Draft Implementation Plan for the WHOIS Online Accuracy and Reporting System

Please note that this is a discussion draft only. GTLD Registrars and Registries should not rely on any of the proposed details of the contained information until after a Final Implementation Plan is published.

Contents

WHOIS Online Accuracy Reporting System Implementation Plan 3

Overview 3
WHOIS Review Team Recommendations 3
NORC's Design of Experiment 2
External Systems Dependency 2
Interaction with Registrars 3
Timeline and Next Steps 3
Additional Information 3

ANNEX A 4

NORC DESIGN OF EXPERIMENT 4

WHOIS Online Accuracy Reporting System Implementation Plan

Overview

On 8 November 2012, the ICANN Board approved a series of improvements to the manner in which ICANN carries out its responsibilities for WHOIS (the current gTLD registration data directory service), in response to the recommendations of the WHOIS Review Team convened under the Affirmation of Commitments (AoC).

The Board's mandate in response to the recommendations of the WHOIS Review Team calls for ICANN to execute its Action Plan for improving WHOIS. In the Action Plan, ICANN committed to:

1) proactively identify potentially inaccurate gTLD data registration information in gTLD registry and registrar services, explore using automated tools, and forward potentially inaccurate records to gTLD registrars for action; and

2) publicly report on the resulting actions to encourage improved accuracy.

To satisfy these requirements, ICANN has been developing a WHOIS Online Accuracy and Reporting System (the System) that will produce reports to be published on the WHOIS website on a periodic basis.

This Implementation Plan describes ICANN's proposed design for determining how to identify a WHOIS record as "inaccurate" for use in the WHOIS Online Accuracy Reporting System. Feedback on this WHOIS Implementation Plan may be provided through the public comment forum open until 22 April 2014.

WHOIS Review Team Recommendations

Although the WHOIS Review Team recommended that:

> "ICANN should take appropriate measures to reduce the number of WHOIS registrations that fall into the accuracy groups "Substantial Failure and Full Failure" (as defined by the NORC Data Accuracy Study, 2009/10)…"

much has changed in the WHOIS environment since 2009, making the original NORC definitions less meaningful as a benchmark of WHOIS accuracy and more difficult to implement through an automated process.

In designing a new WHOIS Online Accuracy Reporting System, it is essential to take into account recent developments in the evolution of WHOIS, such as the newly instituted obligations under the 2013 Registrar Accreditation Agreement, the launch of new gTLDs, and the work of the SSAC examining domain name registration data validation issues as reported in SSAC-58.

NORC's Design of Experiment

ICANN turned to NORC at the University of Chicago, the pioneer of one of the earliest studies into WHOIS accuracy, to assist ICANN in proposing updated definitions of WHOIS accuracy for the WHOIS Online Accuracy Reporting System and to develop a model for implementing the statistical analysis and reporting features.

Taking into account these recent developments, the model examines a WHOIS record for each of the validation perspectives highlighted in SSAC-58, namely, syntactic, operational, and identity validation, and assigns a scoring methodology. These in turn translate into a finding of accuracy labels, based on range of possibilities, such as: No Failure, Minimal Failure, Limited Failure, Substantial Failure, and Full Failure, which will be reported on a regular basis, and tracked over time. The model also describes a sampling methodology to ensure that a statistically significant number of records are examined to provide adequate geographic scope. The model also enables a comparison of accuracy levels associated with new gTLDs and legacy gTLDs, as well as registrars under the 2013 RAA versus the older versions of the RAA.

The NORC proposal (**Annex A**) is published along with this Implementation Plan to solicit feedback from the ICANN community on the proposed methodology suggested by NORC.

External Systems Dependency

Delivery of the WHOIS Online Accuracy and Reporting System is dependent upon several key factors.

One gating element is ICANN's ability to conduct searches to produce the sample sets necessary to deliver statistically significant results for the reporting module. With launch of the WHOIS Online Search Tool in early 2014, ICANN will have the necessary functionality to perform these searches.

