



WHOIS Contact Data Availability and Registrant Classification Study

A Study of the Effects of GDPR and ICANN Policy

by

Greg Aaron

Lyman Chapin

David Piscitello

Dr. Colin Strutt

Interisle Consulting Group, LLC

25 January 2021



Table of Contents

Table of Contents.....	2
Executive Summary.....	3
Sponsorship.....	4
Introduction and Purpose	5
Methodology.....	6
Sample Selection and Margin of Error	7
Contact Data Publication Category	8
Registrant Type Classification	9
GDPR Jurisdiction	12
Data File	13
Analysis and Findings	14
Contact Data Publication Category	15
Use of Privacy/Proxy Services	17
Proxy-Protected Registrants	17
Domain Registrant/User Type Classification.....	18
Reach of GDPR Based on Jurisdiction	20
GDPR Jurisdiction of Redacted Domains	21
Reach of GDPR Based on Jurisdiction <i>and</i> Legal Character	22
Contact Data Publication Category: Cybercrime Domains	23
Conclusion.....	25
Appendix A: TLD Representation in Sample Set	26
Appendix B: Registrar Representation in Sample Set	27
Appendix C: WHOIS Problems	28
About Interisle Consulting Group, LLC.....	30
About the Authors	30

Executive Summary

Domain names are one of the essential components of the Internet. *Domain name registration data* identifies who registered and controls the domain name, and provides technical information that makes the domain name function. This data has always been vital for a variety of legitimate purposes. For those reasons, registration data has always been available in WHOIS, a well-known “registration data directory service” (RDDS) publication system.

The General Data Protection Regulation (GDPR) was adopted by the European Union (EU) and took full effect on 25 May 2018. This important data protection and privacy law had significant impact on how domain name registration data could be published. In response, the Internet Corporation for Assigned Names and Numbers (ICANN) established a new policy, allowing registrars and registry operators to comply with GDPR by redacting (withholding) personally identifiable data from publication in WHOIS. This policy created a tension between ICANN’s two competing goals: allowing compliance with GDPR while also preserving the publication of non-protected data (and the utility of the WHOIS system itself) “to the greatest extent possible.”

To date, ICANN has not performed work to determine key facts about the effects of its policy, such as how much data has been redacted, or under what circumstances. Therefore, there has not been a factual basis for determining if or how the policy is having its intended effects. The purpose of the current study is to establish answers to those questions. This study uses previously established methodologies, so its results can be compared to the results of previous studies and provide historical continuity. It is our hope that this data can be used for fact-based discussion about what effects the GDPR and ICANN’s resulting policy have had.

The major findings of this study are:

ICANN’s GDPR-driven policy has led to the redaction of the contact data for most gTLD domains and has allowed registrars and registry operators to redact (withhold) much more contact data than is required by GDPR—perhaps five times as much as is necessary. While ICANN’s policy has generally protected the data that must be protected per GDPR, ICANN’s policy has also been used to conceal a much larger set of contact data that is *not* subject to GDPR.

1. Before the GDPR went into effect and ICANN changed its registration data policy, the actual identities of about 75.7% of gTLD domain registrants were available in WHOIS. The other 24.3% of domains were under privacy/proxy protection. (*page 15*)
2. At present, only 13.5% of domains have an actual registrant identified in WHOIS. Registrars and registry operators have used ICANN’s post-GDPR policy to redact contact data from 57.3% of all domains. Adding proxy-protected domains, this means that 86.5% of registrants cannot be identified via WHOIS. (*page 15*)
3. The use of privacy/proxy protection has increased over time—from 20.1% of all domains in 2013 to about 29.2% in 2020. (*pages 15, 17*)
4. About 23.1% of gTLD domains are covered by the GDPR’s jurisdictional reach. These are the domains for which the registrant, registrar, registry operator, or registry back-end provider is located in the EU. (*pages 20-21*)
5. However, the data of legal persons is not protected under GDPR. The data suggests that only around 11.5% of domains may belong to natural persons who are subject to GDPR. This 11.5%

may be the percentage of domains that is *necessary to protect* under GDPR. In contrast, registrars and registry operators have redacted contact data from 57.3% of all domains, or five times the amount that may be necessary. (*pages 22-23*)

6. Different registrars are making very different choices about how much contact data they redact. Some display the contact data of registrants outside of GDPR's jurisdictional reach. Others redact the data of all their registrants, whether GDPR applies to them or not. (*pages 15-16*)
7. The data indicates that domain registrars have significant control over (or effective influence on) whether their registrants' contact data is displayed in WHOIS. (*pages 15-17*)
8. More than half of the gTLD namespace—51.7%—is now controlled by unidentifiable parties. These are domains that cannot be attributed to a registrant or user, either via WHOIS or by examining their web site content. This is in contrast to the time before GDPR and ICANN's policy, where only 18% of domains were controlled by unidentifiable parties. (*pages 19-20*)
9. While 23.1% of all domains fall within the GDPR's jurisdictional reach, only 12.5% of domains had a registrant that resided in the EEA. By protecting personal data *processed within* the EU, the GDPR's reach extends protection beyond EU residents to a larger set of domains and registrants. (*pages 22-23*)

Sponsorship

This industry study was funded by contributions from the following organizations. All are participants in the ICANN community. This report is the work of Interisle Consulting Group, and Interisle is wholly responsible for its content. The sponsors had no input into the design of the study, and had no input into or influence on the findings or the content of this report.

Anti-Phishing Working Group (apwg.org)

CAUCE, the Coalition Against Unsolicited Commercial Email (cauce.org)

DomainTools (domaintools.com)

Facebook (facebook.com)

The ICANN Business Constituency (bizconst.org)

Microsoft (microsoft.com)

RiskIQ (riskiq.com)

Introduction and Purpose

Domain names are one of the essential components of the Internet. These identifiers allow web sites, email, and apps to function. *Domain name registration data* includes the information about who registered and controls the domain name, plus technical information that makes the domain name function. This data has always been vital for a variety of legitimate purposes: it establishes who has legal rights to the domain name, is a way of establishing trust online, provides a way to get in touch with domain owners to solve problems, and is highly useful data for fighting cybercrime. For these reasons, domain name registration data has always been available for lookup via WHOIS, a well-known “registration data directory service” (RDDS) system.

The General Data Protection Regulation (GDPR) was adopted by the European Union (EU) and took full effect on 25 May 2018. This important data protection and privacy law had a significant impact on how domain name registration data could be made available. The Internet Corporation for Assigned Names and Numbers (ICANN) has oversight over the generic top-level domains (gTLDs), such as .COM, .ORG, and .ONLINE. ICANN decides what parties manage gTLD registries, ICANN accredits registrars and licenses them to sell gTLD domains, ICANN maintains binding contracts with its registry operators and registrars, and ICANN has mechanisms for making new policies through a multi-stakeholder process.

Before GDPR, ICANN policy and contracts required that domain name contact information be made available via the WHOIS system, allowing anyone to look up the name and contact data of domain registrants and their administrative and technical contacts. But the GDPR restricted the publication of personally identifiable data belonging to natural persons located in the EU, and the data of any natural person whose data is processed within the EU. (A *natural person* is a human being. In contrast, a *legal person* is a non-human entity incorporated under the law of its jurisdiction, such as a company.)

In response, the ICANN Board of Directors established the *Temporary Specification for gTLD Registration Data*¹, effective on 25 May 2018. This “Temp Spec” was affirmed, without notable alterations, in May 2019 under the *Interim Registration Data Policy for gTLDs*². In this policy, ICANN requires that registrars and registry operators protect personal data covered under the GDPR. ICANN’s stated goal was to create a policy to “ensure compliance with the law while preserving the current information contained in WHOIS to the greatest extent possible.”³

Notably this policy also allows registrars and registry operators to redact (withhold) the data of *any* domain contacts they wish, including contact data that is *not* protected by GDPR or a similar data protection law. This created a tension between the two stated goals: protecting data that must be protected by law, but also continuing to make data available (and therefore preserving the utility of the WHOIS system itself) “to the greatest extent possible.”

This was the most significant policy that ICANN had put into place in years. To date, ICANN has not established the exact effects of its data policy, after several years of additional policy development

¹ <https://www.icann.org/resources/pages/gtld-registration-data-specs-en/#temp-spec>

² <https://www.icann.org/resources/pages/interim-registration-data-policy-en>

³ Blog by ICANN CEO Göran Marby, <https://www.icann.org/news/blog/data-protection-privacy-issues-update-an-icann-update-most-frequently-asked-questions>. In the *Temporary Specification* itself, this goal is stated in parallel language: “consistent with ICANN’s stated objective to comply with the GDPR, while maintaining the existing WHOIS system to the greatest extent possible.”

process stretching from 2018 into 2021.⁴ Specifically, ICANN has not established how much data is being redacted, or under what circumstances. Therefore, there has not been a good factual basis for determining the effects of its policy and whether it is having its intended effects.

The purpose of this study is to establish factual answers to those and related questions, using previously established methodologies. By using similar definitions and methods, the results in this study can be compared to the results of previous studies and provide historical continuity. It is our hope that this data can be used as a basis for fact-based discussion and decision-making.

ICANN policy is applicable to generic top-level domains (gTLDs) only. It does not apply to country-code domains (ccTLDs). ccTLDs are beyond the scope of this study.

