WHOIS Accuracy Reporting System (ARS): Phase 2 Cycle 2 Results Webinar | 16 June 2016

ICANN GDD Operations
NORC at the University of Chicago
Phase 2 Cycle 2 – Webinar Agenda

1. WHOIS ARS Background
2. Phase 2 Cycle 2: Timeline and Process
3. Phase 2 Cycle 2: Sample Design and Information
4. Phase 2 Cycle 2: Results and Major Findings
5. Summary & Next Steps
6. Question & Answer Session
WHOIS ARS Background
Pilot
“Proof of Concept”: Tested processes for data collection and validation
Report: Published 23 December 2014
Public Comment Report: Published 3 April 2015

Phase 1: Syntax Accuracy only
Is the record correctly formatted?
Report: Published 24 August 2015

Phase 2: Syntax + Operability Accuracy
Does the email go through, phone ring, mail deliver?
Cycle 1 Report: Published 23 December 2015
Cycle 2 Report: Published 8 June 2016

Phase 3 TBD, if at all: Identity Validation
Is the contacted individual responsible for the domain?
Phase 2 Cycle 2
Process and Timeline
Phase 2 Cycle 2 – Cross-Functional Team

ICANN Team
- GDD OPERATIONS
- CONTRACTUAL COMPLIANCE
- REGISTRAR SERVICES
- LEGAL
- IT & PRODUCT MANAGEMENT

Vendor Team
- NORC
- WHIBSE
- DIGICERT
- UPU

Accuracy Reports
Phase 2 Cycle 2 – Cyclical Timeline of Phase 2

**Cycle 2: Complete**
- Accuracy Criteria Refinement
- Data Collection
- Accuracy Testing
- Data Analysis & Report Development
- Report Publication

**Cycle 3: Begin July 2016**
- Accuracy Criteria Refinement
- Data Collection
- Accuracy Testing
- Data Analysis & Report Development
- Report Publication
Phase 2 Cycle 2 – Contact types, modes, and testing criteria

Registrant
- Email Address
- Telephone Number
- Postal Address

Administrative
- Email Address
- Telephone Number
- Postal Address

Technical
- Email Address
- Telephone Number
- Postal Address

RAA Type

Criteria Examples

**Syntax**: Does the email address contain an “@”?

**Operability**: Did the email bounce back?

**Syntax**: Does the telephone number have a country code?

**Operability**: Did the number ring when dialed?

**Syntax**: Does the postal address include an identifiable country?

**Operability**: Can mail be delivered to the address?

Detailed criteria listed at whois.icann.org/en/whoisars-validation

**GF** = Grandfathered. A domain registered before a registrar changed to the 2013 RAA. Obligated to 2009 RAA requirements.

**NGF** = Non-grandfathered. Obligated to 2013 RAA requirements.
Phase 2 Cycle 2
Sample Design and Testing Criteria
Accuracy Statistics by Subgroup
- Phase 2 Report provides both syntax and operability accuracy rates for:
  - The gTLD space, by region and in total
  - New gTLDs compared to Prior (legacy) gTLDs
  - RAA Type (2009, 2013GF, 2013NGF)
- Data within 95% confidence intervals, \( \leq +/- 5\% \) margin of error

Report identifies reasons for error
- All domains evaluated against 2009 RAA requirements for both syntax and operability
- Appendix C provides results on 2013NGF domains
- Detailed tests enable us to know in what way a record is inaccurate
- Cycle 2 Report contains new section on regional differences in accuracy and reasons for error
# Phase 2 Cycle 2 – Demographics

## gTLD Population At Time of Sample (January 2016)

<table>
<thead>
<tr>
<th>Records in gTLDs</th>
<th>Total gTLDs</th>
<th>2009 RAA*</th>
<th>2013GF RAA*</th>
<th>2013 NGF RAA*</th>
<th>New gTLDs</th>
<th>Prior gTLDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>169.5m</td>
<td>888</td>
<td>4.9m</td>
<td>88.8m</td>
<td>75.8m</td>
<td>870</td>
<td>18</td>
</tr>
</tbody>
</table>

