The Accuracy Framework

Steve Crocker, 15 November 2021

This is a holistic view of the "Accuracy Problem."

The big picture

There are two halves to the big picture.

The first half is the collection process. Registrars collect registration data. As they do so, they may or may not take steps to ascertain the accuracy of the individual data elements. There are multiple possible levels of assurance, including the null level. The level of assurance is determined by the registrar's policy and practice. The registrar's policy reflects both its own preferences and conformance with higher level authorities, i.e., the registry, ICANN policy, and laws of the relevant jurisdictions.

The second half is the use of the data. Authorized requesters receive registration in response to authorized requests. The efficacy of the data depends in part on the accuracy of the data.

The focus of this memo is on the first half, the validation of the data.

The following is a framework consisting of four parts. Part I, the top level, of this framework is the statement of purpose(s). Part II is the specification of the validation requirements. Part III is the measurement of the validation process. Part IV is checking, reporting and enforcement. These are depicted in figure 1.

I. Purpose(s)

A statement of the criteria of utility, i.e., what is the information to be used for?

Examples:

- Contactability

  Question: If the purpose is to be able to contact the registrant, what happens if the contact data is accurate but the registrant fails to respond?

The Validation Framework

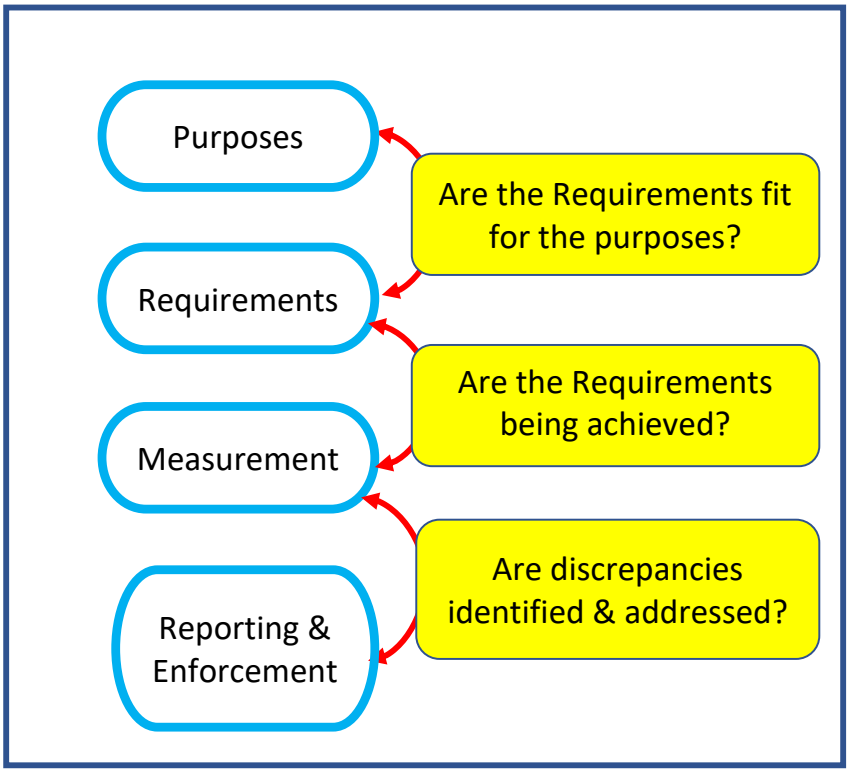

Figure 1: The Validation Framework

- Evidence of contactability suitable for demonstrating that an effort has been made. Useful in transferring risk. "Judge, we notified the registrant the domain name is being used to infringe on our copyrighted material."

- Other use cases?

II. Requirements

Specification of the data to be collected and what level of validation is applied to each data element.

The current dialog seems to have settled on a vocabulary of four levels: no validation, syntactic validation, operational validation, identify verification. In broad terms, operational validation implies the data has been subjected to a test to see if it works, and identity verification indicates an additional test has been performed to confirm the data is correctly associated with the party.

Commented [SC1]: Are these the preferred terms? "Verification" and "validation" are well defined terms in the software development world but seem to be used a bit differently in the ICANN policies and contracts.

The method of validating data at each of these levels depends on the type of data. Most of the dialog has been focused on email addresses and phone numbers. For these types of data, operational validation is understood to mean a message or phone call has been placed and there has been an affirmative response. "Affirmative response" means someone responds and affirms in the response that this is the correct person.

