**Registration Data Accuracy Scoping Team – Gap Analysis**

**References to data sources**

1. **U.S. Government Accountability Office (GAO) (2005), Internet Management: Prevalence of False Contact Information for Registered Domain Names – 2005 (see** [**here**](https://www.gao.gov/products/gao-06-165)**)**

Individuals or organizations seeking to register the names of their Web sites may provide inaccurate contact information to registrars in order to hide their identities or to prevent members of the public from contacting them. Contact information is made publicly available on the Internet through a service known as Whois. Data accuracy in the Whois service can help law enforcement officials to investigate intellectual property misuse and online fraud, or identify the source of spam e-mail, and can help Internet operators to resolve technical network issues. GAO was asked, among other things, to (1) determine the prevalence of patently false or incomplete contact data in the Whois service for the .com, .org, and .net domains; (2) determine the extent to which patently false data are corrected within 1 month of being reported to ICANN; and (3) describe steps the Department of Commerce (Commerce) and ICANN have taken to ensure the accuracy of contact data in the Whois database.

Based on test results, GAO estimates that 2.31 million domain names (5.14 percent) have been registered with patently false data—data that appeared obviously and intentionally false without verification against any reference data—in one or more of the required contact information fields. GAO also found that 1.64 million (3.65 percent) have been registered with incomplete data in one or more of the required fields. In total, GAO estimates that 3.89 million domain names (8.65 percent) had at least one instance of patently false or incomplete data in the required Whois contact information fields. The table below shows the estimated number of instances of patently false data for each of the three types of contact information within each generic top-level domain.

1. **NORC Draft Report for the Study of the Accuracy of WHOIS Registrant Contact Information – Jan 2010 (see** [**here**](https://www.icann.org/en/system/files/newsletters/whois-accuracy-study-17jan10-en.pdf)**)**

Based on accuracy as defined in the [2001](https://www.icann.org/resources/unthemed-pages/raa-2001-05-17-en)/[2009](https://www.icann.org/en/resources/registrars/raa/ra-agreement-21may09-en.htm) Registrar Accreditation Agreement, this study describes the accuracy of WHOIS registrant contact information using a sample of gTLD domains from 16 countries. The study describes the methodology and criteria used for determining a set of scores for domains based on the level of accuracy achieved. The study also discusses various “barriers to accuracy.”

This study was commissioned by ICANN in order to get a baseline measurement of what proportion of WHOIS records are accurate. The methodology of the study was a precursor to the ARS/ARS pilot studies, noted below.

1. **NORC Whois Accurcy Reporting System Pilot Report – Dec 2014 (see** [**here**](https://whois.icann.org/sites/default/files/files/ars-pilot-23dec14-en.pdf)**)**

The WHOIS Pilot Study Report describes the results of a pilot accuracy study conducted by NORC at the University of Chicago in collaboration with private and public sector validation specialists. The Report illustrates the findings and methodology to be deployed in the WHOIS Accuracy Reporting System (ARS) under development by ICANN.

The Pilot Accuracy Study was a precursor to the ARS, designed with the intention of isolating and testing key components of the system. Results from the Pilot Study were used to inform the design of the ARS.

The Pilot study examined accuracy rates from multiple perspectives to give a realistic picture of today's WHOIS. Some notable findings are:

* Operationally, registrars under the 2013 RAA have more accuracy for email addresses than registrars under the 2009 RAA. Registrars under the 2009 RAA have more operational accuracy in telephone numbers while the two groups are similar on postal addresses.
* Email and telephone syntactical accuracy is not statistically different for prior and new gTLDs while syntactical accuracy is better for Prior gTLDs on postal addresses.
* The operational numbers for email addresses indicate that New gTLDs have slightly better accuracy than Prior gTLDs. Prior gTLDs do have more operational accuracy on telephone numbers, but the two groups are equal on operational postal address accuracy.
* The differences in syntactical accuracy between registrars under the 2013 RAA and registrars under the 2009 RAA for email, telephone numbers and postal addresses are not statistically significant.
1. **WHOIS ARS Phase 2 Reporting (see** [**https://whois.icann.org/en/whois-ars-phase-2-reporting**](https://whois.icann.org/en/whois-ars-phase-2-reporting)**)**

WHOIS ARS Phase 2 Reporting

This page provides the results from the Phase 2 Accuracy reports.  Phase 2 examines both syntax and operability accuracy of WHOIS contact information over several dimensions, focusing on rates of conformance by contact mode (Email, Telephone or Post) to the requirements of RAAs (2009 RAA or 2013 RAA).

