



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

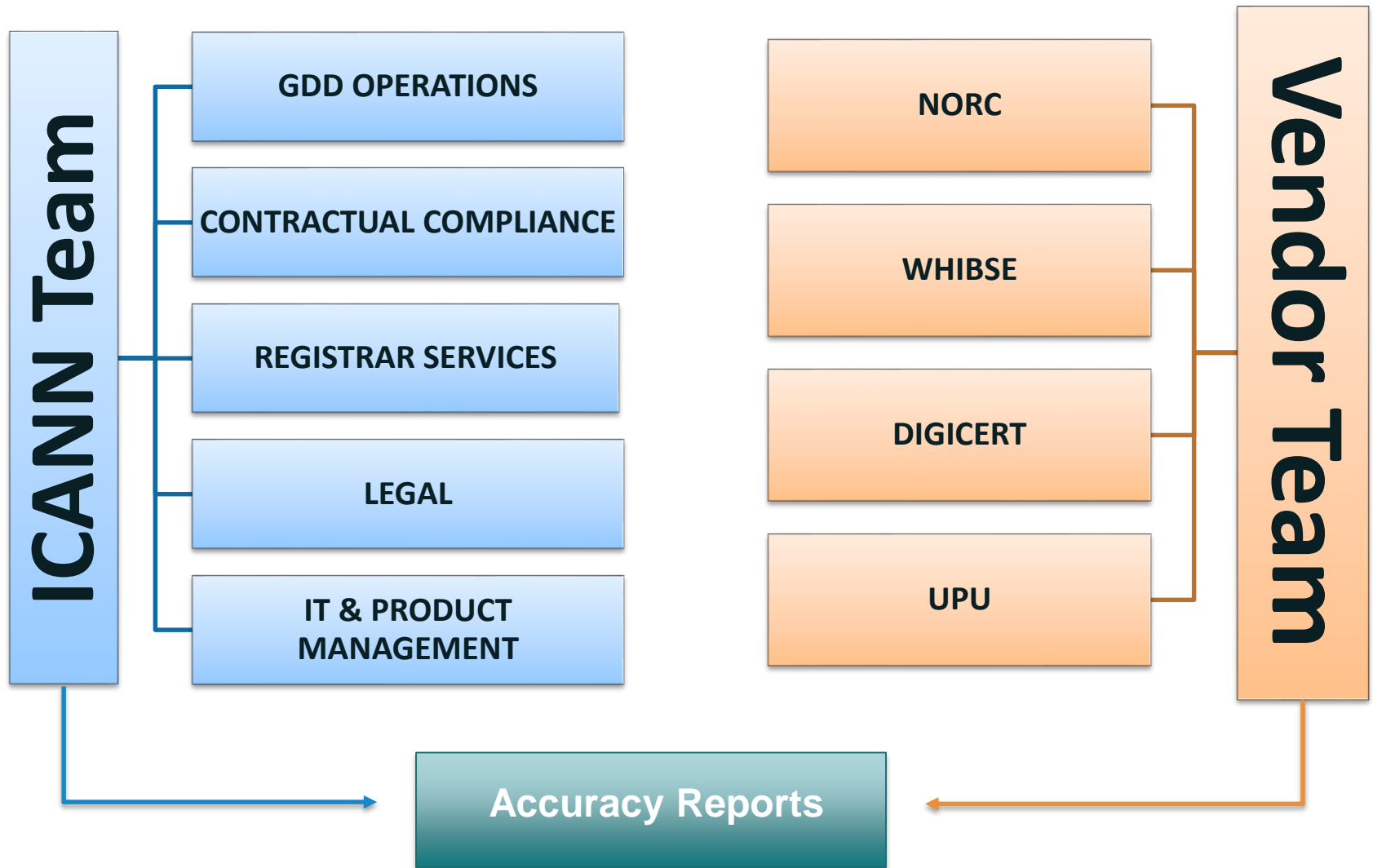
Phase 2: *Syntax + Operability Accuracy*
Does the email go through, phone ring, mail deliver?
Cycle 1 Report: 23 December 2015
Cycle 2 Report: June 2016

3

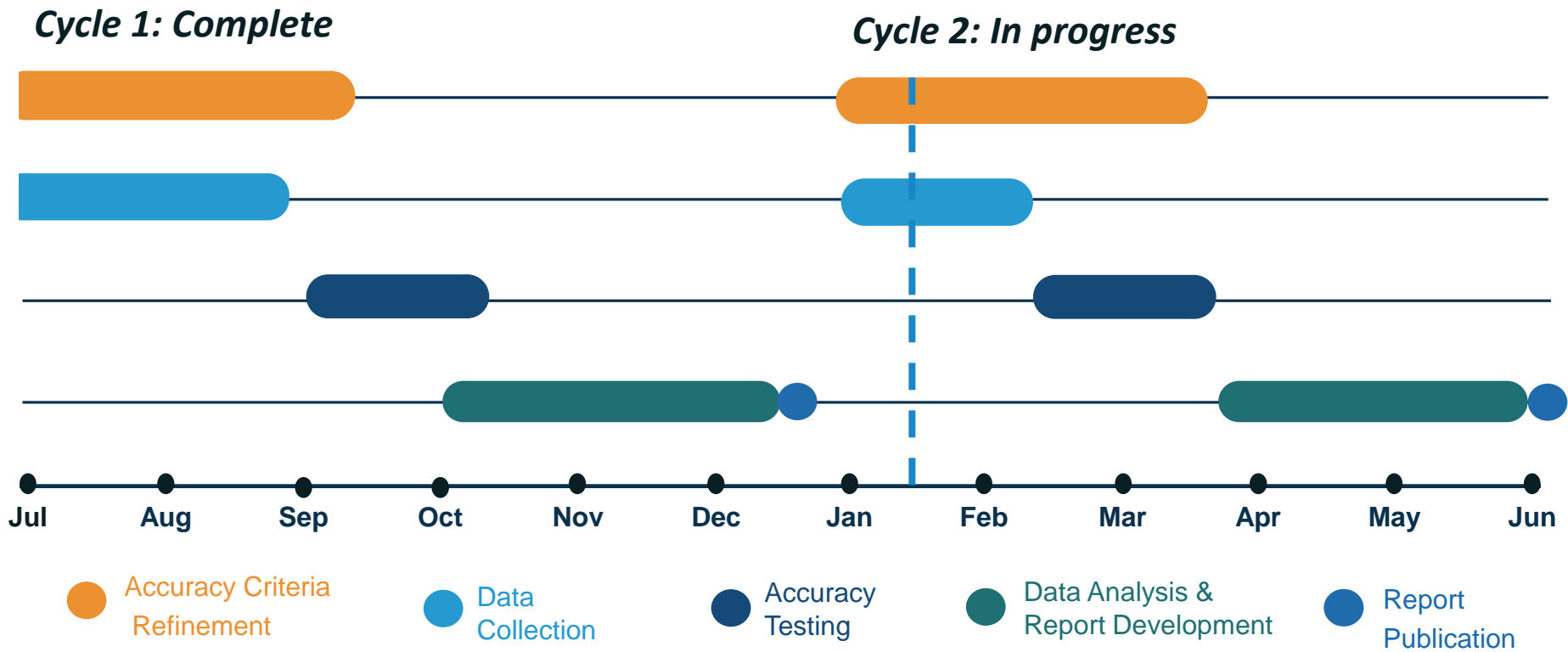
Phase 3 TBD, if at all: *Identity Validation*
Is the contacted individual responsible for the domain?

