

## Report of Public Comments

<b>Title:</b>	<b>WHOIS Accuracy Pilot Study Report</b>																
<b>Publication Date:</b>	3 April 2015																
<b>Prepared By:</b>	Margie Milam																
<table border="1"> <tr> <td colspan="2"><b>Comment Period:</b></td> </tr> <tr> <td>Comment Open Date:</td> <td>23 Dec 2014</td> </tr> <tr> <td>Comment Close Date:</td> <td>13 March 2014</td> </tr> </table>		<b>Comment Period:</b>		Comment Open Date:	23 Dec 2014	Comment Close Date:	13 March 2014	<table border="1"> <tr> <td colspan="2" style="text-align: center;"><b>Important Information Links</b></td> </tr> <tr> <td colspan="2" style="text-align: center;">Announcement</td> </tr> <tr> <td colspan="2" style="text-align: center;">Public Comment Box</td> </tr> <tr> <td colspan="2" style="text-align: center;">View Comments Submitted</td> </tr> </table>		<b>Important Information Links</b>		Announcement		Public Comment Box		View Comments Submitted	
<b>Comment Period:</b>																	
Comment Open Date:	23 Dec 2014																
Comment Close Date:	13 March 2014																
<b>Important Information Links</b>																	
Announcement																	
Public Comment Box																	
View Comments Submitted																	
<b>Staff Contact:</b>	Margie Milam	<b>Email:</b>	<a href="mailto:margie.milam@icann.org">margie.milam@icann.org</a>														
<b>Section I: General Overview and Next Steps</b>																	
<p>The Final Report published for public comment describes the results of a pilot study conducted by NORC at the University of Chicago in collaboration with other public and private sector specialists. The Final Report illustrates the findings and methodology to be deployed in the WHOIS Accuracy Reporting System (ARS) under development by ICANN. By publishing it for public comment, ICANN sought feedback on the approach, methodology, and reporting formats used in the Pilot Study. This feedback will help shape the final design of the ARS.</p> <p>When fully developed, the ARS will produce ongoing periodic capable of tracking trends in accuracy rates, and reporting on the specific factors that affect accuracy. This data may be useful to assess the effectiveness of recent efforts at improving accuracy rates, and to support ongoing policy development activities related to WHOIS.</p> <p>The Pilot Study examines accuracy levels by applying syntactic validation and operation validation tests to a Registrant's postal address, email, and telephone numbers listed in a WHOIS record. Although the study did not attempt to apply identity validation techniques, ICANN sought feedback on the feasibility of including identity validation in subsequent development phases of the ARS.</p> <p>The purpose of the Pilot Study was to test assumptions using real data. The methodology can be adjusted based upon learnings of the Pilot and the public comments received on the Pilot Report. Upon review of these comments and feedback, ICANN will update the design of the ARS and begin development in phases. Phase 1 will focus on reporting accuracy rates from a syntactic perspective, Phase 2 on operational perspective, and Phase 3 (if at all), on identity perspective.</p> <p>The development of Phase 1 of the ARS will be conducted by the Global Domain Division. Upon review and analysis of the comments received, ICANN plans to provide additional information to the Community regarding the details of proposed implementation for Phase 1, in advance of the Buenos Aires Meeting.</p> <p>Separately, the Contractual Compliance Department is conducting a pilot to support the WHOIS ARS Pilot. The</p>																	

Compliance Pilot involves reviewing the email records identified as inaccurate by the WHOIS ARS pilot on a manual basis, and sending WHOIS inaccuracy complaints where appropriate to Registrars through the standard processes. The phased pilot began in December 2014 and is still ongoing. The expected completion date of the Compliance pilot is June 2015. After receiving concerns about volume from the Registrars in the Singapore Meeting, ICANN plans to roll-out the ticket creation process for ARS-related inaccuracies in a manner to minimize impact on the system. Feedback from the Compliance pilot and an implementation advisory group (IAG) regarding a follow up process will be used by the GDD Staff to refine the ARS methodology and to inform the development of the follow-up processes with Registrars.

