

GNSO gTLD Registries Stakeholder Group Statement

Issue: WHOIS Accuracy Pilot Study (APS) Report

Date: 13 March 2015

Ref: <https://www.icann.org/public-comments/whois-ars-pilot-2014-12-23-en>

The Registries Stakeholder Group (RySG) appreciates the opportunity to comment on the [Whois Accuracy Pilot Study Report](#).

Background

ICANN seeks feedback on the approach, methodology and reporting formats documented in the pilot study to help shape the final design of the accuracy reporting system (ARS).

ICANN requests specific community input on:

- Study Design and Methodology;
- Types of Accuracy Reports to be published through the ARS;
- Whether ICANN should conduct Identity Validation in subsequent phases of the ARS Development;
- Whether the methodology should treat registrations under privacy or proxy services differently, and if so, how; and
- Any other aspect of the ARS.

RySG Comment

The RySG provides the following input:

Sample design and methodology

First, the report's authors note that the verification process is relatively slow, and that the timetable for the pilot study was "greatly condensed."

It may be interpreted from this statement that the innately slow nature of Whois data verification and such a condensed timeframe allowed the use of only one sampling method, which, as stated below, is not highly accurate. Accordingly, it's critical that ICANN tests different sampling methods to ensure it eventually employs the one that is most precise.

Second, the report states:

"Different sample sizes result in...uncertainty in the [accuracy] estimates."

This calls into question the methodology of verifying accuracy of an entire Whois entry and demands that ICANN review its approach to eliminate uncertainty.

The report further says:

“The differences in email and telephone syntactical accuracy are not statistically significant for Prior and New gTLDs while syntactical accuracy is better for Prior gTLDs on postal addresses”

This outcome reinforces the need to reassess methodology. Because the syntactic requirements for Whois in the new gTLD Registry Agreement Specification 4 are precise, unlike the Whois RFC, logically the new gTLDs should have a much greater syntactical accuracy than prior gTLDs. The 2013 RAA and the registry agreement for new gTLDs call for compliance with syntactic requirements. However, registrars operating legacy gTLDs under the 2009 RAA may not maintain such precision in Whois syntax because it is not a requirement.

Methodology and resulting data comparing registrars under the 2013 or the 2009 RAA are not meaningful

We do not believe the comparison based on sponsoring registrars’ RAA version is meaningful. The sampled domains that are subject to syntactical requirements and validation requirements represent only a small subset of domains sponsored by registrars under the 2013 RAA. As the report correctly points out:

“although a registrar may have been under the 2013 RAA at the time of sampling, the sampled domain may have been registered when the registrar was under the 2009 RAA. The additional validation and verification requirements of the 2013 RAA are not retroactive to domain names registered through the registrar...”.

In fact, the syntactical requirements became effective when the sponsoring registrar signed the 2013 RAA but the validation requirements only became effective on 1 January 2014. In other words, the date on which a sampled domain name is registered, transferred or has a change in Whois is the differentiator, not the RAA version.

However, the “RAA version population” polled by the pilot study and the resulting data could be taken out of context or could distort the true picture of Whois accuracy. For example, ICANN’s [2014 Annual Report on Whois Improvements](#) claims:

“While the pilot study results are still being confirmed, the preliminary findings revealed that, operationally, Registrars under the 2013 RAA experience better accuracy rates for email addresses than Registrars under prior versions of the RAA..... As more registrars and more gTLD registrations transition to the new requirements, accuracy rates should continue to improve.”¹

Such a claim amounts to a confirmation bias as it focuses only on that part of the findings that may be perceived as “improvements” (i.e., email addresses) and ignores others (i.e., telephone numbers and postal addresses), despite the Pilot Study Report clearly stating:

¹ <http://whois.icann.org/sites/default/files/files/improvements-annual-report-12dec14-en.pdf>, p.8

“Operationally, registrars under the 2013 RAA have more accuracy for email addresses than registrars under the 2009 RAA. Registrars under the 2009 RAA have more operational accuracy in telephone numbers while the two groups are similar on postal addresses. **Further analysis is needed to determine the reason for why registrars under the 2013 RAA experience less operational accuracy on telephone numbers, since the data collected in the Pilot is not sufficient to explore possible reasons.**”² (emphasis added)

Types of Accuracy Reports to be published through the ARS

The RySG is of the view that no accuracy reports should be published through the ARS. The original and most optimal use for ARS is to identify and forward data to registrars for investigation and follow-up, and not for use as a compliance tool. We believe compliance cases are most effectively brought from abuse reports, as originally envisioned.