Development of the definitions and model for examining WHOIS accuracy is also a foundational requirement for the delivery the WHOIS Accuracy and Reporting System. Once the definitions are adopted and the statistical model finalized, ICANN plans to continue developing the platform to deliver the reporting module through the WHOIS Website. ICANN plans to utilize existing technology offered by third party service providers for key aspects of the analysis needed to develop the reports. For example, one or more vendors could provide postal address validation services, WHOIS record parser technology, telephone numbers validation, statistical analysis and reporting services, data visualization and web development, etc.

ICANN plans to release a request for proposal (RFP) after the ICANN Singapore 2014 Meeting to implement the proposed model for determining WHOIS accuracy, as updated following consultation with the Community.

Interaction with Registrars

A key requirement for the WHOIS Online Accuracy and Reporting System is the ability to "*forward potentially inaccurate records to gTLD registrars for action*." As this System is being built, it will be important to work with registrars to determine the best path for delivering these records for follow-up action.

The operational aspects of this functionality may be different than the current WHOIS inaccuracy complaint reporting system. ICANN proposes initiating a dialogue with the Registrar Stakeholder Group to obtain input from registrars on how to create an efficient process for implementing this feature of the WHOIS Online Accuracy Reporting System. Also to be explored is the impact that this new reporting will system will have on ICANN's Contractual Compliance operations.

Timeline and Next Steps

ICANN plans to update the proposed model as appropriate in response to broad based Community feedback in early April. This updated model will be incorporated in an RFP to be released after the ICANN Singapore 2014 Meeting. Once the responses have been evaluated, ICANN will publish an updated implementation plan to keep the public informed of the latest developments and timeline for the WHOIS Online Accuracy and Reporting System.

ICANN also plans to work with Registrars on defining the process for forwarding potentially inaccurate WHOIS records to registrars for action.

Based on these activities, ICANN projects that the first accuracy reports should be available to the public shortly after the ICANN Los Angeles 2014 Meeting.

These dates are subject to change depending on competing priorities and the associated complexities of integrating multiple vendors, which is required in order to complete this work effectively.

Additional Information

More information on the status of the other activities underway to improve WHOIS is available here.

ANNEX A

NORC DESIGN OF EXPERIMENT

Designing Future WHOIS Accuracy Studies

NORC Version 1.1

SECTION 1: OVERVIEW

This report discusses requirements and recommended approaches for future WHOIS accuracy studies. Section 2 of this report summarizes requirements and goals, as defined by various ICANN stakeholders, and presents guidance for conducting future studies.

Expanding on the guidance in the requirements section, Section 3 explains in detail the three accuracy "perspectives" defined by SSAC-58[1]: Syntactic, Operational, and Identity. Accuracy rating scale categories are provided, and discussed in context of the three accuracy perspectives. Examples are provided of how the perspectives and rating scale categories are expected to be incorporated into a framework for assessing the accuracy of WHOIS data.

Next, the report discusses the need for a model which can be used to perform consistent and efficient assessments of accuracy based on the three perspectives, and use of accuracy rating scale categories. An example of a model which contains a complete set of parameters for determining accuracy will be provided and explained in detail. A sample WHOIS record is then examined in order to demonstrate how the model can be used to assess the accuracy of a WHOIS record. The process systematically assesses data using the three aforementioned accuracy perspectives, assigns accuracy rating scale categories to the data, and then applies a quantitative model to assess the accuracy of the sample WHOIS record. An index of use-cases which demonstrate how categories of accuracy may be applied to WHOIS records, is provided in Appendix A of this report.

The report concludes with recommended approaches for sample design. These approaches account for the parameters of WHOIS data extracts, and are recommended on the basis that they can be used in conjunction with the described assessment process to in a way which conforms to the requirements stated in Section 2.

SECTION 2: REQUIREMENTS

This section provides a narrative summary of the issues and recommendations outlined in the WHOIS Accuracy Requirements document. While the issues discussed below are interrelated, they have been separated by stakeholder group in order to demonstrate the diversity of stakeholder input which has been aggregated to create this report.

