WHOIS Accuracy Reporting System (ARS)

Phase 2 Cycle 1 Report: Syntax and Operability Accuracy

Global Domains Division | 21 December 2015

Contents

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 Report[2] (syntax, operability, and identity). Phase 1[3] was completed in August 2015 and analyzed syntax accuracy only. Phase 2, the subject of this report, reviews both the syntax and operability accuracy of WHOIS records and details the leading types of nonconformance, trends and comparisons of WHOIS accuracy across regions, Registrar Accreditation Agreement (RAA) versions and gTLD types.

Accuracy Testing Methods[4]

Syntax and operability accuracy testing were designed to assess the contact information of a WHOIS record by comparing it to the applicable contractual requirements of the RAA. Syntax testing assessed the format of a record (e.g., does the email address contain an "@" symbol?), and operability testing assessed the functionality of the information in a record (e.g., did the email not get bounced back?). Syntax and operability accuracy 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 and operability accuracy for WHOIS contact information across subgroups such as New gTLDs or Prior gTLDs, Region, and RAA type (i.e., 2009 RAA or 2013 RAA[5]).

---

[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] *See* the report here: http://whois.icann.org/en/file/whoisars-phase1-report.
[4] More detailed descriptions of the syntax accuracy tests can be found in Appendix A: Accuracy Criteria.
[5] *See* here for RAA versions: https://www.icann.org/resources/pages/registrars/registrars-en

The accuracy tests were designed in such a way that all records in the analyzed subsample are evaluated against a set of baseline requirements derived from the requirements of the 2009 RAA.[6] While operability requirements differ little between the 2009 and 2013 RAA versions (only that the Registrant email address and telephone number are not required for 2009), the 2013 RAA requires the contact data in a WHOIS record to be more syntactically 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-3166-1[7].

More information on the methodology of this study and the accuracy tests performed can be found in the Study Methods and Approach section in the body of this report and in Appendix A: Accuracy Testing Criteria.

Sample Design

At the time of the initial sample in late 2nd quarter 2015, there were nearly 158 million domains names[8] spread across 442 gTLDs.[9] Nearly 97 percent of the 158 million domains were registered in one of the 18 Prior gTLDs, and about 3 percent were registered in one of the 424 New gTLDs. A two-stage sampling method 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. That is, there was an initial sample followed by a subsample; the initial sample contained approximately 150,000 records and the analyzed subsample contained 10,000 records representing all active gTLDs at the time.[10]

Though an estimated 97 percent of domain names are registered through registrars which have been accredited under the 2013 RAA, a majority of domains are allowed to operate under the WHOIS standards of the 2009 RAA. 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). For this reason, the 2009 RAA criteria is used as the baseline to assess WHOIS accuracy in this report, however all 2013 RAA NGF domains were also tested to the 2013 RAA criteria, the findings are available in Appendix C. Table 1.Ex below shows the breakdown of the initial sample described above. More detailed information can be found in the Sample Design section in the body of this report.

---

[6] The criteria listed in Appendix A are what we have defined as the baseline requirements of contact data to be deemed formatted correctly as well as operable. While the 2009 RAA does not contain explicit syntax requirements, the contact data provided is expected to be formatted correctly and completely

[7] *See* http://www.iso.org/iso/home/standards/country_codes.htm.

[8] Based on information from the gTLD zone files.

[9] At the time of sampling there were 678 delegated gTLDs (18 Prior gTLDs and 660 New gTLDs), 442 of the 678 gTLDs had at least three domains, the others were excluded from sampling.

[10] 424 New gTLDs and 18 Prior gTLDs with at least three domains.

Table 1.Ex: Initial Sample Sizes by Region and RAA

RAA Type	Africa	Latin America and Caribbean	Europe	Asia Pacific	North America	Unknown	TOTAL
2009	21	246	488	761	2,257	46	3,819
2013 GF	438	2,095	15,897	12,035	41,924	361	72,750
2013 NGF	529	3,168	14,227	26,115	26,072	427	70,538
TOTAL	988	5,509	30,612	38,911	70,253	834	147,107

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 2 findings include the rates of both syntax and operability accuracy of WHOIS contact information over several dimensions, focusing on rates of accuracy by contact mode (email address, telephone number, and postal address) to the requirements of RAAs (2009 RAA or 2013 RAA). The results from the analyzed subsample testing are then used to 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 interval[11] 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 is within the 95 percent confidence interval.

Syntax Accuracy

For syntax accuracy, our analysis finds that approximately 99 percent of email addresses, 83 percent of telephone numbers, and 79 percent of postal addresses met all of the baseline syntax requirements of the 2009 RAA for all three contacts.[12] Full syntax accuracy of an entire WHOIS record (all three contact types, for all three contact modes) to the requirements of the 2009 RAA was approximately 67 percent for the gTLD population as a whole. Table 2.Ex below provides the accuracy breakdown by contact mode, presented as 95 percent confidence intervals.

Table 2.Ex: Overall[13] gTLD accuracy to 2009 RAA Syntax Requirements by mode

	Email	Telephone	Postal Address	ALL 3 Accurate
All 3 Contacts Accurate	99.1% ± 0.2%	83.3% ± 0.7%	79.4% ± 0.8%	67.2% ± 0.9%

Operability Accuracy

For operability accuracy, our analysis finds that approximately 87 percent of email addresses, 74 percent of telephone numbers, and 98 percent of postal addresses were found to be operable for all three contacts. Full operability accuracy of an entire WHOIS record was approximately 65 percent for the gTLD

---

[11] 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.

[12] *See* note 6.

[13] Overall accuracy refers here to the entire 158 million domains; See note 9 on confidence intervals and population.

population as a whole. Table 3.Ex below provides the accuracy breakdown by contact mode, presented as 95 percent confidence intervals.

Table 3.Ex: Overall gTLD accuracy to 2009 RAA Operability Requirements by mode

	Email	Telephone	Postal Address	ALL 3 Accurate
All 3 Contacts Accurate	87.1% ± 0.7%	74.0% ± 0.9%	98.0% ± 0.3%	64.7% ± 0.9%

The leading causes of syntax and operability nonconformance in the various subgroups are examined and explained in the [Findings](#) section in the body of this report, as well as in [Appendix B](#) and [Appendix C](#)

Next Steps

Phase 2 Cycle 2

The WHOIS ARS is intended to be a system used for repeatable assessment; the next cycle of the WHOIS ARS Report (Phase 2 Cycle 2) will reprise the syntax and operability review of Phase 2 Cycle 1. The second cycle will be performed like the first, except for those areas discussed in [Challenges and Lessons Learned](#). Phase 2 Cycle 2 will begin in January 2016 with a report targeted for early June 2016.

ICANN Contractual Compliance

As of the publication of the report, the results (i.e., all potentially inaccurate records) of Phase 2 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. As Phase 2 includes both syntax and operability results, compliance follow-up and investigation may be conducted through different processes depending on the type of inaccuracies found within each record, e.g., those records that have been deemed "operable" but with formatting errors will receive a different kind of notice than those records have been deemed "inoperable" with formatting errors. All WHOIS ARS tickets will follow the Contractual Compliance Approach and Process[14] according to the types of issues described in this report. When possible, and in consultation with registrars, ICANN may be able to consolidate multiple WHOIS ARS tickets during processing. WHOIS ARS tickets will be processed alongside other complaints; however, ICANN will continue to give priority to complaints submitted by community members.

---

[14] *See* ICANN Contractual Compliance Approach and Process: https://www.icann.org/resources/pages/approach-processes-2012-02-25-en.
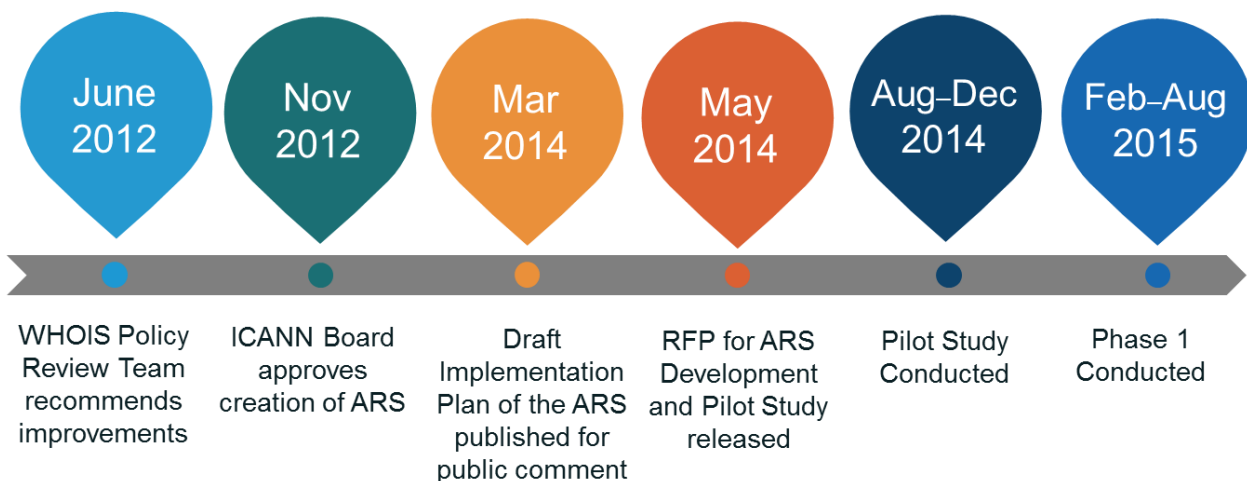
Introduction

WHOIS ARS Background and Approach

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)[15].

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. Figure 1 below illustrates the timeline of events in the development of the WHOIS ARS project.

Figure 1: ARS Background

June 2012	Nov 2012	Mar 2014	May 2014	Aug–Dec 2014	Feb–Aug 2015
WHOIS Policy Review Team recommends improvements	ICANN Board approves creation of ARS	Draft Implementation Plan of the ARS published for public comment	RFP for ARS Development and Pilot Study released	Pilot Study Conducted	Phase 1 Conducted

With input from the community over the course of 2014, ICANN planned for the implementation of the ARS to be in three phases, based on the types of validations described in the SAC058 Report[16]. Phase 1 analyzed the syntax accuracy of WHOIS contact information (i.e., is the contact data complete and formatted correctly?). Phase 2, the subject of this report, assesses the operability of the contact data in the record by combining the syntax tests from Phase 1 with operability tests such as "Does the phone
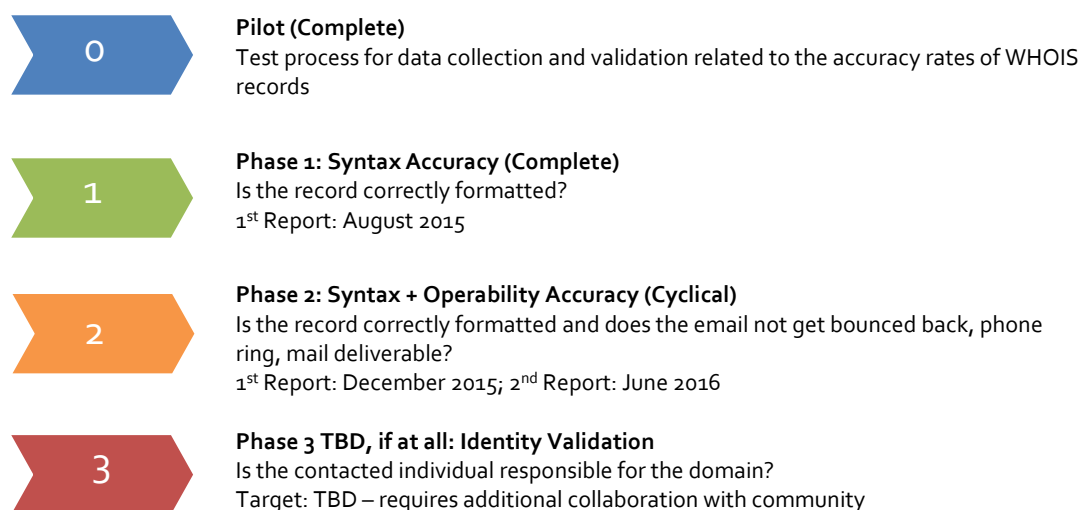
---

[15] *See* https://www.icann.org/resources/pages/aoc-2012-02-25-en.
[16] *See* https://www.icann.org/en/system/files/files/sac-058-en.pdf.

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, the timeline for implementation of Phase 3 has not yet been determined as ICANN will continue to work with the Community to assess 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 2 illustrates this phased approach. ICANN expects to produce new Phase 2 reports every 6 months, with each successive report focusing more and more on the "system results" and improvement of the data over time.

Figure 2: WHOIS ARS Phases

0	**Pilot (Complete)** Test process for data collection and validation related to the accuracy rates of WHOIS records
1	**Phase 1: Syntax Accuracy (Complete)** Is the record correctly formatted? 1st Report: August 2015
2	**Phase 2: Syntax + Operability Accuracy (Cyclical)** Is the record correctly formatted and does the email not get bounced back, phone ring, mail deliverable? 1st Report: December 2015; 2nd Report: June 2016
3	**Phase 3 TBD, if at all: Identity Validation** Is the contacted individual responsible for the domain? Target: TBD – requires additional collaboration with community

Phase 1 Recap

A Phase 1 Report[17] of the WHOIS ARS was published in August 2015 and acted as a follow-up to the Pilot Study conducted in 2014, taking lessons learned from the Pilot Study to refine the ARS testing criteria. The major findings from Phase 1 included:

- Ninety-nine percent of email addresses, 85 percent of telephone numbers and 79 percent of postal addresses met all syntax requirements of the 2009 RAA. Seventy 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.

---

[17] *See* the report here: http://whois.icann.org/en/file/whoisars-phase1-report.

- 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 fewer mistakes on each.

Phase 2 Overview

Phase 2 Objectives

The objective for Phase 2 is to examine both syntax and operability accuracy of WHOIS records. ICANN seeks to determine whether the WHOIS record is meeting the format and content requirements of the applicable RAA and if the contact data provided is contactable. The Phase 2 report builds off of Phase 1 and details the leading types of nonconformance, 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 or inoperable records, leading to investigation, and if needed, correction.

While the report provides comparisons between Phase 1 syntax Accuracy and Phase 2 syntax Accuracy, any improvement in the accuracy of the WHOIS data cannot be directly linked to the ARS. Due to the overlap of Phase 1 and Phase 2 (i.e., Phase 2 data was pulled before the conclusion of Phase 1), ICANN had not yet presented the aggregate data to the ICANN community nor had Contractual Compliance started follow-up with registrars on potential inaccuracies. In light of this, there will be a lag in the potential effect of the ARS, and more likely discussed in later reports. Other factors affecting the data and any changes from Phase 1 to Phase 2 will be discussed in the Findings section.
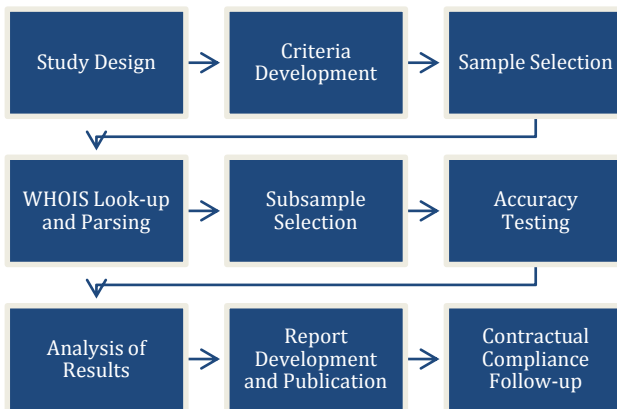
Project Plan, Tasks and Timeline

Phase 2 builds off Phase 1 and adds the examination of the operability of WHOIS contact information. Phase 2 will be conducted in cycles in order to capture trends and improvement in the data. The subject of this report is Phase 2 Cycle 1.

Work on Phase 2 Cycle 1 began in June 2015, overlapping with Phase 1 testing in order to meet the December 2015 report timeframe. Phase 2 was conducted in the same manner as Phase 1: ICANN and the WHOIS ARS vendors approached the work collaboratively[18] and defined together the sampling methodology and accuracy testing criteria. Phase 2, like Phase 1, can be divided up into nine main tasks, which are illustrated below in Figure 3.

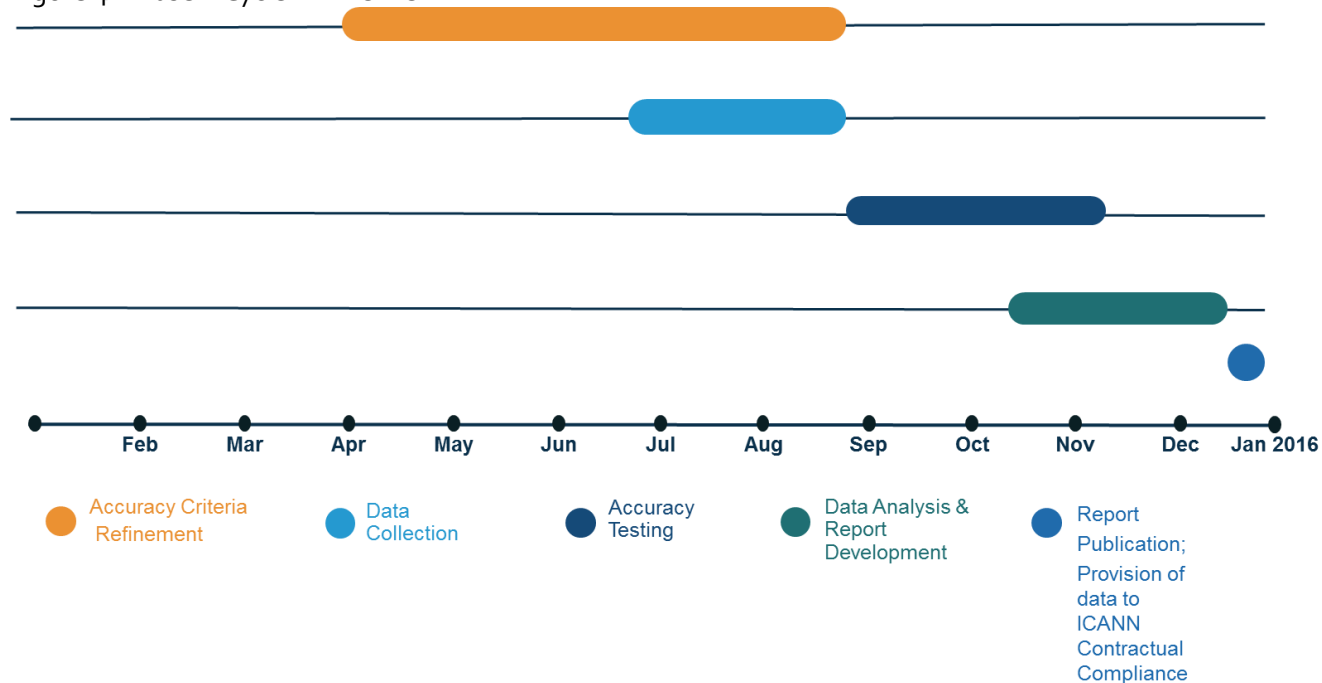[18] Throughout this report, ICANN and the WHOIS ARS team of vendors will be referred to collectively as "the WHOIS ARS team" and, where applicable, a collective "we" will be used to refer in the first person to actions completed by this team.

