

Technical Evolution of the Whois Service

Preliminary Draft, 15 November 2010

Executive Summary

This preliminary discussion paper, prepared by ICANN staff, analyzes the technical shortcomings of the current Whois service¹ and identifies three potential options to address these technical deficiencies. While there may also be other options to consider, in this paper staff specifically examined the following: 1) extending the existing WHOIS protocol; 2) migrating from WHOIS to the IRIS protocol; and 3) migrating from WHOIS to a HTTP-based Representational State Transfer protocol based service (“RESTful Whois Service”, or RWS). We examined each of the options, how each might address the deficiencies, and we list some possible concerns regarding implementation. Note that the paper is intended to initiate a discussion on technical options, and is not intended either as a technical recommendation or as a policy document. We are earnestly seeking feedback from the community on our analysis as well as whether there are other potential technical options for improving WHOIS that should also be considered.

Introduction

When people refer to Whois, they may mean different things. There are, at least, three different uses of the word "Whois" by the ICANN community:

- (1) The WHOIS protocol - RFC 3912.
- (2) The Whois "service" - which provides information via both the WHOIS protocol and web-based interfaces.
- (3) The data collected at registration and made available via the Whois service per the Registrar Accreditation Agreement (RAA) and the gTLD Registry Agreements.

This document *solely* focus on improving (1), the WHOIS protocol.

Created in the 1980s, Whois began as a service used by Internet operators to identify and contact other individuals operating a network resource on the ARPANET. The Whois service has since evolved into a tool used for many purposes, such as determining whether a domain name is available for registration, identifying the registered users of Internet (IP) address allocation blocks, identifying the registrant of a domain name that has been associated with malicious activities, contacting domain name registrants on matters related to trademark protection, verifying online merchants, etc.

As usage of Whois evolved, few changes have been made to the protocol. There are increasing community concerns that the current WHOIS protocol does not meet the community’s current needs. These are noted in recent reports from ICANN’s Security and Stability Advisory Committee (SSAC) [4, 5, 6 and 7], in reports of other ICANN supporting organizations and advisory committees [3] and by external sources [8]. At a high level, these technical deficiencies are:

1. **Lack of standardization:** The WHOIS protocol (RFC 3912 [2]) is very simple. It describes exchanges of queries and messages between a client and a server over TCP in a specific port

¹ “Whois” is used in reference to the service in general and “WHOIS” in caps is used when referring to the RFC 3912 and older protocol.

(43). It does not define query or response formats or encoding, nor does it have a schema for replies and error messages. Such decisions are left to the implementers, e.g., registrars and registries; this often results in different query syntaxes, output formats, output encodings, and error messages. The resulting variability across clients and servers detracts from the quality and usability of Whois.

2. **Lack of support for internationalised registration data and domains:** According to WHOIS protocol specification, “*The WHOIS protocol has not been internationalised. The WHOIS protocol has no 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.*”[2]
3. **Lack of authentication and access control mechanisms:** Users or applications access Whois services anonymously, requiring no identity assertion, credentialing or authentication. The lack of authentication mechanisms inhibits adoption of effective user or group level access controls, auditing, or privacy measures, features that a typical directory system would have [7]. Few methods are used to restrict access to Whois servers listening at port 43 other than IP address-level control.

As a result of these deficiencies, the current Whois services have less than optimal reliability and accuracy properties, and are not as useful as they could be.

Past Efforts to Improve Whois

There have been several attempts in the past to improve the WHOIS protocol.

In 1993, the US National Science Foundation (NSF) created the Internet Network Information Center (InterNIC), giving AT&T a contract to operate directory services. The end goal was to create a “directory of directories,” moving information from Whois and other access protocols to an X.500 directory, but AT&T’s contract with the NSF expired before this ever materialized [8].

In 1994, Network Solutions (now VeriSign) developed the Referral Whois (RWhois) protocol (RFC 1714, RFC 2167), which was designed to address the lack of hierarchy in Whois. RWhois never replaced Whois, but some US ISPs still use it today [8].

In 1995, an IETF working group (Whois and Network Information Lookup Service Working Group (WNILS)), along with the Canadian company Bunyip led an effort to improve on the protocol with Whois++ (RFC 1834). Whois++ expanded and defined the standard for WHOIS types of services, and addressed issues associated with the variations in access and provide a consistent and predictable service across the network. However, Whois++ never saw wide deployment.

