



The Internet Corporation for Assigned Names and Numbers

11 March 2017

Commercial and Business Users Constituency (BC): bc-gnso@icann.org
Intellectual Property Constituency (IPC): ipc-gnso@icann.org
ISPs and Connectivity Providers Constituency (ISPCP): ispcp@icann.org

Dear Commercial Stakeholder Group

In response to the letter dated 5 January 2017 from the ICANN Commercial Stakeholder Group to ICANN's CEO and Chairman of the Board,

First I would like to thank the Commercial Stakeholder Group (CSG) for expressing interest in accessing data sets related to ICANN's mission and apologize for the delay in response. Although this is not the first time such a request has been made, it helps add focus and relevance to the ICANN organization's efforts to make ICANN's data more available.

For background and as the CSG may be aware, the ICANN organization and the ICANN community have been generating and accumulating data related to our mission since ICANN's creation. While the ICANN organization has published significant amounts of data and information in various forms on our website, those data have often been difficult to find, not in a form that is easily consumed in an automated way, and may not have been consistently updated. In response to requests from the community going back as far as 2014, ICANN's Office of the CTO has undertaken a project known as the Open Data Initiative Pilot. There are two components of the initiative that relate to the CSG's letter. First is the need to prioritize implementing open access to data. The CSG's letter helps our evaluation of project priorities and, in particular, which data sets are of interest to the community. Secondly, the list of data sets in which the CSG expresses interest informs the development of a data catalog, that is, an authoritative list of all data sets within ICANN, both the organization and the community. Creating this comprehensive data catalog is obviously one of the most important steps in the Open Data Initiative and an area of active work within the organization.

In addition to the components of the initiative already mentioned, a significant effort is underway to identify the appropriate technology or technologies that can be used to make the identified data sets available. For example, additional software will be required to facilitate data access via public Application Programming Interfaces (APIs) and/or other mechanisms, and some transformation of those data is to be expected. Further, in order to ensure access to data is sustainable over the long term, software will need to be built or obtained that automates the



process of making those data available. Automating this process is critically important, especially to keep data as current as possible, and doing this efficiently and in a cost-effective way depends on having a good technology fit.

Regarding the specific data sets mentioned in the CSG letter, they include both data on hand as well as data that are constrained or not easily obtained. Of the data sets mentioned by the CSG:

- *Historical and ongoing zone file data*

As the CSG is undoubtedly aware, a requirement of top-level domain registries within the new gTLD program as well as some legacy zones is to provide zone data to requesters via the Centralized Zone Data Service (CZDS). As of this writing there are 1,221 TLDs that make their zone data available via the CZDS program. However, many, if not all, registries consider their zone data to be commercially sensitive and require those interested in accessing that data to enter into contractual agreements that specify how those data may be used. Further, the ICANN organization has no mechanism by which we can require ccTLDs to provide their zone data, nor does the ICANN organization have access to zone data beyond the top-level (zones operated by the ICANN organization excepted). Part of the Open Data Initiative Pilot is to explore whether we can make these data available and, if so, under what conditions.

With respect to historical zone data, the ICANN organization has maintained an archive of generic TLD zone data collected from CZDS and other sources since 2014 for internal use and may have access to older zone data for some domains. We are currently investigating the terms and limitations on the non-internal use of that data.

- *Historical and ongoing WHOIS system performance and compliance data*

Some clarification is necessary on what the CSG is specifically referring to in relation to “WHOIS system performance and compliance data.” As the CSG is no doubt aware, the implementation of the servers used for access to registration data represents a widely-distributed database operated by many organizations around the world and the ICANN organization has no direct access to performance data for Whois servers we do not directly operate. In addition, the “WHOIS system” can be interpreted to include Whois servers that provide registration data associated with the Internet numbering system. If the CSG has interest in Whois server performance data outside of domain names, we can enquire with the Regional Internet Registries about the availability of data related to the Whois systems they operate.