2.1. WHOIS Review Team. Recommendations from the WHOIS Review Team[2] focus on establishing a framework which can provide consistent assessments of accuracy, so that potential changes in accuracy might be observed over time. These recommendations stem from an underlying goal of reducing over time, the percentage of WHOIS registrations that can be categorized as "Substantial Failure" or "Complete Failure."

In order to create such a framework, the WHOIS Review Team recommends defining accuracy rating scale categories that can be used to describe the accuracy of WHOIS records during future accuracy studies. The accuracy rating scale categories to be defined are: No Failure, Minimal Failure, Limited Failure, Substantial Failure, and Complete Failure.

[1] SSAC Report on Domain Name Registration Data Validation: http://www.icann.org/en/groups/ssac/documents/sac-058-en.pdf
[2] WHOIS Policy Review Team Recommendation 6 (May 2012)

As part of a consistent framework, the WHOIS Review Team also recommends defining a sampling methodology which can be used to inform consistent sampling techniques across future studies. An adequate methodology will include considerations for statistical significance, geographic stratification of new and existing gTLDs, and global representation of registrars and registrants.

2.2. Board of Directors. ICANN's Board of Directors offered specific actions[3] to build upon or enhance the recommendations provided by the WHOIS Review Team. Chief among these actions is that ICANN determine protocols for proactively identifying potentially inaccurate gTLD registrations information in gTLD registry and registrar services. When potentially inaccurate information is identified, registrars will be alerted and provided with enough specificity about the potential inaccuracies such that they might take corrective action.

The Board New gTLD Committee offers guidance in response to the NGPC Proposal for Implementation of GAC Safeguards Applicable to All New gTLDs[4]. In developing a framework to assess the accuracy of WHOIS data, ICANN is advised to conduct periodic sampling and analysis of WHOIS data across registries in an effort to identify potentially false, incomplete, or inaccurate records. Period sampling and analysis is recommended in intervals of 6 months, or twice per year. The guidance further specifies that ICANN should maintain statistical reports on the numbers of inaccurate WHOIS records identified, focusing on current version of WHOIS requirements but eventually broadening to include WHOIS directory services.

2.3. 2013 Registrar Accreditation Agreement (RAA) WHOIS Accuracy Program Specification.[5] The recommendations of the RAA WHOIS Accuracy Program Specification focus on compliance protocols that place during or immediately after the creation, transfer, or modification of a registration. The intent of these recommendations is to provide proactive checks of accuracy upon initial registration, during the transfer of a registration between registrars, or when the information in a registration is changed.

Further guidance is provided to specify which fields of a registration should be targeted for an assessment of compliance, and on what basis compliance should be assessed. The assessments would include checks on proper formatting of fields based on country or territory, including telephone and postal address fields, and also includes a check on email formats. An additional check would ensure that all required fields contain information. Registrars would verify email and phone accuracy by implementing automated authentication procedures in which a registrant confirms that they can be contacted using the reported email address and phone number.

2.4. Security and Stability Advisory Committee (SSAC) 58 Report.[6] Guidance from SSAC 58 asserts that accuracy should be assessed through three different perspectives: Syntactic, Operational, and Identity. These perspectives illustrate important differences in the way which the accuracy of WHOIS data may be defined and then assessed. These perspectives are briefly defined here, and are further defined in Section 3 of this report.

- Syntactic: assesses accuracy of WHOIS data by examining syntactic elements such as formats, standards, and logical values
- Operational: assesses accuracy of WHOIS data based on practical applicability of information provided in a WHOIS record

[3] Action Plan Appendix to ICANN Board Submission 2012-11-08-01 (1 November 2012)(incorporated into resolution 2012.11.08.02)
[4] NGPC Proposal for Implementation of GAC Safeguards Applicable to All New gTLDs (19 June 2013), adopted in Resolution 2013.06.25.NG02)
[5] www.icann.org/en/resources/registrars/raa/approved-with-specs-27jun13-en.htm#whois
[6] http://www.icann.org/en/groups/ssac/documents/sac-058-en.pdf

- Identity: assesses accuracy of a WHOIS record by determining if the information provided can be used to confirm the identity of the registrant

The SSAC recommends that researchers specifically identify and differentiate the perspectives from which data will be assessed. The recommendation further stipulates that separate samples be considered for each of the three perspectives. A process is depicted in which researchers begin with a sample which is assessed from the syntactic perspective. A smaller sample is then assessed from the operational perspective, and a yet smaller sample is chosen for assessment from the identity perspective.

