



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

WHOIS ARS Implementation

0

Pilot

“Proof of Concept”: Tested processes for data collection and validation

Report: Published 23 December 2014

Public Comment Report: Published 3 April 2015

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: Published 23 December 2015

Cycle 2 Report: Published 8 June 2016

Cycle 3 Report: Expected December 2016

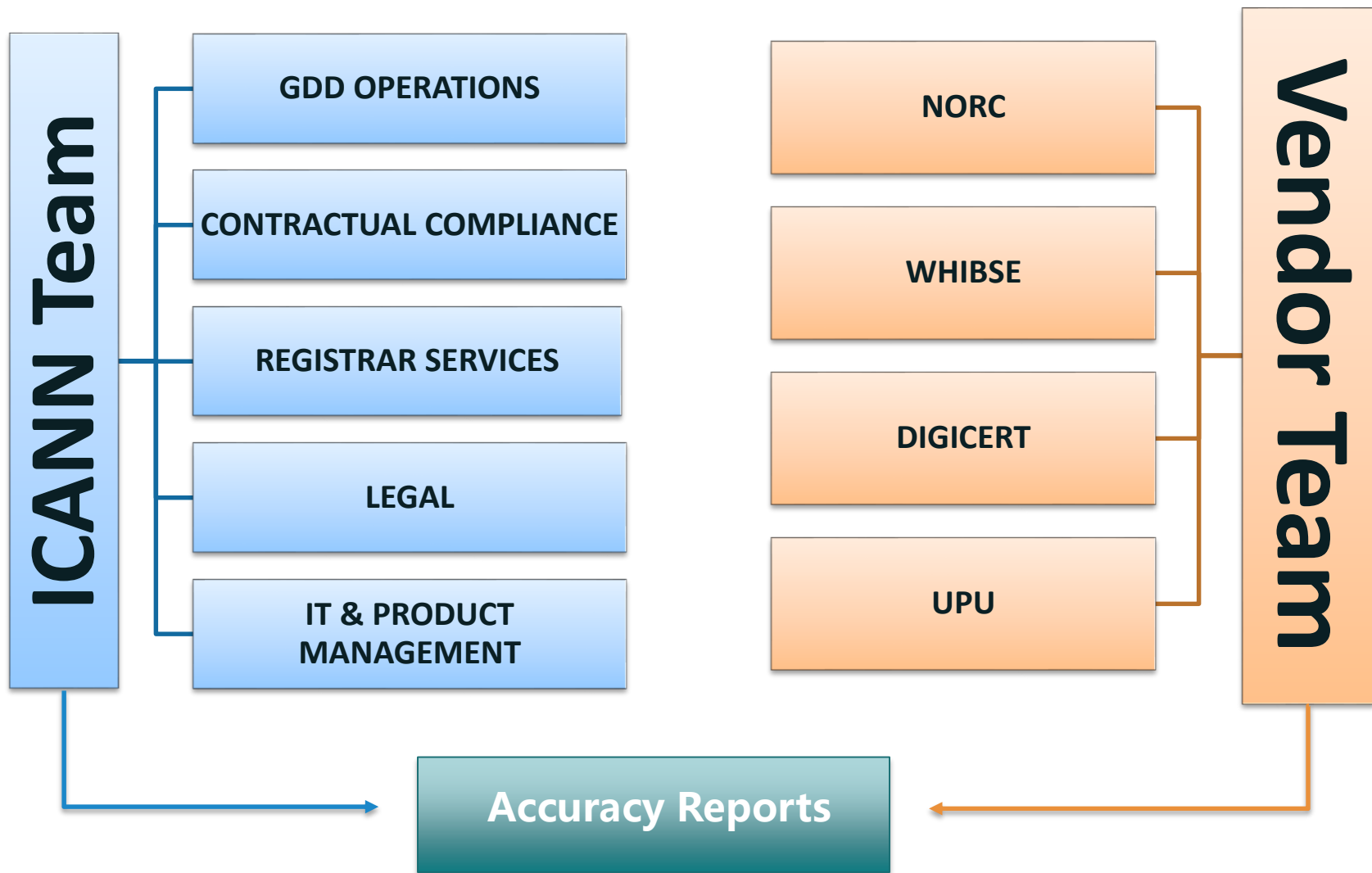
3

Phase 3 TBD, if at all: *Identity Validation*

Is the contacted individual responsible for the domain?