## Section II: Contributors

*At the time this report was prepared, a total of [number] (n) community submissions had been posted to the Forum. The contributors, both individuals and organizations/groups, are listed below in chronological order by posting date with initials noted. To the extent that quotations are used in the foregoing narrative (Section III), such citations will reference the contributor's initials.*

### Organizations and Groups:

Name	Submitted by	Initials
Business Constituency	Steve Del Bianco	SDB/BC
Registrar Stakeholder Group	Michele Neylon	MN/RrSG
Danish Business Authority	Finn Peterson	FP/DBA
United States Government	Ashley Heineman	AH/USG
Intellectual Property Constituency	Greg Shatan	GS/IPC
Non-Commercial Stakeholder Group	Amr Elsadar	AE/NCSG
Registry Stakeholder Group (non-consensus statement)	Paul Diaz	PD/RySG
Universal Postal Union (minority statement to the RySG)		UPU

## Section III: Summary of Comments

### Sample Design & Methodology

Several comments highlighted the need to reduce the rate of false positive and other errors in building out the ARS and to introduce quality assurance checks (RrSG, RySG, BC). Other commentators called on ICANN to identify the error codes that are to be used to determine accuracy levels in the ARS (RrSG, BC).

Other comments recommended that ICANN increase the sample sizes for the future operational checks to ensure that the results produced through the ARS are defensible and statistically significant (USG, IPC, RySg).

Several comments questioned the sampling methodology and specifically, why new gTLDs were oversampled in the methodology (IPC). In response, Staff notes that the ARS is being designed to address additional needs beyond those identified by the WHOIS Review Team, such as to respond to the Board's commitments to the GAC regarding including safeguards for the New gTLD Program.

One commentator recommended that the ARS address the accuracy of all contact addresses (admin, tech), not just the registrant's address (BC).

### Types of Accuracy Reports to be published through the ARS

Several comments recommended that ICANN focus on designing reports that align more closely to the contractual requirements under the RAA to conduct validation/verification on certain contact fields (IPC,

RrSG).

One comment suggested that ICANN report on registrar responsiveness, records corrected, and the tracking of challenges to assuring WHOIS accuracy (BC).

The RySG recommended that ICANN publish no reports through the ARS, but instead build the ARS to be a system that simply checks and forwards inaccuracy complaints to registrars.

ICANN agrees the validation criteria for syntax and operations should closely align to the contractual obligations of the registrar, via the applicable Registrar Accreditation Agreement (RAA), and will work to improve the validation criteria in this manner for the next iteration of the ARS. Based on the direction from the ICANN Board in response to the WHOIS Review Team recommendations and various GAC advice, the ARS will be utilized to produce publicly available reports regarding the accuracy of the WHOIS records in the gTLD namespace.

#### **Whether ICANN Should Conduct Identity Validation in subsequent phases of the ARS Development**

The commentators were mixed on the question of whether the Accuracy Reporting System should be built to include identity related checks for the WHOIS records included in the study sample. Of the responses received, five (5) groups were supportive of exploring identity validation, while three (3) were opposed.

The question of whether to build the ARS to attempt to validate (whether passively through 3<sup>rd</sup> party sources or directly via contact with the registrant) information was widely misunderstood by a number of commentators. Although the Pilot Study Report did not propose validating every record in WHOIS, several commentators erroneously assumed that the methodology intended to do so, or that ICANN would be asking registrars to do so. Instead, ICANN is designing the ARS to validate only those records in the sample size used in the study, and proposes that the validation be done (if at all) through the vendors selected by ICANN.

Several commentators strongly objected to conducting identity validation (NCSG, RrSG, RySG) at all through the ARS. It is important to understand that this issue has been raised by the GAC and the SSAC meriting further exploration. To date, ICANN has not committed to including validation checks in the ARS. Instead, ICANN is exploring whether it is feasible or advisable to do so.

