



WHOIS Accuracy Reporting System (ARS)

Phase 1 Report: Syntax Accuracy
Global Domains Division | 24 August 2015

Contents

Executive Summary	4
Project Background and Goals	4
Syntax Accuracy Testing Methods	4
Sample Design	5
Findings	5
Next Steps	6
Introduction	8
WHOIS ARS Background	8
Lessons Learned from the Pilot Study	9
Phase 1 and the ARS Project Approach	10
Phase 1 Overview	10
Study Methods and Approach	14
Brief Overview	14
Sample Design	14
Syntax Testing Methods	17
Notes on the Analysis of Syntax Accuracy Testing Results	19
Main Findings	22
Summary of Findings	22
Overall Conformance to 2009 RAA Requirements	23
Next Steps	31
Phase 2 Look-Ahead	31
ICANN Contractual Compliance Next Steps	31
Appendix A: Syntax Accuracy Criteria	33
Accuracy Criteria and the RAA	33
Email Addresses	33
Telephone Numbers	36
Postal Addresses	39
Appendix B: Accuracy Results Tables 2009 RAA Requirements	43
2009 RAA Requirements – Reasons for Error	43

Appendix C: Accuracy Result Tables, 2013 RAA Requirements.....	47
Conformance to 2013 RAA Requirements	47
Major Reasons for Errors under 2013 RAA.....	49

Executive Summary

Project Background and Goals

On 8 November 2012, the ICANN Board approved a series of improvements to the manner in which ICANN carries out its oversight of the WHOIS Program, in response to recommendations compiled and delivered by the 2012 WHOIS Review Team, under the Affirmation of Commitments (AoC)¹.

As part of these improvements, ICANN committed to proactively identifying potentially inaccurate generic Top Level Domain (gTLD) WHOIS contact data and forwarding potentially inaccurate records to gTLD registrars for investigation and follow-up. To accomplish these tasks and address Governmental Advisory Committee (GAC) concerns on WHOIS accuracy ICANN initiated the development of the WHOIS Accuracy Reporting System (ARS)—a framework for conducting repeatable assessments of WHOIS accuracy, publicly report the findings, and provide data to the ICANN Contractual Compliance team to follow up on potentially inaccurate records with registrars.

With input from the community, ICANN designed the ARS to be organized into three Phases based on the types of validations described in the SAC058 Report² (Syntax, Operability, and Identity). Phase 1 is centered on syntax accuracy of WHOIS records (i.e., is the contact data complete and formatted correctly?), and is the subject of this report.

Using the requirements of the Registrar Accreditation Agreements (RAA) to derive the syntax accuracy criteria, ICANN sought to determine whether WHOIS records are meeting format and content requirements. ICANN Contractual Compliance will follow-up with registrars on potentially inaccurate records, leading to investigation, and if needed, correction. This report details the leading types of non-conformances, trends and comparisons of WHOIS accuracy across regions, RAA versions and gTLD types. The ARS should lead to improvements over time in the accuracy of WHOIS data, which will be examined in subsequent ARS reports.

Syntax Accuracy Testing Methods³

Syntax accuracy testing was designed to assess the contact information of a WHOIS record by comparing it to the formats specified by the applicable contractual requirements of the RAA. Syntax tests were performed on all nine individual contact information fields in a record (i.e., email address, telephone number, and postal address for the Registrant, Administrative, and Technical Contacts) and compiled as an entire record. The resulting data were analyzed to produce statistics of syntax accuracy for WHOIS contact information across subgroups such as gTLD, Region, and RAA type (i.e., 2009 RAA or 2013 RAA⁴).

¹ <https://www.icann.org/resources/pages/aoc-2012-02-25-en>.

² <https://www.icann.org/en/system/files/files/sac-058-en.pdf>.

³ More detailed descriptions of the syntax accuracy tests can be found in [Appendix A: Syntax Accuracy Criteria](#).

⁴ See here for RAA versions: <https://www.icann.org/resources/pages/registrars/registrars-en>

More information on the methodology of this study can be found in the [Study Methods and Approach](#) section in the body of this report.

The 2013 RAA requires the contact data in a WHOIS record to be more complete and to be formatted per more specific requirements than that of the 2009 RAA. For example, the 2009 RAA requires postal addresses with a valid country, whereas the 2013 RAA requires the country in the address to be formatted per the 2-letter code from ISO-3166-1⁵. The validation criteria were designed in such a way that all records in the analyzed subset would be evaluated against a set of baseline requirements derived from the requirements of the 2009 RAA. Additional testing to the 2013 RAA requirements was applied to only those records where the 2013 RAA is the applicable agreement.

Sample Design

At the time of the initial sample in early 2nd quarter 2015 there were nearly 157 million domain names spread across 610 gTLDs. Nearly 97 percent of these domains were registered in one of the 18 Prior gTLDs⁶, and about 3 percent were registered in one of the 592⁷ New gTLDs. A two-stage sampling method (initial → analyzed subsample) was designed to provide a large enough sample to reliably estimate subgroups of interest, such as ICANN region, New gTLD or Prior gTLD, and RAA type. The initial sample contained nearly 100,000 records and the analyzed subsample contained 10,000 records representing all 610 active gTLDs at the time⁸. More detailed information can be found in the [Sample Design](#) section in the body of this report.

Though an estimated 97 percent of domain names are registered through registrars which have agreed to 2013 RAA, some domains are registered by registrars which are allowed to operate under 2009 RAA standards that were applicable at the time of registration. This could be for one of two reasons: 1) the registrar has not yet signed a 2013 RAA with ICANN and is only subjected to 2009 RAA standards; or, 2) the registrar agreed to 2013 RAA with ICANN but the domain was registered before the effective date of the registrar's 2013 RAA. We refer to the latter group of domains as 2013 RAA Grandfathered (2013 RAA GF) domains. Our analysis thus includes three mutually exclusive RAA subgroups: 2009 RAA, 2013 RAA GF, and 2013 RAA non-grandfathered (referred to as 2013 RAA NGF).

Findings

All 10,000 records in the analyzed subsample were evaluated using the 2009 RAA criteria, and this report uses the 2009 criteria as a baseline to assess the overall accuracy of WHOIS records in gTLDs. Phase 1 findings show the rates of syntax accuracy of WHOIS contact information over several dimensions, focusing on rates of conformance by contact mode (Email, Telephone or Post) to the requirements of RAAs (2009 RAA or 2013 RAA). The results from the analyzed subsample testing are then used to

⁵ http://www.iso.org/iso/home/standards/country_codes.htm.

⁶ Prior gTLDs referring to the 18 gTLDs established before the 2012 round of New gTLDs

⁷ At the time of data collection. See here for latest delegation information: <https://newgtlds.icann.org/en/program-status/delegated-strings>.

⁸ 597 New gTLDs, 18 Prior gTLDs.

estimate the results for the entire gTLD population or the particular subgroup of interest. These data are presented in this report at a 95 percent confidence interval⁹ with an estimated percentage plus or minus approximately two standard errors. Based on sampling error, there is a 95 percent chance that the true parameter (e.g., the percentage of all administrative telephone numbers with a North American registrant that pass all syntax accuracy tests) is within the 95 percent confidence interval. Our analysis finds that approximately 99 percent of email addresses, 85 percent of telephone numbers, and 79 percent of postal addresses met all of the baseline syntax requirements of the 2009 RAA¹⁰. Full accuracy of an entire WHOIS record (all three contact types, for all three contact modes) to the requirements of the 2009 RAA was approximately 70 percent for the gTLD population as a whole. Table 1.Ex below provides the accuracy breakdown by contact mode, presented as 95 percent confidence intervals.

Table 1Ex: Overall¹¹ gTLD conformance to 2009 RAA Requirements by mode

	E-mail	Telephone	Postal Address	ALL 3 Accurate
All 3 Contacts Accurate	99.2% ± 0.2%	85.8% ± 0.7%	79.1% ± 0.8%	70.3% ± 0.9%

The leading causes of syntax nonconformance (including for 2013 RAA requirements) in the various subgroups are examined and explained in the [Findings](#) section in the body of this report, as well as in [Appendix B: Accuracy Result Tables](#). At a high level we can conclude that:

- If an email is provided, it always passed all syntax accuracy tests
- Two-thirds of the telephone numbers that failed at least one syntax accuracy check (13 percent of all telephone numbers) failed the length criteria for the applicable country
- Postal addresses that failed at least one syntax accuracy check (23 percent of all postal addresses) were typically missing at least one required field such as postal code, state, city, or street

Next Steps

Phase 2

In late 2nd quarter 2015, ICANN also began work on Phase 2 of the ARS, which includes both Syntax and Operability Testing. The WHOIS ARS team has collected the initial sample and selected the subsample for analysis and is working to define the Operability criteria. Testing is to be complete by early 4th quarter 2015, and a Phase 2 report is targeted for publication in late 4th quarter 2015.

ICANN Contractual Compliance

⁹ This means that if the population is sampled again, the confidence intervals would bracket the subgroup or parameter (e.g., accuracy by Region) in approximately 95 percent of the cases. For more information on confidence intervals, see here: <http://www.itl.nist.gov/div898/handbook/prc/section1/prc14.htm>.

¹⁰ While the 2009 RAA does not contain explicit syntax requirements, the contact data provided is expected to be formatted correctly and completely. The criteria listed in [Appendix A](#) are what we have defined as the baseline requirements of contact data to be deemed formatted correctly and completely. Throughout the document and appendices, the “syntax requirements of the 2009 RAA” refer to these criteria.

¹¹ Overall accuracy refers here to the entire 157 million domains; See note 9 on confidence intervals and population.

The results of Phase 1 have already been provided to ICANN Contractual Compliance for review and processing. Following the internal review, ICANN Contractual Compliance will assess the types of errors found, as well as the type of follow-up required with registrars. Depending upon the nature of the error, tickets may be processed as WHOIS inaccuracy complaints or WHOIS format complaints. All WHOIS ARS complaints will follow the Contractual Compliance Approach and Process ¹²). However, ICANN Contractual Compliance will continue to give priority to complaints submitted by community members. An advisory panel of registrar-volunteers has also been convened to provide input to ICANN as its practices related to processing of ARS-generated inaccuracy reports evolve.

¹² See here for the ICANN Contractual Compliance Approach and Process:
<https://www.icann.org/resources/pages/approach-processes-2012-02-25-en>.

Introduction

WHOIS ARS Background

On 8 November 2012, the ICANN Board approved a series of improvements to the manner in which ICANN carries out its oversight of the WHOIS Program, in response to recommendations compiled and delivered by the 2012 WHOIS Review Team, under the Affirmation of Commitments (AoC)¹³.

As part of these improvements, ICANN committed to proactively identifying potentially inaccurate generic Top Level Domain (gTLD) WHOIS contact data and forwarding potentially inaccurate records to gTLD registrars for investigation and follow-up. To accomplish these tasks and address Governmental Advisory Committee (GAC) concerns on WHOIS accuracy, ICANN initiated the development of the WHOIS Accuracy Reporting System (ARS) – a framework for conducting repeatable assessments of WHOIS accuracy, publicly reporting the findings, and providing resulting data to compliance for follow up with registrars on potentially inaccurate records.

The design of the ARS was developed in consultation with the ICANN community. Beginning in March, 2014, ICANN posted a Draft Implementation Plan¹⁴ for the ARS, for public comment¹⁵, to solicit feedback on the proposed model¹⁶ that described the methodology and approach. Following the closure of the public comment, ICANN released a Request for Proposal¹⁷ in May, 2014, to seek vendors to provide services to support the ARS development using an updated methodology that took into account the feedback¹⁸ received from the public.

A Pilot Study was then conducted over the course of the 2nd half of 2014 as a “proof-of-concept” for the ARS, designed with the intention of testing key components of the system and ensuring that ICANN is aligned with other activities related to WHOIS, including, examining the effectiveness of the new validation and verification requirements under the 2013 Registrar Accreditation Agreement (RAA), and the various GAC Communiqués that address WHOIS accuracy.

ICANN conducted the Pilot Study to test the proposed methodology and to determine the feasibility of deploying the technology and services of the vendors identified during the RFP process, using real data from gTLD WHOIS records. A Preliminary Findings Paper was published¹⁹ prior to the ICANN 51 Los Angeles Meeting to facilitate consultations with the ICANN community regarding the methodology,

¹³ <https://www.icann.org/resources/pages/aoc-2012-02-25-en>.

¹⁴ <http://whois.icann.org/sites/default/files/files/online-accuracy-reporting-plan-11mar14-en.pdf>.

¹⁵ <https://www.icann.org/public-comments/whois-accuracy-reporting-2014-03-11-en>.

¹⁶ <https://www.icann.org/news/blog/a-model-for-exploring-whois-accuracy>.

¹⁷ <https://www.icann.org/news/announcement-2014-05-19-en>.

¹⁸ <https://www.icann.org/en/system/files/files/report-comments-whois-accuracy-reporting-01oct14-en.pdf>.

¹⁹ <https://www.icann.org/news/announcement-2014-10-10-en>.

findings, and next steps for the development of the ARS. ICANN received input from the GAC²⁰, Registrar Stakeholder Group, the Intellectual Property Constituency and others on the Preliminary Findings Paper. In December 2014, a final Pilot Study report²¹ was released followed by a public comment period which concluded in March 2015.

This Phase 1 Report of the WHOIS Accuracy Reporting System (ARS) acts as a follow-on to the Pilot Study. Phase 1 takes lessons learned from the Pilot Study and operationalizes the ARS, and is the first semi-annual WHOIS accuracy assessment.

Lessons Learned from the Pilot Study

The Pilot Study and subsequent public comment²² provided a means for ICANN to learn from and improve the WHOIS ARS. Specifically, ICANN implemented the following changes in the Phase 1 study based on lessons learned:

- Align accuracy criteria to specific RAA requirements for WHOIS records and account for the record's applicable RAA version (2009 vs. 2013) based on record creation date (e.g., some domains with registrars on the 2013 RAA are still only obligated to meet the WHOIS requirements of the 2009 RAA based on when the domain was registered; this report refers to these as 2013 Grandfathered records or 2013GF)
- Test the Registrant, Technical Contact and the Administrative Contact information listed in WHOIS records for completeness
- Use a series of detailed syntax tests with binary outcomes to determine what types of non-conformances are occurring, and use this information to help pinpoint areas for needed improvement
- Treat Privacy/Proxy records same as all other record types²³

The testing of WHOIS records is more streamlined with the above changes implemented. The records are more easily analyzed and the ARS itself is more effective in reaching its goals of determining potential inaccuracies, facilitating ICANN's Contractual Compliance follow up with registrars and allowing for registrars to investigate and improve the data.