### 200k Sample

<table>
<thead>
<tr>
<th>AFR</th>
<th>LAC</th>
<th>EUR</th>
<th>APAC</th>
<th>N.A.</th>
<th>2009 RAA</th>
<th>2013GF RAA</th>
<th>2013 NGF RAA</th>
<th>New gTLDs</th>
<th>Prior gTLDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3k</td>
<td>8.6k</td>
<td>36.4k</td>
<td>62.8k</td>
<td>85.9k</td>
<td>4.7k</td>
<td>82.3k</td>
<td>109.2k</td>
<td>570</td>
<td>18</td>
</tr>
</tbody>
</table>

### 12k Sub-sample

<table>
<thead>
<tr>
<th>AFR</th>
<th>LAC</th>
<th>EUR</th>
<th>APAC</th>
<th>N.A.</th>
<th>2009 RAA</th>
<th>2013GF RAA</th>
<th>2013 NGF RAA</th>
<th>New gTLDs</th>
<th>Prior gTLDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3k</td>
<td>1.9k</td>
<td>2.6k</td>
<td>2.9k</td>
<td>3.3k</td>
<td>2.2k</td>
<td>4.7k</td>
<td>5.1k</td>
<td>570</td>
<td>18</td>
</tr>
</tbody>
</table>

* Weighted estimates from 200k sample
Phase 2 Cycle 2 – Distribution of RAA type

Current Distribution (as of Jan 16)

The 2009 RAA share is shrinking very slowly; The share of non-grandfathered 2013 RAA domains are growing rapidly.

Change in Distribution across Sample Dates

- **2013 RAA GF**
- **2013 RAA NGF**
- **2009 RAA**

---

- solid line denotes increase
- dotted line denotes decrease
Phase 2 Cycle 2
Overall Results
Overall Accurate
All contact information has passed all tests

Contactable
Any of the nine contacts are operable
Phase 2 Cycle 2 – Overall Syntax and Operability Accuracy by Region

Data as of January 2016
Categorized by ICANN region
Phase 2 Cycle 2 – How contactable are the WHOIS records?

99% of records contactable by at least one method

Only 1% of records were entirely non-contactable
Phase 2 Cycle 2
Results and Major Findings: Syntax, 2009 RAA
Phase 2 Cycle 2 – Changes Over Time: Syntax Accuracy by Contact

Entire gTLD Space
Phase 1 through Phase 2 Cycle 2

- Email Address
  - Phase 1: 99.2%
  - Phase 2, Cycle 1: 99.1%
  - Phase 2, Cycle 2: 99.2%

- Telephone Number
  - Phase 1: 85.8%
  - Phase 2, Cycle 1: 83.3%
  - Phase 2, Cycle 2: 85.3%

- Postal Address
  - Phase 1: 79.1%
  - Phase 2, Cycle 1: 79.4%
  - Phase 2, Cycle 2: 76.4%

Accuracy here also requires accuracy on all 3 contact modes and all 3 contact types

Overall Syn Accuracy
Cycle 1  |  Cycle 2
67.9%  |  67.2%
Δ -0.7%
Phase 2 Cycle 2 – Changes Over Time: Syntax Accuracy by RAA Type

• Accuracy here also requires accuracy on all 3 contact modes and all 3 contact types
Phase 2 Cycle 2 – Overall Syntax Accuracy by Region

Data as of January 2016
Categorized by ICANN region

Entire gTLD Space

<table>
<thead>
<tr>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>67.9%</td>
<td>67.2%</td>
<td>-0.7%</td>
</tr>
</tbody>
</table>

North America

<table>
<thead>
<tr>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>83.9%</td>
<td>82.8%</td>
<td>-1.1%</td>
</tr>
</tbody>
</table>

Latin America/Caribbean Islands

<table>
<thead>
<tr>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>56.9%</td>
<td>64.7%</td>
<td>+7.8%*</td>
</tr>
</tbody>
</table>

Europe

<table>
<thead>
<tr>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>58.8%</td>
<td>60.6%</td>
<td>+1.8%</td>
</tr>
</tbody>
</table>

