3 May 2019

Subject: SSAC2019-02: Registration Data Services Query Reporting

To: Russ Weinstein, Director, Registry Services and Engagement;
Jamie Hedlund, Senior Vice President, Contractual Compliance & Consumer Safeguard
CC: Cyrus Namazi, Senior Vice President, Global Domains Division

Dear Russ and Jamie:

This letter is a follow-up to SSAC's discussion with you about gTLD metrics reporting. As we described, it appears that the WHOIS query statistics provided to ICANN by registry operators as part of their monthly reporting obligations are generally not reliable. Some operators are using different methods to count queries, some are interpreting the registry contract differently, and some may be reporting numbers that are fabricated or otherwise not reflective of reality. Reliable reporting is essential to the ICANN community, especially to inform policy-making. This letter provides our data and observations to you for follow-up as well as some suggestions going forward.

We began with an important question for the ICANN community; "Has ICANN's Temp Spec policy affected the number of WHOIS queries that users have made?" To find out, we downloaded port 43 query statistics from recent public monthly registry reports that are posted on ICANN's Website.[1] As we analyzed the data, it quickly became apparent that the data are problematic, and that our questions cannot be answered using the available data.

We had informal conversations with several registry operators, in which we presented what we saw in the data and learned about their operations in order to better understand the situation.

[1] All gTLD operators must report both port 43 and web-based WHOIS statistics each month. The requirements are found in the Base Registry Agreement, Specification 3, at:
https://newgtlds.icann.org/sites/default/files/agreements/agreement-approved-31jul17-en.html#specification3

The legacy gTLDs have similar requirements. The reports are posted publicly by ICANN at:
https://www.icann.org/resources/pages/registry-reports

These problems have probably existed for many years, and the historical data likely suffer from these problems.

Reporting issues may also be grouped by back-end provider, who provide Registration Data Services (RDS) infrastructure and query counts to their registry operator customers, who then report the metrics to ICANN.

We have observed three phenomena.

Phenomenon #1: Registry-made Queries

Many operators make automated queries to their own port 43 servers, to monitor performance and for Service Level Agreement (SLA) compliance. Such monitoring is a very logical practice. However, some operators are counting these self-queries while others may not be, and so there may not be a common methodology. Some operators told us that self-queries represent a significant percentage of the total port 43 queries reported. No one looking at the reports can know how many queries are "real" – i.e. from users of the service across the Internet – which we believe is the purpose of the statistics. The data often don't represent what most people think they do.

We note that registry operators similarly make Extensible Provisioning Protocol (EPP) test transactions, such as domain create and delete transactions, in order to monitor their EPP servers. However, registry operators do *not* count those queries in their monthly reporting to ICANN, because they are not "real" transactions from accredited registrars, and they are not to be counted as billable transactions.

Phenomenon #2: Identical Data

Some operators are reporting that many of their TLDs *receive the exact same number of queries* in a given month. This is highly improbable, because:

1. The port 43 servers are exposed to the Internet. Activity by outside users should create different numbers for every gTLD.
2. The gTLDs at issue have widely varying numbers of domains in them, and probably widely varying usage. One would expect the larger and busier gTLDs to receive more port 43 queries than small gTLDs. One operator reported that some of its open, generic TLDs (containing hundreds of thousands of domains) and some of its very small .BRAND gTLDs (containing less than ten domains), received the exact same number of WHOIS queries.

One registry employee hypothesized that query data for its gTLDs are identical because *all* the queries to those gTLDs over the month came from registry self-monitoring. This is not plausible, because we know that outside users are making WHOIS queries about domains in *all* the gTLDs.

Why are some operators reporting identical numbers within their TLD portfolios? One explanation is that some operators operate multiple gTLDs on one registry system, and the WHOIS service for those TLDs is provided by a common port 43 server system. *Rather than counting how many queries each of the TLDs actually receives, some operators seem to be*

reporting all of the queries received by the server or system, or are reporting an average of that. Here are two hypothetical scenarios designed to illustrate what might be happening:

Example 1: Operator RegistryCo runs 10 gTLDs. In January, across those 10 gTLDs, RegistryCo received a grand total of 10,000,000 WHOIS queries– 3,000 queries for TLD1, 857,992 queries for TLD2, etc. RegistryCo reports to ICANN that each of the 10 gTLDs received 1,000,000 queries *each* – simply dividing the 10,000,000 queries received at the server by the number of TLDs served (10). This is not a true count for any of the TLDs.

Example 2: Per example 1, operator RegistryCo runs 10 gTLDs, and across them received a grand total of 10,000,000 WHOIS queries in January – 3,000 queries for TLD1, 857,992 queries for TLD2, etc. RegistryCo reports to ICANN that the 10 gTLDs received 10,000,000 queries *each*. This is not a true count for any of the TLDs. And it reports 100,000,000 total queries across the 10 TLDs – not the 10,000,000 total queries that really took place.

At least one operator is evidently of the opinion that this kind of "aggregated reporting" is allowed by the Registry Agreement. We believe that was never the intention of the contract, and such a methodology is clearly problematic. The contract also notes that the "Registry Operator shall provide one set of monthly reports *per gTLD*." [emphasis added]

Phenomenon #3: Unusual Data

There is a case where an operator operates multiple TLDs, served from a common port 43 system. In a given month, the number of reported WHOIS queries for each of the operator's TLDs is different. While some of the TLDs are much larger than others, the WHOIS query totals for them are close to each other. Further statistical analysis on the number of WHOIS queries per TLD revealed that an abnormal distribution. For one month of data for one of the registries, the WHOIS query counts per TLD differed from the mean by about +/- 1%, nearly linearly. This appeared to be highly unusual, especially with TLDs that have different usage patterns and domain counts. There is a chance that the numbers were altered or synthesized. More extensive analysis of these numbers could be warranted.

Follow-Up

We suggest the following:

1. ICANN Org issue guidance to all registry operators, clarifying expectations for reporting port 43 queries and RDAP queries. The guidance should make clear the purposes and goals of the data collection and the contractual obligations.
2. SSAC believes that a purpose of gathering the data is to document queries made by the users (consumers) of the registration data service. Registry operators should exclude the queries they make to their own systems.
3. It is vital that ICANN collect valid, accurate data regarding RDAP queries. The WHOIS query data is unreliable, but the move to RDAP offers an opportunity to get things right.

4. ICANN Org's Global Domains Division and Compliance Department should examine the reporting of other metrics per the registry contracts. In SAC097, the SSAC pointed out irregularities in the reporting of other registry metrics, such as web-based WHOIS queries.[2] The current issue raises the question of whether metrics such as DNS queries are being reported accurately, if they too involve the reporting of self-queries, etc.

We invite you to schedule a meeting with the SSAC at ICANN65 in Marrakech as a follow-up. Please contact us in the meantime if you have any questions.

Best regards,

Rod Rasmussen
Chair, ICANN Security and Stability Advisory Committee

[2] See SAC097: SSAC Advisory Regarding the Centralized Zone Data Service (CZDS) and Registry Operator Monthly Activity Reports at https://www.icann.org/en/system/files/files/sac-097-en.pdf and https://www.icann.org/en/system/files/files/resolutions-board-action-ssac-advice-scorecard-08jun18-en.pdf