The NCSG alleges that ICANN has buried important policy issues in an ICANN Contractor's implementation report. To the contrary, the identity validation issue was highlighted prominently in the announcement for this Public Comment Forum as a specific question singled out for comment. The issue was also highlighted in the Next Steps Document posted for public comment along with the Contractors Report. This issue was further emphasized in several documents and public comment related to the ARS throughout the past year.<sup>1</sup> Numerous sessions at ICANN meetings have explored this issue, including at the All Things WHOIS Sessions held in [Los Angeles](#) and in [Singapore](#).

<sup>1</sup> See: [Public Comment Announcement](#) for the Draft Implementation Plan (11 March 2014), All Things WHOIS – [Los Angeles Meeting](#), All Things WHOIS- [Singapore](#).

In response to the mixed reactions received on the issue of identity validation, ICANN will consider utilizing an implementation advisory group or other mechanism to further the community dialogue on this topic and to explore the advisability and feasibility of including an element of identity validation in future phases of the ARS development.

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

The comments were mixed on the question of whether privacy/proxy services should be treated differently in the methodology for assigning an accuracy score. The RySG noted that it is premature to make a distinction currently, when there are no policies adopted to distinguish these registrations from the other registrations that do not use privacy or proxy services.

After examining the mixed responses, Staff suggests that it is premature to treat registrations covered by privacy/proxy services differently than other registrations, as the GNSO currently has an active PDP underway to develop policies to support the creation of a privacy/proxy services accreditation program. Staff recommends that the PDP also consider examining and proposing additional policies to address whether privacy/proxy registrations should be treated differently than other registrations in the ARS, and if so, how.

**Any other aspect of the ARS**

One commentator (DBA) noted the importance of ensuring that all stakeholders, including governments, are effectively involved in this important work as many of the issues under discussion and analysis have public policy dimensions, including privacy, law enforcement, consumer protection and public safety. Another commentator (USG) agreed and noted that the maintenance, verification, and validation of WHOIS information are an important conduit to public safety, dispute resolution, abuse reporting, and other applications of the WHOIS system.

The NCSG believes that the Pilot Report misrepresents the WHOIS Policy Review Team Final Report and its Recommendations. In its view, the goal of the Whois Review Team was “Contactability” and “Reachability” of the Registrant. To this end WHOIS Policy Review Team Final Report looked “holistically” at the WHOIS record and did not seek the accuracy of each and every element of a Registrant’s Whois record. In response, Staff notes that the ARS is being developed to satisfy additional needs beyond those addressed by the WHOIS Review Team, including Board responses to GAC Communiqué’s related to safeguards for the New gTLD Program, beginning with the Beijing 2013 Communiqué, and the recommendations of the SSAC in [SSAC 58](#).

**Section IV: Analysis of Comments**

*General Disclaimer: This section is intended to provide an analysis and evaluation of the comments*

*received along with explanations regarding the basis for any recommendations provided within the analysis.*

### **Support for/against Including an Identity Validation Component in the ARS**

Several commentators (USG, DBA, IPC, BC, UPU) supported the position that the ARS should include an assessment of whether the identity of the registrant could be confirmed as part of the accuracy rating, while others opposed it (RRSG, RYSG, NCSG). For example, under the Danish Domain Act, which regulates the Danish ccTLD, the .dk WHOIS database must be accurate and in order to achieve this, the administrator of .dk must perform an identity check on the registrant before activating a domain name. *FB on behalf of the Danish Business Authority*. Ultimately, identity validation should be included in the ARS. *SDB, on behalf of the Business Constituency*. It is the view of the USG that evaluating WHOIS data from the identity perspective should be considered for future phases of the ARS development. *AH, on behalf of the USG*. 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 Identity Validation. *GS, on behalf of the IPC*.

