SAC 051
SSAC Report on Domain Name WHOIS Terminology and Structure

A Report from the ICANN
Security and Stability
Advisory Committee
(SSAC)
19 September 2011

Preface

This is a Report of the Security and Stability Advisory Committee (SSAC). The SSAC advises the ICANN community and Board on matters relating to the security and integrity of the Internet's naming and address allocation systems. This includes operational matters (e.g., matters pertaining to the correct and reliable operation of the root name system), administrative matters (e.g., matters pertaining to address allocation and Internet number assignment), and registration matters (e.g., matters pertaining to registry and registrar services). The SSAC engages in ongoing threat assessment and risk analysis of the Internet naming and address allocation services to assess where the principal threats to stability and security lie, and advises the ICANN community accordingly. The SSAC has no official authority to regulate, enforce or adjudicate. Those functions belong to others, and the advice offered here should be evaluated on its merits.

The contributors to this Report, reference to the committee members' biographies and statements of interest, and committee members' objections to the findings or recommendations in this Report, are at end of this Report.

Table of Contents

1.   Introduction

In September 2009, the Internet Corporation for Assigned Names and Numbers (ICANN) signed the Affirmation of Commitments (AoC) [11] with the U.S. Department of Commerce. As part of this agreement, ICANN commits to:

> "…enforcing its existing policy relating to WHOIS, subject to applicable laws. Such existing policy requires that ICANN implement measures to maintain timely, unrestricted, and public access to accurate and complete WHOIS information, including registrant, technical, billing, and administrative contact information." [11]

As the result of the Affirmation of Commitments, the WHOIS review team[1] is leading a renewed effort to improve the WHOIS service by "assessing the extent to which the WHOIS policy is effective and its implementation meets the legitimate needs of law enforcement and promotes consumer trust."

This Report has two parts: In Section 2, we propose a taxonomy that disambiguates terminology that has been used in discussions related to domain name registration data. In Section 3, we identify features the SSAC believes should be considered in future domain name registration data directory services. Section 4 summarizes four key observations related to the stalled progress on WHOIS thus far, and Section 5 offers recommendations for how to move the conversation forward to fulfill the commitment stated above.

The SSAC notes that the observations and recommendations of this document are not new. The deficiencies of the WHOIS service have been known for some time, and several attempts have been made to remedy those deficiencies through protocols, e.g., WHOIS++ (RFC 1834), RWhois (RFC 2167), and CRISP/IRIS (RFC 3707). In particular, the background material provided in RFC 3707 discusses the limitations of the current WHOIS protocol as defined in RFC 3912 at some length. This SSAC Report does not aim to reproduce those discussions; rather, the aim is to facilitate further efforts to improve registration data retrieval by clarifying the taxonomy associated with the WHOIS service, making observations in light of that taxonomy, and offering a set of recommendations aimed at moving forward efforts to improve the WHOIS service. While WHOIS terminology is used to refer to both Internet naming and numbering meta-data, in this document we focus on clarifying terminology related to domain names only, although we hope that the taxonomy structure and principles will be relevant to Internet numbering WHOIS discussions as well.

Finally, this Report complements the following SSAC advisories on WHOIS:

- In 2003, in SAC 003 [3], the SSAC highlighted the need for improvements to the accuracy and consistency of format of WHOIS data maintained at registrars and

---

[1] See WHOIS Policy Review Team Wiki
<https://community.icann.org/display/whoisreview/WHOIS+Policy+Review+Team>.

registries, and recommended that ICANN modify the registry and registrar contracts to implement the improvements.

- In 2007, an SSAC study, ``Is the WHOIS Service a Source for Email Addresses for Spammers?'' (SAC 023 [4]) found that the appearance of email addresses in responses to WHOIS queries virtually assures that spam will be delivered to these email address. The Committee recommended that registries and registrars implement anti-abuse measures such as rate-limiting, CAPTCHA and Acceptable Use Policy (AUP) to protect WHOIS data from automatic collection.

