COMMENTS OF THE INTELLECTUAL PROPERTY CONSTITUENCY ON THE WHOIS ACCURACY REPORTING SYSTEM PILOT STUDY REPORT

MARCH 13, 2015

The Intellectual Property Constituency (IPC) of the GNSO appreciates this opportunity to comment on the WHOIS Accuracy Reporting System ("ARS") Pilot Study Report. See https://www.icann.org/public-comments/WHOIS-ars-pilot-2014-12-23-en.

The IPC acknowledges the efforts made by NORC at the University of Chicago and ICANN in preparing the WHOIS ARS Pilot Study Report (the "Report").

SUMMARY

After reviewing and analyzing the Report, the IPC has identified several issues which need to be addressed by ICANN before completing the design and implementing the ARS. These issues include:

1. The ARS should discount proxy and privacy service records when analyzing the accuracy of WHOIS data. The failure to analyze the effect of proxy services in regard to both accuracy and contactability overstates the accuracy of WHOIS data and undermines the validity of the results set forth in the Report.

2. The use of a sample inappropriately weighted towards the new gTLDs instead of legacy gTLDs, where a majority of domain registrations still occur.

3. The application of an ambiguous category evaluation scale that may overstate the accuracy of WHOIS data.

4. The need to publish the names of all registrars and registries along with their accuracy rankings, and to forward inaccurate registration data to the compliance team for action.

RESPONSE TO ICANN'S QUESTIONS

Sample Design & Methodology

The sample design and methodology for the pilot study are flawed. The accuracy figures presented in the study are overstated because the sample included domain names registered

through proxy and privacy services. Prior studies show that 20-25% of all gTLD registrations use proxy and privacy services.[1] While these registrations are generally syntactically and operationally valid, they are of almost no value in terms of the purpose of WHOIS - promoting contactability[2] - since they do not allow a third party to directly contact the true registrant. The ability to contact the true registrant depends entirely on whether the proxy/privacy service provider will forward ("relay") any message, and on the circumstances in which that provider will disclose ("reveal") the contact information for its customer to a third party. As policies regarding relay and reveal are still completely at the discretion of the proxy/privacy service providers, and vary wildly, no true data about contactability can be extrapolated when these registrations are included in the sample. Registrations granted via proxy/privacy services should not have been included in this study, and should not be included in the ARS, unless such data is segregated from other data.

The Report states that "New gTLDs are a primary focus of the study," but does not say why this choice was made. This focus on new gTLDs significantly skews the results. As noted on page 5 of the Report, over 98% of all gTLD registrations were (and probably still are) in legacy or "prior" gTLDs, yet the sample used in the Pilot Study included 25% new gTLDs. The WHOIS inaccuracy problem flourishes in .com/net/org; but the sample used in the study serves to understate this problem. The IPC would like to know the justification in the WHOIS Review Team report, or in the board's endorsement of the Review Team's recommendations, for drastically overweighting new gTLDs in this study.

A significant majority (currently over 65%) of the new gTLD registrations resolve to parked pages, and likely include a large number of defensive registrations by brand owners. As defensive registrations are more likely to be accurate, but not due to any increased diligence on the part of the Registrar, this further distorts the conclusions of the study. The IPC suggests going forward that the study be revised to account for defensive registrations in order to determine the percentage of accurate WHOIS data.

Types of Accuracy Reports to be published through the ARS

There is no need to have a five category evaluation scale (no failures, minimal failures, limited failures, substantial failure, complete failure) if the majority of instances are either going to comply with the RAA standards, or not comply with the standards.

Further, the five categories of accuracy unnecessarily complicate the findings and create ambiguity. For example, one can easily understand the difference between "No Failure" and "Full Failure" but the substantive differences between the other categories of "Minimal Failure," "Limited Failure," and "Substantial Failure" are difficult to understand. Instead, the study could have used categories based on the field of data missing or inaccurately stated in the WHOIS

---

[1] See, http://gnso.icann.org/en/issues/whois/registrant-identification-summary-06feb13-en.pdf
[2] https://whois.icann.org/en/basics-whois

records, e.g., inaccurate e-mail addresses.   This would allow ICANN, the registries and the registrars to more readily identify the weaknesses in the WHOIS records. Further, the 2013 RAA requires Registrars to validate phone number or email address.[3] If either of these fields is syntactically or operationally inaccurate, this could indicate a complete failure of the Registrar's obligations under the 2013 RAA – and would have been an important finding to flag.

There are a number of other questions about the accuracy rating scale and whether its application tends to overstate the accuracy of WHOIS data. "Recall that a record's field is considered accurate if its rating is limited failure, minimal failure, or no failure" (P. 31). However, these three ratings are then collapsed into one in the "perspective accuracy scoring" process described on the same page. Some examples of records that are considered accurate (at least at the "limited failure" level) include:

(a) Telephone number for a different country than is listed in the postal address (p. 49).

(b) Telephone number with no country code if the country code for the postal address is appended (p. 50).

(In other words, discrepancies between the country address and the phone number only count to make a record accurate, not inaccurate.)

(c) Address that ends in a non-existent country = Limited Failure = "accurate" (p. 60).

Indeed, it is hard to understand any circumstances in which a postal address would be considered "inaccurate," since "No records received a scoring of Substantial Failure (-1) or Full Failure (-2)" (p. 60).   Thus all addresses are "Limited Failure" or above.