In order to create consistent protocols for assessing accuracy, ICANN requires an examination of definitions and standards for what constitutes accurate email, telephone, and postal address information. The parameters for assessing the accuracy of these fields will differ from one perspective to another, but the process for determining when a field has been validated will be constant regardless of perspective. It is expected that validation of these fields will occur during assessment from the operational or the identity perspective.

SECTION 3: ACCURACY PERSPECTIVES AND RATING SCALE CATEGORIES

3.1. Perspectives. The three perspectives by which to assess accuracy, as discussed in the SSAC 58 Report on Domain Name Registration Data Validation, offer a hierarchical approach for evaluating accuracy in WHOIS data. There are value tradeoffs inherent in each perspective, and where one approach offers cost- and volume-efficiency, it compromises by providing a less tangible notion of accuracy.

The first of the assessment perspectives, Syntactic, examines WHOIS data with the intent of determining which data exists or is missing from within WHOIS data fields, and assesses whether data meets specified standards of formatting and structure. For example, a Syntactic check of telephone contact information might first seek to confirm the existence of data in the fields specified for telephone numbers. If data exists in the field, a Syntactic check might further seek to determine whether the characters in the field are numerical values that could indicate a potential phone number. If the field contains a potential phone number, a Syntactic check might then seek to match the format and the number of characters against known standards for phone numbers. Based on the rigor and the sequence of these checks, the accuracy of the phone number in a WHOIS record might be assessed from a Syntactic perspective.

SSAC 58 guidance expects that Syntactic assessment will involve the highest degree of automation and therefore volume-efficiency in assessing accuracy. A limitation of Syntactic assessment is that it does not provide a conclusive assessment of accuracy, however it does provide an ideal mechanism for determining which WHOIS data should be selected for further assessment from the Operational perspective.

The second assessment perspective, Operational, examines WHOIS data that are syntactically accurate (within reasonable bounds) with the intent of determining if the information can be practically used to achieve an intended function. In the case of postal address information, an Operational check of a syntactically correct address—one that is logically consistent with the country in which it is located based on postal address norms for that country—might include a determination of whether the postal address exists in public postal address data. A final check might include an attempt to determine whether the address can receive mail by sending a form letter which attempts to confirm contact with the registrant. The results of this mail check inform the Operational assessment, and if a response is received may also inform the Identity assessment. Similar to Syntactic checks, the effectiveness of assessing accuracy from an Operational perspective will be dependent on the design, sequence, and rigor of the checks

performed. Unlike Syntactic checks, Operational checks may involve gathering and utilizing external data to inform the accuracy assessment.

In contrast to Syntactic checks, Operational checks are expected to include a lower degree of automation, but a more tangible notion of accuracy. A limitation of Operational assessment is that while it may confirm practical applicability of the data, it does not determine if the identifying data belongs to the actual domain registrant.

The third assessment perspective, Identity, seeks to confirm that WHOIS data corresponds to the person or entity being represented as the registrant. If the mail check described above has been deemed operationally accurate (for example, by a delivery confirmation receipt) but did not receive a response, the data might be eligible for an Identity check which includes a second attempt to survey the mail recipient in such a way that confirms the recipient as the registrant of the domain. Similarly, a phone number that has been deemed operationally accurate (by confirming that the line exists and can receive calls) might be eligible for an Identity check that includes calling the phone line and attempting to survey the person who answers the call. Identity checks are performed almost exclusively by gathering external data to inform the accuracy assessment.

In comparison to Syntactic and Operational checks, Identity checks are expected to involve the least degree of automation and the most tangible notion of accuracy. For this reason, only the subset of records that were assessed as operationally accurate should be considered for assessment from the Identity perspective. The most significant limitation of Identity analysis is efficiency. Identity assessment typically gathers external data or performs manual checks by attempting to contact a registrant.