Asia/Australia/Pacific Islands

<table>
<thead>
<tr>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>39.8%</td>
<td>45.0%</td>
<td>+5.5%*</td>
</tr>
</tbody>
</table>

Africa

<table>
<thead>
<tr>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>29.8%</td>
<td>29.3%</td>
<td>-0.5%</td>
</tr>
</tbody>
</table>
Note: A missing telephone number in the Registrant contact type is not a requirement of the 2009 RAA. This graph shows the percentage of overall error types found in the Administrative contact type. The “Unallowable Character” error type has been combined with the “Missing” error type, because unallowable character errors represent less than 0.2% of overall errors.
Phase 2 Cycle 2 – Telephone Syntax Error by Region

Most Common Error (MCE)

- **Europe**
  - **All Tel**: 937 errors
  - **MCE**: Incorrect Length

- **Asia/Australia/Pacific Islands**
  - **All Tel**: 1,088 errors
  - **MCE**: Incorrect Length

- **North America**
  - **All Tel**: 1,074 errors
  - **MCE**: Country Code Missing

- **Latin America/Caribbean Islands**
  - **All Tel**: 845 errors
  - **MCE**: Incorrect Length

- **Africa**
  - **All Tel**: 1,038 errors
  - **MCE**: Incorrect Length
Phase 2 Cycle 2 – Reasons for Postal Syntax Error

Reasons for Postal Address Syntax Error
Administrative Contact Type
5,114 Total Errors

State/Province missing: 33.1% (1,699 errors)
City missing: 27.7% (1,411 errors)
Postal code missing or bad format: 20.4% (1,060 errors)
Street missing: 15.1% (764 errors)
Country code missing or undentifiable: 2.6% (128 errors)
Missing: 1.0% (52 errors)

Percent of all 5,114 errors
Phase 2 Cycle 2 – Postal Syntax Error by Region

**Most Common Error (MCE)**

- **Europe**
  - All Post | 1,765 errors
  - MCE | State/Province Missing

- **Asia/Australia/Pacific Islands**
  - All Post | 2,728 errors
  - MCE | State/Province Missing

- **North America**
  - All Post | 1,112 errors
  - MCE | City Missing

- **Latin America/Caribbean Islands**
  - All Post | 831 errors
  - MCE | Postal Code Missing or Bad Format

- **Africa**
  - All Post | 916 errors
  - MCE | City Missing
Phase 2 Cycle 2
Results and Major Findings: Operability, 2009 RAA
Phase 2 Cycle 2 – Changes Over Time: Operability Accuracy by Contact Mode

Saw mostly improvement from Cycle 1 to Cycle 2

Accuracy here also requires accuracy on all 3 contact modes and all 3 contact types
Phase 2 Cycle 2 – Changes Over Time: Operability Accuracy by RAA

Entire gTLD Space
Phase 2 Cycle 1 through Phase 2 Cycle 2

<table>
<thead>
<tr>
<th>Overall Op Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle 1</td>
</tr>
<tr>
<td>64.7%</td>
</tr>
<tr>
<td>Δ +5.4%</td>
</tr>
</tbody>
</table>

- Saw mostly improvement from Cycle 1 to Cycle 2
- Accuracy here also requires accuracy on all 3 contact modes and all 3 contact types
Phase 2 Cycle 2 – Overall Operability Accuracy by Region

Data as of January 2016
Categorized by ICANN region

<table>
<thead>
<tr>
<th>Region</th>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>73.2%</td>
<td>80.2%</td>
<td>+6.9%</td>
</tr>
<tr>
<td>Latin America/Caribbean Islands</td>
<td>72.7%</td>
<td>71.6%</td>
<td>-1.0%</td>
</tr>
<tr>
<td>Europe</td>
<td>59.8%</td>
<td>63.1%</td>
<td>+3.3%</td>
</tr>
<tr>
<td>Asia/Australia/Pacific Islands</td>
<td>49.4%</td>
<td>57.6%</td>
<td>+8.2%</td>
</tr>
<tr>
<td>Entire gTLD Space</td>
<td>64.7%</td>
<td>70.2%</td>
<td>+5.4%</td>
</tr>
</tbody>
</table>
Phase 2 Cycle 2 – Reasons for Telephone Operability Error