Phase 2 Cycle 1 Process and Timeline

Phase 2 Cycle 1 – Cross-Functional Team

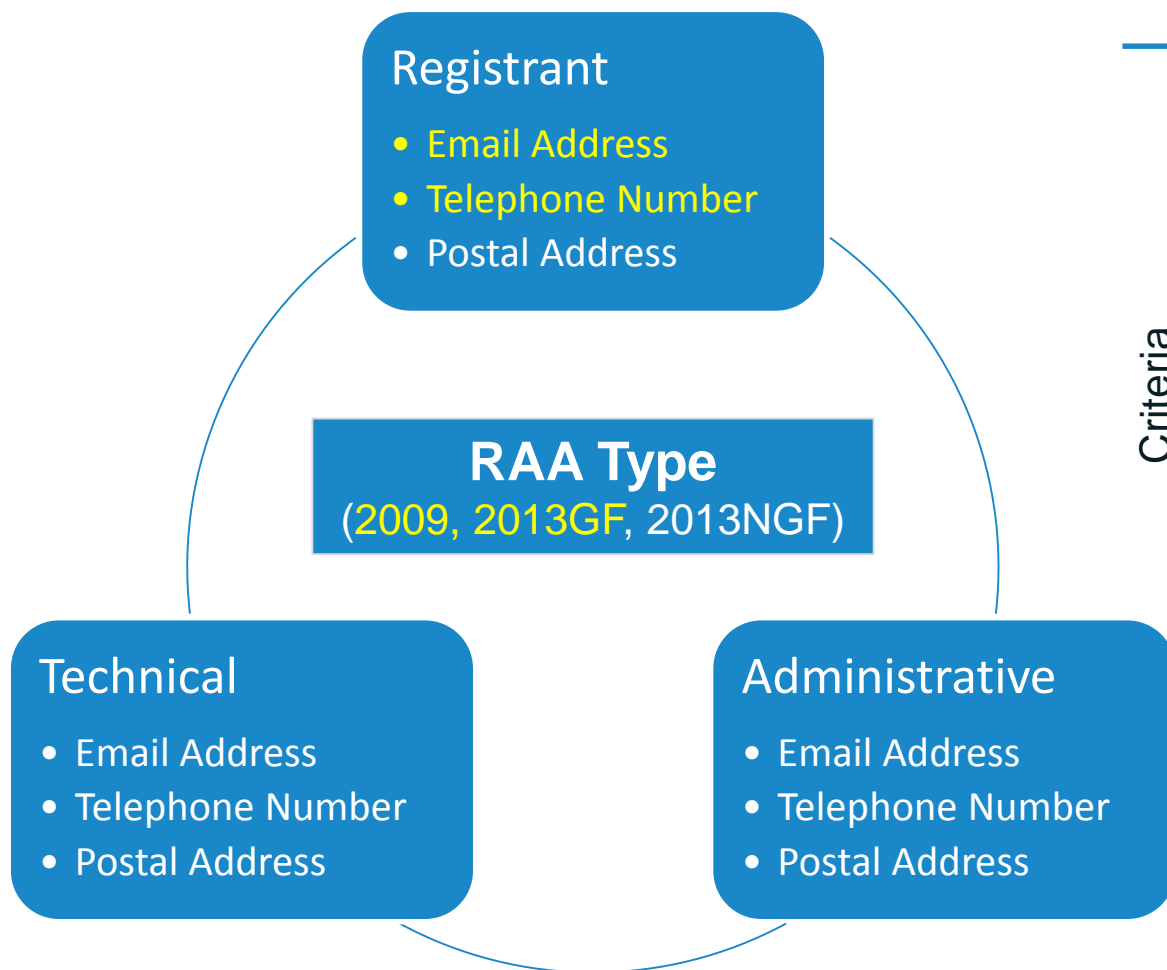


Phase 2 Cyclical Timeline



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

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

Phase 2 Cycle 1 – Report Highlights

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, $\leq \pm 5\%$ 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)

Records in gTLDs	Total gTLDs	2009 RAA*	2013GF RAA*	2013 NGF RAA*	New gTLDs	Prior gTLDs
158m	678	5m	101m	52m	660	18


150k Sample

AFR	LAC	EUR	APAC	N.A.	2009 RAA	2013GF RAA	2013 NGF RAA	New gTLDs	Prior gTLDs
988	5.5k	30.6k	38.9k	70.3k	3.8k	72.7k	70.5k	424	18

10k Sub-sample

AFR	LAC	EUR	APAC	N.A.	2009 RAA	2013GF RAA	2013 NGF RAA	New gTLDs	Prior gTLDs
988	1.8k	2.0k	2.4k	2.7k	2.3k	3.9k	3.7k	424	18

