Email Address Internationalization (EAI) Support Survey tool

Registration Operations Workshop 11

Eduardo Alvarez

GDS Technical Services

ICANN
21 June 2022

ICANN

EAI Support Survey Tool

- **Purpose:**
  Survey Email Address Internationalization (EAI) support on mail exchanger (MX) servers of second level domain names across gTLDs.

- **Input:**
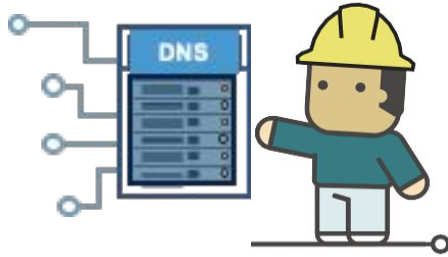 - gTLD zone files
  - public DNS

- **Output:**
  Lists classified per gTLD, country and EAI support verification results of:
  - SLDs
  - MX servers
  - IP addresses

Support Survey Process

1

For each gTLD:

- Get the list of SLDs in the zone.
- Query the DNS for existing MX records in the SLDs and their corresponding IP addresses.
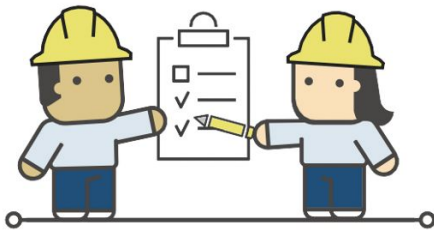
2

For each MX IP address:

- Retrieve IP geolocalization details.
- Test IP address using 2 test IDN addresses:
    - **non-ASCII@a-label-domain**
    - **non-ASCII@u-label-domain**
    - **EHLO** command and look for "SMTPUTF8".
    - **MAIL FROM** command and validate for a "250" response
    - **QUIT**
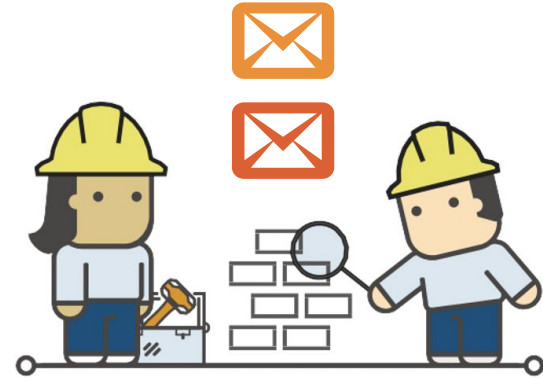
3

Store result for MX EAI support:

- **Full**: All IPs passed.
- **Partial**: Some IPs passed.
- **None**: No IPs passed.

Test Example

Connect to email server

EHLO test

MAIL FROM test (u-label)

Terminate connection

```
C: $ telnet 35.80.244.233 25
   Trying 35.80.244.233...
   Connected to mx1.testo.tel.
   Escape character is '^]'.
S:     220 mx1.testo.tel ESMTP Postfix
C: EHLO client.example.com.
S:     250-mx1.testo.tel
S:     250-PIPELINING
S:     250-SIZE 51200
S:     250-VRFY
S:     250-ETRN
S:     250-ENHANCEDSTATUSCODES
S:     250-8BITMIME
S:     250-DSN
S:     250-SMTPUTF8
S:     250 CHUNKING
C: MAIL FROM: <Почта@испытание.онлайн> SMTPUTF8
S:     250 2.1.0 Ok
C: quit
S:     221 2.0.0 Bye
   Connection closed by foreign host.
```

Tool Implementation

- **Compiling the tool**

  - Java Compiler version 8 or later.
  - Apache Maven version 3.6 or later.

- **Execution**

  - Java JRE version 8 or later.
  - Docker CE version 20.10.7 or later.
  - MariaDB client version 15.1 or later.
  - SnowSQL version 1.2.16 or later.
  - User must be on the docker group.
  - User must be able to run sudo.

