

Whois Study Recommendations to the RySG – 30 March 2011

From: gTLD Registries Stakeholder Group

Background

In the 16 March 2011 GNSO Council meeting in San Francisco, the following motion was deferred to the next Council meeting at the request of the RySG so as to allow time to work on some RySG concerns about the applicable studies [Study 2 (WHOIS Registrant Identification Study), Study 3 (Whois Privacy and Proxy “Abuse” study) and Pre-study survey for Study 4 (Whois Privacy and Proxy “Relay and Reveal” Study)]:

Motion made by John Berard & seconded by Debbie Hughes

Whereas:

In October 2007, the GNSO Council concluded that a comprehensive and objective understanding of key factual issues regarding the gTLD WHOIS system would benefit future GNSO policy development efforts (<http://gns0.icann.org/resolutions/>).

Before defining study details, the Council solicited suggestions from the community for specific topics of study on WHOIS. Suggestions were submitted (<http://forum.icann.org/lists/WHOIS-comments-2008/>) and ICANN staff prepared a 'Report on Public Suggestions on Further Studies of WHOIS', dated 25-Feb-2008 (<http://gns0.icann.org/issues/Whois-privacy/Whois-study-suggestion-report-25feb08.pdf>).

On 28-Mar-2008 the GNSO Council resolved to form a WHOIS Study Working Group to develop a proposed list, if any, of recommended studies for which ICANN staff would be asked to provide cost estimates to the Council (<http://gns0.icann.org/meetings/minutes-gns0-27mar08.shtml>).

The WHOIS Study WG did not reach consensus regarding further studies, and on 25-Jun-2008 the GNSO Council resolved to form a new WHOIS Hypotheses working group to prepare a list of hypotheses from the 'Report on Public Suggestions on Further Studies of WHOIS' and the GAC letter on WHOIS studies (<http://www.icann.org/correspondence/karlins-to-thrush-16apr08.pdf>). The WG reported to the Council on 26-Aug-2008. (https://st.icann.org/Whois-hypoth-wg/index.cgi?Whois_hypotheses_wg#Whois_study_hypotheses_wg_final_report).

On 5-Nov-2008, the Council convened a group of Councilors and constituency members to draft a resolution regarding studies, if any, for which cost estimates should be obtained. The Whois Study Drafting Team further consolidated studies including those from the GAC (<http://www.icann.org/correspondence/karlins-to-thrush-16apr08.pdf>). The Team determined that the six studies with the highest

average priority scores should be the subject of further research to determine feasibility and obtain cost estimates.

On 04-Mar-2009, Council requested Staff to conduct research on feasibility and cost estimates for selected Whois studies and report its findings to Council. (See Motion 3, <http://gns0.icann.org/resolutions/#200903>).

On 23-Mar-2010, Staff presented a report on the feasibility and cost estimates for the Whois “Misuse” and Whois “Registrant Identification” Studies, finding that each study would cost approximately \$150,000 and take approximately one year to complete. (<http://gns0.icann.org/issues/whois/whois-studies-report-for-gns0-23mar10-en.pdf>). The Whois Registrant Identification study would gather info about how business/commercial domain registrants are identified, and correlate such identification with the use of proxy/privacy services.

The ICANN Board approved in Brussels a FY2011 budget that includes at least \$400,000 for WHOIS studies (see <http://www.icann.org/en/minutes/resolutions-25jun10-en.htm#8>).

On 8-September-2010 the GNSO Council approved a resolution requesting staff to proceed with the Whois “Misuse” Study, which would explore the extent to which publicly displayed WHOIS data is misused, <http://gns0.icann.org/resolutions/#201009>.

On 5-October-2010, staff provided feasibility and cost analysis for a Whois Privacy and Proxy “Abuse” study, <http://gns0.icann.org/issues/whois/gns0-whois-pp-abuse-studies-report-05oct10-en.pdf>. This study would compare broad sample of domains registered with a proxy or privacy service provider that are associated with alleged harmful acts with overall frequency of proxy and privacy registrations. This study was estimated to cost \$150,000 and take less than a year to complete.

On 11-February-2011, staff provided a feasibility and cost analysis for a Whois Proxy and Privacy “Relay and Reveal” study, <http://gns0.icann.org/issues/whois/whois-pp-relay-reveal-studies-report-11feb11-en.pdf>, which would analyze relay and reveal requests sent for Privacy and Proxy-registered domains to explore and document how they are processed. The staff analysis concluded that it was premature to conduct a full study, and recommended that a pre-study “survey” be conducted first, to determine if launching a full study is feasible to do.

Resolved:

Council requests ICANN staff to proceed with the WHOIS Registrant Identification Study, as described in Staff's 23-Mar-2010 Report, using the vendor selection process described in Annex of that same report. (<http://gns0.icann.org/issues/whois/whois-studies-report-for-gns0-23mar10-en.pdf>).