- In 2008, the SSAC published SAC 027 [5], a comment to the GNSO reiterating the fact that the limitations of the WHOIS protocol and inconsistencies among WHOIS implementations and services contribute to the poor quality of domain name registration data currently available. The SSAC again recommended that ICANN take "aggressive measures with respect to improving registration data accuracy and integrity," using contractual vehicles and non-compliance penalties to improve WHOIS data accuracy and integrity (e.g., archival and restoration).

- The SSAC further recommended that the ICANN community should adopt an Internet standard directory service as an initial step toward deprecating the use of the WHOIS protocol in favor of a more complete directory service. SAC 033 [6] complements SAC027 by providing rationale and adding clarity to the SSAC's recommendations in SAC 027.

2. Taxonomy of Terms

The term "WHOIS" is overloaded, referring to protocols, services, and data types associated with Internet naming and numbering resources, i.e., domain names, Internet Protocol (IP) addresses, and Autonomous System Numbers (ASNs). The ambiguity in terminology further burdens an already challenging set of discussions intended to resolve conflicts related to the evolution of meta-data for Internet naming and numbering. For example, WHOIS can refer to any of the following:

1. The information that is collected at the time of registration of a domain name or IP numbering resource and subsequently made available via the WHOIS Service, and potentially updated throughout the life of the resource;
2. The WHOIS Protocol itself, which is defined in RFC 3912 [2] (which obsoletes RFCs 812 and 954); or
3. The WHOIS Services that provide public access to domain name registration information typically via applications that implement the WHOIS protocol or a web-based interface.

To support on-going efforts at improving the WHOIS framework for DNS-related data, the SSAC proposes the following terms to better distinguish the components of the WHOIS system related to domain names:

Domain Name Registration Data (DNRD) – refers to the information that registrants provide when registering a domain name and that registrars or registries collect. Some of this information is made available to the public. For interactions between ICANN Accredited Generic Top Level Domain (gTLD) registrars and registrants, the data elements are specified in the current Registrar Accreditation Agreement [8]. For country code Top Level Domains (ccTLDs), the operators of these TLDs set their own or follow their government's policy regarding the request and display of registration information.

Domain Name Registration Data Access Protocol (DNRD-AP) – refers to the elements of a (standard) communications exchange—queries and responses—that make access to registration data possible. For example, the WHOIS protocol (RFC 3912) and Hypertext Transfer Protocol (HTTP) (RFC 2616 and its updates) are commonly used to provide public access to DNRD.

Domain Name Registration Data Directory Service (DNRD-DS) – refers to the service(s) offered by registries and registrars to provide access to (potentially a subset of) the DNRD. ICANN Accredited gTLD registries and registrars are required by contracts to provide the DNRD Directory Services via both port 43 and over the web interface [8,9]. For ccTLDs, the TLD registries determine which service(s) they offer.

Using these terms defined above, we offer the following additional terminology, some already commonly used:

- Domain Name Registration Data (DNRD) is composed of different elements. We refer to them as *Domain Name Registration Data Elements (DNRDe)*.

- Where DNRD can be represented in different languages and scripts, it is referred to as *Internationalized DNRD*.

- Where DNRD contains data other than US-American Standard Code for Information Interchange (US-ASCII) (not just the capacity for it), it is referred to as *Localized DNRD*.

- DNRD refers to the entire data collected from the registrant, but potentially only a subset of this data is made available through the *Domain Name Registration Data Directory Service (DNRD-DS)*. This subset of DNRD is referred to as *DNRD-DS Data (DNRD-DSD)*.

- Policies that apply to the domain name WHOIS service could be related to the WHOIS data itself (*DNRD Policy*) or to the directory service (*DNRD-DS Policy*). Policies related to DRND would include 1) data to be included in the DNRD-DS output, 2) annual WHOIS data reminder policy, and 3) policy that requires Registrars to investigate reports of inaccuracy. Policies applicable to the DNRD-DS (DNRD-DS Policy) might include: 1) acceptable terms of use, 2) service levels (e.g., availability, update frequency, response time), 3) query rate limits, 4) data formats, 5) search capabilities, and 6) pricing models for different service levels.