Figure 3: Work Flow and Tasks

Study Design	→	Criteria Development	→	Sample Selection
WHOIS Look-up and Parsing	→	Subsample Selection	→	Accuracy Testing
Analysis of Results	→	Report Development and Publication	→	Contractual Compliance Follow-up

These tasks[19] were conducted by the team in the timeline illustrated in Figure 4.

Figure 4: Phase 2 Cycle 1 Timeline

Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec   Jan 2016

Accuracy Criteria Refinement

Data Collection

Accuracy Testing

Data Analysis & Report Development

Report Publication; Provision of data to ICANN Contractual Compliance
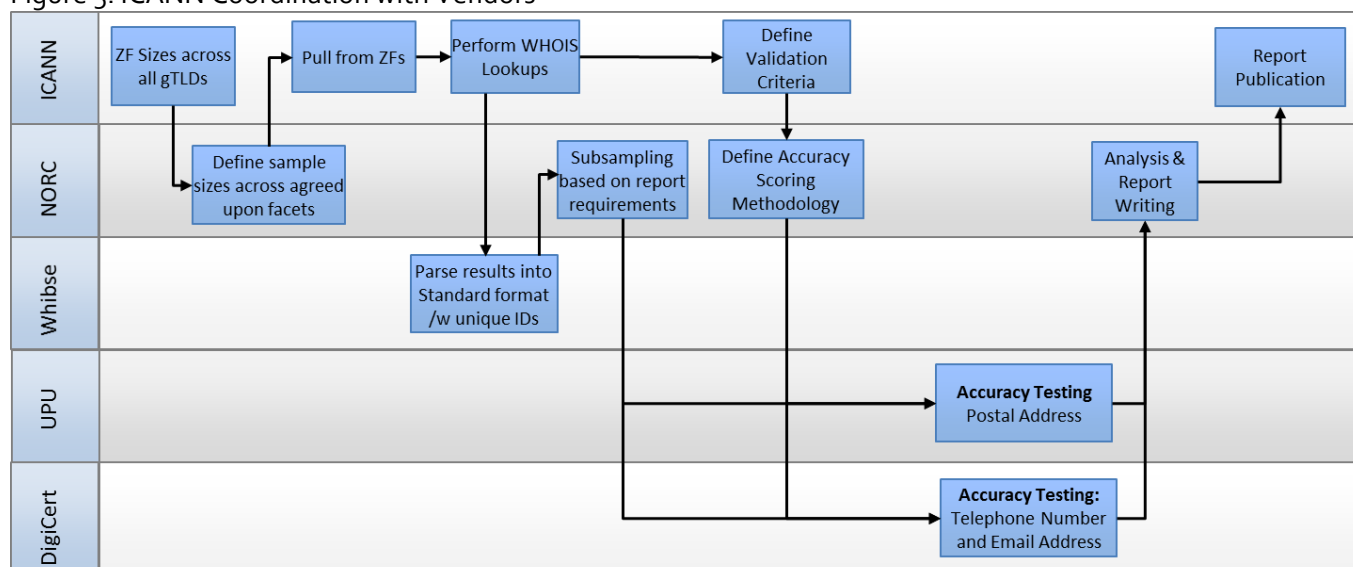
---

[19] With the exception of the Contractual Compliance Follow up, which will begin shortly after publication of this report.

Team Composition

The WHOIS ARS team, listed below, did not change from Phase 1. Figure 5 below illustrates how the team coordinated to develop the Phase 1 report.

- ICANN: Project Management, Accuracy Criteria Definition, WHOIS Lookups

- NORC at the University of Chicago[20]: Study Design, Sample Selection, and Data Analysis

- Whibse, Inc.[21]: Parsing

- DigiCert, Inc.[22]: Email and Telephone Accuracy Testing

- Universal Postal Union[23]: Postal Address Accuracy Testing

Figure 5: ICANN Coordination with Vendors

---

[20] See http://www.norc.org/Pages/default.aspx.

[21] See https://www.whibse.com/.

[22] See https://www.digicert.com.

[23] See http://www.upu.int/en.html.

Study Methods and Approach

Brief Overview

The Phase 2 study first selected a sample of 150,000 WHOIS records from the zone files of 678 gTLDs. Using systematic assessments referred to as syntax Accuracy Testing and operability Accuracy Testing, the contact information from a subsample of 10,000 records was tested for accuracy with syntax standards (i.e., values and formats) based on requirements stipulated within the domain-applicable RAA, and was then tested for accuracy with operability standards (i.e., the information can be used to establish contact). The resulting data were analyzed to produce statistics of syntax and operability accuracy for WHOIS contact information across subgroups such as gTLD Type (Prior or New), ICANN Region, and RAA type. Though an estimated 97 percent of domain names are registered through registrars which have agreed to 2013 RAA, a majority of domains with registrars on the 2013 RAA are only obligated to meet the WHOIS requirements of the 2009 RAA based on when the domain itself 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 150,000 records from gTLD zone files (this number was increased from 100,000 during Phase 1), 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 accuracy estimates for each subgroup. 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 150,000 records, we review the zone file summary data, which indicates how many domains are in each gTLD. At the time of the initial sample in late 2nd quarter 2015, there were nearly 158 million domains names spread across 442 gTLDs.[24] Nearly 97 percent of the 158 million domains were registered in one of the 18 Prior gTLDs, and about 3 percent were registered in one of the 424 New gTLDs with at least three domains.[25]

Similar to the Pilot Study and Phase 1 sample, our Phase 2 sample design oversampled New gTLDs so that 25 percent of the initial sample was from New gTLDs. However, in order to decrease the oversampling necessary in the Analyzed Subsample (see below), we increased the initial sample size from 100,000 to 150,000.

To make sure all 424 New gTLDs with at least three domains[26] were represented, we first selected one from each (this number was 10 in the Pilot Study and Phase 1 samples). The remaining sample was selected proportional to size (more were selected from larger New gTLDs). Similarly, we selected a minimum of one domain from all Prior gTLDs (this number was 30 in both the Pilot Study and Phase 1 samples), with the remaining sample proportional to size. All sampling was done by systematic sampling within gTLD.[27] 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 very 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 150,000 records, but 2,080 (1.9 percent; coincidentally the same percentage as for Phase 1) were dropped because WHOIS queries failed (730 domains no longer existed, 1,339 exceeded the rate limit, and 11 were query failures) and an additional 813 were dropped because they were test records. After these were dropped, we were left with an initial sample of 147,107 domains.

Of the initial sample of 147,107 records, the 2009 RAA subgroup accounted for 2.6 percent of all records, while 2013 GF and 2013 NGF records accounted for 49.5 percent and 47.9 percent of all domains respectively (see Table 1).

---

[24] Based on information from the zone files.
[25] 660 New gTLDs were delegated at time of sampling, only 424 New gTLDs had at least three domains, those gTLDs with less than 3 domains were excluded from the sample.
[26] New gTLDs are required to have at least 1 domain name nic.TLD, for statistical purposes gTLDs with less than 3 domain names (two domains + 1 nic.TLD domain) were excluded from the sample.
[27] 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 166.2 and the skip interval was 300, then the selected records would be the 167th (random start rounded up), the 467th, the 767th, and so on.

Table 1: Initial Sample Sizes by Region and RAA

RAA Type	Africa	Latin America and Caribbean	Europe	Asia Pacific	North America	Unknown	TOTAL
2009	21	246	488	761	2,257	46	3,819
2013 GF	438	2,095	15,897	12,035	41,924	361	72,750
2013 NGF	529	3,168	14,227	26,115	26,072	427	70,538
TOTAL	988	5,509	30,612	38,911	70,253	834	147,107

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 five ICANN regions. Accordingly, we selected the analyzed subsample to maximize the ability of keeping estimates of the accuracy of the data for subgroups of interest with 95 percent confidence intervals of no more than plus or minus five percent. This kind of confidence interval required certain subgroups to be oversampled (even selected with certainty) relative to their representation in the initial sample of 150,000 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 over 400 registrars represented in the subsample. The subsample also did not consider gTLD type (Prior versus New) because the initial sample oversampled New gTLDs. Table 2 below shows the sizes of the analyzed subsample by Region and RAA.[28]

Table 2: Analyzed Subsample Sizes by Region and RAA

RAA Type	Africa	Latin America and Caribbean	Europe	Asia Pacific	North America	Unknown	TOTAL
2009	21	246	488	761	800	16	2,332
2013 GF	438	800	800	800	1,079	9	3,926
2013 NGF	529	800	800	800	800	13	3,742
TOTAL	988	1,846	2,088	2,361	2,679	38	10,000

Records in the 2009 RAA subgroup accounted for 2.6 percent of all initial sample records, while 2013 GF and 2013 NGF records accounted for 49.5 percent and 47.9 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 criteria described above. Table 3 below shows the sample sizes by RAA type in the initial sample of 147,107 and the analyzed subsample of 10,000.

---

[28] 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). 800 was chosen as the goal in order to minimize the size of the confidence intervals in each cell. If a cell had less than 800 in the initial sample, all were selected. We oversampled most other cells to obtain 800 domains in each. Only the North America 2013GF cell has more than 800 domains selected. We sampled the Region Unknown cells at the same proportion as the North America cells.

Table 3: Sample Sizes by RAA Type

RAA Type	Percentage of All Domains	Initial Sample	Analyzed Subsample	Percentage of Subsample
2009 RAA	3.3%	3,819	2,332	23.3%
2013 RAA GF	63.7%	72,750	3,926	39.3%
2013 RAA NGF	33.0%	70,538	3,742	37.4%
TOTAL	**100.0%**	**147,107**	**10,000**	**100.0%**

Syntax Testing Methods

Syntax accuracy testing was designed to assess the contact information of a record by comparing it to formats specified by contractual requirements stipulated in the RAAs. Syntax accuracy tests remained consistent from Phase 1 to Phase 2.

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.[29] The 2009 RAA criteria is used as the baseline to assess WHOIS accuracy in this report because the majority of domains are still obligated to the WHOIS requirements of the 2009 RAA for the reasons explained above. However, all 2013RAA NGF domains in the analyzed subsample were also tested to the 2013 RAA criteria, the findings are available in Appendix C.

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, administrative or technical 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 detail in Appendix A.

---

[29] *See* note 6.

Operability Testing Methods

Operability accuracy testing was designed to assess whether the contact information of a record can be practically used for communication. In consultation with the community, including volunteers from the registrar community, operability accuracy tests were developed to align with RAA requirements. The operability requirements differ little between the 2009 and 2013 RAA versions (only that the Registrant email address and telephone number are not required for 2009).

As with the syntax tests, operability 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. 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. Duplicative data within WHOIS records (e.g., same email address used for all three contact types) and across WHOIS records (e.g., same Registrant contact data used in multiple records) were only tested one time (i.e., duplicates removed). The operability tests for each contact mode are described in brief below.

Email Address Testing

In the 2009 RAA, the presence of an administrative and a technical email address is required, while the presence of a registrant email address is optional.[30] In the 2013 RAA, the registrant, administrative, and technical email addresses are each required to be present.

- Email address operability testing consisted of transmitting one email message to each unique email address in the analyzed subsample (i.e., duplicates removed). The transmitted email contained a standard message[31] which stated that the email was being sent as part of an ICANN study, and that no reply was necessary. If the email transmission did not result in a failure (i.e., bounced email), the email address was considered operational.

Telephone Number Testing

In the 2009 RAA, presence of administrative and technical telephone numbers is required, and presence of a registrant telephone number is optional.[32] In the 2013 RAA, the registrant, administrative, and technical telephone numbers are each required to be present.

- Telephone number operability testing primarily checked to see if a phone number would connect (i.e., ring) when dialed. Similar to email address testing, each unique telephone number was dialed in the analyzed subsample (i.e., duplicates removed).
  - All numbers were dialed internationally (i.e., dialed as if the call originated from a country other than the recipient's country).
  - Dialing tests began by checking if a call could not be completed before all digits were dialed. If a number connected within 60 seconds, the telephone number was determined operable. A

---

[30] However, if the email address is present, it will undergo Operability accuracy testing.
[31] The email message is included in the Appendix as part of the full documentation of operability testing methods.
[32] However, if the telephone number is present it will undergo Operability accuracy testing.

call did not need to be answered for the telephone number to be considered operable. If a dialed call was answered, an automated message[33] stated that the call was part of an ICANN study, and that no reply was necessary.

Postal Address Testing

In both the 2009 RAA and 2013 RAA, presence of a registrant, administrative, and technical postal address is required.

■ Postal address operability testing assesses the likelihood that postal mail can be delivered to the address provided in a record. To accomplish this, the Universal Postal Union (UPU) simulates post office protocols for handling a parcel that is being sent to the postal address provided in the record, but without attempting physical delivery to the destination. In some cases, postal address operability testing can be more forgiving than postal address syntax testing. For example, syntactic accuracy with UPU standards for postal mail requires an abbreviation for state or province (e.g., 'DE' would pass, while 'Delaware' would not), but these syntax elements are not necessarily based on whether the parcel could be delivered.

☐ Verifiable: The address field is not empty, and the address seems generally deliverable because it contains elements of a country, city/locality, and postal code.

☐ Cross-check: These tests assess whether the different elements of the address are compatible with one another. For example, tests to determine if the city or locality exists within the specified country, and (if required for delivery) if the postal code exists in the country, city and state/province.

---

[33] The automated telephone message is included in Appendix A as part of the full documentation of operability testing methods.

Main Findings

Here we present our findings and the statistics related to the outcomes of the syntax and operability accuracy tests. These statistics are organized by contact type (registrant, administrative, and technical) within contact mode (email address, telephone number, and postal address), overall and across the subgroups New vs. Prior gTLDs, 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.[34] Because operability results are similar across RAA versions, separate analysis tables for each set of requirements would be largely redundant. 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 report.[35] 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:
- For syntax accuracy, there was a drop in telephone number accuracy. This change cannot be attributed to effects of Phase 1, however, since Phases 1 and 2 overlapped.
  - □ The syntax reasons for error had very similar distributions to those in Phase 1.
  - □ The drop in telephone number accuracy seems to be due to an increase in missing country codes among the telephone numbers sampled for Phase 2.
  - □ For postal addresses, the vast majority of errors in both Phase 1 and Phase 2 were due to missing fields that were required such as city, state/province, postal code, or street.

- Eighty-seven percent of email addresses, 74 percent of telephone numbers and 98 percent of postal addresses met all operability requirements of the 2009 RAA. Sixty-five percent of domains passed all operability tests for all contact types (registrant, administrative, technical) and contact modes (email address, telephone number, postal address).

  - □ Of those email addresses that failed operability, the majority (10 percent) bounced while very few were simply missing.
  - □ Of the telephone numbers that were present, but failed operability, there were roughly equal numbers that were disconnected, invalid, or that simply did not connect.

---

[34] 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. 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.

[35] In the interest of condensing the findings in this section, many of the analysis tables for 2009 and 2013 results are stored in Appendix B and Appendix C of the report.

- For the small numbers of postal addresses that failed operability testing, almost half did not have an identifiable or easily deduced country.
- Unlike for syntax accuracy, the contact mode with the highest rate of passing all operability tests was postal address. The mode with the lowest rate of passing all operability tests was telephone numbers.

- For over 75 percent of domains, the contact information in the registrant, administrative, and technical contacts is identical for all three contact modes, revealing why accuracy rates among the three contact types are all similar.

Accounting for Common Data across Contact Types

In Phase 2, we tabulated how often the three contact types (registrant, administrative, and technical) were the same for each contract mode (email, telephone, postal address). For all three contact modes, over 75 percent of the domains have the same contact information for all three contact types. Table 4 shows the full distribution of how often the contact information is the same between each contact type.

Table 4: Frequency of common contact information across contact type and mode

Commonality	Email	Telephone	Postal Address
All 3 Exactly Same	76.1% ± 0.8%	79.4% ± 0.8%	76.2% ± 0.8%
Exactly 2 Same, 1 Different	21.4% ± 0.9%	19.4% ± 0.8%	21.4% ± 0.9%
All 3 Different	2.4% ± 0.3%	1.2% ± 0.2%	2.5% ± 0.3%

Table 4 shows that there will not be significant differences between accuracy for the registrant, administrative, and technical contacts because they so often contain the same information. All three contacts are different no more than 2.5 percent of the time. Therefore, while we test and report on all three contact types, it will often be sufficient to simply look at the rates for which "All 3" contact types are accurate. An expanded version of Table 4 can be found as Table B1 in Appendix B.

Syntax Accuracy – 2009 RAA Requirements[36]

The following section reviews the results of the syntax accuracy tests against 2009 requirements by first looking at overall accuracy, then subgroup accuracy, followed by the reasons for error.

Overall Accuracy

First, we look at accuracy to 2009 RAA requirements for all 10,000 domains. The dotted yellow line in Graph 1 below shows that around 67 percent of domains can be said to be syntactically accurate. Ninety-nine percent of

---

[36] Conformance to 2013 RAA Requirements can be found in Appendix C.

email addresses passed the syntax accuracy tests, while 83 percent of telephone numbers and 79 percent of postal addresses passed.