3.2. Rating Scale Categories. An accuracy assessment of WHOIS data, guided by any of the three perspectives, should result in the data being categorized into a rating scale based on the degree to which it can be considered accurate. The rating scale categories established for WHOIS accuracy studies describe accuracy as an inclusive range between No Failure, the highest degree of accuracy, and Complete Failure, which represents the lowest degree of accuracy. There are five rating scale categories in total, which are defined as:

- No Failure: Indicates that data are missing no significant information, and that the data provided are accurate.
- Minimal Failure: Indicates that data are missing no significant information. The record may need clarification or might benefit from additional information, but the data provided are accurate.
- Limited Failure: Indicates that the data are missing at least some significant information, and that the data present are to some degree considered inaccurate.
- Substantial Failure: Indicates that the data are missing significant information, and that the data provided are mostly inaccurate.
- Complete Failure: Indicates that the data are missing, or that data provided are completely inaccurate.

Due to the hierarchical assessment frameworks offered by the three perspectives, a rating scale category will be interpreted differently based on the perspective used to assess accuracy. For this reason, the rating scale categories are sufficiently broad so as to allow interpretation of the category in context to each of the three accuracy assessment perspectives.

It is expected that cases will exist where WHOIS data is categorized as having a high degree of accuracy from one perspective, and a low degree of accuracy from another perspective further down the hierarchy. An example of this would occur when data is analyzed from a Syntactic perspective and is discovered to contain information which has logical character values and proper formatting in all significant fields. The data would likely be categorized as

having No Failure from the Syntactic perspective. This same data when assessed from the Operational perspective, might be found to have some invalid contact information such as non-existent email and postal addresses, but found to have a telephone number which exists in public record and can receive phone calls. Thus, this same record would probably be categorized as having Limited Failure from the Operational perspective. If the same record were to be assessed for accuracy from the Identity perspective and the checks resulted in successful contact with a registrant who confirms their identity, the record might categorized as Minimal Failure from the Identity perspective.

SECTION 4: ACCURACY SCORING MODEL AND DEMONSTRATION

4.1 Accuracy Scoring Model Description. The accuracy rating scale categories discussed in Section 3.2 of this report provide qualitative descriptions of what constitutes in WHOIS data. A chief goal in designing future accuracy studies is to create a model which can determine for a WHOIS record, an *overall* accuracy rating which is informed by assessments from the three accuracy perspectives.

In order to assess overall accuracy ratings in a systematic and consistent fashion, ICANN suggests that a framework be created which can translate an accuracy rating scale category into a quantitative indicator of accuracy. For this example, accuracy rating scale categories have been translated into quantitative indicators in the form of scores ranging from negative two, to positive two, as seen below[7]:

Rating Scale Category		Quantitative Score
No Failure	=	2
Minimal Failure	=	1
Limited Failure	=	0
Substantial Failure	=	-1
Full Failure	=	-2

Quantitative indicators from each of the three perspectives will then be combined into an overall accuracy rating for each WHOIS record. A key challenge will be to design a model which determines how to combine quantitative indicators in way which can provide a meaningful assessment of accuracy within each perspective and *overall* accuracy.

An overall accuracy rating for the record might be determined by taking the sum of all quantitative perspective-based scores. However, such a model would not account for the relative importance of each perspective in determining overall accuracy. As mentioned earlier, the assessment perspectives represent a hierarchal approach to determining accuracy and it can be assumed that operational and identity assessments provide greater tangible value in determining the accuracy of a record. A successful model will account for these differences when combining the perspective-based scores to determine an overall accuracy rating. One potential approach is to apply a weighting scheme to each perspective so that when the perspective scores are combined to determine the overall accuracy score, each perspective has an impact proportional to its value in assessing accuracy.

An example of a model for assessing overall accuracy is provided in Figure A, below. The previously mentioned quantitative indicators would be determined for each field, by

[7] In Appendix A of this report, fifteen use-cases are provided to show examples of WHOIS records that have been assessed from a single perspective and assigned an accuracy rating scale category.

perspective, and recorded in the empty cells of Figure A. An overall "Perspective Score" is calculated as the average of all scores in a given perspective. The weights applied to each perspective are decimal values summing to a value of 1, or 100%, and are seen in bold print next the label for each perspective row. For the purposes of demonstration, there is an assumption data parsing of a WHOIS record can be done without error.