Lastly, the identities of the registries and registrars must be publicly identified if the ARS is going to be useful. Going forward, any Accuracy Report published through the ARS must identify and rank all registries and registrars for their WHOIS accuracy, and inaccurate registration data must be forwarded to the compliance team for action.

Whether ICANN should conduct Identity Validation in subsequent phases of ARS Development

In its April 23, 2014 letter,[4] the IPC stated that an "emphasis on the third level of accuracy may reflect an overall bias toward exhaustive empirical research, and away from practical benefits in terms of improvement to the overall accuracy of the WHOIS data, the primary purpose of which is to enable contact with domain name registrants." The purpose of WHOIS is to provide contactability when an issue arises, whether it is technical or legal.

While the IPC continues to prioritize syntactical and operational validity for practical reasons, the IPC believes that ICANN should explore cost-effective methods of conducting

---

[3] https://www.icann.org/resources/pages/approved-with-specs-2013-09-17-en
[4] http://ipconstituency.org/PDFs/IPC_comments_on_draft_Whois_implementation_plan_042314.pdf

Identity Validation. Without Identity Validation there is a clear roadmap for bad actors to provide totally inaccurate but facially operable contact information. All the registrant has to do is choose existing contact information that is not its own. It is fairly easy to find a real street address, a working phone number and a working email address. This would produce a syntactically and operationally "accurate" WHOIS record – yet would be totally without value with respect to contactability. For instance, a cybersquatter could register a domain name with the following operationally and syntactically valid contact information:

> Registrant Name: Barack Obama
> Registrant Street: 1600 Pennsylvania Ave, NW
> Registrant City: Washington
> Registrant State/Province: DC
> Registrant Postal Code: 20500
> Registrant Country: US
> Registrant Phone Number: +1 202.456.1111
> Registrant Email: potus@whitehouse.gov

The IPC acknowledges there are issues with Identity Validation, including the manual labor required as well as privacy concerns, which could make Identity Validation impractical at this time. However, just because there are logistical issues with Identity Validation does not diminish its value. Going forward, the IPC suggests that ICANN conduct Identity Validation tests for a limited category of registrations -- for example, those which are subjects of WHOIS accuracy complaints and which have been shown to be syntactically and operationally valid.

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

As discussed above, because up to a quarter of registrations considered syntactically and operationally "accurate" do not in fact enable contact with the true registrant (and such contact is left to the discretion of privacy/proxy operators), proxy/privacy registrations should be excluded from this and subsequent studies, at least until clear and enforceable rules are in place to guarantee relay and reveal in appropriate circumstances.[5] Otherwise, the percentage of "accurate" registrations in the Report is significantly overstated. If this methodology is used in the ARS, the results of the ARS would also be significantly overstated, and thus essentially invalid.

The ICANN community should continue its work with regard to privacy and proxy service providers, in order to create a set of rules which ensure reliable contactability of registrants by legitimate third parties.[6] This includes requirements for validation of the underlying contact information of privacy/proxy registrations, and for predictable, consistent,

---

[5] See http://la51.icann.org/en/schedule/tue-ipc/transcript-ipc-14oct14-en.pdf at p. 37.
[6] http://gnso.icann.org/en/group-activities/active/ppsa

and balanced minimum standards for disclosure of this information when domain names are used to infringe third party rights or commit other abuses.

Any other aspect of the ARS

Viewing the study as part of the practical effort to implement the recommendations of the WHOIS Task Force, as approved by the Board (rather than as an abstract academic study), the failure to include any component in which identified inaccurate registrations were forwarded to registrars for appropriate corrective action is a significant flaw) in the design or scope of this study. As noted on page 46 of the Report, this forwarding is intended to be "a key function" of the ARS. Instead of doing so, ICANN is "kicking off a Compliance Pilot" that will "audit the results to determine whether a compliance follow-up is needed." It is unclear when this audit will be completed, and the IPC is concerned that it will then be too late to forward the (likely stale) results to registrars for corrective action.

Results show significant accuracy concerns in new gTLDs

Finally, the IPC wishes to note with surprise and disappointment that, to the extent that this study validly quantifies WHOIS accuracy, the new gTLDs apparently have the same accuracy failures as prior gTLDs. This is especially disappointing given that 2013 Registrar Accreditation Agreement has a WHOIS Accuracy Program Specification[7] and that all Registrars dealing in new gTLDs are required to be a party to the 2013 RAA. The same general amount of errors in the new gTLDs indicates that there could be a high number of cybersquatters who are registering domains in the new gTLDs, and at the very least, the way that registrars are implementing the validation and verification requirements of the 2013 RAA is not resulting in increased accuracy/contactability. The ICANN community, along with ICANN compliance, should continue its work with Registrars and others to determine why the terms of the WHOIS Specification are not being met satisfactorily.

CONCLUSION

The IPC recognizes and appreciates the hard work undertaken by ICANN and NORC at the University of Chicago in preparing the Report. However, the Report has shown that there are several fundamental issues which need to be addressed before the ARS can be fully implemented. While the IPC recognizes that this is a pilot study, and there is room for improvement before the ARS is implemented, there are enough flaws already to cause concern. The IPC highly encourages ICANN to address the above issues soon, before the program moves forward with a model that will generate potentially inaccurate and problematic data. The IPC looks forward to seeing the necessary improvements made and brought back to the community for review and comment.

---

[7] www.icann.org/resources/pages/approved-with-specs-2013-09-17-en#whois-accuracy