Whois Recommendation of the Security and Stability Advisory Committee [SAC003]

1 December 2002
Version 2 - 7 February 2003

[.pdf version]

Whois Recommendation of the Security and Stability Advisory Committee

SAC 003.1
Document 003 Version 2
February 7, 2003

Table Of Contents

Executive Summary

The port 43 Whois protocol has traditionally been used by the Internet community to identify and provide contact information for the person or organization responsible for many Internet resources, for example, a domain name or an IP address. It has been successfully used in a cooperative manner for situations such as informing a person or organization of inappropriate use of their resource (security), or incorrect configuration of their resource (stability). Whois data is thus important for the security and stability of the Internet as the administration and control of Internet resources is widely distributed.

The accuracy of Whois data used to provide contact information for the party responsible for an Internet resource must be improved, both at the time of its initial registration and at regular intervals. Whois records known to be false or inaccurate must be frozen or held until they can be updated or removed. Whois records that have information that can not be validated may be frozen or held until it can be verified.

In order for Whois data to be readily available it must be both accessible and usable by automatic tools. To be accessible the Whois protocol must be updated to support the

recent shift in the architecture to separate the functions of the registry and the registrar. This shift has made it impractical to support searching and frequently makes it difficult to find Whois services. To be usable the data returned by Whois services must in be a common format.

However, being accessible and usable must also protect a registrant's privacy. Many countries require that personal information is protected but in addition registrants may wish to discourage the unintended, undesirable, and otherwise unwanted uses of their Whois data. In particular, it is widely believed that Whois data is a source of email addresses for the distribution of spam. Methods must be developed to discourage the harvesting or mining of Whois information.

1. Introduction. The port 43 Whois protocol is described in RFC 954 and amounts to the following:

```
C:.*\r\n

S:.*\r\n [close socket]
```

This is probably the simplest protocol described by the IETF. While it is the simplest it is also extremely flexible. Three main types of Internet registries use the protocol. The Number registries publish information about IP addresses and Autonomous System numbers, the Routing Registries publish information about routing policy, and the name registries publish information about domain name delegations. This recommendation only applies to the use of the Whois protocol to provide access to contact information for the person or organization responsible for Internet resources.

1. Accuracy

There are two principal reasons to maintain accurate contact information in Whois records: technical and legal. The technical rationale is that if there are problems with or abuse originating from a resource (e.g., a domain name, route, or IP address) the Whois entry for the resource is the only source for finding the responsible party. For legal problems accurate postal addresses are required for serving court papers to the responsible party.

ICANN does require Name Registries and Registrars to publish information about domain name registrations using the port 43 Whois. Unfortunately, the information published by name registries and registrars is often cited as incorrect, invalid or false, or out of date. It has been suggested that there are potentially a significant number of records with addresses that do not exist and telephone numbers that can not exist.

One apparent reason for a registrant to falsify a Whois entry appears to be privacy. In most countries there exist privacy protections but if a private person wants to own an Internet domain name ICANN requires a physical mailing address and a voice telephone number.

While we expect Registries and Registrars to take steps to prevent false information in registrations we must also encourage the development of mechanisms to ensure that a registrant's privacy is protected. When we discuss openness and transparency we should not have the registrant's home address and telephone number in mind.

The IETF PROVREG working group -- its charter can be found here <http://www.ietf.org/html.charters/provreg-charter.html> -- is developing a specification of the requirements for a protocol that enables a registrar to access multiple registries and will develop a protocol that satisfies those requirements. The requirements are currently published as RFC3375 <http://www.ietf.org/rfc/rfc3375.txt>.

2. Timeliness

Unassociated contact data are rarely cleaned from a registrar's database. Indeed some registrars actively marketing their Whois data as a source of revenue have little incentive to remove old and unrelated contact data from their database.