²⁰ <https://gacweb.icann.org/display/gacweb/Governmental+Advisory+Committee>.

²¹ <https://www.icann.org/public-comments/whois-ars-pilot-2014-12-23-en>.

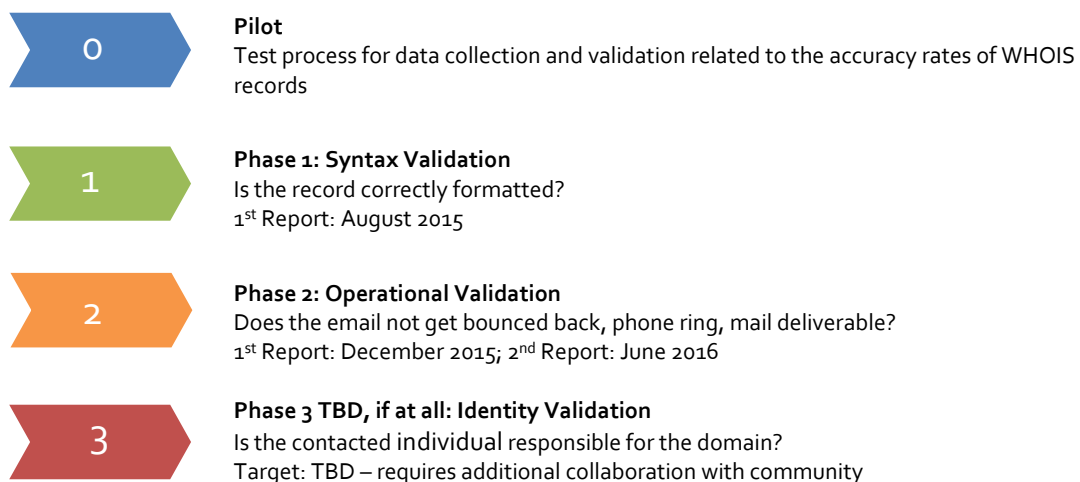
²² <https://www.icann.org/en/system/files/files/report-comments-whois-ars-pilot-03apr15-en.pdf>.

²³ A Privacy/Proxy Accreditation process is currently in the GNSO policy development process. Currently there are no discrete fields in the WHOIS output to indicate a record is utilizing a Privacy or Proxy Service, nor a centralized list of Privacy or Proxy service providers, making it impractical to systematically identify records as Privacy or Proxy.

Phase 1 and the ARS Project Approach

The WHOIS ARS will produce repeatable assessments and reports of WHOIS accuracy over time. With input from the community over the course of 2014, ICANN designed the ARS to be organized into three Phases, based on the types of validations described in the SACo58 Report²⁴. Phase 1 is centered on the syntax accuracy of WHOIS contact information (i.e., is the contact data complete and formatted correctly?). Phase 2 will assess the operability of the contact data in the record by combining the Syntax Testing from Phase 1 with Operability tests such as “Does the phone ring?” and “Does the email go through?” Phase 3 is intended to look at Identity validations, i.e., determining if the individuals listed in a WHOIS record are the responsible individuals for the domain name. However, at this time, Phase 3 is not on an implementation timeline as ICANN and the Community are still assessing if Phase 3 will be implemented at all, and if so, how the validations would occur and what the criteria for success would be. Figure 1 illustrates this phased approach. ICANN expects to produce the 1st Phase 2 Report in December 2015, and produce new Phase 2 reports every 6 months.

Figure 1: WHOIS ARS Phases



Phase 1 Overview

Project Plan, Tasks and Timeline

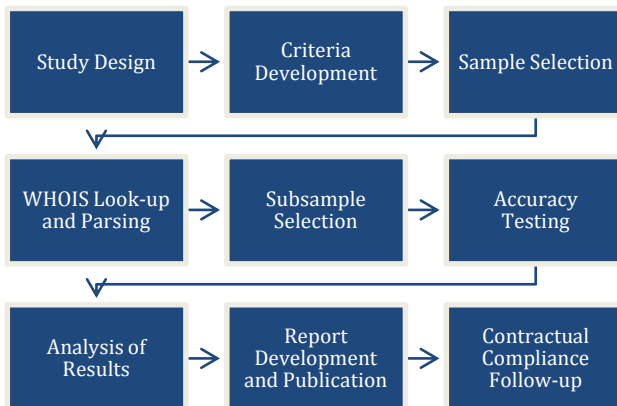
Work on Phase 1 began shortly after the conclusion of the Pilot Study. After a month of pre-planning as well as reviewing public comments and lessons learned, Phase 1 was kicked-off in early 2nd quarter 2015. ICANN and the WHOIS ARS vendors approached the work collaboratively; the WHOIS ARS team²⁵ defined together the sampling methodology (e.g., groups or sectors, size, and selection methodology, and testing approach) based on what levels of data were to be reported. The team also together defined

²⁴ <https://www.icann.org/en/system/files/files/sac-058-en.pdf>.

²⁵ Throughout this report, ICANN and the WHOIS ARS team of vendors will be referred to collectively as “the WHOIS ARS team” and, where applicable, a collective “we” will be used to refer in the first person to actions completed by this team.

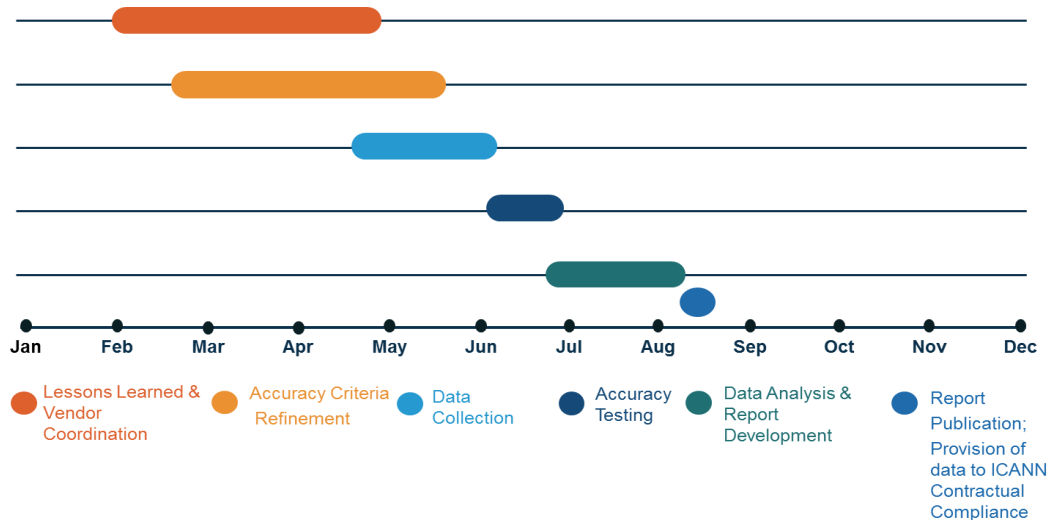
the [accuracy testing criteria](#), ensuring that they both aligned to RAA requirements and that the vendors' accuracy testing systems could accommodate the criteria. Phase 1 can be divided up into nine main tasks, which are illustrated below in Figure 2.

Figure 2: Work Flow and Tasks



These tasks²⁶ were conducted by the team in the timeline illustrated in Figure 3. As is evident from the timeline, Phase 1 included a pre-phase development period based on the outcome from the Pilot Study. These lessons learned helped to shape the accuracy criteria, which were not developed statically but rather throughout the first few months of Phase 1 up until the point of accuracy testing. This time allowed for back-and-forth amongst the members of the WHOIS ARS team to ensure the criteria were correctly mapped to RAA requirements. During this time, the accuracy testing vendors also were able to use the WHOIS data from the Pilot Study to help calibrate their tools for Phase 1.

Figure 3: Phase 1 Timeline



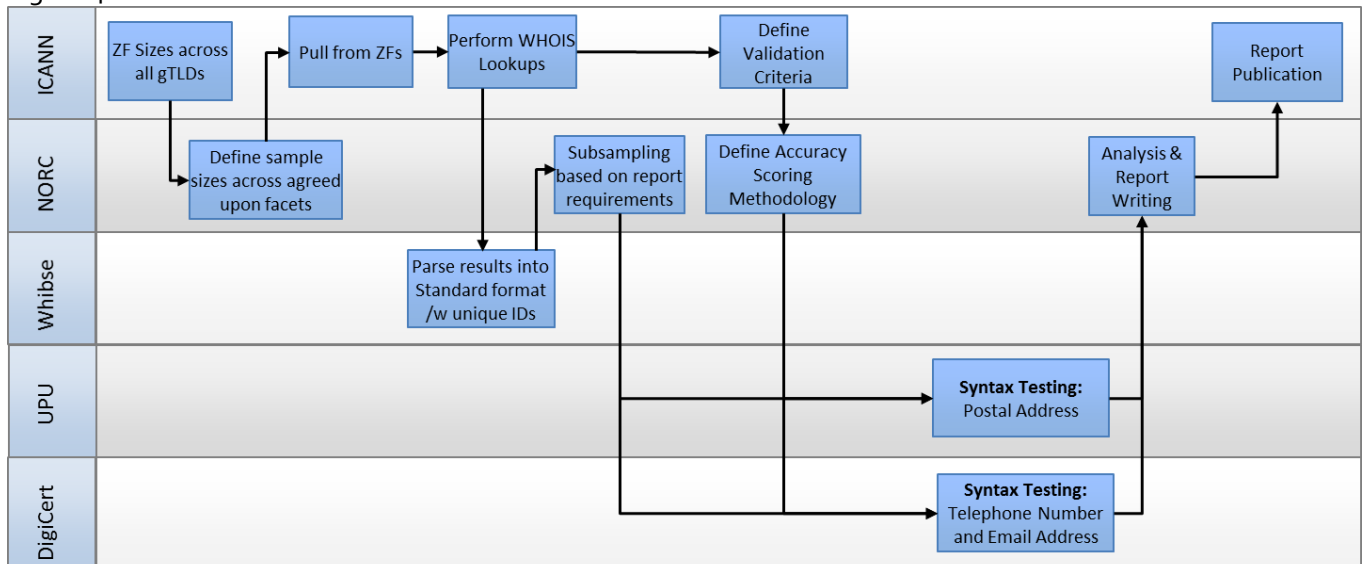
²⁶ With the exception of the Contractual Compliance Follow up, which will be beginning shortly after publication of this report.

Team Composition

As mentioned, the above tasks were carried out by the WHOIS ARS team, the members of which are listed below. Figure 4 below also illustrates further how the team coordinated to develop the Phase 1 report.

- ICANN: Project Management, Accuracy Criteria Definition, WHOIS Lookups
- NORC at the University of Chicago²⁷: Study Design, Sample Selection, and Data Analysis
- Whibse, Inc²⁸: Parsing
- DigiCert, Inc²⁹: Email and Telephone Accuracy Testing
- Universal Postal Union³⁰: Postal Address Accuracy Testing

Figure 4: ICANN Coordination with Vendors



Phase 1 Goals

The overarching goal of the WHOIS ARS is to identify potentially inaccurate records and report on these in such a way to allow for improving WHOIS accuracy over time. The WHOIS ARS provides a mechanism for sampling from the mass amounts of gTLD registration data and estimating the accuracy across the population of 150m+ domain names. ICANN intends the reports and data produced through the WHOIS ARS to inform the community, and help guide the community's future efforts to improve the quality of WHOIS data.

Specifically, for Phase 1, ICANN's goal was to focus on Syntax Accuracy, and using the defined criteria ICANN sought to determine whether the WHOIS record is meeting the format and content requirements

²⁷ <http://www.norc.org/Pages/default.aspx>.

²⁸ <https://www.whibse.com/>.

²⁹ <https://www.digicert.com>.

³⁰ <http://www.upu.int/en.html>.

of the applicable RAA. Likewise, the Phase 1 report is to provide actionable data regarding the syntax of WHOIS records. This report details the leading types of non-conformance, trends and comparisons of WHOIS accuracy across regions, RAA and gTLD types. Finally, the underlying data allows for ICANN Contractual Compliance to follow-up with registrars on potentially inaccurate records, leading to investigation, and if needed, correction. In the end, the ARS should lead to improvement in the WHOIS data over time, which will be examined in subsequent ARS reports (Phase 2).

Study Methods and Approach

Brief Overview

The Phase One study first selected a sample of around 100,000 WHOIS records from the Zone Files of 610 gTLDs. Using a series of systematic assessments, collectively referred to as *Syntax Accuracy Testing*, the contact information from a subsample of 10,000 records was checked for consistency with syntax standards (e.g. values and formats) based on requirements stipulated within the domain-applicable Registrar Accreditation Agreement (RAA). The resulting data were analyzed to produce statistics of syntax accuracy for WHOIS contact information across subgroups such as gTLD Type (Prior or New), ICANN Region, and RAA type. Some domains with registrars on the 2013 RAA are still only obligated to meet the WHOIS requirements of the 2009 RAA³¹ based on when the domain was registered; we refer to such domains as 2013 RAA grandfathered (2013 RAA GF). Domains with registrars on 2013 RAA obligated to meet the WHOIS requirements of the 2013 RAA are referred to as 2013 RAA non-grandfathered (2013 RAA NGF). Thus, analyses considered three RAA Types (2009, 2013GF, and 2013 NGF).

Sample Design

Study data consisted of an initial sample of 100,000 records from gTLD Zone Files, and an analyzed subsample of 10,000 records. This two-stage sample was designed to provide a large enough sample to reliably estimate subgroups of interest, given the technical limitations of collecting study data. The data within gTLD Zone Files is limited, and does not contain the full set of WHOIS information (such as registrant country, registrar RAA version, etc.) necessary for selecting a sample with sufficient size to produce reliable subgroup syntax accuracy estimates. In order to obtain the required information, WHOIS queries are conducted for each record in the initial sample, and the required additional information is then appended to each record. By appending this additional information to records of the initial sample, it is possible to select a subsample that contains adequate representation of the subgroups of interest, in this case 10,000 records. Summary statistics of the initial sample, and the methods for selecting the subsample are described below.

Initial Sample

To select the initial sample of 100,000 records, we have only the Zone File summary data, which indicates how many domains are in each gTLD. At the time of the initial sample in early 2nd quarter 2015 there were nearly 157 million domains names spread across 610 gTLDs³². Nearly 97 percent of the 157 million domains were registered in one of the 18 Prior gTLDs, and about 3 percent were registered in one of the

³¹ See note 10.