Phase 2 Cycle 2 Process and Timeline

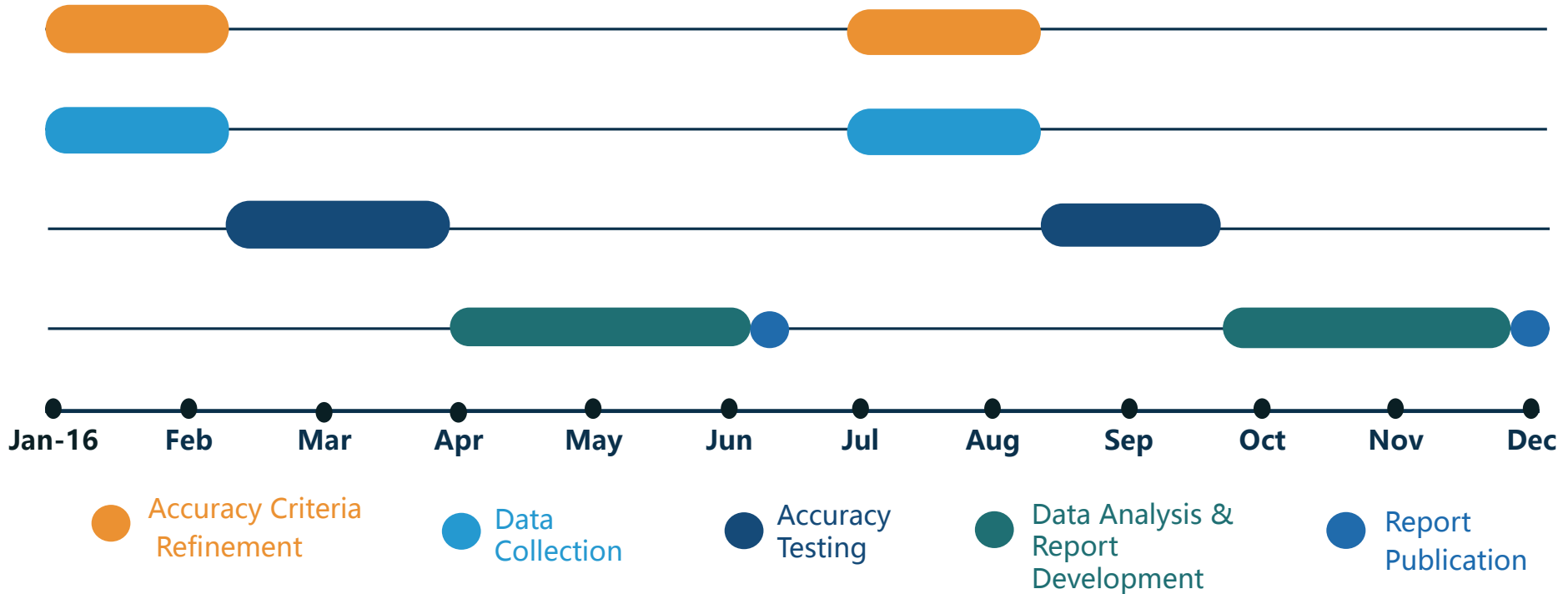
Phase 2 Cycle 2 – Cross-Functional Team



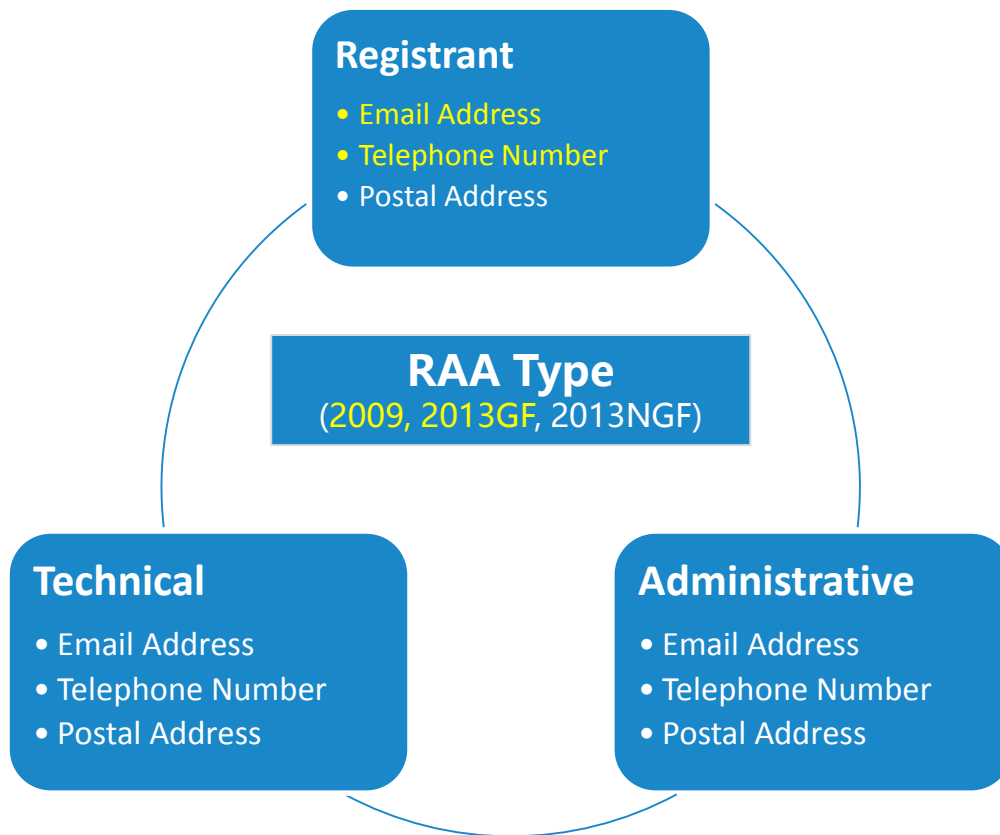
Phase 2 Cycle 2 – Cyclical Timeline of Phase 2

Cycle 2: Complete

Cycle 3: Begin July 2016



Phase 2 Cycle 2 – Contact types, modes, and testing criteria



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

Phase 2 Cycle 2 – 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 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)

Records in gTLDs	Total gTLDs	2009 RAA*	2013GF RAA*	2013 NGF RAA*	New gTLDs	Prior gTLDs
169.5m	888	4.9m	88.8m	75.8m	870	18

200k Sample

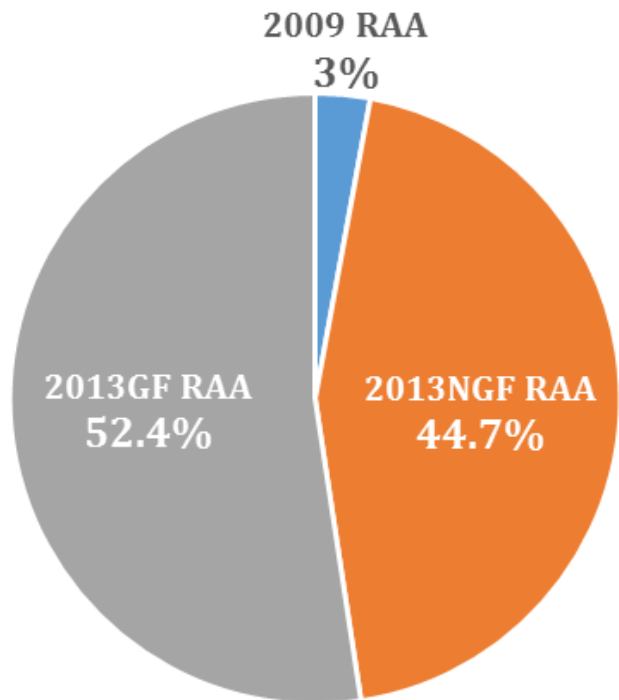
AFR	LAC	EUR	APAC	N.A.	2009 RAA	2013GF RAA	2013 NGF RAA	New gTLDs	Prior gTLDs
1.3k	8.6k	36.4k	62.8k	85.9k	4.7k	82.3k	109.2k	570	18

12k Sub-sample

AFR	LAC	EUR	APAC	N.A.	2009 RAA	2013GF RAA	2013 NGF RAA	New gTLDs	Prior gTLDs
1.3k	1.9k	2.6k	2.9k	3.3k	2.2k	4.7k	5.1k	570	18

* Weighted estimates from 200k sample

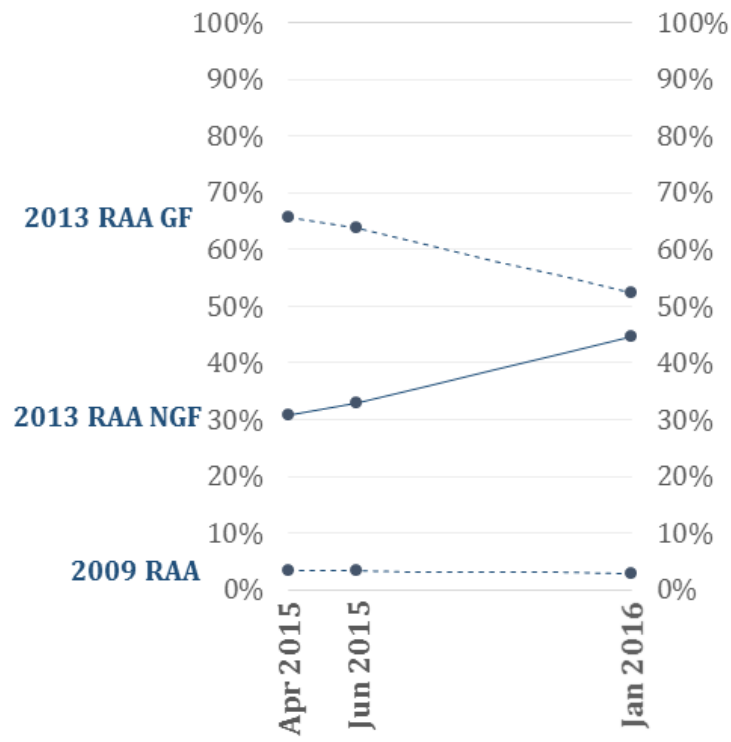
Phase 2 Cycle 2 – Distribution of RAA type



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.