* Weighted estimates from 150k sample



Phase 2 Cycle 1

Results and Major Findings: Syntax

Phase 1 Syntax Accuracy Results Recap

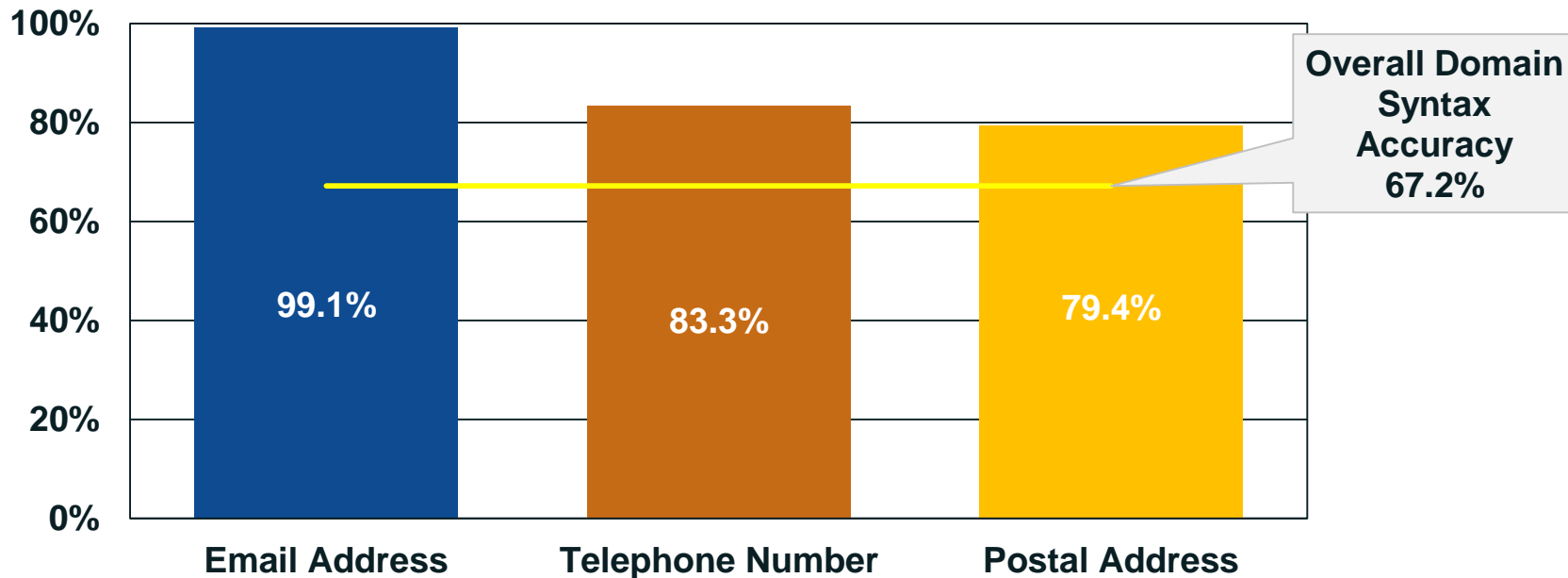
- ⦿ Phase 1 looked at Syntax Accuracy only
- ⦿ Email accuracy very high; postal address lowest (most complex)
 - If an email address is provided, it passed all the tests
 - Two-thirds of telephone number errors due to wrong # of digits
 - Postal address errors: typically missing 1+ required field
- ⦿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

Phase 2 Cycle 1 – Summary of Findings for Syntax

- ⦿ 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.

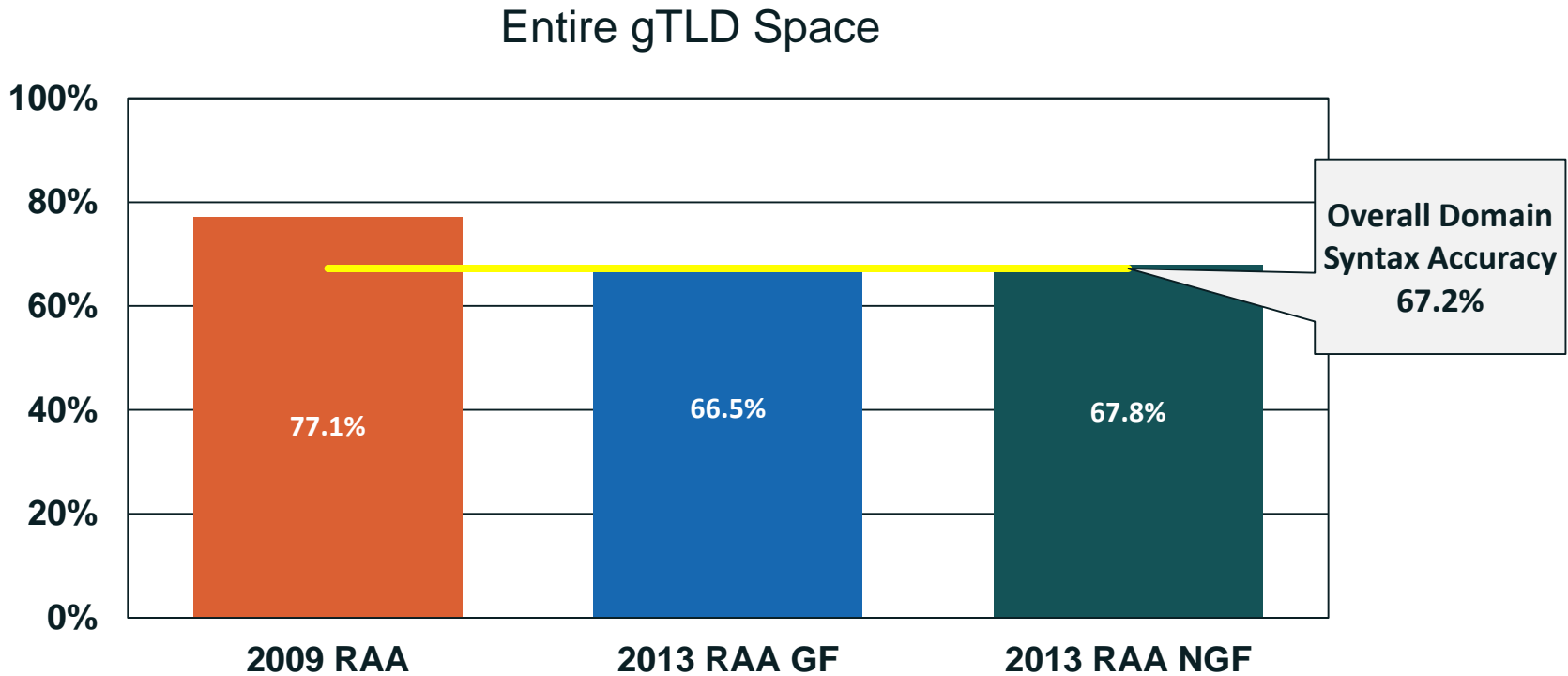
Phase 2 Cycle 1 – Syntax Conformance to 2009 RAA* by Contact Mode

