The DNS Core Census

ICANN Office of the Chief Technology Officer

Edward Lewis
OCTO-019
23 November 2020

TABLE OF CONTENTS

This document is part of ICANN's Office of the Chief Technical Officer (OCTO) document series. Please see the OCTO publication page for a list of documents in the series. If you have questions or suggestions on any of these documents, please send them to octo@icann.org.

Executive Summary

The DNS Core Census is a gathering of information about the global public DNS root zone, its delegations, and delegations from the “arpa” top-level domain. The value of creating the DNS Core Census is in simplifying access to information contained in many other data sources related to these domains.

The DNS Core Census covers a portion of the global public DNS and is designed to enable useful and stable, long-term measurements. The portion covered can be described as the root zone, ccTLDs, gTLDs, and the "reverse map," as well as other selected elements. The embodiment of the DNS Core Census is a set of resources available for analysts to use in automated scripts and programs.

1 Introduction

The DNS Core Census is a gathering of information about the global public DNS root zone, its delegations, and delegations from the *arpa* top-level domain. The value of the DNS Core Census is in simplifying access to information contained in many other data sources related to these domains.

The DNS Core Census is not itself an analysis or statistical study. It is meant to be a reference for other studies. Just like a road map is useful in an automobile journey, the census may highlight interesting points for further detailed study or help label legs of the journey to help clarify what is observed.

Most domain name identifiers are present in the global public DNS, with a few exceptions being names serviced by other (non-DNS) systems. For more information see the section titled "Category 'IETFSpecialUse.'" The inclusion of the exceptions in the DNS Core Census is to acknowledge that such names are explicitly not serviced by the DNS.

The DNS Core Census is designed, first and foremost, to support measurements. Inclusion of a zone or name server or address does not convey any special status. An analogy can be drawn from the concept of a "census designated place," which is used by the United States Census Bureau as a statistical measurement to account for people living in areas that are not defined by a formal boundary. For example, the measurement is applied to people who are not living in an incorporated city.[1] The "census designated place" label provides the ability to draw data-based conclusions without designating any formal authority over the region that is being measured. In this sense, coverage in the census does not imply any other designation.

The DNS Core Census is initially embodied in structured data files. An automated script or program can read the files and have the census available in a data structure to either drive observations of data or to categorize results. In the future, the same information may be available via an API.

Potential use cases include a script intended to observe the current response size from queries for DNSKEY resource record, which is set in use by gTLDs using the census to determine which

[1] “Census designated place,” Wikipedia, https://en.wikipedia.org/wiki/Census-designated_place

queries are needed for a comprehensive survey. Another script might use the census to determine whether elliptic curve cryptography is more popular in gTLDs or ccTLDs as an example of categorizing results.

This document supports ICANN's strategic goal to improve the shared responsibility for upholding the security and stability of the DNS by strengthening DNS coordination in partnership with relevant stakeholders. It is part of ICANN's strategic objective to strengthen the security of the domain name system and the DNS root server system.

1.1 Motivation

The idea behind the DNS Core Census is based on a history of Internet measurement projects. Some are long-running while others are short runs or one-off measurements. What these projects have in common is that they need to rely on external parameters to help interpret the results and guide the analysis.

1.1.1 Example: DNSSEC in TLDs, from 2011 onward

An early clear example of the need for the DNS Core Census comes from measurements of DNSSEC adoption across TLDs. Prior to 2013, it was up to individual TLD operators to deploy DNSSEC; a few did but most did not. Adopting any new technology into operations comes with debate; some people push for innovation and others want to preserve the status quo. Thus, one way to judge the quality of an innovation is to measure its uptake by operators.

Since October 2013, new gTLDs were delegated with a contractual requirement to deploy and maintain DNSSEC. By 2016, these new gTLDs comprised three-quarters of all TLDs, and the metric used to measure DNSSEC adoption showed a meteoric rise.

By graphing the percentage of TLDs signed, as a metric of DNSSEC deployment from 2011 to 2020, it appears that DNSSEC had taken off quickly between 2014 and 2017. This is shown in Figure 1.