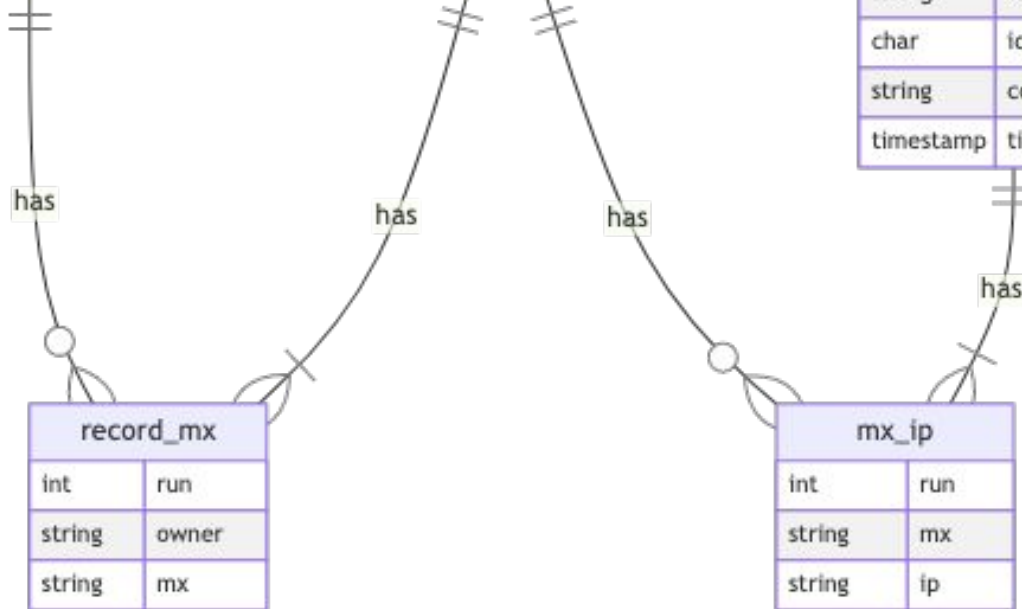- Source code planned to be made available at:
  https://github.com/icann/eai-survey-tool

Data Model

Third Party Services

- The survey tool is configured to run in an AWS virtual environment which allows to keep testing independent from ICANN's internal infrastructure.

- Google's public resolver is used when resolving SLD's MX records in the DNS.

- Maxmind service for IP geolocation data (at the country level) to identify the country associated with a mail server IP address.

- Snowflake is used to persist and process the complete survey results in their cloud data warehousing service.

False Negatives Mitigation

- The server where the survey runs has:

  - Functional SMTP server configured to receive email for the test addresses.

  - Load balancing with 2 IPv4 and 2 IPv6 addresses.

- The survey tool does NOT send any emails.

- IP addresses are tested only once.

- Implemented a maximum number of retries (1) and waiting times (30 sec) for each test performed in an IP.

- Implemented wait times between tests involving IP addresses found within the same /16 IPv4 or /48 IPv6 block.

Known Issues

◉ Automated DNS block lists (e.g. anti-spam "traps")

◉ Email servers with strict security policies that consider any contact from unknown servers as unsolicited traffic causing them to ignore or even report the source IP address as a potentially bad actor.

Survey Results - IP Address

Survey date	18-Oct-2021	06-Jan-2022	07-Apr-2022
Test email script	Arabic	Cyrillic	Han
Processed gTLD zones	1,180	1,173	1,172
Total SLDs	210,811,274	212,989,788	215,885,039
Unique MX servers	34,996,159	35,129,853	35,521,173
Unique IP addresses	2,537,176	2,517,759	2,506,329

- 🟢 Test passed
- 🔴 Test failed
- 🟣 IP could not be tested

18-Oct-2021
- 7.04 %
- 32.33 %
- 60.63 %
- 2,537,176 TOTAL

06-Jan-2022
- 7.24 %
- 32.89 %
- 59.93 %
- 2,517,759 TOTAL

07-Apr-2022
- 7.32 %
- 33.06 %
- 59.61 %
- 2,506,329 TOTAL

Survey Results - MX Server

Survey date	18-Oct-2021	06-Jan-2022	07-Apr-2022
Test email script	Arabic	Cyrillic	Han
Processed gTLD zones	1,180	1,173	1,172
Total SLDs	210,811,274	212,989,788	215,885,039
Unique MX servers	34,996,159	35,129,853	35,521,173
Unique IP addresses	2,537,176	2,517,759	2,506,329

Legend:
- MX Full (green)
- MX Partial (yellow)
- MX None (red)
- Not tested (purple)
- No IPs (dark blue)

18-Oct-2021 chart:
- .02 %
- 6.69 %
- 6.01 %
- 19.52 %
- 34,996,159 TOTAL
- 67.74 %

06-Jan-2022 chart:
- .02 %
- 6.62 %
- 6.33 %
- 19.95 %
- 35,129,853 TOTAL
- 67.05 %

07-Apr-2022 chart:
- .02 %
- 6.64 %
- 6.94 %
- 20.26 %
- 35,521,173 TOTAL
- 66.11 %

Survey Results - Per gTLD

◉ Highest rates per gTLD (with 150,000+ MX servers surveyed)

TLD	Total SLDs	Total MX found	MX EAI full support %	MX with full EAI support	MX with partial EAI support	MX with no EAI support
page.	66,159	174,770	92%	160,787	15	13,968
studio.	95,756	164,959	49%	80,968	45	83,946
dev.	295,330	436,692	41%	180,611	216	255,865
org.	10,581,113	12,573,128	36%	4,526,780	21,923	8,024,425
com.	159,427,821	155,550,297	36%	55,682,494	529,001	99,338,802
design.	121,403	184,557	35%	63,755	136	120,666
app.	625,184	471,945	32%	151,804	216	319,925
email.	112,898	162,234	31%	49,894	664	111,676
tech.	381,665	397,054	28%	110,263	356	286,435
net.	13,177,376	13,056,803	27%	3,535,658	36,174	9,484,971
one.	163,798	264,364	26%	69,126	133	195,105
digital.	119,293	204,214	26%	52,695	115	151,404
life.	237,383	205,275	26%	52,938	316	152,021
fun.	293,173	212,418	24%	51,477	490	160,451
art.	182,831	265,518	23%	62,267	93	203,158
pro.	275,129	281,999	23%	65,750	787	215,462
biz.	1,416,380	1,340,296	23%	302,394	5,728	1,032,174