Entire gTLD Space



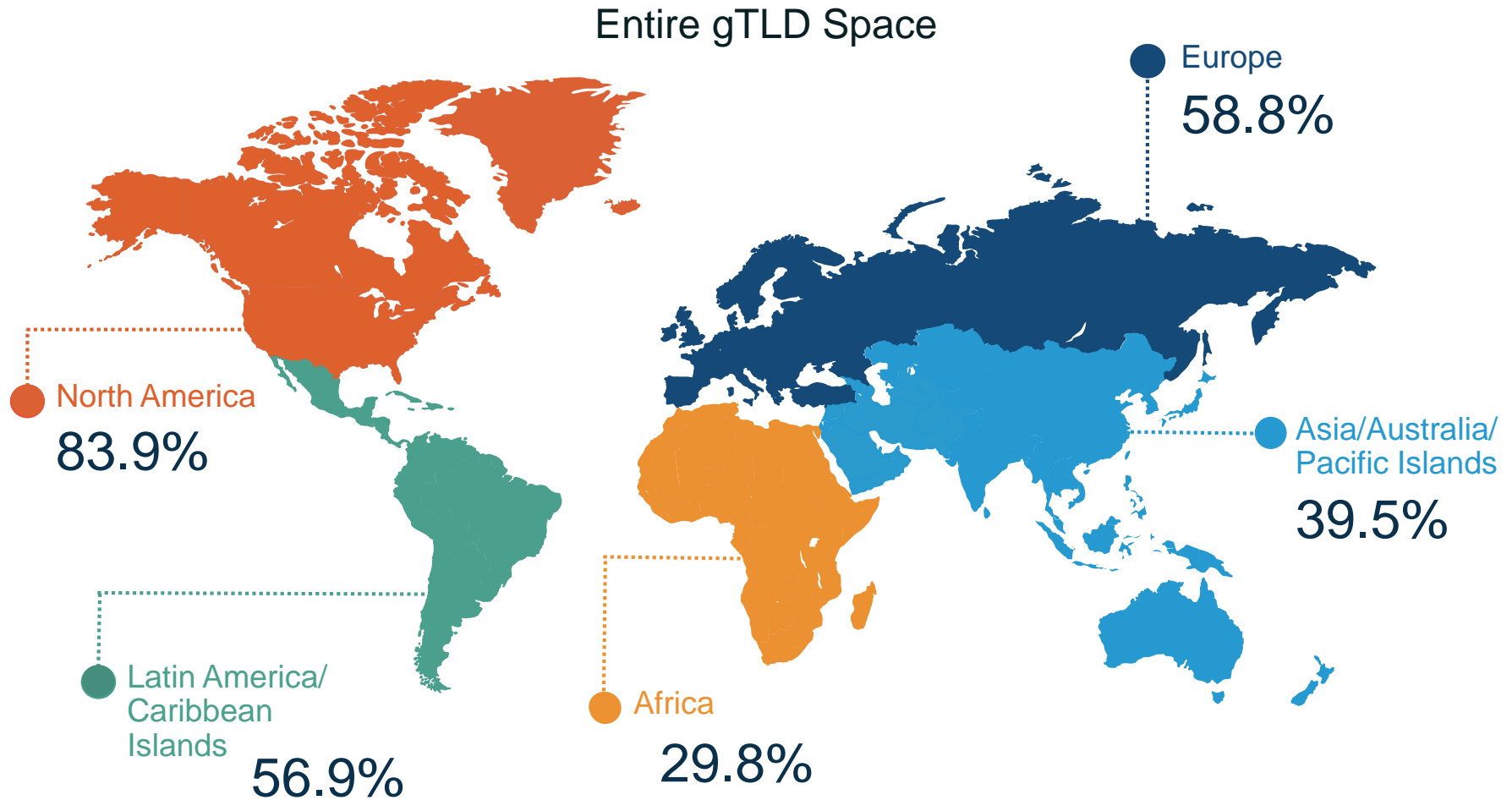
- Contact mode syntax accuracy requires accuracy on all 3 contact types
 - Registrant, Administrative, Technical Contacts
- Overall syntax accuracy requires accuracy on all 3 contact modes and all 3 contact types

Phase 2 Cycle 1 – Syntax Conformance to 2009 RAA* by RAA



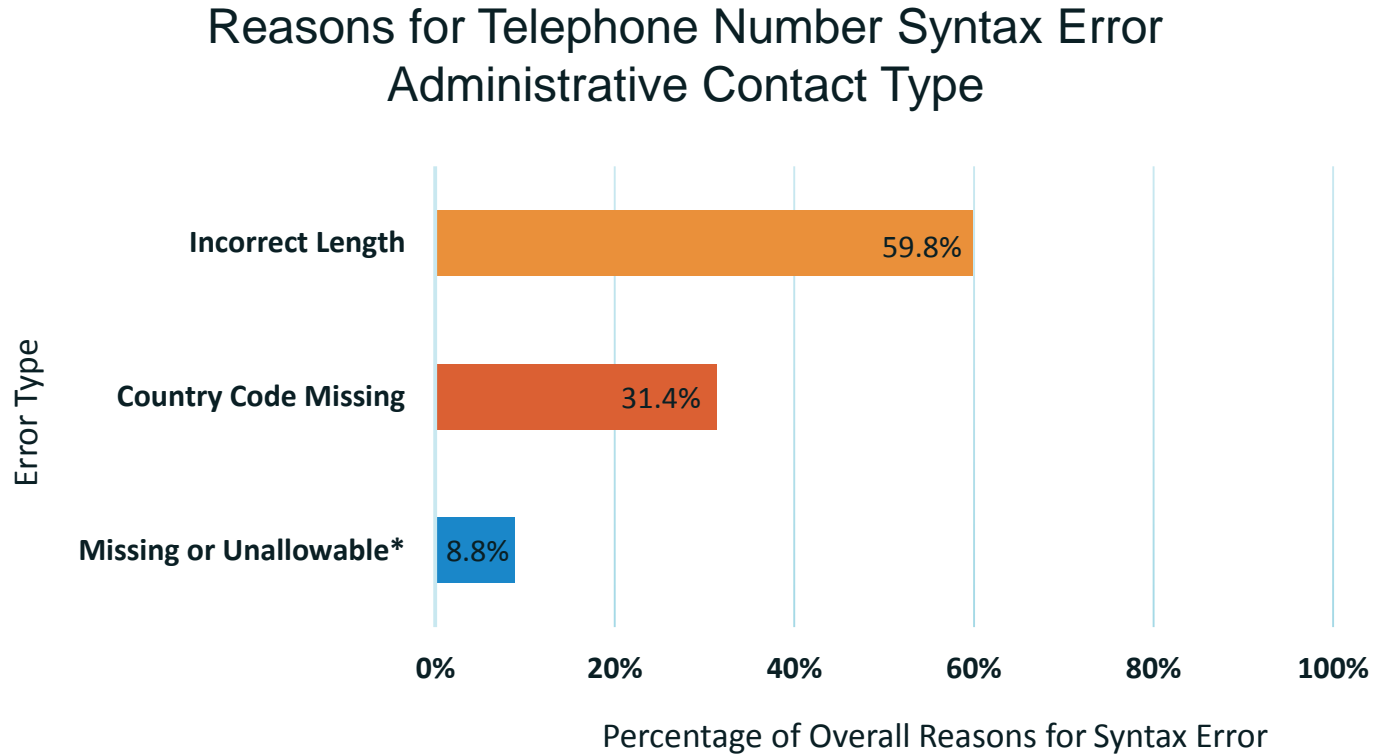
- Syntax accuracy here also requires accuracy on all 3 contact modes and all 3 contact types