³² 597 New gTLDs, 18 Prior gTLDs. Prior gTLDs refers to the gTLDs established before the 2012 round of New gTLDs.

592³³ New gTLDs. Out of the 592 New gTLDs, only 416 had at least one domain³⁴ (176 New gTLDs did not yet have any domains), 26 had exactly one domain, and 51 others had 2-10 domains.

Similar to the Pilot Study, our Phase 1 sample design selected 100,000 records and oversampled New gTLDs so that 25 percent of the initial sample were from New gTLDs. It is expected that New gTLDs will grow over time and we wanted a sufficient number in our sample to study. To make sure all New gTLDs were represented, we first selected 10 from each New gTLD (all if a New gTLD had 10 or less). The remaining sample was selected proportional to size (more were selected from larger New gTLDs). Similarly, we selected a minimum of 30 domains from all Prior gTLDs, with the remaining sample proportional to size. All sampling was done by systematic sampling within gTLD. Based on the sample size determined for each gTLD, a skip interval was determined (total number of domains divided by the desired sample size). Then, a random start between zero and the skip interval was determined. If this random start was 16.6 and the skip interval was 300, then the selected records would be the 17th (random start rounded up), the 317th, and so on. This methodology results in an implicitly stratified sample by any partial or complete sorting within the gTLD Zone File (e.g. newer domains sorting to the top or bottom of the list of records in the Zone file). Our method results in a slight oversampling of smaller gTLDs, while keeping very similar weights among the larger gTLDs, to ensure that variances are not inflated by differential weights.

WHOIS lookups were performed for all 100,000 records, but 1,916 (1.9 percent) were dropped because they were test records (1,212) or domains that no longer existed (704). This left us with an initial sample of 98,084 domains. The initial samples in Phase 2 will exclude gTLDs with exactly one domain in attempt to reduce this issue.

Of the initial sample of 98,084, records in the 2009 RAA subgroup accounted for 2.7 percent of all records, while 2013 grandfathered (2013 GF) and 2013 non-grandfathered (2013 NGF) records accounted for 50.8 percent and 46.4 percent of all domains respectively (see Table 1).

Table 1: Initial Sample Sizes by Region and RAA

RAA Type	Africa	Latin America and Caribbean	Europe	Asia Pacific	North America	Unknown	TOTAL
2009	23	203	258	484	1,688	27	2,683
2013 GF	310	1,451	10,807	8,388	28,592	300	49,848
2013 NGF	376	2,077	10,523	14,497	17,781	299	45,553
TOTAL	709	3,731	21,588	23,369	48,061	626	98,084

Analyzed Subsample

ICANN defined the subgroups of interest for this report to be records with 2009 RAA registrars, records with 2013 RAA registrars, records in New gTLDs, records in Prior gTLDs, and records from each of the 5 ICANN regions. Accordingly, we selected the analyzed subsample to ensure adequate representation

³³ At the time of data collection. See here for latest delegation information: <https://newgtlds.icann.org/en/program-status/delegated-strings>.

³⁴ The domain nic.TLD was excluded from the counts of domains within each gTLD, as this is a mandatory site operated by the registry for all new gTLDs per the New gTLD Registry Agreement.

such that the reporting can estimate the accuracy of the data for subgroups of interest with 95 percent confidence intervals. This required certain subgroups to be over-sampled relative to their representation in the total gTLD population (157M domains). While sampling did not specifically ensure that all registrars were included, sampling by every TLD, RAA type and registrant region did achieve registrar diversity in the analyzed subsample with nearly 400 registrars represented in the subsample.

The subsample also did not consider gTLD type (Prior versus New) because the initial sample oversampled New gTLDs. In selecting the subsample of 10,000 domains that would be analyzed, the goal was to have 800 in each cell of the Region by RAA Type (Table 2). A sample size of 400 gives a confidence interval no larger than ± 5 percent for a binary variable (this maximum of ± 5 percent occurs with an estimate of 50 percent; confidence intervals get smaller as the estimate gets further away from 50 percent). However, a sample size of 400 was not chosen as a goal because the differential weighting does increase the size of the confidence intervals. 800 was chosen as the goal in order to minimize the size of the confidence intervals in each cell.

Six of the RAA by region cells have less than 800³⁵. These are the three Africa cells and the 2009 RAA cells for Latin America and Caribbean, Europe, and Asia Pacific. We kept all domains from these six cells in the 10,000 subsample for analysis. We also oversampled the other cells to obtain 800 domains from each cell with less than 11,000 domains in the initial sample. These are the 2013 RAA cells for Latin America and Caribbean and Europe, plus the 2013 RAA GF cell for Asia Pacific and the 2009 RAA cell for North America. We sampled the Region Unknown 2009 RAA cell at the same proportion as the 2009 RAA North America cell. For the remaining five cells (including the Region Unknown 2013 cells), we selected sample sizes proportional to size for each. Table 2 below shows the sizes of the analyzed subsample by Region and RAA.

Table 2: Analyzed Subsample Sizes by Region and RAA

RAA Type	Africa	Latin America and Caribbean	Europe	Asia Pacific	North America	Unknown	TOTAL
2009	23	203	258	484	800	13	1,781
2013 GF	310	800	800	800	1,643	17	4,371
2013 NGF	376	800	800	833	1,022	17	3,849
TOTAL	709	1,803	1,858	2,117	3,465	47	10,000

Records in the 2009 RAA subgroup accounted for 2.7 percent of all initial sample records, while 2013 grandfathered (2013 GF)³⁶ and 2013 non-grandfathered (2013 NGF) records accounted for 50.8 percent and 46.4 percent of all domains respectively. Because of the small percentage of domains still registered to registrars under the 2009 RAAs, the analyzed subsample contains a disproportionately large subsample of these domains so that estimates related to 2009 RAA domains would meet the reliability

³⁵ Every "Unknown" region cell was sampled at the same rate/percentage as the North America cell.

³⁶ A 2013 "grandfathered" domain is one that was registered with a registrar under the 2013 RAA but who registered prior to a registrar signing the 2013 RAA, e.g., example.exampleTLD registered in January 2010, and the Example Registrar subsequently signed the 2013 RAA. These domains are thus held not to the requirements of the 2013 RAA but to the 2009 RAA for WHOIS accuracy.

criteria described above. Table 3 below shows the sample sizes by RAA type in the initial sample of 98,084 and the analyzed subsample of 10,000.

Table 3: Sample Sizes by RAA Type

RAA Type	Percentage of All Domains	Initial Sample	Analysis Subsample
2009 RAA	3.5%	2,683	1,781
2013 RAA GF	65.7%	49,848	4,371
2013 RAA NGF	30.8%	45,553	3,848
TOTAL	100.0%	98,084	10,000

Syntax Testing Methods

Syntax Testing was designed to assess the contact information of a record by comparing it to formats specified by contractual requirements stipulated in the Registrar Accreditation Agreements (RAAs).

Currently, there are two versions of the RAA in use in the gTLD space, the 2009 version (2009 RAA) and the 2013 version (2013 RAA). Each version of the RAA has requirements for presence, format and operability of specific modes of contact information for the Registrant, Technical, and Administrative contact for each domain name. The 2013 RAA is more specific in its requirements with respect to the contact data in a WHOIS record than the 2009 RAA. For example, the 2009 RAA requires postal addresses with a valid country, whereas the 2013 RAA requires the country in the address to be formatted per the 2-letter code from ISO-3166-1. The syntax accuracy criteria were designed in such a way that all records in the analyzed subset would be evaluated against the requirements of the 2009 RAA. Additional testing to the 2013 RAA requirements was applied to only those records where the 2013 RAA is the applicable agreement, which are those records with a 2013 RAA registrar that were created after 1 January 2014 (the effective date of the WHOIS Accuracy Program Specification of the 2013 RAA for registrars that signed before 31 December 2014) or when the registrar executed the 2013 RAA³⁷.

Syntax tests were performed on all nine individual contact information fields in a record. The fields were categorized by the contact *type*, and by contact *mode*. Contact *type* refers to the designation of the contact information as belonging to the Registrant, Technical (Tech), or the Administrative (Admin) contact, and contact *mode* refers to the communication medium of email address, telephone number, or postal address. Syntax tests were administered via two stages of testing, and the criteria of the tests were specific to each of the three contact modes. Stage one testing verified the presence of contact information, as required by applicable RAA, and stage two involved detailed technical testing of the syntax. The result of each test was recorded as a binary “Yes” or “No” to indicate whether the contact information field met the criteria of the test. The stage one and stage two tests for each contact mode are described in brief, below³⁸.

³⁷ ICANN acknowledges there are other events that would trigger a domain created prior to the effective date of its registrars 2013 RAA to be considered a non-grandfathered record, but to eliminate the possibility of false positives this is the standard used for the ARS.

³⁸ More detailed descriptions of the syntax accuracy tests can be found in [Appendix A: Syntax Accuracy Criteria](#).

Email Address Testing

In the 2009 RAA, the presence of an Admin and a Tech email address is required, while the presence of a Registrant email address is optional.³⁹ In the 2013 RAA, the Admin, Tech, and Registrant email addresses are each required to be present.

- **Stage One Email Address Testing:** Stage one testing looked for the presence of an email address in each of the three email address contact fields. A "No" response for any of the stage one tests, except for an omitted Registrant email address subject to the 2009 RAA requirements, was considered an error for that specific contact field. A missing Registrant email address subject to the 2009 RAA would be noted, but not counted against the domain/registrar (i.e., would not be considered inaccurate). A "Yes" response initiated stage two testing.
- **Stage Two Email Testing:** Syntax tests in stage two were performed on all contact fields that attained a "Yes" from Stage One, including the Registrant email address under the 2009 RAA. Although the Registrant email address under the 2009 RAA is not required, if it is present in the WHOIS output, it must meet applicable RAA requirements. Examples of stage two tests, include testing for:
 - presence of an "@" symbol
 - presence of local, domain, and TLD components (e.g., local@domain.TLD)

Telephone Number Testing

In the 2009 RAA, presence of Admin and Tech telephone numbers is required, and presence of a Registrant telephone number is optional.⁴⁰ In the 2013 RAA, the Admin, Tech, and Registrant telephone numbers are each required to be present.

- **Stage One Telephone Number Testing:** Stage one testing looked for the presence of a telephone number in each of the three telephone number contact fields. A "No" response for any of these tests, except for an omitted Registrant telephone number subject to the 2009 RAA requirements, was considered an error for that contact field. A missing Registrant telephone number subject to the 2009 RAA would be noted, but not counted against the domain/registrar (i.e., would not be considered inaccurate). A "Yes" response initiated stage two testing.
- **Stage Two Telephone Number Testing:** The syntax tests in stage two were performed on all contact fields that attained a "Yes" from stage one above, including the Registrant telephone number under the 2009 RAA. Although the Registrant telephone number under the 2009 RAA is not required, if it is present in the WHOIS output, it must be valid/accurate. If "No" for any of these tests, it was considered an error for that contact field. IDENTIFIERS indicate questions that will determine if tests following the identifier are applicable, so negative answers to IDENTIFIERS do not determine pass/fail. Examples of stage two tests, include testing for:

³⁹ However, if the email address is present, it will undergo Syntax Accuracy Testing, as explained further in this report.

⁴⁰ Similar as with the email address for the Registrant under the 2009 RAA, while not required to be present, if the telephone number is present in the data, it will undergo Syntax Validation.

- ❑ presence of a country code
- ❑ permissible character types

Postal Address Testing

In the 2009 RAA and 2013 RAA, presence of a Registrant, Admin and a Tech postal address is required.

- **Stage One Postal Address Testing:** Stage one testing looked for the presence of a postal address in each of the postal address contact fields. A "No" response for any of these tests was considered an error for that contact field. A "Yes" response initiated Stage Two testing.
- **Stage Two Postal Address Testing:** The syntax tests in stage two were performed on all contact fields that attained a "Yes" from stage one testing. A "No" response for any of these tests was considered an error for that contact field. Examples of stage two tests, include testing for:
 - ❑ presence of a country, whether full country name or an [ISO 3166-1](#) abbreviation
 - ❑ presence of state/province, city, and/or street, if required by the country addressing system per the UPU S42.

Notes on the Analysis of Syntax Accuracy Testing Results

The results of Syntax Testing are analyzed in the section Main Findings. Here, we discuss some of the key points regarding the approach to our analysis.

Firstly, the 2009 RAA was chosen as a baseline against which all 10,000 of the analyzed subsample records were analyzed. The 2013 RAA requirements are stricter than the 2009 Requirements, building off of, and thus encompassing, the 2009 requirements. For example, the 2009 RAA requires an address for each contact, while the 2013 RAA requires the address for each contact to be formatted per the applicable UPU S42 template for a particular country, with each component of the address (street, city, state/province, country/territory name) parsed into its own field in the output.⁴¹ It follows then that any contact field that meets the 2013 RAA requirements would also meet 2009 requirements, and for this reason, the 2009 requirements serve as a baseline off which all records can be compared.

Secondly, all of the 2009 RAA records from the initial sample were included in the analyzed subsample so that estimates would be as reliable as possible. Records in the 2009 RAA subgroup represent an estimated 3 percent of domains. In the analyzed subsample, the 2009 subgroup consisted of 1,781 records. As a smaller subgroup, Syntax Testing data allowed for analysis of 2009 RAA accuracy rates by contact type, and provided insight into the typical reasons that a contact field might fail to meet 2009 RAA syntax requirements.

Thirdly, because the requirements of the 2013 RAA are more detailed than those of the 2009 RAA, our analysis ensures these additional requirements are considered and that the results are appropriately compared to 2009 RAA results. That is, 2013 records were compared to the baseline of 2009 RAA requirements to create estimates of syntax accuracy by gTLD type and region. It should also be noted

⁴¹ For more detail on these requirements, please see [Appendix A: Syntax Accuracy Criteria](#).

that records in the 2013 RAA subgroup represent an estimated 31 percent of domains, and in the analyzed subsample the 2013 subgroup consisted of 3,848 of these records (38 percent) due to the oversampling. Domains registered under the 2013 Grandfathered RAA represent an estimated 66 percent of domains. In the analyzed subsample, the 2013 subgroup consisted of only 4,371 records (44 percent) due to oversampling. The requirements of the 2013 Grandfathered⁴² RAA for WHOIS ARS purposes are equivalent to those under the 2009 RAA.

