WHOIS Accuracy Reporting System (ARS): Phase 2 Cycle 1 Results Webinar | 12 January 2016

ICANN GDD Operations
NORC at the University of Chicago
Webinar Agenda

1. WHOIS ARS Background
2. Phase 2 Cycle 1: Timeline and Process
3. Phase 2 Cycle 1: Sample Design and Information
4. Phase 2 Cycle 1: Results and Major Findings
5. Summary & Next Steps
6. Question & Answer Session
WHOIS ARS Background
WHOIS ARS Implementation

0  Pilot (Public Comment Report Completed March 2015)
   “Proof of Concept”: Tested processes for data collection and validation

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

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

3  Phase 3 TBD, if at all: Identity Validation
   Is the contacted individual responsible for the domain?
Phase 2 Cycle 1
Process and Timeline
Phase 2 Cycle 1 – 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 Cyclical Timeline**

**Cycle 1: Complete**

**Cycle 2: In progress**

Note: Phase 2 Cycle 1 began in June 2015, overlapping with Phase 1
Phase 2 Cycle 1 – Contact types, modes, and testing criteria

Criteria listed in Appendix A

RAA Type

Registrant
• Email Address
• Telephone Number
• Postal Address

Technical
• Email Address
• Telephone Number
• Postal Address

Administrative
• Email Address
• Telephone Number
• Postal Address

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?
Phase 2 Cycle 1
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, ≤+/- 5% margin of error

Report describes why records are inaccurate

- 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 why a record is inaccurate
## Phase 2 Cycle 1 – Demographics

### gTLD Population At Time of Sample (June 2015)

<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>158m</td>
<td>678</td>
<td>5m</td>
<td>101m</td>
<td>52m</td>
<td>660</td>
<td>18</td>
</tr>
</tbody>
</table>

### 150k 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>988</td>
<td>5.5k</td>
<td>30.6k</td>
<td>38.9k</td>
<td>70.3k</td>
<td>3.8k</td>
<td>72.7k</td>
<td>70.5k</td>
<td>424</td>
<td>18</td>
</tr>
</tbody>
</table>

### 10k 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>988</td>
<td>1.8k</td>
<td>2.0k</td>
<td>2.4k</td>
<td>2.7k</td>
<td>2.3k</td>
<td>3.9k</td>
<td>3.7k</td>
<td>424</td>
<td>18</td>
</tr>
</tbody>
</table>

* Weighted estimates from 150k sample
Phase 2 Cycle 1
Results and Major Findings: Syntax
Phase 1 Syntax Accuracy Results Recap

- Phase 1 looked at Syntax Accuracy only

- Email accuracy very high; postal address lowest (most complex)
  - If an email address is provided, it passed all the tests
  - Two-thirds of telephone number errors due to wrong # of digits
  - Postal address errors: typically missing 1+ required field

- 70.3% of all domains meet all 2009 RAA syntax requirements
  - For all three contact types (Registrant, Administrative, Technical)
  - For all three contact modes (email, telephone, postal address)

- Larger initial sample size (increase from 100k to 150k) can help reduce need for oversampling of subgroups
67% of all domains passed all syntax tests for all contact types and modes.

99% of email addresses, 83% of telephone numbers, and 79% of postal addresses met all syntax requirements of the 2009 RAA.
- Syntax reasons for error similar to Phase 1.
- The drop in telephone number accuracy possibly due to an increase in missing country codes.
- For postal addresses, the majority of errors in both Phase 1 and Phase 2 Cycle 1 were due to missing fields such as city, state/province, postal code, or street.

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.
Contact mode syntax accuracy requires accuracy on all 3 contact types
- Registrant, Administrative, Technical Contacts
- Overall syntax accuracy requires accuracy on all 3 contact modes and all 3 contact types
Phase 2 Cycle 1 – Syntax Conformance to 2009 RAA* by RAA

• Syntax accuracy here also requires accuracy on all 3 contact modes and all 3 contact types

Overall Domain Syntax Accuracy 67.2%
Phase 2 Cycle 1 – Syntax Conformance to 2009 RAA by Region

- **European**
  - 58.8%

- **North America**
  - 83.9%

- **Asia/Australia/Pacific Islands**
  - 39.5%

- **Latin America/Caribbean Islands**
  - 56.9%

- **Africa**
  - 29.8%

*Syntax accuracy here also requires accuracy on all 3 contact modes and all 3 contact types*
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 1 – Reasons for Postal Address Syntax Error (2009)

Reasons for Postal Address Syntax Error
Administrative Contact Type

- City missing: 30.7%
- Postal code missing or bad format: 27.2%
- Street missing: 19.3%
- State/Province missing: 18.4%
- Country code missing or undentifiable: 2.5%
- Missing: 1.9%

Percent of Overall Reasons for Syntax Error
Phase 2
Results and Major Findings: Operability
Phase 2 Cycle 1 – Summary of Findings for Operability

- 65% of domains passed all operability tests for all contact types and contact modes.