In 1998, MCI briefly promoted the WHOIS specialization of RFC 2345 for domain name and company name retrieval. The proposed change uses URI structure to locate domain names and company names (e.g. WHO://microsoft.com/ in web browser to find the Whois information for microsoft.com). This proposal never saw wide deployment either.

Finally in 2005, the IETF CRISP working group standardized the Internet Registry Information Service

(IRIS) as a replacement for the WHOIS protocol. IRIS is a directory service that provides additional functionality that the current WHOIS lacks. However, as of this writing, we have seen little adoption of IRIS. In this paper we regard IRIS as one of the alternatives to improve WHOIS, and consulted with IRIS RFC writers about its lack of adoption.

Options to improve Whois:

Staff has identified three potential options to improve Whois: extend the current WHOIS protocol, migrate from the current protocol to the IRIS protocol, or migrate from the current protocol to RWS. We note that there may be additional options to consider and we welcome community ideas in that regard. In this section, we introduce each of the options we've identified to-date, examine how each might address the deficiencies raised above, and list some concerns (where identified) for implementation. Again, this initial analysis is offered as a starting point for further discussion and is not intended as a technical recommendation or a policy document.

Extending the existing WHOIS protocol:

The deficiencies of the WHOIS protocol can be addressed the following way:

1. **Standardization:** A revised and extended WHOIS specification could be developed in the IETF with participation from interested parties. The new specification could include version selection, query types, response formats, etc. The new specification could also standardize error messages. Once the new specification had gone through the IETF standards process, new implementations might gradually replace existing implementations, and the RFC 3912-based legacy protocol could be eventually deprecated.
2. **Support for internationalised registration data and domains:** Using the same approach above, the new specification for WHOIS could include a mechanism for signaling character encodings. For example, a signaling mechanism can be defined to optionally select “legacy” (US-ASCII) or MIME (Multipurpose Internet Mail Extensions) – the approach used to extend SMTP to support email delivery in encodings other than US-ASCII, or some other mechanism.
3. **Authentication and access control mechanisms:** These features could also be added to an extended WHOIS. For example, the protocol could have support for TLS/SSL transport to protect credentials, or it could use a challenge-response process to authenticate users and enable access control mechanisms.

Implementation Considerations related to Extending WHOIS:

- To update the WHOIS protocol interested parties should determine if there is sufficient interest, and then propose the internet-draft for standards track consideration at the IETF. Staff is mindful that the proposed IRIS protocol, discussed further below, has already gone through the IETF process, so any alternative proposal should speak to the rationale behind pursuing other options, whether they be those described in this paper or others that are not yet identified. The proposal should include output schema, mechanisms for signaling character encodings, query format, standardized error messages, support for authentication and authorization.
- Extending the protocol will require a method of signaling “version” so that the current client and server implementations continue to operate while new implementations are deployed (backwards compatibility).
- Extending the WHOIS protocol would require new client and likely obsolete the current client-

base.

Migrating to IRIS:

The Internet Registry Information Service (IRIS) protocol was developed by the IETF as a successor to WHOIS. IRIS is a directory service that provides additional functionality that the current WHOIS lacks and addresses the deficiencies of WHOIS in the following ways:

1. **Standardization:** The IRIS protocol specifies a well-defined structure for query and result sets.
2. **Support for internationalised registration data and domains:** IRIS explicitly supports internationalization and localization. It uses XML for both query and response, which can accommodate different encodings, thus supporting multiple languages.
3. **Authentication and access control mechanisms:** IRIS supports authentication services through its application-transport layer protocols (BEEP, XPC). These protocols define the mechanisms for authentication, message passing, connection and session management, etc.

Implementation Considerations related to adopting the IRIS Protocol:

Although IRIS addresses many of the technical deficiencies of WHOIS identified above, there has been little adoption of IRIS in the five years since its standardization². Staff consulted with IRIS authors, developers, and pilot program operators³ who shared several reasons for the lack of adoption:

1. IRIS is a complex protocol. It has three layers: registry-specific (domain, IP address, etc.), common registry (IRIS), and application-transport (BEEP, IRIS-LWZ, XPC). Each layer may consist of one or more protocols.
2. IRIS requires an application transfer protocol (e.g., BEEP [9], XPC [10]) for correct operation over existing transport protocols (UDP or TCP). The former are not commonly used, which further increases the barrier for implementation.
3. There are no available client implementations of the full IRIS protocol, only DCHK, for domain availability checks exclusively.
4. There is a lack of full IRIS server implementations available for use, from either open-source or commercial developers, only IRIS-DCHK from DENIC.

