5.1.

The second Research Question (*What personal data is still publicly accessible on* WHOIS*?*) is concerned with personal data. In order to be able to answer it, a definition of personal data must first be given. The EU defines it as follows:

> Personal Data is any information that relates to an identified or identifiable living individual. Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data.1 [1]

To better understand how personal data might exist in WHOIS data, an example of a typical WHOIS query output has been given in figure 2. In the case of WHOIS, personal data is only present in the *registrant* fields, as the registrant is the person registering the website. The *registrar* fields, on the other hand, contain data about registrars, which are always large companies. So, they contain no personal data about the people having registered the domain.

For the second, third and fourth Research Questions this research needs to obtain a large amount of WHOIS data. The data used for this research comes from random domains published by OpenDNS [9]. All TLDs with 10 or more domains were looked at in this research. In total, there were 19 of such TLDs, with a median of 30 domains per TLD.
Comparing multiple datasets with 10 elements is usually not representative. In WHOIS it is a bit different. The top-level registries investigated in this research all use a *thick* WHOIS lookup model. Thus, the structure of WHOIS data is the same for all domains within that TLD. This means that there is little to no variation between domains.
Additionally, the top-level registry determines the rules of the WHOIS data. So, a top-level registry – for the Netherlands a top-level registry would be SIDN – enforces the same rules for all registrars. So, when some personal information is available in some domains in the TLD, it is reasonable to suppose that personal information is present in many domains for that TLD.

Finally, the data has to be analysed and personal data needs to be detected. A definition of personal data was already given, but

Figure 2. An example of some WHOIS output from a Danish domain. The personal data has been blurred to protect the person's privacy.

finding it is still not trivial. Determining personal data automatically remains a difficulty. One can look into the "name", email", "address" or "phone" fields of the registrants, but, whereas it is easy to see that `REDACTED_FOR_PRIVACY` is not a personal name, it can be difficult to assess whether a name belongs to a person or company. Also, the given names are sometimes "Domain Administrator", "<company> Support", "Web Master", etcetera. These are clearly no personal names, but can be difficult to automatically detect.

Therefore, the data found was parsed automatically via `pywhois`, but inspected manually to see if it contained personal information.

4. DIFFICULTIES

The two main difficulties encountered during the research were the fixing of `pywhois` and the handling and circumvention of rate limiting.

The first difficulty is explained and answered as Research Question 1. The parsing of the used open-source library was incomplete or inaccurate, for fixes see section Fixing Pywhois.

Another difficulty in the large-scale parsing and analysing is Rate Limiting. Quite a few WHOIS servers belonging to top-level domains enforce rate limits to prevent spam. This is necessary for the proper functioning of the WHOIS servers, but does make data collection for research a bit more difficult. Table 2 shows the final rate limits per TLD.

5. RESULTS

The results will first cover the first research question, the fixing of `pywhois`. Then, the results belonging to personal data will be shown. A comparison between different countries – EU and Non-EU – will be made. Finally, Table 2 shows the Rate Limits encountered during the research.

5.1 Fixing Pywhois

As explained above, `pywhois` was not sufficient for the parsing of WHOIS data. Therefore, some changes needed to be made for the library to improve its accuracy. In total, support for parsing of 3 TLDs were completely added to parse in whois: `.il`, `.si` and `.no`. The United Arab Emirates TLD `.ae` was also added, but this had no effect in the final results since there were not enough domains for the TLD to be added. In addition, 15 TLDs were improved. The library supported parsing for these TLDs but this was either incorrect or incomplete. These improved TLDs also included TLDs such as `.hr` or `.cz`, so also included TLDs that were not in the final dataset.

The following two paragraphs contain examples of the changes made to improve the `pywhois` library. These two changes take different approaches and are more or less representative for the rest of the changes.

Fixing .nl. Since `.nl` has a thick WHOIS model, making a fix for `.nl` domains would have a large difference since the fix would apply for all `.nl` domains. Before, the library only parsed the domain name, status, and the name of the registrar.
After the fix, it also parsed the address, country and the DNSSEC [2] status of the domain. The output of `.nl` WHOIS data was likely changed recently since the names of the registrar fields that `pywhois` expected were wrong.