Finally, syntax accuracy can be analyzed across any of the nine individual contact information fields, and the fields can be grouped by contact type (Registrant, Tech, Admin), by contact mode (email address, telephone number, postal address), or grouped all together to analyze the overall accuracy of a record. All of the calculations in this report can be extrapolated to the entire population of 157 million domains, or to any of the analyzed subgroups.

Major Reasons for Failure

Before presenting the percentages of domains that pass all the Syntax tests, we analyzed the reasons for failure. In the next section of the report, we report on the major reasons for failure, separately by contact mode (email address, telephone number, and postal address). For email addresses, no syntax errors were found except for a missing (blank) email address. However, for telephone numbers and especially postal addresses (which encompasses multiple fields), multiple errors were possible. For telephone numbers, a singular reason for failure could be identified, while for postal addresses, multiple reasons for failure were common.

We provide separate tables reporting the major reasons for failure against the 2009 RAA requirements among all 10,000 analyzed domains and the major reasons for failure against the 2013 RAA⁴³ requirements among the 3,848 2013 NGF subgroup.

Rates of Syntax Accuracy by Contact Type, Contact Mode, and Subgroup

In the next section, we also analyze the syntax accuracy separately by each of the nine individual contact information fields. Each number in these tables refers to the percentage of domains that pass all syntax tests for that contact type and mode. We also include an additional column on the right to indicate the percentage of domains that pass all syntax tests for all three contact modes for that contact type (all email address, telephone number, and postal address tests for the registrant contacts, for example). Similarly, the bottom row for each table indicates the percentage of domains that pass all syntax tests for all three contact types for that contact mode (e.g., all telephone number syntax tests for the registrant, administrative, and technical contacts). Finally, the bottom right cell of each table refers to the percentage of domains that pass all syntax tests for all nine contact information fields.

Just as we did for the Major Reasons for Failure, we first present tables for all 10,000 analyzed domains against the 2009 RAA syntax tests. We also present tables for the 3,848 2013 RAA NGF subgroup against the 2013 RAA syntax tests in [Appendix C](#). For complete information, we do also present the percentages of 2009 RAA and 2013 RAA GF subgroups that pass all the 2013 RAA syntax tests. However, the postal

⁴² See note 35.

⁴³ As mentioned earlier in this section, comparisons to 2013 RAA are included in [Appendix C](#) of this report.

address syntax tests were not carried out for these groups and it is important to remember that these subgroups are not required to fulfill all the requirements of the 2013 RAA.

Finally, one of our main findings is that there is not much difference in the percentages within the same contact mode (e.g., postal address) for who passes all syntax tests for each of the three contact types. In other words, the equivalent percentages are similar for registrant, administrative, and technical contacts. Therefore, we have some of the full tables in an [Appendix B](#), and only include the bottom rows (all three contact types) in the tables within the main body of the report, which allows the gTLD types, regions, and RAA types to be in one table.

Main Findings

Here we present our findings and the statistics related to the outcomes of the syntax accuracy tests. These statistics are organized by contact type (Registrant, Tech, and Admin), contact mode (email address, telephone number, and postal address), across subgroups such as New vs. Prior gTLD as well as ICANN Region, and, finally, RAA requirements. Because the 2009 and 2013 RAA versions have different requirements for valid syntax, we created separate analysis tables for each set of requirements (2009 and 2013), with the 2009 requirements serving as a baseline, as explained [above](#). Finally, the causes of syntax nonconformance are examined by analyzing Syntax Testing outcomes at the level of the binary syntax test (e.g., testing for an “@” symbol in an email address). Additional findings and analysis tables can be found in [Appendix B](#) as well as [Appendix C](#) of this report⁴⁴. We first present a summary of our findings, followed by more detailed statistical analysis.

Summary of Findings

We present here the key takeaways from the findings:

- The results of Syntax Testing showed that 99 percent of email addresses, 85 percent of telephone numbers and 79 percent of postal addresses met all syntax requirements of the 2009 RAA. 70 percent of domains passed all syntax tests for all contact types (Registrant, Administrative, Technical) and contact modes (email address, telephone number, postal address).
- The contact mode with the highest rate of passing all syntax tests was email address, and the mode with the lowest rate of passing all syntax tests was postal address.
 - If an email is provided, it always passed all syntax accuracy tests
 - Two-thirds of the telephone numbers that failed at least one syntax accuracy check (13 percent of all telephone numbers) failed the length criteria for the applicable country
 - Postal addresses that failed at least one syntax accuracy check (23 percent of all postal addresses) were typically missing at least one required component such as postal code, state/province, city, or street
- Accuracy rates among the three contact types are all similar, i.e., it is unlikely that an individual filling in contact information for all three contact types will make different/more or less mistakes on each
 - If a contact type has one contact mode that failed a syntax test, it is much more likely the same mode failed a syntax test in the other two contact types.

⁴⁴ In the interest of condensing the findings in this section, many of the analysis tables discussed herein are stored in the [Appendix B](#) and [Appendix C](#) of the report.

Overall Conformance to 2009 RAA Requirements⁴⁵

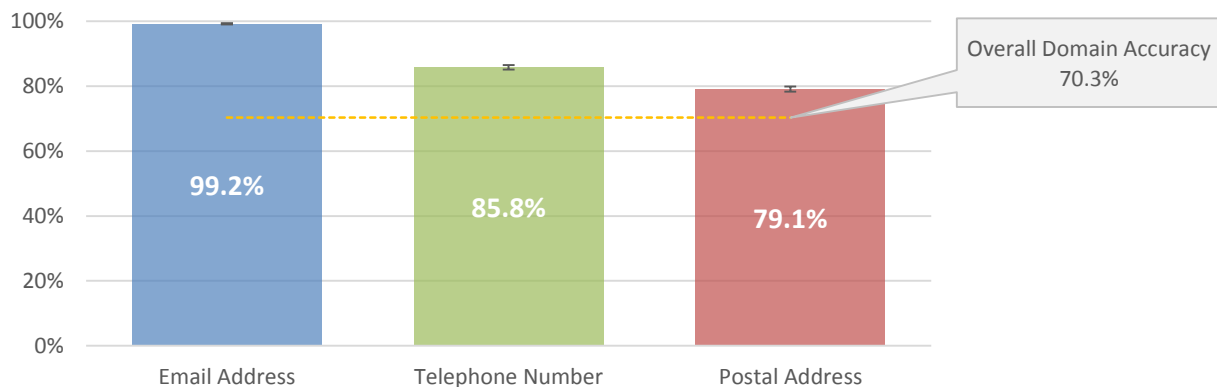
Only the 2013 RAA NGF domains are required to meet the syntax accuracy standards of the 2013 RAA⁴⁶. However, *all* domains are required to meet the syntax accuracy standards of the 2009 RAA. Therefore, we compared all 10,000 domains from the analyzed subsample against the 2009 RAA standards. We also analyzed only the 2013 RAA NGF domains against 2013 RAA standards, this data is available in [Appendix C](#).

We do not combine all 10,000 domains in the analyzed subsample using different accuracy standards into one table. Instead, we have tables that analyze all 10,000 domains under 2009 RAA standards, and other tables that analyze only the 3,848 2013 NGF domains subject to the 2013 RAA standards.

As described above, all 10,000 domains in our analyzed subsample are required to meet the standards of the 2009 RAA, and we describe now what (weighted) percentage of the analysis dataset passes all of the 2009 RAA syntax accuracy tests. Each estimate applies to the population of all gTLD domains (157M), or the appropriate subgroup since we also provide separate estimates by gTLD type, region, and RAA type further on in this section.

As is evident in Graph 1 below, the overall conformance of all gTLD WHOIS records to the criteria of the 2009 RAA among contact modes is highest for email address and lowest for postal address. Nearly 100 percent of all domains had full syntax accuracy for the email addresses of all three contact types.

Graph 1: Summary of all gTLD WHOIS Record Conformance to 2009 RAA Syntax Requirements



⁴⁵ Overall conformance to 2013 RAA Requirements can be found in [Appendix C](#).

⁴⁶ 2013GF records are required to be “upgraded” to the 2013 in the event the registrar becomes aware of an inaccuracy in the data or if the name is transferred to a new registrar or registrant.

In Table 4 below⁴⁷, we can see more clearly the breakdown of overall conformance by both contact mode and type. Even though the 2009 RAA does not require a registrant e-mail address or telephone number, they are required to be correct if provided. In these tables, the bottom row indicates whether all three contact types for the same contact mode pass all of the syntax tests. Because the accuracy rate of “All 3” does not differ greatly from the individual registrant, administrative, and technical success rates, the success of these three contacts can said to be highly correlated; if one fails a syntax check, the other two are much more likely to also fail a syntax check. Because of this, subsequent graphs and tables will only show the “All 3 Accurate” (bottom) row. Complete tables are provided in [Appendix B](#) for 2009 requirements and [Appendix C](#) for 2013 RAA requirements.

Table 4: Summary of all gTLD WHOIS Record Conformance to 2009 RAA Syntax Requirements by Contact Type and Mode

	E-mail	Telephone	Postal Address	ALL 3 Accurate
Registrant	100.0% ± 0.0%	88.0% ± 0.6%	81.5% ± 0.8%	73.0% ± 0.9%
Administrative	99.3% ± 0.2%	86.9% ± 0.7%	81.0% ± 0.8%	72.8% ± 0.9%
Technical	99.2% ± 0.2%	87.5% ± 0.6%	82.4% ± 0.7%	74.5% ± 0.9%
All 3 Accurate	99.2% ± 0.2%	85.8% ± 0.7%	79.1% ± 0.8%	70.3% ± 0.9%

***How to read this table:** The rightmost column indicates whether all three modes for each contact pass all of the syntax tests. This rightmost column has lower numbers than any other column, which indicates the correlation for email address, telephone number, and postal address syntax successes is less than among the three contacts within the same mode. Finally, the bottom right number indicates the percentage of domains for which every syntax test is passed for all nine contacts. For around 70 percent of the domains, every syntax test passed.*

As mentioned above and is evident from the table and graph, email address syntax accuracy is highest and postal address syntax accuracy is lowest⁴⁸, and the differences between the contact mode syntax accuracy percentages are statistically significant. Within a contact mode, there are no statistically significant⁴⁹ differences between registrant, administrative, and technical contact syntax accuracy; these rates are similar enough that any differences are explainable by sampling error. In a later section of the report, we will analyze the most frequent nonconformances for the contact modes.

Analysis by Subgroup: Conformance to 2009 RAA requirements

We now describe whether there are accuracy differences by subgroup when using the 2009 RAA Requirements. Starting with gTLD type, as seen in the Table 5 below, the overall syntax accuracy (all 3 accurate) is slightly higher in Prior gTLDs than New gTLDs, but this difference is not statistically

⁴⁷ Each entry in the table is the 95 percent confidence interval for the weighted data.

⁴⁸ Here “highest” and “lowest” refer not only to sheer numbers, but also statistical significance. This phrase has been left out of most of the narrative for ease of reading. Where there is no statistical significance, it will be mentioned.

⁴⁹ Statistically significant differences can be generalized to the entire sample.

significant. Interestingly New gTLDs scored higher for email addresses, but Prior gTLDs scored higher for postal addresses. The telephone rates are almost the same.

Table 5 Prior versus New gTLDs: Syntax Accuracy using 2009 RAA Requirements for all contact types combined

	E-mail	Telephone	Postal Address	ALL 3 Accurate
Prior gTLD	99.2% ± 0.2%	85.8% ± 0.8%	79.3% ± 0.9%	70.4% ± 1.0%
New gTLD	99.8% ± 0.2%	86.2% ± 1.5%	71.6% ± 2.0%	68.8% ± 2.0%

Grouping by region in Table 6 below, we see larger variances in accuracy rates across the regions for both telephone numbers and postal addresses. Email addresses are the least accurate for Asia-Pacific and North America, but telephone numbers are the most accurate in North America and Europe. Domains in Africa scored lowest by far for both Telephone Numbers and Postal Address Syntax Accuracy.

Table 6 ICANN Region Syntax Accuracy for all contact types combined compared to 2009 RAA Requirements

ICANN Region	E-mail	Telephone	Postal Address	All 3 Accurate
Africa	100% ± 0%	70.3% ± 3.4%	47.2% ± 3.7%	33.7% ± 3.5%
Asia Pacific	99.6% ± 0.3%	80.4% ± 1.7%	55.3% ± 2.1%	48.7% ± 2.1%
Europe	100% ± 0%	87.5% ± 1.5%	66.0% ± 2.2%	58.6% ± 2.2%
Latin America and Caribbean	100% ± 0%	82.4% ± 1.8%	67.1% ± 2.2%	59.2% ± 2.3%
North America	98.7% ± 0.4%	87.7% ± 1.1%	94.3% ± 0.8%	84.3% ± 1.2%

Graph 2 below shows the overall accuracy of records grouped by RAA type when compared to the 2009 RAA Requirements. Table 7 below then details the accuracy rates for the 3 RAA types by contact mode. Looking at accuracy by RAA type, 2009 RAA domains showed higher syntax accuracy than the other RAA domains at the level of the entire record, and also for the contact modes of telephone number and postal address. Grandfathered 2013 RAA domains had the lowest accuracy for telephone numbers, and non-grandfathered 2013 RAA domains had the lowest accuracy for postal addresses.

Graph 2: Overall Syntax Accuracy by RAA Type using 2009 RAA requirements

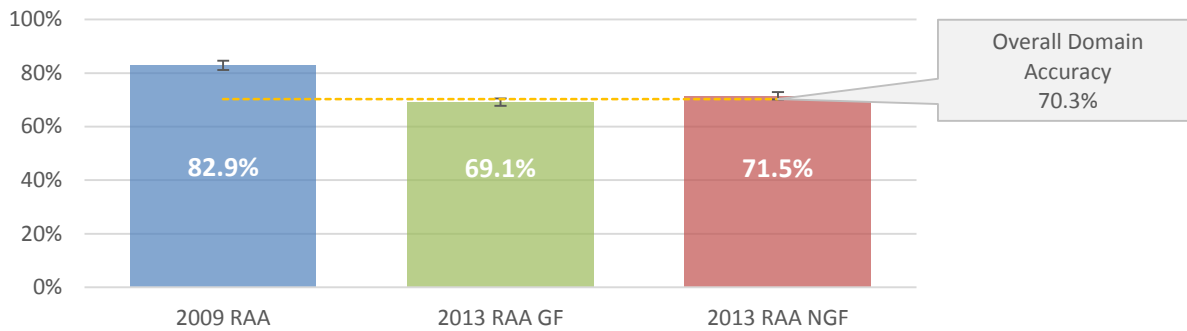


Table 7: RAA Type Syntax Accuracy for all contact modes compared to 2009 RAA Requirements

RAA Type	E-mail*	Telephone	Postal Address	ALL 3 Accurate
2009 RAA	99.1% ± 0.4%	92.5% ± 1.2%	85.8% ± 1.6%	82.9% ± 1.7%
2013 RAA GF	99.4% ± 0.2%	83.8% ± 1.1%	80.8% ± 1.2%	69.1% ± 1.4%
2013 RAA NGF	98.9% ± 0.3%	89.3% ± 1.0%	74.7% ± 1.4%	71.5% ± 1.4%

* Differences on email address syntax accuracy are not statistically significant.