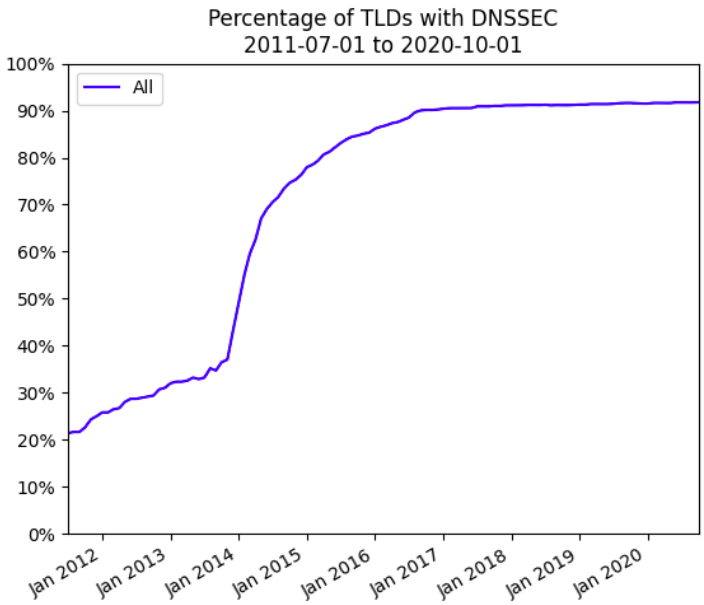

Figure 1

This measure is an overenthusiastic estimation of the acceptance of DNSSEC. Growth is fueled by contractual requirements in the new gTLDs and the large number of new gTLDs. Figure 2 removes the new gTLDs from the calculations (in red) and is arguably a better estimation of DNSSEC acceptance.

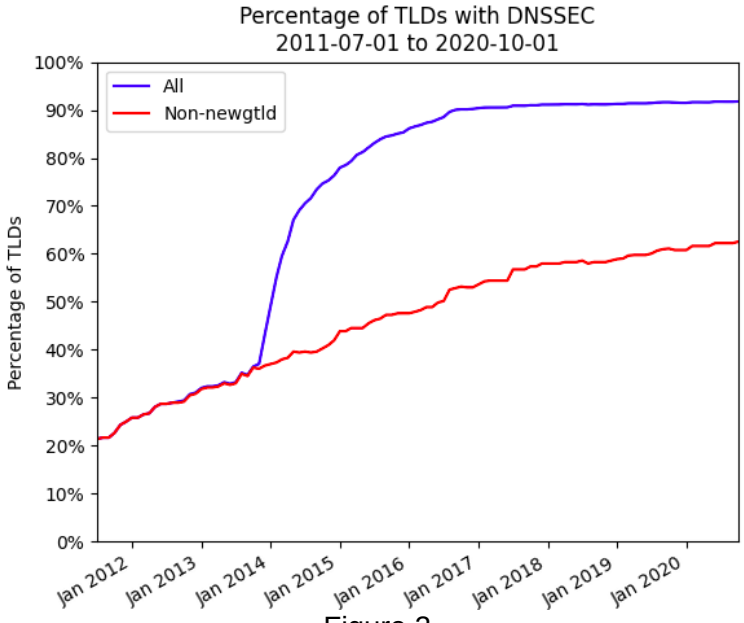

Figure 2

In Figure 2, the red line represents the zones that were not part of the 2012 New gTLD Program. To accomplish this, one needs to know whether a TLD is a "new gTLD" or not. This was the beginning of the need to include external parameters.

1.1.2 Example: Measuring Internationalized Domain Names

It has been long noted that ccTLDs and gTLDs operate in different ways, and the measurements bear this out. There was a time when all ccTLDs were named with two ASCII characters, with everything else being "not a ccTLD." For the most part, distinguishing further could be done manually or with static lists as there were only about 24 other TLDs.