Methodology

To the extent possible, this study used methodologies established in previous studies that were commissioned by ICANN. The methodologies provide objective and statistically valid results, provide continuity, and allow for comparisons over time. Those key studies were:

- The study most similar to the present one is the “WHOIS Registrant Identification Study”⁵ of 2013, and its accompanying “Terms of Reference for WHOIS Registrant Identification Studies”⁶ methodology document. These were designed and executed for ICANN by NORC at the University of Chicago, an independent, non-partisan research institution. They provide methodologies for establishing whether a domain is sponsored by a natural or a legal person, what a privacy/proxy registration is, establishing the percentage of privacy/proxy use among legal persons, and so on.
- Entity classification and privacy/proxy service identification methodologies developed in:
 - The “Study of the Accuracy of WHOIS Registrant Contact Information”⁷, developed for ICANN by NORC at the University of Chicago in 2010.
 - “ICANN Study on the Prevalence of Domain Names Registered using a Privacy or Proxy Service among the top 5 gTLDs”⁸, conducted by ICANN Organization in 2010.
 - The “WHOIS Privacy and Proxy Abuse Study”⁹ and its terms of reference¹⁰ document, developed for ICANN by the National Physical Laboratory of the U.K. in 2014.
 - The “WHOIS Proxy/Privacy Reveal & Relay Feasibility Survey”¹¹, developed for ICANN by Interisle Consulting in 2012.

⁴ The “GNSO Expedited Policy Development Process (EPDP) on the Temporary Specification.” This policy-making process began in 2018 and is continuing as of this writing, into a third phase. See <https://www.icann.org/public-comments/epdp-recs-2019-03-04-en> and <https://gnso.icann.org/en/group-activities/active/gtld-registration-data-epdp-phase-2>

⁵ https://gnso.icann.org/sites/default/files/filefield_39861/registrant-identification-summary-23may13-en.pdf

⁶ https://gnso.icann.org/sites/default/files/filefield_24703/tor-whois-registrant-id-studies-20may11-en.pdf

⁷ <https://www.icann.org/en/system/files/newsletters/whois-accuracy-study-17jan10-en.pdf>

⁸ <https://www.icann.org/en/system/files/newsletters/privacy-proxy-registration-services-study-14sep10-en.pdf>

⁹ <http://whois.icann.org/sites/default/files/files/pp-abuse-study-final-07mar14-en.pdf>

¹⁰ https://gnso.icann.org/sites/default/files/filefield_12392/whois-proxy-abuse-study-18may10-en.pdf

¹¹ https://gnso.icann.org/sites/default/files/filefield_35963/whois-pp-survey-final-report-22aug12-en.pdf

Sample Selection and Margin of Error

NORC's "WHOIS Registrant Identification Study" of 2013 used a data set of 1,600 domains, designed to provide a margin of error at the 95% confidence level of $\pm 5\%$. The NORC study used domains from only five top-level domains (.COM, .NET, .ORG, .INFO, and .BIZ) and was conducted before a thousand-plus new top-level domains were introduced by ICANN beginning in late 2013.

We began by downloading and collating the June 2020 ICANN registry reports for all gTLDs, which were the latest available.¹² This allowed us to determine the size of the gTLD namespace as a whole (a population size of 219,603,320 domains total), the size of every gTLD, and how many domains were sponsored by each registrar.

For the current study we used a set of 3,000 domains. This larger set provides a 98% confidence level with a margin of error of $\pm 2\%$. It also allows a proportional estimate's margin of error at the 95% confidence level of $\pm 5\%$ for any subgroup with 400 or more domains. Our larger set also allowed us to include new top-level domains to represent the expanded namespace.

Following the NORC methodology, we determined how many domains from the study set would be drawn from each gTLD, in order to reflect their share of the gTLD namespace. Our study set contains domains from the 19 largest gTLDs, plus an additional four chosen from the next-largest new gTLDs. The details are presented in *Appendix A: TLD Representation in Sample Set*.

- Together, the 23 TLDs contain 20.6 million domains, or 95.0% of all registered gTLD domains. The selection pool was therefore representative of the gTLD space as a whole.
- The selection of TLDs ensured that the study set contains gTLDs administered by each of the major registry operators, and at least two TLDs serviced by the major back-end providers. This allows the study set to contain domains administered in different geographic regions, including some located in the European Union and therefore subject to GDPR. The selection is also useful because different operators and back-end providers make different choices about how to serve WHOIS data.
- The new gTLDs represent about 13.8% of the gTLD market (30.2 million out of 219.8 million domains). To allow the inclusion of new gTLDs, they are slightly over-represented in our study set, which contains 516 nTLD domains in the study set, or 17.2% of the study set. This was at the expense of .COM, which is slightly under sampled in comparison to its market share; a similar under sampling of .COM was also done in the NORC study.
- Each gTLD had at least 20 domains in the study set. A larger number was not necessary because this study does not seek to make detailed comparisons across gTLDs.

We then obtained the zone files for the 23 gTLDs in early October 20120 and picked the requisite number of domains randomly from each gTLD's zone file. The randomness was obtained by using the PHP library routine `mt_rand`, which generates a random value via the Mersenne Twister Random Number Generator.

We then obtained WHOIS for the 3,000 domains from each registry's port 43 server. This told us which registrar sponsored each domain. We next queried each registrar's port 43 server for its domains. (Domain contact data is primarily held at the registrars.) For the domains registered before April 2018

¹² Monthly registry reports: <https://www.icann.org/resources/pages/registry-reports>

(when registrars began implementing data redaction allowed by ICANN), we also obtained historical WHOIS from DomainTools.

We then evaluated the study set to see if it reasonably reflected gTLD market share by registrar. When randomly drawing 3,000 domains from the pool of 209 million, there can be “draws” that are statistically anomalous. For example, it is possible (but statistically unlikely) to perform a draw in which none of the selected 3,000 domains are sponsored by GoDaddy, which owns 30% of the gTLD market. We determined that the registrars’ proportions of our “draw” reasonably tracked their proportions of market share. There was only one large registrar where its percentage representation in the study set varied from their market share by 0.5%. We were therefore satisfied that the randomly selected study set adequately reflects registrar gTLD market share, and again is representative the gTLD space as whole. For details, please see *Appendix B: Registrar Representation in Sample Set*.

The WHOIS data collection took place immediately after we performed the random selection; our web site visits took place after that, in October and November 2020. Only a handful of domains expired after the WHOIS data was collected but before we performed site visits, and we compensated for those cases by referring to snapshots of those sites at the Internet Archive and in Google cache.

The collected WHOIS data was parsed and, in some cases, manually processed to prepare the data for analysis. There were cases in which registrars withheld data from port 43 WHOIS but we were able to gather it by hand using the registrars’ web-based WHOIS services. Some registrars failed to provide any WHOIS data on their port 43 or their web-based WHOIS lookup pages. In these cases, we counted the data as redacted.

Contact Data Publication Category

Contact data publication category describes what *kind and amount of contact data* is published in WHOIS for a given domain name. In earlier studies, domains only fell into two categories: privacy/proxy, or full contact publication available in WHOIS. However, developments since 2018 have created a third category. What category a domain is in depends on choices made by the registrant and/or its registrar. The three categories are:

- 1. Contact data available:** Contact data for the domain registrant is available in WHOIS and is not privacy/proxy data. The name of the registrant is provided, and there is street address data (and usually phone and email address) data present. Before mid-2018, about 75.7% of gTLD domains were in this category.
- 2. Privacy/proxy:**
 - *Privacy services* offer alternate WHOIS contact information and mail forwarding services, but do not shield the Registered Name Holder’s identity.
 - *Proxy services* register domain names on a third party’s behalf and then license their use so that the provider’s identity and contact information (and not the licensee’s) is published in WHOIS. With proxy services, the identity of the “real” registrant is not revealed.
 - Per ICANN policy, the name and the full contact data for a privacy/proxy service provider must be published in WHOIS, including the provider’s name, street address, and telephone number. That allows domains under proxy protection to be identified reliably.

- In the past, obtaining the actual domain user's identity from WHOIS during any study was likely for Privacy registrations, but not for domains registered by a Proxy provider.¹³ However, it now seems more difficult to identify Privacy registrations. This is because Registrant Name must be redacted in cases when GDPR applies, and because ICANN policy allows registrars to redact Registrant Name at will.¹⁴ It is possible that Privacy registrations are being redacted/masked in WHOIS, making them indistinguishable from other domains with redacted data.
 - Before mid-2018, about 24.3% of gTLD domains were in the privacy/proxy category.
- 3. Contact data redacted:** The contact data is redacted (withheld) from publication in WHOIS, as allowed by ICANN's *Temporary Specification*¹⁵ and *Interim Registration Data Policy for gTLDs*¹⁶. Under those policies, since May 2018, gTLD registrars and registry operators are required to redact personally identifiable data (the contact's name, street address, phone number, and email address) from contact data IF:
- the registrant is in the European Economic Area (EEA), or
 - the registrar or registry operator is in the EEA, or
 - the registrar or registry operator is outside the EEA but processes the data within the EEA. This situation occurs when a registry operator uses a back-end service provider located in the EEA.
- gTLD registrars and registry operators are *allowed* to redact personal data for any other domains they wish, including those not covered by GDPR or any other privacy law.
 - Per the policies above, registrars and registry operators who redact data from a domain record must either publish text in the redacted WHOIS fields substantially similar to "REDACTED FOR PRIVACY", may provide no information in the value section of the redacted field, or may not publish the redacted field at all. Therefore, domains with redacted data can be clearly distinguished from domains in the "Contact Data Available" and "Privacy/Proxy" categories.
 - The fields that can be redacted include Registrant Name, the Street Address and Postal Code fields, and Telephone Number. ICANN policy requires that the State and Country fields and the data in them must *always* be published in WHOIS, because that data is not personally identifiable data. Some registrars show data in the Registrant Organization field if the registrant supplied it, but publication of data in that field is not required by ICANN policy.