2009 RAA Reasons for Syntax Error

Now that we have seen conformance to 2009 RAA standards, we analyze at the level of the accuracy test (for domains under the 2009 RAA), to determine the most common causes of syntax error.

Email Address Errors

Using the 2009 RAA standards, a registrant email address is *not* required, but if it is present, it should be valid. Valid in syntax includes having an “@” character, having a domain, having a local part, having only permissible characters, and having the correct format overall. Table 8 below shows by contact type the total number of email addresses that passed.

Table 8: Total Email Address Errors by Contact Type (2009)

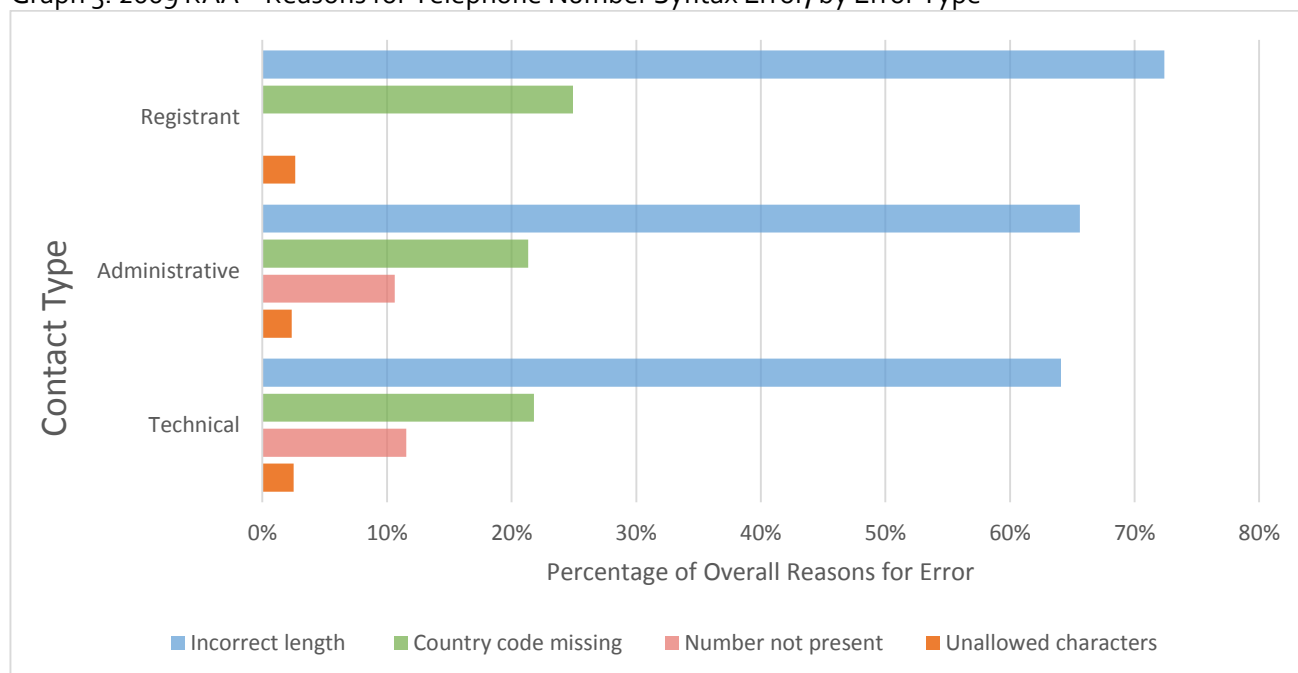
	Administrative	Technical	Registrant	All 3 total
Passed all accuracy tests	9,950	9,954	10,000	29,904
Missing*	50	46	[38]*	96
Total	10,000	10,000	10,000	30,000

No non-blank email addresses failed any of the formatting accuracy tests. The 38 blank registrant e-mail addresses essentially passed the accuracy check since presence of the registrant e-mail address is not required.

Telephone Number Errors

Using the 2009 RAA standards, a registrant telephone number is also *not* required, but if it is present in the data, it should be valid. Accurate syntax in this case means having: the correct length, no unallowed characters, a country code, and, if present, a correctly formatted extension. Graph 3 below shows the total number and type of errors by contact type (further detail on the reasons for telephone error can be found in [Appendix B](#), Table B.1 and Graphs B.1 through B3).

Graph 3: 2009 RAA – Reasons for Telephone Number Syntax Error, by Error Type



The most common error among all contact types was incorrect phone number length, meaning that a telephone number had an unacceptable number of digits (too many or too few) based on the country code. The second most common error was due to an absence of the country code. For technical and administrative contact types, the third most common error was the absence of a phone number. However, according to the regulations of the 2009 RAA, Registrant phone numbers are not required and so while 2.3 percent were missing, this did not count as a syntax error.

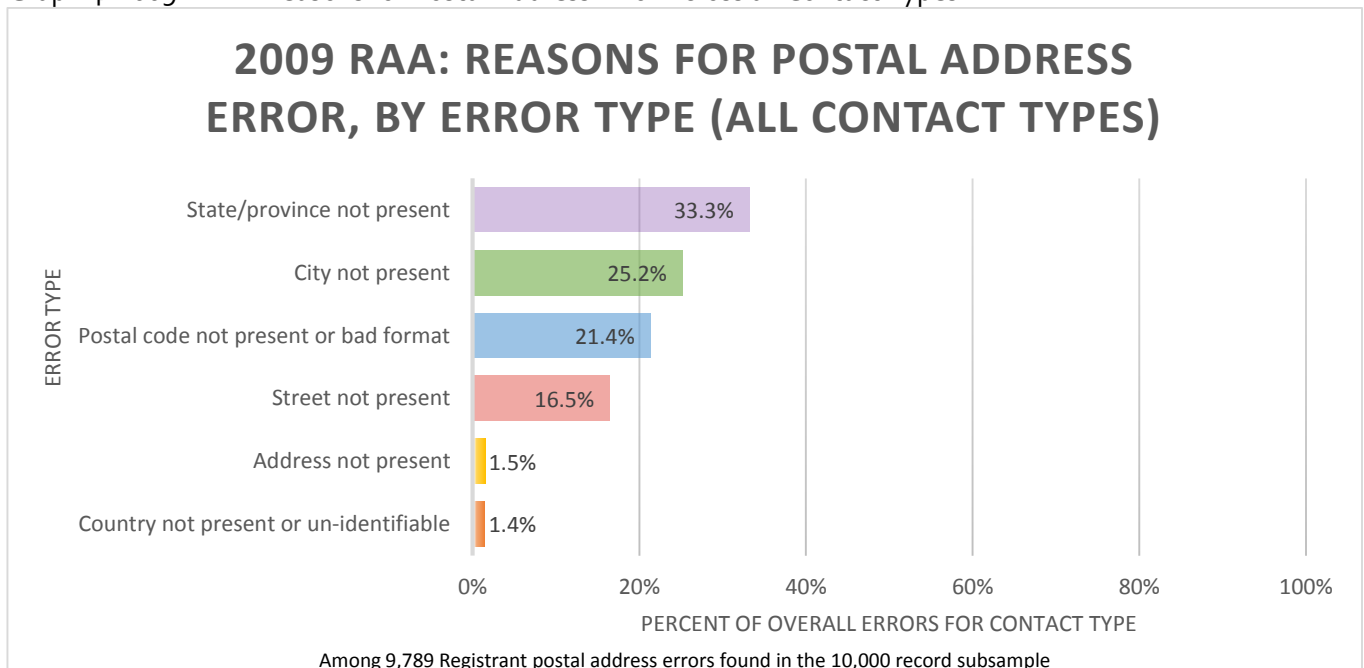
Approximately 87 percent of each of the telephone number contact types passed all 2009 RAA syntax accuracy tests. We were able to assign each domain to one key error that caused the overall syntax test failure. However, some of the domains with unallowed characters overlap with other categories. The "Incorrect Length" category can occur if there are two or more valid lengths (6 or 8, for example), but the

telephone number is between these valid lengths (i.e., not too long or too short, but not correct). It can be seen from the table that two-thirds of the errors are due to incorrect length.

Postal Address Errors – 2009 RAA

Using the 2009 RAA standards, all three postal addresses *are* required. We report here on the errors against the 2009 standards, which require that a country that is identifiable, a postal code is in the right format and that is present in a field, as well as a street, city, and state that are present in a field. Graph 4 below shows the types and percentages of errors found across all contact types (further detail on the reasons for postal address error can be found in [Appendix B](#), Table B.2 and Graphs B.4 through B6).

Graph 4: 2009 RAA – Reasons for Postal Address Error Across all Contact Types



Approximately 75 percent of each of the postal address contact types passed all 2009 RAA syntax accuracy tests, though the Technical postal address success rate was slightly higher at 78 percent. Often, multiple errors occurred within the same contact. Unlike for telephone numbers, it was not possible to isolate one key error. Therefore, the total number of errors are larger than the number of domains with errors. There is almost no overlap, or an occurrence of both errors within the same record, between the missing country code errors and the postal code errors. However, there is overlap among the state/province, city, and street missing errors, as well as between these and the other errors. It can be seen from the graph that almost all the domains with errors were missing at least one of the postal code, state (or province), city, or street. Less than 100 were missing the country code.

Challenges and Lessons Learned

The Phase 1 study was successful as the first “cycle” of the ARS—the study accomplished its goals of determining syntax accuracy of WHOIS records and providing this information to ICANN Contractual Compliance for follow-up. From this data as well as the repetition of the ARS cycles, including follow up by ICANN Contractual Compliance, ICANN expects the overall WHOIS accuracy to improve over time. At this time, however, it is difficult to make firm conclusions about what changes can be made to the WHOIS to improve accuracy. With time we hope to pinpoint more clear implications about the trends in accuracy of the WHOIS and what those mean for future improvement of the WHOIS records.

We would instead like to focus on the challenges and lessons learned with the Phase 1 study to help with learning for future cycles of the ARS:

- **Six months are required to conduct an entire cycle:** Because Phase 1 only included Syntax accuracy testing, it could be shorter, but future cycles need the full six months in order to adequately complete the testing, analysis and report development.
- **A larger sample can help to increase the significance in some instances:** While an initial sample of 100,000 records (domains) is sufficient for statistical significance in most areas of the data, more records from which to draw the sample allow for greater statistical significance for some types of data (e.g., regional). The first cycle for Phase 2 has taken an initial sample of approximately 150,000 records; 10,000 records will again be used for the analyzed subsample size.
- **Phase 2 initial samples will exclude gTLDs with exactly one domain:** Almost two percent of the records in the initial sample were removed from the dataset when it was found that records of gTLDs with exactly one domain were test records or domains that no longer existed. In Phase 2, the initial sample will exclude these gTLDs with exactly one domain so that fewer of the sample records are discarded from the dataset.
- **All records should be tested against all RAA obligation types:** Postal addresses from records on the 2009 RAA and 2013GF RAA were not tested per the 2013 RAA requirements. While such tests are not required, doing so provides more insight into the overall picture of WHOIS accuracy and helps with noting trends in the data.
- **In general, postal address syntax testing is difficult:** While the rules for postal address formatting are relatively straightforward, the country requirements for address data are much more complex (less uniform), complicating accuracy tests and requiring many manual tests (unlike email address and telephone number testing, which is automated). For example, the rules vary widely across countries as to how a state or province should be abbreviated and where it should appear in the data; further, within countries the rules for formatting might be largely ignored, but mail would still be considered formatted adequately (and deliverable). Likewise, the way in which data can be input into the WHOIS record has a great effect on the accuracy of the data (i.e., more options for data entry and presentation in the WHOIS might increase accuracy). Future cycles of the ARS will attempt to automate syntax and operability tests through use of

more specialized tools, but without changes to the structure of WHOIS input options, syntax accuracy may not be improved.⁵⁰

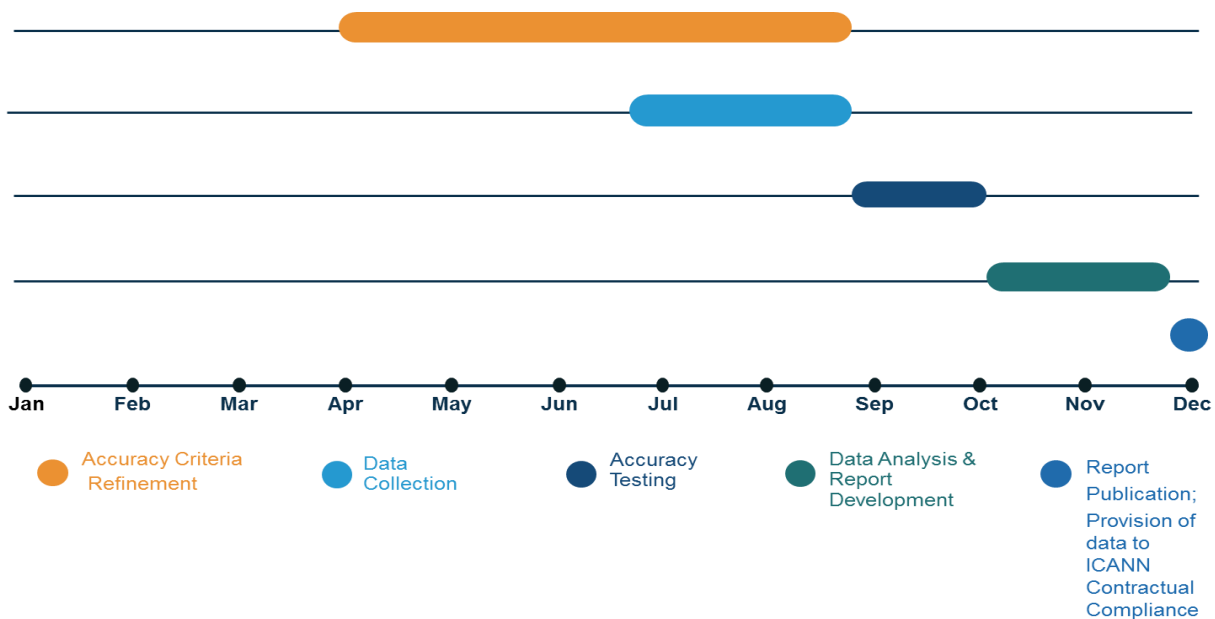
⁵⁰ It is interesting to note that an opposite correlation exists between postal address syntax and operability and email address/telephone number syntax and operability. Where syntax accuracy is an indicator of operability for email address and telephone numbers, it is not for postal addresses. That is, a syntactically inaccurate email address (e.g., without the “@”) will not be operable; a syntactically inaccurate postal address (e.g., where the full name of a country has been entered and not the two digit international code or where one digit is missing from a postal code but every other portion of the postal address is correct) will likely be operable (i.e., deliverable). It is likely that while syntax accuracy was high for email addresses, it could decrease in Phase 2, and postal addresses, low in accuracy in Phase 1, could increase in accuracy in Phase 2.

