WHOIS Accuracy Reporting System (ARS)

Phase 2 Cycle 2 Report: Syntax and Operability Accuracy

Global Domains Division | 8 June 2016

Contents

Executive Summary

Subject of This Report

The WHOIS Accuracy Reporting System (ARS) is a system designed to meet recommendations from the 2012 WHOIS Review Team convened under the Affirmation of Commitments (AOC).[1] Based on these recommendations, 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. The WHOIS ARS was created as part of these improvements and to address Governmental Advisory Committee (GAC) concerns on WHOIS accuracy.

This report is the fourth in a series of reports produced by the WHOIS ARS. The WHOIS ARS has been designed in phases to enable the ICANN community to influence its development. A pilot phase was completed in April 2015, and Phase 1 was completed in August 2015. Phase 2 is ongoing and cyclical – Phase 2 Cycle 1 ("Cycle 1") was completed in December 2015, and Phase 2 Cycle 2 ("Cycle 2"), the subject of this report, started in January 2016. Where Phase 1 examined only syntax accuracy, Phase 2 reports examine both the syntax and operability accuracy of WHOIS records. As with previous reports, this report will again detail the leading types of nonconformance, trends and comparisons of WHOIS accuracy across regions, Registrar Accreditation Agreement (RAA) versions and generic top-level domain (gTLD) types.

Full details on the WHOIS ARS background, as well as results, can be found in previous ARS reports: Pilot Report, Phase 1 Report, Phase 2 Cycle 1 Report.

Accuracy Testing Methods[2]

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.[3] Syntax testing assessed the format of a record (e.g., does the email address contain an "@" symbol?). Operability testing assessed the functionality of the information in a record (e.g., did the email 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[4]).

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

---

[1] *See* https://www.icann.org/resources/pages/aoc-2012-02-25-en.

[2] General information regarding syntax and operability accuracy tests/criteria can be found in Appendix A. More detailed information can be found on the WHOIS ARS webpage: https://whois.icann.org/en/whoisars-validation.

[3] The accuracy tests/criteria are what we have defined as the baseline requirements of contact data to be deemed formatted correctly and operable. While the 2009 RAA does not contain explicit syntax requirements, the contact data provided is expected to be complete and formatted correctly.

[4] *See* here for RAA versions: https://www.icann.org/resources/pages/registrars/registrars-en.

Sample Design

At the time of the initial sample early in the first quarter 2016, there were 169.5 million domains[5] spread across 610 gTLDs.[6] Nearly 94 percent of the 169.5 million domains were registered in one of the 18 Prior gTLDs, and about 6 percent were registered in one of 592 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. The initial sample contained 200,000 records, and the analyzed subsample contained 12,000 records, representing all active gTLDs at the time.[7]

Though an estimated 97 percent of domain names are registered through registrars that have been accredited under the 2013 RAA, a majority of domains are allowed to operate under the WHOIS standards of the 2009 RAA.[8] For this reason, the 2009 RAA criteria are used as the baseline to assess WHOIS accuracy in this report; however, all 2013 RAA non-grandfathered (NGF) domains were also tested to the 2013 RAA criteria, the findings for which are available in Appendix C. Table Ex1 shows the breakdown of the initial sample described above. More detailed information, including why Table Ex1 has only 196,262 domains and how the sample size was determined, can be found in Sample Design.

Table Ex1: Initial Sample Sizes by Region and RAA

RAA Type	Africa	Latin America and Caribbean	Europe	Asia Pacific	North America	Unknown	TOTAL
2009	30	306	619	450	3,258	19	**4,682**
2013 GF	457	2,184	18,275	14,379	46,564	438	**82,297**
2013 NGF	769	6,157	17,571	47,991	36,062	733	**109,283**
TOTAL	**1,256**	**8,647**	**36,465**	**62,820**	**85,884**	**1,190**	**196,262**

Findings

All 12,000 records in the analyzed subsample were evaluated using the 2009 RAA criteria, which acts as a baseline to assess the overall accuracy of WHOIS records in gTLDs. Phase 2 focuses on rates of syntax and operability 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

[5] Based on information from the gTLD zone files.
[6] At the time of sampling, there were 888 delegated gTLDs (18 Prior gTLDs and 870 New gTLDs),). 260 of the 888 gTLDs had zero domains and 40 had exactly one domain. These 300 gTLDs were excluded from sampling.
[7] 552 New gTLDs and 18 Prior gTLDs with at least two domains. Also note that the sample sizes increased from Cycle 1: 150,000 to 200,000 and 10,000 to 12,000.
[8] This could be for one of two reasons: 1) the registrar has not yet signed a 2013 RAA with ICANN and is only subject 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).

data are presented in this report at a 95 percent confidence interval[9] 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 confidence interval.

Ability to Establish Contact

Ninety-nine percent of records had at least one contact mode of the three contact types that met all syntax and operability requirements of the 2009 RAA, which implies that nearly all records contain information that can be used to establish contact. In only 1 percent of records were there no contact modes for any contact type that met syntax or operability requirements.

Syntax Accuracy

The syntax accuracy analysis finds that approximately 99 percent of email addresses, 85 percent of telephone numbers and 77 percent of postal addresses met all of the baseline syntax requirements of the 2009 RAA for all three contacts.[10] 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 Ex2 provides the accuracy breakdown by contact mode, presented as 95 percent confidence intervals.

Table Ex2: Overall[11] gTLD Accuracy to 2009 RAA Syntax Requirements by Mode

	Email	Telephone	Postal Address	All Three Accurate
All Three Contacts Accurate	99.2% ± 0.2%	85.3% ± 0.6%	77.3% ± 0.7%	67.2% ± 0.8%

Operability Accuracy

The operability accuracy analysis finds that approximately 91 percent of email addresses, 76 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 70 percent for the gTLD population as a whole. Table Ex3 provides the accuracy breakdown by contact mode, presented as 95 percent confidence intervals.

Table Ex3: Overall gTLD Accuracy to 2009 RAA Operability Requirements by Mode

	Email	Telephone	Postal Address	All Three Accurate
All Three Contacts Accurate	91.4% ± 0.5%	76.0% ± 0.8%	97.7% ± 0.3%	70.2% ± 0.8%

The leading causes of syntax and operability nonconformance in the various subgroups are examined and explained in Main Findings and in Appendix B and Appendix C.

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

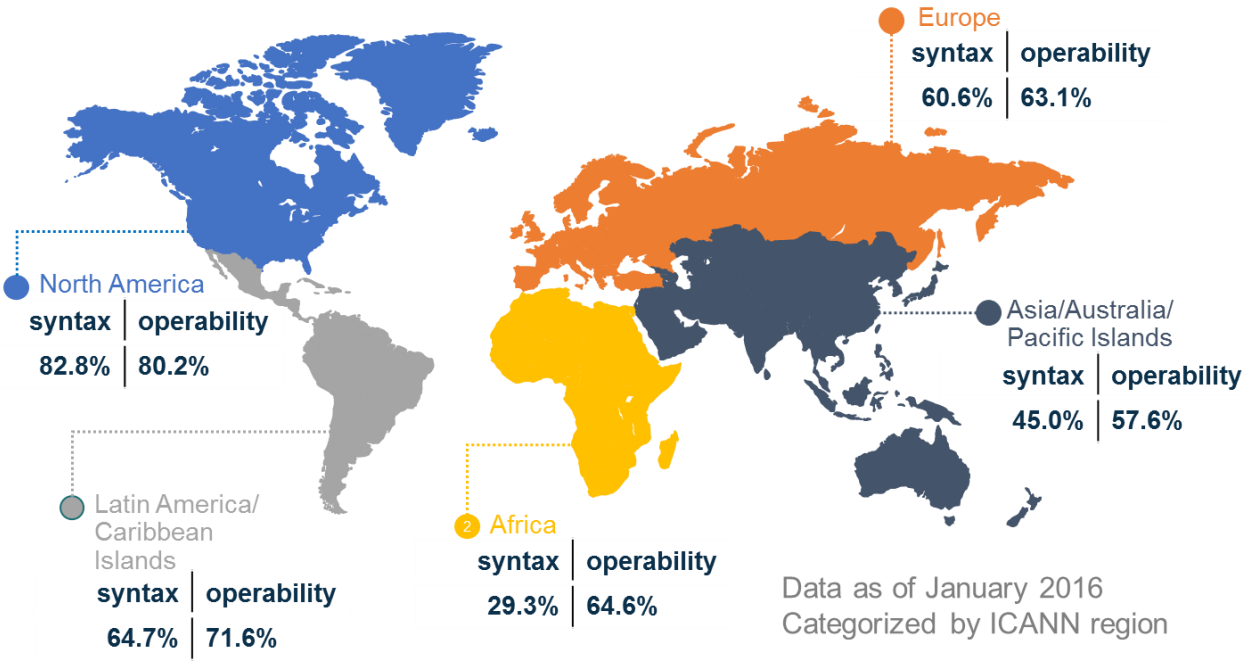
[10] *See* note 3.

[11] Overall accuracy refers to the entire 169.5 million domains. *See* note 12 on confidence intervals and population.

Regional Accuracy

In Cycle 2, we added analyses on regional differences in accuracy and reasons for error. The map in Figure Ex1 shows the overall syntax and operability accuracy of WHOIS records based on ICANN domain region, with syntax accuracy figures on the left, and operability on the right. In Additional Findings, under the heading Regional Analyses, other regional metrics of accuracy and reasons for error can be found.

Figure Ex1: Overall Syntax and Operability Accuracy by ICANN Region, Cycle 2



Note: For each region, syntax and operability accuracy figures are displayed in the format: syntax | operability. Accuracy rates shown are the percentage of records with accurate information in all three contact types, for all three contact modes.

The main body and appendices of the report include additional sub-analyses relating to accuracy rates under the 2013 RAA, trends from Cycle 1 to Cycle 2, differences between New and Prior gTLDs, and also analysis of the scripts used to register domains.

Next Steps

Phase 2 Cycle 3

The WHOIS ARS is intended to be a system used for repeatable assessment; Phase 2 Cycle 3 will reprise the syntax and operability review of Phase 2 Cycles 1 and 2 and will begin in July 2016 with a report targeted for early December 2016.

ICANN Contractual Compliance

As of the publication of this report, the results (i.e., all potentially inaccurate records) of Cycle 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 and the type of follow-up required with registrars. As Cycle 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. For example, records that have been deemed "operable" but with formatting errors will receive a different kind of notice than records have been deemed "inoperable" with formatting errors. All WHOIS ARS tickets will follow the Contractual Compliance Approach and Process[12] 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.

Contractual Compliance continues to present metrics for WHOIS ARS in the Compliance Quarterly Reports (see https://www.icann.org/resources/pages/compliance-reports-2016-04-15-en), and will provide additional information when metrics are generated for the second quarter of 2016. Additionally, metrics will be provided at public ICANN meetings, where appropriate.

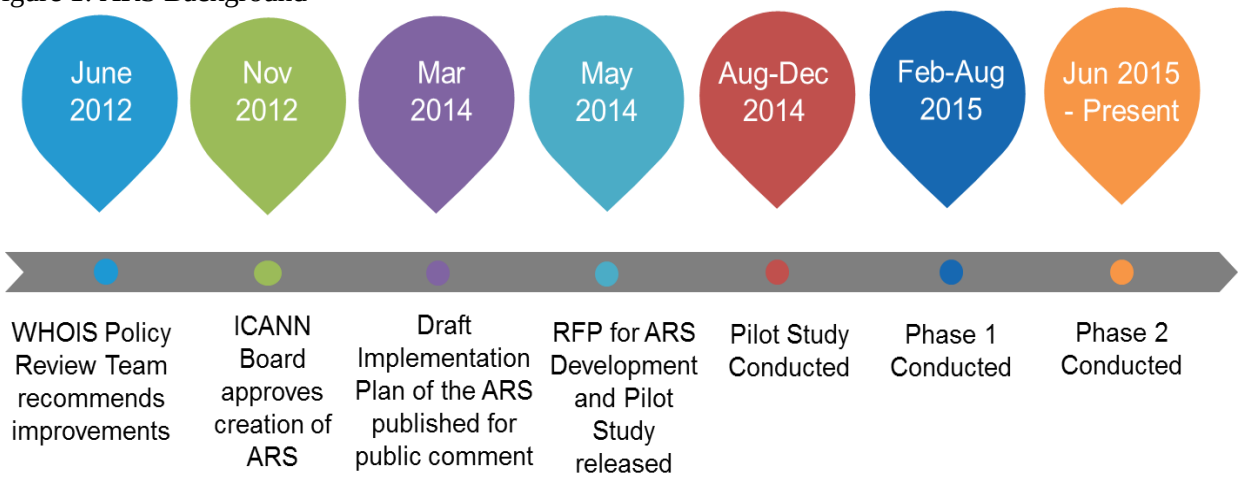
[12] *See* ICANN Contractual Compliance Approach and Process: https://www.icann.org/resources/pages/approach-processes-2012-02-25-en.

Introduction

Subject of This Report

The WHOIS Accuracy Reporting System (ARS) is designed to meet recommendations compiled under the Affirmation of Commitments and delivered by the 2012 WHOIS Review Team.[13] Based on these recommendations, 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. The WHOIS ARS was created as part of these improvements and to address Governmental Advisory Committee (GAC) concerns on WHOIS accuracy. Figure 1 shows a timeline of events surrounding the creation and progress of the WHOIS ARS.

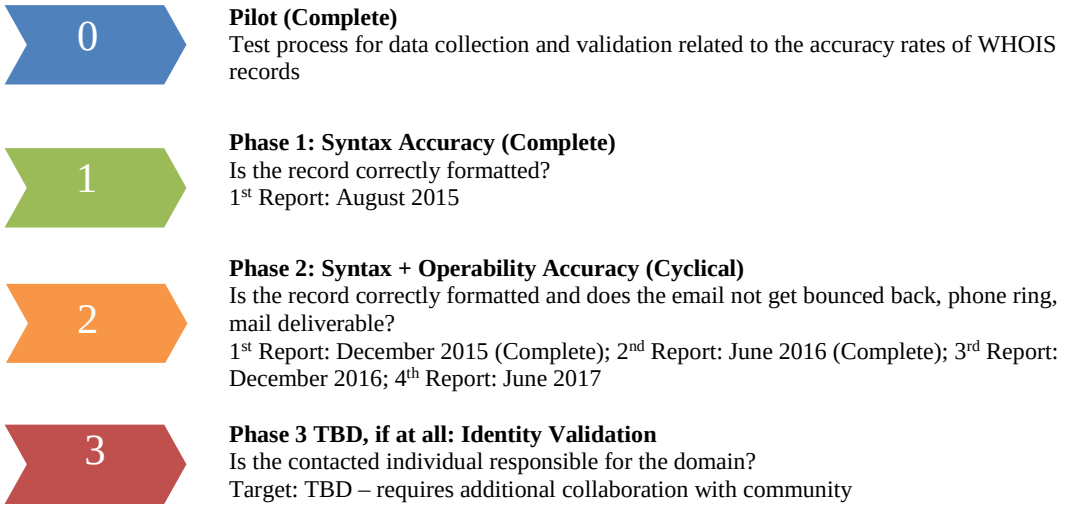
Figure 1: ARS Background



The ARS is divided into three phases, based on the types of validations described in the SAC058 Report[14]. Phase 1 analyzed the syntax accuracy of WHOIS contact information. Phase 2 is ongoing and cyclical and assesses the operability of the contact data in the record by combining the syntax tests from Phase 1 with operability tests. Phase 3 is intended to look at identity validations; however, the timeline for implementation of Phase 3 has not yet been determined. ICANN will continue to work with the Community to assess if Phase 3 will be implemented at all and, if so, how. Figure 2 illustrates this phased approach.

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

Figure 2: WHOIS ARS Phases



This report is the fourth in a series of reports produced by the WHOIS ARS. A pilot phase was completed in April 2015, and Phase 1 was completed in August 2015. Phase 2 is ongoing – Phase 2 Cycle 1 ("Cycle 1") was completed in December 2015. Phase 2 Cycle 2 ("Cycle 2"), the subject of this report, started in January 2016. Full details on the WHOIS ARS background and results can be found in previous ARS reports: Pilot Report, Phase 1 Report, Phase 2 Cycle 1 Report.

Phase 2 Cycle 1 Recap

Phase 2 Cycle 1 ("Cycle 1")[15] of the WHOIS ARS was published in December 2015 and acted as a follow-up to the Phase 1 study conducted from April to August 2015. The major findings from Cycle 1 included:

- For syntax accuracy, there was a drop in telephone number accuracy from Phase 1. The drop in telephone number accuracy seemed to be due to an increase in missing country codes among the telephone numbers sampled for Cycle 1.
- 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).
- Unlike 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.

[15] From this point, Phase 2 Cycle 1 and Phase 2 Cycle 2 will simply be referred to as Cycle 1 and Cycle 2, respectively. Any discussion of Phase 1 or Phase 2 as a whole will be explicitly noted as such to avoid confusion with the cycles. It is also important to note that some Cycle 1 calculations were improved due to correction of some syntax test results related to provinces in the Netherlands. The updated Cycle 1 calculations are used in this document even though the Cycle 1 report has not been updated.

- 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

Phase 2 Cycle 2 Overview

Phase 2 Cycle 2 Objectives

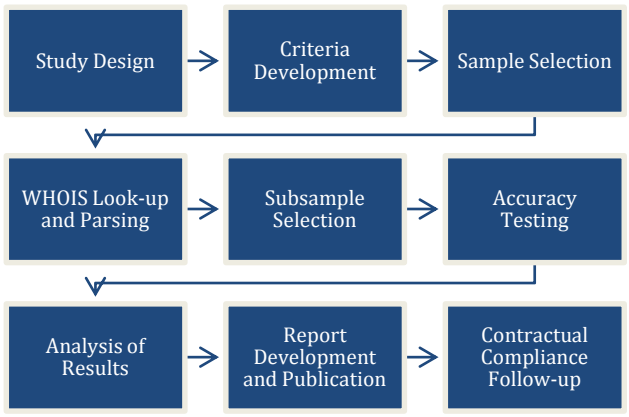
The objective for Phase 2 Cycle 2 ("Cycle 2") is the same as that of Cycle 1: to examine both syntax and operability accuracy of WHOIS records. ICANN seeks to determine whether the WHOIS record is meeting the format and content requirements of the applicable RAA and if the contact data provided is contactable. The Cycle 2 report details the leading types of nonconformance, trends and comparisons of WHOIS accuracy across regions, RAA and gTLD types. One difference from Cycle 1 is that this Cycle 2 report provides more details on regional differences in syntax and operability accuracy.

The underlying data allow 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 does provide comparisons across ARS studies, any improvement in the accuracy of the WHOIS data cannot be said to be directly linked to the ARS. There will be a lag in the potential effect of the ARS due to the timing of data pulls and when ICANN presented the aggregate data to the ICANN community and Contractual Compliance had started follow-up with registrars.

Project Plan, Tasks and Timeline

Cycle 2 was administered in the same manner as Cycle 1. Figures 3 through 5 illustrate the process, timeline and WHOIS ARS team, respectively.

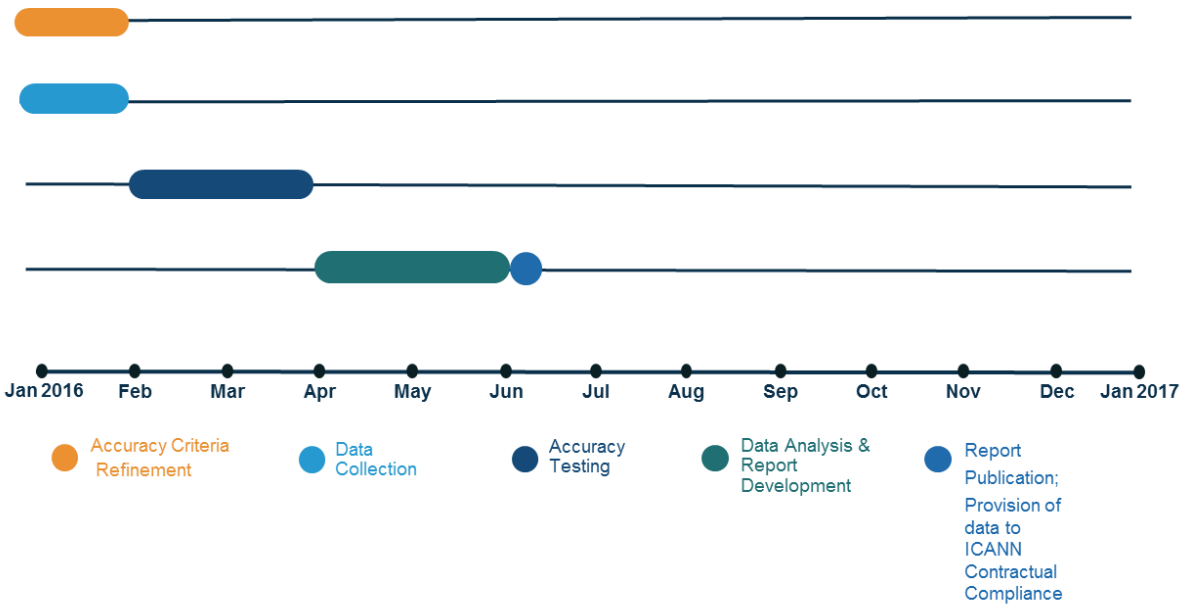
Figure 3: Work Flow and Tasks



These tasks[16] were conducted by the team in the timeline, as illustrated in Figure 4.