Registrant Type Classification

In its 2013 study, NORC used WHOIS data to classify registrants into different types. Domain name registration data is the authoritative record of what party is the *registrant* of record, also known as the "registered domain holder." NORC examined the WHOIS data to assign each registrant an *Apparent*

¹³ Definitions from revised "Terms of Reference for WHOIS Registrant Identification Studies", https://gnso.icann.org/sites/default/files/filefield_24703/tor-whois-registrant-id-studies-20may11-en.pdf

¹⁴ See Appendix A, paragraph 2.6 at <https://www.icann.org/resources/pages/gtld-registration-data-specs-en/#temp-spec>

¹⁵ For the specification, see Appendix A, at <https://www.icann.org/resources/pages/gtld-registration-data-specs-en/#5>

¹⁶ <https://www.icann.org/resources/pages/interim-registration-data-policy-en>

Registrant Type. NORC then examined the web sites on the domains to assign each domain to an *Apparent Domain User Type*. The domain *user* is what party makes use of the web site and evidently controls it. For example, the International Committee of the Red Cross is the domain user of ICRC.ORG, as is evident from the web site.

In most cases, NORC found that the registrant of record and the site user were one and the same. This is logical and unsurprising. NORC found that “there is a strong relationship between apparent registrant type and apparent domain user type, with a p-value for the relationship of less than .0001.”

The registrant or record and the site user can technically be different, but for the purposes of this study the difference may not matter. For example, Company X may hire a web design firm to create a web site for it. Sometimes Company X may be listed as the registrant in WHOIS and is the user of its domain too. Sometimes the web design firm may list itself as the registrant, and Company X is the site user. However, both parties are companies (*i.e.*, legal persons), and so the difference does not matter when assigning registrant type.

Apparent Registrant Type is no longer useful to break out as a separate category for analysis, because in 2020 identifying contact data was available for only a small percentage of domain names. However, some registrants are still clearly identified in WHOIS, and that data can be used to classify their Apparent Registrant Type.

We performed analysis following the methodologies and categories defined in ICANN’s *Revised Terms of Reference for WHOIS Registrant Identification Studies* and used in NORC’s 2013 study. Most of this work consisted of visiting each domain to see if it has a web site, and if so whether the web site reveals the identity of the user. The registrant types are:

- natural person
- legal person
- unclassified

Note that a domain can be privacy/proxy protected, or can have its contact data redacted, but the domain’s web site can reveal the domain user’s identity, and whether it is a natural or legal person.

A *natural person* is a human being. In contrast, a *legal person* is a non-human entity incorporated under the law of its jurisdiction. The distinction is important for ICANN policy, which was designed to allow compliance with GDPR. GDPR is not applicable to the data of legal persons. The text of the GDPR says:

The protection afforded by this Regulation should apply to natural persons, whatever their nationality or place of residence, in relation to the processing of their personal data.

This Regulation does not cover the processing of personal data which concerns legal persons and in particular undertakings established as legal persons, including the name and the form of the legal person and the contact details of the legal person.^{17, 18}

¹⁷ GDPR Recital 14. <https://gdpr.eu/recital-14-not-applicable-to-legal-persons/>

¹⁸ In keeping with the distinction made in the GDPR, the European Commission’s “Proposed Directive on Measures for a High Common Level of Cybersecurity across the Union” states: “TLD registries and the entities providing domain name registration services for them should make publically available domain name registration data that fall outside the scope of Union data protection rules, such as data that concern legal persons.”

Accordingly, we coded the domain user as a *legal person* if one of these cases applied:

- The party identified itself as a company in WHOIS and/or on its web site. In most cases the entity did so with a specific legal form (*Inc., LLC, Limited, Pty, LLP, etc.*), or displayed a business registration number.
- The party is an incorporated not-for-profit entity. For example, we encountered U.S. 401(c)3 entities, churches in dioceses, and U.K. registered trusts and charities.
- The party is a government entity (local, state, or national).
- In some cases, the legal character of the party or its exact legal form was not mentioned in WHOIS data or on the web site, but we were able to confirm it via a reputable database (such as Dun & Bradstreet, the U.K.'s Companies House, or the relevant Secretary of State database), from historical WHOIS data, or from the entity's own social media accounts, often linked to from its web site.

We coded the domain user as a *natural person* if:

- The site was devoted to personal content – for example blogs, wedding sites, hobbies, etc. Or,
- The party appeared to be an unincorporated sole proprietor.

We coded the domain user as *unclassified* if:

- We could not satisfactorily establish the identity of the registrant or user. In most of these cases there was no real registrant data to rely upon, and the web site did not feature any content that could be evaluated.
- We were unable to establish the exact legal character of the registrant. For example, we encountered sites engaged in business activity, but we could not determine whether the owner/user is incorporated or not. This was the case with some individuals engaged in professional activities – for example photographers using their sites to display their professional portfolios, yoga instructors, etc. – who may be incorporated or may be sole proprietors.
- All 3,000 domains were visited, including those marked as “unclassified.”

Following the methodology of the NORC study, our analysts documented rules to be uniformly applied during the site visits, independently classified some of the same sites and then compared the results in order to reconcile discrepancies and reinforce uniformity. They used Google Translate to examine sites in languages other than English. When the analysts visited each domain, they made a note describing what they found. These notes are included in the accompanying data file.

Some legal person registrants include the data of natural persons in their domain name registration records. There are various policy and implementation options for addressing that issue, but they are not a subject for this study. Unfortunately, it is no longer possible to determine how many legal person registrants put the data of natural persons in their domains name records, because ICANN's policy has allowed registrars to redact a great deal of contact data out of WHOIS records. Studying that issue would require a corpus of data obtained from the registrars. That process involves challenges beyond the scope of this study.

<https://ec.europa.eu/digital-single-market/en/news/proposal-directive-measures-high-common-level-cybersecurity-across-union>

We did not code domains as belonging to a natural person based *solely* on WHOIS data. This is because we found examples where the WHOIS Registrant Name and/or Registrant Organization fields list only the name of a natural person, but the site is clearly operated by/dedicated to a legal person. In some such cases we established that the listed natural person was an employee of the legal person site user. Some of these are cases in which the registrant probably should have listed the legal organization as the registrant of legal record, but did not know to, or the registrar did not provide guidance about how to fill in the fields. Also, some registrars decline to publish the Registrant Organization field or the data in it (which is a choice allowed by ICANN policy), and that removes a vital piece of data from consideration.

Our study therefore *reliably establishes a minimum or floor* for the percentage of domains that are registered by/used by legal persons, and by natural person registrants/users.

Overall the methodology is conservative, allowing classification as a legal or natural person only when there is evidence. It acknowledges that many domains may be unclassifiable based on a lack of evidence. We followed the NORC study's principles: "registrant type was based on the evidence that we were able to discover during our investigation, that is, each domain registrant was coded based on what was apparent in the information we found in the dataset. No attempt was made to verify WHOIS accuracy or contact the identified registrant."

We did not use commercial activity as a decisive classification criterion. We did not attempt to classify "business structure" (whether legal persons were for-profit or not-for-profit). Whether a legal entity is for-profit or not-for-profit, or whether a natural person or legal person is engaged in commercial activity or not, do not matter for the purposes of the GDPR or under current ICANN policy.

GDPR Jurisdiction

To measure the effect of the GDPR and ICANN's resulting data policy, we classified whether each domain's registration data is subject to GDPR based on whether a relevant party is within the jurisdiction of the EEA. There are three qualifying cases codified in ICANN's Temporary Specification.¹⁹ The data is subject to GDPR if any of the following cases apply:

1. **The Registrant is located in the European Economic Area (EEA)**²⁰. We based this on the Registrant Country field in WHOIS, which is mandatory for registrars to provide.
2. **The domain's registrar or registry operator is established within the EEA.** We determined this by referring to ICANN's official lists.²¹
3. **The domain's registrar or registry operator is established outside the EEA but processes the data within the EEA.** We therefore included registries that use a *back-end service provider* located in the EEA. A back-end provider provides infrastructure for the registry, including the registry system that processes and holds the domain data, and provides the WHOIS service. An example is back-end provider CentralNIC, which is in the United Kingdom and provides the

¹⁹ Appendix A, section 2. <https://www.icann.org/resources/pages/gtld-registration-data-specs-en/#appendixA>

²⁰ The European Economic Area (EEA) consists of the Member States of the European Union (EU), plus three countries of the European Free Trade Association (EFTA): Iceland, Liechtenstein, and Norway. Although the United Kingdom has "Brexit" the European Union, we counted the U.K. as being under GDPR because the U.K. followed GDPR in 2020, and the U.K. also has its own parallel data protection act that is modelled directly on the GDPR.

