

ICANN Org Memo on the WHOIS Accuracy Reporting System (ARS)

Overview of Background, Issues, and Future Studies

Global Domains & Strategy
January 2022



TABLE OF CONTENTS

EXECUTIVE SUMMARY	3
BACKGROUND ON THE WHOIS ARS	3
Development of the WHOIS ARS	4
WHOIS ARS Sample Design & Testing	5
WHOIS ARS & ICANN Contractual Compliance	5
GDPR & DECISION TO PUT WHOIS ARS ON-HOLD	6
Alternative Options or Data Sources	6
ICANN org Communication to the Board regarding WHOIS ARS Hold	7
ISSUES RELATED TO CONTINUING WHOIS ARS	7
Contracts & Budget	7
Data	7
Legal Environment	8
Purpose & Goal of Measuring Accuracy	8
ICANN Contractual Compliance Processing	9
FUTURE STUDIES REGARDING REGISTRATION DATA ACCURACY	9
Potential Objectives of a Study	9
CLOSING	10

Executive Summary

This paper provides an overview of the WHOIS Accuracy Reporting System (ARS), including background on its development and implementation, impact of GDPR on the ARS, issues related to continuing the ARS, and the potential for a future study of how to measure registration data accuracy. While this paper largely is focused on providing detailed information on these topics, it also answers two initial questions from discussions within the Accuracy Scoping Team:

1. Were there any specific ideas behind the suggestion for a study as mentioned in the ICANN org briefing?
2. In relation to ARS, were any alternatives explored before the program was put on hold?

At the time of making this recommendation in the ICANN Org Briefing on Registration Data Accuracy Requirements and Programs, ICANN org did not have a specific idea as to what the study would be or look like. However, since that time, ICANN org has done some initial brainstorming as to the objectives, which are described in the section [Future Studies regarding Registration Data Accuracy](#).

Additionally, at the time of putting ARS on-hold, no alternative data sources were considered. There have been discussions and conversations regarding other options, such as using escrow data or Bulk Registration Data Access (BRDA), but these have not been thoroughly investigated as viable alternatives.

The GDPR has limited access to gTLD registration data for all parties, including ICANN org. This includes limiting ICANN org's ability to check the accuracy of gTLD registration data. ICANN org has engaged with the European Commission to suggest that the proposed NIS2 legislation could recognize ICANN's legitimate interest in having access to non-public gTLD registration data for the purpose of checking accuracy. Because of the GDPR, however, it is not clear that ICANN org has the ability to access gTLD registration data.

ICANN org understands that the Accuracy Scoping Team may have additional questions and will continue to engage with the Scoping Team to provide any relevant information, as applicable. Should there be a request for an additional study (or studies) regarding accuracy of registration data and/or how to measure accuracy of registration data, ICANN org will examine this request and consider it in terms of available budget and resources. Any study of accuracy of registration data would require substantial review to ensure consistency with all requirements in ICANN policies and agreements, and applicable laws and regulations. If a third-party is required to complete such a study, this may require additional budget and approvals. In some cases, Board approval may be required depending on the cost and scope of the requested study.

Background on the WHOIS ARS

The following section provides information on how the ARS was developed and operated, a brief overview of the ARS accuracy tests, and an overview of how ICANN Contractual Compliance was involved in the ARS process.

Development of the WHOIS ARS

In May 2012, the WHOIS Review Team issued its [Final Report](#), which included recommendations relating to registration data (WHOIS) accuracy. Based on these recommendations, the Board [resolved](#) in August 2012 to:

1. “proactively identify potentially inaccurate gTLD data registration information in gTLD registry and registrar services, explore using automated tools, and forward potentially inaccurate records to gTLD registrars for action; and
2. publicly report on the resulting actions to encourage improved accuracy”

In April 2013, the GAC issued advice in its [Beijing Communique](#) regarding safeguards, which were to be applicable to all new gTLDs, stating that: “Registry operators will conduct checks on a statistically significant basis to identify registrations in its gTLD with deliberately false, inaccurate or incomplete WHOIS data at least twice a year.”

Following this advice, the Board resolved in June 2013, as part of the [New gTLD Program Committee Proposal for Implementing the GAC Safeguards](#) that:

“ICANN (instead of Registry Operators) is well positioned to implement the GAC’s advice that checks identifying registrations in a gTLD with deliberately false, inaccurate or incomplete WHOIS data be conducted at least twice a year. To achieve this, ICANN will perform a periodic sampling of WHOIS data across registries to identify potentially inaccurate records. ICANN will also maintain statistical reports that identify the number of inaccurate WHOIS records identified”.