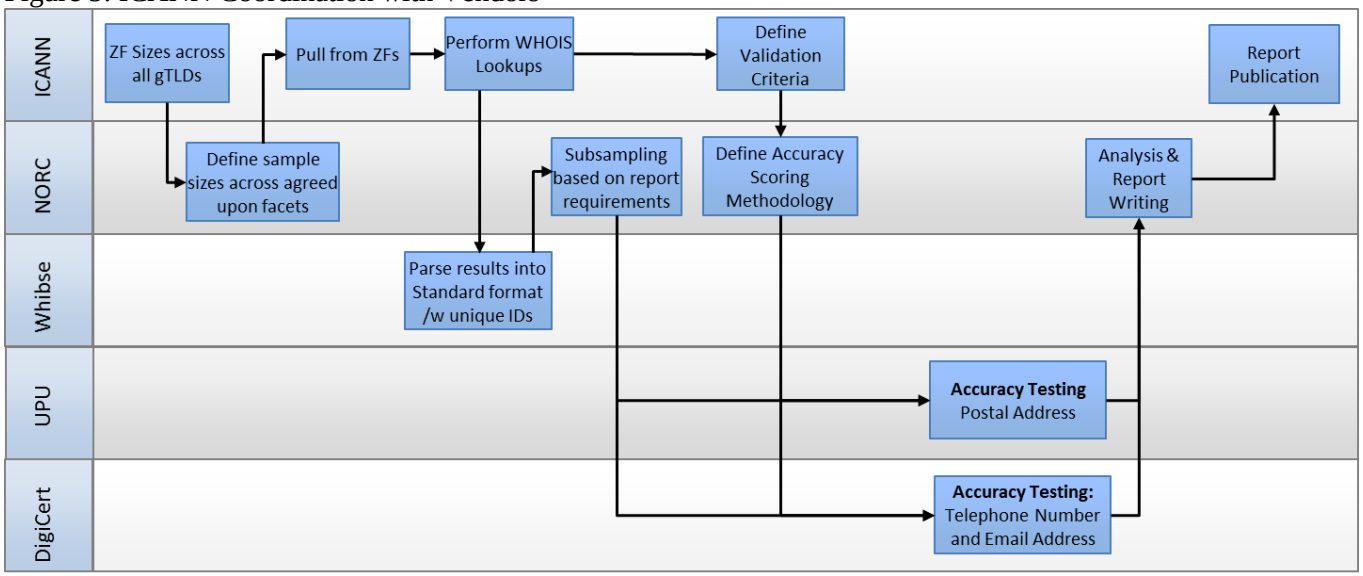
[16] With the exception of the Contractual Compliance follow-up, which will begin after publication of this report.

Figure 4: Cycle 2 Timeline



The WHOIS ARS team did not change from the previous Phases. Figure 5 illustrates how the team coordinated to develop the Cycle 2 report.

Figure 5: ICANN Coordination with Vendors



Study Methods and Approach

Brief Overview

In Cycle 2 we first selected a sample of 200,000 WHOIS records from the zone files of 588 gTLDs. Using systematic testing, the contact information from a subsample of 12,000 records was tested for accuracy with syntax standards (e.g., values and formats) based on requirements stipulated within the domain-applicable Registrar Accreditation Agreement (RAA), and was then tested for accuracy with operability standards (e.g., 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 98 percent of domain names are registered through registrars which operate under the 2013 RAA, a majority of domains with registrars on the 2013 RAA are obligated to meet only 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 that are 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, 2013 GF and 2013 NGF), the distribution for which can be seen in Chart 1.

Chart 1: Proportion of All Registrations in gTLDs, by RAA Status

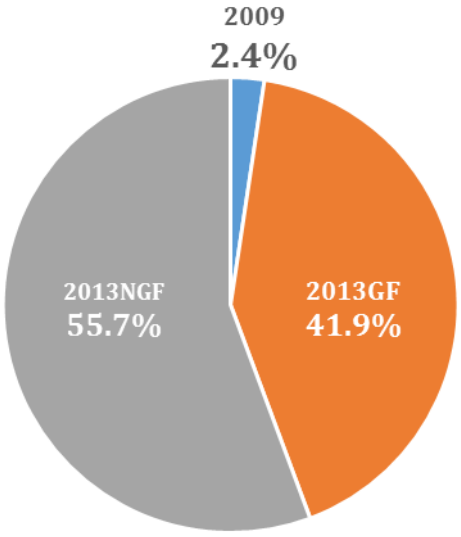
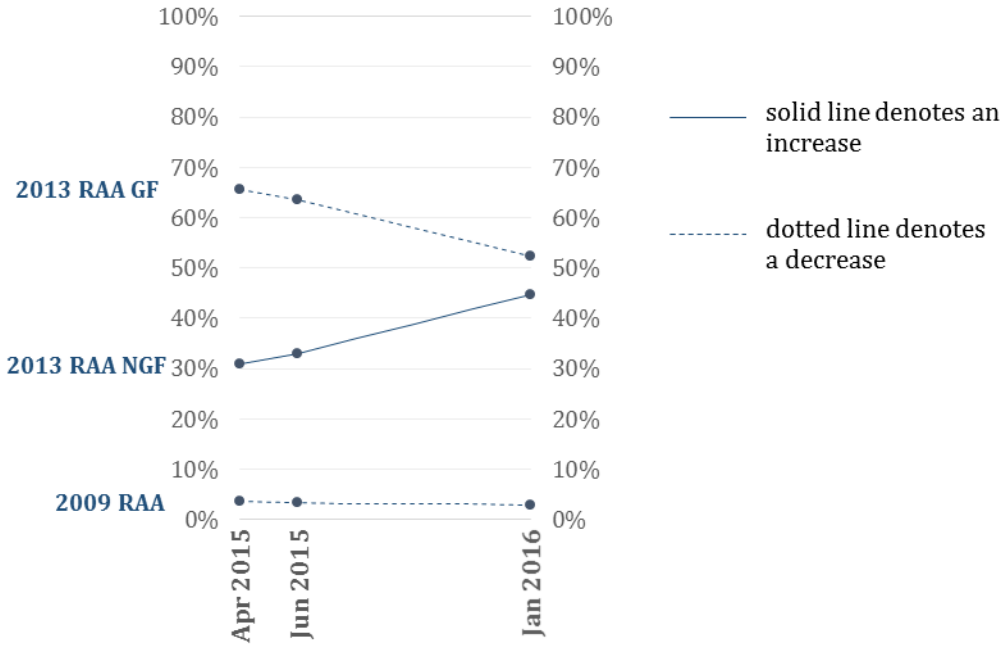


Table 1 and Graph 1 show the change in distribution over time of the three RAA types. They show that the 2009 RAA share is shrinking very slowly, but clearly shows that many grandfathered 2013 RAA domains are being converted to non-grandfathered domains.

Table 1: Distribution of RAA Type, by Sample Date

	2009 RAA	2013 GF RAA	2013 NGF RAA
April 2015 (Phase 1)	3.5%	65.7%	30.8%
June 2015 (Phase 2 Cycle 1)	3.3%	63.7%	33.0%
January 2016 (Phase 2 Cycle 2)	2.9%	52.4%	44.7%

Graph 1: Change in Distribution of RAA Type, by Sample Date[17]



Sample Design

Study data consisted of an initial sample of 200,000 records from gTLD zone files (this number was increased from 150,000 during Cycle 1), and an analyzed subsample of 12,000 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) 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. Summary statistics of the initial sample and the methods for selecting the subsample are described below.

[17] The pilot study sample has not been included here because the pilot did not sample from all domains available at the time.

Initial Sample

To select the initial sample of 200,000 records, we reviewed the zone file summary data, which indicates how many domains are in each gTLD. At the time of the initial sample for Cycle 2, in early January 2016, there were nearly 170 million domains names spread across 888 gTLDs. Approximately 94 percent of the 170 million domains were registered in one of the 18 Prior gTLDs, as compared to 96 percent in June 2015, when data was collected for Cycle 1 (see Table 1). Over 6 percent of domains in January 2016 were registered in New gTLDs, marking a dramatic increase from the 4 percent registered in June 2015. The overall number of New gTLDs also grew substantially, increasing from 660 in June 2015 to 870 in January 2016.

As the total number of delegated gTLDs grows, Prior gTLDs will remain constant at 18, while the number of New gTLDs will continue to increase. Table 2 shows the total number of delegated gTLDs and how many of these gTLDs were Prior vs. New gTLDs at each of the WHOIS ARS sample dates.

Table 2: Total Delegated, Prior and New gTLDs, by Sample Date

	Total Delegated gTLDs	Prior gTLDs	New gTLDs	gTLD Registrations (in millions)
April 2015 (Phase 1)	610	18	592	157
June 2015 (Phase 2 Cycle 1)	678	18	660	158
January 2016 (Phase 2 Cycle 2)	888	18	870	169.5

Out of the 870 New gTLDs, only 610 had at least one domain (260 New gTLDs did not yet have any domains), 40 had exactly one domain (these were excluded from our sample since it is typically an administrative domain for the gTLD) and the remaining 570 others had at least two domains. Adding together the 18 prior gTLDs and the 570 New gTLDs described above, the initial sample represented a total of 588 gTLDs.

Similar to the previous WHOIS ARS[18] study samples, our Cycle 2 sample design oversampled New gTLDs so that 25 percent of the initial sample was from New gTLDs[19]. Based on the lessons learned during the Cycle 1 study, we increased the initial sample size from 150,000 to 200,000 in order to decrease the oversampling necessary in the Analyzed Subsample (see Table 3).

Of the initial sample of 200,000, WHOIS data were gathered and parsed successfully for 196,262 (98.1 percent). Many of the remaining 3,738 domains no longer existed (1,395), but some requests timed out

[18] Previous WHOIS ARS studies include the Pilot Study, the Phase 1 study and the Phase 2 Cycle 1 study.

[19] To make sure all 570 New gTLDs with at least two domains were represented, we first selected one record from each, and the remaining sample was selected proportional to size (and thus more were selected from larger New gTLDs). Similarly, we selected a minimum of one domain from all Prior gTLDs, with the remaining sample proportional to size. All sampling was done by systematic sampling within a gTLD. Based on the sample size determined for each gTLD, a skip interval was determined (total number of domains divided by the desired sample size). Then a random start between zero and the skip interval was determined. If this random start was 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. 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.

repeatedly due to rate limits (2,134) and a few failed queries occurred for various reasons (209). Of the 196,262 domains, records in the 2009 RAA subgroup accounted for 2.4 percent of all records, while 2013 grandfathered (2013 GF) and 2013 non-grandfathered (2013 NGF) records accounted for 41.9 percent and 55.7 percent of all domains, respectively (see Table 3).

Table 3: Initial Sample Sizes by Region and RAA

RAA Type	Africa	Asia Pacific	Europe	Latin America and Caribbean	North America	Unknown	TOTAL
2009	30	450	619	306	3,258	19	**4,682**
2013 GF	457	14,379	18,275	2,184	46,564	438	**82,297**
2013 NGF	769	47,991	17,571	6,157	36,062	733	**109,283**
TOTAL	**1,256**	**62,820**	**36,465**	**8,647**	**85,884**	**1,190**	**196,262**

Analyzed Subsample

ICANN defined the subgroups of interest for this report as: 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, the analyzed subsample was selected in a way that maximizes the potential for estimating subgroups of interest with 95 percent confidence intervals of no more than plus or minus 5 percent. This kind of confidence interval required certain subgroups to be oversampled (or even directly included) relative to their representation in the initial sample of 200,000 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 449 registrars represented in the subsample. The subsample did not consider gTLD type (Prior vs. New) because the initial sample oversampled New gTLDs. Table 4 shows the sizes of the analyzed subsample by Region and RAA.[20]

Table 4: Analyzed Subsample Sizes by Region and RAA

RAA Type	Africa	Asia Pacific	Europe	Latin America and Caribbean	North America	Unknown	TOTAL
2009	30	450	619	306	800	5	**2,210**
2013 GF	457	1,000	1,000	800	1,401	13	**4,671**
2013 NGF	769	1,443	1,000	800	1,085	22	**5,119**
TOTAL	**1,256**	**2,894**	**2,619**	**1,906**	**3,285**	**40**	**12,000**

[20] In selecting the subsample of 12,000 domains that would be analyzed, the goal was to have 800 in each cell of the Region by RAA Type (Table 4). The number 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, and if a cell had greater than 10,000 in the initial sample, 1,000 were selected. Only three cells – North America 2013GF2013 GF, North America 2013 NGF, and Asia Pacific NGF – had more than 1,000 domains selected. We sampled the Region Unknown cells at the same proportion as the North America cells.

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 5 compares the sample sizes by RAA type in the initial sample of 196,262 and the analyzed subsample of 12,000.

Table 5: Sample Sizes by RAA Type

RAA Type	Percentage of All Domains	Initial Sample	Analyzed Subsample	Percentage of Subsample
2009 RAA	2.4%	4,682	2,210	18.4%
2013 RAA GF	41.9%	82,297	4,671	38.9%
2013 RAA NGF	55.7%	109,283	5,119	42.7%
TOTAL	100.0%	196,262	12,000	100.0%

Accounting for Common Data Across Contact Types

For all three contact modes (email, telephone and postal address), over 75 percent of the domains have the same contact information for all three contact types (registrant, administrative and technical). Table 6 shows the full distribution of how often the contact information is the same for each contact type.

Table 6: Frequency of Common Data Across Contact Type and Mode[21]

Commonality	Email	Telephone	Postal Address
All Three Exactly Same	77.6% ± 0.7%	80.3% ± 0.7%	78.2% ± 0.7%
Exactly Two the Same, One Different	20.1% ± 0.7%	18.6% ± 0.7%	19.8% ± 0.7%
All Three Different	2.3% ± 0.3%	1.0% ± 0.2%	2.0% ± 0.3%

The commonality figures in Table 6 indicate 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 2.3 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 three contact types are accurate. An expanded version of Table 6 can be found as Table B1 in Appendix B.

Syntax and Operability Testing Methods

Syntax and operability accuracy tests were designed in such a way that all records in the analyzed subsample would be evaluated against a set of baseline requirements derived from the requirements of the 2009 RAA[22]. Tests were performed on all nine individual contact information fields in a record (i.e., for the three contact modes of email address, telephone number and postal address, within each of the three contact types of registrant, administrative and technical contacts) and then the results were compiled for the entire record. Information on accuracy test criteria and links to more detailed testing information can be found in Appendix A.

[21] An expanded version of Table 4 can be found as Table B1 in Appendix B.
[22] Additional tests to the 2013 RAA requirements are provided in Appendix C of this report.

Syntax Testing Methods

Syntax testing was designed to assess the contact information of a record by comparing it to formats specified by contractual requirements stipulated in the RAAs. Tests were administered in two stages[23] of testing: stage one testing verified the presence of contact information as required by applicable RAA, and stage two involved detailed technical testing of the syntax. Syntax testing criteria have remained consistent across all previous WHOIS ARS studies.

Operability Testing Methods

Operability testing was designed to assess whether the contact information of a record can be used practically for communication. In consultation with the community, including volunteers from the registrar community, operability validation criteria were developed to align with RAA requirements. 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 were removed).

[23] The stage one and stage two syntax tests for each contact mode are described in detail on the WHOIS ARS webpage: https://whois.icann.org/en/whoisars-validation.

Main Findings

Here we present our findings and the statistics related to the outcomes of the syntax and operability accuracy tests. This section of the report includes a summary of the key findings, followed by a detailed statistical analysis of the syntax and operability test outcomes. These statistics are organized by contact type[24] (registrant, administrative and technical) within contact mode (email address, telephone number and postal address), overall and across the subgroups of New vs. Prior gTLDs, RAA type and ICANN region. Further detail on the findings, including analysis tables, can be found in Appendix B.[25]

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[26]. Since operability results are similar across RAA versions, separate analysis tables for each set of requirements would be largely redundant. Analysis tables presenting the outcomes of syntax tests for 2013 RAA requirements can be found in Appendix C.

Summary of Findings

We present here the key takeaways from the findings:

Ability to Establish Contact

- Ninety-nine percent of records had at least one contact mode of the three contact types that met all syntax and operability requirements of the 2009 RAA, which implies that nearly all records contain information that can be used to establish contact. In only 1 percent of records were there no contact modes for any contact type that met syntax or operability requirements.

Operability Accuracy

- Ninety-eight percent of postal addresses, 76 percent of telephone numbers and 91 percent of email addresses met all operability requirements of the 2009 RAA. Seventy percent of domains passed all operability tests for all contact types (registrant, administrative and technical) and contact modes (email, telephone and postal address), which is about a 6% increase from Cycle 1.
  - Regional variations of operability accuracy are greatest for telephone, which ranges from 63.7 percent accurate (Asia-Pacific) to 85.3 percent accurate (North America).

[24] Because the numbers for the registrant, administrative and technical contacts are so similar, we present here subgroup accuracy only for "All Three Accurate", i.e., the registrant, administrative and technical contacts all passed all of the accuracy tests.

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

[26] 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 from, 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 Universal Postal Union S42 template for a particular country. Any contact field that meets the 2013 RAA requirements would also meet 2009 requirements. For this reason, the 2009 requirements serve as a baseline against which all records can be compared.

- 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 the small numbers of postal addresses that failed operability testing, almost 40 percent of those did not have an identifiable or easily deduced country.
- For operability errors for email addresses, about 8.5 percent of the email addresses bounced, compared to less than 1 percent being missing.

Syntax Accuracy:

- Eighty-five percent of telephone numbers met all syntax requirements of the 2009 RAA, increasing slightly from Cycle 1 (83 percent) and aligning more closely with Phase 1 findings (86 percent). The reasons for syntax errors had very similar distributions to those in Cycle 1.
  - Regional variations of syntax accuracy were greatest for postal address, which ranges from 44.6 percent accurate (Africa) to 96.7 percent accurate (North America).
  - The most common reason for telephone syntax error in most regions was incorrect length, but in North America the most common reason for error was a missing country code.
  - For postal addresses, the vast majority of errors in each study have consistently been due to missing fields that were required, such as city, state/province, postal code or street.

Syntax Accuracy – 2009 RAA Requirements[27]

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

Overall Syntax Accuracy

First, we look at accuracy to 2009 RAA requirements for all 12,000 domains in the analyzed subsample. The dotted black line in Graph 2 shows that around 67 percent of domains can be said to be syntactically accurate. Ninety-nine percent of email addresses passed the syntax accuracy tests, while 85 percent of telephone numbers and 77 percent of postal addresses passed.