²¹ <https://www.icann.org/registrar-reports/accreditation-qualified-list.html> and <https://www.iana.org/domains/root/db>

infrastructure for the .XYZ registry operator of record, which is incorporated in the United States.²²

A few domains may have had resellers in the EEA, but the registrant, registrar, registry, and back-end provider were all outside the EU. The number of such domains is probably very small. This is not possible to measure because most registrars choose not to identify their resellers in WHOIS records, as is allowed by ICANN contract.

This study focusses on the reach of the GDPR, which is the major use case that affects the most domain names, and for which ICANN policy was largely developed to address. There are some jurisdictions outside the EEA that may require the protection of personal data in RDDS. However, which jurisdictions those are, and how many domains may be affected, have not been explored or catalogued by ICANN or its EPDP efforts.²³

Data File

In the interest of transparency, the data set is publicly available in an Excel data file. One tab contains a row for each of the 3,000 domains in the study set. For every domain, we have provided the following data, among others:

- Registry WHOIS output (parsed by field)
- Registrar WHOIS output (parsed by field)
- Coding for contact data publication category
- Registrant (user) type classification, and notes
- Whether the domain's registrant, registrar, registry, or registry back-end provider is in an EEA country.

A second tab contains statistics derived from the data sheet. A third tab contains the data for the phishing domains analyzed in this report.

Please note that the domains in the study set feature the breadth of content one may encounter on the web; some sites are “not suitable for work” and some led to malware and other harmful content.

²² <https://www.iana.org/domains/root/db/xyz.html>

²³ For example, the California Consumer Privacy Act does not prevent the publication of personal data in RDDS. At ICANN there has not been a professional evaluation of if or now the Russian Federal Law on Personal Data would prohibit the publication of natural personal data in gTLD RDDS; the Russian law apparently does not protect the data of legal persons.

Analysis and Findings

Analysis indicates that **ICANN's GDPR-driven policy has led to the redaction of the contact data for most gTLD domains and has allowed registrars and registry operators to redact (withhold) much more contact data than is required by GDPR—perhaps five times as much as is necessary. While ICANN's policy has generally protected the data that must be protected per GDPR, ICANN's policy has also been used to conceal a much larger set of contact data that is *not* subject to GDPR.**

1. Before the GDPR went into effect and ICANN changed its registration data policy, the actual identities of about 75.7% of gTLD domain registrants were available in WHOIS. The other 24.3% of domains were under privacy/proxy protection. *(page 15)*
2. At present, only 13.5% of domains have an actual registrant identified in WHOIS. Registrars and registry operators have used ICANN's post-GDPR policy to redact contact data from 57.3% of all domains. Adding proxy-protected domains, this means that 86.5% of registrants cannot be identified via WHOIS. *(page 15)*
3. The use of privacy/proxy protection has increased over time—from 20.1% of all domains in 2013 to about 29.2% in 2020. *(pages 15, 17)*
4. About 23.1% of gTLD domains are covered by the GDPR's jurisdictional reach. These are the domains for which the registrant, registrar, registry operator, or registry back-end provider is located in the EU. *(pages 20-21)*
5. However, the data of legal persons is not protected under GDPR. The data suggests that only around 11.5% of domains may belong to natural persons who are subject to GDPR. This 11.5% may be the percentage of domains that is *necessary to protect* under GDPR. In contrast, registrars and registry operators have redacted contact data from 57.3% of all domains, or five times the amount that may be necessary. *(pages 22-23)*
6. Different registrars are making very different choices about how much contact data they redact. Some display the contact data of registrants outside of GDPR's jurisdictional reach. Others redact the data of all their registrants, whether GDPR applies to them or not. *(pages 15-16)*
7. The data indicates that domain registrars have significant control over (or effective influence on) whether their registrants' contact data is displayed in WHOIS. *(pages 15-17)*
8. More than half of the gTLD namespace—51.7%—is now controlled by unidentifiable parties. These are domains that cannot be attributed to a registrant or user, either via WHOIS or by examining their web site content. This is in contrast to the time before GDPR and ICANN's policy, where only 18% of domains were controlled by unidentifiable parties. *(pages 19-20)*
9. While 23.1% of all domains fall within the GDPR's jurisdictional reach, only 12.5% of domains had a registrant that resided in the EEA. By protecting personal data *processed within* the EU, the GDPR's reach extends protection beyond EU residents to a larger set of domains and registrants. *(pages 22-23)*

Contact Data Publication Category

In early 2018, before the GDPR went into effect and ICANN changed its data policy, about 75.7% of domains had a registrant that was identified in WHOIS; the rest were privacy/proxy-protected. As of late 2020, only 13.5% of domains had a registrant who was identified in WHOIS. The other 86.4% of domain records had redacted contact data or were privacy/proxy-protected.

	2010 ²⁴	2013 ²⁵	April 2018 ²⁶	2020
Contact data available	82%	79.9%	75.7%	13.5%
Privacy/proxy	18%	20.1%	24.3%	29.2%
Contact data redacted	n/a	n/a	n/a	57.3%

In late 2020:

- 13.5% of domains had full registrant contact data available in either the registrar or registry WHOIS (405 out of 3,000). Of those 405 domains, 66 featured contact data because the domain was for sale, and the registrant used WHOIS to advertise the sale.
- 29.2% of domains were under a privacy/proxy service (877 of 3,000).
- 57.3% of domains had their contact data redacted (1,718 of 3,000).

ICANN’s policy resulted in a reduction of the availability of contact data from 75.7% in early 2018 (before GDPR) to 13.5% by late 2020.

Of the twenty largest registrars, thirteen rarely make full registrant contact data available in WHOIS. The majority make it available less than 3% of the time. Of the twenty largest registrars, only two (1&1 IONOS SE and OVH sas) are headquartered in the EEA:

	Registrar	gTLD domains under management	Domains in study	Domains with contact data available	Contact data available %
1	GoDaddy.com, LLC	63,311,867	882	6	0.7%
2	NameCheap, Inc.	10,497,555	152	8	5.3%
3	Tucows Domains Inc.	10,263,753	131	3	2.3%
4	Network Solutions, LLC	6,902,784	100	48	48.0%
5	GMO Internet, Inc. d/b/a Onamae.com	5,238,919	85	26	30.6%
6	Alibaba Cloud Computing (Beijing) Co., Ltd.	6,643,000	81	0	0.0%

²⁴ Source: NORC, University of Chicago, “ICANN Study on the Prevalence of Domain Names Registered using a Privacy or Proxy Service among the top 5 gTLDs <https://www.icann.org/en/system/files/newsletters/privacy-proxy-registration-services-study-14sep10-en.pdf>

²⁵ Source: NORC, University of Chicago, “WHOIS Registrant Identification Study”

²⁶ This is Interisle’s calculation. In our study set of 3,000 domains, 1,420 had been registered before April 2018 (when registrars began preparing for the May 2018 GDPR deadline); 345 out of the 1,420 domains were proxy-protected. This calculation has a 95% confidence level and a margin of error of ±5%. The historical data was obtained from DomainTools.

	Registrar	gTLD domains under management	Domains in study	Domains with contact data available	Contact data available %
7	Alibaba Cloud Computing Ltd. d/b/a HiChina ²⁷	5,200,971	81	19	23.5%
8	eNom, LLC	5,258,420	80	1	1.3%
9	Google LLC	5,020,195	78	0	0.0%
10	PDR Ltd. d/b/a PublicDomainRegistry.com	4,892,251	72	45	62.5%
11	1&1 IONOS SE	4,876,367	58	0	0.0%
12	West263 International Limited	3,306,729	56	0	0.0%
13	Wild West Domains, LLC	2,764,124	41	0	0.0%
14	Chengdu West Dimension Digital Technology Co	3,431,766	38	11	28.9%
15	FastDomain Inc.	2,341,325	37	16	43.2%
16	Xin Net Technology Corp.	3,506,805	37	0	0.0%
17	NameSilo, LLC	3,463,466	35	1	2.9%
18	Eranet International Ltd (TodayNIC)	1,946,035	31	0	0.0%
19	Dynadot, LLC	2,349,300	30	4	13.3%
20	OVH sas	2,308,970	28	0	0.0%

Several large registrars make notably higher percentages of contact data available. The numbers demonstrate how registrars are making very different choices under ICANN policy, even when they do business under similar circumstances. For example:

- Both GoDaddy and Network Solutions are based in the U.S., have registrant bases that are U.S.-centric and mostly outside the EEA, and both cater to small-and-medium sized businesses. GoDaddy has adopted a blanket redaction strategy, redacting contact data for all registrants everywhere (except when registrants opt into publication).²⁸ In contrast, Network Solutions seems to redact all data for registrants in the EEA (both legal and natural persons) but publishes the contact data of registrants outside the EEA.
- Both Tucows and PDR are based outside the EEA, run reseller-based businesses, and their registrants are mostly outside the EEA. Tucows has adopted a blanket redaction strategy, redacting contact data for all registrants everywhere (except when registrants opt into publication).²⁹ In contrast, PDR seems to redact all data for registrants in the EEA (both legal and natural persons) but publishes the contact data of registrants outside the EEA.

The data indicates that registrars have significant control over (or effective influence on) what Contact Data Publication Category its registrants fall into.

For details, see the analysis table in the accompanying data file.

²⁷ The domains sponsored by this registrar were mainly in new gTLDs, and the registrar does not serve WHOIS for new gTLDs. These registrants are identified in registry WHOIS.