- 87% of email, 74% of telephone and 98% of postal met all operability requirements of the 2009 RAA.
  - Of those email addresses that failed operability, 10 percent bounced.
  - Of the telephone numbers that were present, but failed operability, there were roughly equal numbers that were disconnected, invalid, or that simply did not connect.
  - For the small numbers of postal addresses that failed operability testing, almost half did not have an identifiable or easily deduced country.
Phase 2 Cycle 1 – Operability Conformance to 2009 RAA* by Contact Mode

Entire gTLD Space

- Contact mode operability accuracy requires accuracy on all 3 contact types
  - Registrant, Administrative, Technical Contacts
- Overall operability accuracy requires accuracy on all 3 contact modes and all 3 contact types

Overall Domain Operability Accuracy 64.7%
Phase 2 Cycle 1 – Operability Conformance to 2009 RAA* by RAA Type

Whois ARS

Entire gTLD Space

- Operability accuracy here also requires accuracy on all 3 contact modes and all 3 contact types
Phase 2 Cycle 1 – Operability Conformance to 2009 RAA by Region

- **North America**: 73.2%
- **Latin America/Caribbean Islands**: 72.7%
- **Africa**: 57.0%
- **Europe**: 59.8%
- **Asia/Australia/Pacific Islands**: 49.4%

- Operability accuracy here also requires accuracy on all 3 contact modes and all 3 contact types
Phase 2 Cycle 1 – Reasons for Telephone Operability Error (2009)

Reasons for Telephone Number Operability Error
Administrative Contact Type

- Other Not Connected: 36.8%
- Invalid Number: 31.7%
- Number Disconnected: 25.8%
- Not Verifiable (or Missing): 5.7%

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 1 – Reasons for Postal Address Operability Error (2009)

Reasons for Postal Address Operability Error
Administrative Contact Type

Error Type

- Not likely deliverable: 53.8%
- No country: 45.8%
- Unverifiable: 0.4%

Percent of Overall Reasons for Operability Error
Phase 2
Additional Findings
Phase 2 Cycle 1 – Additional Findings

- **Phase 1 vs. Phase 2 Cycle 1 Syntax**
  - Telephone accuracy dropped from 85.8% to 83.3%
  - More missing country codes for Phase 2 Cycle 1 telephone numbers
  - Change may be random, but will be monitored going forward

- **Syntax vs. Operability Accuracy**
  - Email syntax accuracy and postal operability accuracy very high
  - Telephone has > 2% nonconformance for syntax (16.7%) and operability (26.0%)
    - Of those with syntax nonconformance, 75% have Operability nonconformance
    - Of those with operability nonconformance, about half have Syntax nonconformance
ICANN Contractual Compliance Follow-Up
Potentially inaccurate records have been provided to ICANN Contractual Compliance.

- Due to sample weighting and 2013 requirements, the total number of records identified as nonconforming exceeds 30% of the 10k subsample.

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.
Summary & Next Steps
Phase 2 Cycle 1 Summary

- Phase 2 Cycle 1 Report published 23 December 2015
- Subsample of 10k records; Accounted for regions and RAA type
- 67% syntax accuracy rate and 64% operability accuracy rate on all 2009 RAA requirements
- Telephone Syntax Accuracy changed from Phase 1; Email Syntax Accuracy and Postal Operability accuracy very high
- Compliance will conduct follow up on potentially inaccurate records
- Next Cycle underway; Report expected June 2016
Phase 2 Cycle 2 – Implementation Underway

**Cycle 2: In progress**
- Increase sample size to 200,000 records
- Increase subsample size to 12,000 records
- Shift to focus on regional differences in data and reasons for error
- Continued integration with ICANN Contractual Compliance

**Follow-on Cycle**

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

**Timeline**
- Jan
- Feb
- Mar
- Apr
- May
- Jun
- Jul
- Aug
- Sep
- Oct
- Nov
- Dec
Questions & Answers