Fixing .dk. The Denmark WHOIS server does not provide any registrant data by default. However, when one adds `--show-handles` to the request the registrant data is also shown[6]. This was fixed by prepending `--show-handles` to the request that `pywhois` makes when it connects with the Danish NIC Client.

5.2 Personal Data

Concerning research question *What personal data is still publicly accessible on* WHOIS*?*, Table 1 shows the amount of personal data visible depending on the TLD. The column "TLD" refers to the TLD to which the domains belong. "# of domains" means the number of domains that are present of that TLD in the dataset used. Finally, "Personal Data" refers to the percentage of domains which contained some form of personal data, as defined in section Measuring Data.

Of the 13 TLDs in the EU, 3 contained personal information. `.dk` – belonging to Denmark – had full names in the "Registrant" fields. However, many used a privacy service to mask the actual full name. The Italian `.it` domains contained the most personal information of EU domains. This personal information was found in the "Registrant" fields, but also in the Admin and Tech fields ("name", "address", "phone", "email"). Finally, Finnish domains (`.fi`) occasionally contained personal information as well. This included their full name and usually the address and phone number.

Of the 7 TLDs outside the EU, 5 contained personal information. The Canadian `.ca` domains have redacted registrant fields. However, personal information can still be found in the "Admin Name" and "Admin email" fields. For both the Israeli (`.il`) and the Switzerland (`.ch`) domains, personal data was not filtered at all. The full names and addresses were visible in all domains that were not owned by a company. In the case of a company, the registrant name could be "Domain Admin", for example. In the Israeli domains, the phone numbers were also available. The US domains often had "REDACTED FOR PRIVACY" as a replacement for registrant fields. Still, a combination of email address, phone numbers and physical address was often available, either in the Registrant or in the Admin or Tech fields. As for the Korean domains, they were mostly owned by companies and had little personal information.

A large difference can be seen between TLDs belonging to a

[6] `https://github.com/DK-Hostmaster/whois-service-specification#request-4`

Table 1. Different domains and the percentage of domains containing personal data via WHOIS

TLD	# of domains	Personal Data
EU		
.dk	47	26%
.ee	10	0%
.eu	62	0%
.fi	11	18%
.fr	13	0%
.ie	14	0%
.it	74	46%
.nl	153	0%
.no	31	0%
.se	41	0%
.si	10	0%
.sk	14	0%
.uk	146	0%
Non-EU		
.ca	66	21%
.ch	13	77%
.co	19	0%
.il	13	62%
.in	48	0%
.kr	41	27%
.us	41	68%

country within the EU and outside the EU. Taking an average between TLDs is slightly complex. If you would do it naively – 153 .nl domains with 0%, and 11 .fi domains with 18% – you would get an average of 0,11% of domains containing personal data for .nl and .fi domains. This would mean that the .nl data is vastly overrepresented compared to the .fi data. However, as explained in section 3, it is reasonable to assume that 153 .fi domains would have similar results as 11 domains. Therefore, the average of .nl and .fi domains was taken as $\frac{0\%+11\%}{2} = 9\%$.

Via this method, EU domains had personal data on average on 6,89% with a standard deviation of 0,143. The non-EU countries had an average of 36,40% and had a standard deviation of 0,323. Since the data was distributed non-normally, a standard T-Test is not possible. Instead, a Mann-Whitney *U* test is done. The Mann-Whitney *U* test is a statistical test assessing whether two sampled groups are from a single population [7].

The hypothesis is that EU and Non-EU domains are from statistically different populations, concerning the amount of personal data. An additional hypothesis is that the Non-EU domains have a higher percentage of personal data, so we can use a single-sided *U* test. With the current samples, the U-value is 19. For a significance level of $p < .05$, the critical value of *U* is 24. Therefore, the result is significant at $p < .05$. So, Non-EU countries have significantly more personal data on average than EU countries.