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

Graph 2: Overall Accuracy – 2009 RAA Syntax Requirements

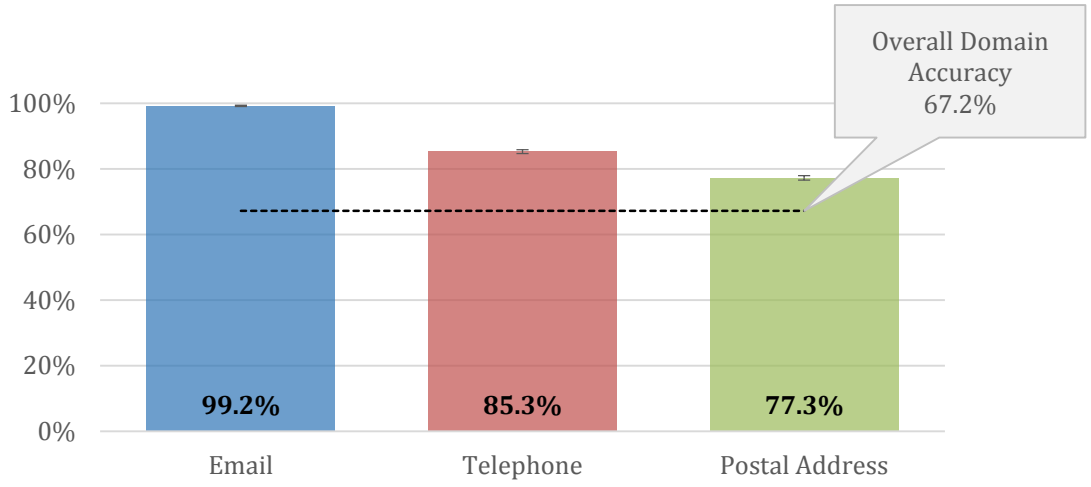


Table 7 shows a more detailed breakdown of the data by contact type. The bottom row of this table shows the rate for which the registrant, administrative and technical contacts all passed syntax tests for a given contact mode (email, telephone or postal address).[28] We will focus on the percentages for all three contact modes passing all accuracy tests (the "All Three Accurate row") in the subgroup analyses.

Table 7: Overall Accuracy by Contact Type and Mode – 2009 RAA Syntax Requirements

	Email	Telephone	Postal Address	All Three Accurate
Registrant	100.0% ± 0.0%	88.1% ± 0.6%	79.0% ± 0.7%	**69.8% ± 0.8%**
Administrative	99.2% ± 0.2%	86.5% ± 0.6%	78.6% ± 0.7%	**69.2% ± 0.8%**
Technical	99.2% ± 0.2%	86.9% ± 0.6%	80.9% ± 0.7%	**71.8% ± 0.8%**
Overall	**99.2% ± 0.2%**	**85.3% ± 0.6%**	**77.3% ± 0.7%**	**67.2% ± 0.8%**

Syntax Accuracy by Prior vs. New gTLDs

Graph 3 and Table 8 show that Prior gTLDs have lower[29] accuracy on email addresses and telephone numbers, but higher accuracy on postal addresses. Prior gTLDs also have a higher rate of having all three contact fields accurate.

[28] The accuracy rates within each contact mode are very similar across contact types due to the high frequency of commonality across the contact data (as shown in Table 4), suggesting that accuracy within each contact mode will be about the same regardless of whether the information is contained in the registrant, administrative or technical contact fields.

[29] Here "higher" and "lower" refer not only to sheer numbers, but also statistical significance. This latter phrase has been left out of most of the narrative for ease of reading.

Graph 3: Accuracy by gTLD Type – 2009 RAA Syntax Requirements

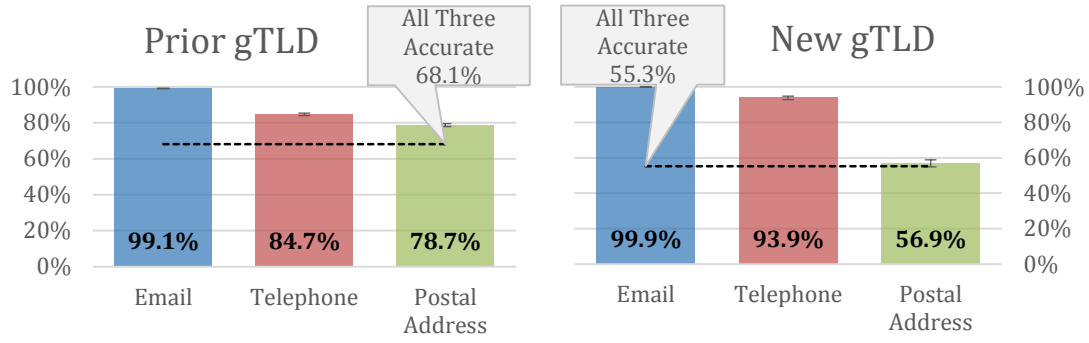


Table 8: Accuracy by gTLD Type – 2009 RAA Syntax Requirements

	Email	Telephone	Postal Address	All Three Accurate
Prior gTLD	99.1% ± 0.2%	84.7% ± 0.7%	78.7% ± 0.8%	**68.1% ± 0.9%**
New gTLD	99.9% ± 0.1%	93.9% ± 1.0%	56.9% ± 2.0%	**55.3% ± 2.0%**
Overall	**99.2% ± 0.2%**	**85.3% ± 0.6%**	**77.3% ± 0.7%**	**67.2% ± 0.8%**

Syntax Accuracy by RAA Status

Next, we look at accuracy rates by RAA status. Graph 4 and Table 9 show a small, yet statistically significant difference in email address accuracy between 2013 GF and 2013 NGF groups, while the 2009 RAA group has similar accuracy to both 2013 groups. The 2013 RAA GF group has the lowest telephone accuracy, while 2009 and 2013 NGF have similar telephone accuracy. Postal address accuracy is highest for the 2009 RAA group, and lowest for the 2013 NGF group. Despite not having the highest accuracy for every contact mode, the 2009 RAA group does have the highest percentage of records in which all three contact modes were accurate.

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

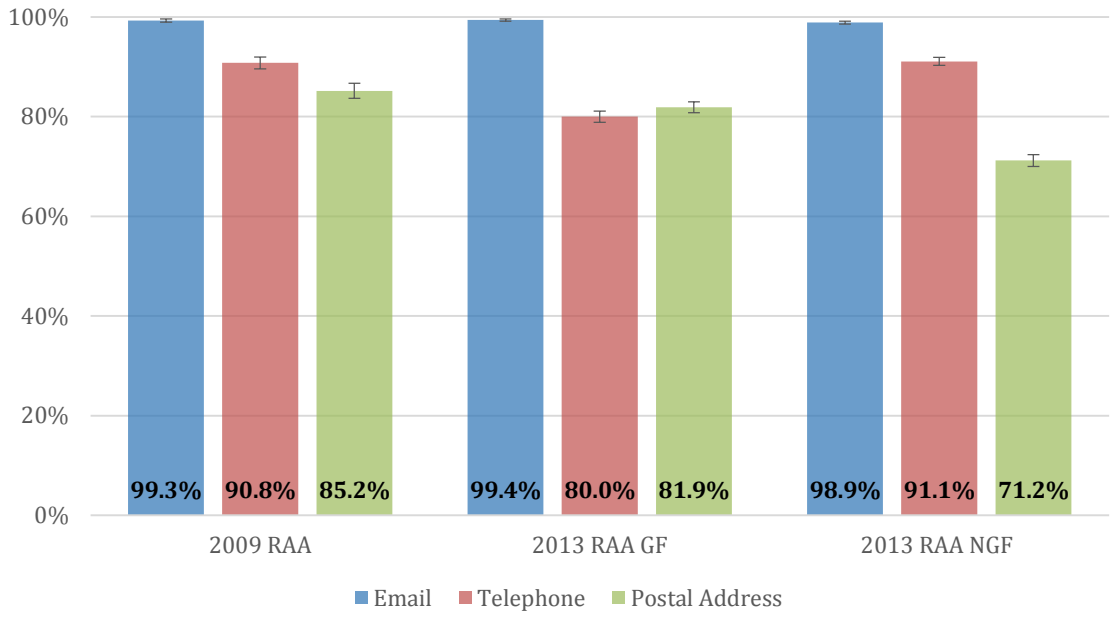


Table 9: Accuracy by RAA Status – 2009 RAA Syntax Requirements

	Email	Telephone	Postal Address	All Three Accurate
2009 RAA	99.3% ± 0.3%	90.8% ± 1.2%	85.2% ± 1.5%	**80.9% ± 1.6%**
2013 RAA GF	99.4% ± 0.2%	80.0% ± 1.1%	81.9% ± 1.1%	**66.8% ± 1.4%**
2013 RAA NGF	98.9% ± 0.3%	91.1% ± 0.8%	71.2% ± 1.2%	**66.9% ± 1.3%**
Overall	**99.2% ± 0.2%**	**85.3% ± 0.6%**	**77.3% ± 0.7%**	**67.2% ± 0.8%**

Syntax Accuracy by ICANN Region

Finally, we look at accuracy by ICANN region. Graph 5 and Table 10 show that the Asia-Pacific region has lower email address accuracy than the other four regions. For telephone numbers, the syntax accuracy for all regions except Africa ranges between 84.3 percent (Latin America/Caribbean) and 88.9 percent (Asia-Pacific). In contrast, during Cycle 1 only two regions (Europe and North America) had a telephone accuracy higher than 80 percent. For postal addresses, results were similar to Cycle 1, with North America showing the highest accuracy, and Africa and Asia-Pacific showing the lowest accuracy. The same pattern was observed for having all three contact modes conform to the RAA 2009 standards, with North America showing the highest for "All Three Accurate," and Africa and Asia-Pacific showing the lowest accuracy. More information on regional accuracy statistics and reasons for error by region, see the section Regional Findings.

Graph 5: Accuracy by ICANN Region – 2009 RAA Syntax Requirements

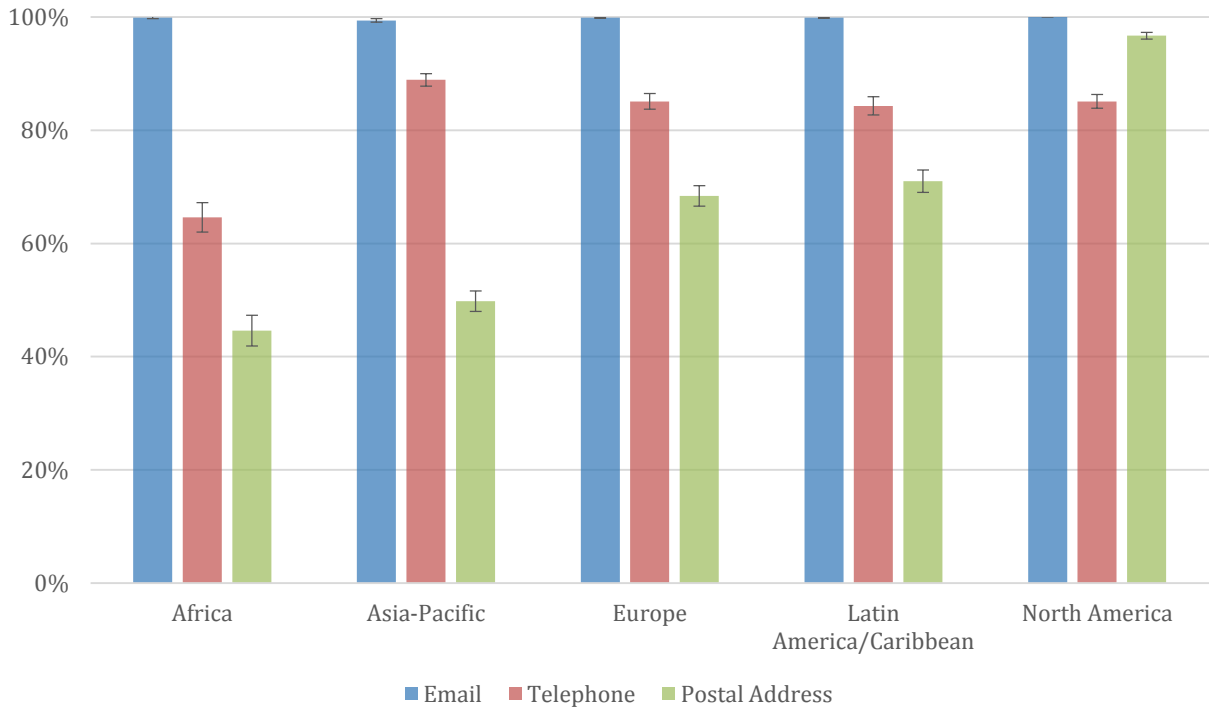


Table 10: Accuracy by ICANN Region – 2009 RAA Syntax Requirements

	Email	Telephone	Postal Address	All Three Accurate
Africa	99.9% ± 0.2%	64.6% ± 2.6%	44.6% ± 2.7%	**29.3% ± 2.5%**
Asia-Pacific	99.4% ± 0.3%	88.9% ± 1.1%	49.8% ± 1.8%	**45.0% ± 1.8%**
Europe	99.9% ± 0.1%	85.1% ± 1.4%	68.4% ± 1.8%	**60.6% ± 1.9%**
Latin America/Caribbean	99.9% ± 0.1%	84.3% ± 1.6%	71.0% ± 2.0%	**64.7% ± 2.1%**
North America	100.0% ± 0.0%	85.1% ± 1.2%	96.7% ± 0.6%	**82.8% ± 1.3%**
Overall	**99.2% ± 0.2%**	**85.3% ± 0.6%**	**77.3% ± 0.7%**	**67.2% ± 0.8%**

Reasons for Error – 2009 RAA Syntax Requirements

We report here on the major reasons for failure for each contact mode separately (email address, telephone number and postal address). For email addresses and telephone numbers, we were able to pinpoint the first test that failed. Because postal addresses require multiple fields, multiple syntax errors were possible. 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 12,000 analyzed domains.[30] In Cycle 1, we showed by

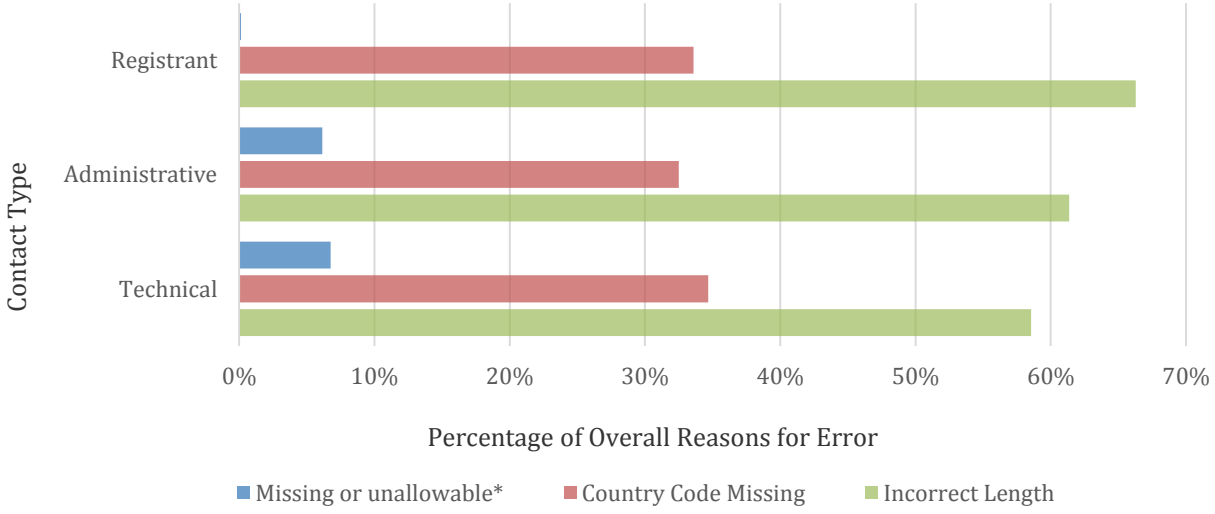
[30] In Appendix C you can find the major reasons for failure against the 2013 RAA requirements among the 2013 NGF subgroup.

contact mode which accuracy tests a record failed. We repeat these graphs for Cycle 2 here. Appendix B shows tabular data from Cycle 1 and Cycle 2 for comparison, and also includes tables for email addresses (which are omitted here because very few syntax errors were found).

Telephone Numbers

Graph 6 shows the reasons for telephone number errors as a percentage of all telephone number errors, by contact type. Similar to Cycle 1, the largest source of errors among telephone numbers was having an incorrect length for the applicable country (around 8.7 percent of all telephone numbers tested). The next most common error was a missing country code (almost 4.7 percent of all telephone numbers tested). Less than 1 percent of telephone numbers were missing.

Graph 6: Reasons for Telephone Number Error – 2009 RAA Syntax Requirements

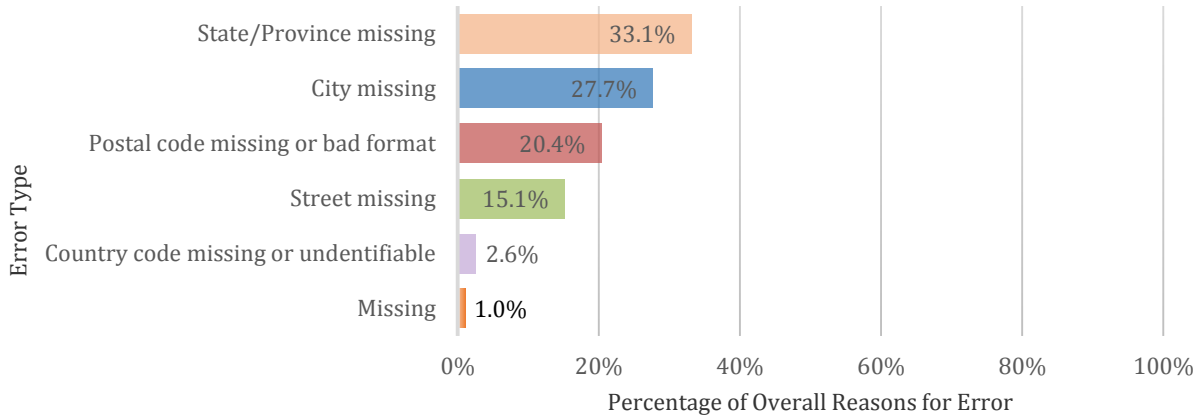


*Note: 2009 RAA does not require presence of a telephone number for the registrant contact type.

Postal Addresses

Graph 7 shows the reasons for postal address errors as a percentage of all postal address errors. Similar to Cycle 1, the majority of postal address syntax errors (96.7 percent) were due to a missing address component, such as a missing state/province (33.1 percent), city (27.7 percent), postal code (20.4 percent) and/or street (15.1 percent). Fewer country codes were missing (only 2.6 percent of all telephone syntax errors) and few telephone numbers were missing (only 1 percent of all telephone syntax errors).

Graph 7: Reasons for Postal Address Error Across All Contact Types – 2009 RAA Syntax Requirements



Operability Accuracy – 2009 RAA Requirements[31]

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. It is important to note here that the only difference between 2013 and 2009 RAA operability requirements is that the 2009 RAA requirements do not require that information be present in the registrant email or telephone number fields, while 2013 RAA requirements do require the presence of information in these fields.

Overall Operability Accuracy

First, we look at accuracy to 2009 RAA requirements for all 12,000 domains. In the syntax test results, we saw that accuracy rates were lowest for postal addresses. However, for operability, accuracy rates are highest for postal addresses, as shown in Graph 8 and Table 11.[32] For the other two contact modes, email and telephone, operability accuracy rates are lower than syntax rates. More details about the operability accuracy testing process can be found in Appendix A.

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

[32] Postal address operability testing can be more forgiving than postal address syntax testing. For example, syntax conformance with Universal Postal Union standards for postal mail requires an abbreviation for state or province (e.g., in the United States, 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 "@" symbol) will not be operable; a syntactically inaccurate postal address can still be operable (i.e., deliverable). *See* also discussion in Challenges and Lessons Learned.

Graph 8: Overall Accuracy – 2009 RAA Operability Requirements

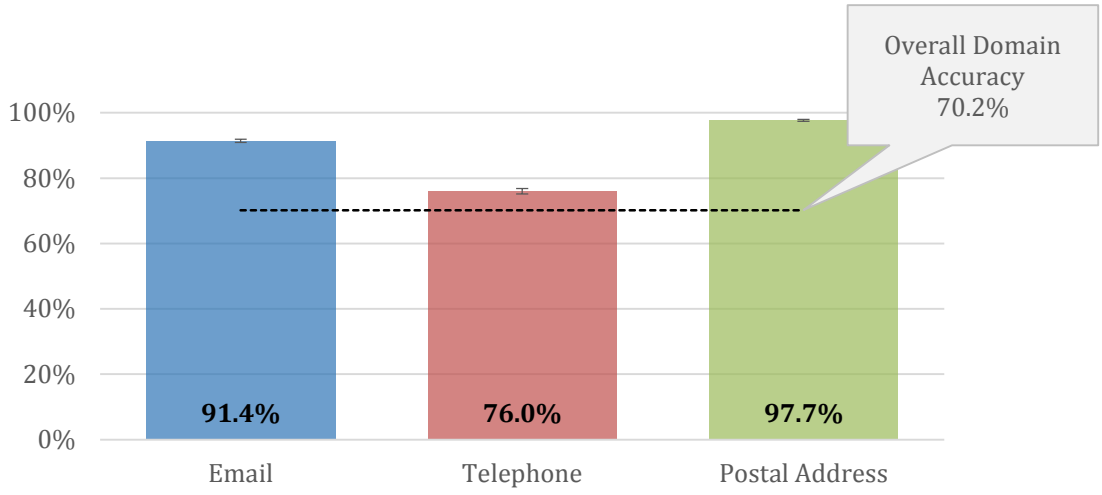


Table 11 shows the breakdown by contact type in more detail.

