

Report on Public Suggestions on Further Studies of WHOIS

Prepared by:

Lorrie Cranor and Liz Gasster

February 25, 2008

On 31 October 2007, the Generic Names Supporting Organization (GNSO) Council concluded that a comprehensive, objective and quantifiable understanding of key factual issues regarding the gTLD WHOIS system will benefit future GNSO policy development efforts, and plans to ask ICANN staff to conduct several studies for this purpose. Before defining the details of these studies, the Council solicited suggestions from the community for specific topics of study on WHOIS that community stakeholders recommend be conducted. The GNSO Council will be considering the public input received on further WHOIS studies, and at an upcoming meeting will direct staff to develop costs and estimated time frames for the study proposals they would like pursued. Following that input, the Council will identify the specific studies, if any, which should be conducted.

Twenty-five suggestions for WHOIS studies were submitted (note that each submission is referred to by the number assigned on the WHOIS public comments chronological index, <http://forum.icann.org/lists/WHOIS-comments-2008/>, except the comment submitted by Steven Metalitz on 14 February, which is referred to in this summary and in the chronological index as the “Metalitz comment”). Three of the suggestions are outside the scope of the current study effort. One of the suggestions does not propose a study but offers relevant comments, which are discussed below.

In the 31 October resolution on WHOIS, the GNSO Council recognized that in the process of selecting studies to conduct, it should consider not only the costs and feasibility of each study, but also the impact that additional data will likely have on policy development efforts. Thus, in reviewing the following, staff suggests that the Council consider whether proposed studies would provide data that is likely to influence decision-making across the various ICANN constituencies. Where there are disagreements over key factual issues relevant to a policy decision, a study that can provide objective, factual information can influence policy makers’ decisions. However, when disputed facts are not central to policy makers’ decisions, additional data may not influence these decisions. Study suggestion #10 suggests that spending money on additional studies will not further inform the debate at this point as the problem is mainly political. The submitter argues that the interests of individual registrants are not given appropriate weight in the policy process. The Council should thus consider the impact that additional data will likely have on policy development efforts.

In this document we group the 21 within-scope study suggestions into seven topical categories and discuss the relative costs and benefits associated with the studies in each category:

1. WHOIS misuse

2. Compliance with data protection laws and registrar accreditation agreements
3. Availability of privacy services
4. Demand and motivation for use of privacy services
5. Impact of WHOIS data protection on crime and abuse
6. Proxy registrar compliance with law enforcement and dispute resolution requests
7. WHOIS data accuracy

The first two categories of studies examine misuse and legal compliance problems that motivate proposals to require WHOIS data protections. The third category examines the availability of privacy services in the marketplace to understand the extent to which services that provide WHOIS data protections are already available at a reasonable cost. The fourth category examines the demand and motivation for use of privacy services to determine the fraction of registrants who would legitimately benefit from them and the extent to which they are currently being used for illegitimate purposes. The fifth category examines the impact of WHOIS data protection on reported crime and abuse incidents. The sixth category examines whether procedures for revealing the identity of an underlying registrant or relaying communications to registrants effectively meet the needs of law enforcement and dispute resolution processes. Finally, the last category examines questions of WHOIS data accuracy.

WHOIS misuse studies

Four proposals (suggestions #1, #14, #15 and #21) suggest that ICANN study misuse of WHOIS data to determine the connection, if any, between WHOIS and illegal activities. These studies will help establish the extent and nature of problems caused by unprotected WHOIS data.

Study Suggestion Number 1: 1) Gather data on WHOIS misuse from consumer protection bureaus and other entities who maintain data on misuse incidents reported by registrants and 2) survey a random sample of registrants in each gTLD and selected ccTLDs.

Study Suggestion Number 14: Create a set of new email addresses, use half of them to register domain names, and monitor all for spam for 90 days to determine how much WHOIS information contributes to spam.

Study Suggestion Number 15: Create a set of new email addresses, use them to register new domain names at registrars that allow and disallow port 43 WHOIS queries, and monitor all for spam to determine the extent to which port 43 WHOIS queries contribute to spam.

Study Suggestion Number 21: Survey registrars and human rights organizations to determine how WHOIS is being used in ways that seem to have no bearing on the security and stability of the DNS.

