Identifier Technologies Health Indicators (ITHI) Retrospective and Proposal

ICANN Office of the Chief Technology Officer

Alain Durand
OCTO-025
8 July 2021

![ICANN logo]

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.

This document supports ICANN's strategic goal to support and grow active, informed, and effective stakeholder participation. It is part of ICANN's strategic objective to improve the effectiveness of ICANN's multistakeholder model of governance.

1   Introduction

The Identifier Health Technology Indicator (ITHI) project is an ICANN organization initiative. It started at the March 2016 Marrakech ICANN meeting as part of the ICANN 2016-2020 Strategic Plan. ITHI was associated with strategic objective 2.1: "Foster and coordinate a healthy, secure, stable, and resilient identifier ecosystem."

Now, in 2021, it is time to take stock of what has been accomplished and propose a new strategy going forward.

The ITHI project is strictly technical. Its role is to define metrics, perform measurements, and assess trends over time in important areas related to the Internet's identifier systems. ITHI helps inform the ICANN community by providing verifiable data on the evolution of the system of identifiers that ICANN helps to coordinate.

Unlike many Internet measurement projects, ITHI is not a one-time effort. It is designed to track a comprehensive series of metrics over a long period of time. As seen in Section 2 of this document, ITHI shows that while some metrics related to the operation of the DNS remain remarkably stable, other metrics have large variations that can be traced back to the action of a few players on the Internet.

2   ITHI Phase 1

Early on, the ITHI project envisioned encompassing all identifiers coordinated by ICANN: names, IP addresses, and protocol parameters. After several discussions, the Internet Engineering Task Force (IETF) decided not to participate. The Regional Internet Registries (RIRs) decided to take the numbering portion of the project as a separate Number Resource Organization (NRO) activity and produced a document describing the indicators they intended to track. ICANN org's Office of the CTO (OCTO) focused on name identifiers, i.e., the Domain Name System (DNS).

In the case of the DNS, the chosen approach was to define a set of problems, then define metrics that could provide relevant indications of "health," and only then get data to measure those metrics. The idea was to collect measurements over a long period of time, defined as a minimum five years. This original approach contrasted with most existing measurement projects in two ways:
- ⊙ OCTO started with the problem definition and ended with collecting data, as opposed to the other way round; and
- ⊙ OCTO did not envision a one-off measurement campaign, but a sustainable system in which a particular measurement at any point in time would not be as interesting as the trend of that measurement over time.

In 2017, OCTO started to build a supporting infrastructure. The first step was implementing the measurement tools and the back-end system to analyze the data, followed by finding partners that would run the tools, collect data, and generate reports, and finally, building a basic website to display results.

The ITHI project had taken privacy concerns into account before the European Union General Data Protection Regulation (GDPR) was enacted. As a result, OCTO had already put in place a number of safeguards to make sure no personal data is entered the system.

The long-term direction had been to publish ITHI measurements through the ICANN's Open Data project, so little effort was put into the preliminary ITHI website beyond basic reporting. Measurements have now been active since January 2018. The ITHI project has a number of partners all over the world, including in Ghana, Argentina, Kazakhstan, and Indonesia. The ITHI website is available at https://ithi.research.icann.org.

Eight metrics are defined and tracked, and each has a number of submetrics:
1. M1: Inaccuracy of Whois Data
2. M2: Domain Name Abuse
3. M3: DNS Root Traffic Analysis
4. M4: DNS Recursive Server Analysis
5. M5: Recursive Resolver Integrity
6. M6: IANA registries for DNS parameters
7. M7: DNSSEC Deployment.
8. M8: DNS Authoritative Servers Analysis

ITHI can help us shed some light on some little known facts about root zone queries. It is well documented that a very large number (over 75%) of queries to the root are for top-level domains that do not exist. What was less understood until recently is why that was the case, or what were the sources of such queries. Multiple reports in 2020 showed that the origins of a large number of those queries were probes sent by the Google Chrome browser and systems derived from Chromium (the open-source software on which Chrome is built). Since then, Chromium has modified its code to reduce the impact. ITHI can help us track this. In early 2020, those probes were making about 45% of the total number of DNS queries sent to the root. In June 2021, ITHI reported that this number was down to about 21%.