5.3 Rate Limits

Table 2 shows the different rate limits enforced depending on the Top-Level Domain (or more precisely, the WHOIS server used of the TLD. For example, for .nl, the WHOIS server to communicate with is whois.domain-registry.nl).

The "Number" column refers to the number of requests that need to be made before being rate-limited. The "Duration" column is relevant when the type of limit is a slowing of the request and means the number of seconds by which the request is slowed. Finally the "Cool-Down" column is the time needed to wait after being rate-limited to "reset" the limit.

6. DISCUSSION

This section will cover the discussion of the research. This includes the limitations of the paper, but also an explanation of the found results.

6.1 Limitations

This research only looked at TLDs with a thick lookup model. A thin lookup model would mean that different registrars have their own rules. This would make drawing conclusions about an entire TLD not that useful any more. But there are some large domains which have a thin lookup model. These were left out in this study, as explained in section 1. For a future study, parsing and analysing thin TLDs might be interesting, but care must be taken to ensure that the parsing is done correctly. For this, inspiration of Liu et al. [6] can be taken.

Additionally, the research took the domains from a random list generated by OpenDNS. This list was not extremely extensive, as there were only 13 TLDs in the European Union with 10 or more domains. For future research, one might make an attempt at creating a larger dataset while ensuring that it remains representative.

There are also large datasets available with open access online, but these are not random or representative at all. One example of such a dataset is a list of malicious domains. These are domains which were flagged for phishing or malware. These are likely skewed since people hosting and registering these websites would have an incentive to enter false data. Datasets like Alexa Top Global Sites[7] also exist. These are datasets which contain the most popular websites of the internet. These popular websites would likely have less personal information than websites hosted or registered by single persons and therefore be invalid to draw conclusions from.

Therefore, gathering a large and random dataset is difficult to do. However, doing so might improve the results.

Finally, determining when the GDPR applies is challenging. When a registrar keeps and shows data about a citizen living in The Netherlands, for example, it is clear that the GDPR applies. This is also the case when a Swiss registrar shows data about a Dutch citizen. But, it does not apply when a Swiss registrar has data about a Swiss citizen. Therefore, it is difficult to draw conclusions about single domains. However, on average, EU registrars have to comply with the GDPR more often than Non-EU registrars, so on average conclusions can be made.

6.2 Improving WHOIS Parsing

As shown in results section 5.1, some changes were made to the open source library pywhois. This made the current research possible by significantly improving the parsing of some TLDs, but future research has multiple options, as explained in section 7.

6.3 Found Results

There is a clear difference between individual Top-Level Domains. Whereas the sample size of some TLDs is rather small, that is less of an issue since all domains within one TLD have to follow the same rules. This means that there is much less variation between domains within a TLD.

EU TLDs

Of the 13 TLDs in the EU, 3 contained personal information: .dk, .it and .fi. This ranged from just full names (.dk) to full names, addresses, phone numbers and email addresses (.it). The 13 TLDs combined contained personal information in 6,89% of the domains (StDev: 0,143). The TLD with the most personal information available is .it. However, the Italian top-level re-

[7] https://alexa.com/topsites

Table 2. Different domains and possible rate-limit methods their WHOIS server uses.

TLD	Type Limit	Number	Duration	Cool-Down	Notes
EU					
.dk	Slow	0	2s	-	Flat 2s delay for every request
.ee	Slow	5	8s	1m	
.eu	Ban	60		1m	
.fi	None	-	-	-	
.fr	Ban	2		10s	
.ie	Ban	100		24h	
.it	Slow	20	2+s	1m	After 20 request, start slowing requests. Delay keeps increasing
.nl	Ban	1		1s	Maximum of 1 request per second
.no	Ban	100		1m	
.se	None	-	-	-	
.si	Ban	10		1m	
.sk	Ban	50		1h	
.uk	Ban	25		5s	Bans large volume requests (~25), lifts ban after slowing requests
Non-EU					
.ca	None	-	-	-	
.ch	Ban	40		10m	
.co	Ban	100		1h	.co, .in and .us cooperate
.il	Ban	30		10m	
.in	Ban	100		1h	.co, .in and .us cooperate
.kr	None	-	-	-	
.us	Ban	100		1h	.co, .in and .us cooperate

gistry Registro did make some attempts to hide personal information, claiming that no personal Whois data is available unless the registrant has expressed consent to the publication of data [10]. This attempt to mask personal information, however, has clearly been ineffective.