1 and 21 propose to survey registrars and other parties who may keep records of misuse incidents. 1 also proposes a survey of registrants. These proposed studies may shed some light on the extent and type of misuse of WHOIS data. However, it will be difficult to gather representative data as not all cases of abuse are reported. In addition, it is not always possible to confirm that misused data was obtained from WHOIS, as this information may be available from other sources. A registrant survey is likely to receive disproportionate responses from registrants who believe their WHOIS information has been abused. Nonetheless, the above studies may result in useful qualitative data about the nature of misuse and provide a rough quantitative estimate of the extent of misuse. Surveying those who already keep track of abuse incidents is likely to be a relatively low-cost approach. The registrant study is likely to be more expensive if done on a large scale, and seems less likely to result in useful data.

14 and 15 focus specifically on spam and propose studies in which new email addresses are created and used to register domains to determine how much WHOIS information contributes to spam. 15 compares the amount of spam received as a result of registering a domain at registrars that allow and prohibit port 43 WHOIS queries. These studies should result in fairly accurate quantitative data. However, 14 is quite similar to the October 2007 SSAC study “Is the WHOIS service a source for email addresses for spammers?” and would not likely contribute new information. If port 43 queries are of interest from a policy perspective, study 15 should provide reliable data to inform that discussion.

Compliance with data protection laws and Registrar Accreditation Agreement

Three proposals (#16, #22, #23) suggest studies of data protection laws and how they relate to ICANN and ccTLD policies for WHOIS and other personal domain name registration data. The results of these studies would provide insights into whether current policies and practices comply with national data protection laws. Study #24 proposes to determine whether registrars’ registration agreements comply with the Registrar Accreditation Agreement. The results of these studies may suggest the need for policy changes or better enforcement measures. These studies involve a legal review and analysis of publicly available documents, and possibly a questionnaire distributed to top ccTLDs. Data collection for these studies should be relatively inexpensive but requires willingness on the part of ccTLD operators to participate. Analysis will require legal expertise.

Study Suggestion Number 16: Conduct legal analysis under the laws of a variety of jurisdictions of the terms of various registrars' registration agreements concerning data collection and disclosure and their process for collecting such data and obtaining consent.

Study Suggestion Number 22: Survey top 25-30 ccTLDs to determine the extent to which ccTLD WHOIS policies reflect national data protection laws and priorities.

Study Suggestion Number 23: Conduct a legal comparison of national data protection laws to determine legal requirements relevant to the protection of registrant information.

Study Suggestion Number 24: Obtain a representative sample of registrars' terms and conditions to determine what percentage of sampled registrars is appropriately obtaining agreement to all of the terms required under Section 3.7.7 of the RAA.

Availability of privacy services

Proposals #2 and #5 would study the current availability of privacy services, determine their costs and extent of use, and correlate use with cost and other features.

Study Suggestion Number 2: 1) Gather data on types of privacy services offered through manual review of websites offering registration services and survey of registrars and 2) attempt to correlate service characteristics (cost and features) with the relative share of eligible registrants who choose to use a given privacy protection service.

Study Suggestion Number 5: Study whether resellers and registrars offer privacy services to differentiate themselves from others, and, if so, whether this is a factor that encourages competition and whether it is available at no charge.

Determining the availability, cost and features of privacy services can be done easily through a survey of the publicly available information on the registrars' web sites, and would provide useful data on whether such services are readily available in the market at little or no cost to consumers. To measure the extent of use and correlate use with cost and other features will require additional data collection, probably through registrar questionnaires. This would provide information about consumer interest in privacy services and willingness to pay for these services. It is unclear how these results will shed light on the question of whether current WHOIS policies encourage competition, or whether these policies ultimately benefit consumers.

Demand and motivation for use of privacy services

Proposals #17, #18, and #19 explore the demand and motivation for use of WHOIS privacy services. They assume that individual registrants who are natural persons would have the greatest demand for using privacy services and that legal entities may not have a legitimate interest in these services, and may even use these services for illegitimate activities. Thus, proposal 19 proposes to determine the fraction of registrants who are natural persons across all domains. Proposal 18 focuses on determining the fraction of proxy-registered domains that appear to be registered by natural persons through an examination of their websites. Proposal 17 suggests an online questionnaire be administered to registrants who use proxy services to determine their motivation for using those services.

Study Suggestion Number 17: Survey proxy/privacy service registrants to determine their reasons for using a proxy service.

Study Suggestion Number 18: Sample proxy service registrants and review their sites to determine what percentage are likely individual registrants concerned about their privacy.

Study Suggestion Number 19: 1) Sample WHOIS records to determine what percentage of registrations are owned by natural persons, legal persons, and proxy services, and 2) survey registrars to gather similar information as well as information about requests to reveal the identity of the registrant.