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. *GS, on behalf of the IPC*.

Several commentators objected strongly to identity validation (NCSG, RrSG, RySG). The Registrar Stakeholder Group believes that ICANN should not conduct Identity Validation in subsequent phases of the ARS Development. Any attempts at "identity validation" would put an unreasonable burden on registrars and registrants and there is no policy or contractual obligation that obliges it. For ICANN Staff to suggest that such actions are desirable would infer that ICANN the corporation is supportive of such a change. This would not only send the wrong message to the broader internet community but would also be a clear example of ICANN Staff circumventing the multi-level stakeholder process by effectively mandating a top-down change on Whois policy which would impact millions of domain name registrants. *ML, on behalf of the Registrar Stakeholder Group*.

The NCSG alleges that ICANN had buried policy issues in an ICANN Contractor's implementation report. The NCSG notes that this follows pages of study "methods and approach" language and sample design which are obscure even to those who follow Whois policy issues on a regular basis. After the many years of heated controversy over this topic, the NCSG believes it is disingenuous at the very least to allow this policy debate to continue its development in this obscure manner. *AE, on behalf of the NCSG*.

### **Suggestions for Implementing Identity Validation Checks**

The UPU suggests 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. *UPU Minority Statement to the RYSG Comments.*

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). *UPU Minority Statement to the RYSG Comments.*

### **Rationale For/Against Identity Validation Assessments**

Performing the Syntactical evaluation and the Operational Evaluation without the Identity evaluation is leaving out a major component as to the relevancy of the data. This study does not look at fraudulent use of a third party’s information. The data if examined would appear to be accurate as it is authentic and even validated contact information, but it is not information that would allow you to contact the individual that registered the domain name. This contravenes the objective of WHOIS. The BC urges the study of including identity confirmation. Fraudulent use of contact data or alleged identity theft should be treated differently when reported through the WHOIS ARS. It does not make sense to ask a registrant who has intentionally provided contact information fraudulently, if the information is correct. *SDB, on behalf of the Business Constituency.*

The USG agreed with the DBA’s recognition of the importance of Identity Validation as necessary to confirm that WHOIS information corresponds to the registrant. Like the GNSO Business Constituency, the USG believes that in the course of moving forward with Identity Validation, "the task and costs of acquiring Identity Validation should be quantified and shared" so that we can arrive at best practices for implementation. *AH, on behalf of the USG.*

The maintenance, verification, and validation of WHOIS information are an important conduit to public safety, dispute resolution, abuse reporting, and other applications of the WHOIS system. The verification of WHOIS information is critical to mitigating the harms created by registrants seeking to conduct nefarious activities, and can aid in vital communication with online service operators. *AH, on behalf of the USG.*

Performing such action [identity validation] would violate the laws of multiple countries. It is the RrSG position that any enforced verification of the identity of internet users carries very concrete risks for the privacy and security of the associated individuals. *ML, on behalf of the Registrar Stakeholder Group.*

Identity validation of those engaged in freedom of expression, publishing and political discussion is a deeply controversial prospect – and one with heartfelt objection and opposition grounded in history and law. *AE, on behalf of the NCSG.*

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. *GS, and behalf of the IPC.*

The RySG believes that such a process would conflict with laws in multiple jurisdictions, would require changes to registry-registrar agreements and would violate privacy rights. *PD, on behalf of the RYSG.*

In the UPU's view, it would definitely be in the interests of ICANN to enhance trust and confidence of Internet users during this new gTLD expansion phase. *UPU Minority Statement to the RYSG Comments.*

### **Identity Validation Raises Policy Issues**

The NCSG notes that at no time has ICANN ever held a Policy Development Processes on Identity Validation. Accordingly, it asks- "where does this guidance from ICANN to its Contractor to explore identity validation implementation come from? For those who attended the public Whois meeting in LA, this issue certainly was not flagged in the discussion; for those who attended the public meeting in Singapore, this issue was introduced and IMMEDIATELY FLAGGED as intensely controversial and divisive." *AE, on behalf of the NCSG.* However, contrary to the NCSG's assertions, the issue was raised in Los Angeles in the discussion related to the Pilot Report, as highlighted in the PowerPoint presentation and transcript for that session.