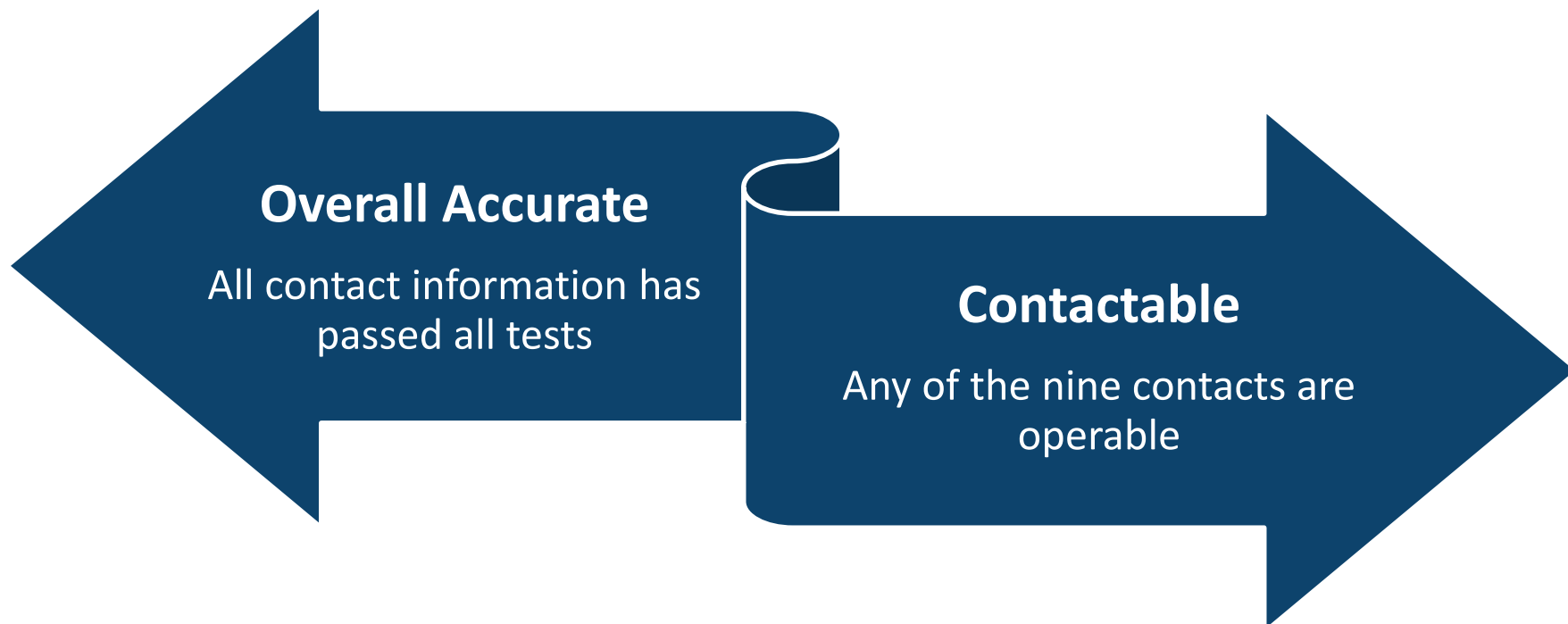
Change in Distribution across Sample Dates



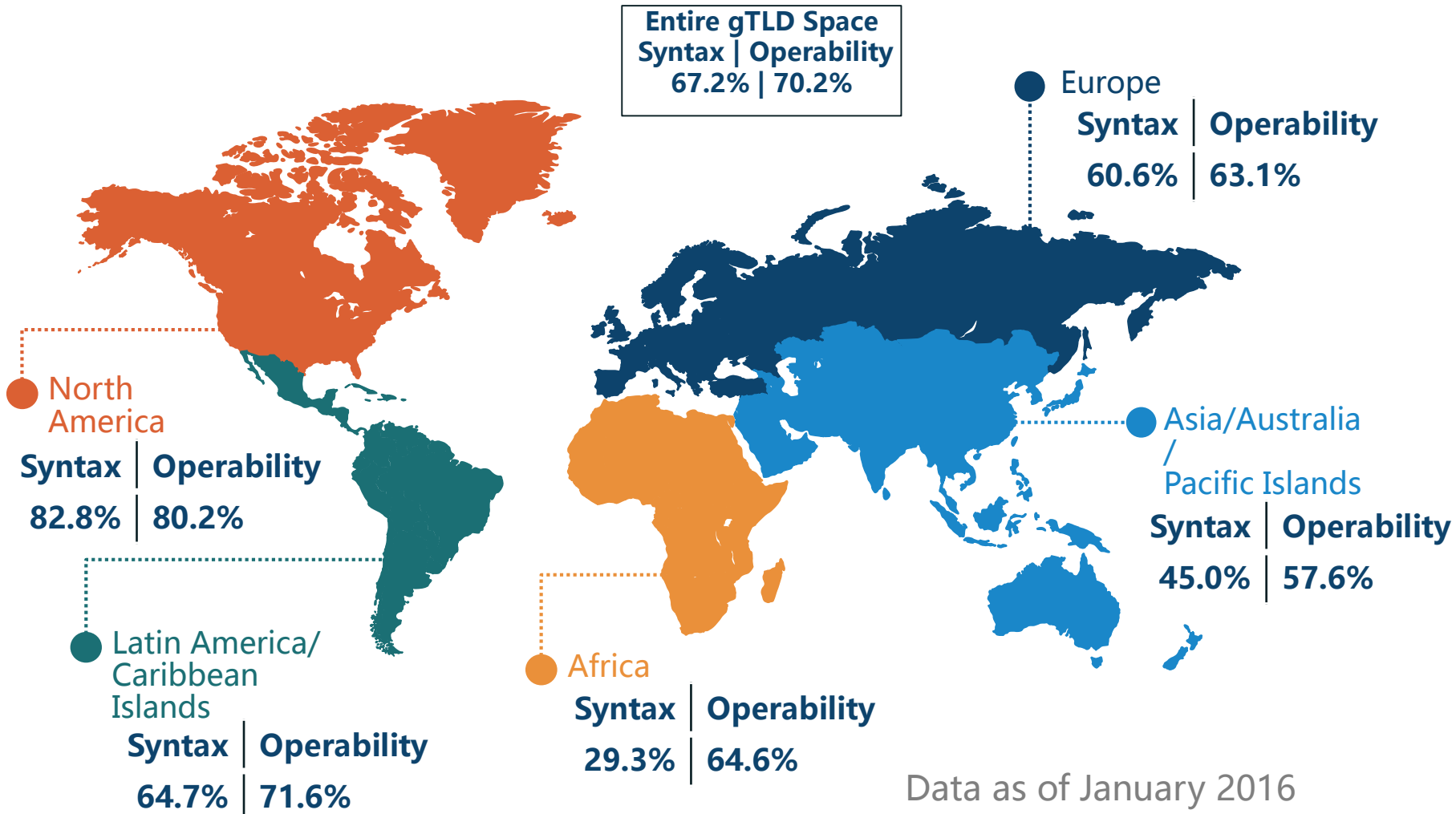
— solid line denotes increase
----- dotted line denotes decrease