Next Steps

Phase 2 Look-Ahead

In late 2nd quarter 2015, ICANN also began work on Phase 2 of the ARS, which includes both Syntax and Operability validation testing. A Phase 2 report is targeted for publication in December 2015. Figure 4 below shows the timeline for Phase 2:

Figure 5: Phase 2 Timeline



ICANN Contractual Compliance Next Steps

As indicated above, one of the major goals of the ARS project is the ability to pass to ICANN Contractual Compliance potential inaccuracies for follow-up and investigation with the Registrars. The results of the Phase 1 study and those potentially inaccurate records have already been provided to ICANN Contractual Compliance and they have begun their review before proceeding to create tickets in the complaint processing system.

ICANN Contractual Compliance approach for Phase 1 follow-up is highly influenced by lessons learned from the Pilot Study. Specifically, because creation of thousands of complaints burdened complaint processing system, improvements to the WHOIS inaccuracy processing script were made (which also increased over system performance). The creation of new tickets based upon WHOIS ARS results will be staggered to minimize system performance issues. WHOIS ARS tickets will be processed alongside other complaints, however ICANN will continue to give priority to give priority to complaints submitted by

community members. WHOIS ARS complaints will be classified as WHOIS format if the error does not comply with the format requirements of the 2013 RAA but the information is otherwise valid and contactable (e.g. a missing +1 county code for a registrant located in the United States). Where the error renders the contact unreachable (e.g. a missing postal address), the WHOIS ARS complaint will be processed as a WHOIS inaccuracy complaint. All WHOIS ARS tickets will follow the Contractual Compliance Approach and Process⁵¹. When possible, and in consultation with registrars, ICANN may be able to consolidate multiple WHOIS ARS tickets during processing.

When ICANN Contractual Compliance sends notices to registrars for WHOIS ARS tickets, the following contractual requirements apply:

- Registrars must investigate and correct inaccurate WHOIS data per:
 - Section 3.7.8 of 2009 and 2013 RAA and
 - WHOIS Accuracy Program Specification of 2013 RAA
- Registrars under 2013 RAA must use WHOIS format and layout required by the Registration Data Directory Service (WHOIS) Specification⁵²

In addition, as with any WHOIS inaccuracy or WHOIS format complaint, failure to respond or demonstrate conformance during the processing of the complaint will result in a Notice of Breach (which will be published on icann.org).

⁵¹ ICANN Contractual Compliance Approach and Process: <https://www.icann.org/resources/pages/approach-processes-2012-02-25-en>.

⁵² <https://www.icann.org/resources/pages/approved-with-specs-2013-09-17-en#whois>.

Appendix A: Syntax Accuracy Criteria

Accuracy Criteria and the RAA

ICANN has attempted to align the syntax accuracy criteria with the contractual requirements of the Registrar Accreditation Agreements (RAAs) and applicable Internet Engineering Task Force (IETF) RFCs. Currently, there are two predominant versions of the RAA in use in the gTLD space: the 2009 version (2009 RAA) and the 2013 version (2013 RAA). Each version of the RAA has requirements for presence, format and operability of specific modes of contact information for the Registrant, the Technical Contact (Tech) and the Administrative Contact (Admin) for each domain name. Each record (i.e., domain name) will be assessed against the criteria of the Registrar's agreement at the time the domain was created. ICANN will account for "grandfathered" records, which are those records (domains) that were created prior 1 January 2014 (the effective date of the WHOIS Accuracy Program Specification for registrars that signed the 2013 RAA before 31 December 2013) or to the effective date of the 2013 RAA for that Registrar. For example:

Record Created	05 Feb 2013
Registrar's 2013 RAA Effective Date	01 Jan 2014
Syntax Accuracy Criteria to be used in testing	2009 RAA Requirements

Record Created	20 Apr 2014
Registrar's 2013 RAA Effective Date	01 Jan 2014
Syntax Accuracy Criteria to be used in testing	2013 RAA Requirements

Below you will find an overview of Phase 1 syntax accuracy criteria for email addresses, telephone numbers and postal addresses to be used by the WHOIS ARS team.

Email Addresses

As identified in the WHOIS ARS Pilot Study⁵³, correct syntax and a verified email addresses do not guarantee email box existence, so while email addresses with incorrect syntax may indicate automatic failures, email addresses with correct syntax should be subject to operability verification in Phase 2.

⁵³ <https://www.icann.org/public-comments/whois-ars-pilot-2014-12-23-en>.

The syntax accuracy tests for email addresses are organized into stages, stage one will verify the presence of an email address, as required by the applicable RAA, and stage two will involve detailed technical testing of the address syntax.

Email Address: Stage One

In the 2009 RAA, the presence of an Admin and a Tech email address is required. The presence of a Registrant email address is optional. In the 2013 RAA, the Admin, Tech, and Registrant email addresses are each required to be present.

A "No" response for any of these tests, except for an omitted Registrant email address subject to the 2009 RAA requirements, will be considered a failure for the contact field. A missing Registrant email address subject to the 2009 RAA will be noted, but not counted against the domain/registrant (i.e., would not be considered inaccurate). A "Yes" response will initiate Stage Two testing.

1. Is there presence of an email address?

(i.e., field is not blank)

a. Registrant email address

- Yes – Pass
- No – Fail: 2013 RAA || Pass: 2009 RAA

b. Tech email address

- Yes – Pass
- No – Fail

c. Admin email address

- Yes – Pass
- No – Fail

Email Address: Stage Two

The syntax tests in Stage Two are performed on all contact fields that attained a "Yes" from Stage One above, including the Registrant email under the 2009 RAA. Although the Registrant email under the 2009 RAA is not required, if it is present in the WHOIS output, it must be valid/accurate.

If "No" for any of these tests, it will be considered a failure for that contact field. Everything with a "Yes" will be subject to subsequent tests.

1. Does the email address only contain permissible characters?

(i.e., as provided for within the RFC 5322⁵⁴)

⁵⁴ <http://tools.ietf.org/html/rfc5322>.

-
- a. Registrant, b. Tech, and c. Admin email addresses**
 - Yes – Pass
 - No – Fail

 - 2. Is there presence of an “@” symbol in the email address?**
 - a. Registrant, b. Tech, and c. Admin email addresses**
 - Yes – Pass
 - No – Fail

 - 3. Is there presence of a domain component?**

(i.e., the characters following the “@” symbol)

 - a. Registrant, b. Tech, and c. Admin email addresses**
 - Yes – Pass
 - No – Fail

 - 4. Is the domain component in a TLD, which is resolvable on the Internet?**

(see IANA’s Root Zone Database: <http://www.iana.org/domains/root/db>)

 - a. Registrant, b. Tech, and c. Admin email addresses**
 - Yes – Pass
 - No – Fail

 - 5. Is the domain component syntactically valid?**

(i.e., the component following the “@” symbol meets requirements)

 - a. Registrant, b. Tech, and c. Admin email addresses**
 - Yes – Pass
 - No – Fail

 - 6. Is there presence of local component?**

(i.e., the characters preceding the “@” symbol)

 - a. Registrant, b. Tech, and c. Admin email addresses**
 - Yes – Pass

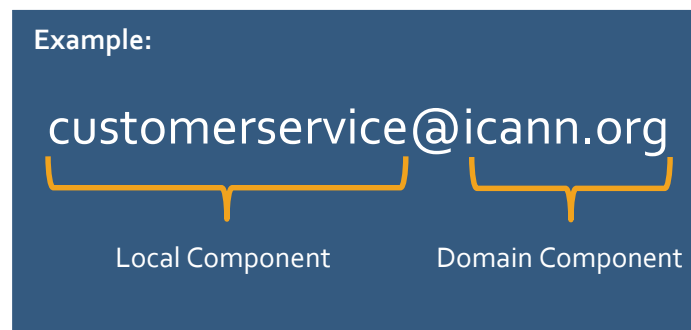
- No – Fail

7. Is the local component syntactically valid?

(i.e., the component preceding the “@” symbol meets requirement)

a. Registrant, b. Tech, and c. Admin email addresses

- Yes – Pass
- No – Fail



Telephone Numbers

As identified in the WHOIS ARS Pilot Study, syntactically correct, verified phone numbers do not guarantee existence or operability of the phone number and incorrect syntax does not guarantee the number is not in operation. All phone numbers will be subject to operability verification in Phase 2.

Telephone Numbers: Stage One

In the 2009 RAA, presence of Admin and Tech telephone numbers is required; presence of a Registrant telephone number is optional. In the 2013 RAA, the Admin, Tech, and Registrant telephone numbers are each required to be present.

A "No" response for any of these tests, except for an omitted Registrant telephone number subject to the 2009 RAA requirements, will be considered a failure for that contact field. A missing Registrant telephone number subject to the 2009 RAA will be noted, but not counted against the domain/registrant (i.e., would not be considered inaccurate). A "Yes" response will initiate Stage Two testing.

1. Is there presence of a phone number?

(i.e., field is not blank)

a. Registrant phone number

- Yes – Pass
- No – Fail: 2013 RAA || Pass: 2009 RAA

b. Tech phone number

- Yes – Pass
- No – Fail

c. Admin phone number

- Yes – Pass
- No – Fail

Telephone Numbers: Stage Two

The syntax tests in Stage Two are performed on all contact fields that attained a "Yes" from Stage One above, including the Registrant telephone under the 2009 RAA. Although the Registrant telephone number under the 2009 RAA is not required, if it is present in the WHOIS output, it must be valid/accurate.

If "No" for any of these tests, it will be considered a failure for that contact field. IDENTIFIERS indicate questions that will determine if tests following the identifier are applicable, so negative answers to IDENTIFIERS do not determine pass/fail.

1. Is there presence of a country code?

(i.e., contains a discernable country code based on the first three digits)

a. Registrant, b. Tech, and c. Admin phone number

- Yes – Pass
- No – Fail

2. Is the country code syntax valid?

(i.e., meets the requirements as specified in RFC5733⁵⁵, +###.)

a. Registrant, b. Tech, and c. Admin phone number

- Yes – Pass
- No – Fail: 2013 RAA || Pass: 2009 RAA

⁵⁵ <http://tools.ietf.org/html/rfc5733>.

-
3. Does the phone number contain at least the minimum allowed digits based on the country code?
 - a. Registrant, b. Tech, and c. Admin phone number
 - Yes – Pass
 - No – Fail

 4. Does the phone number contain, at most, the maximum allowed digits based on the country code?
 - a. Registrant, b. Tech, and c. Admin phone number
 - Yes – Pass
 - No – Fail

 5. Does the phone number contain an appropriate amount of digits based on the country code? (e.g., the number contains 7 digits when only 6 or 8 digits are acceptable based on a country code)
 - a. Registrant, b. Tech, and c. Admin phone number
 - Yes – Pass
 - No – Fail

 6. Does the phone number only contain permissible numbers and formatting characters?
 - a. Registrant, b. Tech, and c. Admin phone number
 - Yes – Pass
 - No – Fail

 7. IDENTIFIER – Is there presence of an extension?
 - a. Registrant, b. Tech, and c. Admin phone number
 - Yes – Proceed to additional extension validation
 - No – Move to next field

 8. Does the extension only contain permissible numbers and formatting characters?
 - a. Registrant, b. Tech, and c. Admin phone number
 - Yes – Pass

- No – Fail

9. Is the extension syntactically valid?

(i.e., "x" to attribute the telephone number extension: RFC5733⁵⁶)

a. Registrant, b. Tech, and c. Admin phone number

- Yes – Pass
- No – Fail: 2013 RAA || Pass: 2009 RAA

Postal Addresses

As identified in the WHOIS ARS Pilot Study, syntactically correct, verified postal addresses do not guarantee existence or operability of the postal address, so while syntactically incorrect postal addresses may indicate some failures, all postal addresses should be subject to operability verification in Phase 2.

Postal Addresses: Stage One

In the 2009 RAA and 2013 RAA, presence of a Registrant, Admin and a Tech postal address is required.

A "No" response for any of these tests, will be considered a failure for that contact field. A "Yes" response will initiate Stage Two testing.

1. Is there presence of a postal address?

(i.e., field is not blank)

a. Registrant, b. Tech, and c. Admin postal address

- Yes - Pass
- No – Fail

Postal Addresses: Stage Two

The Syntax Tests in the Stage Two are performed on all contact fields that attained a "Yes" from the Stage One above.

If "No" for any of these tests, it will be considered a failure for that contact field. IDENTIFIERS indicate questions that will determine if tests following the identifier are applicable, so negative answers to IDENTIFIERS do not determine pass/fail.

1. Is there presence of a country?

⁵⁶ <http://tools.ietf.org/html/rfc5733>.