The ARS also incorporated advice from [SAC058](#), the SSAC Report on Domain Name Registration Data Validation. SAC058 included recommendations for the terminology of validation (e.g., “syntactical” and “operational” validation), which ICANN org ultimately used in its reports. ICANN org also leveraged the taxonomy of validation to implement ARS as a phased approach, i.e., Syntactic (Phase 1), Operational (Phase 2), and Identity (Phase 3) phases (see more below for information on testing).

In August 2015, ICANN org released its first report, the [Phase 1 Report](#), which focused on syntactical accuracy. In December 2015, ICANN org released the first of an iterative series of [Phase 2 Reports](#), in which ICANN org examined both syntactical and operational accuracy. Since that time, ICANN org published a report every six months. The last report published by ICANN org was in June 2018 ([Cycle 6 report](#)).

The ARS used three vendors to complete the analysis described in the reports.

- **NORC at the University of Chicago (NORC):** Provides statistical analysis and parsing of registration data.
- **DigiCert:** Provides syntactical and operational validation services for both email addresses and phone numbers.
- **Universal Postal Union (UPU):** Provides syntactical and operational validation of postal addresses.

See also section on [issues related to vendor contracts and restarting the ARS](#).

WHOIS ARS Sample Design & Testing

With the help of the NORC, ICANN collected a sample of WHOIS records for accuracy testing purposes. For Cycle 6, ICANN org collected an initial sample of 200,000 records from gTLD zone files of 818 gTLDs. From the initial sample, an analyzed subsample of 12,000 records was created. 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. Each ARS report provides a detailed overview of the process for obtaining the sample used for testing.

For testing, and again in cooperation with the vendors noted above, syntax and operability accuracy tests were designed to assess the contact information of a WHOIS record by comparing it to the applicable contractual requirements of the 2009 and 2013 RAAs. Syntax testing assessed the format of a record (e.g., does the email address contain an “@” symbol?), and operability testing assessed the functionality of the information in a record (e.g., did the email not get bounced back?). ICANN org contracted with a group of vendors to conduct the testing. The vendors performed syntax and operability accuracy tests 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). 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).

Regarding findings of the reports, the ARS found in its Cycle 6 Report (June 2018), for example, that approximately 94 percent of email addresses, 60 percent of telephone numbers, and 99 percent of postal addresses were found to be operable (e.g., accurate such that the email address or phone numbers are operational) for all three contacts (administrative, technical, and registrant), according to the requirements of the 2013 RAA. Results from all the ARS reports can be found on the ARS page.

See also section on [issues related to data](#) should the ARS be continued.

WHOIS ARS & ICANN Contractual Compliance

Upon completion of the accuracy testing, ICANN org worked together with the NORC to prepare a report of all the results of the tests that could be provided to ICANN Contractual Compliance. ICANN Contractual Compliance used the report to identify potentially inaccurate records that may require follow-up with registrars, as explained below.

Syntax Inaccuracy Follow-Up

ARS complaints were classified as WHOIS format errors if the error indicated non-compliance with the format requirements of the 2013 RAA, but the information was otherwise valid and contactable (e.g., a missing +1 county code for a registrant located in the United States). Because the 2009 RAA does not include format requirements, WHOIS format errors were not considered for registrars under the 2009 RAA. Where the error rendered the contact unreachable (e.g., a missing postal address), the ARS complaint was processed as a WHOIS inaccuracy complaint.

Operability Inaccuracy Follow-Up

ARS complaints that were generated due to failures of operability were processed as WHOIS inaccuracy complaints. Operability failures indicated 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.

Results

ICANN Contractual Compliance worked with Registries and Registrars to resolve identified issues. Metrics for the ARS are presented in the ICANN Contractual Compliance Performance Reports (see <https://features.icann.org/compliance>) and at ICANN Public Meetings. ICANN publishes additional metrics on the WHOIS ARS Contractual Compliance Metrics page (see <https://whois.icann.org/en/whoisars-contractualcompliance-metrics>).

See also section on [issues identified by ICANN Contractual Compliance](#) related to ARS complaints.

GDPR & Decision to Put WHOIS ARS On-Hold

Between December 2015 to June 2018, ICANN org published an ARS report every six months. However, since June 2018, ICANN org has not conducted any further data collection or analysis, and the last report published by ICANN org was the June 2018 Cycle 6 Report. ICANN org made the decision to pause further reports based on consultation with Legal following the GDPR being implemented and subsequent adoption of the Temporary Specification. Additionally, inquiries made by registrars as to whether it is permissible to provide certain registration data to ICANN in response to a WHOIS inaccuracy ticket issued by ICANN Contractual Compliance because of the ARS caused ICANN org to reconsider continuing with the ARS.