Phase 2 Cycle 2 Overall Results

Phase 2 Cycle 2 – Accurate and Contactable Definitions



Phase 2 Cycle 2 – Overall Syntax and Operability Accuracy by Region



Data as of January 2016
Categorized by ICANN
region



**99% of records
contactable by at
least one method**



**Only 1% of records
were entirely non-
contactable**

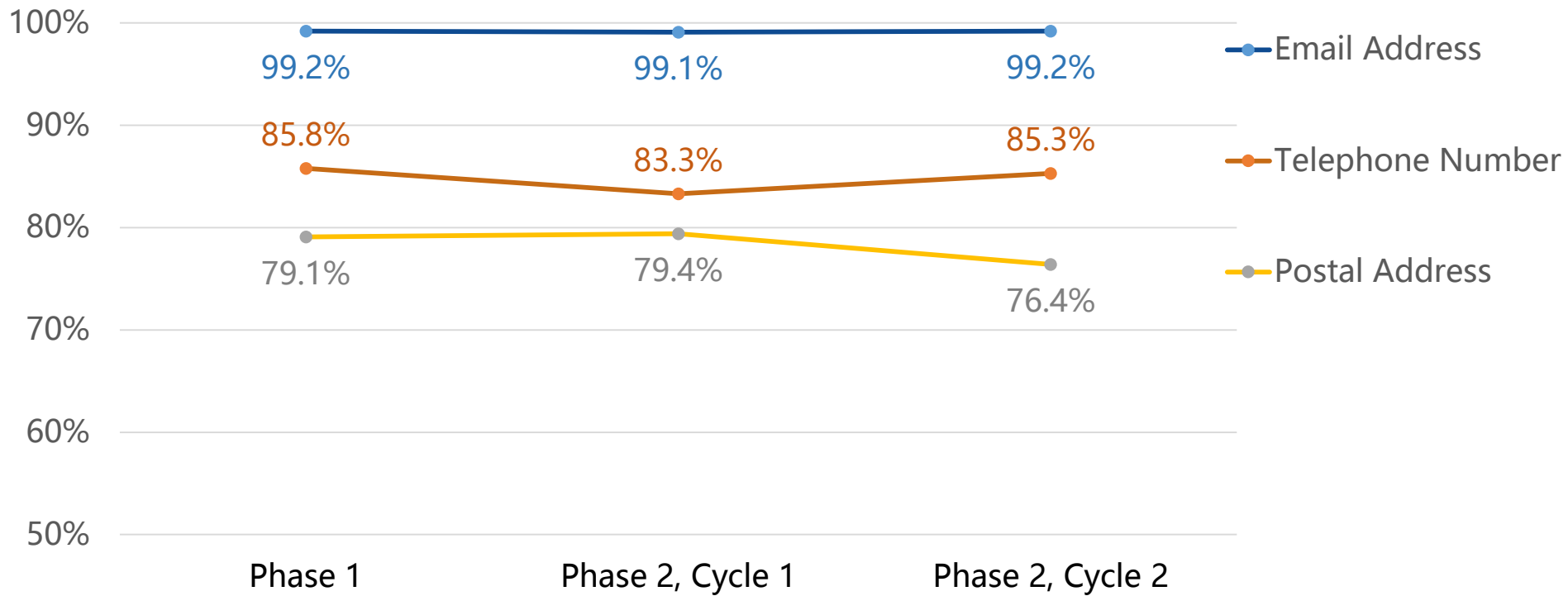
Phase 2 Cycle 2

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

Overall Syn Accuracy
Cycle 1 | Cycle 2
67.9% | 67.2%
Δ -0.7%

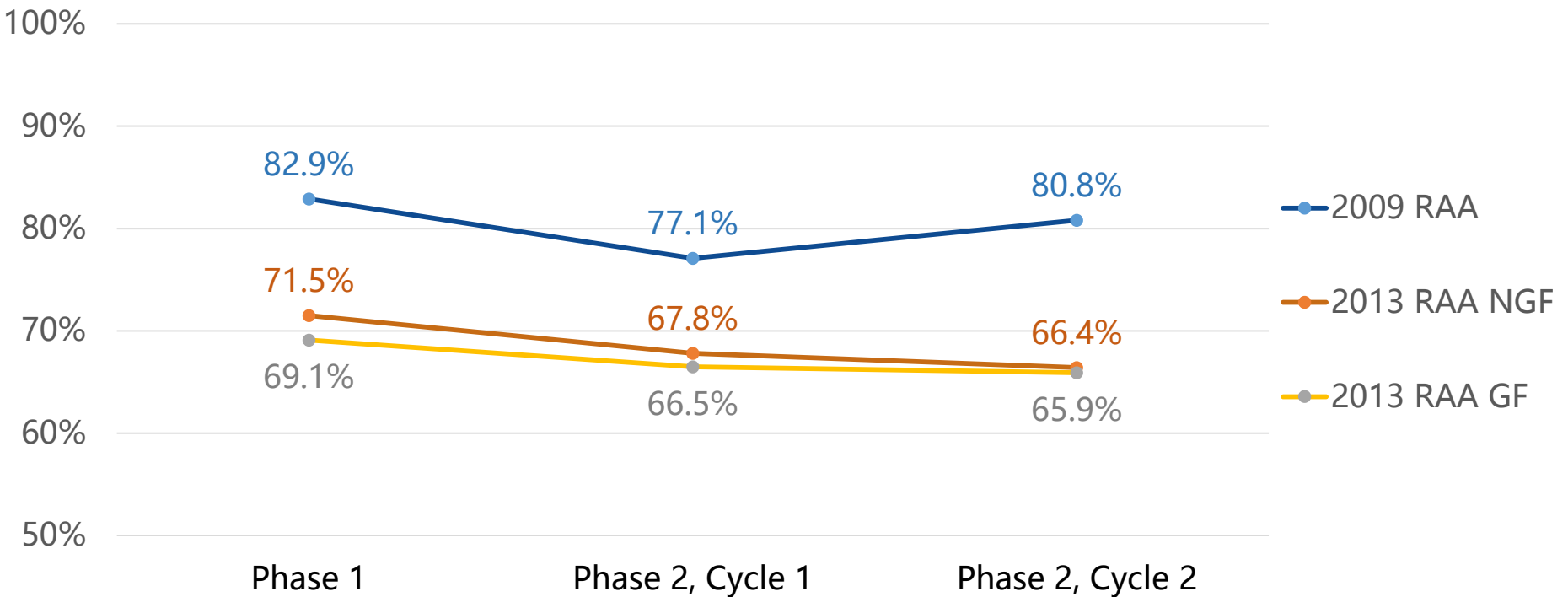


- Accuracy here also requires accuracy on all 3 contact modes and all 3 contact types

Phase 2 Cycle 2 – Changes Over Time: Syntax Accuracy by RAA Type

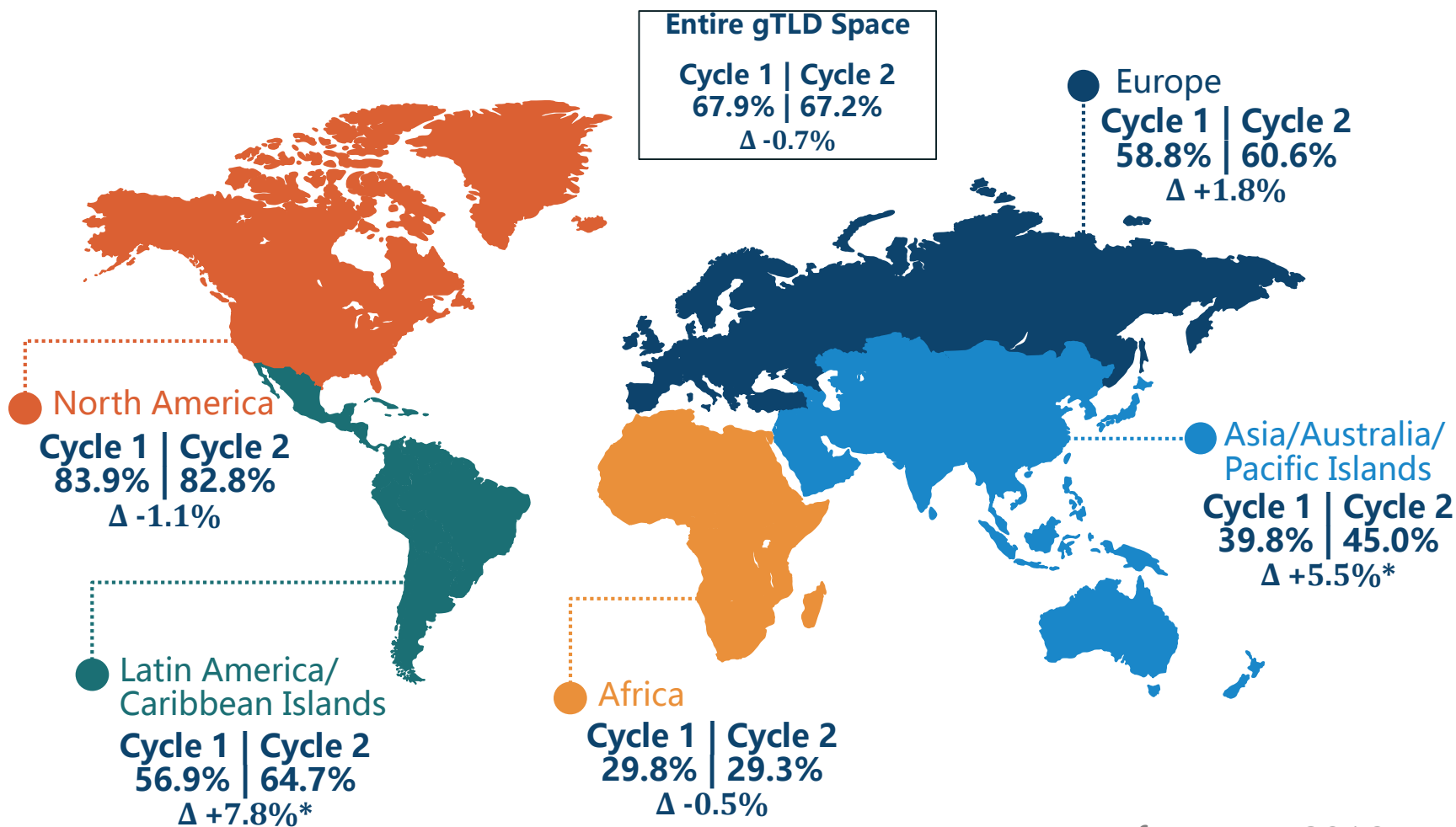
Overall Syn Accuracy
Cycle 1 | Cycle 2
67.9% | 67.2%
Δ -0.7%