Further resolved, that the Council requests ICANN staff to proceed with the Whois Privacy and Proxy “Abuse” study, as described in staff’s 5-October-2010 report, using the vendor selection process described in that same report, <http://gnso.icann.org/issues/whois/gnso-whois-pp-abuse-studies-report-05oct10-en.pdf>.

Further resolved, that the Council requests ICANN staff to proceed with the Whois Privacy and Proxy “Relay and Reveal” pre-study survey, as proposed in staff’s 11-February-2011 report, <http://gnso.icann.org/issues/whois/whois-pp-relay-reveal-studies-report-11feb11-en.pdf>.

Further resolved, that the Council request that the Board authorize additional funding for FY 2012 for Whois studies, to make up the shortfall of \$130,000 between the amount of “at least \$400,000” that was allocated for Whois studies in FY 2011 (and remains unspent), and the total amount needed to conduct the Whois Misuse Study (\$150,000); the Whois Registrant Identification Study (\$150,000); the Proxy/Privacy “Abuse” Study (\$150,000); and the Proxy and Privacy “Pre-study” (\$80,000), total of \$530,000.

Further resolved, in recognition that there is a substantial amount of coordination needed to direct this research, that staff be given the discretion to manage the studies serially or in parallel, with a goal of expediting completion of the studies as efficiently as possible.

In the same meeting, Chuck Gomes offered to coordinate a small group of GNSO participants to work on possible solutions to the RySG concerns. Subsequently the following individuals volunteered to work on this effort: Steve DelBianco from Net Choice and the Commercial Business Constituency, Kathy Kleiman from PIR and the RySG, and Don Blumenthal from PIR and the RySG. Liz Gasster (ICANN Staff) and Lisa Phifer (ICANN consultant), both of whom lead the efforts with regard to Whois studies, also participated actively in the group and made critical contributions. The group held three teleconference calls and worked online via email to develop a common understanding of the RySG concerns as related to the study designs and discussed possible ways of mitigating the RySG concerns. From that work, Kathy, Don and Chuck developed the following recommendations for RySG consideration in its meeting on 30 March 2011.

Recommendations & Rational

RySG recommend the following:

Study 2 (WHOIS Registrant Identification Study)

- Recommendation: Defer consideration of this study until the 28 April 2011 Council meeting

- Rationale: Because Study 2 contained the biggest RySG concerns and because of a personal emergency, the small group referenced above did not have sufficient time to adequately address the RySG concerns and consider mitigation steps.

Study 3 (Whois Privacy and Proxy “Abuse” study)

- Recommendation: Approve proceeding with Study 3 with some minor amendments as describe below
- Rationale: The leading candidate for performance of this study actually proposed an approach that dealt with most of the RySG concerns but it was felt that documenting key elements of that approach would better ensure that the study was handled appropriately.

Pre-study survey for Study 4 (Whois Privacy and Proxy “Relay and Reveal” Study)

- Recommendation: Approve proceeding with the pre-study survey at an estimated cost of \$80,000, which would cover including translation needed for international participation
- Rationale: After gaining a better understanding of the reasons why no RFP responses were received for Study 4, it became clear to the small group that it would be very difficult, if not impossible, to determine the feasibility of making the study doable without a fairly significant effort; to help others understand, Liz and Lisa prepared the document attached below. (See ‘Whois Privacy and Proxy Relay & Reveal - Pre-Study Feasibility Survey’ below.)

GNSO Council Motion Amendments

Study 2 (WHOIS Registrant Identification Study)

Change the first resolved clause from “Council requests ICANN staff to proceed with the WHOIS Registrant Identification Study, as described in Staff’s 23-Mar-2010 Report, using the vendor selection process described in Annex of that same report. (<http://gns0.icann.org/issues/whois/whois-studies-report-for-gns0-23mar10-en.pdf>)” to “Council defers consideration of the WHOIS Registrant Identification Study until the 28 April 2011 meeting and requests that any applicable motions in that regard be submitted not later than 20 April 2011.”

Study 3 (Whois Privacy and Proxy “Abuse” study)

Change the second resolved clause from “Further resolved, that the Council requests ICANN staff to proceed with the Whois Privacy and Proxy ‘Abuse’ study, as described in staff’s 5-October-2010 report, using the vendor selection process described in that same report, <http://gns0.icann.org/issues/whois/gns0-whois-pp-abuse-studies-report-05oct10-en.pdf>.” to “Further resolved, that the Council requests ICANN staff to **amend the study to include the RySG proposed changes (insert link) and to** proceed with the Whois Privacy and Proxy ‘Abuse’

study, as described in staff's 5-October-2010 report [and as amended](#), using the vendor selection process described in that same report, <http://gnso.icann.org/issues/whois/gnso-whois-pp-abuse-studies-report-05oct10-en.pdf> . " (See Study 3 amendments below.)

Pre-study survey for Study 4 (Whois Privacy and Proxy "Relay and Reveal" Study)

No amendment is necessary for the third resolved clause.