The NCSG notes that the GAC has not issued policy in this area. According to the "Brief Overview" provided by ICANN as introduction to this Contractor Report and this public comment period, the GAC "asked for an assessment of the feasibility, costs and benefits of conducting identity validation as part of the development of the ARS." *AE, on behalf of the NCSG.*

As always, policy development must proceed implementation. The NCSG calls on ICANN to take this discussion out of the recesses of a Contractor report, and into the light of the policy development process. *AE, on behalf of the NCSG.*

### **Treatment of Privacy/Proxy Registrations in the ARS**

The survey results do not contain information related to the total number of domains that used privacy services vs. unique registrant data. One may argue the domains with higher concentration of privacy may skew the report results since the WHOIS output for private registration services is typically a Registrar specific template. The "geographic region" aspect of the report does not appear to attempt any correlation between the region of the registrant and the region of the registrar. *ML, on behalf of the Registrar Stakeholder Group.*

The email address methodology should be different for privacy or proxy services, since previous studies have shown that these registrations comprise at least 25% of the domain name registrations – and growing. The work of the PPSAI has shown that many of the email addresses provided by a privacy or proxy service are technically operational, but do not serve the purpose of WHOIS because emails may not be delivered to the actual registrant and operator of a domain name. Many privacy and proxy vendors allow their customer to choose not to receive any email sent to the proxy privacy specified email address for their domain name. The email address may be shown to be operational but not relevant. To avoid overstating the quality of the WHOIS data for meeting the core objective of WHOIS, these emails should be treated separately in any accuracy study. *SDB, on behalf of the Business Constituency.*

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. *GH, and behalf of the IPC.*

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. *GH, and behalf of the IPC.*

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 reexamine methodology. *PD, on behalf of the RySG.*

### **Support for Quantifying the Tasks & Costs of Identity Validation**

The BC believes that the task and costs of acquiring Identity Validation should be quantified and shared, and that the level of assurance should be quantified and shared, so that the community is able to make a qualified judgment about moving forward with it. *SDB, on behalf of the Business Constituency.*

### **WHOIS Implicates Public Policy Considerations of Importance to Governments**

It is of the greatest importance to ensure that all stakeholders, including governments, are effectively involved in this important work as many of the issues under discussion and analysis have public policy dimensions, including privacy, law enforcement, consumer protection and public safety. *FB on behalf of the Danish Business Authority.*

If the GAC deems the verification of identity of domain owners to be desirable, verification of identities should first be incorporated into the legal structures within their respective countries. While this may already be the case for some ccTLDs such as China or Russia, most countries rightly believe such measures are undesirable. It should also be noted that not all countries have GAC representation. *ML, on behalf of the Registrar Stakeholder Group.*

### **Contact Methods Examined by the ARS Design**

The design appropriately accounts for the most important contact methods in domain registration records. *SDB, on behalf of the Business Constituency.*

### **Email Related Checks**

The report does not include details about how recipient validity is detected for domains that do not give a definitive response through an SMTP conversation (e.g., Yahoo, Hotmail, AOL, etc.). *SDB, on behalf of the Business Constituency.*

ICANN should list safeguards included in the policy, setup, testing, and reporting. For example, with testing, larger validation data sets increase the possibility of validators being blocked at sites like gmail.com, yahoo.com, and MSN, among others. *SDB, on behalf of the Business Constituency.*

### Postal Address Related Checks

The report conflates “ease of validation” with “validity”. It is significantly easier to validate physical addresses in countries such as the USA where the postcode system is well established. The report confuses this ease of automated validation with actual accuracy. *ML, on behalf of the Registrar Stakeholder Group*