Entire gTLD Space
Phase 1 through Phase 2 Cycle 2



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

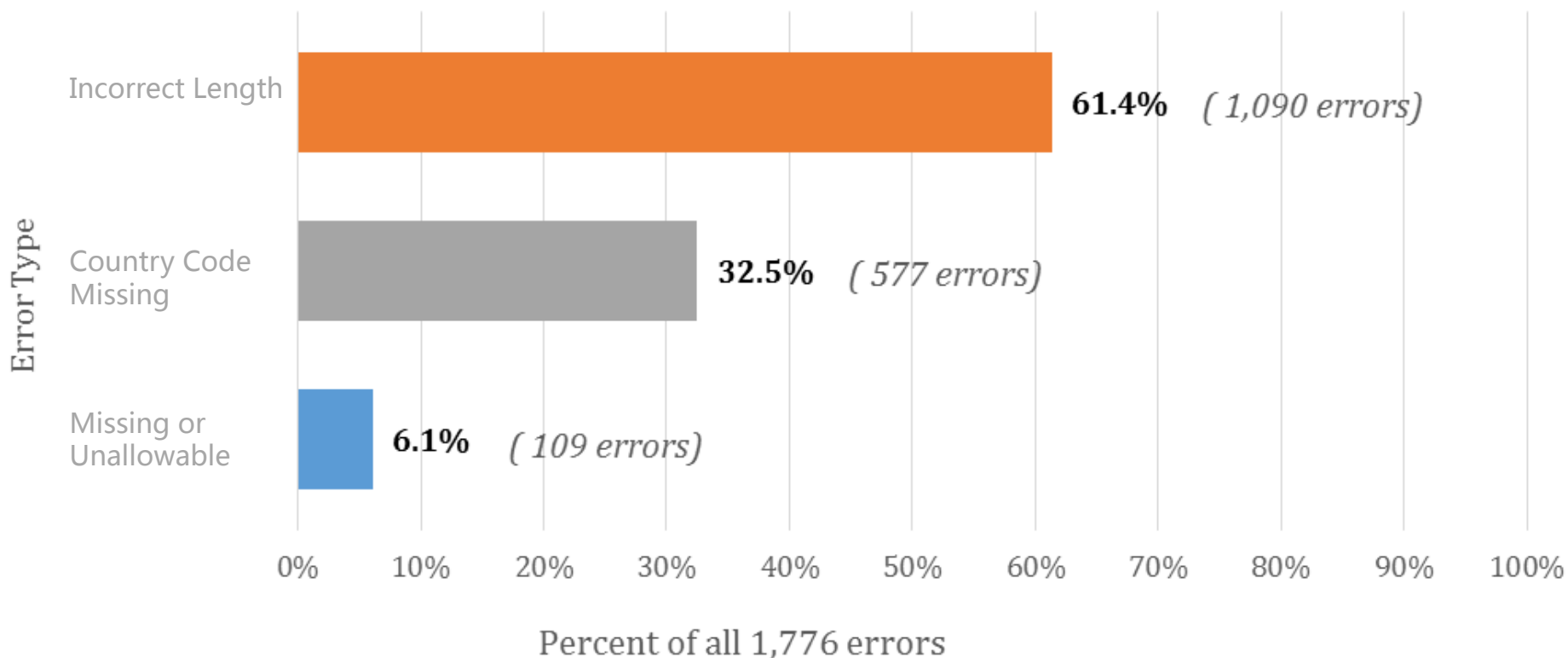
Phase 2 Cycle 2 – Overall Syntax Accuracy by Region



Data as of January 2016
Categorized by ICANN
region

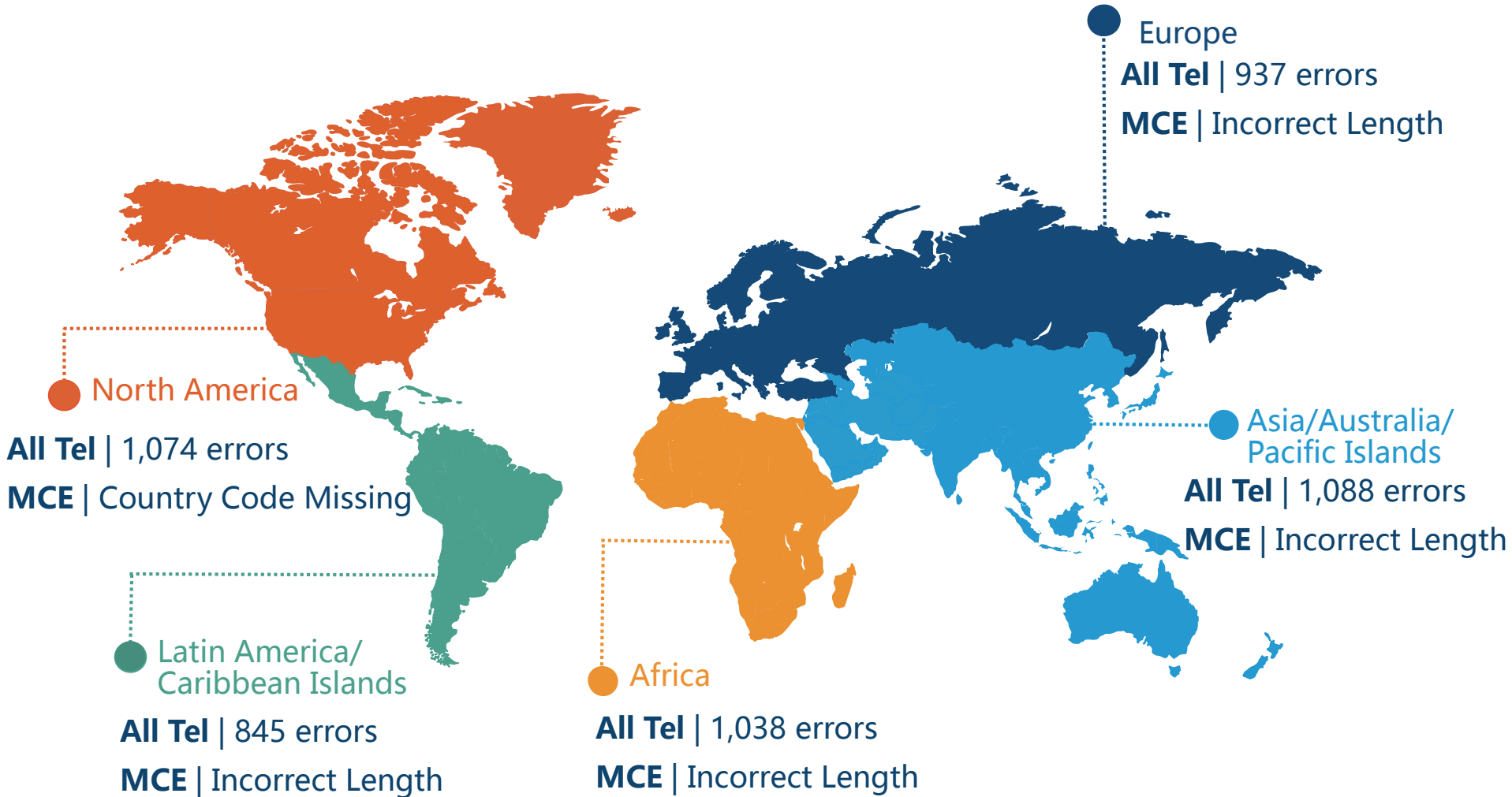
Phase 2 Cycle 2 – Reasons for Telephone Syntax Error

Reasons for Telephone Number Syntax Error Administrative Contact Type 1,776 total 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. 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.