Phase 2 Cycle 1 – Syntax Conformance to 2009 RAA by Region



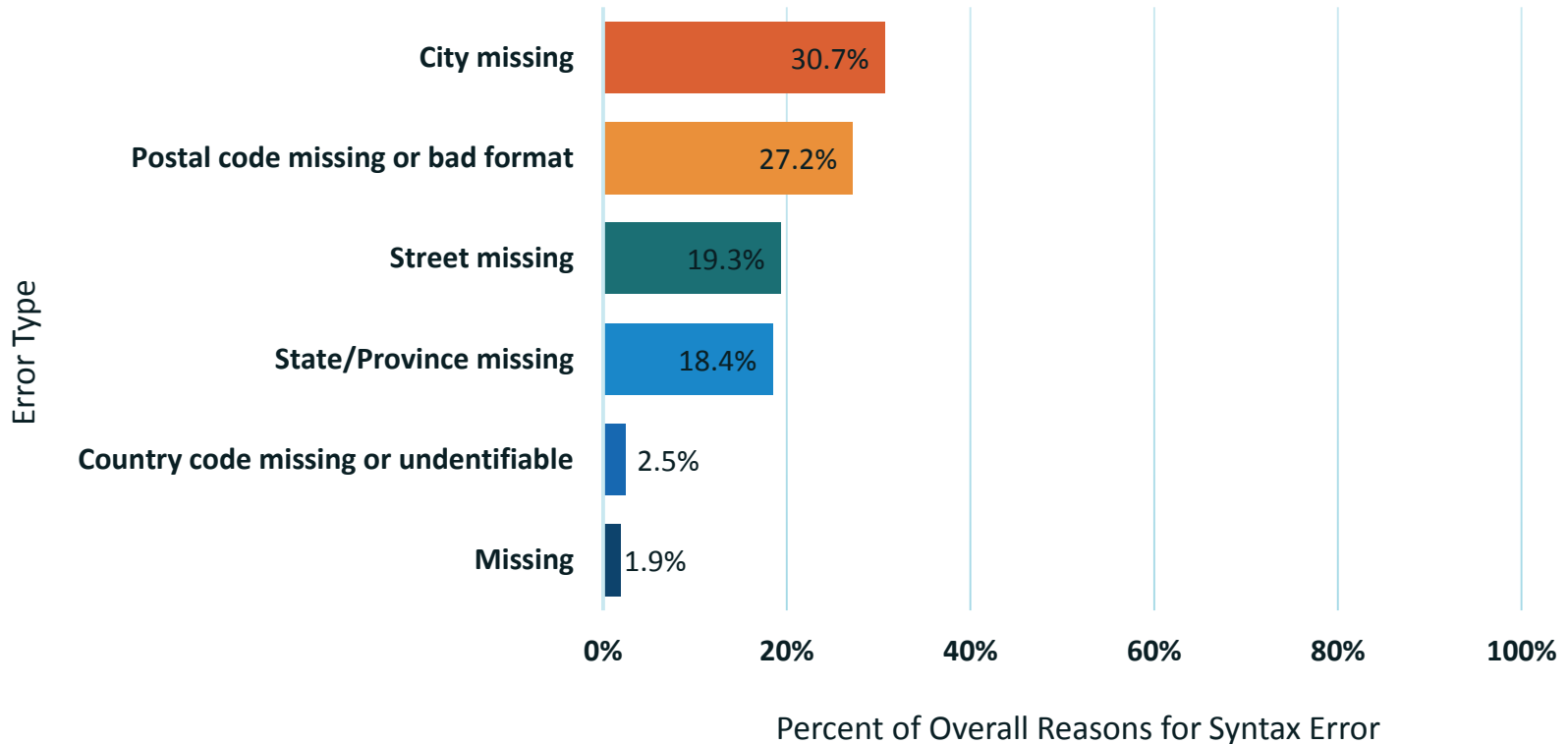
- Syntax accuracy here also requires accuracy on all 3 contact modes and all 3 contact types

Phase 2 Cycle 1 – Reasons for Telephone Syntax Error (2009)



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.