These studies appear to be getting at two distinct questions: What fraction of registrants have legitimate interests in privacy services? What fraction of the registrants who currently use privacy services are using them for illegitimate purposes? Suggestions 18 and 19 are reasonable approaches to estimating the fraction of natural persons among samples of registrants. However, it is not clear that only natural persons have a legitimate interest in privacy services, as businesses may also wish to use these services—for example to reduce spam on corporate networks or to protect the privacy of their employees. A better metric may be needed to determine which registrants are using proxy registrations for illegitimate purposes. The online survey of registrants proposed in suggestion 17 seems unlikely to result in accurate self-reported data.

Impact of WHOIS data protection on crime and abuse

Studies #6 and #13 propose to examine the impact of WHOIS data protection on crime and abuse. While the previous category of studies proposed to gather data to provide insights into the fraction of proxy registrants who might be using data protection services for illegitimate purposes, proposals in this category study reported abuse incidents. Study 6 proposes to compare crime/abuse levels across ccTLDs with more restrictive WHOIS access than ICANN's gTLDs. Study 13 proposes to determine whether phishers are using privacy services and, if so, how this impacts shut-down times for phishing sites.

Study Suggestion Number 6: Study whether more restrictive WHOIS data policies lead to more crime and abuse by comparing crime/abuse levels on a percentage basis across two or more ccTLDs with different and/or more restrictive WHOIS access than ICANN's gTLDs.

Study Suggestion Number 13: Conduct analysis of APWG phishing web site data to determine whether phishing web sites tend to be hosted on private/proxy domains and to understand how shut down times of phishing sites are impacted by proxy/private WHOIS registrations.

Both of these studies seem like reasonable approaches to studying the impact of WHOIS data protection on crime and abuse. APWG has volunteered use of their phishing data set for this study.

Proxy registrar compliance with law enforcement and dispute resolution requests

Three proposals (#3, #20, and Metalitz) suggest studies that will determine whether proxy registrar procedures for revealing the identity of an underlying registrant or relaying communications to registrants effectively meet the needs of law enforcement and dispute resolution processes. These suggestions propose to gather data by reviewing stated policies, collecting data from requesting parties, and submitting inquiries and measuring response time. Study 13 in the previous category also includes a similar component.

Before deciding whether or not to conduct a study in which requests are submitted to measure response time, it would be useful to first determine what data on response time is available from requesting parties. A small study might be useful to independently verify reports from requesting parties. A larger study might be needed if little or no data is available from requesting parties.

A related study suggestion, #12, proposes to inventory law enforcement requirements and determine how to balance them with individual privacy needs. Such a requirements analysis, combined with results of the other studies in this category, would provide insights into whether law enforcement needs are currently being met.

Study Suggestion Number 3: 1) Review stated policies of registrars and privacy protection services to determine whether they comply with the RAA and 2) determine actual compliance through a) reports from requesting parties and consumer protection agencies and b) submitting properly constructed inquiries and measuring response time.

Study Suggestion Number 20: Survey proxy registrars, brand owners and law enforcement officials and/or conduct a study to determine timeliness of proxy services in relaying communications to registrants and/or revealing the identity of underlying registrants per RAA 3.7.7.3.

Metalitz Comment: Collect data on UDRP cases brought against registrants who used proxy or private registration services to determine the extent to which a registrant's use of a proxy/private registration service reduced the registrant's ability to contest a UDRP proceeding.

Study Suggestion Number 12: Inventory privacy requirements and law enforcement requirements related to WHOIS.

WHOIS data accuracy

Two proposals (suggestions #8 and #11) suggest that ICANN study certain questions of WHOIS data accuracy. These proposals are orthogonal to questions about the extent to

which WHOIS data should be protected. In conducting these studies, researchers would first need to establish objective criteria to determine whether records have been falsified or whether accuracy and readability have been impaired. As a large sample size is needed, the expense of these studies is dependent on the extent to which the analysis can be automated.

Study Suggestion Number 8: Sample WHOIS data from domains at several registrars and check records for valid combinations of address and phone information to determine whether registrars are tolerating systematic abuse of WHOIS records.

Study Suggestion Number 11: Examine whether IDN (non-ASCII) characters in TLDs will impair the accuracy and readability of WHOIS records displaying the domain name, email address, and name server addresses.

Out of scope proposals

Three proposals (suggestions #4, #7 and #9) appear out of scope.

Study Suggestion Number 4: Study best approach to offering transport layer security for WHOIS queries.

Study Suggestion Number 7: Find a way to stop domain name thieves or protect domain owners from unscrupulous providers.

Study Suggestion Number 9: Survey webmasters to determine whether they have observed the top 10 registrars locking domains after a domain name search.