EAI Support Rate per Country

◉ Results from 07 Apr 2022 show the following support rates of surveyed addresses per country associated to the IP address

Legend:
- 0 - 4.09 %
- 4.25 - 7.87 %
- 8.07 - 12.65 %
- 13.84 - 19.71 %
- 27.54 + %
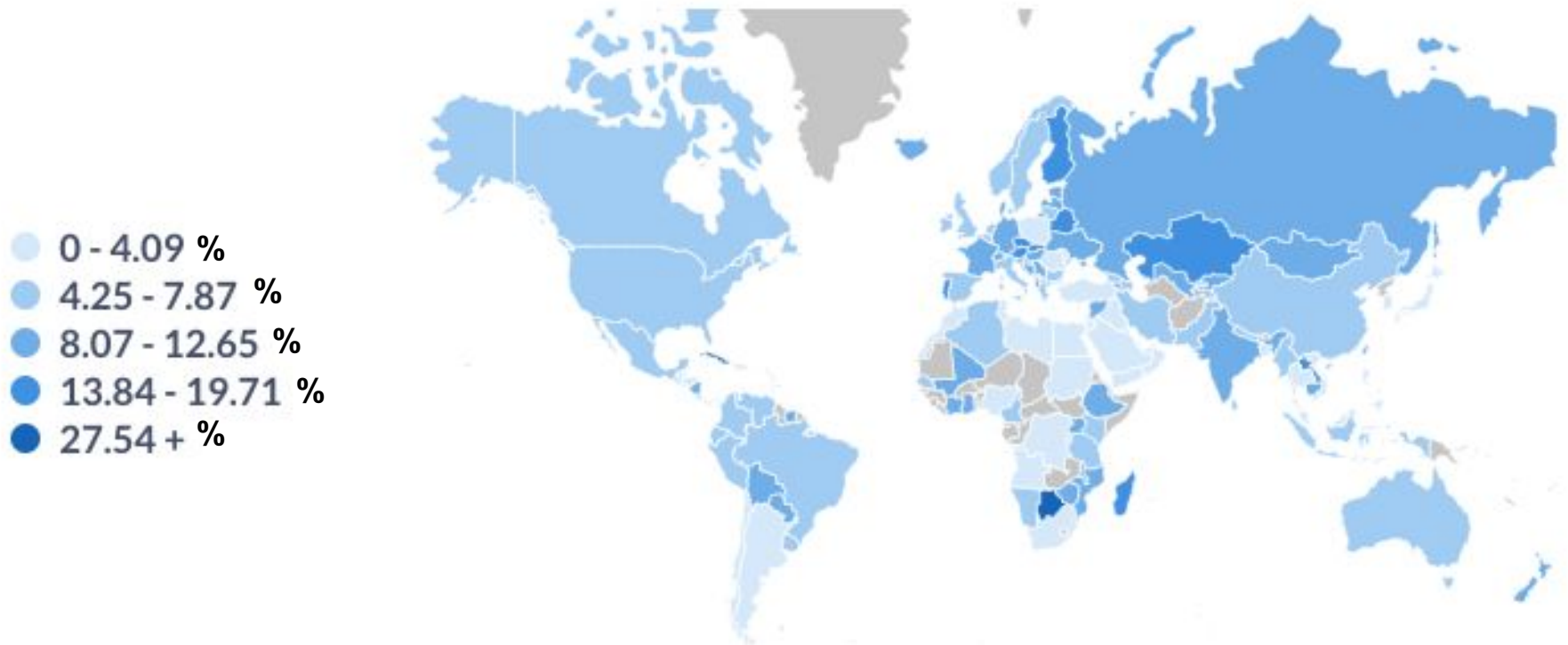
Next Steps (1/2)

- Integration with the "Identifier Technology Health Indicators" (ITHI) metrics https://ithi.research.icann.org/

- Identify reserved IP addresses with specific result codes.

- Support exclude lists for surveyed IP addresses.

- Identifying the mail server software based on the test results.

- Persist if the 8BITMIME header is present or not (new check of EHLO response)

- Send internal notification when execution ends or is interrupted

Next Steps (2/2)

◉ Log the type of connection error when there is no server response

◉ Configure DKIM in testo.tel server

◉ Configure DMARC in testo.tel server

◉ Support SMTP over TLS (port 465) (RFC 7672)

◉ Support for STARTTLS (RFC 3207)

◉ Add logic to retry connections that fail due to transient errors (e.g. ratelimit, concurrent connections, etc.)

Q&A

Thank You and Questions