Figure A: Example of an Accuracy Scoring Model

	Name	Postal Address	Country	Email Address	Telephone Number	Perspective Score	
Syntactic **0.17**						X	(average Syntactic score)
Operational **0.33**						Y	(average Operational score)
Identity **0.50**						Z	(average Identity score)
Total Score						0.17X + 0.33Y + 0.5Z	(overall accuracy score)

Another consideration for assessing overall accuracy is the categorization of contact fields contained in a WHOIS record. The three contact categories of a WHOIS record are labeled as: Registrant, Administrative, and Technical. It is suggested that the Registrant contact category is, with all other factors being equal, the most significant contact category for assessing accuracy. Information in the Administrative and Technical contact categories should only be assessed to provide supplemental information for judging Registrant accuracy. That is, if data in the Registrant category is missing or inaccurate, data which is present or accurate within the Administrative and Technical categories may have an impact on an assessment of accuracy. For example, if the name field of both the Registrant and Administrative contact is the same, but the Registrant address is missing, the Administrative address, if present, could be used. A successful model will account for the relative value of information in one contact category based on the presence and accuracy of data in the other two categories.

4.2. Accuracy Scoring Model Demonstration. To further demonstrate how this example model might function, a fictional WHOIS record is provided below in Figure B, and assessed from each perspective. The assessment scores are then applied to the model in Figure A, and an overall accuracy rating is determined. The following narrative represents the process of

assessing and rating the accuracy of a WHOIS record. Assume that all data in this analysis is taken from the contact category of Registrant.

Figure B: Sample WHOIS Record

Sample WHOIS Record

Registrant:
John Doe john.doe@example.web
1234 Main Street
New York, NY 12345
United States
(555) 555-5555 Fax –
Registered through: Go Daddy
Domain Name: EXAMPLE.COM
Created on: 1-Mar-14
Expires on: 1-Mar-15
Last updated on: 1-Mar-14
Administrative Contact:
Doe, John john.doe@example.web
1234 Main Street
New York, NY 12345
United States
(555) 555-5555 Fax –
Technical Contact:

Domain servers in listed order:
A123.EXAMPLE.COM
A456.EXAMPLE.COM

An assessment of the accuracy of the record in Figure B, above, begins with the Syntactic perspective. Automated checks determine that data in the administrative category is present, and that it uses expected characters and formats. Based on these facts, all of these fields would be assessed as having No Failure from the Syntactic perspective, and will each receive the score of 2.

Examining this record from an Operational perspective reveals more about the accuracy of the record. Checks of the postal address use publicly available data to determine that a matching street address exists in New York, NY, which is located in the United States (Country score of 2). However, checks of the zip code determine that "12345" is not an existing zip code. The postal address could be described as mostly accurate but in need of clarification or additional information (easily obtained cross-referencing a public address database), which meets the definition for Minimal Failure and earns a quantitative score of 1. A check of the email address reveals that it does not use any known domains and may be invalid, and a test email is returned as undeliverable (Email Address score of -2). A check of the telephone

contact information reveals that it is a valid telephone number (Telephone Number score of 2) and the publicly listed owner of that phone number is "J.P. Doe" (Name score of 1).

When assessing this record from the Identity perspective, for the sake of efficiency, accuracy checks should begin with data fields that contain potentially correct information. In this case the email address has been determined as invalid, and the postal address has been determined to contain at least some inaccurate information[8]. However, the telephone number has been assessed as potentially accurate through both the Syntactic and Operational perspectives. An Identity check on the telephone number is performed by calling the telephone number and attempting to confirm that the number can be used to contact the registrant or a representative of the registrant. The first call is not answered, and does not lead to a voicemail. The second attempt to call leads to contact with a person who identifies themselves as Jane Doe. Jane states the domain registration belongs to her husband John Doe (Name score of 2), but that John cannot be contacted at this number during business hours. While the telephone number is technically accurate, it is not ideal for contacting the owner of the registration without further clarification or additional information, and earns a score of 1. Jane confirms that the address in the record is correct except for the zip code (Postal Address score of 1), and confirms that the address is in the United States (Country score of 2). Finally, Jane states that the email address in the record is inaccurate, and she is unwilling to confirm the correct email address, resulting in a score of -2 for the Email field.