Non-EU TLDs

Of the 7 TLDs outside the EU, 5 contained personal information. The (thin) TLDs that did not belong to a specific country (such as .com, .net, etc.) were left out of the analysis, as explained in section 3.

The 5 TLDs which contained personal information were .ca, .il, .ch, .kr and .us. Between these domains, there was a large variation in the types of personal data to be found as well. In total, the Non-EU TLDs had personal information in 36,40% of domains, with a standard deviation of 0,323.

One possible limitation of the dataset was the randomness of Korean domains. These domains were mostly owned by companies, and perhaps a more representative dataset would not have this issue.

Comparison

The comparison of different TLDs attempts to answer the third research question: *Is there a difference between the amount of personal data available via* EU TLD*s and Non-*EU TLD*s?*

As seen in section 5.2, TLDs of Non-EU countries have significantly more personal data in their domains than TLDs of EU countries. This means registries of countries which do not follow the GDPR are more likely to contain personal information about the registrants.

6.4 Summary

The second research question was *What personal data is still publicly accessible on* WHOIS*?*

As the research has shown, 8 out of 20 TLDs contained personal information on WHOIS. So, there is at least some personal information still available via WHOIS, mainly full names, addresses and phone numbers.

The fourth research question was *Is there a difference between the amount of personal data available within* EU TLD*s?*.

As seen in Table 1, there are some differences between TLDs within the European Union. Whereas 10 TLDs did not contain any personal information, 3 TLDs did. A future, larger dataset could include more of the 28 EU member states, but clear differences between member states can already be seen in this dataset.

6.5 Rate Limiting

In general, most domains had some form of rate limiting. These all banned or slowed a single IP address.

An interesting find was that .co, .in and .us (belonging to Colombia, India, and the United States, respectively) collaborate with rate limiting. For example, making a lot of requests on the .co WHOIS server (whois.nic.co) leads to being rate-limited on the Colombian WHOIS server, but also on the US server (whois.nic.us) even after no US queries were done.

7. FUTURE WORK

With a proper parser – such as one similar to Liu et al. [6] developed –, one might be able to do similar research on TLDs with a *thin* WHOIS lookup model. This would mean that there would be quite a lot of extra data to parse, from .com and .net for example.

Furthermore, a recommendation for a future study would definitely be to reproduce this research on a larger dataset. Care must be taken to ensure the dataset is representative – so popular datasets such as lists of malicious domains or Alexa Top Sites would not be suitable –, but if one can get a random and large dataset, that would be an improvement for future research.

8. CONCLUSION

This study looked at the prevalence of personal information found in WHOIS data. WHOIS is a protocol to store information about a domain, including the registrant of the domain. The registrant of the domain is the person, company or organization that requested the domain. A domain always belongs to a Top-Level Domain,

or TLD, such as `.nl`. In some cases, personal information resides in the publicly available WHOIS data. Everyone with an internet connection can request WHOIS data (see figure 2 for an example of WHOIS output).

In order to properly parse WHOIS data, the state of the art of WHOIS parsing first had to be improved. This is related to the first research question: *In what way can the current public way of parsing* WHOIS *data be improved?*
The open-source library `pywhois` was chosen to parse WHOIS, but some fixes had to be made. In total, 18 TLDs were improved. `pywhois` already parsed some data of 15 of those TLDs, but this parsing was either incorrect or incomplete. The other 3 TLDs had to be added from scratch. Some of the improved TLDs, however, did not have enough domain entries in the used dataset, so these were irrelevant for the final results. Section 5.1 shows some improvements added to `pywhois`.