Alternative Options or Data Sources

At the time of putting ARS on-hold, no alternative data sources were seriously considered. There have been discussions/conversations regarding other options, such as using escrow data or Bulk Registration Data Access (BRDA), but these have not been thoroughly investigated as viable alternatives.

The GDPR has limited access to gTLD registration data for all parties, including ICANN org. This includes limiting ICANN org's ability to check the accuracy of gTLD registration data. ICANN org has engaged with the European Commission to suggest that the proposed NIS2 legislation could recognize ICANN's legitimate interest in having access to non-public gTLD registration data for the purpose of checking accuracy. Because of the GDPR, however, it is not clear that ICANN org has the ability to access gTLD registration data.

Any study of the accuracy of registration data would require substantial review to ensure consistency with all requirements in ICANN policies and agreements, and applicable laws and regulations.

ICANN org Communication to the Board regarding WHOIS ARS Hold

ICANN org has made the Board aware that the ARS is on hold via its twice-annual CEO reports to the Board. ICANN org first noted in its [January 2019](#) report that “[t]he cycle 7 report of the ARS has been paused as we consider updates to the process based upon GDPR and changes to available public registration data as a result of Registry and Registrar implementation of the Temporary Specification.” And further in April 2019:

“The ARS remains paused as ICANN org assesses the effects of General Data Protection Regulation (GDPR). Based on the lack of predictable publicly available registration data and given the community work from the GNSO’s Expedited Policy Development Process (EPDP) on Temporary Specification for gTLD Registration Data, ICANN org believes it may be prudent to continue to pause and consider the impact of the EPDP efforts and assess our ability to effectively administer ARS.”

ICANN org also briefed the Board on the topic of the ARS ahead of the ICANN68 meeting. In its [discussion](#) with the GAC at ICANN68, the Board provided information on many of the issues surrounding the ARS and why it is on hold.

Issues Related to Continuing WHOIS ARS

ICANN org has identified the following additional issues with continuing and/or re-starting the ARS:

Contracts & Budget

- **Vendor contracts have expired.** Since pausing the ARS, the contracts for all three ARS vendors (NORC, DigiCert, and UPU) have expired. To restart ARS, ICANN org would require approximately 12-18 months of development time to either engage existing and/or any new service providers as well as assess what changes are may be required (i.e., in sampling, testing, reporting, and the interaction with Compliance) in order to continue with the ARS. Due to the expense (potentially between \$300,000 - \$500,000 annually), Board approval may also be required.

Data

- **Many of the fields currently tested for accuracy will no longer be required following implementation of the EPDP recommendations.** In line with EPDP [Phase 1](#) recommendations, registrars will no longer be required to collect information for many of the nine fields noted above. In fact, only the registrant email, address, and phone will be required. The information for the administrative contact will no longer be required at all, and the name, phone and email for the technical contact are optional, while the address will not be required.
- **Continuing the ARS with publicly available registration data may not be useful.** ICANN org has expressed to the GNSO Council concerns with continuing the ARS using publicly available data, which it has relied on to measure accuracy. There is a question

as to whether publicly available data will provide useful results in terms of the overall accuracy of registration data; indeed, any results may be biased toward those contracted parties who do publish contact details in registration data and/or those registrants who consent to publication.

Legal Environment

- **The legal environment has changed since ARS started.** While ICANN org could restart ARS using public registration data (despite the issues noted above), ICANN org does not have the contractual ability to require the contracted parties to provide access to non-public registration data to ensure that the ARS is collecting a representative sample of registrations (i.e., not simply domains for which registration data is publicly available).
- **It is not clear that the “Purpose 2” would make the processing of non-public registration data GDPR-compliant.** The new “Purpose 2” recommended by the EPDP Phase 2 team (“Contribute to the maintenance of the security, stability, and resiliency of the Domain Name System in accordance with ICANN's mission”) identifies a purpose for ICANN's processing of personal data that could be of relevance in this context. However, this new statement of purpose, alone, wouldn't make ICANN's processing of non-public registration data for ARS GDPR-compliant. In order for such processing to comply with the GDPR, the processing (i.e., collection, analysis, retention, and eventual destruction) must be necessary for the purposes of a legitimate interest pursued (such as contributing to the maintenance of the security, stability, and resiliency of the DNS in accordance with ICANN's mission), and this interest must not be outweighed by the data subjects' interests and fundamental rights and freedoms. Further, even if the requirements for legitimate interest processing can be demonstrated, the processing must comply with other GDPR requirements, including implementation of any required safeguards for cross-border transfers of personal data and ensuring compliance with data subject rights, including notice and opportunity to object. Therefore, significant changes to prior ARS procedures would be required.