Another example of the results from ITHI is a study on DNS resolver concentration, or how many distinct DNS resolvers it takes to serve 50% or 90% of all Internet users. The lower the number of DNS resolvers, the higher the concentration. Tracking the evolution of this concentration was important in the context of the deployment of technologies such as encrypted DNS DoH (DNS over HTTPS, RFC 8484) and DoT (DNS over TLS, RFC7858). ITHI metrics M5.6.1 and M5.6.2 show that it takes about 200 distinct DNS resolvers to serve 50% of all Internet users and 2,000 of them to serve 90% of all Internet users. More interestingly, these numbers have been relatively stable over the last two years.

3   ITHI Phase 2

3.1   Known issues

There are some known issues with ITHI phase 1:
- The health idea has not caught on in the community, since that can imply a value judgement of "healthy" or "sick."

- The web site lacked modern design elements. This was fixed in early 2021.
- The number of partners involved in collecting statistics at recursive resolvers and top-level domains is not large enough to have confidence in the data.
- The community's understanding is evolving regarding what is worth tracking. There are other interesting data, beside the original eight metrics, that could be included.
- The data aggregation level is sometimes too coarse, making fine grain analysis difficult.

OCTO is proposing to evolve the ITHI project in three directions:
- Redesign the website
- Find more partners in order to collect metrics related to recursive resolvers and top-level domains (TLDs) authoritative servers
- Define and track new metrics

3.2 ITHI website

Although the ingestion of data for the current metrics is almost entirely automatic, the process to add additional data sets is manual and should be simplified. The web site should include pre-built interactive data visualization tools. Such tools could, for example, allow zooming in to specific date ranges or filtering out some values.

3.3 Outreach

Outreach is a critical element for this project, both to increase knowledge of the project inside and outside of ICANN, but also to help find new partners to collect data. This is especially important for measurements at recursive resolvers and measurements at TLD authoritative servers.

Until recently, outreach efforts have been limited mainly to various conferences attended by OCTO staff. OCTO plans to increase the level of its outreach efforts and leverage all the different channels ICANN has to engage with the community. In particular, the following organizations can be leveraged to engage with potential new partners, Internet service providers (ISPs) and TLDs:
- ICANN org's Global Stakeholder Engagement (GSE) team, Government Engagement (GE) team, and its executive team
- ICANN TLD operator groups
- ICANN Internet Service Providers and Connectivity Providers (ISPCP)
- ICANN Generic Names Supporting Organization (gNSO)
- ICANN Country Code Names Supporting Organization (ccNSO)

One particular idea to explore is to create and support a new community that combines active partners' contributing ITHI data with active researchers making use of the data.

3.4 New metrics

Some new metrics are suggested through work undertaken in phase 1: while analyzing some of the data about root servers, OCTO discovered that there was a very large variability in the data collected from the 200+ ICANN Managed Root Server (IMRS) instances. For example, the bulk of the traffic at some IMRS instances is made of Google Chrome induced traffic, while some

other instances see mostly traffic from resolvers that do not perform any caching. New metrics could be defined based on this study.

Similar variability studies could also be done for other metrics, in particular the ones computed for recursive resolvers and for TLDs authoritative servers.

There are other potential areas of interest in which new, or more refined, measurements could be done. Here is a non-exhaustive list:
- ⊙ Domain name generation algorithms (DGA)
  - Can those be detected at the root?
- ⊙ Domain search lists
  - What is the impact on the root of such configuration options?
- ⊙ DNS concentration (e.g. how many independent pieces of infrastructure are supporting 50% or 90% of all domain names)
  - This would be extending the study about DNS resolver concentration to other pieces of DNS infrastructure
- ⊙ Domain Name System Security Extensions (DNSSEC)
  - Understanding the usage of DNSsec better, not just its deployment.
- ⊙ Extension Mechanisms for DNS (EDNS)
  - Prevalence of such options in DNS queries
- ⊙ Qname minimization
  - Tracking the adoption of such a technique
- ⊙ Encrypted DNS services (DoH/DoT)
  - Tracking the adoption of those technologies

This project can provide a unique value by observing these trends from multiple points of view: the root, authoritative servers, recursive resolvers, individual clients, and the publication of data in DNS records.

4   Conclusion

The ITHI project is a valuable tool for the community. It enables important trends to be tracked, such as resolver concentration or the impact of Chromium probes on root traffic. However, experiences gained in Phase I have suggested that a second phase would increase the tool's value. This new phase will enable both the collection of data from new partners and the opportunity to develop the current set of metrics.