Reasons for Telephone Number Operability Error
Administrative Contact Type
2,494 total errors

- Other Not Connected: 47.0% (1,171 errors)
- Invalid Number: 28.1% (702 errors)
- Number Disconnected: 20.5% (511 errors)
- Not Verifiable or Missing: 4.4% (110 errors)

Note: A missing telephone number in the Registrant contact type is not a requirement of the 2009 RAA. This graph shows the percentage of overall error types found in the Administrative contact type.
Phase 2 Cycle 2 – Most Common Reason for Telephone Operability Error by Region

**Most Common Error (MCE)**

- **Overall Tel**: 1,765 errors
  - MCE: Not Connected

- **Overall Tel**: 2,728 errors
  - MCE: Invalid #

- **Europe**
  - Overall Tel: 1,765 errors
    - MCE: Not Connected

- **Asia/Australia/Pacific Islands**
  - Overall Tel: 2,728 errors
    - MCE: Invalid #

- **North America**
  - All Tel: 1,112 errors
    - MCE: Not Connected

- **Latin America/Caribbean Islands**
  - Overall Tel: 831 errors
    - MCE: Not Connected

- **Africa**
  - Overall Tel: 916 errors
    - MCE: Not Connected
Phase 2 Cycle 2
Additional Findings
Phase 2 – Cycle 1 to Cycle 2 Changes

**Email**
- Syntax accuracy consistent at 99%.
- Operability accuracy increased from 87.1% to 91.4%

**Telephone**
- Syntax and operability accuracy both increased, from 83.3% to 85.3% and from 74.0% to 76.0% respectively

**Postal**
- Syntax accuracy decreased from 80.2% to 77.3%
- Operability accuracy consistent at 98%
In Phase 2 Cycle 2, records were analyzed to determine which script language was used to register the domain.

The majority of domains across all regions are registered using Latin script (~98%), but some are registered using scripts such as Diacritic marking (1.4%), Hanzi (0.5%), or Arabic (0.02%).

All records with Hanzi script were registered in the Asia-Pacific region. Records with Diacritic marking were typically registered in either the Europe, Latin America/Caribbean, or Asia-Pacific region.

*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 for the differences between any script types (e.g., difference between accuracy of Latin and Arabic).*
Phase 2 Cycle 2
ICANN Contractual Compliance
Follow-Up
Potential records have been provided to ICANN Contractual Compliance.

Registrars must investigate and correct inaccurate WHOIS data:
- Section 3.7.8 of 2009 and 2013 RAA (and WHOIS Accuracy Program Specification)
- Failure to respond or demonstrate compliance during complaint processing will result in a Notice of Breach

Registrars under 2013 RAA must use WHOIS format and layout required by Registration Data Directory Service Specification.

WHOIS inaccuracy and format complaints will follow the Contractual Compliance Approach and Process.

ICANN will continue to give priority to complaints submitted by community members.

For WHOIS ARS Phase 1, 3,168 tickets were created in the compliance system. 1,942 were closed before first notice because the domain names either had already been suspended or the current WHOIS data was not consistent with the sampled data. The relevant registrars addressed the remaining 1,226 tickets per Contractual Compliance process.

For WHOIS ARS Phase 2, Cycle 1, 3,689 tickets were created and 1,663 tickets remain to be processed. Metrics will be available in the Compliance Quarterly Reports (see https://www.icann.org/resources/pages/compliance-reports-2016-04-15-en).
Summary & Next Steps
Phase 2 Cycle 2 Summary

Phase 2 Cycle 2 Report published 8 June 2016

Subsample of 12k records; Accounted for regions and RAA type

67% syntax accuracy rate and 70% operability accuracy rate on all 2009 RAA requirements

Increase in overall Op accuracy; Email Syntax Accuracy and Postal Operability accuracy again very high

Compliance will conduct follow up on potentially inaccurate records

Next Cycle begins July 2016; Report expected December 2016
Questions & Answers