

ICANN Study on the Prevalence of Domain Names Registered using a Privacy or Proxy Service among the top 5 gTLDs

Executive Summary

ICANN-accredited registrars and some gTLD registries are required to make available the identity and contact information of domain name registrants, administrative and technical contacts, and in some cases, billing contacts by way of web-based and plain-text Whois services.

Domain names can be registered in ways that help limit the amount of users' personal information that is made public via registrar and registry Whois services. Through use of Whois privacy services, registered name holders are typically listed in Whois as the registrants of record, but alternate, valid contact information (such as a mail forwarding service address) is published in place of, for example, the registrant's home address. Users may also elect to use a Whois proxy service. The proxy service is the registered name holder of record, and its identity and contact information are published in Whois. The beneficial user (or "licensee") of the registration holds rights to use the name, subject to its agreement with the registered name holder/proxy service.

The results indicate that at least 18% of domain names registered under the top 5 gTLDs are likely to have been registered using a privacy or proxy service.

Background

To ensure that the community can identify who is responsible for a domain name, a registered name holder is required to provide and update, as needed, its contact information with its registrar of record. Registrars are required by ICANN to collect and provide free public access to the name of the registered domain name and its name servers and registrar, the date the domain was created and when it expires, and the contact information for the registered name holder, the technical contact, and the administrative contact. Similarly, many gTLD registries are required by ICANN to make these data elements, and in some cases, contact information for billing contacts, available via their own Whois services.

Registered name holders can limit the amount of personal information that is made available to the public through Whois queries of registrar domain name databases. To do so, they can use a privacy service. In other cases, the beneficial user of a registration can utilize a "proxy service" that acts as registered name holder while licensing use of the registration to the beneficial user of the domain.

For the purpose of this study, the definitions of privacy and proxy services, as they relate to their use in the domain name system, are as follows:

- A **privacy** service provider offers alternate contact information that the registered name holder may choose to have listed in a Whois record instead of the registrant's other addresses, telephone numbers, or email addresses.
- A **proxy** service provider acts as registered name holder of record and licenses the use of a domain name to the customer or beneficial user of the domain. The contact information in a Whois record for a domain name registered with a proxy service is that of the proxy service provider / registered name holder.

The ICANN community has expressed a considerable amount of interest in and raised questions about domain names registered using a privacy and proxy service. However, there is an absence of factual information on the prevalence of their use, which is why ICANN conducted this exploratory study.

Objective

The objective of this study was to establish an approximation of the percentage of domain names (through a statistical sampling plan) contained in gTLD registries that used (1) a privacy or (2) proxy registration service.

Conclusion

Domain names can be registered using a Whois privacy or proxy service, which helps limit the amount of users' personal information that is made public via registrar and registry Whois services. The sample of domain names registered under the top 5 gTLDs indicates that about 18% of them used this type of service. Among these, Whois proxy service registrations were the most common.

Statistical Sampling

The sample was randomly selected in accordance with a design advised by statisticians from NORC, who used commonly known and well accepted statistical methodologies. The sample was a systematic random sample drawn within strata, or categories, formed by each of the top five gTLDs, with the number of selections in each strata strictly proportional to the number in the entire domain name universe. For instance, since 75% of domain names within the top five gTLDs have been registered under .com, 75% of domain names contained in the sample were also associated with the .com top-level domain. The absence of clustering or any further sub-sampling in this design made the sample self-weighting, meaning that whatever percentage of the sample was found to be of domain names registered using a privacy or proxy service, no further adjustment to that statistic was needed and it could be inferred that an equal percentage of the broader population of domain names was also linked to a privacy or proxy service.

Data Source and Methodology

To conduct this study, ICANN used a statistically based sample of 2400 domain names drawn in March 2009 for the Whois Accuracy Study, conducted by the National Opinion Research Center (NORC), a research firm affiliated with the University of Chicago. A full description of that sample can be found in the *Draft Report for the Study of the Accuracy of WHOIS Registrant Contact Information* dated 17 January 2010 (accessible here: <http://www.icann.org/en/compliance/reports/whois-accuracy-study-17jan10-en.pdf>). The two studies shared the sample of domain names.

The following table compares the distribution of the top five gTLDs at the time of sample selection with the distribution used in the microcosm of 2400.

Table 1. Summary of global Top-Level Domains (gTLDs) of interest to ICANN

Rank	Top-Level Domain	Total Domains in universe	Percentage of Domains in universe	Microcosm of 2400	Percentage of Domains in microcosm
1	.com	75,785,462	75%	1,801	75%
2	.net	11,478,837	11%	273	11%
3	.org	6,840,493	7%	167	7%
4	.info	5,092,053	5%	114	5%
5	.biz	2,029,143	2%	45	2%
	TOTAL	101,225,988	100%	2,400	100%

Although 2,400 is only a 0.002% sample from the universe, it is sufficient for estimates based on a sample of 2400 to be made with reasonable precision, that is, with a margin of error of less than 2 percentage points for a 95% confidence interval. Another way of describing this is to say that had we drawn 1000 samples of 2400 from the universe and examined the percentage with characteristic x, we would almost certainly find that 95% of them produced an estimate between 48% and 52% (i.e., 50% +/- 2% is the 95% confidence interval).