²⁸ <https://domainnamewire.com/2020/06/08/big-news-godaddy-starts-redacting-whois-information/>

²⁹ <https://www.tucowsdomains.com/help/whois-use-and-information/tiered-access-directory-gated-whois/>

Use of Privacy/Proxy Services

The use of privacy/proxy services apparently rose after the GDPR and ICANN's Temporary Specification went into effect—from perhaps 24.3% in early 2018 to 29.2% in late 2020.

Proxy protection possibly rose because some registrars began giving it to their customers for free. A prominent example is NameCheap, the industry's second-largest registrar. NameCheap began giving away its WhoisGuard service for free in May 2018, the month that GDPR went into effect.³⁰ Top-twenty registrars Google Domains³¹, NameSilo³², 1&1 IONOS SE³³, and Dynadot³⁴ also give away proxy protection for free, usually as the default at registration.

Several of the industry's largest registrars now have proxy protection on the majority of the gTLD domains they sponsor. Examples from the data set include:

- NameCheap, Inc.: 86.8% proxy-protected (132 out of 152 domains).
- Google LLC: 98.7% proxy-protected (77 out of 78 domains).
- NameSilo: 97.1% proxy-protected (34 out of 35 domains).
- GMO Internet Inc.: 69.4% proxy-protected (59 out of 85 domains).

Again, the data highlights that registrars have significant control over (or effective influence on) what Contact Data Publication Category its registrants fall into.

The numbers suggest that some registrars are using proxy protection as a form of effective GDPR contact redaction, achieving the same goal by another means.

Virtually all of the protected domains we identified were under *proxy* services. We were unable to clearly identify more than a few *privacy*-protected domains, where the registrant's real name was identified in the contact data.

For complete by-registrar statistics, see the analysis table in the accompanying data file.

Proxy-Protected Registrants

The study set contained 877 proxy-protected domains. Of those:

- 28.4% (249) of the domain users identified themselves on their web sites.
- 175 were legal persons, as confirmed by their web sites.
- 56 were natural persons, as confirmed by their web sites.
- 559 did not resolve or had no content to evaluate (parked, etc.), and therefore could not be classified as natural or legal persons.
- 18 identified themselves on their web sites but we could not classify them as natural or legal persons.

³⁰ <https://www.namecheap.com/blog/free-whoisguard-forever/>

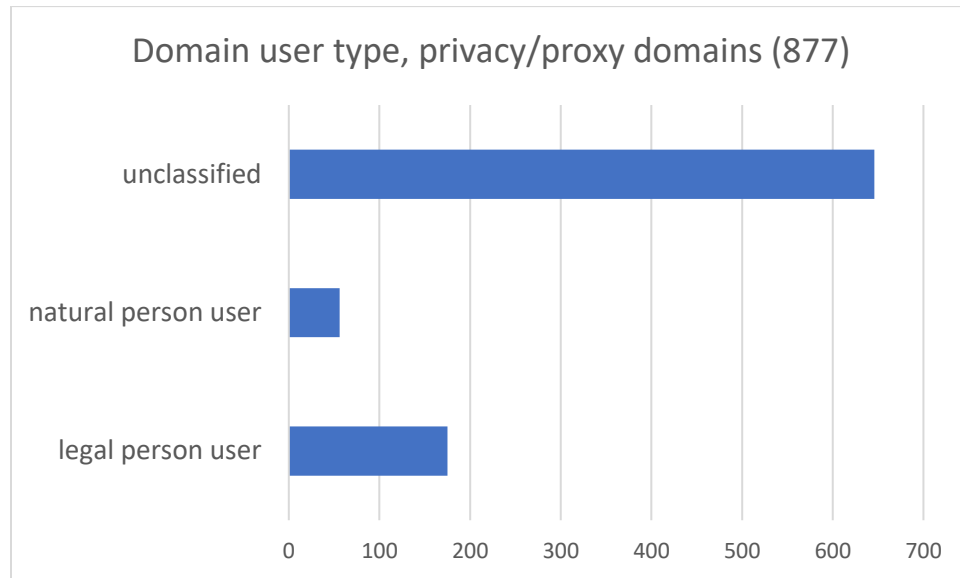
³¹ <https://support.google.com/domains/answer/3251242?hl=en>

³² <https://www.namesilo.com/Support/WHOIS-Privacy>

³³ <https://www.ionos.com/help/domains/preventing-spam-with-private-domain-registration/enabling-private-registration-for-a-11-ionos-domain/>

³⁴ <https://www.dynadot.com/domain/privacy.html>

- 67 had content on their web sites but the users did not identify themselves. The majority of these were dedicated to gambling or pornography.



In most cases the legal persons identified themselves on their web sites with specific contact data such as street address and phone number. Many gave the names of company personnel. Many included links to their social media accounts, especially at LinkedIn, Facebook, Instagram, and Twitter.

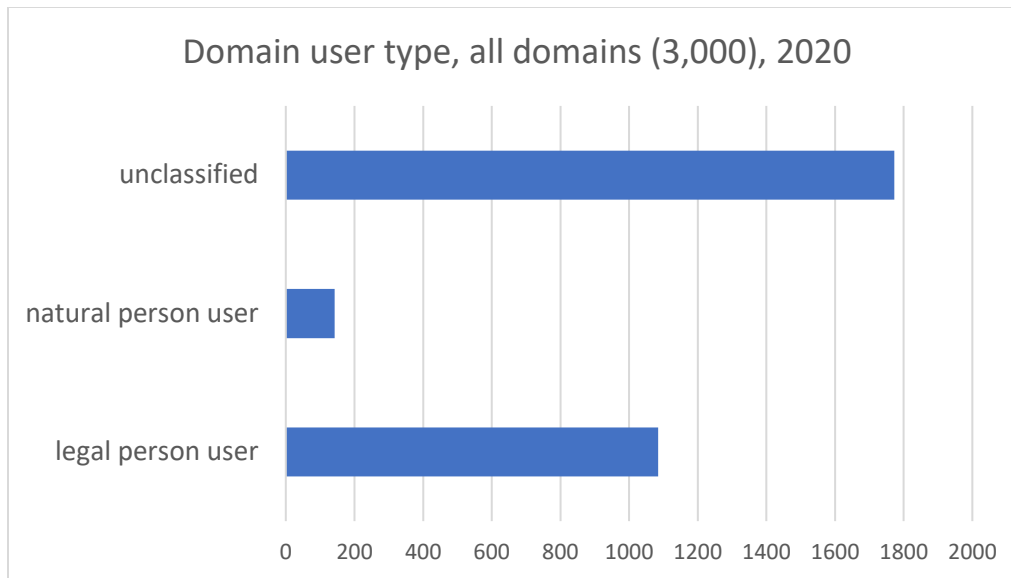
Domain Registrant/User Type Classification

After visiting all 3,000 domains and examining the WHOIS contact data that was available:

- 36.3% of domains were attributed to legal persons. (1,088 out of 3,000.)
- 4.7% of domains were attributed to natural persons. (142 out of 3,000.)
- The remaining 59.0% were unclassified as either legal or natural persons. (1,770 out of 3,000.) In these cases, the registrant/user was unidentifiable, or the registrant/user was identifiable but its legal/natural status could not be determined.

Our findings closely matched NORC's findings from 2013, and are well within NORC's margin of error:

Domain User Type	2013 (NORC)	2020 (Interisle)
Legal person	36.6%	36.3%
Natural person	5.4%	4.7%
Unclassified	56.0%	59.0%



Why were so many domains unclassified? In the set of 3,000 domains, 1,551 domains (51.7%) were unclassifiable because they had no web site content to evaluate, *and* their registrant data was redacted or under proxy protection. Those unclassified domains were:

- 638 domains did not resolve (NXDOMAIN). (21.3% of all 3,000 domains)
- 635 resolved to parking pages. (21.2% of all domains)³⁵
- 152 resolved but offered no content to evaluate (blank pages, “under construction” placeholders, Apache default pages, blank WordPress templates, etc.) (5.1% of all domains)
- 126 resulted in server and browser errors (such as 403, 404, and 502 errors) or were blocked. (4.2% of all domains)

Put another way, more than half of the gTLD namespace is controlled by unidentifiable parties. 51.7% of gTLD domains cannot be attributed to a registrant or site operator/user, either via WHOIS or by site content. That is a significant shift from 2018, before GDPR and ICANN’s Temporary Specification went into effect, when around 18% of domains were controlled by unidentifiable parties.³⁶

The loss of so much contact data may have had an effect on a range of consumers. A 2015 survey conducted for ICANN by Nielsen found that significant percentages of registrants and general Internet consumers use WHOIS to identify the creators of web sites.³⁷

³⁵ The NORC study of 2013 found that 20.6% of domains were parked. NORC, University of Chicago, “WHOIS Registrant Identification Study”, Appendix A page 20.

³⁶ See “Contact Data Publication Category” above. Pre-GDPR, unidentifiable registrants were those that held privacy/proxy-protected domains (about 25% of all domains) minus those privacy/proxy domain holders who identified themselves in web site content. According to our current analysis, about 28% of privacy-proxy registrants identify themselves on their web sites. Assuming that registrants have not changed their general web presence practices over the past few years, this means that 25% x .28 or 18% of domains were unidentifiable before GDPR and ICANN’s Temporary Specification went into effect.