In the remainder of this report the SSAC will use the terminology proposed above, and use the term *WHOIS* to refer only to the WHOIS protocol as defined by RFC 3912.

3. Features for Consideration in Future DNRD-DS

Originally, Internet operators used the DNRD-DS to identify individuals or entities responsible for the operation of a network resource. At present, "its deployments cover a much broader range of information services" (RFC 3912). Today it serves the needs of a variety of stakeholders, including domain name registrants, law enforcement agents, intellectual property and trademark owners and their representatives, businesses, and individual users, as documented in SAC 025.

Many users of DNRD-DS assume that the services provide access to a single centrally managed database that contains registration data for all domain names. In reality, DNRD are stored at disparate locations, especially for "thin" TLD registries, i.e., those that maintain minimal registrant information along with a reference to the registrar that possesses more complete information. Today, DNRD-DS are provided by hundreds of parties who each manage a subset of the registration data and set their own policies regarding access, conventions, and service levels. No single administration defines policy, service, data format, or access requirements for the entirety of TLD registries (gTLD and ccTLD) and registrars, which contributes to considerable and problematic variability in the services offered.

The WHOIS protocol, as defined by RFC 3912, describes a simple exchange of free-form query and response messages between a client and a server over a specific protocol (TCP) and port (43). The only constraint the specification imposes on query and message formats is that they must be terminated with ASCII line feed and carriage return characters. The specification does not define standard formats or mechanisms to signal encodings, which are crucial to support Internationalized Domain Name Registration Data. The WHOIS protocol thus lacks many of the features that a directory service would need to implement the recommendations in SAC 003, SAC 023 and SAC 027.

SSAC believes that the DNRD-DS using the WHOIS protocol today provides only rudimentary functionality, heavily constrained by the limits of the WHOIS protocol and the limitations of the data model. In this section we identify three areas that developers and operators of future directory services should consider.

3.1 Making DRND Representation Consistent

Today, there is considerable variability across various DNRD-DS in data labels, representation, and data format. For example:

1. **Data labels**: Currently there is no consistent labeling system for identifying particular DNRDe's across directory services offered by different entities.

2. **Representation format**: Representation format differs from service to service. For gTLD registries, most use a <label>: <data> format [9]; however, most registrars do not use

similar representation formats.[2] Note that ICANN has proposed a data format for the new gTLD registries and registrars to escrow their DNRD [1].

3. **Data format:** Currently there is no common format for many of the DNRDe. For example, a data element can be presented in any of the following ways by various Domain Name Registration Data Directory Services: 04-May-2008 00:00:00 UTC, 04-May-08, 2008-05-04, or Sun, May 4, 2008. Similarly, a telephone or fax number can be presented as (123) 456-7890, +1.1234567890, 123-456-7890, or +1 123 4567890.

The SSAC thinks that the above variability should be addressed through the specification and implementation of a standards-based, structured, and extensible data model. Such a model would also improve the end user experience when it comes to both internationalized and non-internationalized DNRD. For example:

- A standard data model with consistent data labels would enable end user applications to better localize the output.
- A structured data model that includes additional meta-information (e.g., language, script, etc.) would provide relevant information for translation and/or transliteration tools for Localized DNRD.
- An extensible data model would provide flexibility for registries/registrars who may wish to provide the information in more than one language or script.

3.2   Add Support in DNRD-AP to Accommodate Query and Display of Internationalized Domain Name Registration Data

Operators that provide DNRD-DS using the WHOIS protocol today do not provide standardized support for character sets other than US-ASCII. According to RFC 3912, the WHOIS protocol "lacks [a] mechanism for indicating the character set in use … This inability to predict or express text encoding has adversely impacted the interoperability (and, therefore, usefulness) of the WHOIS protocol."

