**SAC058, Report on Domain Name Registration Data Validation (2013)**

***Possible Requirements***

*Summarized by Maryan Rizinski*

*June 6, 2016*

# Description of the document:

This document is a report to the ICANN Board from the Security and Stability Advisory Committee (SSAC) concerning the issue of domain name registration data quality. In this report, the SSAC examines the feasibility and suitability of improving registration data accuracy through validation.

# List of possible requirements from the document:

## Data Accuracy (DA):

* Identify validation techniques that can be automated and develop policies that incent the development and deployment of those techniques. (Page 4)
* To improve registration data accuracy, there needs to be 1) an incentive for the registrant to submit accurate data, or 2) efforts by registry / registrar to follow up and check the accuracy of the submitted data; or 3) both. (Page 5)
* Syntactic Validation: Assess data with the intent to ensure that they satisfy specified syntactic constraints, conform to specified data standards, and are transformed and formatted properly for their intended use. (Page 7)
* Operational Validation: Assess that data correspond to the intended use in their routine functions (e.g. check that an email address or phone number can receive email or phone calls, check that a postal address can receive postal mail, etc.). (Page 8)
* Identity Validation: Assess that data corresponds to the real world identity of the registrant entity. (Page 8)
* Determine the length of time before the validation of changes to contact information must be repeated. (Page 8-9)

*Note: In the section on “Implementing Validation”, the SSAC considers the feasibility of validation of four types of contact information elements. They are name, postal address, email address, and telephone and fax number. The requirements for these four types of contact information elements are given as follows.*

### Name

* Syntactic Validation: To achieve effective syntactic validation of a name as one of the contact information elements, the script (or writing system) used for a name element must be known. If it is, confirming that the syntax conforms to the script is possible and can be automated. However, the language of a name cannot be determined precisely as many languages share the same script. (Page 9)
* Operational Validation: Create exception lists for auditing purposes in order to facilitate the process of operational validation of names because names in the world are diverse and it may not be possible to operationally verify a name automatically. (Page 10)
* Identity Validation: Require the submission of physical documentation issued by a government authority to verify that registration data contact information corresponds to a real world entity. (Page 10)

### Email Address

* Syntactic Validation: Syntax for a valid email address (defined as per RFC 5322) and syntax for a valid internationalized email address (defined as per PRFs 6530-33) should be checked automatically. (Page 10)
* Operational Validation: Having in mind that an email address is defined as a string composed of a Left Hand Side (LHS) and Right Hand Side (RHS) separated by the at-symbol (@), verify that an email address is operational implementing several checks (e.g. with respect to the RHS check that the domain name exist in the DNS while with respect to the LHS check that the endpoint SMTP accepts an email message for the recipient specified at the LHS). (Page 10)
* An effective verification technique of an email address is to attempt to deliver an email message that requires explicit user action. In this technique, an email address should not be considered valid until the user receives and performs some action described in the email, such as clicking on a web link or replying to the message in a specified way. Note that sometimes anti-spam measures could still block these verification emails. (Page 11)
* Identity Validation: In order to verify that an email address is used exclusively by a particular registrant, contact the registrant using an out-of-band method, i.e., contacting the registrant without using email (e.g. two possibilities are using the postal information or the telephone information to contact the registrant). (Page 11)

### Telephone number

* Syntactic Validation: Perform automatic checks to determine if a telephone number complies with the E.164 standard (E.164 is an ITU-T recommendation that defines the international public telecommunication numbering plan used in the PSTN and some other data networks). (Page 11)
* Operational Validation: Verify E.164 formatted PSTN addresses (telephone numbers) by leveraging PSTN databases. (Page 11)
* Use the Short Message Service (SMS) to verify a phone number (works only for cellular numbers). (Page 12)
* Identity Validation: In order to verify that a telephone number is used exclusively by a particular registrant, contact the registrant using an out-of-band method, i.e., contacting the registrant without using the telephone number (e.g. two possibilities are using the postal information or the email address to contact the registrant). (Page 12)

### Postal address

* Syntactic Validation: The EPP standard defines an opaque container and loose constraints that can support internationalized postal addresses.
* Operational Validation: Verify postal addresses by leveraging postal databases. There are about 200 such databases in the world with about 20 (G20 major economies) being highly accurate. (Page 13)
* Deliver a postal message to a postal address in order to verify with a high level of certainty that the postal address is valid. (Page 13)
* Identify Validation: In order to verify that a postal address is used exclusively by a particular registrant, contact the registrant using an out-of-band method, i.e., contacting the registrant without using the postal address (e.g. two possibilities are using the telephone number or the email address to contact the registrant). (Page 13)