Migrating to RWS:

Representational State Transfer based Whois service [1] (RWS) is offered over the HTTP protocol and conforms to the REST architectural approach [11]. The choice of HTTP as a transport is intentional, so that the services built on top of HTTP⁴ can leverage popular web server infrastructures and administrative experience involved in web applications. REST describes well-known and widely adopted design patterns and architecture, for example the Atom Publishing Protocol for publishing to

² Verisign and RIPE NCC created pilot reference implementations that are not available anymore. DENIC has a working implementation for IRIS-DCHK.

³ We consulted Andy Newton and Marcos Sanz the authors of the IRIS RFCs and pilot implementers. We also consulted with Vincent Levigueron from AFNIC (.FR), leader implementer of an IRIS-DCHK implementation.

⁴ By adopting HTTP we are not suggesting that access to Whois would be exclusively done through a web browser; in particular, client applications or automation can use command line tools such as *curl* or *wget* to formulate queries and process responses.

blogs.

RWS supports several features the community considers necessary or beneficial for Whois services in the following way:

1. **Standardized output and error format:** The base response output format is XML, which when paired with a well-defined schema would allow for automated processing.
2. **Support for internationalisation:** RWS has complete support for internationalised registration data, as well as IDNs with U-labels, by using the XML data format, which contains inherent capability to support multiple character encodings to support internationalization.
3. **Authentication and access control:** HTTP, the transport for RWS already supports authentication, and by means of using these capabilities, RWS makes technically possible to implement granular permissions over registration data if required.

RWS offers the following additional benefits:

- **Similar requirement:** Current gTLD registries and registrars are required to, and already offer Web-based Whois; therefore requiring RWS would not be a new service for them, but a new specification of an already required service. It is also worth to mention that RWS specification will allow a variety of possible implementations that can be adapted to the reality of each registry, just as it is with most Web services.
- **Addressable Whois Service:** RWS requires the use of a URI/URL standard structure for each object/resource. This has the additional benefit of providing a widely recognized manner to refer unambiguously to objects in Whois.
- **Increased Usability:** Some of the inherent capability of the HTTP protocol (such as redirects) can be used to provide additional functionality such as automatic referrals to more specific WHOIS data sources without requiring specialized parsing by the client.
- **Authenticity of Origin:** RWS provided over HTTPS offers confidence in the origin of the information.
- **Leverage existing infrastructure and expertise:** RWS is HTTP-based and can be supported using popular web server infrastructures. Web administration is a skill-set and resource likely already commonplace inside registries and registrars. Similarly, RWS can benefit from existing technology to implement load-balance servers, cache answers to minimize network traffic, etc.
- **RWS Proxy:** Another possibility that the use of HTTP allows which may be interesting to explore, is the use of proxy capabilities. Registries and registrars may be able to allow a trusted third party (maybe with previous agreement), e.g., ICANN to offer proxied access to the RWS content. Registries and registrars would still be able to apply their own access controls, even IP-address level, to restrict those clients being proxied by the third party. For ICANN or another third party this would mean the possibility of offering a one-stop shop for Whois information for all TLDs that have an agreement with the third party.
- **Integrated Whois service:** As mentioned before, currently all the gTLD registries and registrars offer WHOIS and web-based Whois; two services. RWS allows the possibility to integrate the two.
- **Existing implementations:** Currently ARIN has a production quality RWS, RIPE and ICANN have pilot implementations.

Implementation Considerations related to **implementing RWS**:

Staff has identified the following potential issues related to implementing RWS. We are seeking further input from community experts:

- At this moment RWS is not standardized yet, as a result, various implementations may have differing specifications.
- It is unclear whether there is sufficient stakeholder interest to pursue development of a technical standard, even on an exploratory or experimental basis. It will be important to socialize this option with the technical community, including those with insight into standards development work that has occurred with WHOIS to-date.