The Internet Architecture Board also highlighted this issue in RFC 4690:

"In addition to their presence in the DNS, IDNs introduce issues in other contexts in which domain names are used. In particular, the design and content of databases that bind registered names to information about the registrant (commonly described as "WHOIS" databases) will require review and updating. For example, the WHOIS protocol itself has no standard capability for handling non-ASCII text: one cannot search consistently for, or report, either a DNS name or contact information that is not in ASCII characters. This may provide some additional impetus for a switch to IRIS but also raises a number of other questions about what information, and in what languages and scripts, should be included or permitted in such databases." [10]

In the absence of a standardized protocol specification, various registries/registrars have adopted

---

[2] See, for example, "whois –h whois.godaddy.com icann.org".

ad hoc solutions to support Internationalized Domain Name Registration Data. Continued deployment of ad hoc solutions will inevitably lead to widespread inconsistent user experience and interoperability conflicts, further reducing the utility of DNRD.

The SSAC observed in SAC 037 [7] that as the adoption of IDNs becomes more prevalent, Internet users will expect to be able to register domain names as well as domain owner names and addresses in their native languages, using familiar scripts (character sets). This adoption is already well underway, increasing the priority of defining a standard means to accommodate the submission and display of Internationalized Domain Name Registration Data.

3.3   Add Support for Authentication and Authorization

Current directory services generally offer public and largely anonymous access to DNRD, requiring no identity assertion or authentication before allowing access to the data.

A standardized authentication and authorization access control framework would support the evolving range of needs for access to different granularities of data, as well as help prevent data-mining abuses [6]. An authentication framework allows DNRD-DS providers to use stronger authentication methods before granting access to personal data of natural persons and simpler authentication methods for access to DNRD associated with business entities. Fine-grained (per DNRDe) access controls can support differentiated access to registration data, e.g., making sponsoring registrar data elements public but restricting access to contact data to authenticated users.

Such a framework would provide mechanisms to define and implement a range of verification methods, credential services, and access control capabilities, enabling those who develop and implement policies to select, configure, and apply these mechanisms.

4.   Summary Observations

As the ICANN community invests effort in meeting the WHOIS-related requirements in the Affirmation of Commitments, including accommodating the legitimate needs of law enforcement and promoting consumer trust. We summarize four key observations to guide subsequent discussions, development, and standardization efforts:

1. The use of ambiguous terminology has for many years inhibited productive discussion of DNRD-related issues and pursuit of consensus on how to improve the DNRD-DS.

2. There is currently no uniform data model that exists for DNRD which unduly complicates data submission, handling, exchange, and access.

3. The DNRD directory services implemented using the WHOIS protocol today do not support character sets other than US-ASCII in a standard manner. As the adoption of IDNs increases, continued deployment of ad hoc solutions will inevitably lead to widespread inconsistent user experience and interoperability conflicts, further reducing

the utility of DNRD and associated services.

4. Current directory services generally offer public and largely anonymous access to DNRD, making it difficult to satisfy an evolving range of legitimate needs for access to different granularities of data.

5.   Recommendations

Recommendation 1: The ICANN community should adopt the terminology outlined in this report in documents and discussions, in particular:

- **Domain Name Registration Data (DNRD).** The data that domain name registrants provide when registering a domain name and that registrars or registries collects.

- **Domain Name Registration Data Access Protocol (DNRD-AP).** The components of a (standard) communications exchange—queries and responses—that specify the access to DNRD.

- **Doman Name Registration Data Directory Service (DNRD-DS).** The service(s) offered by domain name registries and registrars to implement the DNRD-AP and to provide access to DNRD-DSD.

Additional terminology includes "DNRDe," "DNRD Policy," "DNRD-DS Policy," "Internationalized DNRD," and "Localized DNRD." The term "WHOIS" should only be used when referring to the protocol as currently specified in RFC 3912.

Recommendation 2: The ICANN community should evaluate and adopt a replacement domain name registration data access protocol that supports the query and display of Internationalized DNRD as well as addressing the relevant recommendations in SAC 003, SAC 027 and SAC 033.