Reasons for Postal Address Syntax Error Administrative Contact Type



Phase 2

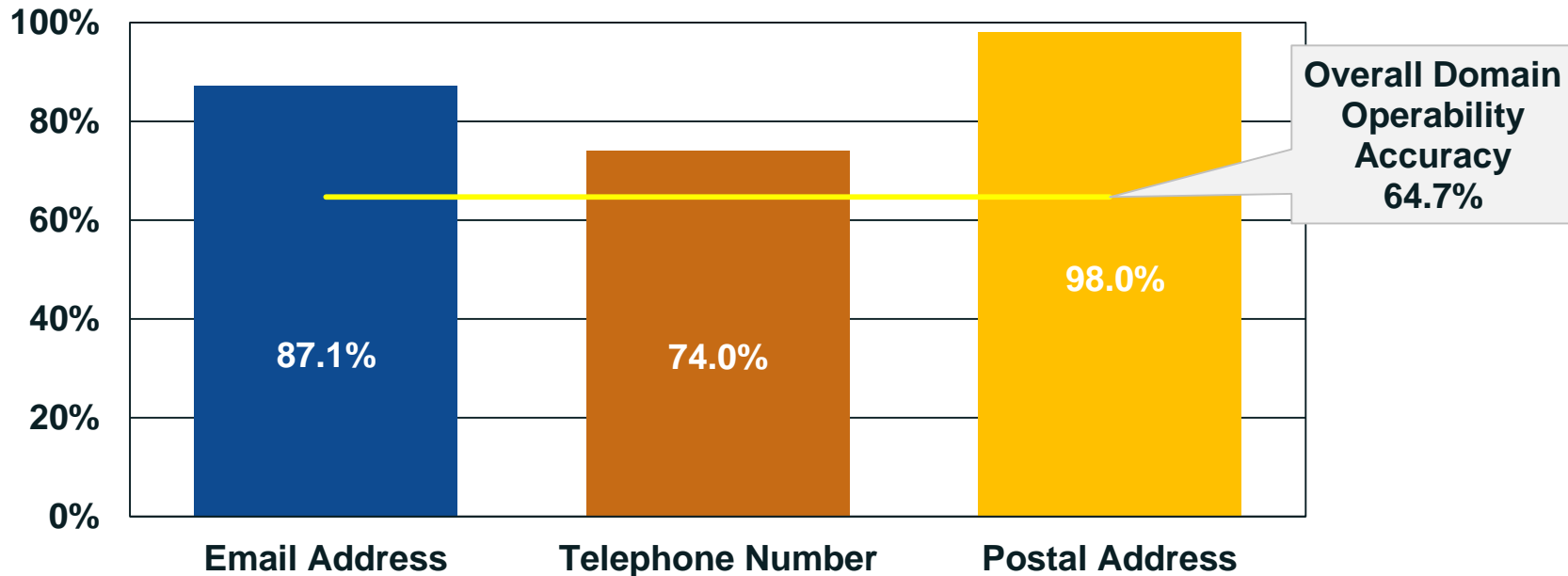
Results and Major Findings: Operability

Phase 2 Cycle 1 – Summary of Findings for Operability

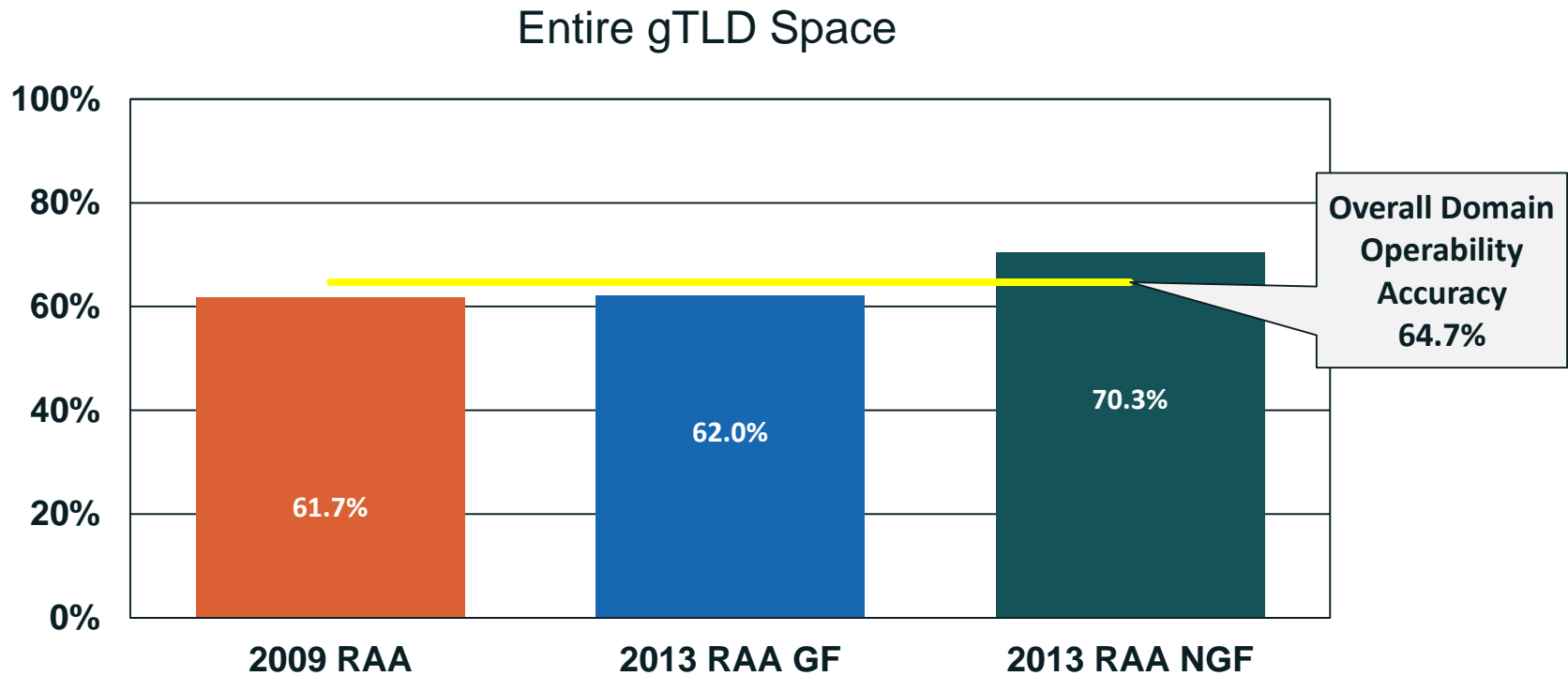
- ⦿ 65% of domains passed all operability tests for all contact types and contact modes.
- ⦿ 87% of email, 74% of telephone and 98% of postal met all operability requirements of the 2009 RAA.
 - Of those email addresses that failed operability, 10 percent bounced.
 - Of the telephone numbers that were 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

