2022-02-24 Registration Data Accuracy Scoping Team- Meeting #19

The Registration Data Accuracy Scoping Team call will take place on **Thursday, 24 February 2022 at 14:30 UTC for 60 minutes**.

For other places see: https://tinyurl.com/2p98t5fz

ⓘ **PROPOSED AGENDA**

1. Welcome & Chair Updates (5 minutes)
2. Measurement of accuracy (20 minutes)
    a. Review remaining input received on google doc [docs.google.com], see page 25 - How and by whom can it be measured whether current goal(s) of existing accuracy requirements are met?
    b. Confirm next steps
3. Accuracy working definition / construct (30 minutes)
    a. Review input received (see https://docs.google.com/document/d/1JGdNLjOjhmJnj-iDaGvcVnv1gZ5tQu5C/edit [docs.google.com])
    b. Confirm next steps
4. Review of existing data sources (30 minutes)
    a. See compilation developed by Staff Support Team
    b. Team to consider these existing data sources and role in assessing existing current situation & identifying possible gaps
    c. Confirm next steps
5. Confirm action items & next meeting (Thursday 3 March at 14.00 UTC)

BACKGROUND DOCUMENTS

ⓘ **RECORDINGS**

Audio Recording

Zoom Recording

Chat Transcript

GNSO transcripts are located on the GNSO Calendar

✓ **PARTICIPATION**

CRM Attendance

Apologies:Olga Cavalli

⚠ **Notes/ Action Items**

---

1. Welcome & Chair Updates (5 minutes)

- Due to conflict with Policy Update this meeting is scheduled for 60 minutes.
- Thanks to all for doing homework in preparation for this meeting.

2.Measurement of accuracy (20 minutes)

    a. Review remaining input received on google doc [docs.google.com], see page 25 - How and by whom can it be measured whether current goal(s) of existing accuracy requirements are met?

    b. Confirm next steps

- See BC input in the google doc

Comments:

- Input is based on assumptions not facts.
- Problematic to have ICANN org review the data as that would require data transfer to the US which may be legally difficult.
- Focus is on a very small sub-set of registrations – majority of registrations are not abusive. How can that be demonstrated? If there is data available it should be provided.
- Either there is a general problem that needs to be addressed or not. Focusing on a sub-set shouldn't be the objective.
- Transfers to US could be prevented by using ICANN office in Brussels.
- If policy needs to be changed, there is a need to demonstrate that there is a problem, not the other way around.

See ARS Cycle 6 Report (https://whois.icann.org/sites/default/files/files/whois-ars-phase-2-report-cycle-6-15jun18_0.pdf [whois.icann.org])

From the report: We present here the key takeaways from the findings:

Ability to Establish Immediate Contact

- *Ninety-eight percent of records had at least one email or phone number meet all operability requirements of the 2009 RAA, which implies that nearly all records contain information that can be used to establish immediate contact. Only two percent of records had all contact information that met neither email nor or phone operability requirements.*

- 'only' two percent of records still means over 4 million registrations (based on Q3 2021 registration numbers) that do not have contact information that meets email nor or phone operability requirements.
- Does the group consider that an acceptable number? If not, what could be suggested to allow for further consideration to improve this number?

Operability Accuracy

- *Ninety-nine percent of postal addresses, 60 percent of telephone numbers and 92 percent of email addresses met all operability requirements of the 2009 RAA. Fifty-six percent of domains passed all operability tests for all contact types (registrant, administrative and technical) and contact modes (email, telephone and postal address), which is a 7 percent drop from Cycle 5 findings.*
  - *Regional variations of operability accuracy are greatest for telephone, which ranges from 35 percent accurate (Asia Pacific) to 81 percent accurate (North America).*

- *The contact mode with the highest rate of passing all operability tests was postal address, with 99 percent passing all tests. The mode with the lowest rate of passing all operability tests was telephone numbers, with 60 percent passing all tests.*
  - *The majority of email operability errors occurred when an email address bounced (98 percent of errors), compared to the error of a missing email address (2 percent of errors).*
  - *The majority of telephone operability errors were from invalid numbers (58 percent), while most of the remaining errors were from disconnected telephone numbers (40 percent) and other issues preventing connection (2 percent). Less than one percent were missing or not verifiable.*