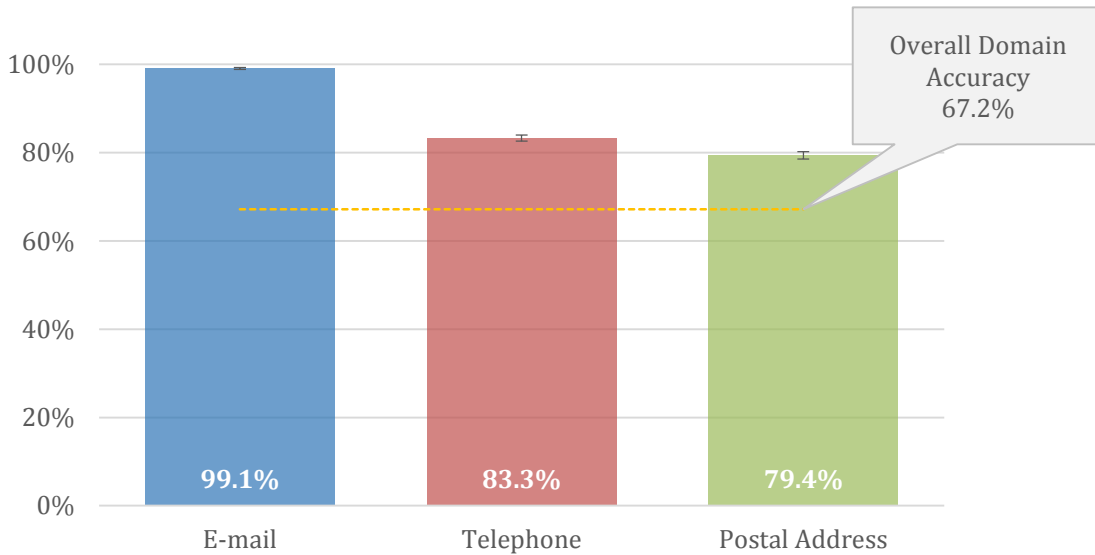Graph 1: Overall Accuracy – 2009 RAA Syntax Requirements

Table 5 below shows in more detail the breakdown by contact type. The bottom row shows the rate for which registrant, administrative and technical contacts all pass. All four rows are very similar because of the commonality of the contact data, suggesting that looking at the registrant, administrative and technical contact accuracy rates separately is unnecessary. We will focus on the percentages that all three contacts pass all tests (the "All 3 Accurate row") in the subgroup analyses.

Table 5: Overall Accuracy by Contact Type and Mode – 2009 RAA Syntax Requirements

	Email	Telephone	Postal Address	ALL 3 Accurate
Registrant	100.0% ± 0.0%	85.6% ± 0.7%	81.0% ± 0.8%	**69.6% ± 0.9%**
Administrative	99.2% ± 0.2%	84.2% ± 0.7%	80.8% ± 0.8%	**69.1% ± 0.9%**
Technical	99.2% ± 0.2%	85.0% ± 0.7%	82.5% ± 0.7%	**71.3% ± 0.9%**
All 3 Accurate	**99.1% ± 0.2%**	**83.3% ± 0.7%**	**79.4% ± 0.8%**	**67.2% ± 0.9%**

Accuracy by Subgroup

Next, we look at subgroups in Phase 2, starting with Prior versus New gTLDs. Because the numbers for the registrant, administrative, and technical contacts are so similar[37], we present here subgroup accuracy only for "All 3 Accurate", i.e., the registrant, administrative, and technical contacts all passed all of the accuracy tests.

Subgroup 1: Prior vs. New gTLDs

Graph 2 and Table 6 below both show that Prior gTLDs have lower[38] accuracy on email addresses and telephone numbers, but higher accuracy on postal addresses. The two groups are not statistically different[39] on having "All 3 Accurate".

Graph 2: Accuracy by gTLD type – 2009 RAA Syntax Requirements
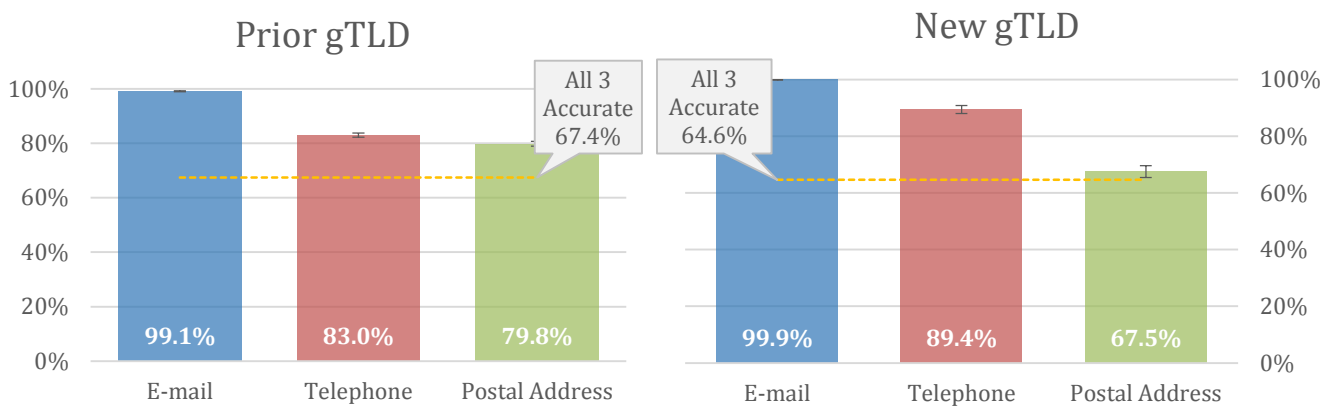
Table 6: Accuracy by gTLD type – 2009 RAA Syntax Requirements

	Email	Telephone	Postal Address	ALL 3 Accurate
Prior gTLD	99.1% ± 0.2%	83.0% ± 0.8%	79.8% ± 0.9%	**67.4% ± 1.0%**
New gTLD	99.9% ± 0.1%	89.4% ± 1.4%	67.5% ± 2.1%	**64.6% ± 2.2%**
All 3 Accurate	**99.1% ± 0.2%**	**83.3% ± 0.7%**	**79.4% ± 0.8%**	**67.2% ± 0.9%**

---

[37] *See* section Accounting for Common Data across Contact Types.

[38] Here "higher" and "lower" 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.

[39] Where there is no statistical significance between subgroups, this is how it will be mentioned.

Subgroup 2: ICANN Regions

Next, we look at accuracy by ICANN region. Again, we present subgroup accuracy for the registrant, administrative, and technical contacts all passing the accuracy tests. Graph 3 and Table 7 below show that there are no statistical differences by region for email addresses. For telephone numbers, Europe and North America are highest and Africa is lowest, just as in Phase 1. For postal addresses, North America is highest and Africa and Asia-Pacific are lowest. North America is also highest for having all three contact modes conform to the RAA 2009 standards and Africa and Asia-Pacific are lowest.

Graph 3: Accuracy by ICANN Region – 2009 RAA Syntax Requirements
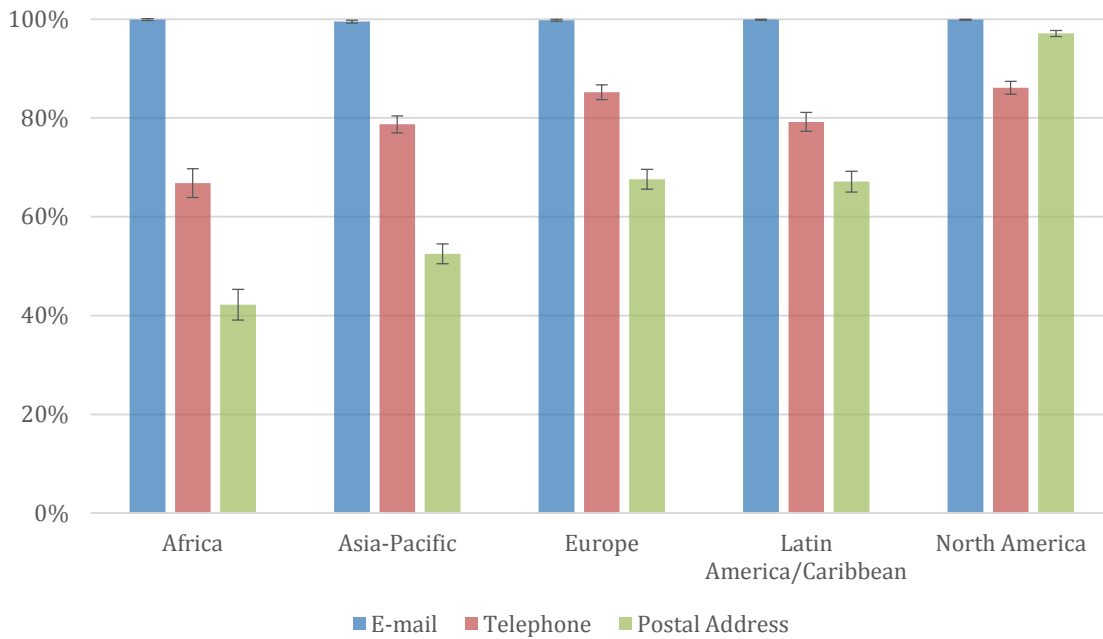
Table 7: Accuracy by ICANN Region – 2009 RAA Syntax Requirements

	Email	Telephone	Postal Address	ALL 3 Accurate
Africa	99.9% ± 0.2%	66.8% ± 2.9%	42.2% ± 3.1%	**29.8% ± 2.9%**
Asia-Pacific	99.5% ± 0.3%	78.7% ± 1.7%	52.5% ± 2.0%	**39.5% ± 2.0%**
Europe	99.8% ± 0.2%	85.2% ± 1.5%	67.6% ± 2.0%	**58.8% ± 2.1%**
Latin America/Caribbean	99.9% ± 0.1%	79.2% ± 1.9%	67.1% ± 2.1%	**56.9% ± 2.3%**
North America	99.9% ± 0.1%	86.1% ± 1.3%	97.1% ± 0.6%	**83.9% ± 1.4%**
All 3 Accurate	**99.1% ± 0.2%[40]**	**83.3% ± 0.7%**	**79.4% ± 0.8%**	**67.2% ± 0.9%**

---

[40] This number is lower than all subgroups because there is a sixth subgroup ("Unknown Region") with only 9 percent conformance.

Subgroup 3: RAA Status

Finally, we look at accuracy rates by RAA group. Again, we present subgroup accuracy for "All 3 Accurate". Graph 4 and Table 8 below show that there are no statistical differences in email address accuracy despite the variability. The 2013 RAA GF group has lower telephone accuracy while the 2013 RAA NGF group has lower postal address accuracy. Since the 2009 RAA was not lower for any mode, this group has a higher percentage of all nine contacts being accurate.

Graph 4: Accuracy by RAA Status – 2009 RAA Syntax Requirements
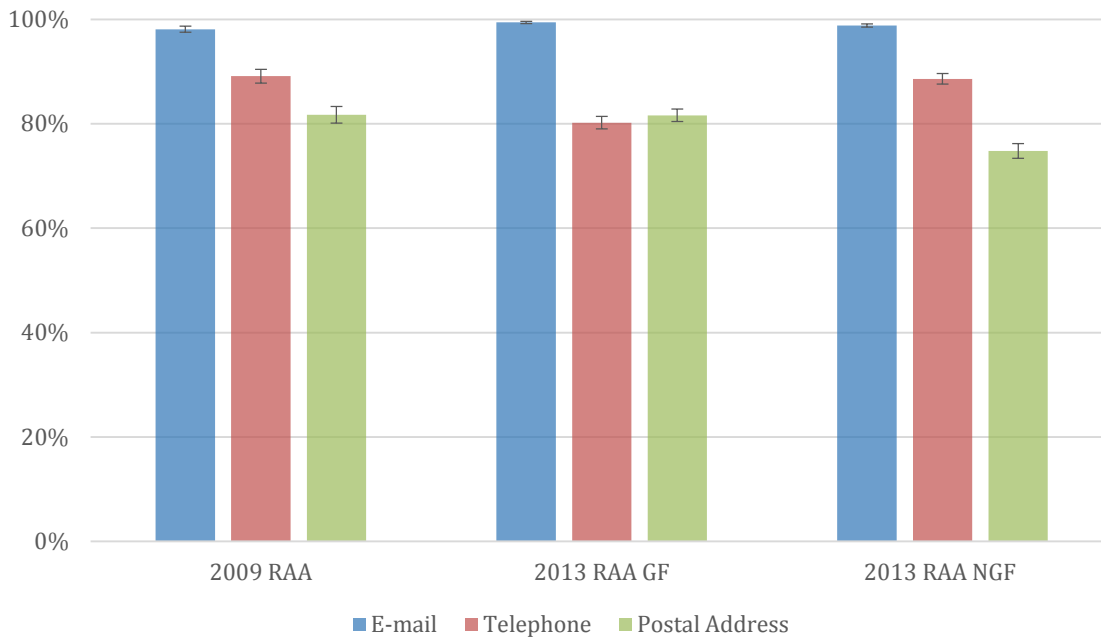
Table 8: Accuracy by RAA Status – 2009 RAA Syntax Requirements

	Email	Telephone	Postal Address	ALL 3 Accurate
2009 RAA	98.1% ± 0.6%	89.1% ± 1.3%	81.7% ± 1.6%	**77.1% ± 1.7%**
2013 RAA GF	99.4% ± 0.2%	80.2% ± 1.2%	81.6% ± 1.2%	**66.5% ± 1.5%**
2013 RAA NGF	98.8% ± 0.3%	88.6% ± 1.0%	74.8% ± 1.4%	**67.8% ± 1.5%**
All 3 Accurate	**99.1% ± 0.2%**	**83.3% ± 0.7%**	**79.4% ± 0.8%**	**67.2% ± 0.9%**

Reasons for Error – 2009 RAA Syntax Requirements

We report here on the major reasons for failure separately by contact mode (email address, telephone number, and postal address). For email addresses and telephone numbers, we were able to pinpoint the first test failed. Because postal addresses require multiple fields, multiple errors were possible for syntax. As we do for the accuracy testing results, we also provide separate tables reporting the major reasons for

failure against the 2009 RAA requirements among all 10,000 analyzed domains.[41] In Phase 1, we showed which accuracy tests were failed by each contact. We repeat these graphs for Phase 2 here. Appendix B shows tabled data from Phase 1 and Phase 2 for comparison, and also includes tables for email addresses, where very few syntax errors were found.

Graph 5: Reasons for Telephone Number Error – 2009 RAA Syntax Requirements
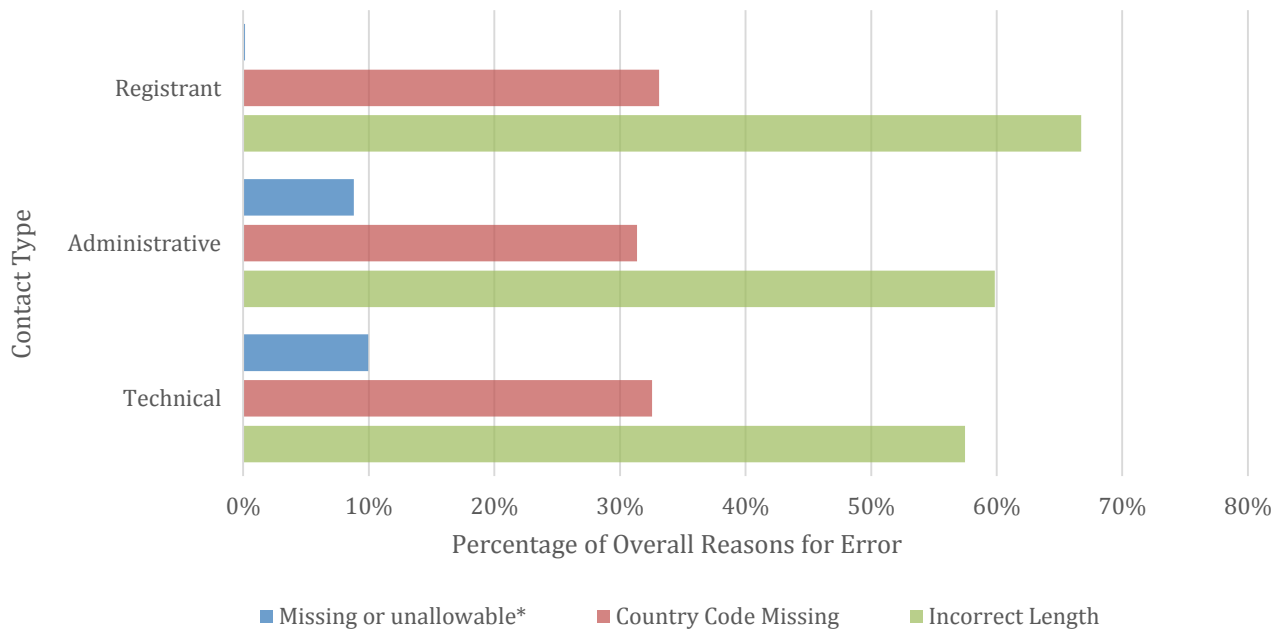
Note: 2009 RAA does not require presence of a telephone number for the Registrant contact type.

Similar to Phase 1, the largest source of errors among telephone numbers was having an incorrect length for the applicable country (around 9 percent of telephone numbers). The next most frequent error was a missing country code (almost 5 percent of telephone numbers). Less than 2 percent of telephone numbers were missing.

---

[41] In Appendix C you can find the major reasons for failure against the 2013 RAA requirements among the 2013 NGF subgroup.