Table 11: Overall Accuracy by Contact Type and Mode – 2009 RAA Operability Requirements

	Email	Telephone	Postal Address	All Three Accurate
Registrant	93.3% ± 0.4%	81.2% ± 0.7%	97.9% ± 0.3%	**74.8% ± 0.8%**
Administrative	92.5% ± 0.5%	80.1% ± 0.7%	97.8% ± 0.3%	**74.6% ± 0.8%**
Technical	93.1% ± 0.5%	78.6% ± 0.7%	97.8% ± 0.3%	**73.5% ± 0.8%**
Overall	**91.4% ± 0.5%**	**76.0% ± 0.8%**	**97.7% ± 0.3%**	**70.2% ± 0.8%**

Operability Accuracy by Prior vs. New gTLD

Graph 9 and Table 12 show that Prior gTLDs have lower operability accuracy on email and postal addresses, but higher accuracy on telephone numbers. These results are similar to the findings of Cycle 1.

Graph 9: Accuracy by gTLD Type – 2009 RAA Operability Requirements

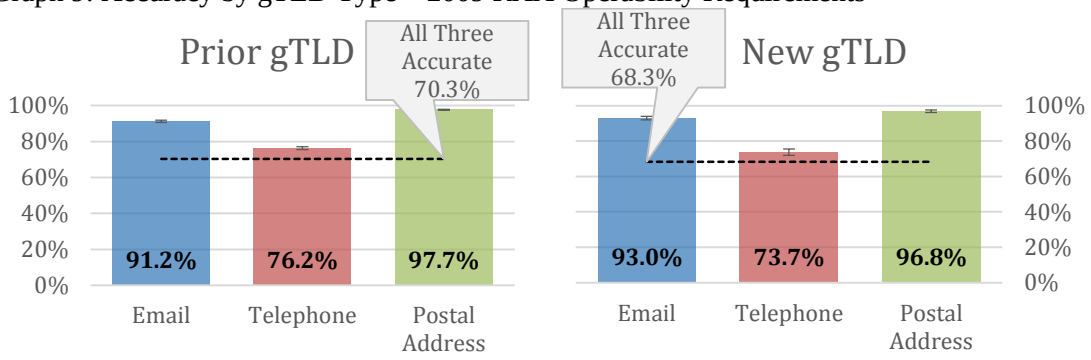


Table 12: Accuracy by gTLD Type – 2009 RAA Operability Requirements

	Email	Telephone	Postal Address	All Three Accurate
Prior gTLD	91.2% ± 0.6%	76.2% ± 0.8%	97.7% ± 0.3%	**70.3% ± 0.9%**
New gTLD	93.0% ± 1.0%	73.7% ± 1.8%	96.8% ± 0.7%	**68.3% ± 1.9%**
Overall	**91.4% ± 0.5%**	**76.0% ± 0.8%**	**97.7% ± 0.3%**	**70.2% ± 0.8%**

Operability Accuracy by RAA Status

Finally, we look at accuracy by RAA status. Graph 10 and Table 13 show that the 2013 RAA GF and 2013 RAA NGF groups both have higher email address accuracy than the 2009 RAA group, while the 2009 RAA group has higher telephone accuracy than the other two groups. The 2013 RAA NGF group has the highest email accuracy and does not have lowest accuracy for any of the contact modes, and thus has the highest percentage of all three modes being accurate.

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

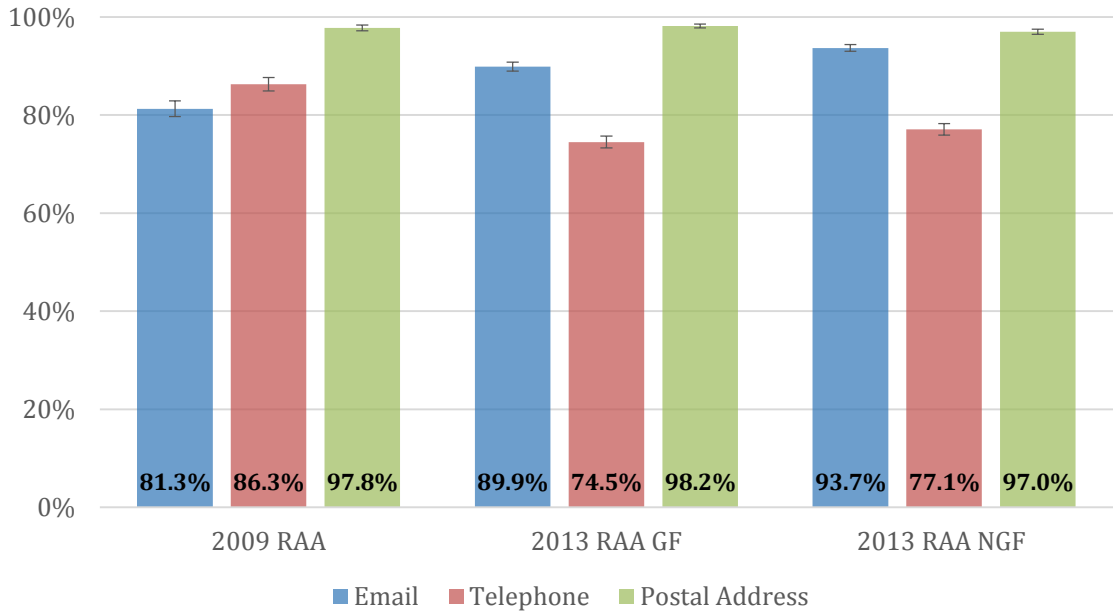


Table 13: Accuracy by RAA Status – 2009 RAA Operability Requirements

	Email	Telephone	Postal Address	All Three Accurate
2009 RAA	81.3% ± 1.6%	86.3% ± 1.4%	97.8% ± 0.6%	**69.7% ± 1.9%**
2013 RAA GF	89.9% ± 0.9%	74.5% ± 1.2%	98.2% ± 0.4%	**68.4% ± 1.3%**
2013 RAA NGF	93.7% ± 0.7%	77.1% ± 1.2%	97.0% ± 0.5%	**72.3% ± 1.2%**
Overall	**91.4% ± 0.5%**	**76.0% ± 0.8%**	**97.7% ± 0.3%**	**70.2% ± 0.8%**

Operability Accuracy by ICANN Region

Next, we look at accuracy by ICANN region. Graph 11 and Table 14 show that for email addresses, African, Asian-Pacific, and North American domains have higher operability accuracy rates. 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. With regard to all nine contacts passing all accuracy tests, Latin American/Caribbean and North American domains have higher rates, and Asian-Pacific domains have a lower rate. More information on regional accuracy statistics and reasons for error by region, see the section Regional Findings.

Graph 11: Accuracy by ICANN Region – 2009 RAA Operability Requirements

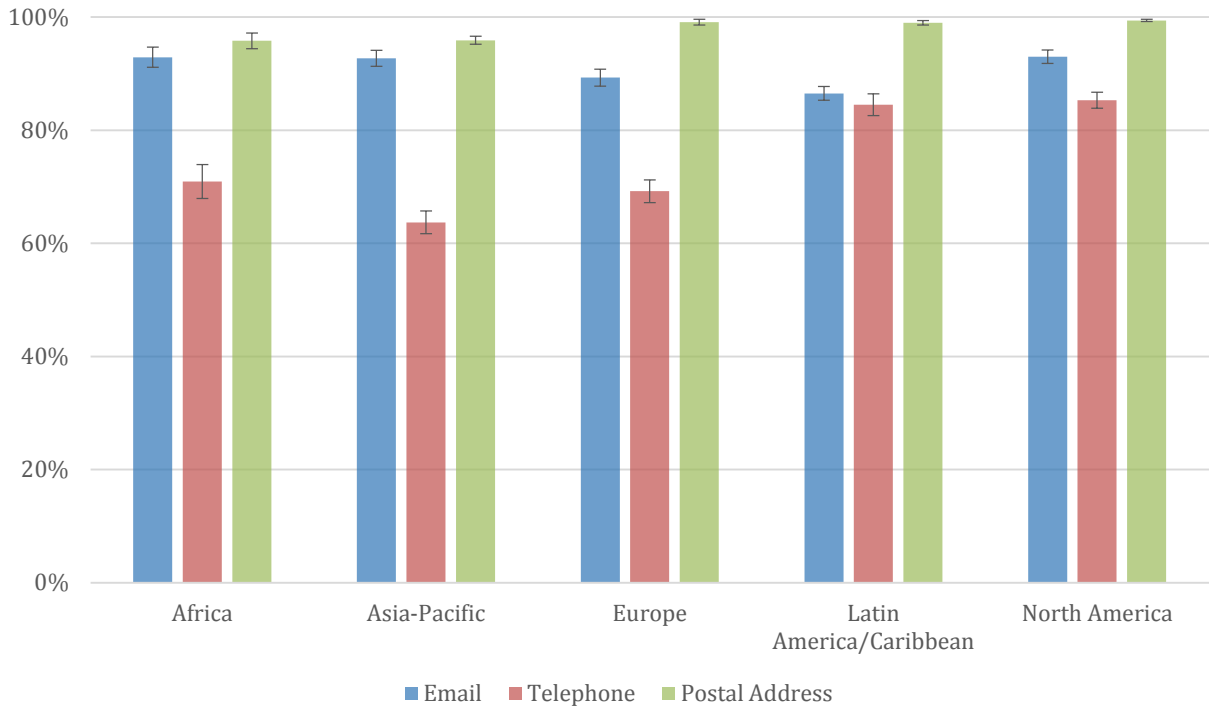


Table 14: Accuracy by ICANN Region – 2009 RAA Operability Requirements

	Email	Telephone	Postal Address	All Three Accurate
Africa	92.9% ± 1.4%	70.9% ± 2.5%	95.8% ± 1.1%	**64.6% ± 2.6%**
Asia-Pacific	92.7% ± 0.9%	63.7% ± 1.8%	95.9% ± 0.7%	**57.6% ± 1.8%**
Europe	89.3% ± 1.2%	69.2% ± 1.8%	99.1% ± 0.4%	**63.1% ± 1.8%**
Latin America/Caribbean	86.5% ± 1.5%	84.5% ± 1.6%	99.0% ± 0.4%	**71.6% ± 2.0%**
North America	93.0% ± 0.9%	85.3% ± 1.2%	99.4% ± 0.3%	**80.2% ± 1.4%**
Overall	**91.4% ± 0.5%**	**76.0% ± 0.8%**	**97.7% ± 0.3%**	**70.2% ± 0.8%**

Reasons for Error – 2009 RAA Operability Requirements[33]

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.

[33] To find more information on how the tests were conducted and how the errors map to the tests, *see* Appendix A or the WHOIS ARS webpage: https://whois.icann.org/en/whoisars-validation.

Email Addresses

Table 15 shows that around 8.5 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 contact fields.

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

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	10,995	10,943	10,898	**32,836**
Not Verifiable (or Missing)	130*	49	51	**100**
Email Bounced	1,005	1,008	1,051	**3,064**
Total	**12,000**	**12,000**	**12,000**	**36,000**

* Registrant email is not required under the 2009 RAA.

Telephone Numbers

Graph 12 and Table 16 show the operability errors for telephone numbers. Once again, a registrant telephone number is not required under the 2009 RAA. If a verifiable telephone number was given, it was dialed. Table 16 shows that approximately 4.5 percent of the numbers were disconnected, another 6.1 percent were invalid and another 9.4 percent did not connect. The percentage of required telephone numbers that were missing was less than 1 percent.

Graph 12: Telephone Number Errors by Contact Type – 2009 RAA Operability Requirements

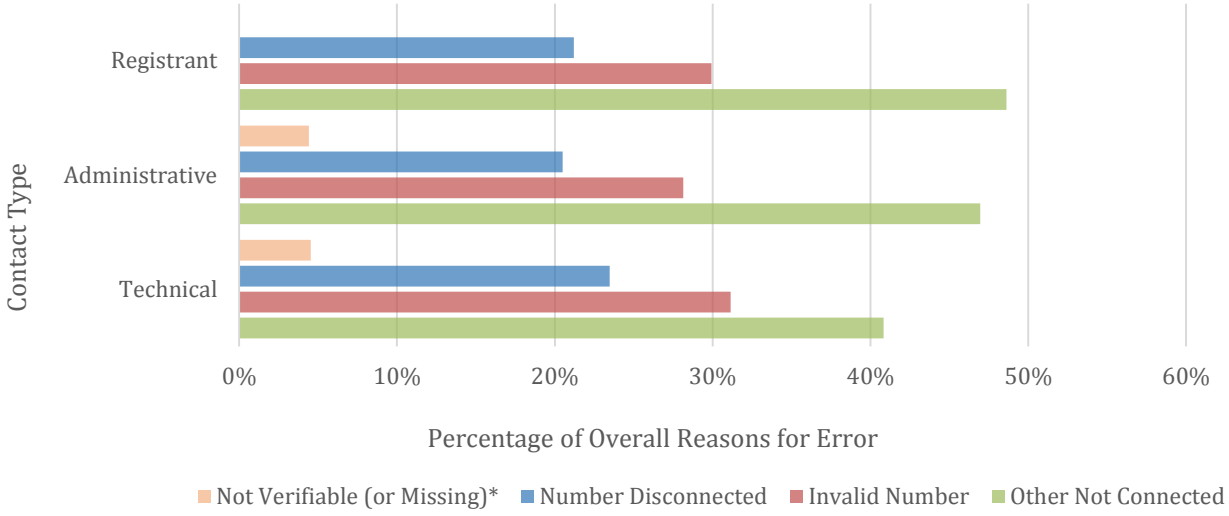


Table 16: Telephone Number Errors by Contact Type – 2009 RAA Operability Requirements

	Administrative	Technical	Registrant	Total
Passed All Accuracy tests	9,618	9,506	9,444	**28,568**
Not Verifiable (or Missing)	188*	110	116	**232**
Number Disconnected	505	511	600	**1,616**
Invalid Number	713	702	796	**2,211**
Other Not Connected	1,158	1,171	1,044	**3,373**
Total	**12,000**	**12,000**	**12,000**	**36,000**

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

Postal Addresses

Finally, Graph 13 and Table 17 show the postal address errors for operability. The postal addresses are first coded for operability automatically by the Universal Postal Union's operability testing tool.[34] V (verified as is), C (corrected and verified) and P3 (probably deliverable) are all considered operable postal addresses.[35] The inoperable postal addresses are categorized as P2 (might not be deliverable), P1 (probably not deliverable), N1 (country unknown) and N2 (unverifiable address due to country standards not available). However, a manual process[36] is carried out for the P1 and P2 addresses, allowing most postal addresses to be identified as operable. Table 17 shows that 92.9 percent of the P2 addresses and 72.0 percent of the P1 addresses were determined to be operable using a manual process. Graph 13 shows only those addresses still determined to be inoperable after this manual process. Table 17 shows more detail, including how many in each code were determined to be operable by the manual process.

[34] The Universal Postal Union, 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, and does so without attempting physical delivery to the destination. Information regarding accuracy testing can be found in Appendix A and on the WHOIS ARS webpage: https://whois.icann.org/en/whoisars-validation.

[35] *See* note 34 and discussion in Challenges and Lessons Learned.

[36] This manual process is also intended to simulate post office protocols that – for an address that is technically incorrect – can sometimes determine an operable address by analyzing components of the address data (assuming at least some address data were provided). As can be seen in the table, a correct address cannot always be determined, (i.e., deemed operable).

Graph 13: Postal Address Errors Across All Contact Types – 2009 RAA Operability Requirements

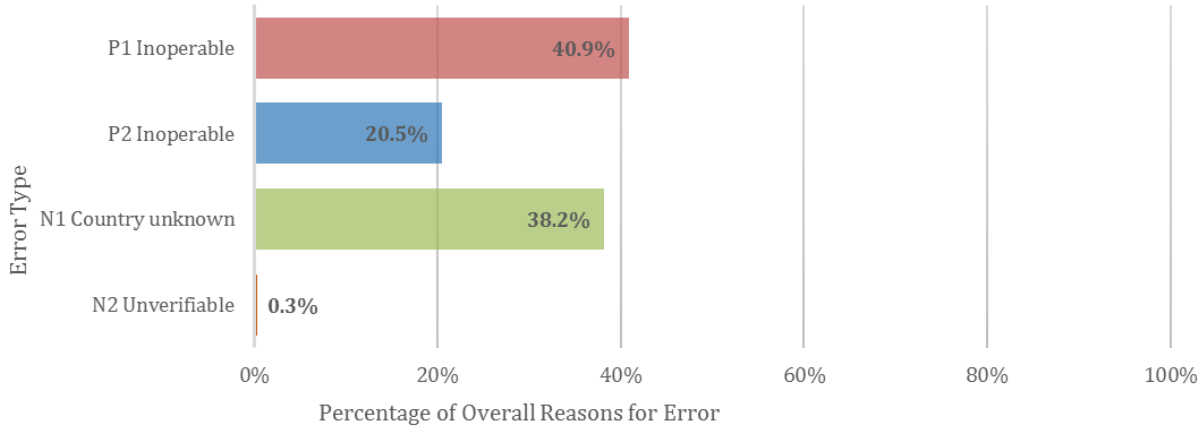


Table 17: Postal Address Errors by Contact Type – 2009 RAA Operability Requirements

	Administrative	Technical	Registrant	Total
Operable	10,432	10,445	10,480	31,357
Operable P2	829	809	823	2,461
Operable P1	439	447	409	1,295
TOTAL OPERABLE	11,700	11,701	11,712	35,113
Inoperable P2	63	61	58	182
Inoperable P1	122	124	117	363
N1, Country unknown	114	113	112	339
N2, Unverifiable	1	1	1	3
TOTAL INOPERABLE	300	299	288	887
OVERALL TOTAL	12,000	12,000	12,000	36,000

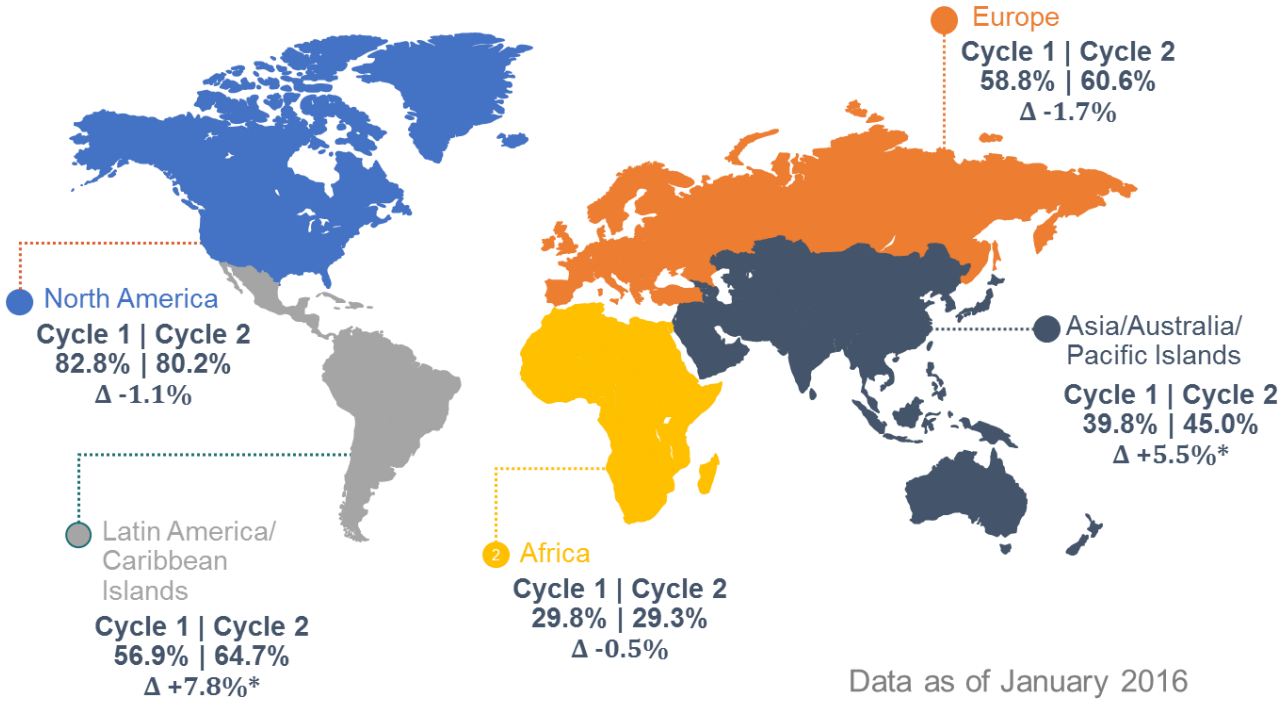
Regional Findings – Analysis of Accuracy and Reasons for Error by Region

Here we report on additional regional statistics, such as changes by region in overall syntax and operability accuracy between Cycle 1 and Cycle 2, reasons for syntax and operability errors in Cycle 2 by region, and finally, syntax and operability accuracy by the script language used to register a domain.