For postal address validation, the difference between success and failure should be determined only by deliverability of mail to the address. As long as the postman can find it, it is a valid address. However, the opposite need not be true. Valid addresses can result in mail being returned due to failures of the delivering entities. The report rightly states that quality of databases varies from country to country. Therefore, if the validation process produces a result of failed verification the result should be verified by sending actual mail to the address and examining the returns due to inability to deliver. *ML, on behalf of the Registrar Stakeholder Group.*

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. *ML, on behalf of the Registrar Stakeholder Group.*

### Suggested Improvements to the ARS Design

The study did not contain the detailed requirements (referencing either the Registrar Agreement, and/or ICANN policy) that determined the framework for the Accuracy Result Study. ARS design should only determine whether a record matches the minimum requirements of the registration agreement and ICANN policies. For example "catch all" email addresses should be considered as accurate as they provide a functional means for email communications. *ML, on behalf of the Registrar Stakeholder Group.*

The Business Constituency notes that it would be helpful for FAX, and its elimination from the validation set, to be addressed. *SDB, on behalf of the Business Constituency.*

The ARS should address accuracy of all contact addresses, not just the registrant's address, and suggests that they should be treated the same, and asks that ICANN provide detail on treatment of Admin, Billing, Tech, and Registrant contacts, if they are treated in different ways. *SDB, on behalf of the Business Constituency.*

Presently, an email address "Full Failure (-2)" may result from other conditions not mentioned in the report, such as a server temporarily down or server timeouts. *SDB, on behalf of the Business Constituency.*

If email is intended to be used in the future as part of identity validation methods, a defined policy and method as it relates to WHOIS validation will be helpful in guarding against false negatives. For example, these methods might be useful toward guaranteeing the most accurate results: testing via DMARC (Domain-based Message Authentication, Reporting & Conformance); testing via VBR (Vouch by Reference); use of IANA Email Authentication Parameters and SPF and DKIM records when sending; accounting for milers/MTA's, receivers and relays using iprev. *SDB, on behalf of the Business Constituency.*

The application of an ambiguous category evaluation scale that may overstate the accuracy of WHOIS data. *GS, on behalf of the IPC.*

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. *GS, on behalf of the IPC.*

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 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. *GS, on behalf of the IPC.*

The report’s authors note that the verification process is relatively slow, and that the timetable for the pilot study was “greatly condensed.” Accordingly, it’s critical that ICANN tests different sampling methods to ensure it eventually employs the one that is most precise. *PD, on behalf of the RySG.*

Another problem with current methodology is that it does not sufficiently take into account discrepancies in: Results from different vendors, 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. *PD, on behalf of the RYSG.*

### **Sample Sizes to be used in the ARS methodology**

The USG finds measuring operational verification of WHOIS as a key component to ensuring accuracy. As such, Phase 2 of the ARS needs to include a more appropriate and robust sampling size (greater than 1,000) for operational validation of telephone and email to provide a true representation of the market. *AH, on behalf of the USG.*

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 2 through proxy and privacy services. *GS, on behalf of the IPC.*

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. *PD, on behalf of the RySG.*

### **Information on Resources Needed to Build the ARS**

Understanding resources anticipated to be required for ARS will be helpful in setting timeline expectations. *SDB, on behalf of the Business Constituency.*

### **Focus on New gTLDs vs. Legacy gTLDs**

It is not clear why the study focuses on new gTLDs to the exclusion of legacy gTLDs (.com; .net; .org) where inaccurate data is known to be an issue. *SDB, on behalf of the Business Constituency.*

The use of a sample inappropriately weighted towards the new gTLDs instead of legacy gTLDs, where a majority of domain registrations still occur. *GS, on behalf of the IPC.*

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. *GS, on behalf of the IPC.*