Graph 6: 2009 RAA – Reasons for Postal Address Error Across all Contact Types – Phase 2
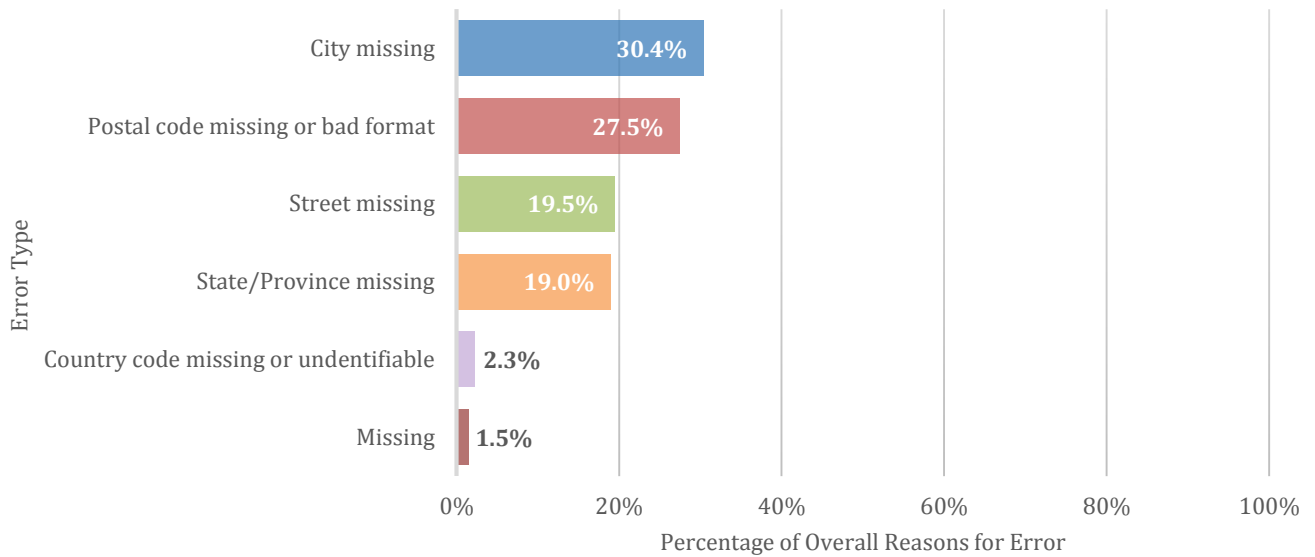
Similar to Phase 1, the largest source of errors among postal addresses was that the address was missing at least one required component such as city (30.4 percent of all telephone syntax errors), postal code (27 percent), street (19.5 percent), and/or state/province (19.0 percent). Fewer country codes were missing (only 2.3 percent of all postal address syntax errors) and few postal addresses were altogether missing (only 1.5 percent of all postal address syntax errors).

Operability Accuracy – 2009 RAA Requirements

The following section reviews the results of the operability accuracy tests against 2009 requirements by first looking at overall accuracy, then subgroup accuracy, followed by the reasons for error.

Overall Accuracy – 2009 RAA Operability Requirements

First, we look at accuracy to 2009 RAA requirements for all 10,000 domains. Above, we saw that syntax accuracy rates were lowest for postal address. However, for operability, accuracy rates are highest for postal address, shown below in Graph 7 and Table 9.[42] Operability accuracy rates are lower than syntax accuracy rates for email addresses and telephone numbers. More details about the operability accuracy testing can be found in Appendix A.

Graph 7: Overall Accuracy – 2009 RAA Operability Requirements

---
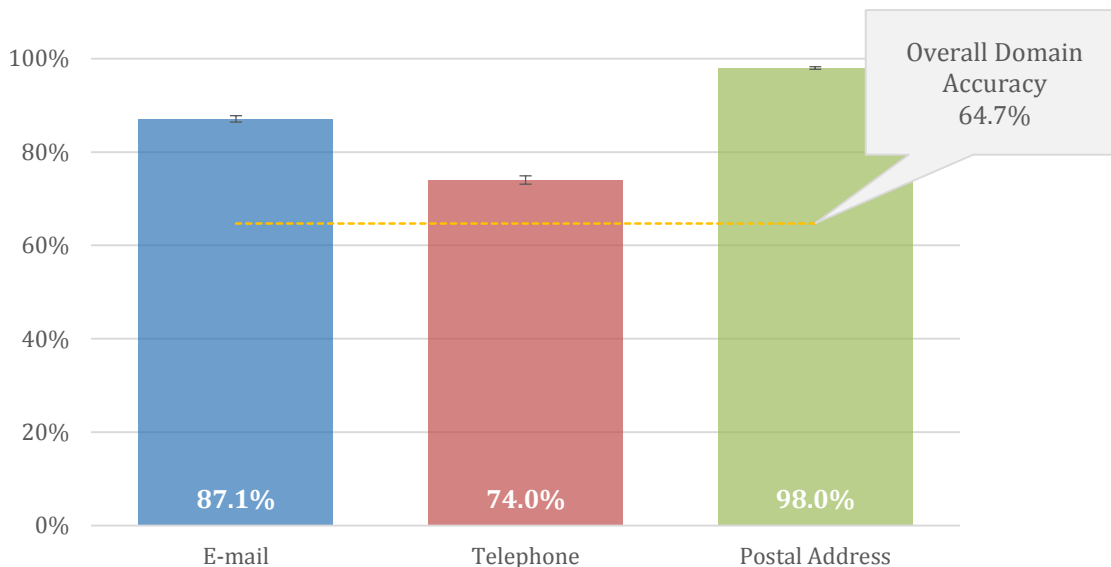
[42] As stated above, in some cases, postal address operability testing can be more forgiving than postal address syntax testing. For example, syntactic conformance with UPU standards for postal mail requires an abbreviation for state or province (e.g., 'DE' would conform, while 'Delaware' would not), but these syntax elements are not necessarily based on whether the parcel could be delivered. That is, where syntax accuracy is an indicator of operability for email address and telephone numbers, it is not for postal addresses. A syntactically inaccurate email address (e.g., without the "@") will not be operable; a syntactically inaccurate postal address will likely be operable (i.e., deliverable).

Table 9 below shows in more detail the breakdown by contact type. Similar to Table 5, the bottom row shows the rate for which the registrant, administrative, and technical contacts all pass. Just as it was for syntax, all four rows are very similar because of the commonality of the contact data, suggesting that looking at the registrant, administrative and technical accuracy rates separately is unnecessary. We will again focus on the percentages that all three contacts pass all tests (the "All 3 Accurate row") in the subgroup analyses.

Table 9: Overall Accuracy by Contact Type and Mode – 2009 RAA Operability Requirements

	Email	Telephone	Postal Address	ALL 3 Accurate
Registrant	88.9% ± 0.6%	78.4% ± 0.8%	98.2% ± 0.3%	**68.4% ± 0.9%**
Administrative	89.3% ± 0.6%	77.0% ± 0.8%	98.2% ± 0.3%	**69.0% ± 0.9%**
Technical	89.9% ± 0.6%	77.2% ± 0.8%	98.3% ± 0.3%	**69.7% ± 0.9%**
All 3 Accurate	**87.1% ± 0.7%**	**74.0% ± 0.9%**	**98.0% ± 0.3%**	**64.7% ± 0.9%**

Subgroup Accuracy – 2009 RAA Operability Requirements
Next, we look at subgroups for operability in Phase 2, starting with Prior versus New gTLDs. Because the numbers for registrant, administrative, and technical are so similar[43], we present here subgroup accuracy only for "All 3 Accurate", i.e., the registrant, administrative, and technical contacts all passed all of the accuracy tests.

Subgroup 1: Prior vs. New gTLDs
Graph 8 and Table 10 below show that Prior gTLDs have lower accuracy on email addresses, but higher accuracy on telephone numbers for operability. Prior gTLDs also have a higher rate of having "All 3" contact fields accurate.

---

[43] *See* section Accounting for Common Data across Contact Types.

Graph 8: Accuracy by gTLD type – 2009 RAA Operability Requirements
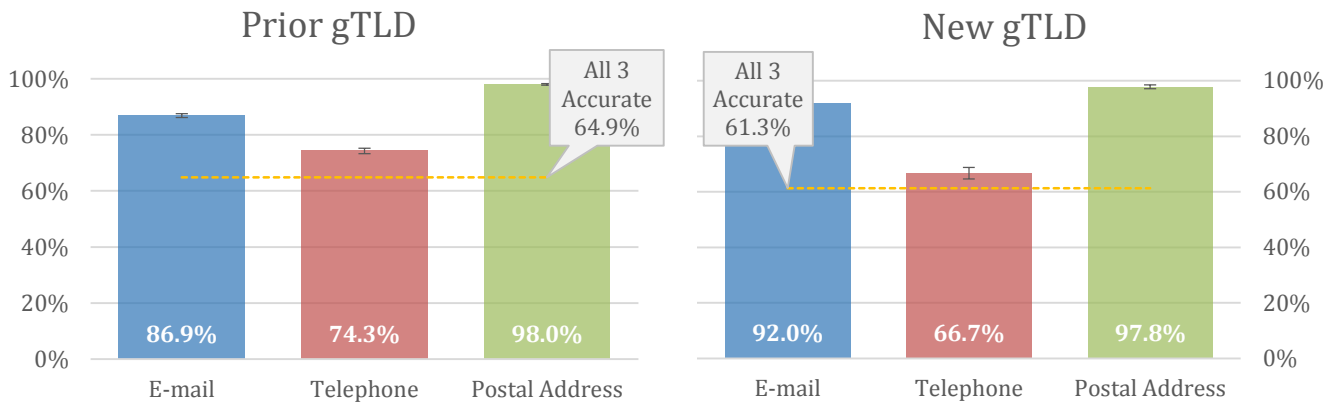
Table 10: Accuracy by gTLD type – 2009 RAA Operability Requirements

	Email	Telephone	Postal Address	ALL 3 Accurate
Prior gTLD	86.9% ± 0.7%	74.3% ± 1.0%	98.0% ± 0.3%	**64.9% ± 1.0%**
New gTLD	92.0% ± 1.2%	66.7% ± 2.1%	97.8% ± 0.7%	**61.3% ± 2.2%**
All 3 Accurate	**87.1% ± 0.7%**	**74.0% ± 0.9%**	**98.0% ± 0.3%**	**64.7% ± 0.9%**

Subgroup 2: ICANN Region
Next, we look at accuracy by ICANN region. Again, we present subgroup accuracy for the registrant, administrative, and technical contacts all passing the operability accuracy tests. Graph 9 and Table 11 below show that for email addresses, African and Latin America/Caribbean domains have higher operability accuracy rates than the other regions. For telephone numbers, Latin American/Caribbean and North American domains have higher operability accuracy rates. For postal addresses, African and Asian-/Pacific domains have lower operability accuracy rates than the other regions. With regard to all nine contacts passing all accuracy tests, Latin American/Caribbean and North American domains have higher rates among the regions, and Asian-/Pacific domains have a lower rate.

Graph 9: Accuracy by ICANN Region – 2009 RAA Operability Requirements

- ■ E-mail ■ Telephone ■ Postal Address
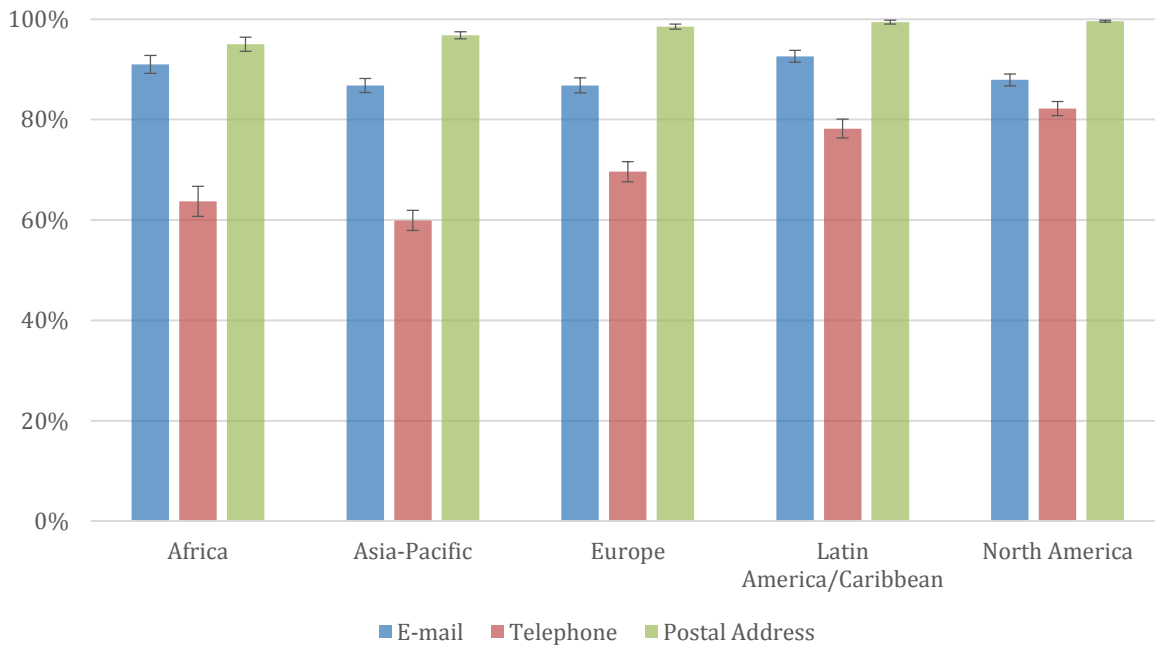
Table 11: Accuracy by ICANN Region – 2009 RAA Operability Requirements

	Email	Telephone	Postal Address	ALL 3 Accurate
Africa	91.0% ± 1.8%	63.7% ± 3.0%	95.0% ± 1.4%	**57.0% ± 3.1%**
Asia-Pacific	86.8% ± 1.4%	59.9% ± 2.0%	96.8% ± 0.7%	**49.4% ± 2.0%**
Europe	86.8% ± 1.5%	69.6% ± 2.0%	98.5% ± 0.5%	**59.8% ± 2.1%**
Latin America/Caribbean	92.6% ± 1.2%	78.2% ± 1.9%	99.4% ± 0.4%	**72.7% ± 2.0%**
North America	87.9% ± 1.2%	82.2% ± 1.4%	99.6% ± 0.2%	**73.2% ± 1.7%**
All 3 Accurate	**87.1% ± 0.7%**	**74.0% ± 0.9%**	**98.0% ± 0.3%**	**64.7% ± 0.9%**

Subgroup 3: RAA Status
Finally, we look at accuracy by RAA status. Again, we present subgroup accuracy for "All 3 Accurate". Graph 10 and Table 12 below show that, compared to the other groups, the 2013 RAA NGF group has higher email address accuracy while the 2009 RAA group has lower email address accuracy. The 2009 RAA group has higher telephone accuracy while the 2013 RAA GF group has lower telephone number accuracy compared to the other RAA Status groups. The 2013 RAA GF group has higher postal address accuracy while the 2009 RAA group has lower postal address accuracy. Since this group is highest in email and not lowest in any, the 2013 RAA NGF group has a higher percentage of "all 3" being accurate.

Graph 10: Accuracy by RAA Status – 2009 RAA Operability Requirements

Table 12: Accuracy by RAA Status – 2009 RAA Operability Requirements

	Email	Telephone	Postal Address	ALL 3 Accurate
2009 RAA	77.2% ± 1.7%	82.2% ± 1.6%	95.9% ± 0.8%	**61.7% ± 2.0%**
2013 RAA GF	85.4% ± 1.1%	72.0% ± 1.4%	98.4% ± 0.4%	**62.0% ± 1.5%**
2013 RAA NGF	91.2% ± 0.9%	76.9% ± 1.3%	97.4% ± 0.5%	**70.3% ± 1.5%**
All 3 Accurate	**87.1% ± 0.7%**	**74.0% ± 0.9%**	**98.0% ± 0.3%**	**64.7% ± 0.9%**

Reasons for Error – 2009 RAA Operability Requirements

For operability, the reasons for error were straightforward because the tests for email addresses, telephone numbers, and postal addresses were all sequential. If a test failed, operability failed. If a test succeeded, the contact information passed onto the next test.

Email Addresses

Table 13 below shows that around 10 percent of the records received a "bounced back" email, revealing that the email address was not operable. Once again, a registrant email address is not required under the 2009 RAA. If a verifiable email address was given, an attempt to send an email was made. If the connection succeeded, the email address was deemed operable. The required emails were not given only about 0.3 percent of the time for the administrative and technical fields.

Table 13: Email Address Errors by Contact Type – 2009 RAA Operability Requirements

	Administrative	Technical	Registrant	All 3 total
Passed all accuracy tests	8,937	8,999	8,939	**26,875**
Not Verifiable (or Missing)	53	65	[125]*	**118**
Email Bounced	1,010	936	1,061	**3,007**
Total	**10,000**	**10,000**	**10,000**	**30,000**

* Registrant email is not required under the 2009 RAA.

Telephone Numbers

Table 14 shows the operability errors for telephone number. Once again, a registrant telephone number is not required under the 2009 RAA. If a verifiable telephone number was given, it was dialed. Table 14 shows that approximately 6.3 percent of the numbers were disconnected, another 7.5 were invalid, and another 8.7 percent did not connect. The percentage of required telephone numbers that were missing was only 1.4 percent.

Table 14: Telephone Number Errors by Contact Type – 2009 RAA Operability Requirements

	Administrative	Technical	Registrant	All 3 total
Passed all accuracy tests	7,558	7,689	7,713	**22,960**
Not Verifiable (or Missing)	139	146	[202]*	**285**
Number Disconnected	630	623	632	**1,885**
Invalid Number	774	729	746	**2,249**
Other Not Connected	899	813	909	**2,621**
Total	**10,000**	**10,000**	**10,000**	**30,000**

* Registrant telephone number is not required under the 2009 RAA.

Postal Addresses

Finally, Table 15 shows the postal address errors for operability. As explained in Appendix A, the postal addresses are first coded for operability automatically by the UPU's operability testing tool.[44] V (Verified as is), C (Corrected and Verified), and P3 (Probably Deliverable) are all considered operable postal addresses. P2 (Might or Might Not be Deliverable), P1 (Probably Not Deliverable), N1 (Country Unknown), and N2 (Unverifiable Address due to Country Standards Not Available) are all considered inoperable postal addresses. However, a manual process[45] is carried out for the P1 and P2 addresses, allowing most postal addresses to be identified as operable. Table 15 shows that 93.3 percent of the P2 addresses and 88.9 percent of the P1 addresses were determined to be operable manually. The one N2 record was considered to be inoperable.

---

[44] The UPU, accuracy testing vendor for postal addresses, simulates post office protocols for handling a parcel that is being sent to the postal address provided in the record, but without attempting physical delivery to the destination.

[45] This manual process is also intended to simulate post office protocols, in which it may be attempted to determine the correct address from the correct components provided in an address that on a whole might be considered incorrect. As can be seen in the table, not all addresses falling in this category may be corrected, i.e., deemed operable.

Table 15: Postal Address Errors by Contact Type – 2009 RAA Operability Requirements

	Administrative	Technical	Registrant	All 3 total
Operable	8,537	8,619	8,533	**25,689**
Operable P2	777	743	780	**2,300**
Operable P1	448	404	447	**1,299**
TOTAL OPERABLE	**9,762**	**9,766**	**9,760**	**29,288**
Inoperable P2	56	52	57	**165**
Inoperable P1	72	64	80	**216**
N1, No Country	109	118	102	**329**
N2, Unverifiable	1	0	1	**2**

Additional Findings

Statistical comparisons can be made between Phase 1 syntax accuracy and Phase 2 syntax accuracy[46], but any changes (i.e., improvements) cannot be said to be directly correlated to the WHOIS ARS due to the timing of when Phase 2 data was pulled (i.e., in the middle of Phase 1). However, we present the comparisons below for informational purposes, as looking at changes from phase to phase is still useful in better understanding the entire domain name universe. In addition, from the above findings, some general observations can be made about the relationship between syntax and operability accuracy. The following section details these observations.