Internationalized Domain Names (IDN) TLDs changed that. One can not "look" at an IDN TLD and determine whether it is an IDN ccTLD or an IDN gTLD. Early methods to distinguish between the two involved looking for the name in a list of gTLD contracts published by ICANN. Alternatively, the IANA website was used, but this was slow due to the need to parse textual web pages. Eventually, a direct reference to the IANA TLD database was developed. This work was the direct predecessor to developing the DNS Core Census as the wealth of other available meta-data about each TLD emerged.

1.1.3 Example: Counting Route Origin Attestations

A study of the deployment of Route Origin Attestation (ROA) certificates has drawn attention to the benefit of expanding studies beyond the zones delegated from the DNS root zone. The focus in this study is the routing layer, which supports more than the TLDs listed in the root zone. This study benefits by also investigating the domains and zones that are managed by TLDs as part of the same registry, as well as examining the DNS that is called the "reverse map," the (sub) roots of which are operated by the Regional Internet Registries (RIRs).

Names and zones below a TLD name that are operated as part of the TLD registry are called many things. They may be called sub-TLDs, affiliated, or federated registries. For the DNS Core Census they are labelled sub-ccTLDs and sub-gTLDs. Because of the lack of access to zone files for many TLDs, discovering these zones is done with a heuristic approach, assuming that for any TLD, the TLD substructure uses at least one of the name servers as the TLD top zone.

The RIRs maintain databases of IP addresses and delegate based on these databases and DNS zones whose primary use is to map an address to a hostname. This is colloquially known as "the reverse map" as it is the reverse of mapping hostnames to addresses. Despite the variety of data entered, the reverse map portion of the DNS functions in the same way as the names, also known as "the forward map," and deserves the same consideration in measurements.

These realizations caused the idea of the DNS Core Census to expand beyond the TLDs of the DNS root zone in order to provide a more complete picture of the DNS.

1.1.4 Where Does This Stop?

For the DNS Core Census to be manageable and useful, it has to have a boundary. It has to have a size that is easy to handle and understand as a basis of informative measurements and analysis.

The boundary is labelled "the commercial registration boundary." A precise definition is hard to give, but it is more or less where a delegation is obtained for a fee. Note, there may also not be a monetary fee.

There may be boundaries beyond the commercial registration boundary, with a "beyond" from the perspective of the root zone. These boundaries are not considered to be in the DNS Core, and thus they are not present in the census. The existence of these boundaries and their exclusion from the census is an example of why the census is not a replacement for the PSL[2], a resource used in the development of the census and which is mentioned later in the document.

In the forward map, the boundary is where a registrant would purchase a name from a TLD registry or one of its affiliates registrars. This could be a "name.example" or "name.edu.example" or even "name.school.city.prefecture.example." The level depends on the substructure of the ccTLD or gTLD.

In the reverse map, the boundary is essentially between an RIR and a service provider or enterprise managing a block of addresses. There are many scenarios, however, which make this definition far from clear. The census discovery process relies on examining the operators of the name servers to draw the line.

The notion of a commercial registration boundary is potentially useful. It identifies one boundary of authority in the DNS, but not the only boundaries.

1.2 Criteria for the DNS Core Census

The membership of the DNS Core Census is determined via a discovery process with these goals in mind:

- Closely related to the DNS root zone
- Long-term stability, short-term flexibility
- Unbiased, representative inclusion
- Lightweight compilation
- Repeatability

Closely related to the global DNS root zone places a focus on the portion of the DNS that is shared the widest. Starting with the DNS root zone, and descending through delegations to other zones, a view into the most widely shared section is gained. Associated with these zones are name servers that publish the zones and the addresses of the name servers.

Choosing to focus the DNS Core Census on the DNS root zone and the zones close to it is based on the assumption that DNS operations for these elements will exhibit a higher bar of performance than general DNS elements. Work to date has upheld this assumption.

Long-term stability means that elements of the DNS Core Census ought to be included or excluded consistently over the elements' lifetimes. This should help ensure that comparisons, to as great an extent as is possible, can be made over long periods of time.