Syntax Accuracy:

- *More than 92 percent of telephone numbers met all syntax requirements of the 2009 RAA, increasing from Cycle 5 (90 percent).*
  - *Regional variations of syntax accuracy were greatest for postal address, which ranged from 66 percent accurate (Africa) to 97 percent accurate (North America).*
  - *The most common reason for telephone syntax error in most regions was incorrect length, but in North America the most common reason for error was a missing country code.*
  - *For postal addresses, the vast majority of errors in each study have consistently been due to missing fields that were required, such as city, state/province, postal code or street.*

- *In Cycles 5 and 6 the 2009 RAA group had the highest percentage of records in which all three contact modes were accurate. This is a change from Cycle 4, where the 2013 NGF RAA group had the highest percentage of records in which all three contact modes were accurate. Note that the 2009 RAA group contained only 130 records in Cycle 6.*

- Team to consider whether based on this data, the outcomes are 'good enough' or whether there are issues that could/should be addressed. In addition, reviewing this information may help inform what further study or research should be undertaken.
- Team could also consider whether a sub-set of data could / should be reviewed if the team suspects that there may be specific issues with that sub-set such as abusive domain names.
- Team to document the viewpoints on how looking at this issue – some view that there is no issue / problem, while others consider that there is an issue.
- Can we correlate the 2% of the bad actors or just administrative errors? That is a gap in information.
- Is this data comparable to actual use experience? Testing seems to be a proxy for what actual use is.
- Agree that there is no current data and note that there are opposing points of views on whether there is a problem or there isn't and document that as part of the team's findings. Need to stop repeating the argument that because there is no data there is no problem or there is a problem because the team is not going to agree on this.
- Need to look at the relevance of this data in today's world – if data is not publicly available, it is not of use.
- Concerning that 40% of phone numbers were disconnected.
- Numbers don't speak for themselves and are open to interpretation.
- Need to look at the trends at previous ARS.
- Less invasive method of contact have a better result – need to factor that in and need to consider what is behind that. Having phone number publicly posted can be considered invasive.
- This is not about access or display, this is about accuracy. If data is provided, it is accurate. Need to avoid comingling topics.
- Measurement is a snap shot in time – need to also consider how issues were addressed.
- See also https://whois.icann.org/en/whoisars [whois.icann.org] - in 83% a ticket was closed out because the domain was suspended, not because underlying accuracy of data was considered. Trend of higher rate of suspended domain names towards later ARS reports. Is it considered there is 'no issue' because tickets are closed because domain name is suspended (but no review of the underlying data to determine inaccuracy)? Should also consider correlating data.
- We agree that accuracy is difficult to measure. We do not agree on whether it is important to continue such measurement attempts.
- Suspension does not mean the data was incorrect. It only means the registrant did not respond in time or did not provide evidence.
- Consider pulling out of the table specific suggestions that can be further considered and discuss these further by the team (pros and cons). Consider inviting ICANN Compliance to provide input on suggestions such as whether a Registrar Audit on accuracy could be carried out.
- Would it be worth considering sending out a survey to registrars to better understand the steps they take to implement requirements and whether there is data that they would be willing to share? If this is voluntary, how useful will the data be as only 'good' actors may respond?
- Will be limited what can be done until we get a clear confirmation in relation to ICANN org's role as controller or processor. How long before the factual determination takes place?

3. Accuracy working definition / construct (30 minutes)

    a. Review input received (see https://docs.google.com/document/d/1JGdNLjOjhmJnj-iDaGvcVnv1gZ5tQu5C/edit [docs.google.com])
    b. Confirm next steps

Deferred to the next meeting

4. Review of existing data sources (30 minutes)

    a. See compilation developed by Staff Support Team
    b. Team to consider these existing data sources and role in assessing existing current situation & identifying possible gaps
    c. Confirm next steps

Deferred to the next meeting – see also notes above in relation to ARS

5. Confirm action items & next meeting (Thursday 3 March at 14.00 UTC)