Changes in Overall Accuracy by Region

Syntax Accuracy
Chart 2 shows that changes in syntax accuracy were most pronounced in the Asia-Pacific and Latin America/Caribbean regions, which increased by 5.5 percent and 7.8 percent, respectively. Overall syntax accuracy across all regions decreased from Cycle 1 to Cycle 2 by 0.7 percent to 67.2 percent (see below for trend analysis).

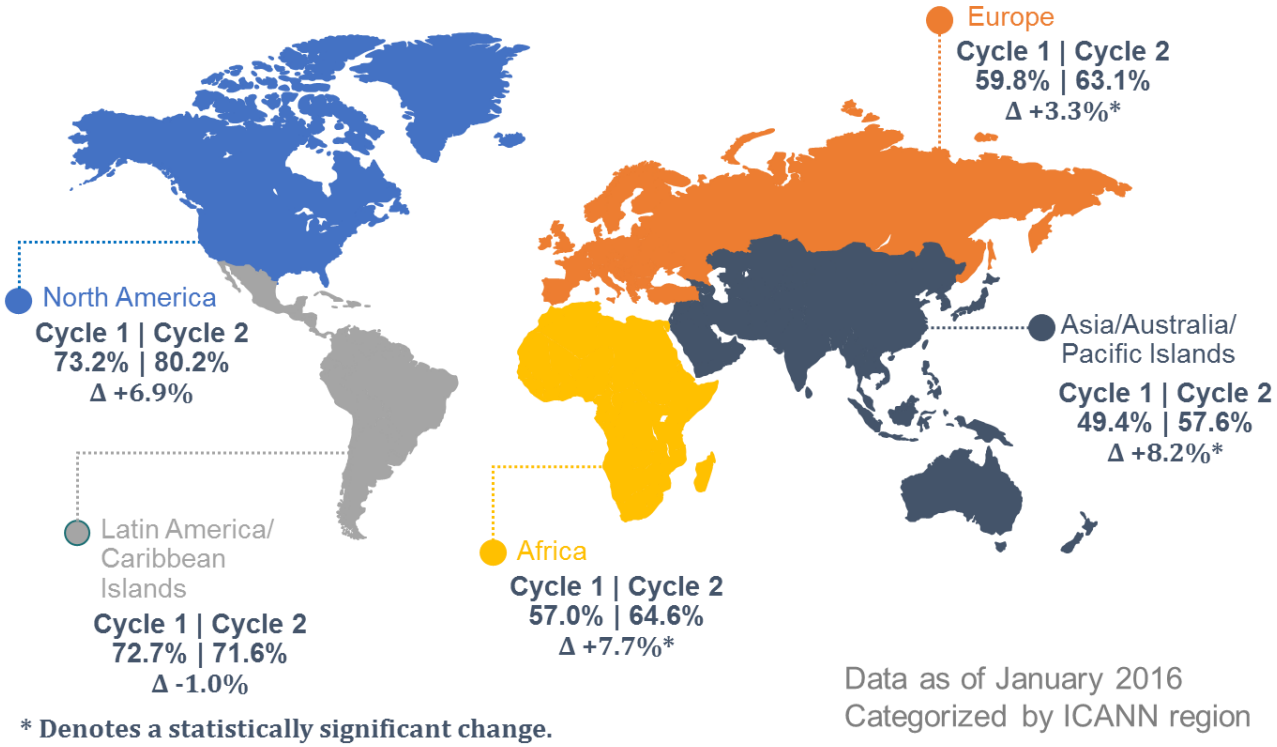
Chart 2: Change in Overall Syntax Accuracy for ICANN Regions, by ARS Cycle – 2009 RAA Requirements



*** Denotes a statistically significant change.**

Operability Accuracy
Chart 3 shows that operability accuracy increased in every region except for Latin America/Caribbean, ranging from a 3.3 percent increase in the Europe region to an 8.2 percent increase in the Asia-Pacific region. Overall operability accuracy across all regions increased between Cycle 1 and Cycle 2 by about 5.4 percent to 70.2 percent (see below for trend analysis).

Chart 3: Change in Overall Operability Accuracy for ICANN Regions, Cycle 1 to Cycle 2 – 2009 RAA Requirements



*** Denotes a statistically significant change.**

Reasons for Error by Region

We report here the major reasons for syntax and operability testing errors by region, separated by contact mode (email address, telephone number and postal address). For email addresses and telephone numbers, we report the first test failed. Because postal addresses require multiple fields, multiple errors were possible.

Reasons for Email Syntax and Operability Error by Region – 2009 RAA
The syntax errors in Table 18 show significant variability by region. However, it is important to remember that the actual number of syntax errors for email addresses is very small. Most of the errors are the result of missing email addresses.

Table 18: Reasons for Email Syntax Error by Region – 2009 RAA Requirements

Error	Africa	Asia Pacific	Europe	Latin America and Caribbean	North America	Unknown	All Regions
Missing	0.0%	75.9%	100.0%	0.0%	100.0%	100.0%	86.8%
Characters Not Allowed	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
@ Symbol Missing	0.0%	3.4%	0.0%	40.0%	0.0%	0.0%	2.6%
Not Resolvable	100.0%	20.7%	0.0%	60.0%	0.0%	0.0%	10.5%
Overall Accuracy for Region – Email Syntax	99.9% ± 0.2%	99.4% ± 0.3%	99.9% ± 0.1%	99.9% ± 0.1%	100.0% ± 0.0%	N/A	**99.2% ± 0.2%**

Note: This table should be read as follows: Of the errors in X region, Y% were for Z reason (e.g., of the reasons for syntax error among email addresses from Africa, 100% of the errors were due to a non-resolvable address). The "Overall Email Syntax Accuracy for Region" is not a total of the percentages above it, but is included rather to provide additional context for the errors. That is, 100% of the email syntax errors in Africa were due to non-resolvable addresses, but overall, Africa's email address syntax accuracy was 99.9%, meaning the actual number of errors was very small.

The operability errors in Table 19 show that email addresses have two main categories of operability errors: missing/non-verifiable, or an email address that bounces. In every region, email errors were largely due to bounced emails, but when the region was unknown, almost all errors were due to missing or unverifiable information.

Table 19: Reasons for Email Operability Error by Region – 2009 RAA Requirements

Error	Africa	Asia Pacific	Europe	Latin America and Caribbean	North America	Unknown	All Regions
Not Verifiable (or Missing)	0.0%	4.0%	0.2%	0.3%	0.4%	98.6%	3.2%
Email Bounced	100.0%	96.0%	99.8%	99.7%	99.6%	1.4%	96.8%
Overall Accuracy for Region – Email Operability	92.9% ± 1.4%	92.7% ± 0.9%	89.3% ± 1.2%	86.5% ± 1.5%	93.0% ± 0.9%	N/A	**91.4% ± 0.5%**

See note in Table 18 for how to read this table.

Reasons for Telephone Syntax and Operability Error by Region – 2009 RAA

Once again, Table 20 shows that the when the region is unknown, telephone syntax errors were due to information that was missing or not allowed. Among the regions, North America had the largest percentage of missing country codes, while the Asia-Pacific region had the largest percentage of missing or telephone numbers that are not allowed.

Table 20: Reasons for Telephone Syntax Error by Region – 2009 RAA Requirements

Error	Africa	Asia Pacific	Europe	Latin America and Caribbean	North America	Unknown	All Regions
Incorrect Length	74.4%	58.1%	76.8%	63.6%	44.4%	0.0%	62.0%
Country Code Missing	25.2%	32.7%	21.0%	35.5%	54.4%	0.0%	33.6%
Missing or Not Allowed	0.4%	9.2%	2.1%	0.9%	1.2%	100.0%	4.4%
Overall Accuracy for Region – Telephone Syntax	64.6% ± 2.6%	88.9% ± 1.1%	85.1% ± 1.4%	84.3% ± 1.6%	85.1% ± 1.2%	N/A	**85.3% ± 0.6%**

See note in Table 18 for how to read this table.

Table 21 shows that the regions did not differ much in their distributions of telephone operability. Between 0.8 and 3.8 percent of the inoperable telephone numbers were missing or unverifiable for each region, between 12.8 and 28.3 percent were disconnected, between 22.9 and 34.1 percent were invalid numbers and the remaining 33.9 to 56.7 percent of the inoperable telephone numbers failed to connect for another reason.

Table 21: Reasons for Telephone Operability Error by Region – 2009 RAA Requirements

Error	Africa	Asia Pacific	Europe	Latin America and Caribbean	North America	Unknown	All Regions
Not Verifiable (or Missing)	0.8%	3.8%	1.1%	1.0%	1.3%	100.0%	3.1%
Number Disconnected	12.8%	28.3%	21.8%	15.6%	19.2%	0.0%	21.7%
Invalid Number	32.2%	34.1%	26.5%	31.9%	22.9%	0.0%	29.7%
Other Not Connected	54.3%	33.9%	50.7%	51.5%	56.7%	0.0%	45.4%
Overall Accuracy for Region – Telephone Operability	70.9% ± 2.5%	63.7% ± 1.8%	69.2% ± 1.8%	84.5% ± 1.6%	85.3% ± 1.2%	N/A	**76.0% ± 0.8%**

See note in Table 18 for how to read this table.

Reasons for Postal Address Error by Region – 2009 RAA

From Table 22, you can see that across all regions the majority of postal address syntax errors were due to missing address components such as city or state/province. Once again, when the region was unknown, the error was due to a lack of any information having been provided ("missing").

Table 22: Reasons for Postal Address Syntax Error by Region – 2009 RAA Requirements

Error	Africa	Asia Pacific	Europe	Latin America and Caribbean	North America	Unknown	All Regions
Missing	0.1%	0.4%	0.0%	0.0%	1.0%	100.0%	1.0%
Country Code Missing	0.4%	2.3%	8.1%	0.3%	3.4%	0.0%	2.6%
Street Missing	22.6%	14.8%	11.2%	12.1%	13.1%	0.0%	15.1%
Postal Code Missing or Bad Format	23.8%	15.3%	13.7%	39.5%	19.2%	0.0%	20.4%
City Missing	29.3%	26.6%	33.0%	22.7%	39.1%	0.0%	27.7%
State/Province Missing	23.8%	40.7%	34.0%	25.3%	24.1%	0.0%	33.1%
Overall Accuracy for Region – Postal Syntax	44.6% ± 2.7%	49.8% ± 1.8%	68.4% ± 1.8%	71.0% ± 2.0%	96.7% ± 0.6%	N/A	**77.3% ± 0.7%**

See note in Table 18 for how to read this table.

Table 23 shows that almost all of the postal address operability errors are coded as P1 (probably not deliverable) in the Africa region and the Latin America and Caribbean region. Excluding the "unknown" region cases, the Asia-Pacific and Europe regions have the highest percentages of N1 "country unknown" errors.

Table 23: Reasons for Postal Address Operability Error by Region – 2009 RAA Requirements

Error	Africa	Asia Pacific	Europe	Latin America and Caribbean	North America	Unknown	All Regions
P1 Inoperable[37]	81.5%	29.9%	51.9%	79.4%	32.7%	0.0%	40.9%
P2 Inoperable	12.1%	28.5%	14.8%	14.3%	44.9%	0.0%	20.5%
N1 Country Unknown	6.4%	40.9%	33.3%	6.3%	22.4%	100.0%	38.2%
N2 Unverifiable	0.0%	0.7%	0.0%	0.0%	0.0%	0.0%	0.3%
Overall Accuracy for Region – Postal Operability	95.8% ± 1.1%	95.9% ± 0.7%	99.1% ± 0.4%	99.0% ± 0.4%	99.4% ± 0.3%	N/A	**97.7% ± 0.3%**

See note in Table 18 for how to read this table.

[37] *See* page 31 for an explanation of the reasons for Error for Postal Address Operability errors in postal address operability.

Script Analysis

We report here the syntax and operability accuracy of records by contact type and by the script language that was used to register the domain. The majority of domains across all regions are registered using Latin script (see Table 24), but some are registered using scripts such as Hanzi, Arabic or Diacritic marking. Records with Hanzi script were typically registered in the Asia-Pacific region, while records with Diacritic marking were typically registered in either the Europe, Latin America/Caribbean or Asia-Pacific regions. Tables containing region-specific script analysis can be found in Appendix D. As this analysis is new to the report, we are not yet able to provide trends across cycles.

Table 24: Overall Syntax and Operability Accuracy by Contact Type and Script Language

	Script Language	Count	Syntax Percent Accurate	Operability Percent Accurate
Registrant	**Latin**	11,957	65.8% ± 0.9%	75.0% ± 0.8%
	Arabic	2	0.0% ± 0.0%	97.1% ± 23.3%
	Chinese (Hanzi)	66	10.7% ± 7.5%	41.6% ± 11.9%
	Diacritical Marking	174	57.5% ± 7.3%	66.2% ± 7.0%
Administrative	**Latin**	11,948	65.7% ± 0.9%	75.2% ± 0.8%
	Arabic	2	0.0% ± 0.0%	97.1% ± 23.3%
	Chinese (Hanzi)	65	11.1% ± 7.6%	39.7% ± 11.9%
	Diacritical Marking	174	58.3% ± 7.3%	63.9% ± 7.1%
Technical	**Latin**	11,943	67.0% ± 0.8%	74.1% ± 0.8%
	Arabic	2	0.0% ± 0.0%	97.1% ± 23.3%
	Chinese (Hanzi)	65	6.3% ± 5.9%	34.9% ± 11.6%
	Diacritical Marking	141	70.4% ± 7.5%	69.5% ± 7.6%

As we saw in Main Findings, accuracy rates are similar across the three contact types. Operability accuracy rates for Latin script are similar to rates of overall operability accuracy in every region, which should be expected given the high prevalence of records registered using Latin. By contrast, syntax accuracy for Latin script is higher than overall rates of syntax accuracy for all three contact types. When viewed at a regional level, syntax accuracy for Latin script is similar to overall accuracy in every region except Asia-Pacific (see Appendix D). Since there are so few records registered in other script languages, it is important not to focus on estimates of accuracy for less prevalent script types, or on the differences between any script types (i.e., difference between accuracy of Latin and Arabic).

Comparisons Between Cycles

Statistical comparisons of syntax and operability accuracy can be made between Cycle 1 and Cycle 2 findings. We present the comparisons below for informational purposes, and to explore what general observations can be made about the relationship between syntax and operability accuracy.

Comparisons of Accuracy Between Cycles – 2009 RAA Syntax Requirements

In Main Findings, we presented syntax accuracy of records against 2009 RAA Requirements for Cycle 2. Here, we compare the Cycle 2 syntax accuracy results to the results from Cycle 1.

Change in Overall Accuracy

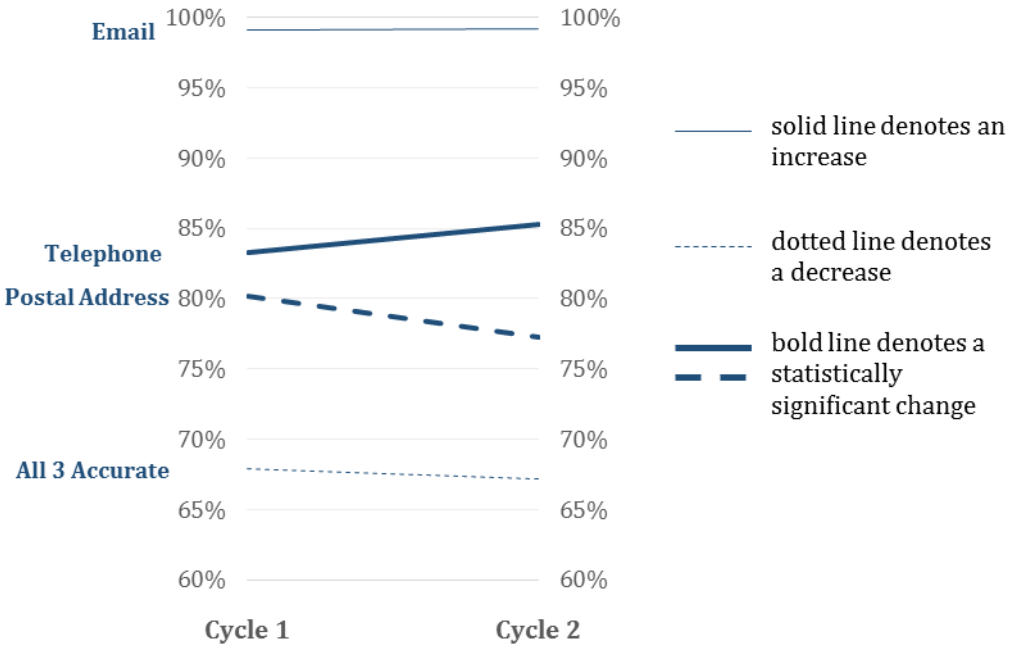
Table 25 and Graph 14 show that email accuracy rates are very similar across phases, but that telephone accuracy was lower in Cycle 1, and postal address accuracy is lower in the Cycle 2. The most likely explanation for the differences in telephone and postal addresses is random variation.[38] The rate at which all modes were accurate has decreased between each phase.

Table 25: Overall Accuracy by Cycle – 2009 RAA Syntax Requirements

	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	99.1% ± 0.2%	83.3% ± 0.7%	80.2% ± 0.8%	**67.9% ± 0.9%**
Cycle 2	99.2% ± 0.2%	85.3% ± 0.6%	77.3% ± 0.7%	**67.2% ± 0.8%**
Change (C2–C1)	**0.0% ± 0.2%**	**2.0% ± 1.0%**	**-3.0% ± 1.1%**	**-0.7% ± 1.2%**

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

Graph 14: Overall Accuracy by Cycle – 2009 RAA Syntax Requirements



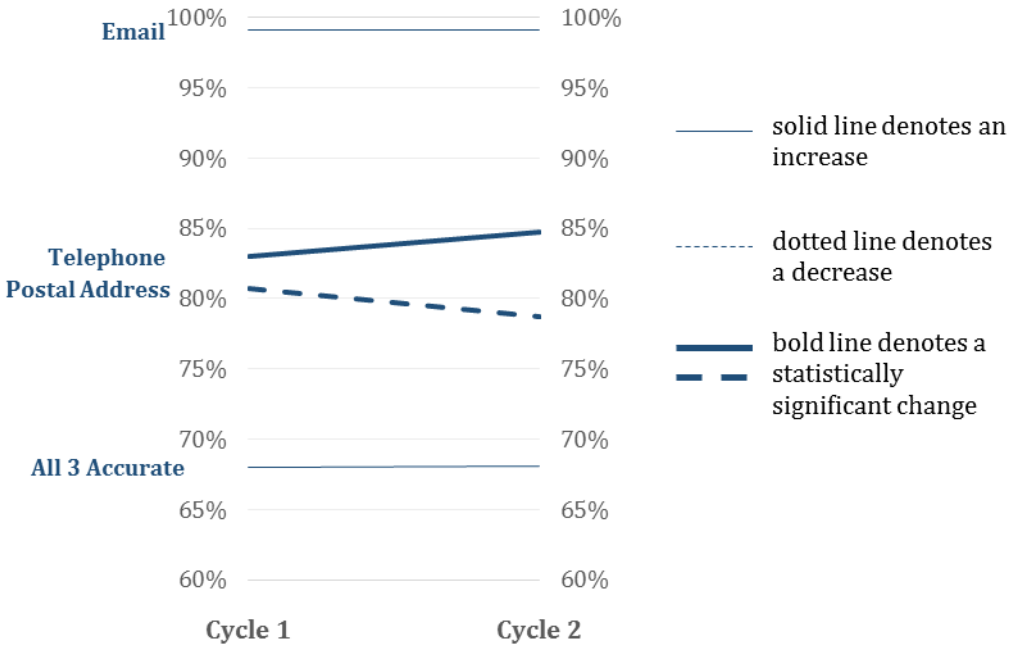
Change in Prior gTLDs

Since most of the domains in the domain universe are from Prior gTLDs, the patterns for the Prior gTLDs seen in Table 26 and Graph 15 are similar to the pattern for overall accuracy rates that appear above in Table 25. That is, the data for Prior gTLDs shows the same decrease in overall accuracy of Cycle 1 telephone numbers and Cycle 2 postal addresses. Similarly, there was no change in email accuracy.