### **Importance of Process to Forward Inaccurate Records**

The core objective of this effort is to improve the WHOIS data so that responsible parties can be contacted. Therefore, when inaccuracies are found, even in this pilot study phase, a process should be established to forward those inaccuracies to registrars for appropriate corrective action. *SDB, on behalf of the Business Constituency.*

### **Suggestion for a Pilot focused on Identity Validation**

Because the pilot did not include identity confirmation, it may be useful to organize a pilot for just this scope, if it is determined that this will be included in the scope of the ARS. *SDB, on behalf of the Business Constituency.*

### **Additional Reports Suggested for the ARS**

The ARS should include a scorecard of registrar responsiveness, records corrected, and tracking of challenges to assuring WHOIS accuracy. *SDB, on behalf of the Business Constituency.*

According to the report, a record "received a scoring of "Full Failure (-2)" if the username or mailbox was rejected as a valid email address, the email was invalid according to RFC specifications, and the email was identified as a spam trap." In the future, having a breakdown of the occurrence of these individual conditions may produce useful information. *SDB, on behalf of the Business Constituency.*

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. *GS, on behalf of the IPC.*

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. *GS, on behalf of the IPC.*

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. *GS, on behalf of the IPC.*

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). *GS, on behalf of the IPC.*

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. *PD, on behalf of the RySG.*

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). *PD, on behalf of the RySG.*

#### **Suggestion to Rename to "Domain Name WHOIS Accuracy"**

Because WHOIS pertains to domain names, IP addresses, and autonomous system numbers, a more appropriate formal name for WHOIS Accuracy should include the words "domain name". *SDB, on behalf of the Business Constituency.*

#### **Process for Documenting Email Rejections**

A documented process should exist for processing rejected or bounced mail from servers running SPF with HELO (or EHLO) identity checks, and rejected or bounced mail from servers running SenderID or other controls in the receiver's Administrative Management Domain (ADMD). *SDB, on behalf of the Business Constituency.*

#### **Eliminating False Positives/Quality Assurance Checks**

The verification results do not contain any details of the steps taken to reduce or eliminate false positives in the reported inaccuracies. It is unclear as to what (if any) verification steps were taken by the reporting contractors to ensure findings were accurate and false positives were identified and removed from the study results. The absence of these proactive measures may have led to inaccurate results and statistics. *ML, on behalf of the Registrar Stakeholder Group.*

The email analysis conducted by Strikelron has produced results already known to be partially false. It has been determined by review of Whois inaccuracy reports sent by ICANN compliance to the sponsoring registrar that a fully functional, accurate and previously verified sample email address was reported as being inaccurate by the study. Since the analysis methodology of the study was automated, it must be noted that manual validation of any "inaccurate" results would be advisable. *ML, on behalf of the Registrar Stakeholder Group*

Q&A check with multiple requests using known invalid addresses. *SDB, on behalf of the Business Constituency.*

Each allegedly inaccurate set of domain registration details contained within the report should include documented details of the associated key findings including determining factors that led to such a conclusion. By providing the actual data for each domain, participating Registrars would be able to reach a better understanding of the published statistical summary. If a domain is deemed as being inaccurate, then additional measures such as cross-referencing could occur with alternative verification methods. This would allow for improved quality of the data results prior to referencing for the purposes of statistical analysis. *ML, on behalf of the Registrar Stakeholder Group.*

The RysG believes the proposed operational email validation methodology should be thoroughly reviewed to ensure checks performed are not creating false positives. *PD, on behalf of the Registry Stakeholder Group.*

### **Role of ICANN Contractual Compliance Dept.**

Whois Inaccuracy Reports issued by ICANN compliance based on ARS reports contain a high rate of false positives results. It has been noted that the experience of one particular Registrar has concluded a very high error rate in the data output, i.e. reporting accurate data as inaccurate. Based on this information, the quality of the reported results is questionable. Therefore, unless the accuracy rates of the data contained within the report increases, ICANN compliance is not the appropriate venue for WHOIS records that have been categorized as inaccurate. All reports should be manually verified before compliance-related actions are required. *ML, on behalf of the Registrar Stakeholder Group.*