Operational validation of other types of data, e.g., names, organization, address is less well defined.

Identity verification also depends on the type of data. Further, the NIST and EU validation scales distinguish between two levels of identity validation. In broad terms, the lower level in each system (NIST IAL2; EU "Substantial") permit remote verification, and the higher level in each system (NIST IAL3; EU "High") require in person verification of the data elements.

Criteria: If the requirements are satisfied, is data fit for purpose? This is an essential question. It is not uncommon in the design and deployment of a system that it may fail to meet the needs of the users but satisfy the contractual requirements.

III. Measurement of Accuracy

1. Classification of Conformance

   Comment: A categorization, apparently in the Whois Accuracy Pilot Study Report, lists three levels of conformance between No Failure and Full Failure: Minimal Failure, Limited Failure, Substantial Failure. It is not clear how the definitions associated with these terms relate to whether the data does or does not serve the intended use(s). What is the impact regarding achievement of the intended purpose(s)?

2. Method measurement

3. Target Levels of Performance

   Question: What are the target levels of conformance?
   Question: How does the target relate to achievement of the purpose(s)?

IV. Checking, Reporting and Enforcement

Description of the methods for checking, reporting inaccuracies, and enforcing both required practice and correction of errors.

Question: How effective are these methods in achieving the specified target levels of conformance?

Assurance Levels

Within the DNS registration community, as shown in figure 2, it appears there are four distinguishable levels of assurance used in requirements and/or used in discussion of what the requirements should or might be. The lowest level, V0, indicating no validation, is not usually mentioned but is implied if no validation is required.

The details of how to do syntactic validation, operational validation and identity validation depend on the data element being validated. Therefore, these descriptions convey the intent and top level of the validation process. Additional details are required to apply each of these levels to individual data elements.

V0	**No validation**. Whatever the registrant provides.
V1	**Syntactic validation**. Syntactic validation depends on the particular data element, e.g. text@domain_name for an email address, a limited string of digits for a phone number, etc.
V2	**Operational validation**. An operational test to see if the data is operational. The details vary with the type of data. Of particular interest are phone numbers and email addresses. In both cases, an operational test consists of using the data and waiting for an affirmative response.
V3	**Identity validation**. This requires verification that the data element is properly associated with the registrant. This often requires legal documentation or third-party attestation.

Figure 2: Validation Levels

The highest level, Identity Validation, is rarely used in DNS registration processes. Identity Validation is, of course, used in other registration processes such as banking. Further, many applications distinguish between two levels of Identity Validation. The lower level allows for remote validation. The higher one generally requires in person validation. The NIST and EU assurance levels refer to these as IAL2 or “Substantial” for the lower level and IAL3 or “High” for the higher level. The text from the NIST and EU Assurance Levels in the Appendix.

Appendix: Text from the NIST and EU Assurance Levels

NIST Levels

Assurance in a subscriber's identity is described using one of three IALs:[1]

IAL1: There is no requirement to link the applicant to a specific real-life identity. Any attributes provided in conjunction with the subject's activities are self-asserted or should be treated as selfasserted (including attributes a CSP asserts to an RP). Self-asserted attributes are neither validated nor verified.

IAL2: Evidence supports the real-world existence of the claimed identity and verifies that the applicant is appropriately associated with this real-world identity. IAL2 introduces the need for either remote or physically-present identity proofing. Attributes could be asserted by CSPs to RPs in support of pseudonymous identity with verified attributes. A CSP that supports IAL2 can support IAL1 transactions if the user consents.

IAL3: Physical presence is required for identity proofing. Identifying attributes must be verified by an authorized and trained CSP representative. As with IAL2, attributes could be asserted by CSPs to RPs in support of pseudonymous identity with verified attributes. A CSP that supports IAL3 can support IAL1 and IAL2 identity attributes if the user consents.

EU Levels

The three levels of assurance are as follows:[2]

- **Low**: for instance, enrolment is performed by self-registration in a web-page, without any identity verification;
- **Substantial**: for instance, enrolment is performed by providing and verifying identity information, and authentication by using a user name and a password and a one-time password sent to your mobile phone;
- **High**: for instance, enrolment is performed by registering in person in an office, and authentication by using a smartcard, like a National ID Card.

[1] https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-63a.pdf, section 2.2, page 3
[2] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eIDAS+Levels+of+assurance