## Cost (CS):

* The actual cost of validation is dependent on many factors that need to be considered at the same time. Some of these factors are the cost of developing and deploying automation where applicable, the cost of a single validation, the cost of repeating the validation, and the cost of maintaining the information and infrastructure necessary to support the process of validation. (Page 9)
* Verifying whether or not an E.164 conformant phone number can be called requires attempting to connect to it using either the PSTN or the Signaling System No. 7 (SS7) network. Both methods may incur charges. (Page 12)
* When a cellular number is verified with the use of the Short Message Service (SMS), having a registrant call from a particular number may pose problems for those that use corporate direct inward dialing (DID) lines where outbound calls are automatically mapped to the main corporate number, frequently without the knowledge of the person making the call. Both may incur charges for either the sender or receiver or both. (Page 12)
* Within the G20 major economies, about eight have highly accurate address information. While the information is available it is expensive and each country has a different procedure for normalizing an address, which must be done before it can be checked against a postal address database. (Page 13)
* There is a large upfront cost in the beginning as nothing is validated. As registrants are validated the number of unverified registrants drops significantly, and thus costs for subsequent years might be more directly related to the validity periods, i.e., the frequency at which data must be revalidated. (Page 14)
* There are economies of scale for validation: costs of per contact data element validation drops as more contacts are validated. (Page 14)
* In EPP registries, registrars are free to create and manage multiple contact objects that refer to the same individual. Thus, the cost of validating the contact data associated with a domain name may be the cost of validating each contact object. However, from an operational cost and registrant experience perspective, validation of a registrant associated with multiple domains might not require each domain’s contact data elements to be re-validated if the registrant’s contact data elements are the same for each domain name. (Page 14)

## Benefits (BE):

* From a technical perspective, certain verification measures can be taken to reduce unintentional errors by registrants; for example, a formal data structure and strong typing of data (e.g., this field must be Arabic numbers only, this field must be alphabetical characters only) can reduce certain typographical errors. Enforcing mandatory submission of data for key data fields may reduce cases where users omit information. (Page 14)
* The use of automated techniques may necessitate an initial investment but the long-term improvement in the quality and accuracy of registration data will be substantial. (Page 15)

## Risks (RS):

* Registration data often contain "stale" contact information and that this problem can cause difficulties when registrants seek to renew a domain name or modify DNS information. Stale information may prevent registrars from notifying a registrant that a domain registration is about to expire or that changes, possibly unauthorized, have been made to his domain registration. Failure to update information may result in domain hijacking or a dispute over the "ownership" of a domain. (Page 6)
* Since current access to registration data is public and anonymous, some individuals and businesses submit incorrect information because they do not wish their contact information to be collected and used by miscreants as targets for spam and other attacks. (Page 6)
* Some people intentionally submit false information because they do not wish to disclose personal contact information that can be accessed publicly and anonymously. (Page 6)
* Miscreants intentionally provide false information to obfuscate identification by law enforcement or parties that investigate malicious use of domains. (Page 7)
* Current registration requirements take a minimalist approach to validation. Unless credit verification measures are stringently applied for all levels of payment, little or no additional proof of identity and verification of contact information is required when a user registers a domain name. (Page 8)
* Users may mistype when registering domain names. The current validation processes can overlook errors. (Page 8)
* Users may not understand the consequences of the registration data accuracy program and annual obligation to maintain accurate and complete registration data. They also may refuse to take time to check that their contact information is current, or reject the notion that they will forfeit a domain registration simply because some registration data are inaccurate. (Page 8)
* If an email address is verified requiring explicit user action upon receiving a verification email (such as clicking on a web link or replying to the message in a specific way), the timing of the verification email message will need to be carefully considered as to how it affects the overall registration process. Sending the verification email as an integral part of the registration process would alter the business process and may affect registration costs. Sending the verification email after registration would risk being ignored by the registrant or could introduce an attack vector. A miscreant, knowing that these verification emails will be sent, could initiate various types of man-in-the-middle attacks. Past security research has shown that such spear-phishing attacks are highly effective. (Page 11)
* Existence of a postal address in a database does not guarantee that the physical address exists (e.g. apartment numbers in the United State Postal Service address database are indicated as a range. As a result, an address may validate as accurate and complete when in fact it is undeliverable). (Page 13)