In 2018, the EU General Data Protection Regulation, or GDPR was introduced. Since the GDPR, personal information from citizens from the European Union is protected. One of the effects is that personal WHOIS data cannot publicly be shown any more. A question which comes up when looking at WHOIS and GDPR is the second research question of this paper, *What personal data is still publicly accessible on* WHOIS*?*
In total, this research looked at 20 different TLDs. 13 of those were within the European Union, versus 7 outside. 3 of the 13 EU country TLDs had personal information available via WHOIS. `.it`, the country-code TLD for Italy, was the TLD with the most personal information of the TLDs within the EU. This personal information included full names, email addresses, physical addresses and phone numbers.
Of the 7 TLDs outside the EU, TLDs, 5 contained personal information. So yes, personal information is still publicly available via WHOIS.

This leads to the third Research Question: *Is there a difference between the amount of personal data available via* EU TLD*s and Non-*EU TLD*s?*
In this research, a significant difference was found between the number of domains containing personal information with the Non-EU TLDs and with the TLDs belonging to EU countries. Non-EU domains have personal information in 36,40% of the cases. Within EU TLDs, however, the figure is 6,89%. With the data gathered, Non-EU TLDs have more personal information than EU TLDs. This is significant with a significance level of $p < .05$.

The fourth research question is: *Is there a difference between the amount of personal data available within* EU TLD*s?*
As there were still countries whose top-level registry did not comply with the GDPR, there was also a difference between EU countries. 10 of the 13 EU TLDs had no personal information available via WHOIS, and 3 did. These were `.dk`, `.fi` and `.it`.

As a final note, countries belonging to the European Union have significantly less personal information available via WHOIS than Non-EU countries. This could mean that the GDPR has reached its goal – at least for WHOIS data: improving privacy of registrants in the EU.

References

[1] *2018 reform of EU data protection rules*. European Union, 4th May 2018. URL: `https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en` (visited on 12/05/2019).

[2] ICANN. *DNSSEC - What Is It and Why Is It Important?* 5th Mar. 2019. URL: `https://www.icann.org/resources/pages/dnssec-what-is-it-why-important-2019-03-05-en` (visited on 26/06/2019).

[3] ICANN. *Temporary Specification for gTLD Registration Data*. 25th May 2018. URL: `https://www.icann.org/resources/pages/gtld-registration-data-specs-en` (visited on 11/06/2019).

[4] ICANN. *Thick WHOIS*. 7th May 2019. URL: `https://www.icann.org/resources/pages/thick-whois-2016-06-27-en` (visited on 26/06/2019).

[5] Joanna Kulesza. 'Balancing Privacy and Security in a Multistakeholder Environment. ICANN, WHOIS and GDPR'. In: *Future of Europe: Security and Privacy in Cyberspace* (Dec. 2018). URL: `https://www.researchgate.net/profile/Tanja_Porcnik2/publication/330194322_Future_of_Europe_Security_and_Privacy_in_Cyberspace/links/5c3334ac299bf12be3b4cdcf/Future-of-Europe-Security-and-Privacy-in-Cyberspace.pdf#page=54` (visited on 04/05/2019).

[6] Suqi Liu et al. 'Who is. com?: Learning to parse whois records'. In: *Proceedings of the 2015 Internet Measurement Conference*. ACM. 2015, pp. 369–380.

[7] Patrick E McKnight and Julius Najab. 'Mann-Whitney U Test'. In: *The Corsini encyclopedia of psychology* (2010).

[8] Norid. *New data model - implementation project*. 6th June 2018. URL: `https://teknisk.norid.no/en/registrar/system/videreutvikling/data-model-implementation-project/` (visited on 26/06/2019).

[9] OpenDNS. *Public Domain Lists*. 6th Nov. 2014. URL: `https://github.com/opendns/public-domain-lists` (visited on 27/06/2019).

[10] Registro.it. *Registry Facts and Figures, News and Events of First Quarter of 2018*. 19th Aug. 2018. URL: `https://www.nic.it/en/news/2018/registry-facts-and-figures-news-and-events-first-quarter-2018` (visited on 27/06/2019).

[11] Karin Vink. *SIDN and the GDPR*. 27th Mar. 2018. URL: `https://www.sidn.nl/en/news-and-blogs/sidn-and-the-gdpr` (visited on 21/06/2019).