³⁷ According to a 2015 survey conducted by Nielsen on behalf of ICANN, 55% of domain name registrants “have tried to identify the creator of a website. When asked how they did so, the most common specific mention is

Reach of GDPR Based on Jurisdiction

In the study set, we determined that **23.1% of the domains (693 out of 2,994) were subject to GDPR based on the jurisdictions of the parties involved.**³⁸

- 373 domains, or 12.5% of all domains, had a registrant in the EEA.³⁹
- 323 domains, or 10.8% of all domains, had their registrar in the EEA.
- 89 domains, or 3.0% of all domains, had their registry operator in the EEA. Of the TLDs in our set, only one registry operator (.ICU) is legally domiciled in the EEA.
- 250 domains, or 8.4% of all domains, had their back-end provider in the EEA.⁴⁰

The following chart illustrates the subsets and overlaps. The first row indicates that in there were 2,301 domains that were completely outside the GDPR's jurisdictional reach. The second row indicates there were 134 domains where *only* the registrant was in the EEA (and the registrar, registry, and back-end provider were *outside* the EEA). The last row indicates that no domains had *all four* parties in the EEA.

Domain has Registrant in EEA	Domain has Registrar in EEA	Domain has Registry Operator in EEA	Domain has Registry Back-end in EEA	TOTAL Domains (out of 2,994)	% of GDPR-covered domains (693)	% of all domains
				2301		76.9%
✓				134	19.3%	4.5%
	✓			85	12.3%	2.8%
		✓		0	0.0%	0.0%
			✓	141	20.3%	4.7%
✓	✓			224	32.3%	7.5%
✓		✓		0	0.0%	0.0%
✓			✓	6	0.9%	0.2%
	✓	✓		0	0.0%	0.0%
	✓		✓	5	0.7%	0.2%
		✓	✓	89	12.8%	3.0%
✓	✓	✓		0	0.0%	0.0%
	✓	✓	✓	0	0.0%	0.0%
✓		✓	✓	0	0.0%	0.0%
✓	✓		✓	9	1.3%	0.3%
✓	✓	✓	✓	0	0.0%	0.0%

WHOIS (22%), while some form of Internet search is mentioned by 28%." And 31% of consumers have tried to identify the creator of a website; their "prevalent methods for doing so include doing an Internet search or looking on the website itself." ICANN Global Registrant Survey, September 2015, pages 28 and 35.

<https://www.icann.org/news/announcement-2015-09-25-en>

³⁸ There were 6 domains in the set of 3,000 where we could not discern the country of the domain's registrant or user.

³⁹ The last data set that ICANN published, in 2017, showed that 14.3% of registrants resided in the EEA countries. See <https://www.icann.org/en/system/files/files/cct-metric-2-11-01nov17-en.xlsx>

⁴⁰ For a list of registries and back-end providers in the EU/EEA, see *Appendix A: TLD Representation in Sample Set*.

While 23.1% of the domains (693 out of 2,994) were subject to GDPR, only 12.5% of all domains (373 domains out of 2,994) had a registrant that resided in the EEA. This illustrates how the GDPR's jurisdictional reach extends far beyond EEA residents. By protecting personal data processed within the EEA, the GDPR extends protection to almost twice as many domains as have registrants who live in the EEA.⁴¹

Also of note:

- 134 domains (4.5% of all 3,000 domains) had the registrant in the EEA, but the registrar, registry operator, and back-end were outside the EEA.
- 233 domains (7.8% of all domains) had both the registrant and the registrar in the EEA. This was a third (33.6%) of the GDPR-covered domains.
- There were 140 domains (4.7%) where the registrant was in the EEA but used a registrar outside the EU. Registrants in the EEA used registrars in the EEA 63% of the time.
- 85 domains (2.8% of all domains) had the registrar in the EEA, but the registrant, registry, and registry back-end were outside the EEA. This demonstrates that registrars in the EEA did not have many registrants outside the EEA.
- 141 domains (4.7% of all domains) had only the back-end provider in the EEA, while the registrant, registrar, and registry operator were outside the EU. This is mainly because CentralNIC acts as the back-end provider for several gTLDs, and the registrants in those TLDs are mostly outside the EEA.
- There were zero domains (0.0%) that had *just* the registry in the EEA. The one TLD in our set that is domiciled in the EEA is .ICU, but it uses CentralNIC (counted in the EEA) as its back-end.

Of the 693 domains subject to GDPR, 47 had identifiable (non-proxy) contact data present in WHOIS. Of those, 33 were new gTLD domains where the registrants and registrars were outside the EEA, data was present in the Registrant Organization field, and only the back-end provider (CentralNIC) was in the EEA. Of the 373 domains with registrants in the EEA, 9 had identifiable (non-proxy) contact data in WHOIS. These might be registrants who opted into WHOIS publication – an option required by ICANN policy.⁴²

GDPR Jurisdiction of Redacted Domains

57.3% of the domains in the study set had their contact data redacted (1,718 of 3,000). Of the 1,718 redacted domains, one third of them (572, or 33.3%) were subject to GDPR based on the location of the registrant, registrar, registry operator, or registry back-end provider. The other 1,146 redacted domains (66.7%) were not subject to GDPR. Most of those domains had their registrants in the United States (534 domains) and China (273 domains).

In general, based on the above, it appears that contact data that requires GDPR protection is being protected from exposure in WHOIS/RDDS.

⁴¹ Some registrants register more than one domain name, so there is not a one-to-one ratio between registrants (data subjects) and domains. Because so little contact data is now available in WHOIS, is not possible to establish what the ratio may be.

⁴² See paragraph 7.2.1, at <https://www.icann.org/resources/pages/gtld-registration-data-specs-en>

Reach of GDPR Based on Jurisdiction *and* Legal Character

To be covered by the GDPR, data must fall within the GDPR's jurisdictional scope, *and* the data subject must be a natural person. How many domains qualify for GDPR protection under these criteria?

The GDPR "does not cover the processing of personal data which concerns legal persons and in particular undertakings established as legal persons, including the name and the form of the legal person and the contact details of the legal person."⁴³ For this reason, the European Commission is currently considering the "Proposed Directive on Measures for a High Common Level of Cybersecurity across the Union." This proposed legislative act states: "TLD registries and the entities providing domain name registration services for them should make publically available domain name registration data that fall outside the scope of Union data protection rules, such as data that concern legal persons."⁴⁴

A total of 693 domains in our study set of 3,000 are subject to GDPR based on jurisdictional scope, *i.e.*, on the location of the registrant, registrar, registry operator, or the back-end provider. Of those 693 domains, 271 (39.1%) were with legal persons, 39 (5.6%) were with natural persons, and the other 383 (55.3%) were unclassifiable as either legal or natural person users.

Of the 693 domains, 74 were under privacy/proxy protection. If the privacy/proxy domains are set aside, the set consists of 619 domains, where 260 (42.0%) were with legal persons, 36 (5.8%) were with natural persons, and 323 (52.2%) that were unclassifiable as either legal or natural person users.

	Registrant in EEA	Registrar in EEA	Registry Operator in EEA	Registry back-end in EEA	Domains completely outside EEA jurisdiction
legal person user	206	175	5	34	817
natural person user	30	25	0	5	103
unclassified user	137	123	84	211	1,387

Notes: Includes privacy/proxy domains. A user can fall into more than one of the first four categories above – for example, a registrant can reside in the EEA and also use a registrar in the EEA.

Some of the unclassifiable domains were certainly registered by legal persons. How many is unknown, and the only way to confirm how many is to obtain redacted data from the registrars. That process involves several challenges beyond the scope of this study.

For the sake of discussion, let us make a conservative assumption that 50% of domains are registered by legal persons. As described above, 23.1% of gTLD domains are subject to GDPR based on jurisdictional scope. If half of those domains are registered by legal persons and are ineligible for GDPR protection, then **the percentage of gTLD domains eligible for GDPR protection is around 11.5%**. That set consists of the natural person registrants located with the EU, and natural person registrants who have their domain's registrar, registry, or registry back-end provider that process data within the EU.

In contrast, 57.3% of domains currently have redacted contact data. **By this measure, ICANN policy has allowed the redaction of around five times as much domain contact data as is required by the GDPR.**

⁴³ GDPR Recital 14.

⁴⁴ Paragraph 62, pp. 24-25, at https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=72166. See also <https://ec.europa.eu/digital-single-market/en/news/proposal-directive-measures-high-common-level-cybersecurity-across-union>

If ICANN were to update its policies to require that parties “make publically available domain name registration data that fall outside the scope of Union data protection rules, such as data that concern legal persons,” as the proposed European Commission act states, then redaction should fall from 57.3% to about 11.5% of all gTLD domains names. The implementation of such would require a policy and procedure to address the situation in which legal persons list the personally identifiable data of natural persons in their domain registration records.

As noted above, the GDPR’s reach extends far beyond EEA residents. By protecting personal data processed within the EEA, the GDPR extends protection to almost twice as many domains as have registrants who live in the EEA.⁴⁵ Extrapolated across the gTLD space, the GDPR’s jurisdictional reach protects about 23.2 million domains where the registrant does not live in the EEA.⁴⁶ If consideration is only applied to natural persons, that number would be reduced by half or more, to around 11.6 million domains. For perspective, that is more domains that are sponsored by the second-largest registrar in the world, NameCheap, which has 10.5 million gTLD domains names under management.