Recommendation 3: The ICANN community should develop a uniform and standard framework for accessing DNRD that would provide mechanisms to define and implement a range of verification methods, credential services, and access control capabilities.

6.   Acknowledgments, Statements of Interests, and Objections and Withdrawals

In the interest of greater transparency, these sections provide the reader information on three aspects of our process. The Acknowledgments section lists the members who contributed to this particular document. The Statements of Interest section points to the biographies of the Committee members and any conflicts of interest, real, apparent, or potential, that may bear on the material in this document. The Objections and Withdrawals section provides a place for individual members to disagree with the content of this document or the process for preparing it.

6.1   Acknowledgments

The committee wishes to thank the following SSAC members for their time, contributions, and review in producing this Report.

Bruce Tonkin
Dave Piscitello
David Conrad
Frederico A C Neves
Jeff Bedser
Jeremy Hitchcock
Jim Galvin
KC Claffy
Mark Kosters
Mark Seiden
Merike Kaeo
Patrik Fältström
Ram Mohan
Rick Wesson
Sarmad Hussain
Steve Crocker
Steve Sheng
Warren Kumari

6.2   Statements of Interest

SSAC member biographical information and Statements of Interest are available at:
http://www.icann.org/en/committees/security/biographies-25mar11-en.htm.

6.3   Objections and Withdrawals

There were no objections or withdrawals.

7.   References

1. Arias, F. and S. Noguchi. (2011) Domain Name Data Escrow Specification. Internet Draft <draft-arias-noguchi-registry-data-escrow-02>.

2. Daigle, L. (2004) "WHOIS Protocol Specification", RFC 3912.

3. ICANN Security and Stability Advisory Committee (SSAC). (2003) *WHOIS Recommendation of the Security and Stability Advisory Committee* (SSAC publication No. 003). Retrieved from http://www.icann.org/en/committees/security/sac003.pdf.

4. ICANN Security and Stability Advisory Committee (SSAC). (2007) *Is the WHOIS Service a Source for email Addresses for Spammers?* (SSAC publication No. 023). Retrieved from http://www.icann.org/en/committees/security/sac023.pdf.

5. ICANN Security and Stability Advisory Committee (SSAC). (2008a) *SSAC Comment to GNSO regarding WHOIS studies* (SSAC publication No. 027). Retrieved from http://www.icann.org/en/committees/security/sac027.pdf.

6. ICANN Security and Stability Advisory Committee (SSAC). (2008b) *Domain Name Registration Information and Directory Services* (SSAC publication No. 033). Retrieved from http://www.icann.org/en/committees/security/sac033.pdf.

7. ICANN Security and Stability Advisory Committee (SSAC). (2009a) *Display and usage of Internationalized Registration Data: Support for characters from local languages or scripts* (SSAC publication No. 037). Retrieved from http://www.icann.org/en/committees/security/sac037.pdf.

8. Internet Corporation for Assigned Names and Numbers (ICANN). (2009) *Registrar Accreditation Agreement.* Marina Del Rey, CA: ICANN. Retrieved September 14, 2011, from http://www.icann.org/en/registrars/ra-agreement-21may09-en.htm#2.

9. Internet Corporation for Assigned Names and Numbers (ICANN). (2011) *Registry Agreements.* Marina Del Rey, CA: ICANN. Retrieved September 14, 2011, from http://www.icann.org/en/registries/agreements.htm.

10. Klensin, J., Faltstrom, P., Karp, C., and IAB, "Review and Recommendations for Internationalized Domain Names (IDNs)", RFC 4690, September 2006.

11. United States Dept. of Commerce and Internet Corporation for Assigned Names and Numbers (ICANN) (2009). Affirmation Of Commitments By The United States Department Of Commerce And The Internet Corporation For Assigned Names And Numbers. Available at< http://www.icann.org/en/documents/affirmation-of-commitments-30sep09-en.htm>.