Comparison of Options

In this section, we compare the three options discussed above based on available features, cost of implementation, available resources, and extensibility.

Available features: In summary, IRIS seems to offer the most features compared with RWS or extending the Whois protocol. IRIS is already an IETF standard. RWS can address all the deficiencies in Whois, but it needs to be standardized. Extending WHOIS can address some of the deficiencies identified, but would require significant protocol change to add authentication and access control capabilities.

Cost: We have not conducted a formal cost analysis or comparison, but we offer the following anecdotal assessment. Due to the lack of available client and server implementations for IRIS and the complexity of the protocol, implementing IRIS is likely to be costly for registrars and registries. RWS is likely to be less costly than IRIS, due to the wide availability of clients, using a well-known and widely adopted architectural standard. Recently staff implemented a pilot domain registry RWS. The pilot service is a prototype-quality implementation. It was developed in about 80 man-hours, including the design of the basic elements of the service. IANA has also done a RWS pilot implementation for the root zone, ARPA and .INT, with similar costs to staff's pilot implementation. However, part of the cost for wide RWS deployment is the standardization process, and we have not assessed the cost to turn existing implementations into production quality software. Finally, the cost of extending WHOIS appears to be approximately equal if not greater than RWS, as it would need to go through standardization process, as well as updating clients and servers.

Extensibility: IRIS is a layered protocol and each layer can be extended. A similar case can be made for RWS. Both IRIS and RWS are based on XML schemas and support versioning, so the data model can be easily extended. In comparison, extending WHOIS is more difficult.

Readily Available Resources: RWS can use the web browser and command-line programs as *curl* or *wget* as client, it can also benefit from existing technology to implement load-balance servers, cache answers to minimize network traffic, etc. There are no IRIS clients available, IRIS uses non widely-used protocols for transport; therefore few people would know how to write a client. The existing WHOIS client is likely to be made obsolete once the protocol is updated, therefore requiring new or updated clients.

In summary, IRIS has the most features and easily extensible, but it is costly to implement and there are no readily available resources. Extending Whois does not address all the technical deficiencies, is not extensible. From our assessment, RWS has a number of features that appear to address the deficiencies of WHOIS, would be extensible to accommodate future improvements, which seem to be achievable at a reasonable low cost.

Next steps

Staff is seeking feedback from registries, registrars, RIRs, users and other interested parties regarding this preliminary draft discussion paper. In particular, we would like to hear feedback on the following questions:

- 1) Have we correctly summarize the problems of WHOIS protocol? Are there any problems of the protocol that we missed?
- 2) Have we correctly identified the potential solution space? Are there any other viable solutions that we have not identified?
- 3) For the solutions that we identified, is our analysis correct? Are there any other factors we did not identify?
- 4) Which of the three identified options is the most adequate and why?

For more information:

Please contact Francisco Arias (Francisco.arias@icann.org) or Steve Sheng (steve.sheng@icann.org)

Your input is important and would be greatly appreciated.

References

1. American Registry for Internet Numbers (ARIN). (2010) WHOIS-RWS API Documentation. Retrieved October 21, 2010, from https://www.arin.net/resources/whoisrws/whois_api.html
2. Daigle, L. (2004) WHOIS Protocol Specification, RFC 3912.
3. ICANN Generic Names Supporting Organization (GNSO). (2010) *Inventory of Whois Service Requirement Final Report*. Marina Del Rey, CA: ICANN. Retrieved October 21, 2010, from <http://gns0.icann.org/issues/whois/whois-service-requirements-draft-final-report-31may10-en.pdf>
4. 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>
5. 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>
6. 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>
7. 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>
8. Newton, A. (2006) *Replacing the WHOIS Protocol: IRIS and the IETF's CRISP Working Group*. Internet Computing, IEEE Volume: 10 Issue: 4 July-Aug. 2006 Page(s): 79-84
9. Newton, A. and M. Sanz, "Using the Internet Registry Information Service (IRIS) over the Blocks Extensible Exchange Protocol(BEEP)", RFC 3983, January 2005.
10. Newton, A., "XML Pipelining with Chunks for the Information Registry Information Service (XPC)", RFC 4992, August 2007.
11. Roy Thomas Fielding (2000). *Architectural Styles and the Design of Network-based Software Architectures* (chapter 5). Dissertation: University of California, Irvine. Accessed from: <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>