Short-term flexibility refers to allowing elements to be created, new reverse map zones created and deleted, and gTLDs retired.

[2] “Public Suffix List,” https://publicsuffix.org/

Unbiased, representative inclusion means that there is a consistent application of criteria, and that it includes the major divisions over the years, namely gTLDs, ccTLDs, and the reverse map. There are other categories beyond these that are included mainly for consistency, even if an included element is not heavily used in operations.

Lightweight collection means that the process of maintaining the census can keep it up to date without undue burden on the global public Internet. This goal clashes somewhat with the goal of long-term stability due to the rare occurrences of misconfigured systems.

The clash of goals is based on this observation. To heuristically verify that a sub-TLD candidate is truly a subTLD, a query is sent to one of the TLD name servers. So long as all of the TLD name servers are synchronized, one query is sufficient and quickly leads to a correct determination. If the name server queried is not properly synchronized, the heuristic test may fail and result in fluctuations in the list of zones included from run to run. The only way to ascertain whether the queried name server is properly synchronized is to repeat the query to other name servers of the TLD, which is opposed to a goal of lightweight compilation.

Based on observations that the incidence of out-of-synchronization servers is small, choosing a lightweight compilation may make sense. In addition, focusing on the fluctuating names can pin-point name servers deserving some operational attention.

Repeatability refers to making use of data that is publicly available, so that anyone can generate a copy of the DNS Core Census. There are a few zone files that are used by special arrangements.

1.3 Organizations Administering Elements of the DNS Core Census

While not used as a starting-point, it can be noted that the DNS Core Census consists of zones or registries administered by the following organizations, along with name servers, addresses, and routes managed directly or indirectly via various business relationships.

- The Internet Engineering Task Force (IETF)
- The Internet Corporation for Assigned Names and Numbers (ICANN)
- The Number Resource Organization (NRO)
- A collection of organizations operating ccTLDs
- The Internet Architecture Board (IAB)
- A handful of other top-level name operators and legacy appointments

Zones used in the operation of the organizations themselves are not included in the DNS Core Census. For example, icann.org and ietf.org are zones for operating ICANN and IETF, respectively, but are not included in the DNS Core Census.

2 Elements of the DNS Core Census

As of July 2020, the DNS Core Census contains approximately 4,500 domains, 4,700 name servers and makes use of 7,300 addresses. These numbers will fluctuate from day to day.

2.1 DNS Zones

The global public DNS root zone is the top zone in the DNS name hierarchy. The root zone is a product of the root registry, which is a database operated by the IANA Functions Operator. The root registry is directly modified by updates to it and is indirectly influenced by the Special-Use Domain Names Registry[3] managed by the IETF.

2.1.1 Top Level Zones

Beginning with the DNS root zone, all delegations from it are considered part of the DNS Core Census. The .arpa zone is special: all of its delegations are also considered part of the core, and these prominently include the top of the IPv4 and IPv6 reverse map trees.

2.1.2 Subordinate Top Level Zones and Names

Within some ccTLD zones and a few gTLD zones there are substructures to the zone. One of the most familiar examples is "co.uk." under "uk." Many cases of this are fairly simple. Within a TLD there could be names for "com.TLD," "edu.TLD" and so on, emulating the structure found in the root zone. There are a few cases where the substructure is extensive, perhaps listing names for all of the municipalities in a jurisdiction.

Within the IPv4 and IPv6 reverse map regions, zones managed by the RIRs, National Internet Registries (NIRs), and ICANN are included. The RIRs are identified as operators via the names listed in the "Responsible Name" field of Start of Authority (SOA) resource records.

The census discovery process attempts to include this substructure. Because of the lack of the zone files for many of the cases, the substructure is discovered heuristically in the PSL and through other research. There are alternative approaches for example to obtain, opportunistically, zone files not currently put to use.