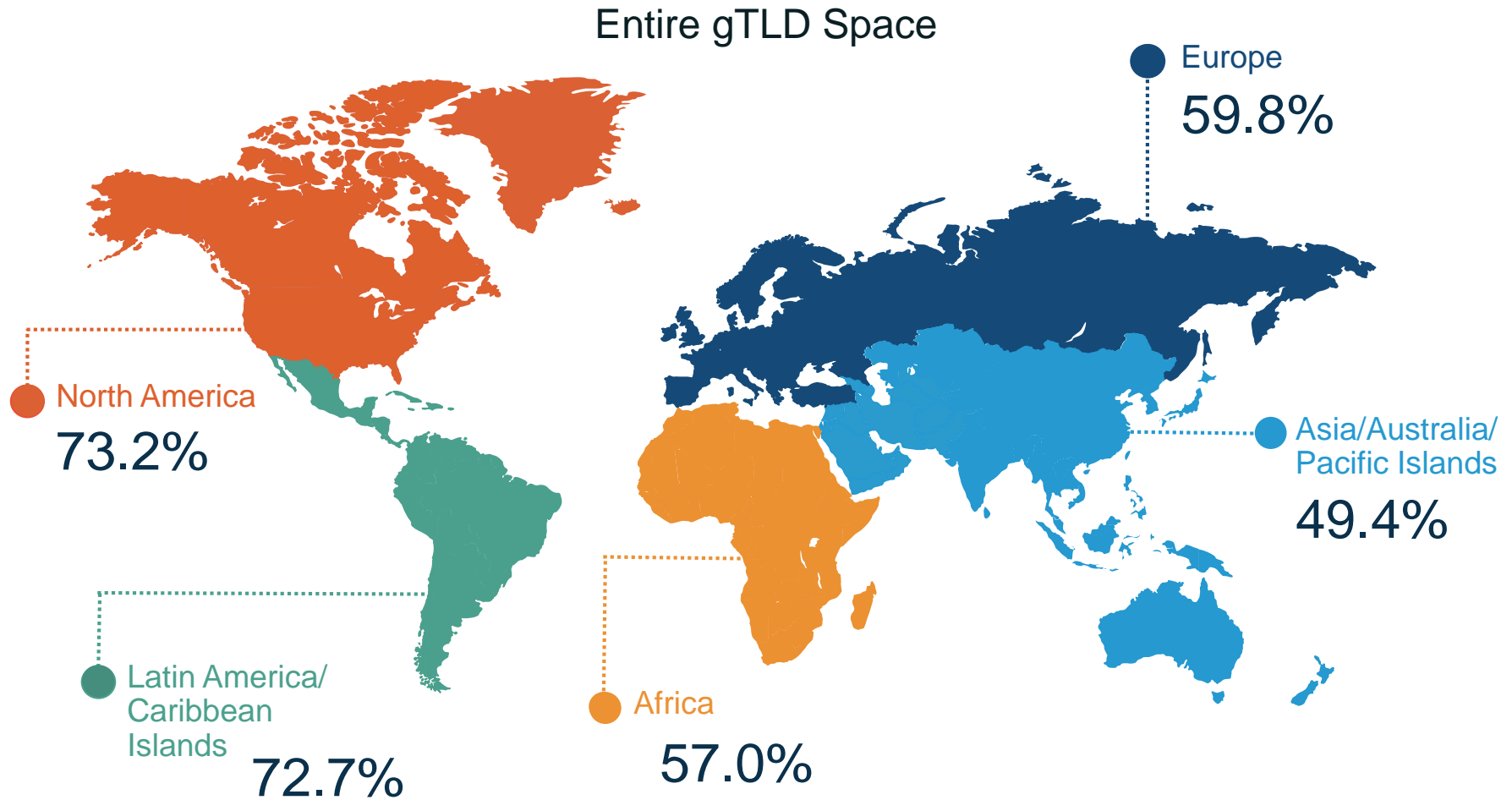


- Contact mode operability accuracy requires accuracy on all 3 contact types
 - Registrant, Administrative, Technical Contacts
- Overall operability accuracy requires accuracy on all 3 contact modes and all 3 contact types

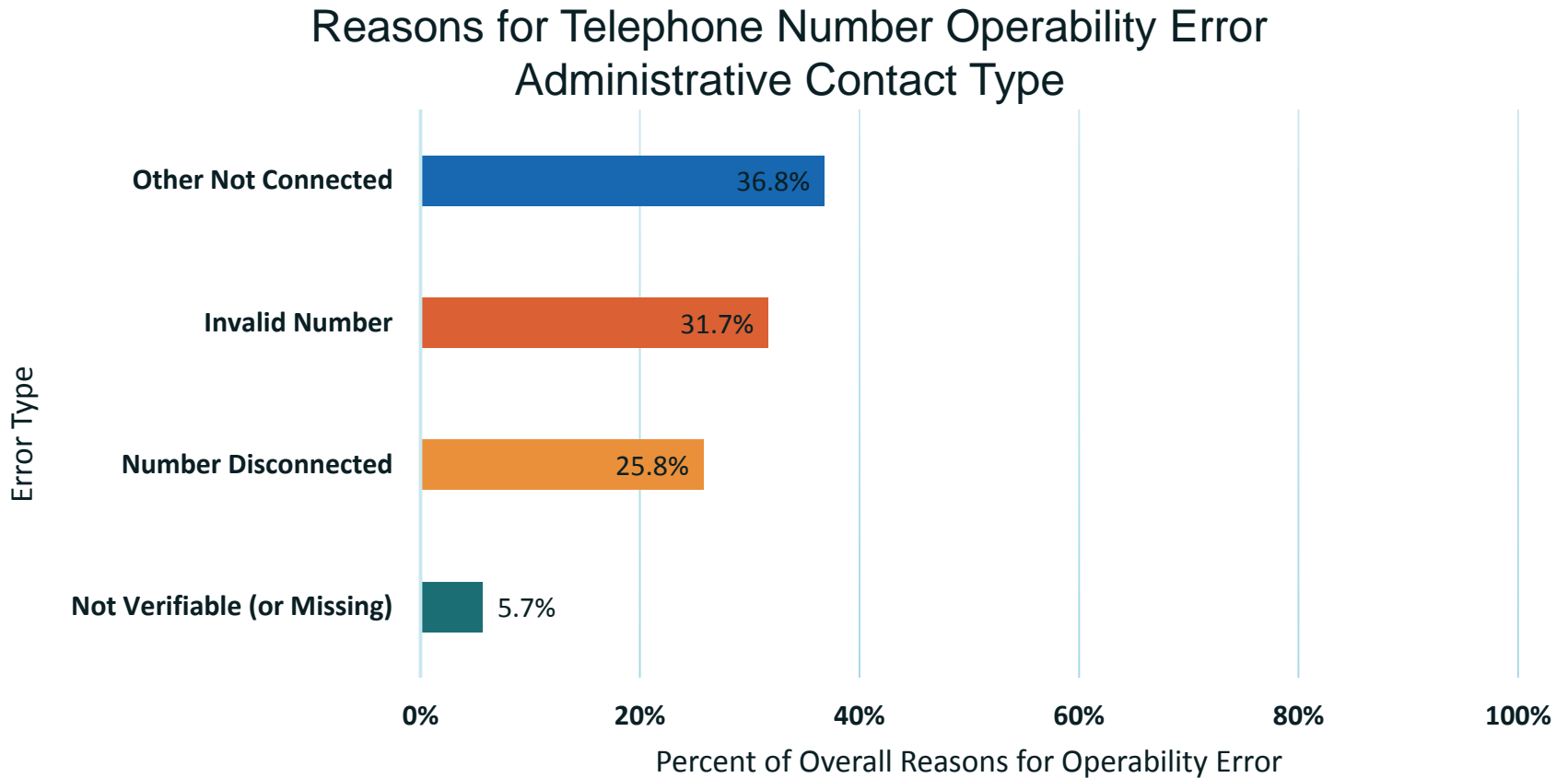


- Operability accuracy here also requires accuracy on all 3 contact modes and all 3 contact types