Users of Whois data need to know both how current the information presented is when it is received and the process by which the information was validated or confirmed. Contact information must include a "Last Verified Date" that reflects the last point in time at which the information was known to contain valid data and a reference to the process by which the data is both initially and regularly verified. The process should be readily available on the web site of the Registry and Registrar.

3. Searchability

In 1999, before the introduction of the SRS, the domain name registry had indexes for all kinds of elements available in the Whois. The registry could answer questions about searches for contact names, hosts, or domain names because the registry had the actual data and could create indices for many types of queries.

With the advent of ICANN and the separation of the Registry and Registrar functions, the best, centralized service a Registry can provide is basic referrals to the Registrar, which resulted in a fractured Whois space. Without the data no centralized index can be created and without the index no centralized search can be performed.

The IETF CRISP working group -- its charter can be found here <http://www.ietf.org/html.charters/crisp-charter.html> -- is developing requirements for a revised Whois-like service that will support this disjoint Whois space and distributed indices. A URL to the current version of their requirements can be found in their charter. As of this writing the current version is <http://www.ietf.org/internet-drafts/draft-ietf-crisp-requirements-00.txt>.

4. Machine Readability and SPAM

It is widely believed that the Whois data is a source of email addresses for the delivery of SPAM and other unsolicited and otherwise unwanted email messages. Consequently, many Registrars have started offering their Whois data in random formats to deter harvesting. This is unfortunate because a common format is necessary to ensure that the data is readily accessible and understandable when it is needed. We must encourage not only the use of a common format but the development of mechanisms to prevent the harvesting and mining of Whois data.

2. Recommendations

- The accuracy of Whois data used to provide contact information for the party responsible for an Internet resource must be improved, both at the time of its initial registration and at regular intervals. Whois records known to be false or inaccurate must be frozen or held until they can be updated or removed. Whois records that have information that can not be validated may be frozen or held until it can be verified.
- A standard format for Whois data must be developed.
- Whois data must contain a "Last Verified Date" that reflects the last point in time at which the information was known to contain valid data. It must also contain a reference to the data verification process.
- A Whois service that supports searching in the current architecture of distributed indices and separated registry and registrar services must be developed.
- A publicly available list of publicly available Whois servers must be available using a widely known and available resource, e.g., a web page or DNS SRV records.
- Whois services must provide mechanisms to protect the privacy of registrants.
- A Whois service must discourage the harvesting and mining of its data.

3. Implementation Plan

- ICANN should modify the Registry and Registrar contracts to require the recommendations as described in the previous section.
- Registrars, registries, and all interested parties should be encouraged to support and participate in the activities of the CRISP and PROVREG working groups of the IETF.

4. Progress Measurement

Annually ICANN should publicly report the following:

- Report on Updated contracts for compliance to the recommendations.
- Report on Whois accuracy measurements from a statistical survey and complaints received from Internic.net reports.

5. Acknowledgements

Special thanks go to Rick Wesson who led the effort to produce this document with the members of the Security and Stability Advisory Committee.

Alain Aina (Consultant)

Jaap Akkerhuis (SIDN)

Doug Barton (Yahoo!)

Steve Bellovin (ATT)

Rob Blokzijl (RIPE)

David Conrad (Nominum)

Steve Crocker (Shinkuro), Chair

Mark Kosters (VeriSign)

Allison Mankin (ISI)

Ram Mohan (Afilias)

Russ Mundy (Network Associates Laboratories)

Jun Murai (Keio University)

Frederico Neves (registro.br)

Ray Plzak (ARIN)

Doron Shikmoni (ForeScout, ISOC-IL)

Ken Silva (VeriSign)

Bruce Tonkin (Melbourne IT)

Paul Vixie (ISC)

Rick Wesson (Alice's Registry)

Support for the committee is provided by Jim Galvin (eList eXpress).

Prior draft:

Version 1 - 1 December 2002

Comments concerning the layout, construction and functionality of this site should be sent to webmaster@icann.org.

Page Updated 21-Mar-2003