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail

2. Is the country identifiable?

(i.e., full country name or an [ISO 3166-1 abbreviation](#))

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail

3. Is the country provided in the Country field?

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail: 2013 RAA || Pass: 2009 RAA

4. Is the country syntactically valid?

(i.e., meets [ISO 3166-1: Alpha 2-code](#) format)

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail: 2013 RAA || Pass: 2009 RAA

5. IDENTIFIER – Does the country use a postal code system?

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Proceed to additional postal code validation
- No – Appropriately left blank, move to next field (i.e., Test g)

6. Is there presence of a postal code?

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail

7. Is the postal code in the Postal Code field?

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail: 2013 RAA || Pass: 2009 RAA

8. Is the Postal Code syntactically valid based on the country?

(i.e., format of postal code meets length, alpha/numeric formats of country)

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail

9. IDENTIFIER – Does the country require states/provinces in its addressing system?

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Proceed to additional State/Provide validation
- No – Appropriately left blank, move to next field (i.e., Test 13)

10. Is there presence of a state/province?

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail

11. Is the state/province in the State/Province field?

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail: 2013 RAA || Pass: 2009 RAA

12. Is the State/Province syntactically valid?

(i.e., full name or abbreviation depending on country addressing system)

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail: 2013 RAA || Pass: 2009 RAA

13. Is there presence of a city?

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail

14. Is the city in the City field?

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail: 2013 RAA || Pass: 2009 RAA

15. Is there presence of a street?

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail

16. Is the street in the Street field?

a. Registrant, b. Tech, and c. Admin postal address

- Yes – Pass
- No – Fail: 2013 RAA || Pass: 2009 RAA

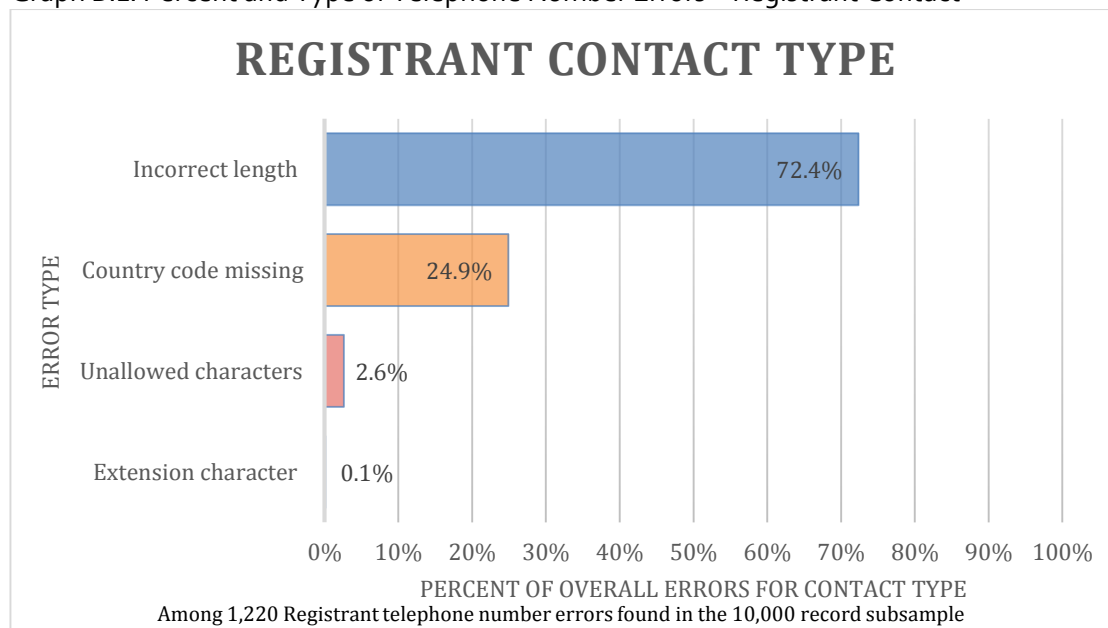
Appendix B: Accuracy Results Tables 2009 RAA Requirements

2009 RAA Requirements – Reasons for Error

Table B.1: Total and Type of Telephone Number Errors by Contact Type (2009)

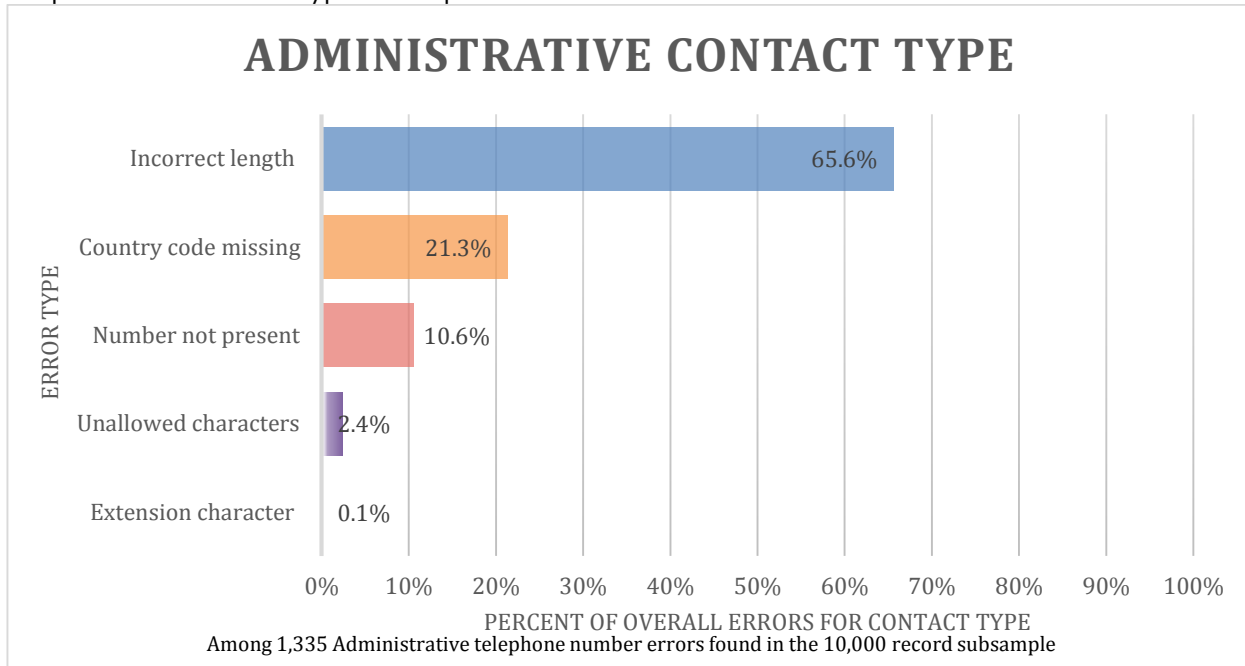
	Administrative	Technical	Registrant	ALL 3 TOTAL
Passed all accuracy tests	8,645	8,719	8,780	26,144
Not present	144	148	[234]*	292
Country code missing	289	279	304	872
Incorrect length	889	821	883	2,593
Unallowed characters	33*(49)	33*(49)	33*(54)	97*(150)
Total	10,000	10,000	10,000	30,000

Graph B.1: Percent and Type of Telephone Number Errors – Registrant Contact



Note: 2009 RAA does not require the presence of a Registrant telephone number.

Graph B.2: Percent and Type of Telephone Number Errors – Administrative Contact



Graph B.3: Percent and Type of Telephone Number Errors – Administrative Contact

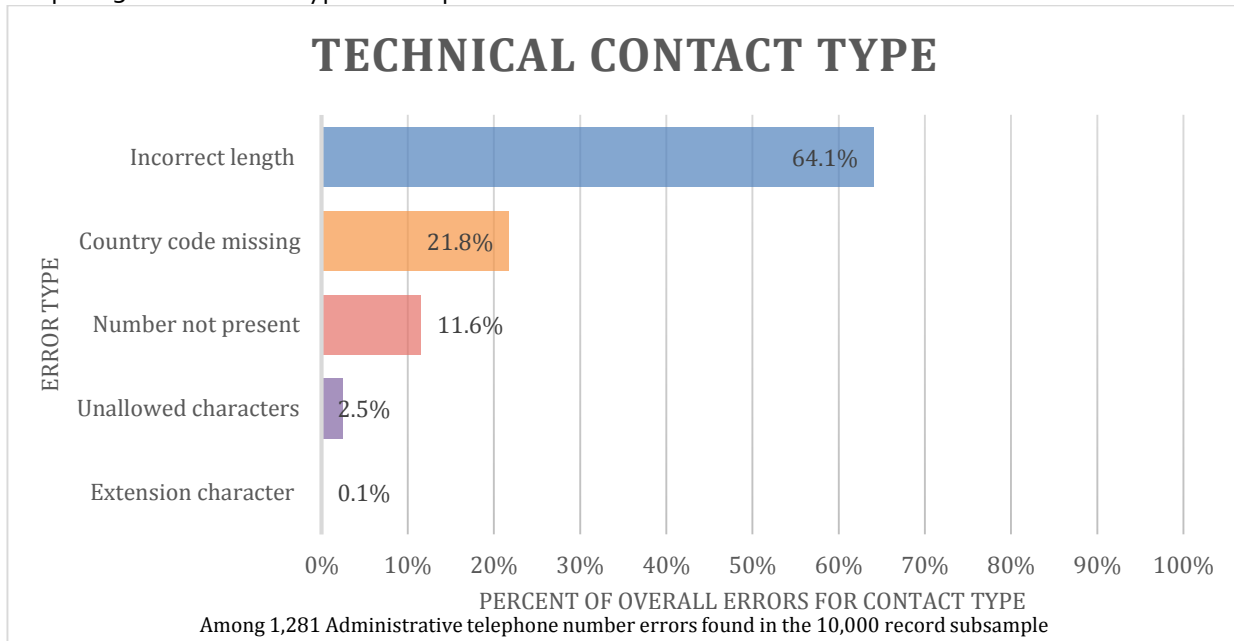
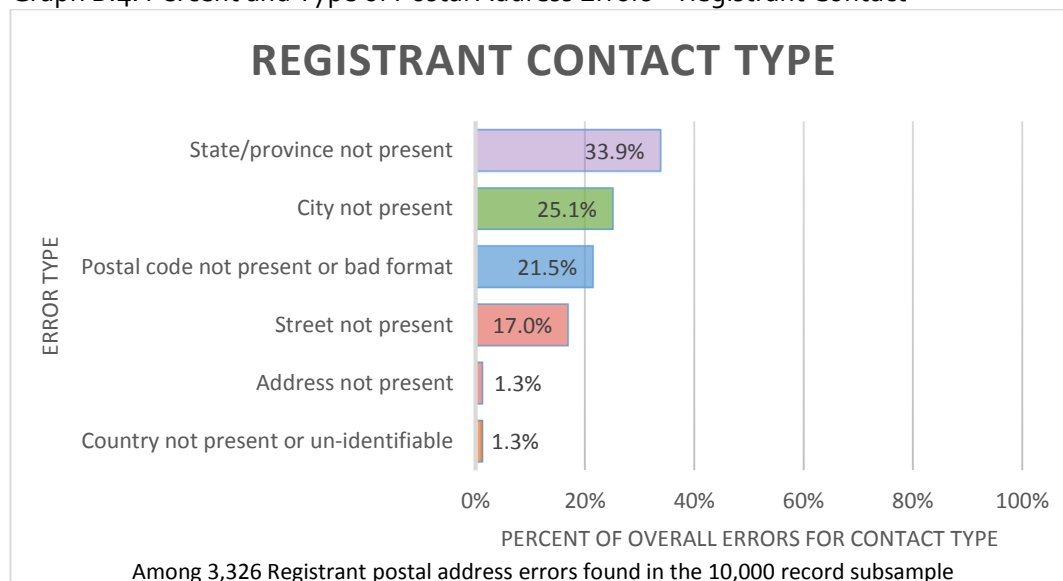


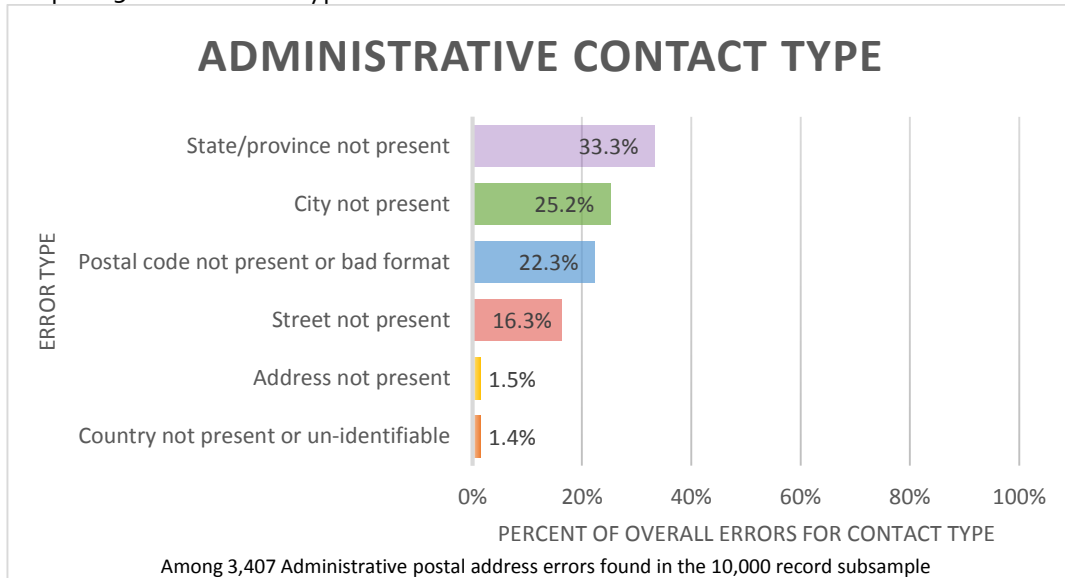
Table B.2: Total and Type of Postal Address Errors by Contact Type (2009)

	Administrative	Technical	A + T	Registrant	All 3 total
Passed all accuracy tests	7,570	7,826	15,396	7,582	22,978
Missing	50	56	106	42	148
Country missing	22	22	44	18	62
Country un-identifiable	26	27	53	24	77
Postal code missing	736	665	1,401	691	2,092
Postal code bad format	24	20	44	25	69
State missing	1,134	995	2,129	1,126	3,255
City missing	858	777	1,635	836	2,471
Street missing	557	494	1,051	564	1,615
Total	10,000	10,000	20,000	10,000	30,000
Total errors	3,407	3,056	6,463	3,326	9,789
Total Domains w/ Errors	2,430	2,174	4,604	2,418	7,022