Comparisons of Syntax Accuracy between Phases

We have presented above the syntax accuracy of records against 2009 RAA Requirements for Phase 2, Cycle 1. Here, we compare the Phase 2 results to those from Phase 1.

Table 16 below shows that Phase 2 has a lower[47] overall telephone accuracy rate, but that the email and postal address accuracy rates are very similar between the phases. The most likely explanation for the difference between Phase 1 and Phase 2 is random variation.[48] It is also plausible that the distribution has changed, but not much time has passed between Phase 1 and Phase 2. A change in distribution would imply that the newest domains not subjected to Phase 1 sampling have lower accuracy in terms of telephone numbers.

Table 16: Overall Accuracy by Phase – 2009 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	99.2% ± 0.1%	85.8% ± 0.7%	79.1% ± 0.8%	70.3% ± 0.9%
Phase 2	99.1% ± 0.2%	83.3% ± 0.7%	79.4% ± 0.8%	67.2% ± 0.9%
Change (2 - 1)	**-0.1% ± 0.2%**	**-2.5% ± 1.0%**	**0.3% ± 1.1%**	**-3.1% ± 1.3%**

Since most of the domains in the domain universe are from Prior gTLDs, the patterns for the Prior gTLDs are the same as "overall", as seen below in Table 17. That is, the same drop in telephone number accuracy is seen for Prior gTLDs as in Table 16. Similarly, there is no change in email and postal address accuracy.

---

[46] As a reminder, we cannot compare operability accuracy from Phase 1 to Phase 2 because Phase 1 focused solely on syntax accuracy. We plan to compare operability accuracy from Phase 2 Cycle 1 to Phase 2 Cycle 2 in the next WHOIS ARS report, expected in June 2016.

[47] *See* footnotes 38 and 39.

[48] Since we are using 95 percent confidence intervals, there is still a 5 percent chance we would show a significant difference even if there is no difference.

Table 17: Prior gTLDs Accuracy by Phase – 2009 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	99.2% ± 0.2%	85.8% ± 0.8%	79.3% ± 0.9%	70.4% ± 1.0%
Phase 2	99.1% ± 0.2%	83.0% ± 0.8%	79.8% ± 0.9%	67.4% ± 1.0%
Change (2 - 1)	-0.1% ± 0.3%	-2.8% ± 1.1%	0.5% ± 1.3%	-3.0% ± 1.4%

In Table 18, the New gTLDs show different patterns. Again, there is no change for email addresses, but Phase 2 shows an increase in telephone number accuracy and a decrease in postal address accuracy. Phase 2 also shows a decrease in the percentage of domains in New gTLDs that pass all accuracy tests for all nine contacts. While it is difficult to pinpoint an exact reason for this difference, one possible explanation could be a different geographical spread of the domains in New gTLDs versus those in the Prior gTLDs.[49]

Table 18: New gTLDs Accuracy by Phase – 2009 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	99.8% ± 0.2%	86.2% ± 1.5%	71.6% ± 2.0%	68.8% ± 2.0%
Phase 2	99.9% ± 0.1%	89.4% ± 1.4%	67.5% ± 2.1%	64.6% ± 2.2%
Change (2 - 1)	0.1% ± 0.2%	3.2% ± 2.1%	-4.1% ± 2.9%	-4.2% ± 3.0%

Relationship between Syntax and Operability Accuracy – 2009 RAA Requirements

We have above presented syntax and operability accuracy for all 10,000 domains based on the 2009 RAA standards. The following three tables show the relationship between syntax and operability accuracy against 2009 RAA standards. Each table seeks to answer the question "What rate of records pass the operability accuracy tests that also passed the syntax accuracy tests?" or vice versa.

Table 19 shows that email addresses that do not pass the syntax accuracy tests also fail the operability accuracy test (i.e., zero percent fail syntax and then pass operability accuracy tests). However, almost all of the email addresses that fail operability accuracy tests (12.9 percent of all domains) pass the syntax accuracy tests (12.0 percent of the 12.9 percent). Another way to look at Table 19 is that 12 percent of email addresses pass the syntax accuracy tests, but are inoperable.

---

[49] See Appendix B or the Phase 1 report for more information on results, especially by region.

Table 19: Syntax and Operability Accuracy for Email Addresses – 2009 RAA Requirements

		Operability		
		Pass	Fail	TOTAL
Syntax	Pass	87.1% ± 0.7%	12.0% ± 0.6%	**99.1% ± 0.2%**
	Fail	0.0% ± 0.0%	0.9% ± 0.2%	**0.9% ± 0.2%**
	TOTAL	**87.1% ± 0.7%**	**12.9% ± 0.7%**	**100%**

Table 20 shows that 16.7 percent of telephone numbers fail the syntax accuracy tests while 26.0 percent fail the operability accuracy tests. However, these groups do not fully overlap. Of the telephone numbers that do not pass the syntax accuracy tests, (16.7 percent total), most also fail the operability test (12.3 percent out of the 16.7 percent, or 74 percent). Of those that fail the operability test, however (26.0 percent total), about half (13.7 percent out of the 26.0 percent, or 53 percent) pass the syntax accuracy tests.

Table 20: Syntax and Operability Accuracy for Telephone Numbers – 2009 RAA Requirements

		Operability		
		Pass	Fail	TOTAL
Syntax	Pass	69.6% ± 0.9%	13.7% ± 0.7%	**83.3% ± 0.7%**
	Fail	4.4% ± 0.4%	12.3% ± 0.6%	**16.7% ± 0.7%**
	TOTAL	**74.0% ± 0.9%**	**26.0% ± 0.9%**	**100%**

Table 21 shows that postal addresses that do not pass operability accuracy tests also fail the syntax test (i.e., zero percent fail operability accuracy tests but pass syntax accuracy tests). However, almost all of the postal addresses that fail syntax accuracy tests (20.6 percent of all domains) pass the operability accuracy tests (18.6 percent of the 20.6 percent). Another way to look at Table 21 is that 18.6 percent of postal addresses fail the syntax accuracy tests, but pass the operability accuracy tests.

Table 21: Syntax and Operability Accuracy for Postal Addresses – 2009 RAA Requirements

		Operability		
		Pass	Fail	TOTAL
Syntax	Pass	79.4% ± 0.8%	0.0% ± 0.0%	**79.4% ± 0.8%**
	Fail	18.6% ± 0.8%	2.0% ± 0.3%	**20.6% ± 0.8%**
	TOTAL	**98.0% ± 0.3%**	**2.0% ± 0.3%**	**100%**

These tables show that syntax accuracy is not the same as operability accuracy. On the one hand, for syntax, accuracy is highest for email addresses and lowest for postal addresses. On the other hand, for operability, accuracy is highest for postal addresses and lowest for telephone numbers. For email

addresses, syntax accuracy is necessary for operability accuracy, but is not a guarantee of operability. For postal addresses, syntax accuracy is not a necessary condition for operability accuracy. However, syntax accuracy seems to guarantee operability accuracy. For telephone numbers, the relationship is more complicated since 18 percent (13.7 percent plus 4.4 percent) are non-conforming in syntax *or* operability, but not both. We can thus conclude that where syntax accuracy is an indicator of operability for email address and postal addresses, it is not for telephone numbers. That is, a syntactically inaccurate email address (e.g., without the "@") will not be operable and a syntactically accurate postal address will be operable (i.e., deliverable), but the syntax accuracy of a telephone number could have operability accuracy or not.

Challenges and Lessons Learned

Continually seeking ways to improve the ARS and looking ahead to subsequent ARS reports, this section provides background on the issues that created challenges in Phase 2 Cycle 1 and how those issues can be avoided or improved upon in subsequent reports. Similarly, we pose a few questions that the WHOIS ARS team will be tackling as it moves on to Phase 2 Cycle 2.

■ **Adjustments required for Email Operability Testing:**

A number of technical tests were devised to check the operability of the email addresses in our subsample. One test, for example, checked whether an MX record existed for a particular email address; others attempted to verify server existence. In the end, it was determined the most reliable test was the "bounce-back" test and in the midst of the data analysis stage of Phase 2 Cycle 1, re-testing was conducted in order to ensure accurate results based on this test only. In addition, attempts to circumvent "spam traps" proved unsuccessful and adjustments were made to prevent this issue.

■ **Adjustments required for Telephone Operability Testing:**

As with email operability, a number of technical tests were devised to check the operability of telephone numbers by attempting to establish a connection. In some cases, issues emerged with the wait time required to make a connection to some numbers, which may have been related to the telephone service provider. Re-testing was conducted with an extended wait time to ensure adequate time to make a connection.

■ **Challenges related to Postal Address testing:**

As discussed in the Phase 1 report, syntax and operability testing of postal addresses pose unique challenges. As the WHOIS does not allow for separate fields for different address components based on a particular country, many times an address will be included all in one line. This result makes both automated syntax and operability testing difficult. Secondly, the syntactically inaccurate address may look operationally inaccurate for this reason, but in reality, it is likely the address is deliverable. Knowing this relationship, we devised operability tests that used both automated and manual checks to ensure that all potentially deliverable addresses were marked as such.[50]

■ **Increasing the sample size (again) will continue to improve statistical significance:**

The initial sample size increase from 100,000 to 150,000 helped with the smallest subgroups (Africa and 2009 RAA). A further increase to 200,000 is being considered. Further, the subgroups

---

[50] Indeed, the operability accuracy rates were much higher than syntax accuracy rates, thus showing that while formatting requirements are strict and the WHOIS allows for little variation by country, creating opportunity for many errors, many of the addresses in the WHOIS are likely to be deliverable as is.

of interest are the three RAA subgroups crossed with the five regions. An analysis subsample of 10,000 is barely enough to include 800 in each cell, so an increase to 12,000 is being considered.[51]

■ **Analyzing all three contact types may be redundant:**

From examining the commonality across contact types, we found in Phase 2 that the registrant, administrative, and technical contacts are identical in over 75 percent of the records. The accuracy testing vendors had already accounted for duplicates in their testing in order to reduce inefficiencies and redundancies, but going forward, perhaps testing only one contact (e.g., registrant) may be even more efficient. We could also test the contacts on a rotating basis (e.g., each cycle looks at a different contact), but this would result in comparability issues from cycle to cycle. While testing only one contact could potentially reduce inefficiency, it might be best to continue testing all three contact types, but account for the duplication in the analysis.
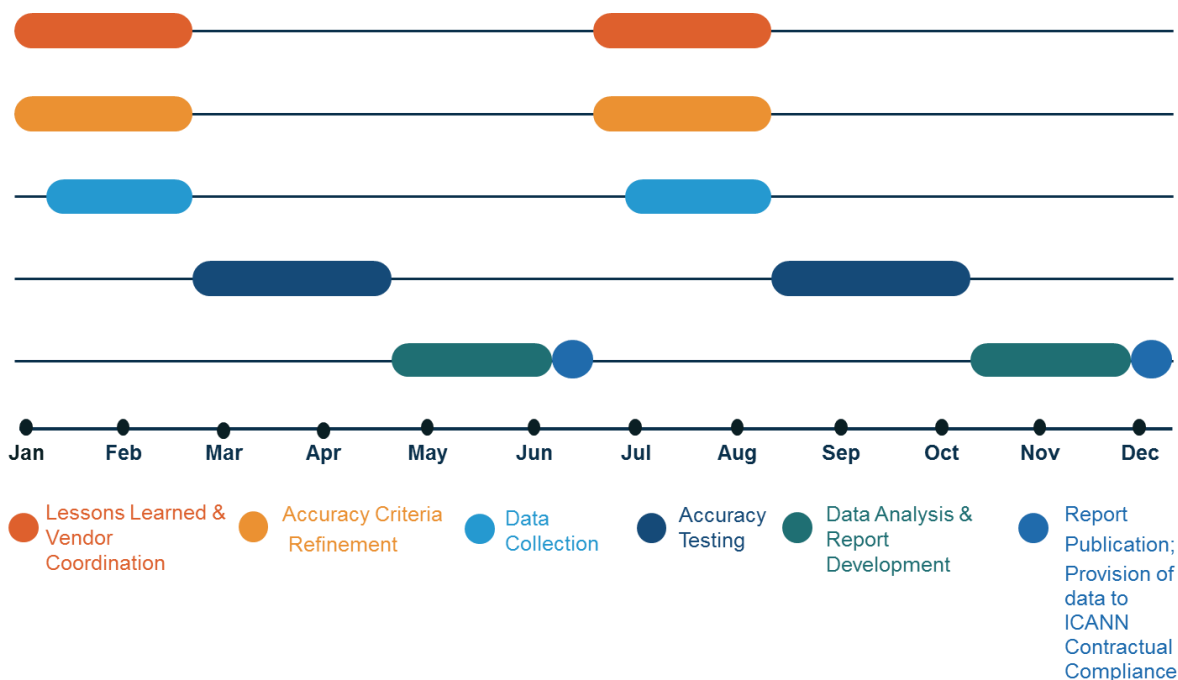
---

[51] *See* the section on Study Methods and Approach for more information on this issue.

Next Steps

Phase 2 Cycle 2 Look-Ahead

The next cycle of the WHOIS ARS Report (Phase 2 Cycle 2) will reprise the syntax and operability review of Phase 2 Cycle 1. The cycle will be performed in the same manner as the previous, except for those areas discussed in Challenges and Lessons Learned. Phase 2 Cycle 2 is scheduled to begin in January 2016, and the timeline below shows the stages for Phase 2 Cycle 2 up to report publication:

Figure 5: Phase 2 Cycle 2 (and follow-on)

| Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |

Legend:
- Lessons Learned & Vendor Coordination
- Accuracy Criteria Refinement
- Data Collection
- Accuracy Testing
- Data Analysis & Report Development
- Report Publication; Provision of data to ICANN Contractual Compliance

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 2 study and those potentially inaccurate records have already been provided to ICANN Contractual Compliance and are currently under review. The processing of new tickets based upon WHOIS ARS results will be staggered to minimize system performance issues and impact on registrars. WHOIS ARS tickets will be processed alongside other complaints; however, ICANN will continue to give priority to complaints submitted by community members.

As Phase 2 Cycle 1 includes both syntax and operability results, Compliance follow-up and investigation may be conducted through different processes depending on the type of inaccuracies found within the

record, e.g., those records with only formatting errors but that have been deemed "operable" will receive a different kind of notice than those records that have formatting errors and have been deemed "inoperable."

All WHOIS ARS tickets will follow the Contractual Compliance Approach and Process[52] according to the types of issues described below. When possible, and in consultation with registrars, ICANN may be able to consolidate multiple WHOIS ARS tickets during processing.

Syntax Inaccuracy Follow-up
WHOIS ARS complaints will be classified as WHOIS format errors if the error concerns non-compliance 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. WHOIS format errors will not be forwarded to registrars under the 2009 RAA.

Operability Inaccuracy Follow-up
WHOIS ARS complaints that are generated due to failures of operability will be processed as WHOIS inaccuracy complaints. While format issues may not require contact with registered name holders, operability failures indicate substantive inaccuracies that require registrars to take reasonable steps to investigate, and where applicable, correct the alleged inaccuracies under the 2009 and 2013 RAAs. Additionally, the WHOIS Accuracy Program Specification (WAPS) of the 2013 RAA has additional requirements, including validating format requirements and suspending a domain name for failure of the Registrant to respond timely to the WHOIS inaccuracy complaint.

Contractual Requirements
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) Specification [53] and Advisory: Clarifications to the Registry Agreement, and the 2013 Registrar Accreditation Agreement (RAA) regarding applicable Registration Data Directory Service (WHOIS) Specifications.[54]

---

[52] *See* ICANN Contractual Compliance Approach and Process: https://www.icann.org/resources/pages/approach-processes-2012-02-25-en.

[53] *See* https://www.icann.org/resources/pages/approved-with-specs-2013-09-17-en#whois.

[54] *See* https://www.icann.org/resources/pages/registry-agreement-raa-rdds-2015-04-27-en.

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

Appendix A: Accuracy Testing Criteria

ICANN has attempted to align the accuracy testing criteria with the contractual obligations 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 elements of contact information for the registrant, the technical contact (technical) and the administrative contact (administrative) 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 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
Validation criteria to be in testing	2009 RAA Requirements

Record Created	20 Apr 2014
Registrar's 2013 RAA Effective Date	01 Jan 2014
Validation criteria to be in testing	2013 RAA Requirements

Below you will find an overview of syntax and operability accuracy testing criteria for email addresses, telephone numbers and postal addresses. These criteria will be used by the validation vendors supporting the WHOIS ARS project.

Email Addresses

The syntactical criteria 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. Syntactically correct, verified email addresses do not guarantee email box existence, so while syntactically incorrect email addresses may indicate automatic failures, syntactically correct email addresses will also be subjected to operability testing, explained further below.

Email Address Syntax: Stage One
In the 2009 RAA, the presence of an administrative and a technical email address is required. The presence of a registrant email address is optional. In the 2013 RAA, the administrative, technical, 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/registrar. 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. **Technical email address**

      ▪ Yes – Pass

      ▪ No – Fail

   c. **Administrative email address**

      ▪ Yes – Pass

      ▪ No – Fail

Email Address Syntax: Stage Two

The Syntactical 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 5322)

   a. **Registrant,** b. **Technical, and** c. **Administrative email addresses**

      ▪ Yes – Pass

      ▪ No – Fail

2. **Is there presence of an "@" symbol in the email address?**

   a. **Registrant,** b. **Technical, and** c. **Administrative email addresses**

      ▪ Yes – Pass

      ▪ No – Fail

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

(i.e., the characters following the "@" symbol)

a. Registrant, b. Technical, and c. Administrative 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. Technical, and c. Administrative 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. Technical, and c. Administrative email addresses

- Yes – Pass
- No – Fail

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

(i.e., the characters preceding the "@" symbol)

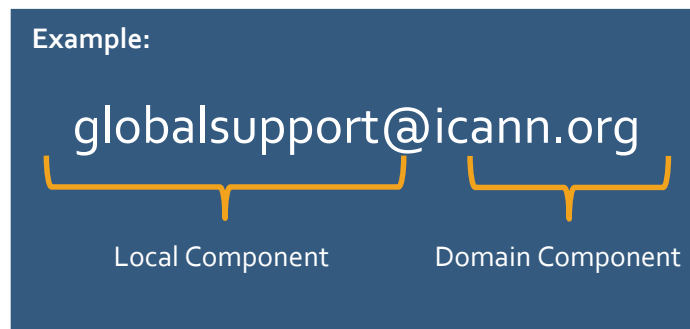**a. Registrant, b. Technical, and c. Administrative email addresses**

- Yes – Pass
- No – Fail

7. **Is the local component syntactically valid?**

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

   a. **Registrant,** b. **Technical, and** c. **Administrative email addresses**

   - Yes – Pass
   - No – Fail

<div style="background-color:#1f4e6b; color:white; padding:20px;">

Example:

globalsupport@icann.org

Local Component   Domain Component

</div>

Email Address Operability[55]
The operational criteria tests for email address are a combination of direct (i.e., sending an email) and indirect (i.e., server communication) methods to validate an email address.