2.1.3 Metadata of DNS Zones

The metadata of a zone includes the category (gTLD, ccTLD, reverse map), the status (active or not), the jurisdiction of the zone (usually designated by a two-letter ISO code if applicable), dates of registration, and so forth. This information is most useful when looking for measurement patterns across the DNS core; for example, whether DNSSEC adoption is higher among gTLDs or ccTLDs. All told, there are dozens of metadata fields, with the details of this left to deeper technical documents as the field names may change and as the DNS Core Census matures.

2.1.4 Configuration of DNS Zones

Configuration parameters include the name in the SOA resource record identifying the responsible party of the zone, the set of name servers, DNSSEC chain-of-trust (DS and

[3] “Special-Use Domain Names,” https://www.iana.org/assignments/special-use-domain-names/special-use-domain-names.xhtml

DNSKEY) resource records, and signatures. This information allows for some superficial examinations of DNS core operations.

2.1.5 Commercial Registration Boundary of DNS Zones

Within a DNS zone, a name may be delegated to another person or organization that is unrelated to the zone. This is common in many DNS core zones as they are open to registrations. But there are cases where some names remain internal to the zone or to the registry.

Some registry operators are open to "third-level" registrations, such as "agency.gov.example." Typically, these are seen in ccTLDs. Sometimes the names like "gov.example" are delegated to separate zones and sometimes they are managed as part of the same DNS zone. This difference may be significant, but not always.

In the reverse map zones, particularly related to IPv6, the commercial registrations, representing address space allocated to a connectivity provider, may include many levels. The "longest" chain of labels defines the commercial registration boundary.

Within a DNS zone, lists are maintained for sub zones delegated "back" to the registry's own name servers, undelegated sub domain names, and names with a CNAME or DNAME resource record set. This enables drawing of the commercial registration boundary.

The significance of these names is that the commercial registration boundary may be deeper than just the top label in a TLD name. The commercial registration boundary is not a replacement for the PSL because the PSL is far more flexible.

2.2 Domain Names *not in* the global public DNS

Names in the Special-Use Domain Names registry are included in the DNS Core Census and marked as special use. These names are included because research may come across them and need to categorize them appropriately.

2.3 Name servers

Name servers are included in the DNS core despite playing only a supporting role in the identifier system. There are a number of motivations for this.

Name servers are listed twice at each delegation point, in the parent zone at the cut point of the naming authority and in the child zone's apex. The set of name servers in each place are supposed to be the same, but this is not always the case. Inconsistencies are of interest in measurements examining operational performance.

Name servers, at the name level, may be shared by many zones. A registry operator who runs a collection of TLDs might use the same named servers for all zones, in fact the largest operator in counting zones uses the same four name-server records for each. By including this, it is possible to see the degree of consolidation of registry operations and how this plays a role in the adoption of changed technologies or practices.

By tracking name servers in addition to zones, the complexity of relationships within the DNS core begins to emerge. Many zones are owned or registered to one organization who then subcontracts to a registry back-end operator to maintain the technical aspects. In turn, the back-end operator might contract for DNS hosting services to run name servers and eventually for an Internet service provider (ISP) to provide network address and routing.

2.4 Addresses

Addresses also play a supporting role, as sometimes, an operator of name servers might map many name servers on to the same address. Like name servers, this can indicate a consolidation of operations.

Associated with addresses are route origins, a route origin being a Border Gateway Protocol (BGP) route prefix and a last-hop autonomous system number. A route origin refers to the pairing of a prefix and an autonomous system number. Many addresses may be serviced by a single route origin, indicating a shared configuration.

3 The Categories within the DNS Core Census

Viewing the DNS Core Census begins with zones and, in particular, each zone's category. This section provides an overview of the various categories. The categories of the DNS Core Census do change over time. The numbers or examples cited are correct at the time of writing.

3.1 Category "ccTLD"

In this category, there are zones associated with jurisdiction and either follow the ISO 3166 two-letter assignments or are IDN representations. For example, CZ represents a region in Europe, and XN--90A3AC represents the same region as RS, but in Cyrillic (СРБ). There are 317 of these zones; this count includes some retired ccTLDs.