Table 26: Prior gTLDs Accuracy by Cycle – 2009 RAA Syntax Requirements

	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	99.1% ± 0.2%	83.0% ± 0.8%	80.7% ± 0.9%	68.0% ± 1.0%
Cycle 2	99.1% ± 0.2%	84.7% ± 0.7%	78.7% ± 0.8%	68.1% ± 0.9%
Change (C2–C1)	**0.0% ± 0.3%**	**1.6% ± 1.1%**	**-2.0% ± 1.2%**	**0.0% ± 1.4%**

Graph 15: Prior gTLDs Accuracy by Cycle – 2009 RAA Requirements



Change in New gTLDs

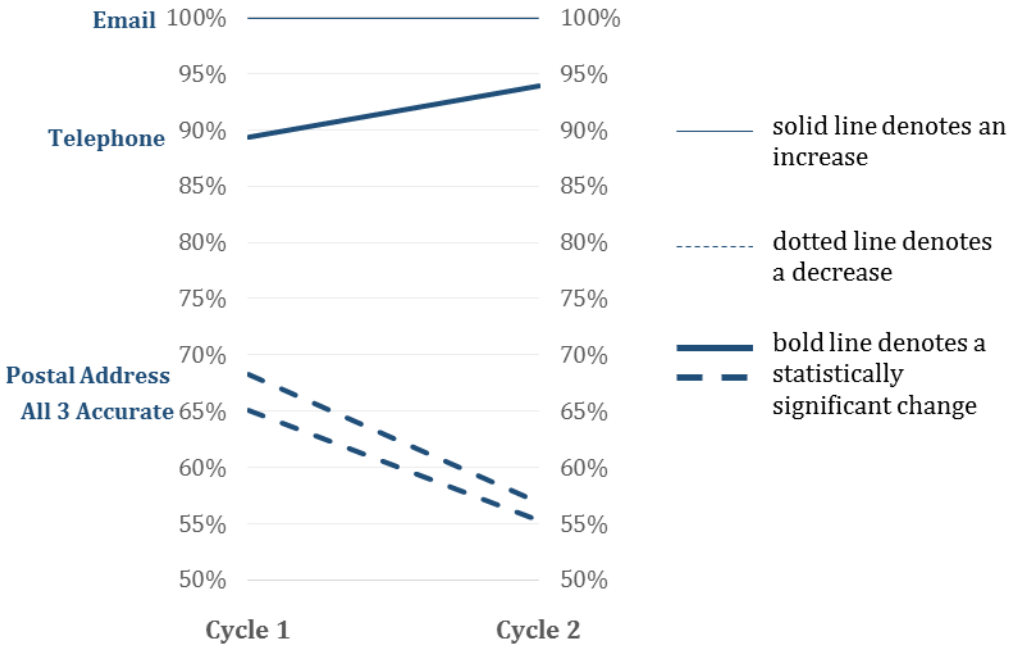
In Table 27 and Graph 16, the New gTLDs show the same pattern as Prior gTLDs. Again, there is no change for email addresses, while the rates for telephone accuracy increased across cycles, and the rates for postal address accuracy decreased across cycles. The percentage of domains in New gTLDs that pass all accuracy tests for all nine contacts also decreased across cycles.[39]

Table 27: New gTLDs Accuracy by Cycle – 2009 RAA Syntax Requirements

	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	99.9% ± 0.1%	89.4% ± 1.4%	68.3% ± 2.1%	65.1% ± 2.1%
Cycle 2	99.9% ± 0.1%	93.9% ± 1.0%	56.9% ± 2.0%	55.3% ± 2.0%
Change (C2–C1)	**0.0% ± 0.2%**	**4.6% ± 1.7%**	**-11.4% ± 2.9%**	**-9.8% ± 3.0%**

[39] *See* Appendix B or the Cycle 1 report for more information on results, especially by region.

Graph 16: New gTLDs Accuracy by Cycle – 2009 RAA Syntax Requirements



Comparisons of Accuracy Between Cycles – 2009 RAA Operability Requirements

Change in Overall Accuracy

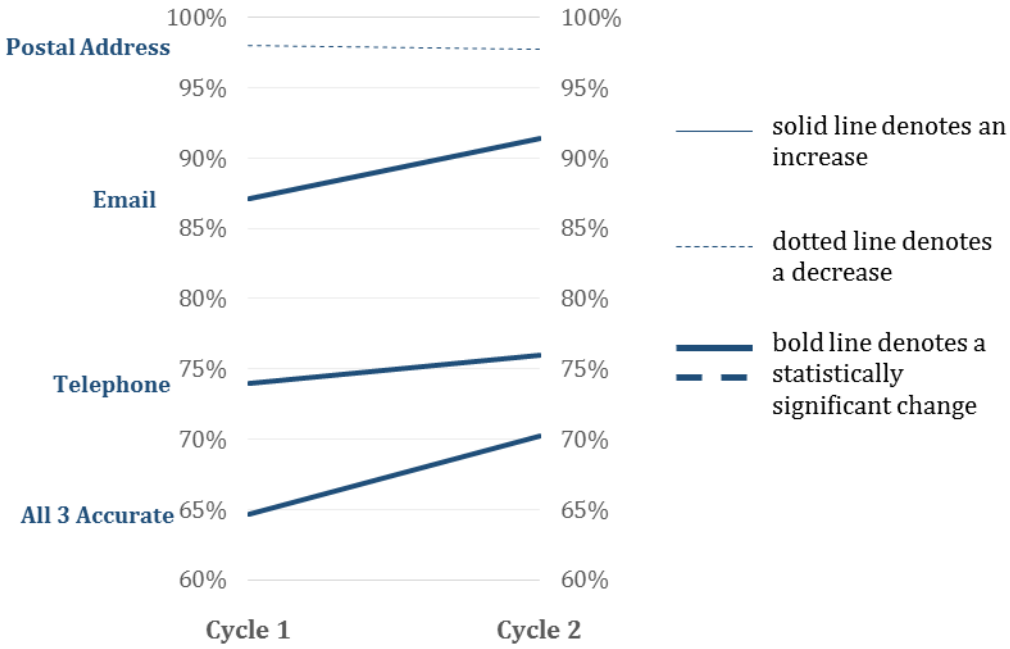
Table 28 and Graph 17 show that Cycle 2 has higher overall email and telephone accuracy rates, but that postal address accuracy rates are very similar between the cycles. The most likely explanation for the difference between Cycle 1 and Cycle 2 is random variation.[40] It is also plausible that the distribution has changed, but not much time has passed between Cycle 1 and Cycle 2 for such a change to take place. A change in distribution would imply that the newest domains not subjected to Cycle 1 sampling have higher accuracy in terms of email addresses and telephone numbers. The rate of records with all modes accurate increased in Cycle 2.

Table 28: Overall Accuracy by Cycle – 2009 RAA Operability Requirements

	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	87.1% ± 0.7%	74.0% ± 0.9%	98.0% ± 0.3%	64.7% ± 0.9%
Cycle 2	91.4% ± 0.5%	76.0% ± 0.8%	97.7% ± 0.3%	70.2% ± 0.8%
Change (C2–C1)	**4.3% ± 0.8%**	**2.1% ± 1.2%**	**-0.4% ± 0.4%**	**5.4% ± 1.2%**

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

Graph 17: Overall Accuracy by Cycle – 2009 RAA Operability Requirements

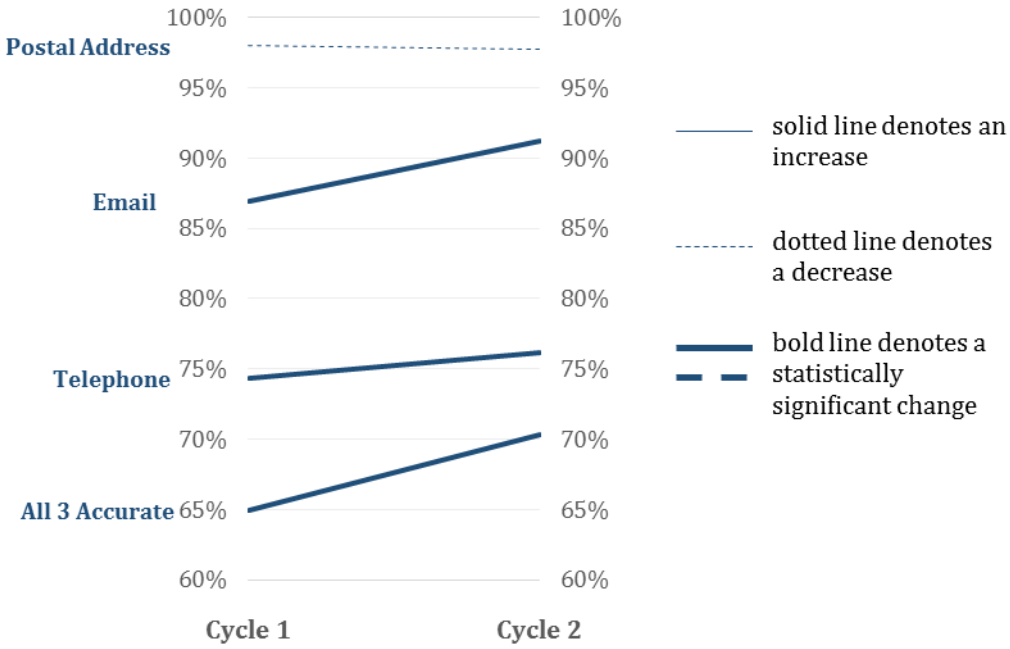


The same increases in overall accuracy that were observed for email, telephone and all-mode accuracy can be seen in the data below for Prior gTLDs. Similarly, postal address accuracy showed no change between cycles.

Table 29: Prior gTLDs Accuracy by Cycle – 2009 RAA Operability Requirements

	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	86.9% ± 0.7%	74.3% ± 1.0%	98.0% ± 0.3%	64.9% ± 1.0%
Cycle 2	91.2% ± 0.6%	76.2% ± 0.8%	97.7% ± 0.3%	70.3% ± 0.9%
Change (C2–C1)	**4.4% ± 0.9%**	**1.9% ± 1.3%**	**-0.3% ± 0.4%**	**5.4% ± 1.4%**

Graph 18: Prior gTLDs Accuracy by Cycle – 2009 RAA Operability Requirements



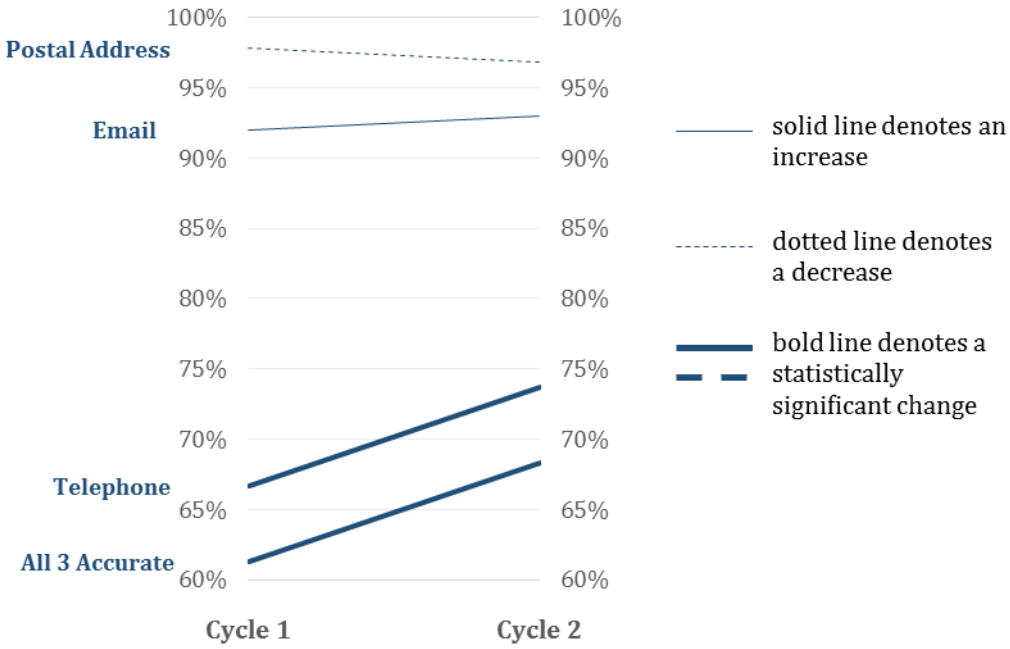
In Table 30 and Graph 19, the New gTLDs show similar patterns. Again, email, telephone and all-mode accuracy show increases in Cycle 2, while there is little or no change for postal address accuracy. Among New gTLDs, however, the increase in email operability is not significant.[41]

Table 30: New gTLDs Accuracy by Cycle – 2009 RAA Operability Requirements

	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	92.0% ± 1.2%	66.7% ± 2.1%	97.8% ± 0.7%	61.3% ± 2.2%
Cycle 2	93.0% ± 1.0%	73.7% ± 1.8%	96.8% ± 0.7%	68.3% ± 1.9%
Change (C2–C1)	**0.9% ± 1.6%**	**7.0% ± 2.8%**	**-1.0% ± 1.0%**	**6.9% ± 2.9%**

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

Graph 19: New gTLDs Accuracy by Cycle – 2009 RAA Operability Requirements



Relationship Between Syntax and Operability Accuracy

In Main Findings, we presented syntax and operability accuracy for all 12,000 domains based on the 2009 RAA standards. Tables 31 through 33 show the relationship between syntax and operability accuracy against 2009 RAA standards. They examine the rate at which records that pass one of the two testing methods would also pass the other testing method (e.g., what percentage of records that pass operability testing also pass syntax testing, and vice versa).

Email Address Syntax and Operability Accuracy

Table 31 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). This pattern is logically consistent because certain syntax failures – for example, an email address missing the "@" symbol (or any another critical component) – indicate that the email address is not operable. The opposite is true for email addresses that fail operability accuracy tests (8.6 percent of all domains); most of these email addresses actually pass the syntax accuracy tests (7.8 percent of the 8.6 percent). This pattern is also logically consistent because certain operability failures – for example, email bounce-backs resulting from an email address that is no longer in use – will occur even when the syntax is accurate. Another way to describe the figures in Table 31, is to say that 7.8 percent of email addresses pass syntax accuracy tests, but are found to be inoperable.

Table 31: Syntax and Operability Accuracy for Email Addresses – 2009 RAA Requirements

		Operability		
		Pass	Fail	**TOTAL**
Syntax	Pass	91.4% ± 0.5%	7.8% ± 0.5%	99.2% ± 0.2%
	Fail	0.0% ± 0.0%	0.8% ± 0.2%	0.8% ± 0.2%
	TOTAL	**91.4% ± 0.5%**	**8.6% ± 0.5%**	100%

Telephone Number Syntax and Operability Accuracy

Table 32 shows that 14.7 percent of telephone numbers fail the syntax accuracy tests, while 24.0 percent fail the operability accuracy tests. However, these groups do not fully overlap. Unlike for email, failing syntax is not an indicator that the telephone number will fail operability – there are some telephone numbers that can fail syntax testing, but pass operability testing.[42] Of the telephone numbers that fail the syntax accuracy tests (14.7 percent total), more than half also fail the operability test (9.0 percent out of the 14.7 percent, or 61.2 percent). Similarly, of those that fail the operability test (24.0 percent total), over half (15.0 percent out of the 24.0 percent, or 62.5 percent) pass the syntax accuracy tests.

[42] For example, a telephone number may be missing a country code, but if the country is readily available in the WHOIS record, the telephone number is operable.

Table 32: Syntax and Operability Accuracy for Telephone Numbers – 2009 RAA Requirements

		Operability		
		Pass	Fail	**TOTAL**
Syntax	Pass	70.3% ± 0.8%	15.0% ± 0.6%	**85.3% ± 0.6%**
	Fail	5.8% ± 0.4%	9.0% ± 0.5%	**14.7% ± 0.6%**
	TOTAL	**76.0% ± 0.8%**	**24.0% ± 0.8%**	100%

Postal Address Syntax and Operability Accuracy
Table 33 shows that postal addresses that fail 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 (22.7 percent of all domains) pass the operability accuracy tests (20.4 percent of the 22.7 percent).[43] Another way to look at Table 33 is that 20.4 percent of postal addresses fail the syntax accuracy tests, but pass the operability accuracy tests.

Table 33: Syntax and Operability Accuracy for Postal Addresses – 2009 RAA Requirements

		Operability		
		Pass	Fail	**TOTAL**
Syntax	Pass	77.3% ± 0.8%	0.0% ± 0.0%	**77.3% ± 0.7%**
	Fail	20.4% ± 0.7%	2.3% ± 0.3%	**22.7% ± 0.7%**
	TOTAL	**97.7% ± 0.3%**	**2.3% ± 0.3%**	100%

Takeaways Regarding Syntax and Operability Accuracy
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 20.8 percent (15.0 percent plus 5.8 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 "@" symbol) 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.

[43] *See* note 34 and discussion in Challenges and Lessons Learned.

Challenges and Lessons Learned

We have continually sought ways to improve the ARS and are looking ahead to subsequent ARS reports. This section provides background on the issues that created challenges in Cycle 2 – and how those issues can be avoided or improved upon in subsequent reports.

First, to follow up on issues raised in the last report:

- **Increasing the sample size continues to improve statistical significance.**

  As discussed in the Cycle 1 report, increasing the initial sample and subsample sizes helps with the smallest subgroups (e.g., increases the chance that a domain from Africa will be sampled). For Cycle 2, we increased the sample size to 200,000 and the subsample size to 12,000. As discussed in the Main Findings section, this change allowed for stronger estimates regarding the smaller subgroups.[44] As ICANN has determined with the WHOIS ARS vendors that marginal increases in the sample and subsample sizes has little effect on resource requirements, we are again considering what kind of additional increase in sample sizes is both feasible and beneficial to the ARS study.

- **Analyzing all three contact types is useful for a full picture of the data.**

  From examining the commonality across contact types, we found in Cycle 1 and Cycle 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 to reduce inefficiencies and redundancies, but we did consider whether further efficiency could be achieved by analyzing only one contact type, e.g., registrant. We determined, however, that it was more beneficial to have complete consistency across the reports to allow for comparisons. Similarly, by testing only one of the contact types, we would be unable to speak as accurately about all of the WHOIS ARS data, as some inaccuracies would be missed.

Second, we would like to address the issue of the relationship between the contact modes (email, telephone, postal) as it relates to *contactability* (i.e., the likelihood that a contact type can be reached via a specific contact mode).

- **Postal address operability is high, but is perhaps not the best indicator of contactability.**

  As discussed in previous reports, syntax and operability testing of postal addresses pose unique challenges (e.g., syntactically inaccurate postal addresses may look operationally inaccurate, but in reality it is likely postal mail can be delivered to that address). Working with the Universal Postal Union we devised operability tests that used both automated and manual checks to ensure that all potentially deliverable addresses were marked as such. This testing method resulted in very high operability accuracy rates, both in Cycle 1 and Cycle 2. While the vendor's tests seek to mimic actual delivery, they cannot be 100 percent accurate. This doubt does not exist with email and telephone testing: we typically do not put something in the postal mail to check deliverability, but we do place phone calls and send emails to see if they work.[45] Because of this fact, we caution

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

[45] This does not speak to identity validation, however, which is not the subject of this report, nor was it conducted by the WHOIS ARS vendors. As previously stated, the plan for identity validation has not yet been determined.

against relying too heavily on the high accuracy rates of postal address operability to reflect contactability; the real numbers may be somewhat lower. In addition, we feel that it is unlikely that a person attempting to reach out to a domain registrant, administrative or technical contact will first try postal delivery. Rather, it is more likely that the individual will first choose to send email or place a telephone call. For this reason, we feel that a better indicator of contactability of a record is demonstrated by the operability accuracy rates of either email or telephone. As discussed above, both are still very high, around 94 percent.