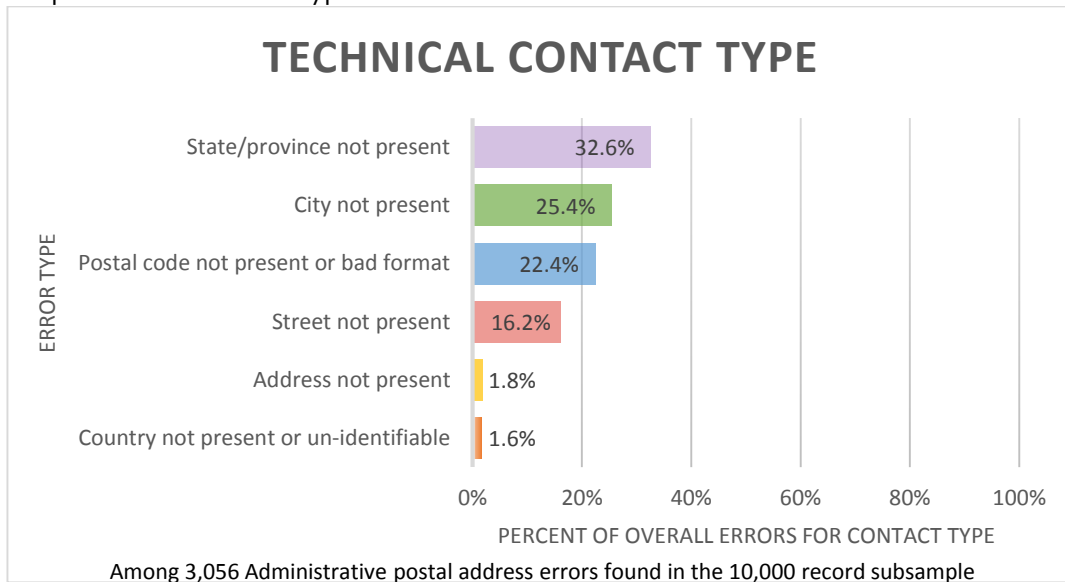
Graph B.4: Percent and Type of Postal Address Errors – Registrant Contact



Graph B.5: Percent and Type of Postal Address Errors – Administrative Contact



Graph B.6: Percent and Type of Postal Address Errors – Administrative Contact

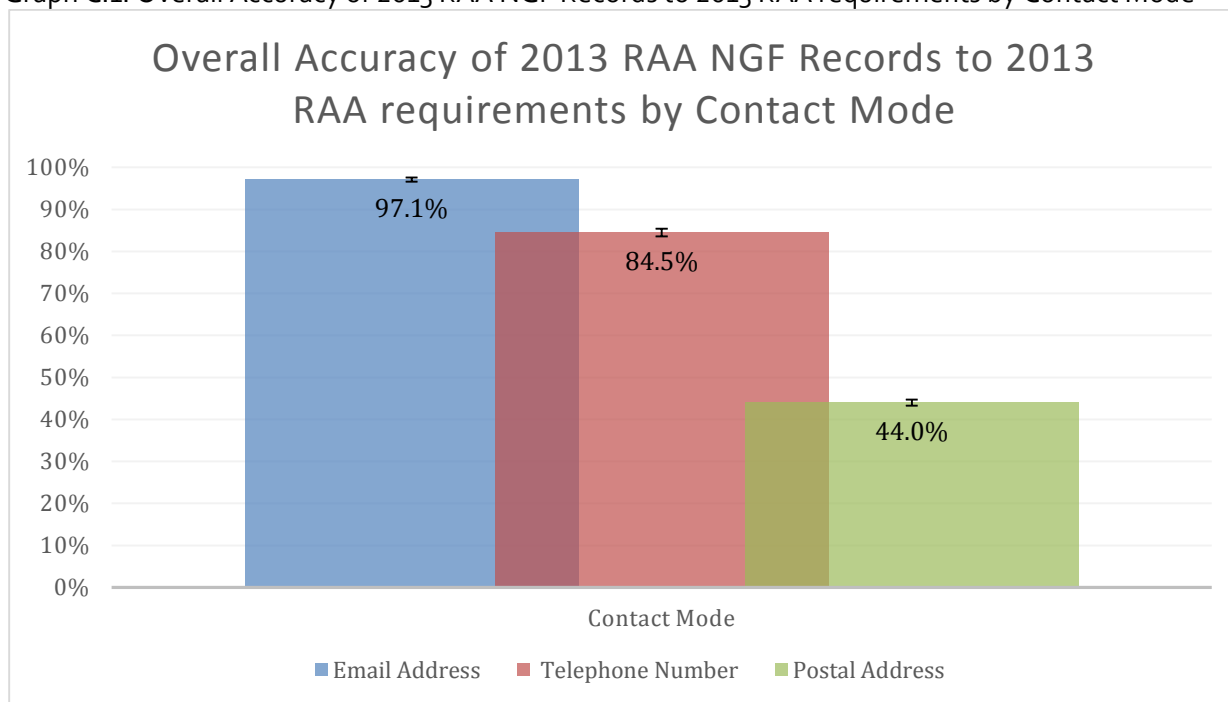


Appendix C: Accuracy Result Tables, 2013 RAA Requirements

Conformance to 2013 RAA Requirements

There are 3,848 domains with 2013 RAA registrars that are not grandfathered because they have been registered when their registrar was already under the 2013 RAA. Graph C.1 below shows the overall accuracy by contact mode of 2013 RAA domains.

Graph C.1: Overall Accuracy of 2013 RAA NGF Records to 2013 RAA requirements by Contact Mode



The 2013 RAA adds many additional requirements, especially to postal addresses. Among 2013 RAA domains, less than 50 percent passed all postal address syntax accuracy tests. Requiring registrant e-mail addresses and telephone numbers are also new requirements compared to the 2009 RAA. Conformance lags for these two contacts, in contrast to our finding that there are no differences between registrant, administrative, and technical syntax accuracy when assessing 2009 RAA requirements. The many new requirements for postal addresses reduce the percentage of domains passing all syntax accuracy tests below 50 percent.

Table C.1: Overall Conformance of 2013 NGF Records to 2013 RAA by Contact Type and Mode

SYNTAX	E-mail	Telephone	Postal Address	ALL 3 Accurate
Registrant	97.2% ± 0.5%	84.9% ± 1.1%	46.2% ± 1.6%	39.7% ± 1.5%
Administrative	98.9% ± 0.3%	86.1% ± 1.1%	46.0% ± 1.6%	40.3% ± 1.5%
Technical	98.9% ± 0.3%	86.6% ± 1.1%	46.4% ± 1.6%	41.0% ± 1.6%
ALL 3 accurate	97.1% ± 0.5%	84.5% ± 1.1%	44.0% ± 1.6%	37.8% ± 1.5%

Most of the 3,848 domains being checked for 2013 RAA syntax accuracy are in New gTLDs, but there are enough in Prior gTLDs domains to make the comparisons shown in Table C.2 below. There are no significant differences for telephone numbers, but the domains in New gTLDs were significantly more likely to pass all 2013 RAA syntax accuracy tests for email and postal addresses.

Table C.2: Prior versus New gTLDs: Syntax Accuracy for all contact types combined

gTLD Type	E-mail	Telephone	Postal Address	ALL 3 Accurate
Prior gTLD	96.8% ± 0.8%	84.2% ± 1.7%	42.0% ± 2.2%	35.3% ± 2.2%
New gTLD	99.8% ± 0.2%	86.4% ± 1.5%	61.5% ± 2.1%	59.4% ± 2.2%

We can see from Table C.3 that Europe and North America are lower in registrant email address syntax accuracy, and North America is significantly lower in email address syntax accuracy for the Administrative and Technical contacts as well. Telephone syntax accuracy is lowest for the Africa and Asia-Pacific regions, while Europe has significantly higher syntax accuracy than Latin America/Caribbean for the administrative and technical telephone numbers. Africa is significantly lower in postal address syntax accuracy than all other regions. Europe and Latin America/Caribbean are significantly higher in postal address syntax accuracy than all other regions. Asia-Pacific is significantly different from all other regions (second lowest), as is North America (third lowest/third highest).

Table C.3: ICANN Region Syntax Accuracy for all contact types combined (2013)

ICANN Region	E-mail	Telephone	Postal Address	All 3 Accurate
Africa	98.7% ± 1.2%	74.8% ± 4.4%	26.1% ± 4.4%	20.3% ± 4.1%
Asia Pacific	99.7% ± 0.4%	72.4% ± 3.0%	36.6% ± 3.3%	27.3% ± 3.0%
Europe	94.4% ± 1.6%	88.3% ± 2.2%	54.6% ± 3.5%	48.4% ± 3.5%
Latin America and Caribbean	99.8% ± 0.3%	87.5% ± 2.3%	61.5% ± 3.4%	58.7% ± 3.4%
North America	95.9% ± 1.2%	91.4% ± 1.7%	43.9% ± 3.0%	39.4% ± 3.0%

For completeness and interest, we do compare all three RAA groups on 2013 criteria, though it is important to note that only one of the three groups can be held to these standards. It was hoped that this group (2013 RAA Non-Grandfathered domains) would have the highest accuracy by these standards.

Most noticeably, none of the 2009 RAA or 2013 RAA Grandfathered domains were subjected to syntax address accuracy tests for the 2013 RAA. We will carry these accuracy tests out in the next phase of this work (which will include syntax and operability accuracy tests). The 2013 RAA non-grandfathered domains also have the highest rate of passing all the syntactic telephone number accuracy tests for the 2013 RAA. A significantly higher percentage of 2013 RAA grandfathered domains pass syntactic telephone number accuracy tests for the 2013 RAA than for 2009 RAA domains. There are no significant differences among email addresses except that a significantly lower percentage of 2013 RAA non-grandfathered domains pass all email address syntactic accuracy tests than the other two groups.

Table C.4: ICANN RAA Type Syntax Accuracy for all contact types combined – 2013 RAA

RAA Type	E-mail	Telephone	Postal Address	ALL 3 Accurate
2009 RAA	98.9% ± 0.5%	71.1% ± 2.1%	n/a	n/a
2013 RAA GF	97.8% ± 0.4%	74.2% ± 1.3%	n/a	n/a
2013 RAA NGF	97.1% ± 0.5%	84.5% ± 1.1%	44.0% ± 1.6%	37.8% ± 1.5%

Major Reasons for Errors under 2013 RAA

As discussed in the body of the report, the 2013 RAA has stricter requirements for content and formatting of contact information in a WHOIS record than the 2009 RAA. We provide similar tables of error breakdowns for the 3,848 2013 RAA non-grandfathered domains. Each domain was checked against all 2013 RAA requirements (which include the 2009 RAA requirements).

Email Address Errors

There are no additional tests beyond the 2009 RAA standards required for email addresses under the 2013 RAA. However, where it is not required under the 2009 RAA, a registrant email address *is* required under the 2013 RAA. Accurate syntax again means having an “@” character, a domain part of the email address, a local part, no unallowed characters, and correct overall format. Table C.5 below shows by contact type the total number of email addresses that passed.

Table C.5: Total Email Address Errors by Contact Type (2013)

	Administrative	Technical	A + T	Registrant	All 3 total
Passed all accuracy tests	3,829	3,830	7,659	3,802	11,461
Missing	19	18	37	46	83
Total	3,848	3,848	7,696	3,848	11,544

Again, no email addresses failed any of the second-stage syntax accuracy tests. However, there were more missing registrant e-mail addresses (46, or 1.2 percent) than were administrative or technical (37, or 0.5 percent). Regardless, the total number of missing e-mail addresses (83, or 0.7 percent) is miniscule.

Telephone Number Errors

Using the 2013 RAA standards, a registrant telephone number is now required. Table C.6 shows the total number and type of errors by contact type.

Table C.6: Total and Type of Telephone Number Errors by Contact Type (2013 RAA)

	Administrative	Technical	Registrant	All 3 total
Passed all accuracy tests	3,362	3,389	3,336	10,087
Missing	102	103	121	326
Country code missing	70	65	76	211
<i>Country code bad format</i>	<i>90</i>	<i>91</i>	<i>91</i>	<i>272</i>
Incorrect length	223	199	223	645
Unallowed characters	1* (47)	1*(46)	1*(48)	3*(141)
Total	3,848	3,848	3,848	11,544

Note: Italics indicate new 2013 RAA requirements

Similar to the 2009 RAA, approximately 87 percent of each telephone number contact type passed all 2013 RAA syntax accuracy tests. The 2013 RAA standards have two additional tests beyond the 2009 RAA standards: country code and extension (if appropriate) must be both present and in the correct format. The requirement of the extension did not result in further errors, while the requirement for the country code did result in additional syntax test errors. We were able to identify a singular syntax test that resulted in a failure for each domain. All domains with unallowed characters also failed another test, except one domain with "NA" as the extension for all three contact types. About one-third of the reasons for failure was too many digits; an additional one-fourth was missing entirely.

Postal Address Errors – 2013 RAA

The 2013 RAA standards add a number of requirements. There are additional accuracy tests that the country, postal code, city, state, and street all appear in the correct field, as well as that the country must also be in ISO alpha 2 format, and the state must be in the proper format per country requirements. Table B.7 shows the total number and type of errors by contact type.

Table C.7: Total and Type of Postal Address Errors by Contact Type (2013 RAA)

	Administrative	Technical	Registrant	ALL 3 TOTAL
PASSED ALL ACCURACY TESTS	2,186	2,239	2,187	6,612
Missing	19	19	16	54
Country Missing	3	3	2	8
Country Un-Identifiable	10	10	10	30
<i>Country in Wrong Field</i>	<i>811</i>	<i>811</i>	<i>810</i>	<i>2,432</i>
<i>Country not ISO alpha 2</i>	<i>1</i>	<i>1</i>	<i>1</i>	<i>3</i>
Postal Code Missing	255	250	233	738
Postal Code Bad Format	8	8	8	24
<i>Postal Code in Wrong Field</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>
State Missing	140	124	140	404
<i>State in Wrong Field</i>	<i>357</i>	<i>336</i>	<i>344</i>	<i>1,037</i>
<i>State Bad Format</i>	<i>83</i>	<i>79</i>	<i>84</i>	<i>246</i>
City Missing	283	255	278	816
<i>City in Wrong Field</i>	<i>165</i>	<i>186</i>	<i>165</i>	<i>516</i>
Street Missing	242	224	249	715
<i>Street in Wrong Field</i>	<i>31</i>	<i>28</i>	<i>30</i>	<i>89</i>
TOTAL	3,848	3,848	3,848	11,544
Total Errors	2,408	2,334	2,370	7,112
Total Domains w/ Errors	1,662	1,609	1,661	4,932

Note: Italics indicate new 2013 RAA requirements

With the additional 2013 accuracy tests, the percentage of domains passing drops to around 57 percent for all three contact types. The total number of errors are larger than the number of domains with errors because it is possible to have more than one error (i.e., there is no overlap between the missing and country code errors. However, there is overlap among the postal code, state, city, and street missing errors, as well as between these and the country code errors.