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/registrar.

1. **Does the email address' domain have at least one MX record?**

   a. **Registrant,** b. **Technical, and** c. **Administrative email addresses**

   - Yes – Pass
   - No – Fail

2. **Is a viable connection made to the mail server?**

   a. **Registrant,** b. **Technical, and** c. **Administrative email addresses**

   - Yes – Pass
   - No – Fail

---

[55] Tests 1 and 2 were later determined to be not as useful indicators as Test 3. Though they were still used for informational purposes, Test 3 was the sole indicator for email operability in the final results. See Lessons Learned for more information.

3. **Is there absence of a bounce email within one (1) day after email issuance?**

   a. **Registrant,** b. **Technical, and** c. **Administrative email addresses**

   - Yes – Pass
   - No – Fail

Email Address Operability Template:

Please do not respond to this email. This email is part of a survey by the Internet Corporation for Assigned Names and Numbers (ICANN) related to contact information in the WHOIS records of domain names. No action is required. For more information please visit, whois.icann.org.

Thank you,
WHOIS Accuracy Reporting System Project

Telephone Numbers

The syntax accuracy tests for telephone numbers are organized into stages: stage one will verify the presence of a telephone number, as required by the applicable RAA, and stage two will involve detailed technical testing of the telephone number syntax. Syntactically correct, verified telephone numbers do not guarantee existence or operability of the phone number, and incorrect syntax does not guarantee the number is not in operation, so all telephone numbers will be subjected to both syntax and operability testing, explained further below.

Telephone Number Syntax: Stage One
In the 2009 RAA, presence of administrative and technical telephone numbers is required; presence of a registrant telephone number is optional. In the 2013 RAA, the administrative, technical, 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/registrar. 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. **Technical phone number**
      - Yes – Pass
      - No – Fail

   c. **Administrative phone number**
      - Yes – Pass
      - No – Fail

Telephone Number Syntax: Stage Two
The Syntactical 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 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. Technical, and c. Administrative phone number

   - Yes – Pass
   - No – Fail

2. **Is the country code syntactically valid?**

   (i.e., meets the requirements as specified in RFC5733, +###.)

   a. Registrant, b. Technical, and c. Administrative phone number

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

3. **Does the phone number contain at least the minimum allowed digits based on the country code?**

   a. Registrant, b. Technical, and c. Administrative 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. Technical, and c. Administrative 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. Technical, and c. Administrative phone number

   - Yes – Pass
   - No – Fail

6. **Does the phone number only contain permissible numbers and formatting characters?**

   a. **Registrant,** b. **Technical, and** c. **Administrative phone number**

   - Yes – Pass
   - No – Fail

7. **IDENTIFIER – Is there presence of an extension?**

   a. **Registrant,** b. **Technical, and** c. **Administrative 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. **Technical, and** c. **Administrative phone number**

   - Yes – Pass
   - No – Fail

9. **Is the extension syntactically valid?**

   (i.e., "x" to attribute the telephone extension: RFC5733)

   a. **Registrant,** b. **Technical, and** c. **Administrative phone number**

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

Telephone Number Operability

As previously mentioned, all telephone numbers will be subjected to both syntax and operability testing. Telephone number operability testing involves a series of tests to determine if the connection exists and is working. The operability test is performed by attempting to place a call to the number listed in the WHOIS record.

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 the contact field. A missing Registrant telephone number subject to the 2009 RAA will be noted, but not counted against the domain/registrar. A "Yes" response will initiate Stage Two testing.

1. **Does the phone number connect (i.e., provide a ring tone, busy signal, or an answer)**
   a. **Registrant,** b. **Technical, and** c. **Administrative phone number**
   - Yes – Pass
   - No – Fail

2. **Is there absence of a disconnected message?**
   a. **Registrant,** b. **Technical, and** c. **Administrative phone number**
   - Yes – Pass
   - No – Fail

3. **Is there absence of an invalid number error?**
   a. **Registrant,** b. **Technical, and** c. **Administrative phone number**
   - Yes – Pass
   - No – Fail

Telephone Number Operability Response Message, if answered

"Hello. We're working with ICANN, the Internet Corporation for Assigned Names and Numbers to evaluate the functionality of phone numbers in WHOIS records for domain names. No action is required on your part. Thank you for verifying the phone number as operational."

Postal Addresses

The syntax accuracy tests for postal addresses are organized into stages: stage one will verify the presence of a postal address, and stage two will involve detailed technical testing of the postal address syntax. Syntactically correct postal addresses do not guarantee operability of the postal address, and reversely, syntactically incorrect postal addresses do not guarantee inoperability. All postal addresses will thus be subjected to both syntax and operability testing, explained further below.

Postal Address Syntax: Stage One

In the 2009 RAA and 2013 RAA, presence of a registrant, administrative and a technical 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. **Technical, and** c. **Administrative postal address**

   - Yes - Pass
   - No – Fail

Postal Address Syntax: Stage Two

The Syntactical 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?**

   a. **Registrant,** b. **Technical, and** c. **Administrative 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. **Technical, and** c. **Administrative postal address**

   - Yes – Pass

- No – Fail

3. **Is the country provided in the Country field?**

    a. **Registrant,** b. **Technical, and** c. **Administrative 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. **Technical, and** c. **Administrative postal address**

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

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

    a. **Registrant,** b. **Technical, and** c. **Administrative postal address**

    - Yes – Proceed to additional postal code validation
    - No – Appropriately left blank, move to next field (i.e., Test 9)

    6. **Is there presence of a postal code?**

        a. **Registrant,** b. **Technical, and** c. **Administrative postal address**

        - Yes – Pass
        - No – Fail

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

        a. **Registrant,** b. **Technical, and** c. **Administrative 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. **Technical, and** c. **Administrative postal address**

        - Yes – Pass

- No – Fail

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

    a. **Registrant,** b. **Technical, and** c. **Administrative 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. **Technical, and** c. **Administrative postal address**

        - Yes – Pass
        - No – Fail

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

        a. **Registrant,** b. **Technical, and** c. **Administrative 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. **Technical, and** c. **Administrative postal address**

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

13. **Is there presence of a city?**

    a. **Registrant,** b. **Technical, and** c. **Administrative postal address**

    - Yes – Pass
    - No – Fail

14. **Is the city in the City field?**

    a. **Registrant,** b. **Technical, and** c. **Administrative postal address**

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

15. **Is there presence of a street?**

    a. Registrant, b. Technical, and c. Administrative postal address

- Yes – Pass
- No – Fail

16. **Is the street in the Street field?**

    a. Registrant, b. Technical, and c. Administrative postal address

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

Postal Address Operability

As mentioned above, all postal addresses will be subjected to both syntax and operability testing. Operability testing is mostly automated through the use of a tool employed by the Universal Postal Union (UPU), ICANN's postal address validation vendor. No mail will be sent as part of the testing.

The tool uses available country reference data to determine whether an address, based on the given components as provided in the WHOIS record, might be deliverable. The tool returns a "deliverability" code specifying how likely deliverable (if at all) an address is. Any address that receives a "likely deliverable" code is considered a "Pass" for operability. Any address without a country is considered a "Fail" for operability, as a country is required to check against the appropriate reference data. If the tool may return a "not likely deliverable" or "no reference data for given country" code, the UPU has developed rules for complementary manual tests to further assess the deliverability of an address. Any address that is neither verified by the tool nor the manual tests will be considered a "Fail" for operability.

Note: There is no difference between 2009 and 2013 RAA verification requirements. All records will be given the same tests.

Postal Address Operability: Automated Stage

The UPU attempts to verify the deliverability of the address automatically via the tool.

A "Yes" for Stage One will be considered a "Pass" for operability (i.e., deliverable). A "No" for Stage One will trigger either a "Fail" or a follow-on testing scenario.

1. **Is the tool able to automatically verify deliverability?**

    (i.e., returns a "deliverable" code)

    a. Registrant, b. Technical, and c. Administrative postal address

- Yes – Pass

- No – Fail: No country present in address

- No – Possible Fail (A): Tool has no reference data for given country; Proceed to Manual Stage (A)

- No – Possible Fail (B): Ambiguity or incomplete data; Proceed to Manual Stage (B)

Postal Address Operability: Manual Stage (A)

The operability test in Manual Stage (A) is performed on all contact fields that attained a "Possible Fail (A)" from the Automated Stage above. Here the UPU conducts manual testing of the given address using its own complementary data on the given country to determine if the address is likely to be deliverable.

A "Yes" response for Manual Stage (A) will be considered a "Pass" for operability (i.e., deliverable). A "No" for Manual Stage (A) will be considered "Indeterminate" for operability (i.e., unable to verify automatically or manually the address deliverability)

1. **Based on UPU's manual check of the address, is the address likely deliverable?**

   a. **Registrant,** b. **Technical, and** c. **Administrative postal address**

   - Yes – Pass

   - No – Indeterminate

Postal Address Operability: Manual Stage (B)

The operability test in Manual Stage (B) is performed on all contact fields that attained a "Possible Fail (B)" from the Automated Stage above. Here the UPU checks the available reference data for the given address components to determine if the address is likely to be deliverable. Two components state/province (if applicable) and city, are essential to the UPU's ability to complete this test; in the absence of these two components, an address will not be considered deliverable.

A "Yes" response to all three criteria for Manual Stage (B) will be considered a "Pass" for operability (i.e., deliverable). A "No" response to at least one will be considered a "Fail" for operability (i.e., not deliverable).

1. **Is there presence of a city in the given address?**

   (i.e., field is not blank)

   a. **Registrant,** b. **Technical, and** c. **Administrative postal address**

   - Yes – Pass

   - No – Fail

2. **If Postal Address Syntax Identifier Test #9 = yes, is there presence of a state/province in the given address?**

   (i.e., field is not blank)

   a. **Registrant,** b. **Technical, and** c. **Administrative postal address**

   - Yes – Pass

- No – Fail

3. **Based on UPU's manual check of the given address components, is the address likely to be deliverable?**

   a. Registrant, **b.** Technical, and **c.** Administrative postal address

   - Yes – Pass
   - No – Fail

Appendix B: Additional Analysis – 2009 RAA Requirements

Commonality of Contact Data

Table B1 shows that when two of the three contact types are identical (and one is different), it is most likely to be the registrant and administrative that agree, and least likely to be the registrant and technical that agree.

Table B1: Frequency of common contact information across contact type and mode

Commonality	Email	Telephone	Postal Address
All 3 Exactly Same	76.1% ± 0.8%	79.4% ± 0.8%	76.2% ± 0.8%
Registrant=Administrative	14.7% ± 0.7%	14.5% ± 0.7%	14.3% ± 0.7%
Registrant=Technical	0.7% ± 0.2%	0.5% ± 0.1%	0.5% ± 0.1%
Administrative=Technical	6.0% ± 0.5%	4.4% ± 0.4%	6.6% ± 0.5%
All 3 Different	2.4% ± 0.3%	1.2% ± 0.2%	2.5% ± 0.3%

2009 RAA Reasons for Syntax Error in Phases 1 and 2

The main results section contains the Phase 2 results, but here in Appendix B, we present Phase 1 and 2 results side-by-side for comparison.

Table B2: Total Email Address Errors by Contact Type (2009 RAA) – Phase 1

	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**

* Registrant email is not required under the 2009 RAA.

Table B3: Total Email Address Errors by Contact Type (2009 RAA) – Phase 2

	Administrative	Technical	Registrant	All 3 total
Passed all accuracy tests	9,945	9,933	9,997	**29,875**
Missing*	51	62	[124]*	**113**
@ Missing	2	3	2	**7**
Not Resolvable	2	2	1	**5**
Total	**10,000**	**10,000**	**10,000**	**30,000**

* Registrant email is not required under the 2009 RAA.

* The missing registrant email addresses passed the accuracy check since presence of the registrant email address is not required.

Seen in Tables B2 and B3, in Phase 1, no errors were ever found in the email addresses except if a required email address was missing (the registrant email address is not required under the 2009 RAA). In Phase 2, however, a very small number of email addresses were found to either have the "@" symbol missing or to not be resolvable into a valid-looking email address. Also, the number of missing registrant email addresses is much larger than in Phase 1 (even though it is not required).

Table B4: Total Telephone Number Errors by Contact Type (2009 RAA) – Phase 1

	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	33	33	**97**
Total	**10,000**	**10,000**	**10,000**	**30,000**

* Registrant telephone number is not required under the 2009 RAA.

Table B5: Total Telephone Number Errors by Contact Type (2009 RAA) – Phase 2

	Administrative	Technical	Registrant	All 3 total
Passed all accuracy tests	8,409	8,523	8,569	**25,501**
Missing*	137	144	[199]*	**281**
Country Code Missing	499	481	474	**1,454**
Incorrect Length	952	849	955	**2,756**
Unallowed Characters	3	3	2	**8**
Total	**10,000**	**10,000**	**10,000**	**30,000**

* Registrant telephone number is not required under the 2009 RAA.

As indicated above, there is a decrease in accuracy for telephone numbers in Phase 2, seen here in Table B5. The largest difference is in the country code missing. In Phase 1, there were 872 (2.9 percent) missing country codes, but this has increased to 1,454 (4.8 percent) in Phase 2. There is a decrease in the number of domains with unallowed characters from 97 (0.3 percent) in Phase 1 to 8 in Phase 2. There was a slight decrease in missing registrant telephone numbers from 234 (0.8 percent) to 199 (0.7 percent).

Table B6: Total Postal Address Errors by Contact Type (2009 RAA) – Phase 1

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

Table B7: Total Postal Address Errors by Contact Type (2009 RAA) – Phase 2

	Administrative	Technical	Registrant	All 3 total
Passed all accuracy tests	7,151	7,511	7,150	**21,812**
Missing	54	63	41	**158**
Country Code Missing	53	52	59	**164**
Country Not Identifiable	27	30	23	**80**
Postal Code Missing	144	128	154	**426**
Postal Code Format	901	768	853	**2,522**
State/Province Missing	709	607	720	**2,036**
City Missing	1,126	1,010	1,125	**3,261**
Street Missing	723	637	731	**2,091**
TOTAL	**10,000**	**10,000**	**10,000**	**30,000**
Total Errors	**3,737**	**3,295**	**3,706**	**10,738**
Total Domains w/ Errors	**2,849**	**2,489**	**2,850**	**8,188**

As indicated above, the estimated percentage of conforming postal addresses is unchanged overall. However, more errors were detected (10,738 in Phase 2 compared 9,789 in Phase 1), as seen in Table B7. There have been increases in missing country codes from 62 (0.2 percent) in Phase 1 to 164 (0.5 percent) in Phase 2; in postal code errors from 2,169 (7.2 percent) in Phase 1 to 2,948 (9.8 percent) in Phase 2; in missing cities from 2,471 (8.2 percent) in Phase 1 to 3,261 (10.9 percent) in Phase 2; and in missing streets from 1,615 (5.4 percent) in Phase 1 to 2,091 (7.0 percent) in Phase 2. There was a decrease in missing state or provinces from 3,255 (10.6 percent) in Phase 1 to 2,036 (6.8 percent) in Phase 2.

Additional Comparisons of Syntax Accuracy between Phases (by Region and RAA group)

Table B8: African Domains Accuracy by Phase – 2009 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	100% ± 0%	70.3% ± 3.4%	47.2% ± 3.7%	33.7% ± 3.5%
Phase 2	99.9% ± 0.2%	66.8% ± 2.9%	42.2% ± 3.1%	29.8% ± 2.9%
Change (2 - 1)	-0.1% ± 0.2%	-3.5% ± 4.5%	-5.0% ± 4.8%	-3.9% ± 4.5%

Table B8 shows that African accuracy had sizable drops in Phase 2 for telephone numbers and postal addresses, but neither change was statistically significant due to the small African sample sizes.

Table B9: Asia-Pacific Domains Accuracy by Phase – 2009 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	99.6% ± 0.3%	80.4% ± 1.7%	55.3% ± 2.1%	48.7% ± 2.1%
Phase 2	99.5% ± 0.3%	78.7% ± 1.7%	52.5% ± 2.0%	39.5% ± 2.0%
Change (2 - 1)	-0.1% ± 0.4%	-1.7% ± 2.4%	-2.8% ± 2.9%	-9.2% ± 2.9%

Seen in Table B9, accuracy also dropped in Phase 2 for Asia-Pacific domains for telephone numbers and postal addresses, but again, these changes were not statistically significant. However, the decrease in the percentage of all nine contacts conforming to the 2009 RAA is statistically significant.

Table B10: European Domains Accuracy by Phase – 2009 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	100% ± 0%	87.5% ± 1.5%	66.0% ± 2.2%	58.6% ± 2.2%
Phase 2	99.8% ± 0.2%	85.2% ± 1.5%	67.6% ± 2.0%	58.8% ± 2.1%
Change (2 - 1)	-0.2% ± 0.2%	-2.3% ± 2.1%	1.6% ± 3.0%	0.2% ± 3.0%

In Table B10, there is only one significant change in accuracy for European domains: there is a drop in telephone accuracy in Phase 2.

Table B11: Latin/Caribbean Domains Accuracy by Phase – 2009 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	100% ± 0%	82.4% ± 1.8%	67.1% ± 2.2%	59.2% ± 2.3%
Phase 2	99.9% ± 0.1%	79.2% ± 1.9%	67.1% ± 2.1%	56.9% ± 2.3%
Change (2 - 1)	**-0.1% ± 0.1%**	**-3.2% ± 2.6%**	**0.0% ± 3.0%**	**-3.7% ± 3.3%**

In Table B11, for Latin American/Caribbean domains, there is again a drop in telephone accuracy in Phase 2. There is also a decrease in the percentage of all nine contacts conforming to the 2009 RAA in Phase 2.