Finally, we would like to discuss one last issue as it relates to the accuracy testing of postal addresses.

- **Postal address testing rules continue to be improved.**

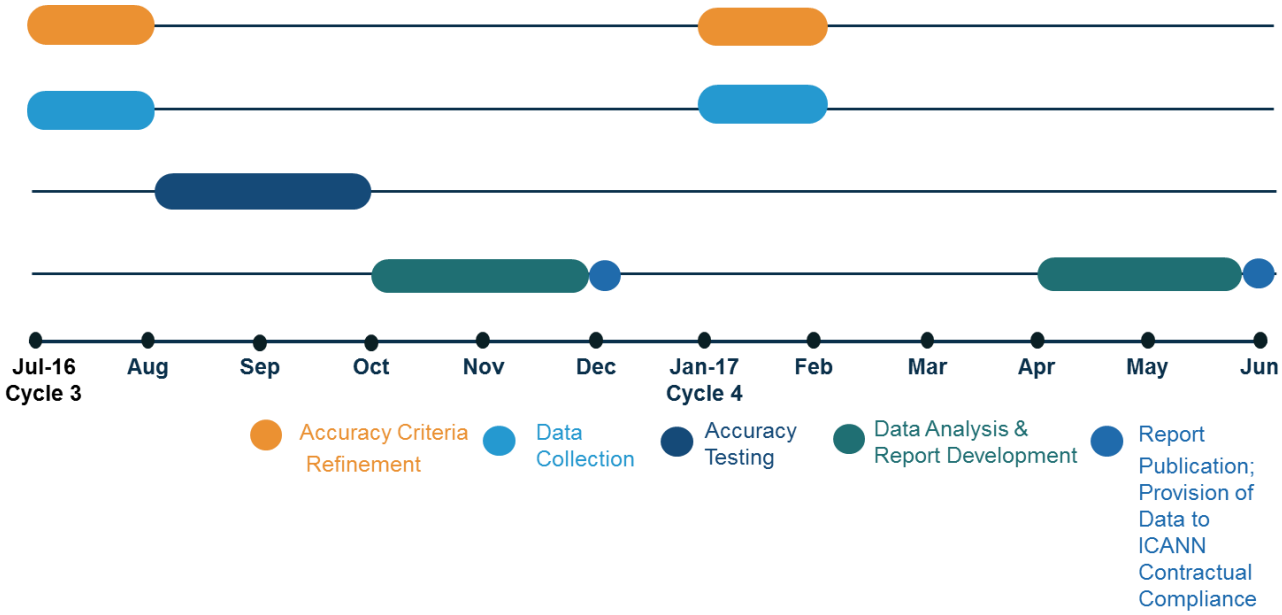
  As noted above and in previous reports, postal address testing poses challenges for numerous reasons. One additional reason is that the rules for syntax accuracy in a country (i.e., country formatting requirements) can have exceptions. Throughout the life of the ARS project we have received feedback from registrars that certain items that have been marked as a formatting error or inaccuracy are actually allowed within a country. When we receive such feedback, we report this information to the Universal Postal Union, which then notes it for future testing. Our intent is to be flexible where the country is also flexible with its postal addressing rules. We welcome this kind of feedback in the future as well.

Next Steps

Looking Ahead to Phase 2 Cycle 3

Phase 2 Cycles 3 and 4 will continue the syntax and operability testing of WHOIS records. Cycle 3 is scheduled to begin in July 2016. The timeline in Figure 6 shows the stages for Phase 2 Cycles 3 and 4 up through report publication.

Figure 6: Phase 2 Cycles 3 and 4



Next Steps for ICANN Contractual Compliance

As indicated above, one of the major goals of the ARS project is the ability to pass to ICANN Contractual Compliance any potential inaccuracies that the registrars can investigate and follow up on. The results of the Phase 2 Cycle 2 study is that those potentially inaccurate records have already been given 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 2 includes both syntax and operability results, Contractual Compliance review and follow-up may be conducted through different processes, depending on the type of inaccuracies found within the record. For example, those records with only formatting errors but that have been deemed "operable" will receive a different kind of follow-up than those records that have been deemed "inoperable."

All WHOIS ARS tickets will follow the Contractual Compliance Approach and Process[46] 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 may be classified as WHOIS format errors if the error indicates 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. These requirements include validating format requirements and suspending a domain name for failure of the registrant to respond in a timely manner 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:
  - Registration Data Directory Service (WHOIS) Specification"[47] and
  - Advisory: Clarifications to the Registry Agreement, and the 2013 Registrar Accreditation Agreement (RAA) regarding applicable Registration Data Directory Service (WHOIS) Specifications.[48]

In addition, as with any WHOIS inaccuracy or WHOIS format complaints, 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).

Phase 2 Results

Compliance continues to present metrics for WHOIS ARS in the Compliance Quarterly Reports (see https://www.icann.org/resources/pages/compliance-reports-2016-04-15-en), and will provide additional

[46] *See* ICANN Contractual Compliance Approach and Process: https://www.icann.org/resources/pages/approach-processes-2012-02-25-en.
[47] *See* https://www.icann.org/resources/pages/approved-with-specs-2013-09-17-en#whois.
[48] *See* https://www.icann.org/resources/pages/registry-agreement-raa-rdds-2015-04-27-en.

information when metrics are generated for the second quarter of 2016. Additionally, metrics will be provided at ICANN Public Meetings, where appropriate.

Appendix A: Accuracy Testing Criteria

ICANN has attempted to align the accuracy testing criteria with the contractual obligations of the Registrar Accreditation Agreements (RAA) and applicable Internet Engineering Task Force Requests for Comments. Currently, there are two predominant versions of the RAA in use in the gTLD space, the 2009 version and the 2013 version. Each version of the RAA has requirements for presence, format and operability of specific elements of contact information for the registrant, the technical contact and the administrative contact 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 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

You can find an overview of criteria for syntax and operability accuracy testing for email addresses, telephone numbers and postal addresses at https://whois.icann.org/en/whoisars-validation. The criteria listed there were used by the validation vendors supporting the WHOIS ARS project.

Appendix B: Additional Analyses - Accuracy to 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 contact that match, and least likely to be the registrant and technical contact that match.

Table B1: Frequency of Common Contact Information Across Contact Type and Mode

Commonality	Email	Telephone	Postal Address
All Three Exactly the Same	77.6% ± 0.7%	80.3% ± 0.7%	78.2% ± 0.7%
Registrant=Administrative	14.0% ± 0.6%	14.0% ± 0.6%	13.2% ± 0.6%
Registrant=Technical	0.4% ± 0.1%	0.3% ± 0.1%	0.4% ± 0.1%
Administrative=Technical	5.7% ± 0.4%	4.3% ± 0.4%	6.2% ± 0.4%
All Three Different	2.3% ± 0.3%	1.0% ± 0.2%	2.0% ± 0.3%

2009 RAA Reasons for Syntax Error in Cycle 1 and Cycle 2

The Main Findings section contains the Cycle 2 results, but we present sequentially the results from all three prior ARS studies (Phase 1, Cycle 1 and Cycle 2).

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

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	10,000	9,950	9,954	29,904
Missing*	[38]*	50	46	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) – Cycle 1

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

* Registrant email is not required under the 2009 RAA.

Table B4: Total Email Address Errors by Contact Type (2009 RAA) – Cycle 2

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	11,994	11,947	11,945	35,886
Missing*	128*	48	51	99
@ Symbol Missing	2	1	0	3
Not Resolvable	4	4	4	12
Total	12,000	12,000	12,000	36,000

* Registrant email is not required under the 2009 RAA.

Table B5: Total Telephone Number Errors by Contact Type (2009 RAA) – Phase 1

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	8,780	8,645	8,719	26,144
Not Present*	[234]*	144	148	292
Country Code Missing	304	289	279	872
Incorrect Length	883	889	821	2,593
Characters Not Allowed	33	33	33	97
Total	10,000	10,000	10,000	30,000

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

Table B6: Total Telephone Number Errors by Contact Type (2009 RAA) – Cycle 1

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	8,569	8,409	8,523	25,501
Missing*	[199]*	137	144	281
Country Code Missing	474	499	481	1,454
Incorrect Length	955	952	849	2,756
Characters Not Allowed	2	3	3	8
Total	10,000	10,000	10,000	30,000

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

Table B7: Total Telephone Number Errors by Contact Type (2009 RAA) – Cycle 2

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	10,398	10,224	10,316	30,938
Missing*	182*	107	113	220
Country Code Missing	538	577	584	1,699
Incorrect Length	1,062	1,090	986	3,138
Characters Not Allowed	2	2	1	5
Total	12,000	12,000	12,000	36,000

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

Table B8: Total Postal Address Errors by Contact Type (2009 RAA) – Phase 1

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	7,582	7,570	7,826	22,978
Missing	42	50	56	148
Country Missing	18	22	22	62
Country Not Identifiable	24	26	27	77
Postal Code Missing	691	736	665	2,092
Postal Code format	25	24	20	69
State Missing	1,126	1,134	995	3,255
City Missing	836	858	777	2,471
Street Missing	564	557	494	1,615
Total	**10,000**	**10,000**	**10,000**	**30,000**
Total Errors	3,326	3,407	3,056	9,789
Total Domains with Errors	2,418	2,430	2,174	7,022

Table B9: Total Postal Address Errors by Contact Type (2009 RAA) – Cycle 1

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

Table B10: Total Postal Address Errors by Contact Type (2009 RAA) – Cycle 2

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	8,407	8,383	8,815	25,605
Missing	43	52	57	152
Country Code Missing	71	58	53	182
Country Not Identifiable	65	70	64	199
Postal Code Missing	953	1,039	920	2,912
Postal Code Format	23	21	20	64
State/Province Missing	1,676	1,699	1,463	4,838
City Missing	1,398	1,411	1,235	4,044
Street Missing	786	764	662	2,212
TOTAL	12,000	12,000	12,000	36,000
Total Errors	5,015	5,114	4,474	14,603
Total Domains with Errors	3,593	3,617	3,185	10,395

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

Table B11: African Domains Accuracy by Cycle – 2009 RAA Syntax Requirements

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	99.9% ± 0.2%	66.8% ± 2.9%	42.2% ± 3.1%	29.8% ± 2.9%
Cycle 2	99.9% ± 0.2%	64.6% ± 2.6%	44.6% ± 2.7%	29.3% ± 2.5%
Change (C2–C1)	0.0% ± 0.3%	-2.1% ± 4.0%	2.4% ± 4.1%	-0.5% ± 3.8%

Table B12: Asia-Pacific Domains Accuracy by Cycle – 2009 RAA Syntax Requirements

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	99.5% ± 0.3%	78.7% ± 1.7%	52.5% ± 2.0%	39.5% ± 2.0%
Cycle 2	99.4% ± 0.3%	88.9% ± 1.1%	49.8% ± 1.8%	45.0% ± 1.8%
Change (C2–C1)	-0.1% ± 0.4%	10.2% ± 2.0%	-2.7% ± 2.7%	5.5% ± 2.7%

Table B13: European Domains Accuracy by Cycle – 2009 RAA Syntax Requirements

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	99.8% ± 0.2%	85.2% ± 1.5%	72.2% ± 1.9%	62.3% ± 2.1%
Cycle 2	99.9% ± 0.1%	85.1% ± 1.4%	68.4% ± 1.8%	60.6% ± 1.9%
Change (C2–C1)	0.1% ± 0.2%	-0.0% ± 2.0%	-3.8% ± 2.6%	-1.7% ± 2.8%

Table B14: Latin/Caribbean Domains Accuracy by Cycle – 2009 RAA Syntax Requirements

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	99.9% ± 0.1%	79.2% ± 1.9%	67.1% ± 2.1%	56.9% ± 2.3%
Cycle 2	99.9% ± 0.1%	84.3% ± 1.6%	71.0% ± 2.0%	64.7% ± 2.1%
Change (C2–C1)	0.0% ± 0.2%	5.1% ± 2.5%	3.8% ± 3.0%	7.8% ± 3.1%

Table B15: North American Domains Accuracy by Cycle – 2009 RAA Syntax Requirements

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	99.9% ± 0.1%	86.1% ± 1.3%	97.1% ± 0.6%	83.9% ± 1.4%
Cycle 2	100.0% ± 0.0%	85.1% ± 1.2%	96.7% ± 0.6%	82.8% ± 1.3%
Change (C2–C1)	0.1% ± 0.1%	-1.1% ± 1.8%	-0.4% ± 0.9%	-1.1% ± 1.9%

Table B16: 2009 RAA Domains Accuracy by Cycle – 2009 RAA Syntax Requirements

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	98.1% ± 0.6%	89.1% ± 1.3%	82.0% ± 1.6%	77.1% ± 1.7%
Cycle 2	99.3% ± 0.3%	90.8% ± 1.2%	85.2% ± 1.5%	80.9% ± 1.6%
Change (C2–C1)	1.2% ± 0.7%	1.7% ± 1.7%	3.2% ± 2.2%	3.8% ± 2.4%

Table B17: 2013 RAA GF Domains Accuracy by Cycle – 2009 RAA Syntax Requirements

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	99.4% ± 0.2%	80.2% ± 1.2%	82.7% ± 1.2%	67.3% ± 1.5%
Cycle 2	99.4% ± 0.2%	80.0% ± 1.1%	81.9% ± 1.1%	66.8% ± 1.4%
Change (C2–C1)	-0.0% ± 0.3%	-0.2% ± 1.7%	-0.8% ± 1.6%	-0.6% ± 2.0%

Table B18: 2013 RAA NGF Domains Accuracy by Cycle – 2009 RAA Syntax Requirements

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	98.8% ± 0.3%	88.6% ± 1.0%	75.3% ± 1.4%	68.2% ± 1.5%
Cycle 2	98.9% ± 0.3%	91.1% ± 0.8%	71.2% ± 1.2%	66.9% ± 1.3%
Change (C2–C1)	0.1% ± 0.4%	2.4% ± 1.3%	-4.1% ± 1.9%	-1.3% ± 2.0%

Appendix C: Additional Analyses – Accuracy to 2013 RAA Requirements

Domains registered in the 2013 RAA now represent nearly 56 percent of all domains. In this appendix, we look at accuracy rates based on 2013 RAA requirements. As stated previously in this report, the 2009 RAA was chosen as a baseline against which all 12,000 of the analyzed subsample records were analyzed. The 2013 RAA requirements are stricter than the 2009 requirements, building from, 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 Universal Postal Union 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 against which all records can be compared.

Graph C1: Overall Accuracy – 2013 RAA Syntax Requirements

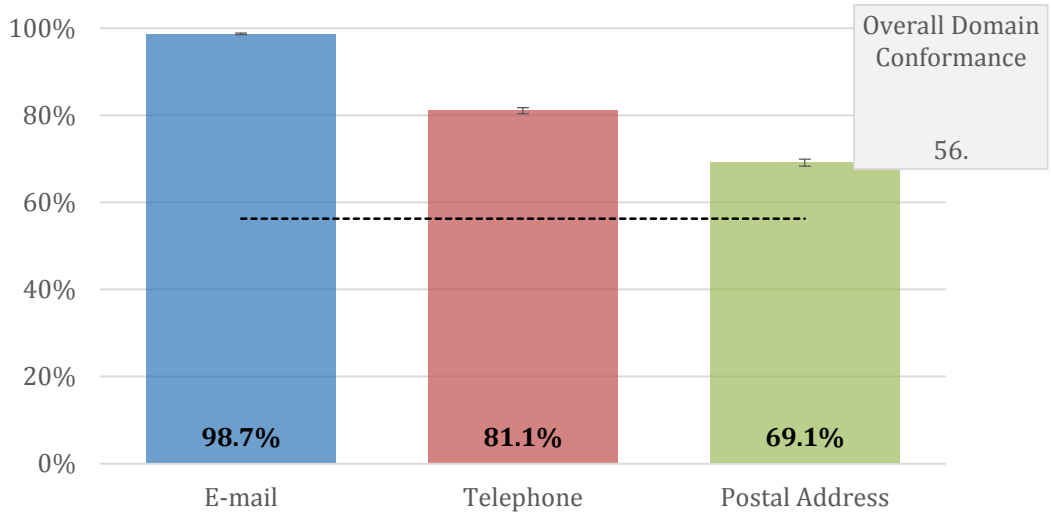


Table C1: Overall Accuracy by Contact Type and Mode – 2013 RAA Syntax Requirements

	Email	Telephone	Postal Address	All Three Accurate
Registrant	98.8% ± 0.2%	83.3% ± 0.7%	71.0% ± 0.8%	58.9% ± 0.9%
Administrative	99.2% ± 0.2%	82.9% ± 0.7%	70.8% ± 0.8%	59.0% ± 0.9%
Technical	99.2% ± 0.2%	82.7% ± 0.7%	71.1% ± 0.8%	59.3% ± 0.9%
Overall	**98.7% ± 0.2%**	**81.1% ± 0.7%**	**69.1% ± 0.8%**	**56.3% ± 0.9%**

Subgroup Accuracy – 2013 RAA Syntax Requirements

Next, we look at subgroups in Cycle 2, starting with Prior vs. New gTLDs. Since the numbers for registrant, administrative and technical contacts 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 that all passed the accuracy tests.

Subgroup 1: Prior vs. New gTLD

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

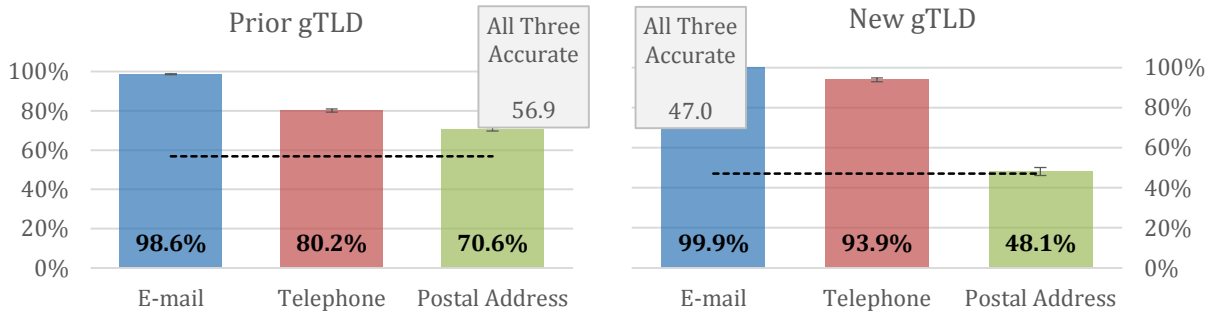


Table C2 shows that the New gTLDs have higher email and telephone syntax accuracy, but lower postal address syntax accuracy.

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

	Email	Telephone	Postal Address	All Three Accurate
Prior gTLD	98.6% ± 0.2%	80.2% ± 0.8%	70.6% ± 0.9%	**56.9% ± 1.0%**
New gTLD	99.9% ± 0.1%	93.9% ± 1.0%	48.1% ± 2.0%	**47.0% ± 2.0%**
Overall	**98.7% ± 0.2%**	**81.1% ± 0.7%**	**69.1% ± 0.8%**	**56.3% ± 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 that all passed the accuracy tests.