**Cycle 1 Summary - Dec 2015***(records as of Jul 2015)*

* For syntax accuracy, there was a drop in telephone number accuracy from Phase 1. The drop in telephone number accuracy seemed to be due to an increase in missing country codes among the telephone numbers sampled for Cycle 1.
* 87 percent of email addresses, 74 percent of telephone numbers and 98 percent of postal addresses met all operability requirements of the 2009 RAA.
* 65 percent of domains passed all operability tests for all contact types (registrant, administrative, technical) and contact modes (email address, telephone number, postal address).

**Cycle 2 Summary - Jun 2016***(records as of Jan 2016)*

* 99 percent of records had at least one contact mode (email, phone, or postal address) that can be used to establish contact.
* 98 percent of postal addresses, 76 percent of telephone numbers and 91 percent of email addresses met all operability requirements of the 2009 RAA.
* 70 percent of domains passed all operability tests for all contact types (registrant, administrative and technical) and contact modes (email, telephone and postal address)

**Cycle 3 Summary - Dec 2016***(records as of Jul 2016)*

* 97 percent of records had at least one email or telephone number meet all operability requirements of the 2009 RAA, which implies that nearly all records contain information that can be used to establish immediate contact.
* 97 percent of postal addresses, 72 percent of telephone numbers and 90 percent of email addresses met all operability requirements of the 2009 RAA.
* 65 percent of domains passed all operability tests, which is about five percent less than Cycle 2. Besides natural sample variation, one possible reason could be market changes and the growth of domains in certain regions.

**Cycle 4 Summary - Jun 2017***(records as of Jan 2017)*

* 98 percent of records had at least one email or telephone number meet all operability requirements of the 2009 RAA, which implies that nearly all records contain information that can be used to establish immediate contact.
* 97 percent of postal addresses, 69 percent of telephone numbers and 94 percent of email addresses met all operability requirements of the 2009 RAA.
* 65 percent of domains passed all operability tests, which is on par with Cycle 3. Across all 4 cycles of Phase 2, full operability has remained steady between 65 and 70 percent.

**Cycle 5 Summary - Dec 2017***(records as of Jul 2017)*

* 98 percent of records had at least one email or telephone number meet all operability requirements of the 2009 RAA, which implies that nearly all records contain information that can be used to establish immediate contact.
* 98 percent of postal addresses, 67 percent of telephone numbers and 94 percent of email addresses met all operability requirements of the 2009 RAA.
* 63 percent of domains passed all operability tests, a slight decrease from Cycle 4. Across all 5 cycles of Phase 2, full operability has been between 63 and 70 percent.

**Cycle 6 Summary - Jun 2018***(records as of Jan 2018)*

* 98 percent of records had at least one email or telephone number meet all operability requirements of the 2009 RAA, which implies that nearly all records contain information that can be used to establish immediate contact.
* 99 percent of postal addresses, 60 percent of telephone numbers and 92 percent of email addresses met all operability requirements of the 2009 RAA.
* 56 percent of domains passed all operability tests, a decrease from Cycle 5. Across all 5 cycles of Phase 2, full operability has been between 56 and 70 percent.
1. **Interisle WHOIS Contact Data Availability and Registrant Classification Study – January 2021 (see** [**here**](https://www.interisle.net/ContactStudy2021.pdf)**)**

To date, ICANN has not performed work to determine key facts about the effects of its policy, such as how much data has been redacted, or under what circumstances. Therefore, there has not been a factual basis for determining if or how the policy is having its intended effects. The purpose of the current study is to establish answers to those questions. This study uses previously established methodologies, so its results can be compared to the results of previous studies and provide historical continuity. It is our hope that this data can be used for fact-based discussion about what effects the GDPR and ICANN’s resulting policy have had.

The major findings of this study are:

ICANN’s GDPR-driven policy has led to the redaction of the contact data for most gTLD domains and has allowed registrars and registry operators to redact (withhold) much more contact data than is required by GDPR—perhaps five times as much as is necessary. While ICANN’s policy has generally protected the data that must be protected per GDPR, ICANN’s policy has also been used to conceal a much larger set of contact data that is *not* subject to GDPR.

1. Before the GDPR went into effect and ICANN changed its registration data policy, the actual identities of about 75.7% of gTLD domain registrants were available in WHOIS. The other 24.3% of domains were under privacy/proxy protection. *(page 15)*
2. At present, only 13.5% of domains have an actual registrant identified in WHOIS. Registrars and registry operators have used ICANN’s post-GDPR policy to redact contact data from 57.3% of all domains. Adding proxy-protected domains, this means that 86.5% of registrants cannot be identified via WHOIS. *(page 15)*