ICANN should refrain from using ICANN compliance (and thereby registrars) as a validation tool to check whether a potential inaccuracy is an actual inaccuracy. Performing inaccuracy verification under the Whois inaccuracy reporting system is an onerous task for both registrars and compliance that must be undertaken manually, thereby needlessly incurring substantial costs and work-load. *ML, on behalf of the Registrar Stakeholder Group.*

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. *GS, on behalf of the IPC.*

### **Analysis of the Study Results**

The IPC is surprised and disappointed that 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<sup>7</sup> 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. *GS, on behalf of the IPC.*

The report 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. *PD, on behalf of the RYSG.*

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. *PD, on behalf of the RYSG.*

It’s clear that the differences of approaches among vendors and the resulting lack of precision, make it unfair to post reports that could paint an unjust picture of Whois accuracy. *PD, on behalf of the RYSG.*

### **Other Observations**

The NCSG is deeply concerned that this Report or this Proceeding appears to be – a process to seek permission from the ICANN Community for the: a) wholesale checking of the physical addresses of online speakers across the world (whether using domain names for political speech, personal speech, or religious, ethnic or sexual minority expression) thus creating an unprecedented inextricable link between a speaker and her physical location, and b) the radical new concept of Identity Validation for each and every domain name Registrant to the ICANN Community, a concept with inconceivable implications for political, ethnic and religious minorities worldwide, as well as entrepreneurs, emerging organizations and those operating today without identities who seek to create them. *AE, on behalf of the NCSG.*

### **The Report Misrepresents the WHOIS Review Team Final Report**

The Current NORC Study (2014) and its accompanying ICANN Staff Summary accompanying this NORC’s Pilot Report misrepresent the WHOIS Policy Review Team Final Report and its Recommendations. The goal of the Whois Review Team was “Contactability” and “Reachability” of the Registrant. To this end WHOIS Policy Review Team Final Report looked “holistically” at the Whois record and did not seek the accuracy of each and every element of a Registrant’s Whois record. *AE, on behalf of the NCSG.*

The Whois Review Team called for ICANN to significantly reduce the number of “Full Failure” and “Substantial Failure” Whois Records --- Avoidance of “No Failure” was not a goal at all. *AE, on behalf of the NCSG.*

**Next Steps**

ICANN may want to consider conducting two independent analyses of the report findings based on multiple data sources in order to compare the quality of the results and identifying validation gaps. The RrSG notes that a high error rate within the determinations of the study would render any conclusion drawn from these determinations useless, therefore all care should be taken to avoid incorrect determinations. *ML, on behalf of the Registrar Stakeholder Group.*

ICANN staff should consult with the RrSG prior to launching the next phase to ensure the study methodology meets real-world requirements and expectations. *ML, on behalf of the Registrar Stakeholder Group.*

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. *GS, on behalf of the IPC.*

The RySG urges ICANN to provide the community with an opportunity to review any new methodologies proposed, before they are employed. *PD, on behalf of the RYSG.*

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. *PD, on behalf of the RYSG.*

The NCSG suggests that ICANN add Policy Staff and freedom of expression and data Protection expertise to work on this project. The NCSG asks that an ICANN Staff deeply steeped in data protection and freedom of expression laws and rights be brought on to work on the development of these address and identity issues. *AE, on behalf of the NCSG.*

One thing the Whois Review Team did note in its Final Review is the need for clear and concerted outreach on issues that impact the WHOIS. That has clearly not happened here – when so much of substance is buried so deeply in the back of a report. When will ICANN be undertaking clear, robust global Outreach on these important freedom of expression and privacy issues and implications? *AE, on behalf of the NCSG.*