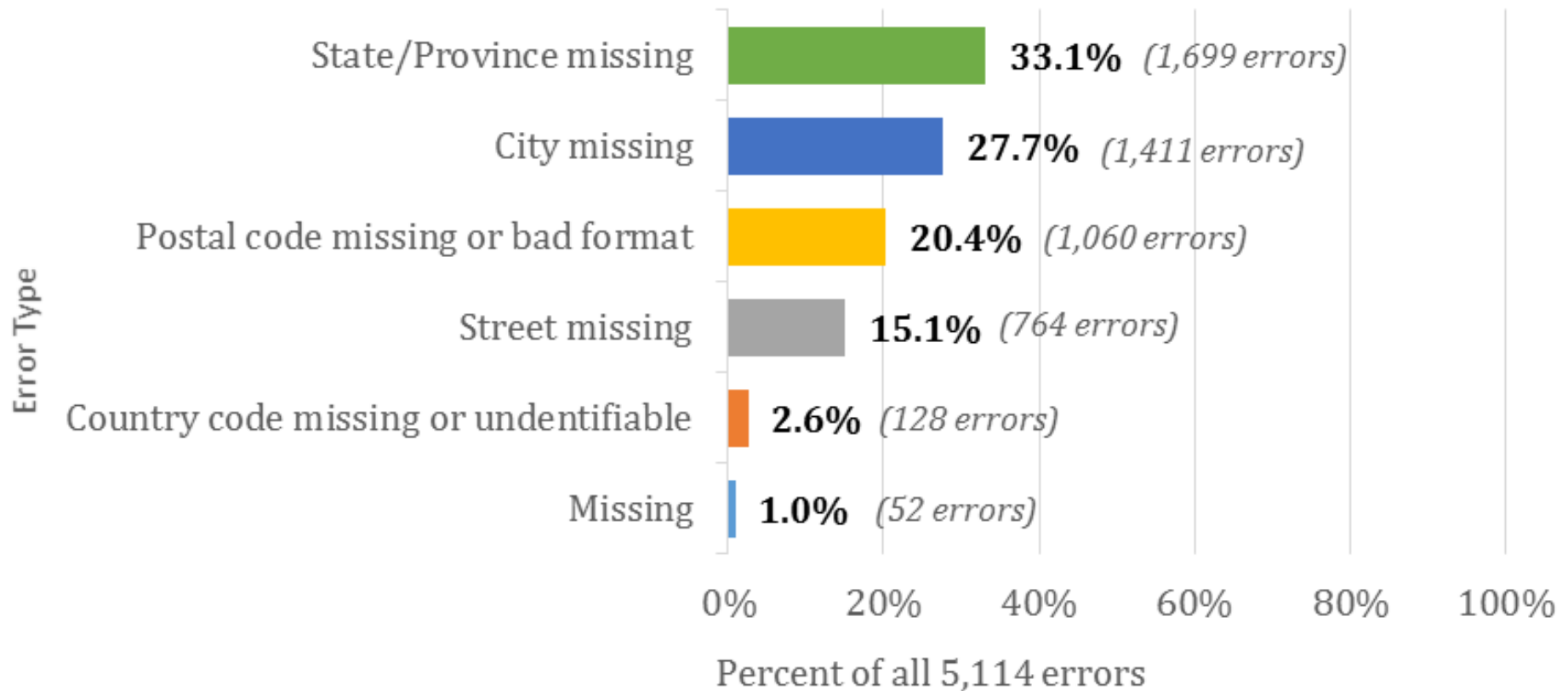
Phase 2 Cycle 2 – Telephone Syntax Error by Region



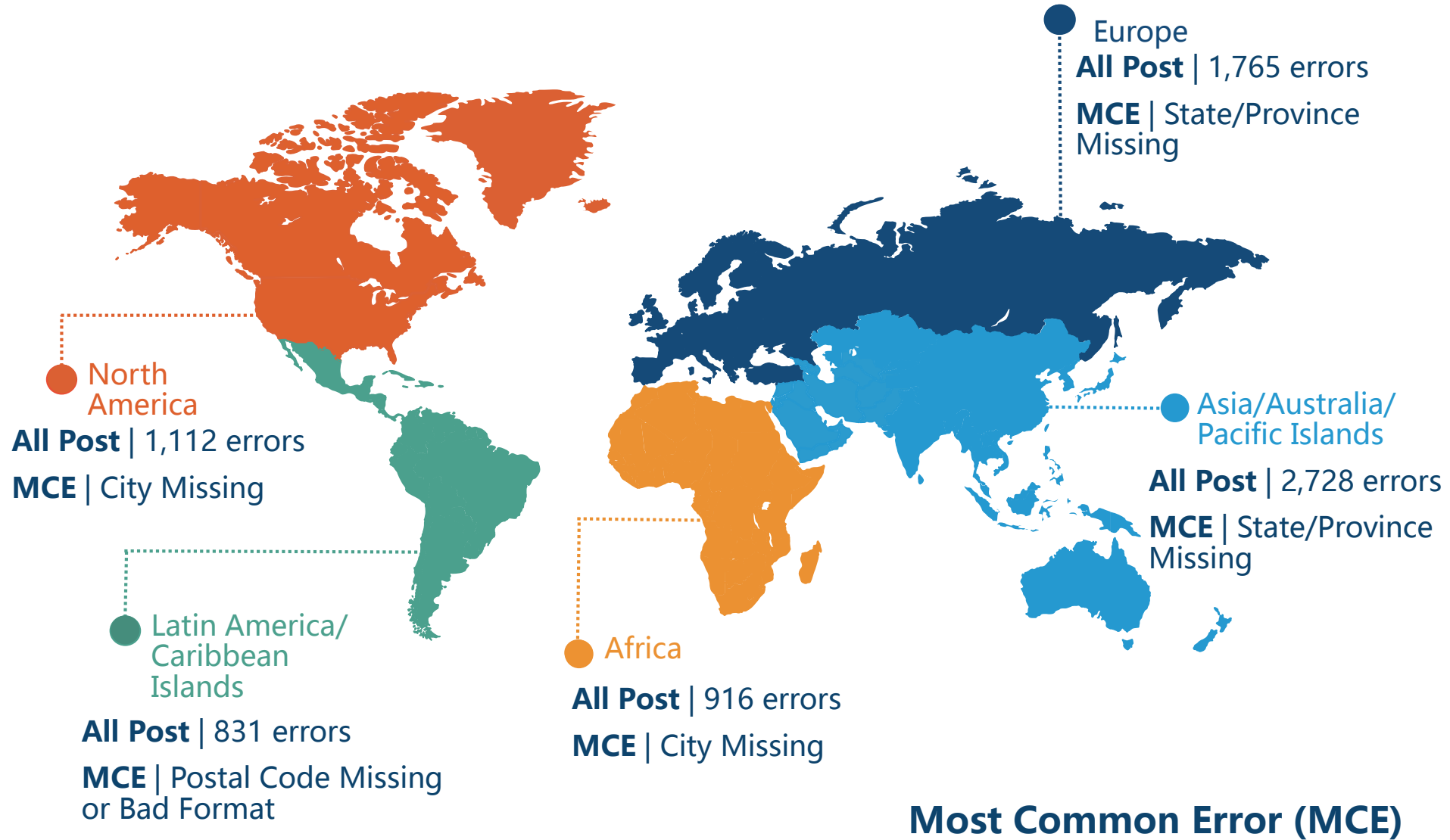
Most Common Error (MCE)