The Whois information for each domain name contained in the sample was extracted by ICANN and provided to NORC.

Once the sample was drawn, there were three stages of work:

1. an initial classification, based on registrant and administrative contact details, to identify a subset of cases for more detailed examination;
2. a detailed examination of each of these cases, including research into apparent privacy/proxy service provider, to classify the case as privacy/proxy or not; and
3. if classified as privacy/proxy, a tentative classification of the case as either privacy or proxy depending on the content of the Whois record.

Each of these stages is described in more detail below.

Stage 1: Initial classification

Among the 2400 sampled cases, nearly one quarter (580 cases) were identified in this stage as being potentially privacy or proxy service providers, although only two thirds of these were classified as “highly likely”, with the remaining one third including various borderline cases. A wide net was deliberately cast since the next stage was designed to be more precise.

This initial classification for potential privacy or proxy domain name registrations was done by searching the Whois record for the following clues:

1. where the registrant name contained the following words or phrase 'privacy', 'proxy', 'registration', 'registration service', 'identity', 'shield', 'guard', 'private', 'buy', 'rare', 'names', 'whois', 'value', 'domain', and 'secure'.
2. where multiple domains contained the same registrant name, registrant organization, or registrant address.
3. where the registrant name or registrant organization may correspond to the name of a privacy service, proxy service, or multiple domain name holder. Registrant Name Examples: DowntownWebsites.com and DNSADMIN.

The majority of cases identified in the “highly likely” category satisfied at least two, if not three, of these criteria.

Over time, as more cases were verified as part of the Whois accuracy study, an additional 48 “potential” cases were added to this group.

Stage 2: Final classification

Since the majority of potential privacy/proxy cases were part of multiple listings – i.e., the same registrant name or address was used for multiple cases, NORC started with those cases and tried to find a single contact point for each set was identified. For other cases, NORC approached them individually, using whatever information was in Whois, supplemented by additional phone book and online searches. For those records that were also in the Accuracy study, NORC was also making contact as part of the process to ascertain whether the registrant acknowledged ownership of the domains listed to them. As part of that contact, NORC would ask the service provider whether it was providing a privacy/proxy service or if it held multiple domain names for its own benefit (as it would if it were a domain name investor).

If NORC was unable to establish contact, or if the status of the service was still in question, several additional tests were considered:

1. NORC would search for mention of the provision of a privacy/proxy service among any of the organizations' literature;
2. NORC would search business records and/or their online presence to see if there was an indication that they were primarily a service provider of another kind. For example, many attorneys and web developers provide proxy or privacy type services to clients, but only as a byproduct of their main service to the client, and there is no evidence they provide privacy or proxy services independent of their other services.

Unless NORC was fairly certain it had a privacy/proxy service, it coded the case as a multiple domain name holder, organization, or person, as most appropriate.

This process yielded 18% of domain names to be represented by a privacy or proxy service. The majority of the cases which were in the initial classification but which were deemed not to be Privacy or Proxy in the final classification were domains belonging to holders of multiple domain names. Below are the final counts and associated percentages:

Final classification - Privacy or Proxy service	429	18%
Final classification - neither	1971	82%
<i>Total domains</i>	<i>2400</i>	<i>100%</i>

Because this is a sample, our 95% confidence interval for the percentage of Privacy or Proxy services among the top five gTLDs is 16% -20%.

Extrapolating this to the full count of 101 million domains in the top five gTLDs at the time of sampling, it is likely that the total number of domains represented by Privacy or Proxy services was between 17.7 and 18.4 million.

The true proportion of privacy/proxy registrations at the time of writing may well be higher than the percentage quoted above, due to several factors:

- industry trends suggesting that increasing proportions of domain name holders are taking advantage of privacy and proxy services; and
- at the time of extraction, information on 47 of the 2400 records was unable to be extracted. Thus by default, all these cases for this Privacy/Proxy study were deemed to be NOT privacy/proxy (using our broad approach of classifying cases as NOT privacy/proxy unless sufficient evidence proved otherwise). Should all 47 have turned out to be Privacy/Proxy, the 429 above would increase to 476, and the associated percentage increase to 20%.

This, combined with the growth in registration numbers to over 115 million domain names as of January 2010, means that the actual number of privacy/proxy registrations among the top five gTLDs is likely to be substantially higher than 18 million.

Stage 3: Classifying as privacy or proxy

The last stage was to take the 429 cases found to be a Privacy/Proxy case, and classify as either Privacy or Proxy. NORC based its classification strictly on the definitions listed above in the background section of this report.

The majorities of these domain name registrations, close to 91% of them, had no information to identify the beneficial user, and so were classified as Proxy registrations. Therefore, 9% were found to be Privacy domain name registrations.