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

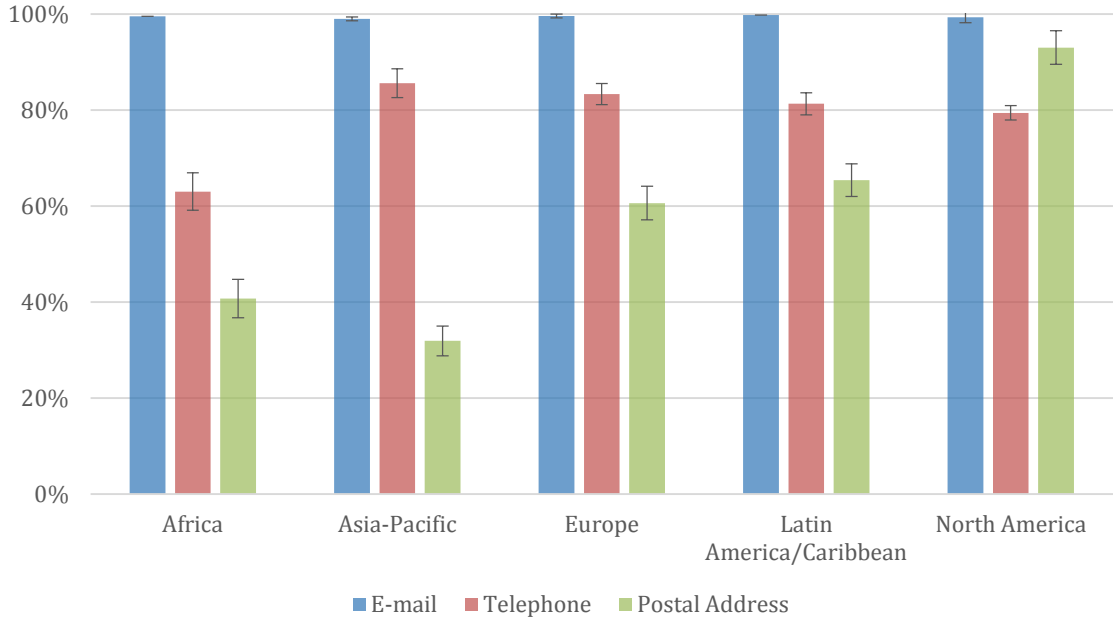


Table C3: Accuracy by ICANN Region – 2013 RAA Syntax Requirements

	Email	Telephone	Postal Address	All Three Accurate
Africa	99.5% ± 0.4%	63.0% ± 2.7%	40.7% ± 2.7%	**26.3% ± 2.4%**
Asia-Pacific	99.0% ± 0.4%	85.6% ± 1.3%	31.9% ± 1.7%	**27.3% ± 1.6%**
Europe	99.6% ± 0.2%	83.3% ± 1.4%	60.6% ± 1.9%	**53.1% ± 1.9%**
Latin America/Caribbean	99.8% ± 0.2%	81.3% ± 1.8%	65.4% ± 2.1%	**59.8% ± 2.2%**
North America	99.3% ± 0.3%	79.4% ± 1.4%	93.0% ± 0.9%	**73.2% ± 1.5%**
Overall	**98.7% ± 0.2%**	**81.1% ± 0.7%**	**69.1% ± 0.8%**	**56.3% ± 0.9%**

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

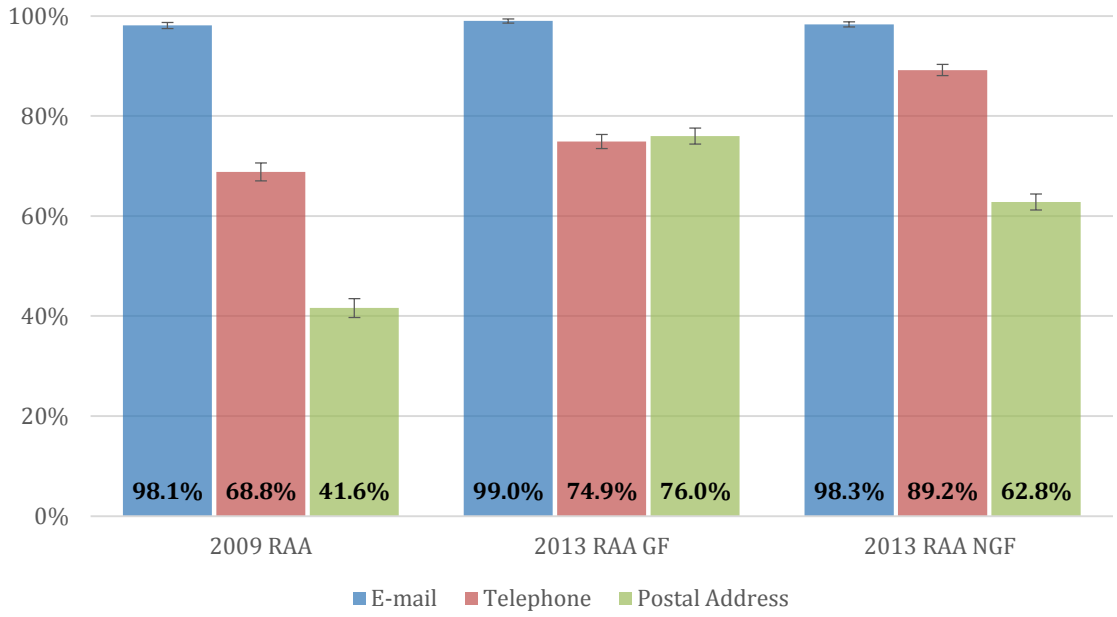


Table C4: Accuracy by RAA Status – 2013 RAA Syntax Requirements

	Email	Telephone	Postal Address	All Three Accurate
2009 RAA	98.1% ± 0.6%	68.8% ± 1.9%	41.6% ± 2.1%	20.2% ± 1.7%
2013 RAA GF	99.0% ± 0.3%	74.9% ± 1.2%	76.0% ± 1.2%	57.6% ± 1.4%
2013 RAA NGF	98.3% ± 0.4%	89.2% ± 0.9%	62.8% ± 1.3%	57.1% ± 1.4%
Overall	98.7% ± 0.2%	81.1% ± 0.7%	69.1% ± 0.8%	56.3% ± 0.9%

Comparisons Between Phases – 2013 RAA Syntax Requirements

Above, we presented the syntax accuracy to 2013 RAA requirements for Cycle 2. Here, we compare the Cycle 2 results to those from Cycle 1 for the 5,119 domains required to conform to these requirements.

Overall Accuracy

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

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	97.7% ± 0.5%	85.5% ± 1.1%	41.6% ± 1.6%	36.8% ± 1.5%
Cycle 2	98.3% ± 0.4%	89.2% ± 0.9%	62.8% ± 1.3%	57.1% ± 1.4%
Change (C2–C1)	0.5% ± 0.6%	3.7% ± 1.4%	21.2% ± 2.1%	20.3% ± 2.1%

Prior vs. New gTLDs

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

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	97.5% ± 0.7%	85.0% ± 1.6%	39.8% ± 2.2%	34.6% ± 2.1%
Cycle 2	98.0% ± 0.5%	88.4% ± 1.2%	65.3% ± 1.7%	58.8% ± 1.8%
Change (C2–C1)	0.5% ± 0.9%	3.4% ± 2.0%	25.5% ± 2.8%	24.1% ± 2.8%

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

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	99.6% ± 0.3%	89.4% ± 1.4%	56.0% ± 2.3%	53.8% ± 2.3%
Cycle 2	99.9% ± 0.1%	93.9% ± 1.0%	48.1% ± 2.1%	47.0% ± 2.1%
Change (C2–C1)	0.3% ± 0.3%	4.5% ± 1.7%	-7.9% ± 3.1%	-6.8% ± 3.1%

ICANN Regions

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

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	100.0% ± 0.0%	70.3% ± 3.9%	32.8% ± 4.0%	24.4% ± 3.7%
Cycle 2	99.8% ± 0.3%	69.2% ± 3.3%	36.3% ± 3.4%	24.8% ± 3.1%
Change (C2–C1)	-0.2% ± 0.3%	-1.1% ± 5.1%	3.4% ± 5.2%	0.4% ± 4.8%

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

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	99.7% ± 0.4%	74.3% ± 3.0%	26.9% ± 3.1%	21.8% ± 2.9%
Cycle 2	99.5% ± 0.4%	90.8% ± 1.5%	27.8% ± 2.3%	25.5% ± 2.2%
Change (C2 - C1)	-0.2% ± 0.5%	16.5% ± 3.4%	1.0% ± 3.8%	3.7% ± 3.6%

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

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	99.6% ± 0.4%	88.7% ± 2.2%	55.7% ± 3.4%	49.2% ± 3.5%
Cycle 2	100.0% ± 0.0%	90.2% ± 1.8%	59.3% ± 3.0%	56.0% ± 3.1%
Change (C2–C1)	0.4% ± 0.4%	1.5% ± 2.9%	3.5% ± 4.6%	6.8% ± 4.6%

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

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	100.0% ± 0.0%	87.2% ± 2.3%	59.7% ± 3.4%	55.9% ± 3.4%
Cycle 2	100.0% ± 0.1%	91.3% ± 2.0%	74.3% ± 3.0%	72.4% ± 3.1%
Change (C2–C1)	-0.0% ± 0.1%	4.0% ± 3.0%	14.6% ± 4.6%	16.4% ± 4.6%

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

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	97.6% ± 1.1%	94.8% ± 1.5%	46.6% ± 3.5%	42.3% ± 3.4%
Cycle 2	98.5% ± 0.7%	89.7% ± 1.8%	95.2% ± 1.3%	84.9% ± 2.1%
Change (C2–C1)	0.9% ± 1.3%	-5.1% ± 2.4%	48.6% ± 3.7%	42.7% ± 4.0%

RAA Status

Finally, Tables C13 through C15 show the changes from Cycle 1 to Cycle 2 by contact mode and RAA group.

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

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	97.4% ± 0.6%	70.8% ± 1.8%	33.7% ± 1.9%	17.7% ± 1.6%
Cycle 2	98.1% ± 0.6%	68.8% ± 1.9%	41.6% ± 2.1%	20.2% ± 1.7%
Change (C2–C1)	0.7% ± 0.9%	-2.0% ± 2.7%	7.9% ± 2.8%	2.4% ± 2.3%

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

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	98.4% ± 0.4%	74.6% ± 1.4%	50.5% ± 1.6%	44.7% ± 1.6%
Cycle 2	99.0% ± 0.3%	74.9% ± 1.2%	76.0% ± 1.2%	57.6% ± 1.4%
Change (C2–C1)	0.6% ± 0.5%	0.3% ± 1.8%	25.6% ± 2.0%	13.0% ± 2.1%

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

Cycle	Email	Telephone	Postal Address	All Modes Accurate
Cycle 1	97.7% ± 0.5%	85.5% ± 1.1%	41.6% ± 1.6%	36.8% ± 1.5%
Cycle 2	98.3% ± 0.4%	89.2% ± 0.9%	62.8% ± 1.3%	57.1% ± 1.4%
Change (C2–C1)	0.5% ± 0.6%	3.7% ± 1.4%	21.2% ± 2.1%	20.3% ± 2.1%

2013 RAA Reasons for Syntax Error

In all prior WHOIS ARS studies we showed which accuracy tests were failed by each contact. We repeat these tables from Cycle 1, and also show the same data for Cycle 2.

Email Addresses

Table C16 shows that 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 required under the 2013 RAA.)

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

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	3,802	3,829	3,830	11,461
Missing	46	19	18	83
Total	3,848	3,848	3,848	11,544

Table C17: Total Email Address Errors by Contact Type (2013 RAA) – Cycle 1

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	3,713	3,725	3,726	11,164
Missing	27	15	14	56
@ Symbol Missing	1	1	1	3
Not Resolvable	1	1	1	3
Total	3,742	3,742	3,742	11,226

Table C18: Total Email Address Errors by Contact Type (2013 RAA) – Cycle 2

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	5,083	5,095	5,095	15,273
Missing	35	23	23	81
Not Resolvable	1	1	1	3
Total	5,119	5,119	5,119	15,357

Telephone Numbers

Table C18: Total Telephone Number Errors by Contact Type (2013 RAA) – Phase 1

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	3,336	3,362	3,389	10,087
Missing	121	102	103	326
Country Code Missing	76	70	65	211
Country Code Format	91	90	91	272
Incorrect Length	223	223	199	645
Characters Not Allowed	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) – Cycle 1

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	3,251	3,267	3,298	9,816
Missing	82	83	83	248
Country Code Missing	82	79	82	243
Country Code Format	*45*	*44*	*42*	*131*
Incorrect Length	282	269	237	788
Characters Not Allowed	0	0	0	0
Total	**3,742**	**3,742**	**3,742**	**11,226**

Note: Italics indicate new 2013 RAA requirements.

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

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	4,606	4,590	4,586	13,782
Not Present	60	62	63	185
Country Code Missing	87	86	100	273
Country Code Format	*53*	*56*	*67*	*176*
Incorrect Length	313	325	303	941
Characters Not Allowed	0	0	0	0
Total	5,119	5,119	5,119	15,357

Note: Italics indicate new 2013 RAA requirements.

Postal Addresses

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

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	2,213	2,209	2,258	6,680
Missing	16	19	19	54
Country Code Missing	2	3	3	8
Country Not Identifiable	10	10	10	30
Country in Wrong Field	811	812	812	2,435
Country Not ISO Alpha 2	0	0	0	0
Postal Code Missing	233	255	250	738
Postal Code Format	8	8	8	24
State/Province Missing	456	472	440	1,368
State/Province in Wrong Field	37	37	34	108
State/Province Format	47	46	45	138
City Missing	278	283	255	816
City in Wrong Field	165	165	186	516
Street Missing	249	242	224	715
Street in Wrong Field	30	31	28	89
TOTAL	**3,848**	**3,848**	**3,848**	**11,544**
Total Errors	2,342	2,383	2,314	7,039
Total Domains with Errors	1,635	1,639	1,590	4,864

Note: Italics indicate new 2013 RAA requirements.

Table C21: Total Postal Address Errors by Contact Type (2013 RAA) – Cycle 1

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	2,010	2,011	2,051	6,072
Missing	15	17	17	49
Country Code Missing	1	2	2	5
Country Not Identifiable	13	13	14	40
Country in Wrong Field	865	865	864	2,594
Postal Code Missing	270	283	270	823
Postal Code Format	7	8	8	23
Postal Code in Wrong Field	0	0	0	0
State/Province Missing	459	468	409	1,336
State/Province in Wrong Field	33	27	24	84
State/Province Format	62	62	71	195
City Missing	366	365	337	1,068
City in Wrong Field	244	250	304	798
Street Missing	312	305	286	903
Street in Wrong Field	38	38	28	104
TOTAL	**3,742**	**3,742**	**3,742**	**11,226**
Total Errors	2,685	2,703	2,634	8,022
Total Domains with Errors	1,732	1,731	1,691	5,154

Note: Italics indicate new 2013 RAA requirements.

Table C21: Total Postal Address Errors by Contact Type (2013 RAA) – Cycle 1

	Registrant	Administrative	Technical	Total
Passed All Accuracy Tests	3,041	3,037	3,093	9,171
Missing	23	25	25	73
Country Code Missing	1	1	1	3
Country Not Identifiable	33	33	32	98
Country in Wrong Field	*61*	*61*	*61*	*183*
Country Not ISO Alpha 2	0	0	0	0
Postal Code Missing	362	389	386	1,137
Postal Code Format	12	11	11	34
Postal Code in Wrong Field	*0*	*0*	*0*	*0*
State/Province Missing	893	914	805	2,612
State/Province in Wrong Field	*45*	*44*	*42*	*131*
State/Province Format	*96*	*98*	*116*	*310*
City Missing	568	582	538	1,688
City in Wrong Field	*389*	*386*	*483*	*1,258*
Street Missing	425	415	386	1,226
Street in Wrong Field	*109*	*108*	*95*	*312*
TOTAL	5,119	5,119	5,119	15,357
Total Errors	3,017	3,067	2,981	9,065

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 became required fields. Results for Accuracy to 2013 RAA requirements for operability would be very repetitive, and are thus not presented in this report.

Appendix D: Additional Analyses – Scripts and Accuracy by Region

Table D1: Africa Region Syntax and Operability Accuracy, by Contact Type and Script Language

	Script Language	Count	Syntax Percent Accurate	Operability Percent Accurate
Registrant	**Latin**	1,256	34.4% ± 2.6%	68.4% ± 2.6%
	Arabic	1	0.0% ± 0.0%	0.0% ± 0.0%
	Chinese (Hanzi)	0	NA ± NA	NA ± NA
	Diacritical Marking	14	42.9% ± 25.9%	71.4% ± 23.7%
Administrative	**Latin**	1,255	35.4% ± 2.6%	68.9% ± 2.6%
	Arabic	1	0.0% ± 0.0%	0.0% ± 0.0%
	Chinese (Hanzi)	0	NA ± NA	NA ± NA
	Diacritical Marking	13	46.2% ± 27.1%	69.2% ± 25.1%
Technical	**Latin**	1,254	40.7% ± 2.7%	70.7% ± 2.5%
	Arabic	1	0.0% ± 0.0%	0.0% ± 0.0%
	Chinese (Hanzi)	0	NA ± NA	NA ± NA
	Diacritical Marking	14	42.9% ± 25.9%	71.4% ± 23.7%

Table D2: Asia-Pacific Region Syntax and Operability Accuracy, by Contact Type and Script Language

	Script Language	Count	Syntax Percent Accurate	Operability Percent Accurate
Registrant	**Latin**	2,891	47.1% ± 1.8%	63.3% ± 1.8%
	Arabic	1	0.0% ± 0.0%	100.0% ± 0.0%
	Chinese (Hanzi)	66	15.3% ± 8.7%	73.9% ± 10.6%
	Diacritical Marking	7	19.5% ± 29.3%	61.1% ± 36.1%
Administrative	**Latin**	2,882	45.9% ± 1.8%	62.8% ± 1.8%
	Arabic	1	0.0% ± 0.0%	100.0% ± 0.0%
	Chinese (Hanzi)	65	15.8% ± 8.9%	39.7% ± 11.9%
	Diacritical Marking	7	19.5% ± 29.3%	61.1% ± 36.1%
Technical	**Latin**	2,882	51.8% ± 1.8%	60.3% ± 1.8%
	Arabic	1	0.0% ± 0.0%	100.0% ± 0.0%
	Chinese (Hanzi)	65	12.7% ± 8.1%	34.9% ± 11.6%
	Diacritical Marking	5	20.0% ± 35.1%	60.0% ± 42.9%

Table D3: Europe Region Syntax and Operability Accuracy, by Contact Type and Script Language

	Script Language	Count	Syntax Percent Accurate	Operability Percent Accurate
Registrant	Latin	2,619	65.1% ± 1.8%	74.1% ± 1.7%
	Arabic	0	NA ± NA	NA ± NA
	Chinese (Hanzi)	0	NA ± NA	NA ± NA
	Diacritical Marking	123	64.4% ± 8.5%	66.6% ± 8.3%
Administrative	Latin	2,619	65.0% ± 1.8%	74.2% ± 1.7%
	Arabic	0	NA ± NA	NA ± NA
	Chinese (Hanzi)	0	NA ± NA	NA ± NA
	Diacritical Marking	125	65.0% ± 8.4%	64.2% ± 8.4%
Technical	Latin	2,619	71.4% ± 1.7%	71.2% ± 1.7%
	Arabic	0	NA ± NA	NA ± NA
	Chinese (Hanzi)	0	NA ± NA	NA ± NA
	Diacritical Marking	99	81.1% ± 7.7%	71.1% ± 8.9%

Table D4: Latin America/Caribbean Region Syntax and Operability Accuracy, by Contact Type and Script Language

	Script Language	Count	Syntax Percent Accurate	Operability Percent Accurate
Registrant	**Latin**	1,905	66.2% ± 2.1%	73.1% ± 2.0%
	Arabic	0	NA ± NA	NA ± NA
	Chinese (Hanzi)	0	NA ± NA	NA ± NA
	Diacritical Marking	30	19.7% ± 14.2%	63.7% ± 17.2%
Administrative	**Latin**	1,906	66.8% ± 2.1%	73.6% ± 2.0%
	Arabic	0	NA ± NA	NA ± NA
	Chinese (Hanzi)	0	NA ± NA	NA ± NA
	Diacritical Marking	29	20.3% ± 14.6%	61.4% ± 17.7%
Technical	**Latin**	1,906	68.5% ± 2.1%	74.4% ± 2.0%
	Arabic	0	NA ± NA	NA ± NA
	Chinese (Hanzi)	0	NA ± NA	NA ± NA
	Diacritical Marking	23	26.2% ± 18.0%	59.8% ± 20.0%

Table D5: North America Region Syntax and Operability Accuracy, by Contact Type and Script Language

	Script Language	Count	Syntax Percent Accurate	Operability Percent Accurate
Registrant	**Latin**	3,286	85.0% ± 1.2%	82.4% ± 1.3%
	Arabic	0	NA ± NA	NA ± NA
	Chinese (Hanzi)	0	NA ± NA	NA ± NA
	Diacritical Marking	0	NA ± NA	NA ± NA
Administrative	**Latin**	3,286	84.4% ± 1.2%	82.1% ± 1.3%
	Arabic	0	NA ± NA	NA ± NA
	Chinese (Hanzi)	0	NA ± NA	NA ± NA
	Diacritical Marking	0	NA ± NA	NA ± NA
Technical	**Latin**	3,282	84.1% ± 1.3%	82.3% ± 1.3%
	Arabic	0	NA ± NA	NA ± NA
	Chinese (Hanzi)	0	NA ± NA	NA ± NA
	Diacritical Marking	0	NA ± NA	NA ± NA