Contact Data Publication Category: Cybercrime Domains

Are domains registered to perpetrate cybercrime treated differently in WHOIS? In at least one category of cybercrime, the answer is “yes.” **An unusually high percentage of the domain names registered to perpetrate phishing are registered under proxy services, to obscure the perpetrator's identity.**

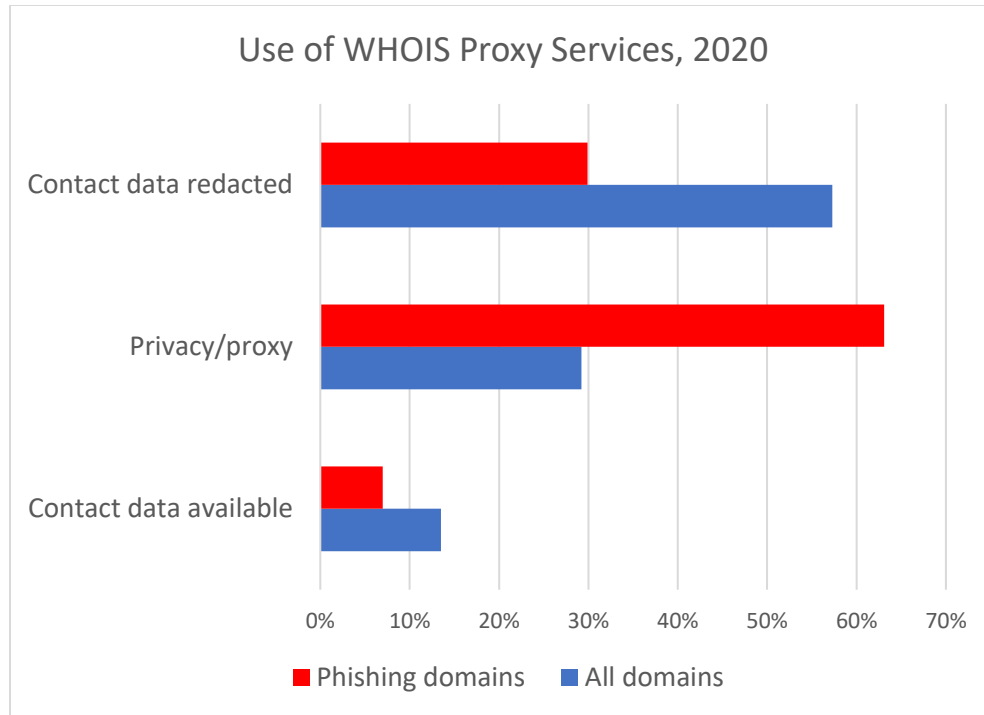
We studied a set of 1,000 “maliciously registered” domains that were registered by phishers, to perpetrate phishing attacks. These domains are “beyond the pale” for two reasons. First, each domain name contained a deceptive string designed to fool victims– the name of a targeted company, or a term such as “login” or “security.”⁴⁷ Second, phishing usually occurred on these domains within a short time of registration, mostly three days or less; this is an indication of bad intent.

The phishing domains were under proxy protection at more than twice the usual rate:

⁴⁵ Some registrants register more than one domain name, so there is not a one-to-one ratio between registrants (data subjects) and domains. Because so little contact data is now available in WHOIS, is not possible to establish what the domains-to-registrant ratio may be.

⁴⁶ If 23.1% of domains are protected by GDPR, and 12.5% have a registrant in the EEA, then 10.6% of the protected domains are outside the EEA. The gTLD space consists of 219.6 million domains, and 219.6 million x 10.6% = 23.2 million.

⁴⁷ To learn more about these domains, see “Phishing Landscape 2020: A Study of the Scope and Distribution of Phishing,” <http://www.interisle.net/PhishingLandscape2020.html>



	Contact data available	Privacy/proxy	Contact data redacted
All domains	13.5%	29.2%	57.3%
Phishing domains	7.0%	63.1%	29.9%

In the set of 1,000 phishing domains, 631 were under proxy protection, and 299 had contact data redacted. Only 70 had contact data available in WHOIS.

These results are very different from those found in the *WHOIS Privacy and Proxy Abuse Study*⁴⁸ of 2014, conducted for ICANN by the National Physical Laboratory of the U.K. In that study six years ago, only 31.2% of maliciously registered phishing domains were under privacy/proxy protection – half the rate in 2020.

The WHOIS data for these domains, plus the specific URLs of the phishing attacks, the dates of the phishing reports, and the deceptive strings, are available in the data file that accompanies this report.

⁴⁸ <http://whois.icann.org/sites/default/files/files/pp-abuse-study-final-07mar14-en.pdf>, page 6

Conclusion

ICANN's stated goal was to "ensure compliance with the law while preserving the current information contained in WHOIS to the greatest extent possible." The major finding of this study is that ICANN has not met this goal. The data indicates that perhaps 11.5% of domain name registrations fall legitimately within the legal scope of the GDPR. However, ICANN's GDPR-driven policy has allowed registrars and registry operators to redact (withhold) the contact data for the majority of gTLD domains—perhaps five times as much as is required under GDPR. While ICANN's policy has generally protected the data that must be protected per GDPR, it has also been used to conceal a much larger set of contact data that is not subject to GDPR. ICANN has not preserved the current information contained in WHOIS to the greatest extent possible and has thus deprived parties of data they need to legitimately help maintain a secure and interoperable Internet. The current situation was not inevitable and is mainly the result of ICANN policy rather than the GDPR itself.

Appendix A: TLD Representation in Sample Set

Registry operators or back-end providers located in the EU/EEA are noted below.

TLD	TLD type	Domains in registry, June 2020	Market share, of all gTLDs	Domains in study set	% of sample set	Registry operator	Back-end registry provider
.com	legacy	152,319,470	69.4%	2,060	68.7%	Verisign	Verisign
.net	legacy	13,700,522	6.2%	187	6.2%	Verisign	Verisign
.org	legacy	10,670,839	4.9%	146	4.9%	PIR	Afilias
.icu	nTLD	6,502,844	3.0%	89	3.0%	Shortdot	CentralNIC
.info	legacy	4,732,996	2.2%	65	2.2%	Afilias	Afilias
.top	nTLD	3,683,753	1.7%	50	1.7%	Jiangu Bangning	ZDNS
.xyz	nTLD	3,173,653	1.5%	43	1.4%	XYX.COM LLC	CentralNIC
.site	nTLD	2,008,616	0.9%	27	0.9%	Radix FCZ	CentralNIC
.online	nTLD	1,630,394	0.7%	26	0.9%	Radix FCZ	CentralNIC
.biz	legacy	1,523,941	0.7%	26	0.9%	Neustar	Neustar
.club	nTLD	1,442,760	0.7%	25	0.8%	Neustar	Neustar
.wang	nTLD	1,390,601	0.6%	25	0.8%	Zodiac Registry	ZDNS
.vip	nTLD	1,243,622	0.6%	25	0.8%	MMX	Nominet
.app	nTLD	875,823	0.4%	25	0.8%	Charleston Road (Google)	Charleston Road (Google)
.shop	nTLD	749,783	0.3%	20	0.7%	GMO Registry	GMO Registry
.live	nTLD	722,465	0.3%	20	0.7%	Donuts	Donuts
.work	nTLD	690,771	0.3%	20	0.7%	MMX	Nominet
.buzz	nTLD	621,719	0.3%	20	0.7%	Dotstrategy	Neustar
.fun	nTLD	522,689	0.2%	20	0.7%	Radix FCZ	CentralNIC
.dev	nTLD	233,566	0.1%	20	0.7%	Charleston Road (Google)	Charleston Road (Google)
.life	nTLD	217,971	0.1%	20	0.7%	Donuts	Donuts
.tokyo	nTLD	190,474	0.1%	20	0.7%	GNO Registry	GMO Registry
.world	nTLD	138,225	0.1%	20	0.7%	Donuts	Donuts
TOTAL		208,987,497	95.2%	3,000			

The registry operators and back-end providers legally established in the EEA are in **boldface**, below:

- .FUN (registry operator DotSpace, Inc. in India, back-end **CentralNIC in U.K.**)
- .ICU (registry operator **ShortDot in Luxembourg**, back-end **CentralNIC in U.K.**)
- .ONLINE (registry operator DotOnline Inc. in the United Arab Emirates, back-end **CentralNIC in U.K.**)
- .SITE (registry operator DotSite Inc. in the United Arab Emirates, back-end **CentralNIC in U.K.**)
- .VIP and .WORK (registry operator Mind + Machines in British Virgin Islands, back-end **Nominet in U.K.**)
- .XYZ (registry operator XYZ.COM LLC in U.S., back-end CentralNIC in U.K.)

The European Economic Area (EEA) consists of the Member States of the European Union (EU), plus three countries of the European Free Trade Association (EFTA): Iceland, Liechtenstein, and Norway. Although the United Kingdom has “Brexit” the European Union, we counted the U.K. as being under GDPR because the U.K. followed GDPR in 2020, and the U.K. also has its own parallel data protection act that is modelled directly on the GDPR.

Appendix B: Registrar Representation in Sample Set

Below is data for the twenty largest registrars. A complete list may be found in the data file that accompanies this report.