Confidence in results is unwarranted under current methodology

Further, publication is unwarranted under the current methodology because of low confidence intervals—in other words, confidence in the accuracy of the produced result has not been earned (e.g., 72% accuracy percentage with only a 69% confidence interval—nearly one third of the time, as reported in the study, the result cannot be relied upon). Since the standard error and confidence intervals vary widely across accuracy types and measurements, it would be unfair to all concerned to use the ARS, as currently designed, to report Whois data accuracy.

If reports are to be generated under current methodology, the most meaningful would be a recurring report on syntactical accuracy for “Prior and New gTLDs” (as only domain names in new gTLDs are subject to the stringent syntactical requirements under the 2013 RAA).

Dissimilarities in regions, requirements, vendors contributes to inconsistent results

Another problem with current methodology is that it does not sufficiently take into account discrepancies in:

- Dissimilar syntax for phone numbering and postal address data across multiple regions;
- Requirements by the current new gTLD RA and the 2013 RAA for greater operational accuracy, vs. the RA for legacy gTLDs and the 2009 RAA; and
- Results from different vendors (one each for syntactic, operational telephone and operational email accuracy), each of which were given wide latitude and no guidance from the community on the application of its own technology and methodology to categorizing accuracy. (It’s not unreasonable to conclude those who favor a particular outcome may have an aberrant incentive to interpret “inaccuracy” and repeatedly discover bad data, or vice versa.)

It’s clear, therefore, that such differences, and the resulting lack of precision, make it unfair to post reports that could paint an unjust picture of Whois accuracy. At a minimum, a determination of accuracy should be subject to guidelines provided by the community, including those responsible for operating the Whois system and enforcing its accuracy.

² <http://whois.icann.org/sites/default/files/files/ars-pilot-23dec14-en.pdf>, p.7

Whether ICANN should conduct Identity Validation in subsequent phases of the ARS development

The RySG is not in favor of identity validation. We believe that such a process would conflict with laws in multiple jurisdictions, would require changes to registry-registrar agreements and would violate privacy rights.

Whether the methodology should treat registrations under privacy or proxy services differently, and if so, how

The RySG does not believe the methodology should treat such registrations differently. Such differentiation is unnecessary due to the acceptance of privacy and proxy service use. Should policy be developed by the GNSO regarding the use of these services, it may become appropriate to re-examine methodology.

Any other aspect of the ARS

According to the report, "ICANN's Contractual Compliance Department is in the process of auditing the results of the Pilot Study as part of a compliance pilot to determine whether a compliance follow-up is needed." As a matter of transparency, ICANN should publish the audit methodology and the audit results as soon as possible.

Further, the RySG:

- requests that ICANN conduct a cost-benefit analysis to determine the efficacy of requirements under the 2013 RAA prior to proceeding with introducing any further operational requirements;
- believes the proposed operational email validation methodology should be thoroughly reviewed to ensure checks performed are not creating false positives; and
- urges ICANN to provide the community with an opportunity to review any new methodologies proposed, before they are employed.

Conclusion

Thank you for the opportunity to comment on this important matter.

Minority Statement

The UPU provides the following input:

Whether ICANN should conduct Identity Validation in subsequent phases of the ARS development

The UPU would like to share with you some elements of reflection regarding this statement.

In particular, we would like to suggest to the RySG that identity validation may be considered as a two-step approach combining elements of both automatic and manual processes. The deployment of a validation tool (automatically) checking in real time against “reference data” (e.g. postal addresses) would constitute an easy and cost-savvy solution to implement (on both Registry and Registrar sides), with the aim of providing a basic (yet sufficient for most cases) level of assurance.

In that regards, we are exploring some routines for assistance for faster contact address data entry against UPU reference data (POST*CODE DB and address formats/fields by country).

In case there is a need for increased levels of assurance, and especially in order to address concerns raised by some governments regarding highly-regulated markets and sensitive strings, an “in-person” check is a solution that some Registries may want to envisage to provide increased trust and confidence to Internet users, using for example the global postal network with its 660,000 post offices as a validation agent, or any other suitable validation agent with similarly extensive reach that may exist in countries (so as to ensure equality across all Internet users).

In our view, it would definitely be in the interests of ICANN (including the RySG) to enhance trust and confidence of Internet users during this new gTLD expansion phase.