Phase 2 Cycle 2 – Reasons for Postal Syntax Error

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



Phase 2 Cycle 2 – Postal Syntax Error by Region



Phase 2 Cycle 2

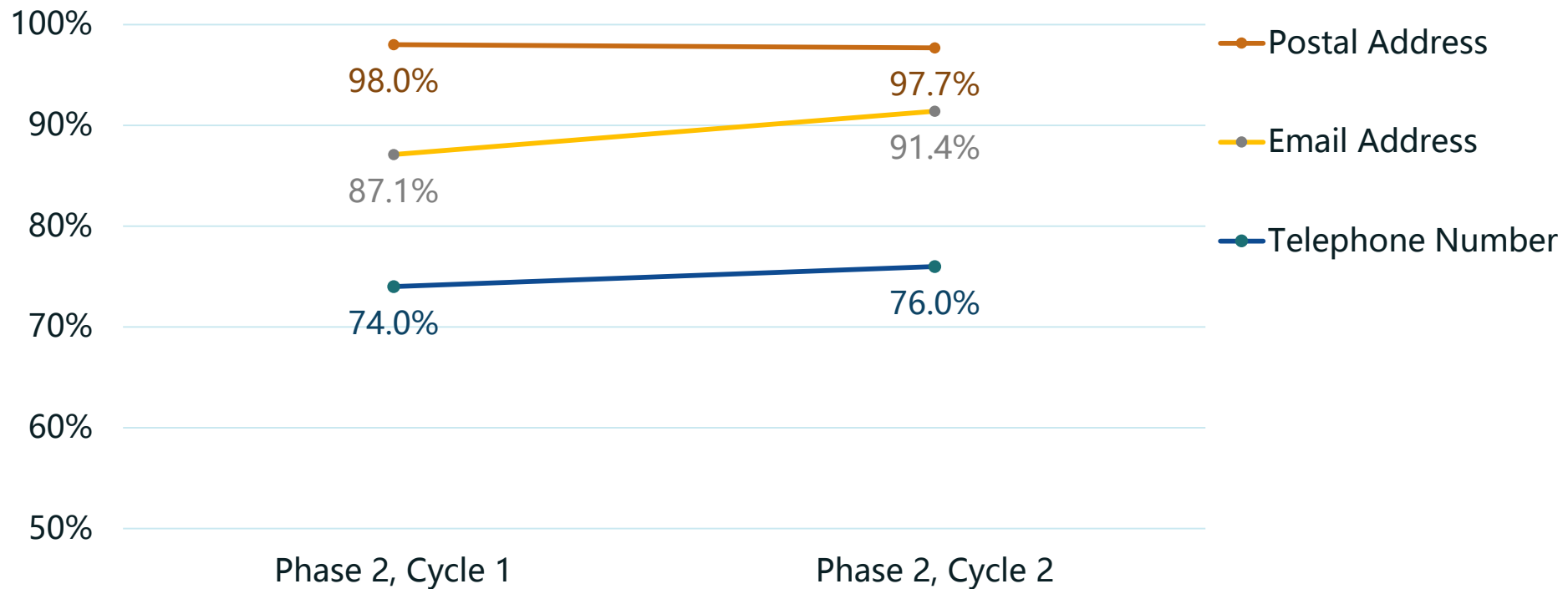
Results and Major Findings: Operability, 2009 RAA

Phase 2 Cycle 2 – Changes Over Time: Operability Accuracy by Contact Mode

Overall Op Accuracy

Cycle 1 | Cycle 2
64.7% | 70.2%
Δ +5.4%

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

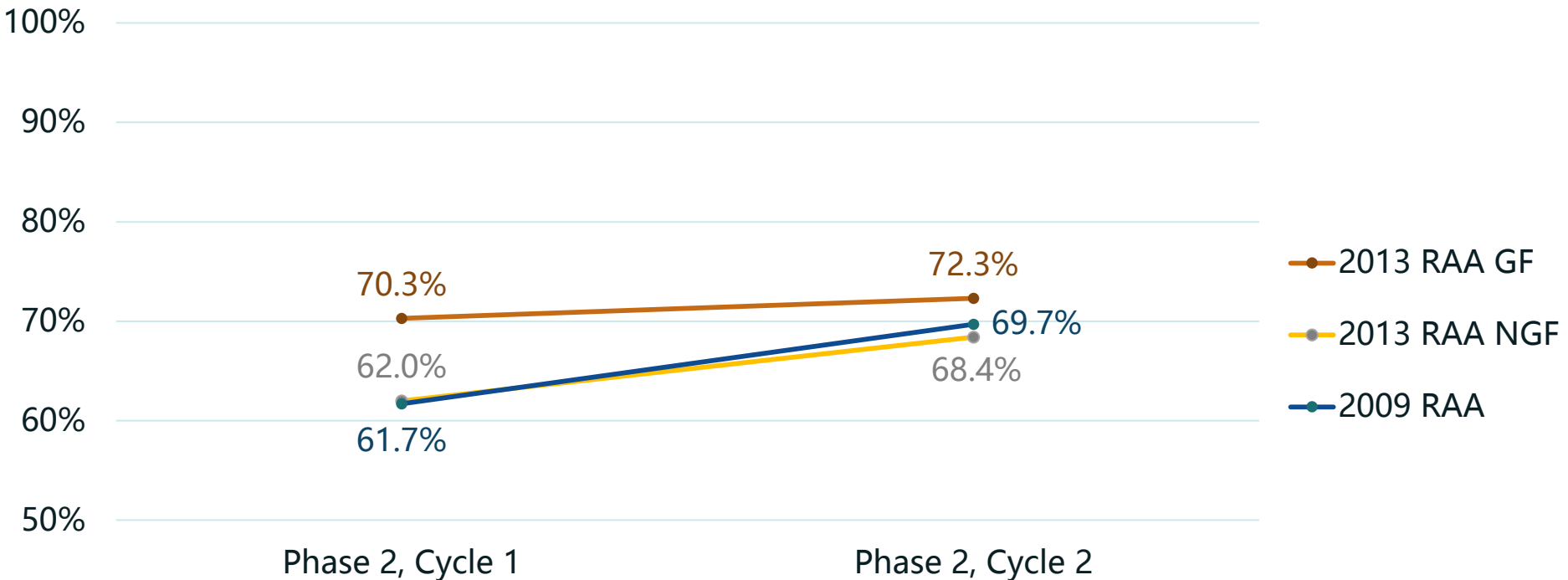


- **Saw mostly improvement from Cycle 1 to Cycle 2**
- Accuracy here also requires accuracy on all 3 contact modes and all 3 contact types