The scores assessed from each perspective, as described above, are now incorporated into the model as seen below in Figure C. An average score is computed for each perspective and recorded in the perspective score column. The average of each perspective is then multiplied by the given perspective weight, and these three values are added together to determine an overall quantitative accuracy score. This quantitative score may then be used to determine the rating scale category of a record.

Figure C: Accuracy Scoring Model Demonstration

	Name	Postal Address	Country	Email Address	Telephone Number	Perspective Score	
Syntactic **0.17**	2.00	2.00	2.00	2.00	2.00	2.00	(average Syntactic score)
Operational **0.33**	1.00	1.00	2.00	-2.00	2.00	0.80	(average Operational score)
Identity **0.50**	2.00	1.00	2.00	-2.00	1.00	0.80	(average Identity score)
Total Score						1.00	(overall accuracy score)

The overall accuracy score for the sample record in Figure B is calculated at 1, by assigning the values from each perspective score to the formula described in figure A (.17×2.00) + (.33×0.80) + (.5×0.80) = 1. According to the range of quantitative scores established for this model, the score of 1 can be approximated to the category of Minimal Failure. This example demonstrates a process by which a record might be assessed and scored from each perspective, and how a scoring model might be used to determine an overall accuracy rating for the record.

SECTION 5: SAMPLE DESIGN

The ICANN WHOIS Accuracy studies are envisioned as a semi-annual series of studies to track trends in the accuracy of WHOIS records. Each of the series will examine WHOIS data

[8] The inaccuracy is minor, so proceeding with an Identity assessment of the address is a reasonable option.

for a representative sample of domain names. In studying accuracy, there are several possible sub-groups that are of interest: registry operators, generic Top-Level Domains (gTLDs), countries (or regions), and registrars. It is only possible for ICANN to select stratified samples by gTLD. Country and registrar cannot be identified for the population, and must be identified for a sample. Therefore, a two-phase sampling process will be necessary. As described below, registry operators are tied to gTLD, so if we stratify by gTLD, we are stratifying by registry operator as well.

All analysis subgroups should have a sufficient sample size for estimates to have a 95 percent confidence interval for a binary percentage of 50 percent of plus or minus 5 percent. It is permissible to have one or more combined or "Other" categories as analysis subgroups. If the sample design involves clustering or oversampling, the design effects due to clustering and weighting must be taken into account. As described below, there can be different sample sizes for all three different accuracy steps. Sample sizes will be smallest for the Identity step, so the number of possible analysis subgroups will be smallest. We suggest that contractors start their design work with the sample size for the Identity step. Sample sizes for the Syntactic step can be large (the original previous study used 2,400, but larger should be possible with mostly automatic checks). Then, the sample size for the Operational Accuracy step can be customized depending on how the contractor will carry out this step.

As described above, the ICANN WHOIS Accuracy studies will occur in three steps. At the first step, Syntactic accuracy will be checked with mostly automatic checks. This step can process a larger number of domain names because it is largely automatic. At the second step, Operational accuracy will be checked with some automatic checks, but may involve gathering and utilizing external data to inform the accuracy assessment. Therefore, it will be more expensive to check operational accuracy on a large number of records. It may be desirable to carry out operational accuracy checks on a subsample of the domain names processed for Syntactic accuracy. At the third step, Identity accuracy will be checked by contacting the registrant. Since this is a manual process, the number of domain names that can be checked at this stage is even smaller than for operational accuracy to keep expenses under control. In addition, since these studies will involve direct contact with the sampled registrants, and registrants in this universe come from all countries on the globe, it would be cost prohibitive to use a totally unrestricted sample. Therefore, the sample needs to be clustered by country so the number of countries (and languages and systems) involved would be restricted to a manageable count. Therefore, the sample must be smallest for the Identity accuracy step, can be larger for the Operational accuracy step, and can be even larger for the Syntactic accuracy step. Each further step of accuracy checks can subsample the domain names from the previous step with a sampling strategy that meets the analysis goals.