Phase 2 Cycle 1 – Operability Conformance to 2009 RAA by Region

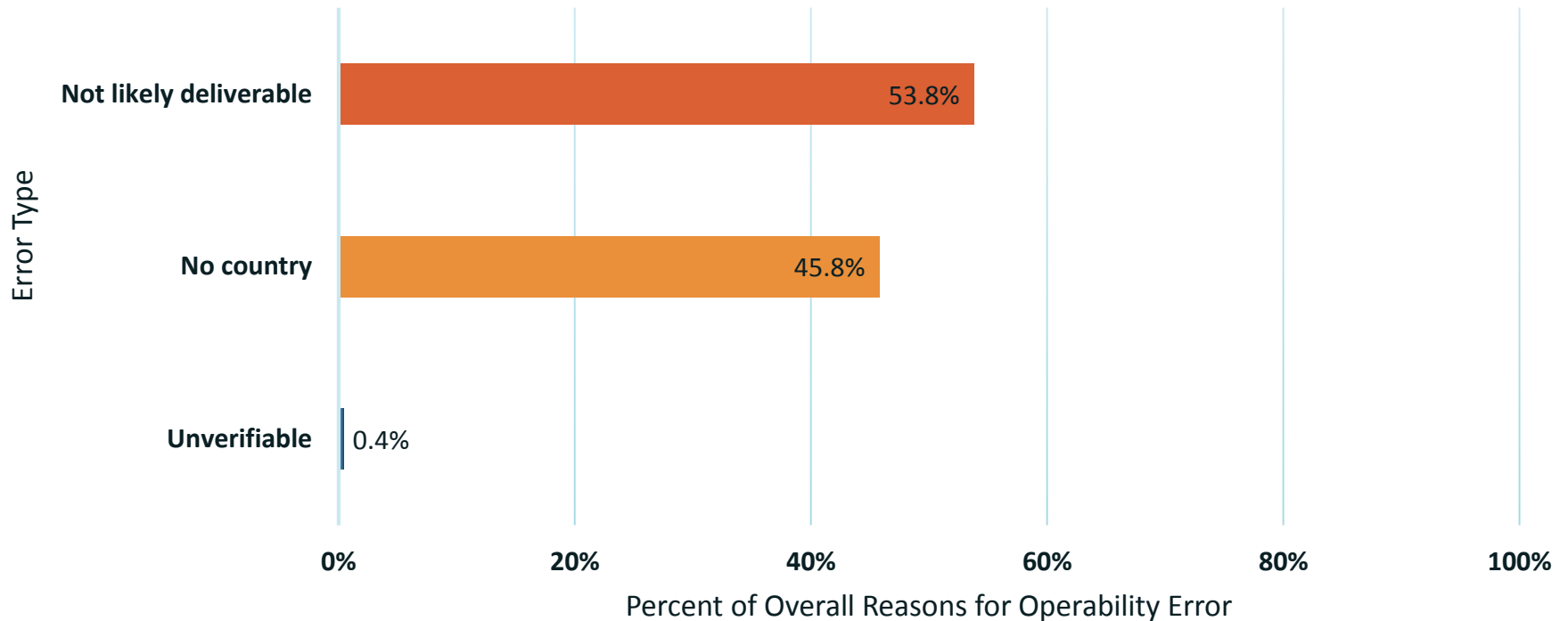


- Operability accuracy here also requires accuracy on all 3 contact modes and all 3 contact types



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.

Reasons for Postal Address Operability Error Administrative Contact Type





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

Phase 2 Cycle 1 – ICANN Contractual Compliance

- ⊙ Potentially inaccurate records have been provided to ICANN Contractual Compliance
 - ⊙ Due to sample weighting and 2013 requirements, total number of records identified as nonconforming exceeds 30% of 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**

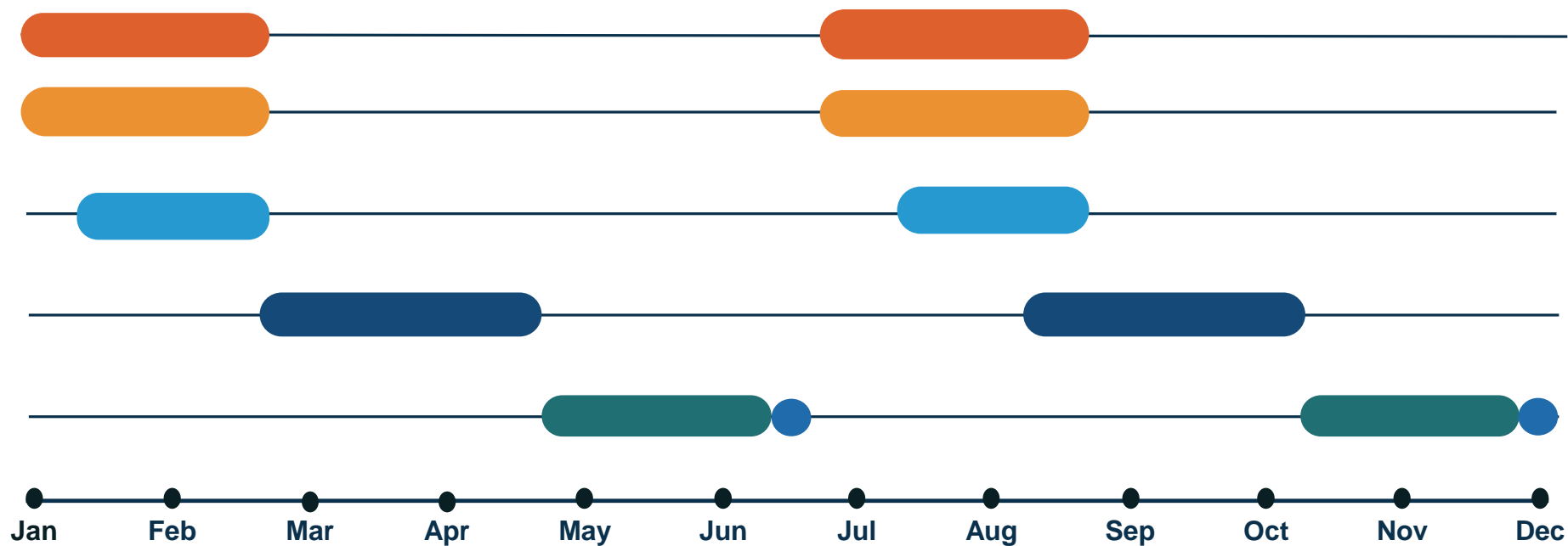
**Compliance will
conduct follow
up on potentially
inaccurate
records**

**Next Cycle
underway;
Report expected
June 2016**

Phase 2 Cycle 2 – Implementation Underway

Cycle 2: In progress

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

- 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

Questions & Answers