Table B12: North American Domains Accuracy by Phase – 2009 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	98.7% ± 0.4%	87.7% ± 1.1%	94.3% ± 0.8%	84.3% ± 1.2%
Phase 2	99.9% ± 0.1%	86.1% ± 1.3%	97.1% ± 0.6%	83.9% ± 1.4%
Change (2 - 1)	**1.2% ± 0.5%**	**-1.6% ± 1.7%**	**2.8% ± 1.0%**	**-0.4% ± 1.8%**

In Table B12, North American domains show increases in accuracy for email addresses and postal addresses. There is also an increase in the postal address accuracy in Phase 2.

Table B13: 2009 RAA Domains Accuracy by Phase – 2009 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	99.1% ± 0.4%	92.5% ± 1.2%	85.8% ± 1.6%	82.9% ± 1.7%
Phase 2	98.1% ± 0.6%	89.1% ± 1.3%	81.7% ± 1.6%	77.1% ± 1.7%
Change (2 - 1)	**-1.0% ± 0.7%**	**-3.4% ± 1.8%**	**-4.1% ± 2.3%**	**-5.8% ± 2.4%**

Table B14: 2013 RAA GF Domains Accuracy by Phase – 2009 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	99.4% ± 0.2%	83.8% ± 1.1%	80.8% ± 1.2%	69.1% ± 1.4%
Phase 2	99.4% ± 0.2%	80.2% ± 1.2%	81.6% ± 1.2%	66.5% ± 1.5%
Change (2 - 1)	**0.0% ± 0.3%**	**-3.6% ± 1.6%**	**0.8% ± 1.7%**	**-2.6% ± 2.1%**

Table B15: 2013 RAA NGF Domains Accuracy by Phase – 2009 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	98.9% ± 0.3%	89.3% ± 1.0%	74.7% ± 1.4%	71.5% ± 1.4%
Phase 2	98.8% ± 0.3%	88.6% ± 1.0%	74.8% ± 1.4%	67.8% ± 1.5%
Change (2 - 1)	**-0.1% ± 0.4%**	**-0.7% ± 1.4%**	**0.1% ± 2.0%**	**-3.7% ± 2.1%**

Table B13 shows that the accuracy for 2009 RAA domains has decreased in all categories for Phase 2. The only other difference in Tables B13-B15 is that telephone accuracy has decreased for telephone numbers for 2013 GF domains in Phase 2.

Appendix C: Syntax Accuracy – 2013 RAA Requirements

Here we look at accuracy against 2013 RAA requirements for the 3,742 domains required to conform to these requirements. As an additional comparison, we do compare the three RAA groups on their accuracy against the 2013 RAA requirements.
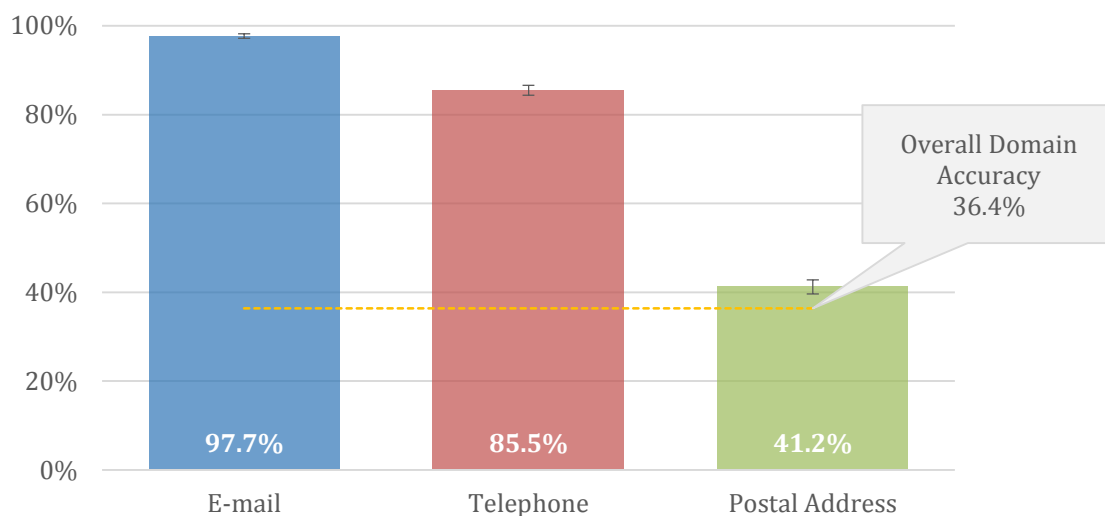
Graph C1: Overall Accuracy – 2013 RAA Syntax Requirements

Table C1: Overall Accuracy by Contact Type and Mode – 2013 RAA Syntax Requirements

	Email	Telephone	Postal Address	ALL 3 Accurate
Registrant	97.8% ± 0.5%	85.9% ± 1.1%	42.3% ± 1.6%	**37.7% ± 1.6%**
Administrative	98.8% ± 0.3%	86.0% ± 1.1%	42.3% ± 1.6%	**37.6% ± 1.6%**
Technical	98.8% ± 0.3%	86.6% ± 1.1%	42.6% ± 1.6%	**38.5% ± 1.6%**
All 3 Accurate	**97.7% ± 0.5%**	**85.5% ± 1.1%**	**41.2% ± 1.6%**	**36.4% ± 1.5%**

Table C1 shows that the email address accuracy is lower for Registrant email addresses, but that there are no other differences.

Subgroup Accuracy – 2013 RAA Syntax Requirements

Next, we look at subgroups in Phase 2, starting with Prior versus New gTLDs. Since the numbers for Registrant, administrative, and Technical are so similar (since they have the same information more than three-quarters of the time), we present subgroup accuracy for the registrant, administrative, and technical contacts all passing the accuracy tests.

Subgroup 1: Prior vs. New gTLD

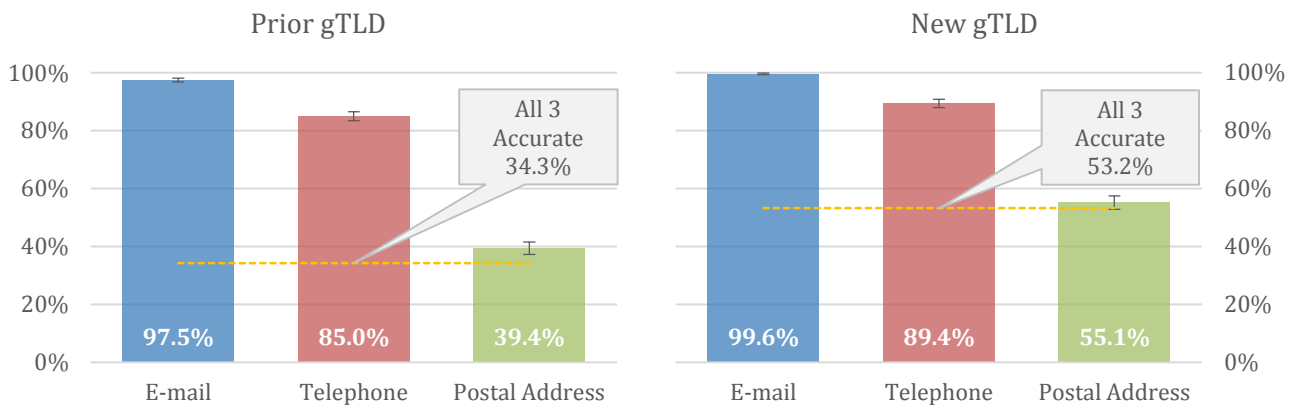Graph C2: Accuracy by gTLD Type – 2013 RAA Syntax Requirements

Table C2. Accuracy by gTLD Type – 2013 RAA Syntax Requirements

	Email	Telephone	Postal Address	ALL 3 Accurate
Prior gTLD	97.5% ± 0.7%	85.0% ± 1.6%	39.4% ± 2.2%	**34.3% ± 2.1%**
New gTLD	99.6% ± 0.3%	89.4% ± 1.4%	55.1% ± 2.3%	**53.2% ± 2.3%**
All 3 Accurate	**97.7% ± 0.5%**	**85.5% ± 1.1%**	**41.2% ± 1.6%**	**36.4% ± 1.5%**

New gTLDs have significantly higher accuracy on all three contact modes.

Subgroup 2: ICANN Region

Next, we look at accuracy by ICANN Region. Again, we present subgroup accuracy for the registrant, administrative, and technical contacts all passing the accuracy tests.

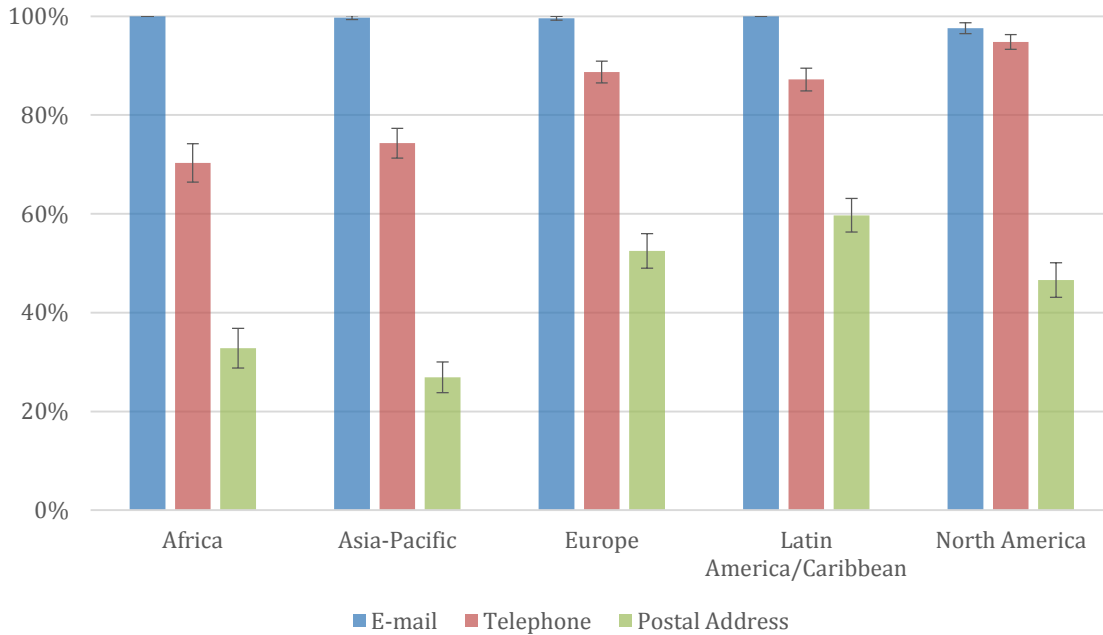Graph C3: Accuracy by ICANN Region – 2013 RAA Syntax Requirements

Table C3: Accuracy by ICANN Region – 2013 RAA Syntax Requirements

	Email	Telephone	Postal Address	ALL 3 Accurate
Africa	100.0% ± 0.0%	70.3% ± 3.9%	32.8% ± 4.0%	**24.4% ± 3.7%**
Asia-Pacific	99.7% ± 0.4%	74.3% ± 3.0%	26.9% ± 3.1%	**21.8% ± 2.9%**
Europe	99.6% ± 0.4%	88.7% ± 2.2%	52.5% ± 3.5%	**46.6% ± 3.5%**
Latin America/Caribbean	100.0% ± 0.0%	87.2% ± 2.3%	59.7% ± 3.4%	**55.9% ± 3.4%**
North America	97.6% ± 1.1%	94.8% ± 1.5%	46.6% ± 3.5%	**42.3% ± 3.4%**
All 3 Accurate	**97.7% ± 0.5%**	**85.5% ± 1.1%**	**41.2% ± 1.6%**	**36.4% ± 1.5%**

North American domains have lower accuracy for email addresses, as seen in Graph C3 and Table C3. For telephone numbers and postal addresses, Africa and Asia-Pacific are lowest. North America is highest in accuracy for telephone numbers, but Latin American/Caribbean domains are highest in accuracy for postal addresses. The patterns for having all three contact modes conform to the RAA 2009 standards are the same as for postal addresses.

Subgroup: RAA Status

Finally, we look at accuracy by RAA Status. Only the 2013 RAA NGF group is required to meet the standards of the 2013 RAA, so we should expect that this group has the highest accuracy.

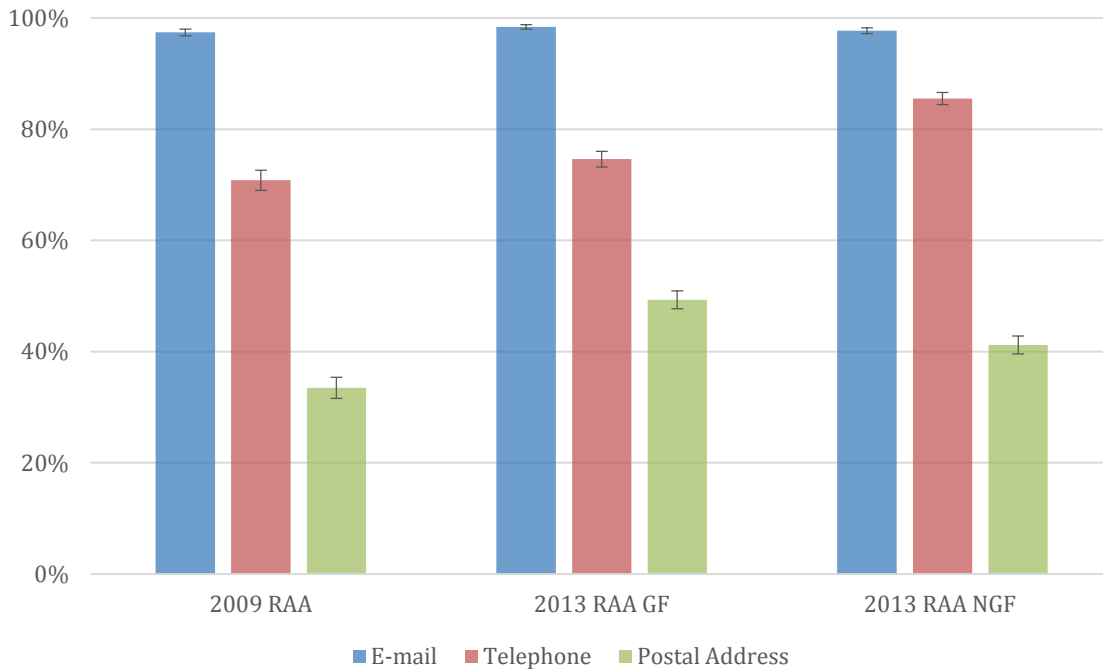Graph C4: Accuracy by RAA Status – 2013 RAA Syntax Requirements

Table C4: Accuracy by RAA Status – 2013 RAA Syntax Requirements

	Email	Telephone	Postal Address	ALL 3 Accurate
2009 RAA	97.4% ± 0.6%	70.8% ± 1.8%	33.5% ± 1.9%	**17.7% ± 1.6%**
2013 RAA GF	98.4% ± 0.4%	74.6% ± 1.4%	49.3% ± 1.6%	**43.9% ± 1.6%**
2013 RAA NGF	**97.7% ± 0.5%**	**85.5% ± 1.1%**	**41.2% ± 1.6%**	**36.4% ± 1.5%**

Graph C4 and Table C4 show that there are no significant differences in email address accuracy despite the variability. The 2009 RAA group has a lower accuracy for telephone numbers and postal addresses. The 2013 NGF group has a higher accuracy for telephone numbers while the 2013 RAA GF group has a higher accuracy for postal addresses. The 2009 RAA also has a lower percentage of all nine contacts being accurate, while the 2013 RAA GF has a higher percentage of all nine contacts being accurate.

Comparisons between Phase 1 and Phase 2 Accuracy – 2013 RAA Syntax Requirements

We have above presented the syntax accuracy to 2013 RAA Requirements for Phase 2 Cycle 1. Here, we compare the Phase 2 results to those from Phase 1.

Overall Accuracy
Table C5: Overall Accuracy by Phase – 2013 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	97.1% ± 0.5%	84.5% ± 1.1%	44.0% ± 1.6%	37.8% ± 1.5%
Phase 2	97.7% ± 0.5%	85.5% ± 1.1%	41.2% ± 1.6%	36.4% ± 1.5%
Change (2 - 1)	0.6% ± 0.7%	1.0% ± 1.6%	-2.8% ± 2.3%	-1.4% ± 2.1%

Table C5 shows that Phase 2 has a lower overall postal address accuracy rate, but that the email address and telephone number accuracy rates are very similar between the phases.

Prior vs. New gTLDs
Table C6: Prior gTLDs Accuracy by Phase – 2013 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	96.8% ± 0.8%	84.2% ± 1.7%	42.0% ± 2.2%	35.3% ± 2.2%
Phase 2	97.5% ± 0.7%	85.0% ± 1.6%	39.4% ± 2.2%	34.3% ± 2.1%
Change (2 - 1)	0.7% ± 1.1%	0.8% ± 2.3%	-2.6% ± 3.1%	-1.0% ± 3.0%

Table C7: New gTLDs Accuracy by Phase – 2013 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	99.8% ± 0.2%	86.4% ± 1.5%	61.5% ± 2.1%	59.4% ± 2.2%
Phase 2	99.6% ± 0.3%	89.4% ± 1.4%	55.1% ± 2.3%	53.2% ± 2.3%
Change (2 - 1)	-0.2% ± 0.4%	3.0% ± 2.1%	-5.6% ± 3.1%	-6.2% ± 3.2%

For Prior gTLDs, there are no differences between Phase 1 and Phase 2 seen in Tables C6 and C7. For New gTLDs, there is an increase in telephone number accuracy in Phase 2 and a decrease in postal addresses in Phase 2. Due to these decreases in Phase 2, the percentage of domains with all nine contacts passing all accuracy tests is lower in Phase 2.

ICANN Regions
Table C8: African Domains Accuracy by Phase – 2013 RAA Syntax Requirements

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	98.7% ± 1.2%	74.8% ± 4.4%	26.1% ± 4.4%	20.3% ± 4.1%
Phase 2	100.0% ± 0.0%	70.3% ± 3.9%	32.8% ± 4.0%	24.4% ± 3.7%
Change (2 - 1)	1.3% ± 1.2%	-4.5% ± 5.9%	6.7% ± 5.9%	4.1% ± 5.5%

Per table C8, African accuracy increased in Phase 2 for email addresses and postal addresses.