3.2 Category "sub-ccTLD"

In this category, there are zones delegated from a ccTLD whose name servers overlap with the same ccTLD. Not all ccTLDs have affiliated sub-ccTLDs, some have many. The total number, across all ccTLDs of these zones is 2,394. These zones are detected via probing, a process which relies on some heuristics and, thus, may change over time.

3.3 Category "gTLD"

In this category, there are zones operated either under contract from ICANN or via other historic arrangements. Examples include COM, EDU, IDNs that are not ccTLDs, and members of the 2012 round of the new gTLD program. There are 1,275 of these zones, and not all are active.

3.4 Category "sub-gTLD"

Eleven gTLD zones have structures mirroring "sub-ccTLD" zones.

3.5 Category "revMap"

This category includes the top delegations in the reverse map zones. IPv4 delegations begin with in-addr.arpa, and IPv6 begins with ip6.arpa. There are a total of 271 zones included in this category, although this number is poised to grow as IPv6 is further allocated.

IPv4's top delegations correspond to the first "quad" in the address – for example, in "192.0.2.5" the first quad is "192" and there is a zone 192.in-addr.arpa. Related to the history of IPv4 address allocations, some of the first-quad zones are not run by the Number Resource Organization, RIRs, NIRs, or ICANN.

IPv6's top delegations depend on the original allocations of the address space. Besides the NRO, RIRs and affiliates, there are a handful of delegations to others, again for reasons of history.

3.6 Category "noncore-revMap"

These are zones of the reverse map that are delegated to entities outside the NRO and RIR system; for instance, organizations that have "old class A" space in IPv4. There are 21 of these zones, but this number may slowly shrink over time. Although these zones are not run by the NRO system, they are included for any studies that may want to examine the delegations.

3.7 Category "IETFSpecialUse"

Names marked here are not necessarily zones but may be. There are 37 entries. Some are names designated to be used external to the DNS and some are names in the DNS that behave in a unique or unexpected manner. An example of this category would be “onion.”

3.8 Categories related to phone numbers

The categories "enum" and "noncore-enum" designate zones below the .arpa zone set aside for a particular protocol. There is one "enum" zone, e164.arpa, managed by one of the RIRs; the rest are delegated to managers of appropriate spaces, delegated in a manner related to telephone numbering. These zones are included because of their structural similarity to the reverse map zones and for consistency across .arpa delegations.

3.9 Category "tTLD"

There are eleven zones dedicated to testing that are still present in the Root Registry, none of which have been actively delegated in recent memory; hence they may be otherwise forgotten. These zones were used to test the IDN mechanisms prior to the use of IDN for ccTLDs and

gTLDs. The designation of "tTLD" stands for a test TLD and may be reused if there are more tests conducted.

3.10 Infrastructure Categories

There is one zone marked "root" and seven marked "arpa." The DNS root zone is specially marked for its uniqueness. The zones marked as "arpa" include the “arpa” zone itself and the delegations to zones that are neither the reverse map nor related to the enum, such as home.arpa.

3.11 Name servers

There is approximately the same number of name-server entries as zone entries, which may be a surprise. This is due to name server sharing among zones.

3.12 Addresses

There are approximately 7,000 addresses in the DNS Core Census. The census does not explicitly differentiate between IPv4 and IPv6; this is left to analysis programs to parse.

4 Closing

The DNS Core Census is meant to be a measurement construct and does not convey any other designation to the elements listed within the core. The DNS core is not meant to capture the critical elements of the Internet nor the busiest and most referenced elements. The goal is to help measure the structure of the DNS and put those measurements in some context.

The concept is still in its infancy; selections of what is in and what is out of the DNS Core Census as well as what is represented in the census are still open to discussion.

5 Availability

For the time being, availability will be limited to a prearranged basis. Contact DNSCoreCensus@icann.org to request access to the DNS Core Census files and documentation. In the early stages, names (keys) of fields and values of fields may be updated as well as what is included or omitted, based on feedback.