5.1. Stratification by global Top-Level Domain. There are 18 gTLDs managed by ICANN with active domains according to the November 2013 Registry Operator Monthly Reports at http://www.icann.org/en/tlds/monthly-reports/. Excluded from eligibility are .edu, .mil, and .gov, which are not administered by ICANN, and are therefore out of scope. The responsibility for operating each gTLD (including maintaining a registry of the domain names within the gTLD) is delegated to a particular organization. These organizations are referred to as "registry operators" or "sponsors". Currently, each gTLD has its own sponsor, except that Verisign, Inc. is the sponsor for three gTLDs: *.com, *.name, and *.net. Therefore, except for Verisign, Inc., the subgroups of sponsors are subgroups of gTLDs. Organizing the sample by gTLD allows estimates by gTLD as well as estimates by registry operator. Previous studies have been limited to the top five most common gTLDs (*.com, *.net, *.org, *.info, and *.biz). All gTLDs are of interest, but many are too small to be analysis subgroups on their own.

5.2. Cluster sample of countries and regions. As stated above, the identity sample must be clustered by country to limit the number of languages and systems involved to a manageable count. However, it is not possible to select from the ICANN databases by anything other than gTLD. Therefore, the sample must be in at least two phases. First, a larger sample of domain names needs to be selected, which we will refer to as the "microcosm" (it is a smaller sample

that represents the whole ICANN domain name universe). Variables such as country and region must be determined for all of the domain names in the microcosm. This can be done by extracting the corresponding WHOIS records for each of these domains, and the country of the registrant must then be coded for each domain name directly from the WHOIS record. A larger microcosm allows more flexibility in the second phase of sampling, but increases costs because variables such as country and region must be identified for more domain names. The syntactic accuracy checks should be done for the entire microcosm. It will be necessary to subsample for the identity accuracy checks, and it may be necessary and/or desirable to also subsample for the operational accuracy checks.

5.3. Analyzing Accuracy by Registrars. ICANN is interested in tracking the accuracy by registrar, so one of the variables that must be filled in for every domain name in the microcosm is registrar. It will not be possible for every registrar to be an analysis subgroup, but contractors may consider oversampling and undersampling certain registrars to allow the most efficient and cost-effective tracking of accuracy for the largest number of registrars.

5.4. Implementation Walkthrough. Each ICANN WHOIS Accuracy Study microcosm sample will be drawn by ICANN with the sample specifications designed by the contractor with ICANN input and agreement. ICANN can stratify the sample according to gTLD prior to sample selection. The contractor can specify the sample size within each gTLD. Within gTLD, no stratification can be done; instead, ICANN's selection will be equivalent to a simple random sample by gTLD. It is possible to stratify by gTLD because each gTLD is covered by a different zone file that ICANN uses to sample domain names. However, there are no other useful variables within the zone file for sorting; by location, for example. Additionally, the registrar information cannot be obtained without WHOIS data extraction.

Along with the microcosm sample of domain names, the contractor should assume that ICANN will deliver the domain names as well as the correctly parsed WHOIS record from the ICANN WHOIS database. The format for delivery of the microcosm sample file of domain names can be any convenient file type to ICANN and the contractor (text, Excel, SAS). The contractor will then conduct the syntactic accuracy checks on the entire microcosm sample. The contractor will then conduct the operational accuracy checks on the entire microcosm sample or a subsample. Prior to the identity accuracy checks, a clustered sample must be drawn by country or region. The contractor will have coded and standardized country or region for the entire microcosm sample. The contractor must specify a strategy for selecting the sample for identity accuracy checks with respect to registrars and keeping the ability to track accuracy by gTLD. Then, the contractor will conduct the identity accuracy checks on this clustered sample, analyze all three accuracy check results, and report on the findings.