Phase 2 Cycle 2 – Changes Over Time: Operability Accuracy by RAA

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

Overall Op Accuracy
Cycle 1 | Cycle 2
64.7% | 70.2%
Δ +5.4%



- **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

Entire gTLD Space

Cycle 1 | Cycle 2
64.7% | 70.2%
Δ +5.4%

● Europe
Cycle 1 | Cycle 2
59.8% | 63.1%
Δ +3.3%

● Asia/Australia/
Pacific Islands

Cycle 1 | Cycle 2
49.4% | 57.6%
Δ +8.2%

● Africa
Cycle 1 | Cycle 2
57.0% | 64.6%
Δ +7.7%

● Latin America/
Caribbean Islands

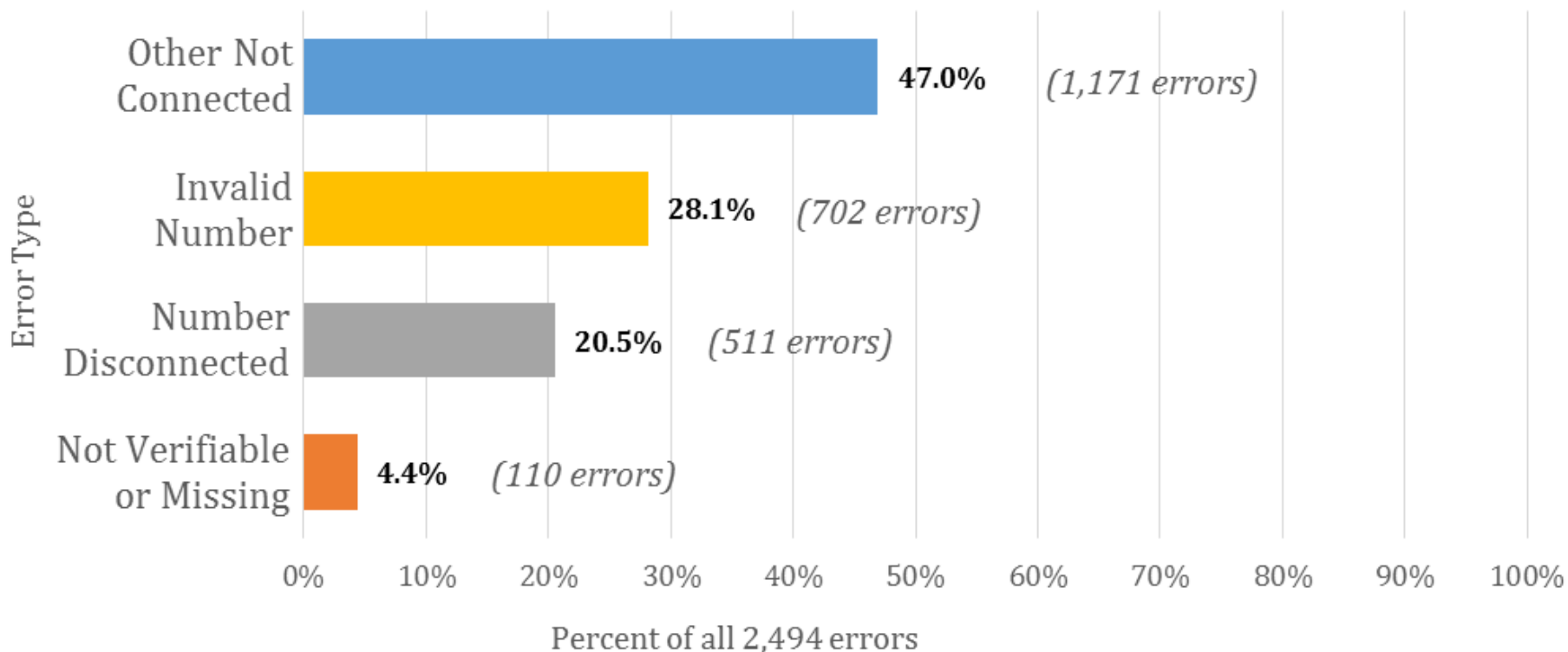
Cycle 1 | Cycle 2
72.7% | 71.6%
Δ -1.0%

● North America
Cycle 1 | Cycle 2
73.2% | 80.2%
Δ +6.9%

Data as of January 2016
Categorized by ICANN
region

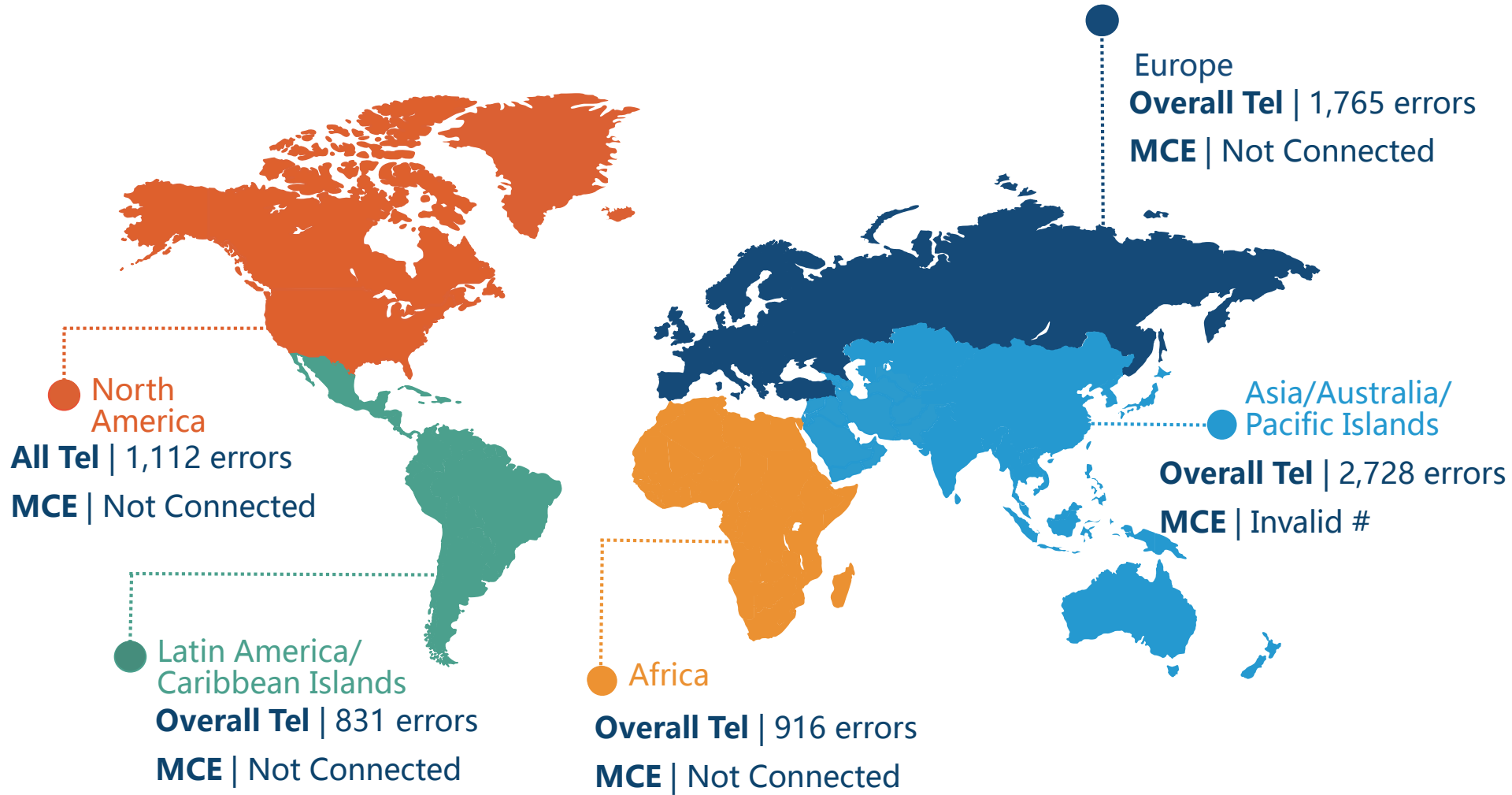
Phase 2 Cycle 2 – Reasons for Telephone Operability Error

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



Most Common Error (MCE)



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%

Phase 2 Cycle 2 – Script Analysis

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

Phase 2 Cycle 2 – ICANN Contractual Compliance

- ⦿ Potentially inaccurate 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