Registrar IANA ID	Registrar	Registrar DUM	Domains in study set	% of study	% market
146	GoDaddy.com, LLC	63,311,867	882	29.4%	28.8%
1068	NameCheap, Inc.	10,497,555	152	5.1%	4.8%
69	Tucows Domains Inc.	10,263,753	131	4.4%	4.7%
2	Network Solutions, LLC	6,902,784	100	3.3%	3.1%
49	GMO Internet, Inc. d/b/a Onamae.com	5,238,919	85	2.8%	2.4%
420	Alibaba Cloud Computing (Beijing) Co., Ltd.	6,643,000	81	2.7%	3.0%
1599	Alibaba Cloud Computing Ltd. d/b/a HiChina (www.net.cn)	5,200,971	81	2.7%	2.4%
48	eNom, LLC	5,258,420	80	2.7%	2.4%
895	Google LLC	5,020,195	78	2.6%	2.3%
303	PDR Ltd. d/b/a PublicDomainRegistry.com	4,892,251	72	2.4%	2.2%
83	1&1 IONOS SE	4,876,367	58	1.9%	2.2%
1915	West263 International Limited	3,306,729	56	1.9%	1.5%
440	Wild West Domains, LLC	2,764,124	41	1.4%	1.3%
1556	Chengdu West Dimension Digital Technology Co., Ltd.	3,431,766	38	1.3%	1.6%
1154	FastDomain Inc.	2,341,325	37	1.2%	1.1%
120	Xin Net Technology Corporation	3,506,805	37	1.2%	1.6%
1479	NameSilo, LLC	3,463,466	35	1.2%	1.6%
1868	Eranet International Limited	1,946,035	31	1.0%	0.9%
472	Dynadot, LLC	2,349,300	30	1.0%	1.1%
433	OVH sas	2,308,970	28	0.9%	1.1%

Appendix C: WHOIS Problems

Interisle is a world-class expert at using the WHOIS system. Even so, Interisle had difficulty obtaining about 10% of the WHOIS data that was required to perform this study and had to go to extended efforts to obtain the necessary data. Current practices in the domain name industry make performing this kind of study very challenging.

Some of the problems prevented access to data that ICANN requires its registries and registrars to make available publicly. These problems were:

1. Several registrars had **broken or non-responsive port 43 WHOIS services**. Some were non-responsive intermittently, others not at all. Others had non-responsive web-based WHOIS lookup tools.
2. **Rate-limiting**. Many registrars and registry operators have decided to restrict the number and frequency of queries that users can make to WHOIS and RDAP servers, a practice known as rate-limiting. To avoid rate-limiting, Interisle made its port 43 WHOIS queries at a slow rate, performing no more than two queries per minute to any given registrar, but we still ran afoul of some registrars' limits. A few registrars were so restrictive that they rejected our *initial* queries to their servers. For more about rate-limiting, see *SAC101v2: SSAC Advisory Regarding Access to Domain Name Registration Data*.⁴⁹
3. **Contractual loophole regarding service method**: ICANN policy requires registrars to serve certain mandatory data fields, which are specified in the *Temporary Specification* and the 2013 Registrar Accreditation Agreement. However, some registrars have decided to not serve the required data elements via WHOIS port 43 service or RDAP. Instead, they serve some required data elements only via single, manual Web-based lookups. This practice eliminates the ability of legitimate users to access important data via WHOIS port 43 or RDAP, and via automation. This loophole gives registrars latitude to control the release of even non-sensitive, non-personal domain data.
4. **Contractual loophole regarding TLD**. ICANN only requires registrars to provide WHOIS for the .COM and .NET domains they sponsor.⁵⁰ This means that for all other TLDs, the *registry* is the only source of WHOIS data. However, ICANN's *Interim Registration Data Policy for gTLDs* allows registry operators to redact (not provide) most contact data fields in their WHOIS services. The result is that for all the TLDs other than .COM and .NET, there is usually no way to see most contact data fields, even when the registrant opts into publication.

Interisle was eventually able to compensate for the above problems by either:

- performing web-based WHOIS lookups – a laborious process that cannot be automated—or,
- through multiple port 43 attempts separated by days to weeks, or
- via RDAP (a new RDDS service that ICANN is in the process of rolling out, but which requires entirely different methods to parse than WHOIS), or
- via historical WHOIS data, if the web site demonstrated that the registrant had not changed between 2018 and 2020.

⁴⁹ <https://www.icann.org/en/system/files/files/sac-101-v2-en.pdf>

⁵⁰ Paragraph 3.3.1, at <https://www.icann.org/resources/pages/approved-with-specs-2013-09-17-en>

Once the data was gathered, we noticed problems with some of it.

We found that sometimes *the registry and the registrar provided different Registrant Country data for the same domain name*. This is a fundamental discrepancy that should never occur. When this problem occurred, we used web site data to confirm the correct country, and if that was not possible, we used the Registrant Country provided by the registrar.

Some registrars simply did not provide Registrant Country data, which is vital for determining which registrants were subject to GDPR. The *Interim Registration Data Policy for gTLDs* requires that registrars provide Registrant Country data, since it is not personally identifiable data, and therefore GDPR does not come into play.

Some registrars' WHOIS services say that their domains in Redemption Grace Period (RGP) "do not exist." However, these domains do exist, as stated in registry WHOIS, and are eligible for redemption. While ICANN policy says that the registrar can change the contact data on domains names in RGP⁵¹, ICANN 's 2013 Registrar Accreditation Agreement states that registrars must provide WHOIS for all .COM and .NET domains under their sponsorship⁵², regardless of lifetime phase. We filled in data for such domains using recent historical WHOIS records.

We also found discrepancies between the domain Create dates and domain Expiration Dates provided by the registry versus the dates provided by the sponsoring registrar. This is a fundamental discrepancy that should never occur.

⁵¹ Paragraph 1, at <https://www.icann.org/resources/pages/errp-2013-02-28-en>

⁵² Paragraph 3.3, at <https://www.icann.org/resources/pages/approved-with-specs-2013-09-17-en>

About Interisle Consulting Group, LLC

Interisle's principal consultants are experienced practitioners with extensive track records in industry and academia and world-class expertise in business and technology strategy, Internet technologies and governance, financial industry applications, and software design. For more about Interisle, please visit: www.interisle.net

About the Authors

Greg Aaron is an internationally recognized authority on the use of domain names for cybercrime, and is an expert on domain name registry operations, DNS policy, and related intellectual property issues. He has been a member of ICANN's EPDP Working Groups, which have been creating registration data access policies and examining the requirements of GDPR as they apply to domain name data. As a member of ICANN's Security and Stability Advisory Committee (SSAC), he advises the international community regarding the domain name and numbering system that makes the Internet function. Mr. Aaron is Senior Research Fellow for the Anti-Phishing Working Group. He works with industry, researchers, and law enforcement to investigate and mitigate cybercrime, and is also a licensed private detective. He was the senior industry expert on a team that evaluated the policy and technical qualifications of more than one thousand new TLD applications to ICANN in 2012-2013. He has managed multiple top-level domains around the world, including .INFO, .ME, and .IN. He is President of Illumintel, Inc., a consulting company. Mr. Aaron is a *magna cum laude* graduate of the University of Pennsylvania.

Lyman Chapin has contributed to the development of technologies, standards, and policy for the Internet since 1977, and is widely recognized and respected as a leader in the networking industry and the Internet community. Mr. Chapin is a Life Fellow of the IEEE and has chaired the Internet Architecture Board (IAB), the ACM Special Interest Group on Data Communication (SIGCOMM), and the ANSI and ISO standards groups responsible for Network and Transport layer standards. Mr. Chapin was a founding trustee of the Internet Society, and a Director of the Internet Corporation for Assigned Names and Numbers (ICANN). He currently chairs ICANN's Registry Services Technical Evaluation Panel (RSTEP), which is responsible for assessing the impact of new Domain Name System (DNS) registry services on the security and stability of the Internet, and the DNS Stability Panel, which evaluates proposals for new Internationalized Domain Names (IDNs) as country code top-level domains (ccTLDs). He is also a member of ICANN's Security and Stability Advisory Committee (SSAC). He has written many other papers and articles over the past 40 years, including the original specification of the Internet standards process operated by the IETF. Mr. Chapin holds a B.A. in Mathematics from Cornell University.

David Piscitello has been involved in Internet technology and security for more than 40 years. Until July 2018, Mr. Piscitello was Vice President for Security and ICT Coordination at ICANN, where he participated in global collaborative efforts by security, operations, and law enforcement communities to mitigate Domain Name System abuse. He also coordinated ICANN's security capacity-building programs and was an invited participant in the Organisation for Economic Co-operation and Development (OECD) Security Expert Group. Dave is an Associate Fellow of the Geneva Centre for Security Policy. He served on the Boards of Directors at the Anti-Phishing Working Group (APWG) and Consumers Against Unsolicited Commercial Email (CAUCE). He is the recipient of M3AAWG's 2019 Mary Litynski Award,

which recognizes the lifetime achievements of individuals who have significantly contributed to making the Internet safer.

Dr. Colin Strutt has published and spoken extensively on networking technology, name collisions, enterprise management, eBusiness, and scenario planning, and has represented the interests of Digital Equipment, Compaq, and the Financial Services Technology Consortium in national and international industry standards bodies. He holds six patents on enterprise management technology and brings more than thirty five years of direct experience with information technology, as a developer, architect, and consultant, with recent work including design and operation of a regional public safety network, providing technical expertise relating to patents, and analysis of world-wide Internet use. Dr. Strutt holds a B.A. (with First Class Honours) and Ph.D. in Computer Science from Essex University (UK).