Table C9: Asia-Pacific Domains Accuracy to 2013 RAA Syntax Requirements by Phase

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	99.7% ± 0.4%	72.4% ± 3.0%	36.6% ± 3.3%	27.3% ± 3.0%
Phase 2	99.7% ± 0.4%	74.3% ± 3.0%	26.9% ± 3.1%	21.8% ± 2.9%
Change (2 - 1)	0.0% ± 0.6%	1.9% ± 4.2%	-9.7% ± 4.5%	-5.5% ± 4.2%

Per table C9, accuracy dropped in Phase 2 for Asia-Pacific domains for postal addresses, and this drop also resulted in a drop in the Phase 2 percentage of domains with accuracy for all nine contacts.

Table C10: European Domains Accuracy to 2013 RAA Syntax Requirements by Phase

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	94.4% ± 1.6%	88.3% ± 2.2%	54.6% ± 3.5%	48.4% ± 3.5%
Phase 2	99.6% ± 0.4%	88.7% ± 2.2%	52.5% ± 3.5%	46.6% ± 3.5%
Change (2 - 1)	5.2% ± 1.6%	0.4% ± 3.1%	-2.1% ± 4.9%	-1.8% ± 4.9%

European accuracy against 2013 RAA syntax standards increased for email addresses, as seen in table C10. In Phase 1, Europe had the lowest accuracy rate, but in Phase 2, the accuracy is very high like all the regions except North America, which remains lowest.

Table C11: Latin/Caribbean Domains Accuracy to 2013 RAA Syntax Requirements by Phase

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	99.8% ± 0.3%	87.5% ± 2.3%	61.5% ± 3.4%	58.7% ± 3.4%
Phase 2	100.0% ± 0.0%	87.2% ± 2.3%	59.7% ± 3.4%	55.9% ± 3.4%
Change (2 - 1)	0.2% ± 0.3%	-0.3% ± 3.3%	-1.8% ± 4.8%	-2.8% ± 4.8%

There are no changes in accuracy for Latin American/Caribbean domains, evidenced by Table C11.

Table C12: North American Domains Accuracy to 2013 RAA Syntax Requirements by Phase

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	95.9% ± 1.2%	91.4% ± 1.7%	43.9% ± 3.0%	39.4% ± 3.0%
Phase 2	97.6% ± 1.1%	94.8% ± 1.5%	46.6% ± 3.5%	42.3% ± 3.4%
Change (2 - 1)	1.7% ± 1.6%	3.4% ± 2.3%	2.7% ± 4.6%	2.9% ± 4.5%

In Table C12, North American domains show increases in accuracy for email addresses and telephone numbers, but North America is still lowest for email addresses.

RAA Status

Finally, Tables C13-C15 below shows the changes from Phase 1 to Phase 2 by contact mode and RAA group, though the postal syntax tests were only carried out for the 2013 RAA NGF group in Phase 1.

Table C13: 2009 RAA Domains Accuracy to 2013 RAA Syntax Requirements by Phase

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	98.9% ± 0.5%	71.1% ± 2.1%	n/a	n/a
Phase 2	97.4% ± 0.6%	70.8% ± 1.8%	33.5% ± 1.9%	17.7% ± 1.6%
Change (2 - 1)	-1.5% ± 0.8%	-0.3% ± 2.8%	n/a	n/a

Table C14: 2013 RAA GF Domains Accuracy to 2013 RAA Syntax Requirements by Phase

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	97.8% ± 0.4%	74.2% ± 1.3%	n/a	n/a
Phase 2	98.4% ± 0.4%	74.6% ± 1.4%	49.3% ± 1.6%	43.9% ± 1.6%
Change (2 - 1)	0.6% ± 0.6%	0.4% ± 1.9%	n/a	n/a

Table C15: 2013 RAA NGF Domains Accuracy to 2013 RAA Syntax Requirements by Phase

Phase	Email	Telephone	Postal Address	ALL Modes Accurate
Phase 1	97.1% ± 0.5%	84.5% ± 1.1%	44.0% ± 1.6%	37.8% ± 1.5%
Phase 2	97.7% ± 0.5%	85.5% ± 1.1%	41.2% ± 1.6%	36.4% ± 1.5%
Change (2 - 1)	0.6% ± 0.7%	1.0% ± 1.6%	-2.8% ± 2.3%	-1.4% ± 2.1%

The 2009 RAA accuracy has significantly decreased for email addresses, while the 2013 RAA NGF accuracy has significantly decreased for postal addresses.

2013 RAA Reasons for Syntax Error
In Phase 1, we showed which accuracy tests were failed by each contact. We repeat these tables from Phase 1 and also show the same data for Phase 2 for comparison.

Email Addresses
In Phase 1, in Table C16, no errors were ever found in the email addresses except if a required email address was missing (the Registrant email address is required under the 2013 RAA). From Table C17, we see that in Phase 2, however, a very small number of email addresses were found to either have the "@" symbol missing or to not be resolvable into a valid-looking email address. However, the number of missing registrant email addresses is less than in Phase 1.

Table C16: Total Email Address Errors by Contact Type (2013 RAA) – Phase 1

	Administrative	Technical	Registrant	All 3 total
Passed all accuracy tests	3,829	3,830	3,802	**11,461**
Missing	19	18	46	**83**
Total	**3,848**	**3,848**	**3,848**	**11,544**

Table C17: Total Email Address Errors by Contact Type (2013 RAA) – Phase 2

	Administrative	Technical	Registrant	All 3 total
Passed all accuracy tests	3,725	3,726	3,713	**11,164**
Missing	15	14	27	**56**
@ Missing	1	1	1	**3**
Not Resolvable	1	1	1	**3**
Total	**3,742**	**3,742**	**3,742**	**11,226**

Telephone Numbers
As seen below in Graph C5 and Tables C18 and C19, there is no significant change in accuracy for telephone numbers between Phase 1 and Phase 2. The errors found did change, however. There were more telephone numbers that were too long in Phase 2 (573, 5.1 percent) than in Phase 1 (431, 3.8 percent). There were fewer telephone numbers that were missing in Phase 2 (248, 2.2 percent) than in

Phase 1 (326, 2.9 percent). There were fewer telephone numbers with an invalid country code format in Phase 2 (131, 1.2 percent) than in Phase 1 (272, 2.4 percent).

Graph C5: Reasons for Telephone Number Syntax Error, by Contact Type – 2013 RAA Requirements
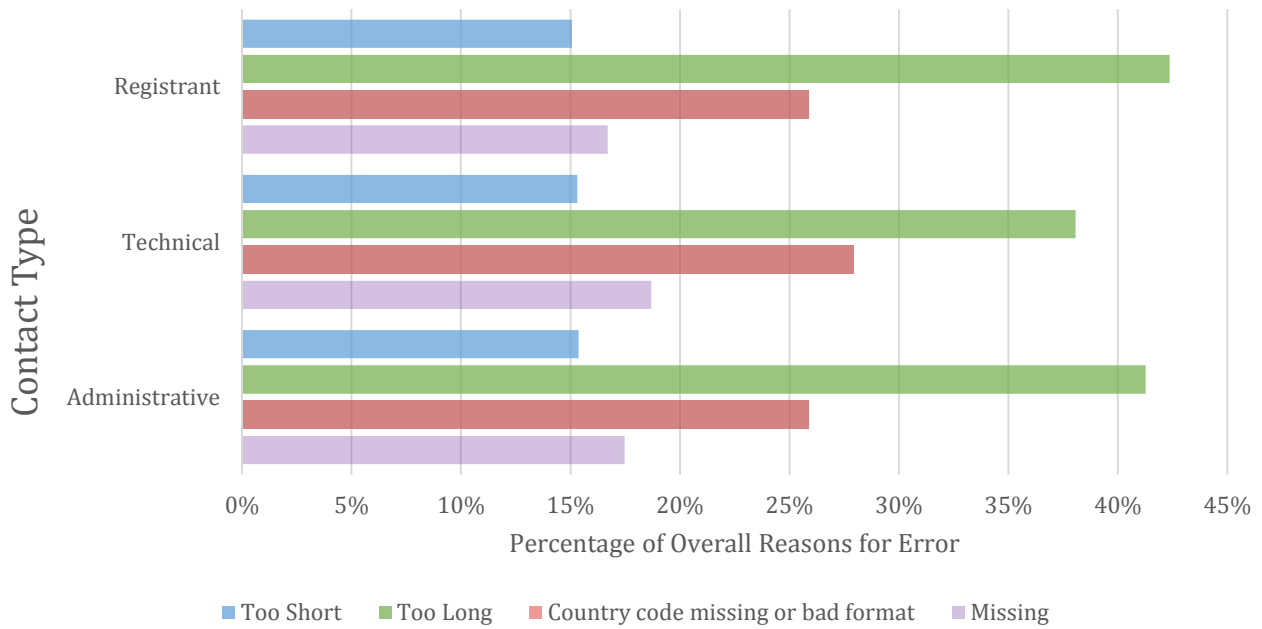
Graph C6: Reasons for Telephone Number Syntax Error, by Error Type – 2013 RAA Requirements
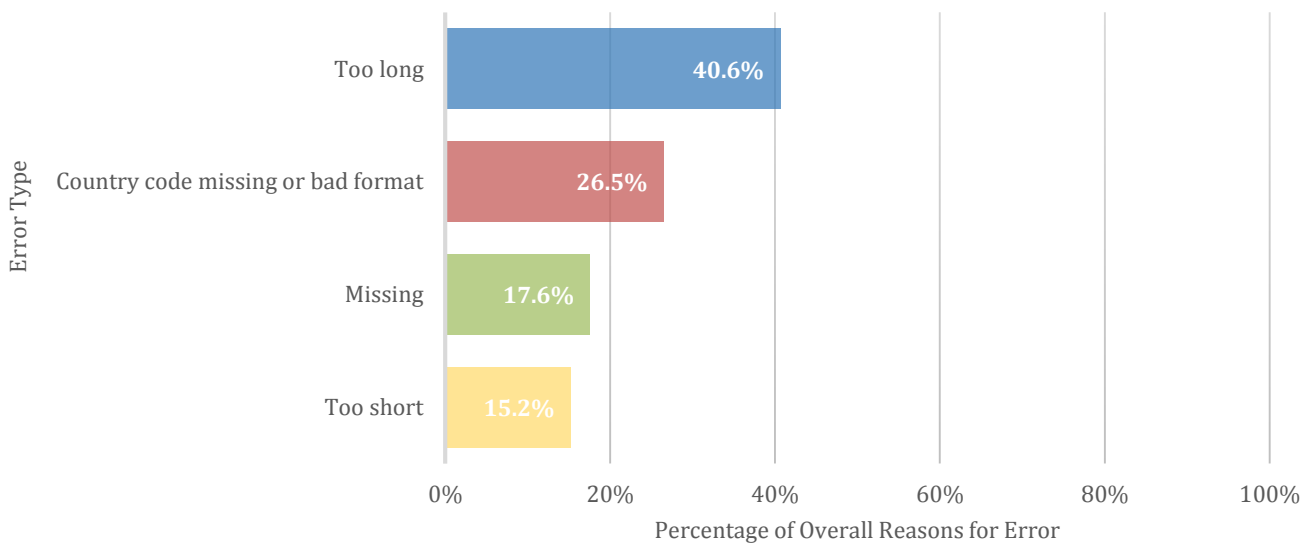
Table C18 Total Telephone Number Errors by Contact Type (2013 RAA) – Phase 2

	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**
Country Code Format	90	91	91	**272**
Incorrect Length	223	199	223	**645**
Unallowed Characters	1	1	1	**3**
Total	**3,848**	**3,848**	**3,848**	**11,544**

Note: Italics indicate new 2013 RAA requirements

Table C19 Total Telephone Number Errors by Contact Type (2013 RAA) – Phase 2

	Administrative	Technical	Registrant	All 3 total
Passed all accuracy tests	3,267	3,298	3,251	**9,816**
Missing	83	83	82	**248**
Country Code Missing	79	82	82	**243**
Country Code Format	*44*	*42*	*45*	*131*
Incorrect Length	269	237	282	**788**
Unallowed Characters	0	0	0	**0**
Total	**3,742**	**3,742**	**3,742**	**11,226**

Postal Addresses

As indicated below in Graph C6 and Tables C20 and C21, the estimated percentage of conforming postal addresses dropped from Phase 1 to Phase 2. However, more errors were detected in Phase 1 (9,594 in Phase 1 compared to 7,112 in Phase 2). There have been increases in missing postal codes from 396 (3.4 percent) in Phase 1 to 738 (6.4 percent) in Phase 2 and in state/provinces formats from 86 (0.7 percent) in Phase 1 to 246 (2.1 percent) in Phase 2. There have been decreases in postal code formats from 823 (7.1 percent) in Phase 1 to only 24 (0.2 percent) in Phase 2; in state/provinces missing from 1,101 (9.5 percent) in Phase 1 to 404 (3.5 percent) in Phase 2; in state/provinces given in the wrong field from 1,408 (12.2 percent) in Phase 1 to 1,037 (9.0 percent) in Phase 2; in missing cities from 1,068 (9.3 percent) in Phase 1 to 816 (7.1 percent) in Phase 2; in cities in the wrong field from 798 (6.9 percent) in Phase 1 to 516 (4.5 percent) in Phase 2; and in missing streets from 903 (7.8 percent) in Phase 1 to 715 (6.2 percent) in Phase 2.

Graph C7: Reasons for Postal Address Errors Across all Contact Types – 2013 RAA Requirements
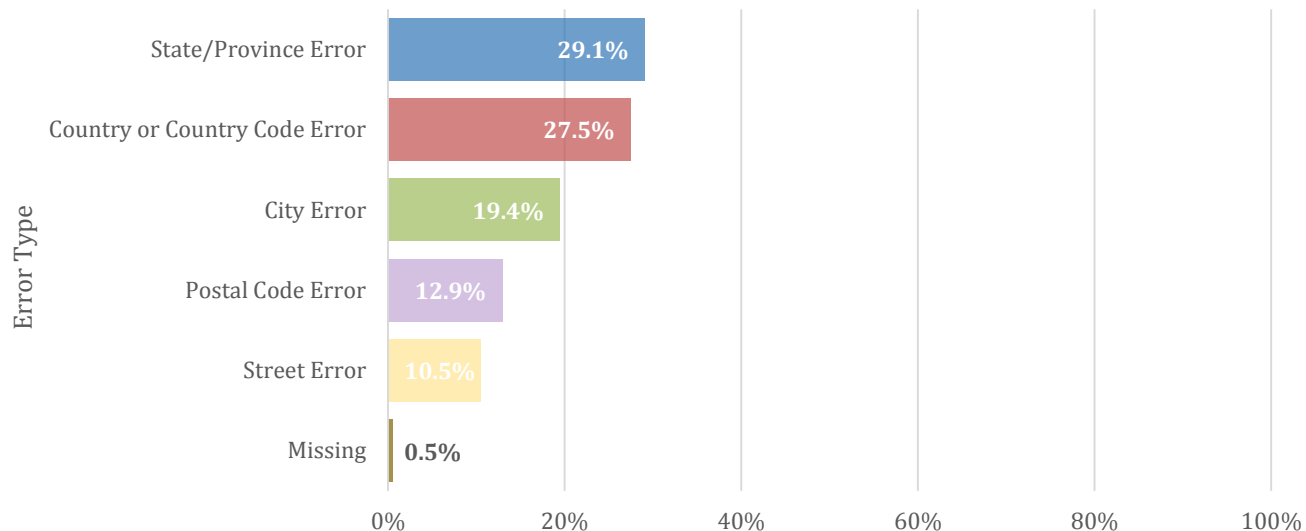
Error Type	Percentage
State/Province Error	29.1%
Country or Country Code Error	27.5%
City Error	19.4%
Postal Code Error	12.9%
Street Error	10.5%
Missing	0.5%

Table C20: Total Postal Address Errors by Contact Type (2013 RAA) – Phase 1

	Administrative	Technical	Registrant	All 3 total
Passed all accuracy tests	2,186	2,239	2,187	**6,612**
Missing	19	19	16	**54**
Country Code Missing	3	3	2	**8**
Country Not Identifiable	10	10	10	**30**
Country in Wrong Field	*811*	*811*	*810*	*2,432*
Country not ISO alpha 2	*1*	*1*	*1*	*3*
Postal Code Missing	255	250	233	**738**
Postal Code Format	8	8	8	**24**
State/Province Missing	140	124	140	**404**
State/Province in Wrong Field	*357*	*336*	*344*	*1,037*
State/Province Format	*83*	*79*	*84*	*246*
City Missing	283	255	278	**816**
City in Wrong Field	*165*	*186*	*165*	*516*
Street Missing	242	224	249	**715**
Street in Wrong Field	*31*	*28*	*30*	*89*
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

Table C21: Total Postal Address Errors by Contact Type (2013 RAA) – Phase 2

	Administrative	Technical	Registrant	All 3 total
Passed all accuracy tests	1,985	2,031	1,983	**5,999**
Missing	17	17	15	**49**
Country Code Missing	2	2	1	**5**
Country Not Identifiable	13	14	13	**40**
Country in Wrong Field	865	864	865	**2,594**
Postal Code Missing	141	114	141	**396**
Postal Code Format	283	270	270	**823**
Postal Code in Wrong Field	8	8	7	**23**
State/Province Missing	373	347	381	**1,101**
State/Province in Wrong Field	494	428	486	**1,408**
State/Province Format	90	96	96	**282**
City Missing	365	337	366	**1,068**
City in Wrong Field	250	304	244	**798**
Street Missing	305	286	312	**903**
Street in Wrong Field	38	28	38	**104**
TOTAL	**3,742**	**3,742**	**3,742**	**11,226**
Total Errors	**3,244**	**3,115**	**3,235**	**9,594**
Total Domains w/ Errors	**1,757**	**1,711**	**1,759**	**5,545**

Note: Italics indicate new 2013 RAA requirements

Analysis by Subgroup: Accuracy to 2013 RAA requirements - Operability

For operability, the only additional requirement for the 2013 RAA is that registrant email addresses and telephone numbers become required. Therefore, results for Accuracy to 2013 RAA requirements for operability would be very repetitive, and are not repeated in this report.