The Internet Corporation for Assigned Names and Numbers

However, making the assumption that the CSG is referring to data collected by the ICANN “Service Level Agreement Monitoring” (SLAM) system, the ICANN organization currently maintains approximately two months’ data from that system on an ongoing basis (that is, data older than two months are discarded). We intend to explore which of these data are useful to be made available via the Open Data Initiative Pilot and are working to maintain more than two months’ worth of that data.

- *Historical and ongoing Compliance data*

By “Compliance data”, we assume the CSG is referring to data collected by the ICANN organization’s Contractual Compliance department. While a large cohort of data related to the operation of the Contractual Compliance department is publicly available on the Contractual Compliance section of the ICANN.ORG website, the format of those data is typically not ideal for the purposes of the Open Data Initiative Pilot. Conversion of those data into a more useful format is being investigated. In addition, as part of the data cataloging effort, we intend to identify other Contractual Compliance data to determine what can be made public, either in full or in redacted form, as part of the Open Data Initiative Pilot.

- *Anonymized pricing data*

We assume the CSG is interested in pricing data associated with obtaining second-level domain names. The ICANN organization does not have data related to pricing of these domains. The Office of the CTO is exploring potential ways in which pricing data can be obtained via public, unbiased, and repeatable means. If a sustainable mechanism can be identified by which pricing data can be obtained, it will be included in the Open Data Initiative Pilot.

- *Domain block list data*

The ICANN organization does not collect, compose, create, or maintain its own domain block list data. Within the Office of the CTO, we make use of block list data that are published via the DNS or made available as repositories or data feeds from third-party providers for research purposes. Some of these data are public or open resources whereas other data feeds are offered under commercial (for fee) licenses. It would be useful if the CSG clarified whether they believe the ICANN organization should collect and archive these data, at least the public resources, with the understanding that the data archived may be voluminous and will require ongoing operational expenditures to ensure upkeep.



- *DNSSEC implementation/deployment data beyond the top-level*

Presuming this is referring to the scale of DNSSEC implementation and deployment across the entire DNS tree from the root to the leaf nodes, without availability of zone data, it is difficult to determine how many DNS zones are signed. While the ICANN

organization has access to the root zone and most top-level domains, access to zone data further down the DNS tree is challenging. The Office of the CTO has done and is doing some preliminary research into DNSSEC deployment beyond the top-level and is investigating ways in which DNSSEC implementation and deployment below the top-level can be measured in a sustainable, repeatable, and unbiased way. Assuming we are able to identify efficient and sustainable ways of collecting these data, they will be included in the Open Data Initiative Pilot.

- *Domain abuse data*

As with Domain block list data, the ICANN organization does not actively collect domain abuse data but instead relies on commercial or open sources for reputation or abuse data. The Office of the CTO contracts with outside parties that do collect and maintain that data, however the ability to re-distribute those data is often constrained by the legal agreements under which the data has been obtained. The Office of the CTO is actively working on developing a reporting system that will provide analytics or reports on security threats mentioned in the GAC Beijing Communique, i.e., “such as pharming, phishing, malware, and botnets”, as well as other threats, including spam. We are exploring how to make those data publicly available.

More generally, for any data set that is collected by the ICANN organization (and unencumbered by commercial or other re-use licenses), the long-term goal is to make that data publicly accessible in a straightforward, machine-readable form and in accordance with the general definition of Open Data efforts. Limitations to that goal obviously exist, as is noted in the CSG letter, including privacy concerns and other data protections, as well as policy-based restrictions and constraints based on agreements, both between ICANN and the vendors of data as well as imposed by Registry and Registrar agreements, and these limitations must be honored. As such, it will take time and resources to implement this goal, but the order in which data sets are made public will be influenced by the community.



The Internet Corporation for Assigned Names and Numbers

In summary, as may be suggested by the ICANN organization undertaking the Open Data Initiative, we agree with the sentiment that making data available in an open way can help inform the community about the activities ICANN, both the organization and community, are engaging in and can improve ICANN's transparency and accountability. We look forward to working with the CSG and other stakeholders in both identifying useful data sets as well as prioritizing access to those data sets via the Open Data Initiative.

Best regards,

Göran Marby
President & CEO, ICANN