Purpose & Goal of Measuring Accuracy

- **Continued higher-level discussions on accuracy.** ICANN org believes it is important to view the question of measuring registration data accuracy in light of ongoing conversations on data protection, such as addressed in Göran Marby's September 2020 letter to the Governmental Advisory Committee (GAC), October 2020 letter to the European Commission, subsequent response from the European Commission, and an ICANN org December 2020 blog regarding ICANN's GDPR-related efforts. The discussion of accuracy measurement should not be solely focused on the ARS but should encompass the wider range of issues related to the GDPR and data protection. Additionally, developments in the EU regarding accuracy, such as the proposal for a revised Directive on Security of Network and Information Systems (NIS2 Directive), as discussed in ICANN org's December 2020 blog, should also be taken into account.

-
- **The ARS is focused on a static moment of the accuracy of registration data but not necessarily on how to improve it.** While one of the intended functions of the ARS is to provide information on registration data inaccuracies to ICANN Contractual Compliance for follow-up with registrars, which could in turn lead to improvement of accuracy of registration data, the ARS itself is generally focused on a snapshot of accuracy, not on ways to improve accuracy over time. It should be considered whether this method of reviewing accuracy meets the needs and demands of the ICANN community, or whether, at this time, a different mechanism should be considered for reviewing and improving accuracy of registration data.

ICANN Contractual Compliance Processing

- **Many of the accuracy tests create false positives.** Many of the tickets created from the ARS contain data incorrectly marked as inaccurate (i.e., inoperable). For example, “standard format” is somewhat relative; address formats can vary from country to country and, depending on the source for the format used by the testing vendor, it could label some data as inaccurate when it is in fact accurate.
- **Compliance often receives tickets with already out-of-date registration data.** Because of the time required to conduct the accuracy analysis of the registration data, ICANN Contractual Compliance receives the results up to six months after the registration data was originally sampled. This delay means that the information may in turn be obsolete and the complaint thus no longer valid. For example, of the over 5,000 tickets generated from the ARS report in June 2018, over 1,100 were closed before 1st notice because the data was already outdated from the time of sampling.

Future Studies regarding Registration Data Accuracy

In the context of the issues noted above, ICANN org noted in its Briefing on accuracy, that it may be beneficial to commission a study on how accuracy of registration data might be measured, whether using publicly available data or with access to non-public registration data (in lieu of continuing with the ARS as-is). ICANN org noted in the Briefing that any such study would need to be done with input from the community as well as potentially an agreement with contracted parties. ICANN org further recommended that it--together with the GNSO Council--develop a framework for a study on how to measure accuracy and/or obtain a snapshot of accuracy as it stands now, and that this be presented to the ICANN community for review and input.

At the time of making this recommendation, ICANN org did not have a specific or concrete plan for the study or how it might be designed. However, since that time, ICANN org has done some initial brainstorming as to potential objectives of such a study, which are described below.

Potential Objectives of a Study

Based on the above context for the ARS and an apparent community desire for continued study of gTLD registration data accuracy, the objectives of such a study could be to:

-
- Determine what gTLD registration data is currently publicly available and to what extent this data might be used for a study of gTLD registration data accuracy.
 - Based on the findings related to the above, produce a snapshot of accuracy of publicly available gTLD registration data.
 - Determine the legal requirements and procedures for accessing non-public gTLD registration data, including any required involvement of contracted parties.
 - Produce—if it is possible to access non-public gTLD registration data—a snapshot of accuracy of non-public gTLD registration data as well as provide a comparison to the accuracy of publicly available gTLD registration data.
 - Determine how the higher-level conversation on accuracy (e.g., considering NIS2 Directive) affects a study of gTLD registration data accuracy going forward.

Closing

ICANN org understands that the Accuracy Scoping Team may have additional questions and will continue to engage with the Scoping Team to provide any relevant information, as applicable. Should there be a request for an additional study (or studies) regarding accuracy of registration data and/or how to measure accuracy of registration data, ICANN org will examine this request and consider it in terms of available budget and resources. As noted above, any study of accuracy of registration data would require substantial review to ensure consistency with all requirements in ICANN policies and agreements, and applicable laws and regulations. If a third-party is required to complete such a study, this may require additional budget and approvals. In some cases, Board approval may be required depending on the cost and scope of the requested study.



One World, One Internet

Visit us at icann.org



[@icann](https://twitter.com/icann)



facebook.com/icannorg



youtube.com/icannnews



flickr.com/icann



linkedin/company/icann



soundcloud/icann



instagram.com/icannorg