Fourth Resolved Clause

Change the fourth resolved clause from "Further resolved, that the Council request that the Board authorize additional funding for FY 2012 for Whois studies, to make up the shortfall of \$130,000 between the amount of "at least \$400,000" that was allocated for Whois studies in FY 2011 (and remains unspent), and the total amount needed to conduct the Whois Misuse Study (\$150,000); the Whois Registrant Identification Study (\$150,000); the Proxy/Privacy "Abuse" Study (\$150,000); and the Proxy and Privacy "Pre-study" (\$80,000), total of \$530,000." to "Further resolved, that the Council request that the Board authorize additional funding for FY 2012 for Whois studies, to make up ~~the any~~ ~~shortfall of \$130,000~~ between the amount of "at least \$400,000" that was allocated for Whois studies in FY 2011 (and remains unspent), and the total amount needed to conduct: the Whois Misuse Study (\$150,000); the Whois Registrant Identification Study (\$150,000) [if subsequently approved](#); the Proxy/Privacy "Abuse" Study (\$150,000); and the Proxy and Privacy "Pre-study" (\$80,000)~~total of \$530,000.~~"

Fifth Resolved Clause

No change.

Motion Amendments for Study 3

RySG recommend that Study #3 be amended so that:

- 1) Study #3 expressly reference in its introduction data from another ICANN Study. Since Study #3 references only proxy/privacy use by "bad guys," we ask that the study clearly and in a prominent early area place the results of the 2010 ICANN Study on Whois Proxy/Privacy use overall. Namely:

"Because this is a sample, our 95% confidence interval for the percentage of Privacy or Proxy services among the top five gTLDs is 16% -20%."

This provides context and background for this Whois study of bad actors having an overall context of reference within the Whois system.

- 2) A control group be included in the study (as recommended by one leading research group), but clearly note that the control group work has many limitations. Specifically, the work would analyze a randomly-selected set of domain names for lawful websites, chosen to mirror sites involved in illegal/harmful activities to be studied. This

approach has some potential for useful results. We expect it would lead to a comparison of pharmaceutical websites, both legitimate and fraudulent sellers, and typo squatters, both domains being used to direct browsers to advertising pages and also those being used to direct browsers to web pages that deliver malware.

But, it also has significant problems that the research group must account for carefully. These include that some “bad actors” simply will not have corresponding “good actors” (online crime, for example, or websites entirely devoted to the distribution of malware). Further, while the control group concept focuses on website comparison, not all Internet activity or domain use involves websites. Some behavior specified in Study #3 specifically would not need a site, such as Denial-of-Service attacks and DNS Cache Poisoning. In addition, it is hard to imagine a legitimate use of those tactics. Additional types of malicious behavior also would not be amenable to comparison.

Clearly all of these limitations and disclaimers should be clearly accounted for in the body of the final report to allow the GNSO Council and Community to fully understand the context and limitations of the data being received.

- 3) It should be noted in the study that many legitimate proxy/privacy users will not be touched – they are outside the scope of Whois Study #3.

Specifically, neither the study nor its control group will reach legitimate noncommercial and commercial use that has no significant illegal counterpart – namely, non-commercial organizations, school groups, scouting organizations, human rights organizations, and the millions of individuals with websites online. They generally have no illegal counterpart per se.

Accordingly, the study should clearly and expressly, within the body, note that the purpose is only to review bad actors and, notwithstanding the control groups, many legitimate proxy/privacy actors will be unaccounted for within its scope.

Whois Privacy and Proxy Relay & Reveal - Pre-Study Feasibility Survey

Prepared by Liz Gasster

March 28, 2011

This report briefly describes the Whois Privacy/Proxy Relay & Reveal Pre-Study Feasibility Survey referenced by the following GNSO Council Motion dated 07-Mar-2011:

<http://gns0.icann.org/mailling-lists/archives/council/msg10868.html>

Background

In 2010, terms of reference for a **Whois Privacy/Proxy Relay & Reveal** study were drafted, based on proposals by members of the ICANN community, including GAC Data Set 1 (to what extent are legitimate uses of Whois data curtailed by use of proxy or privacy services). The study explores a sample of actual relay and reveal requests for Privacy/Proxy-registered domain names to document how they are processed and identify factors that may promote or impede timely communication and resolution.

An RFP inviting interested researchers to bid on conducting that study was posted on 29-Sep-2010:

<http://www.icann.org/en/announcements/announcement-29sep10-en.htm>

No bids were received and potential bidders expressed significant concerns with study feasibility. The following assessment report was delivered to the GNSO Council, summarizing perceived barriers:

<http://gns0.icann.org/issues/whois/whois-pp-relay-reveal-studies-report-11feb11-en.pdf>

Staff concluded that it was premature to conduct a full study, and recommended that a pre-study survey be conducted first, to determine if launching a full study is in fact feasible. The following sections summarize the goals, benefits, and costs associated with conducting this pre-study survey.

Study Barriers

The RFP proposed collecting a broad sample of relay and reveal requests from volunteers, including individuals, businesses, first responders, complaint centers, and law enforcement agencies. Privacy/Proxy providers and Registrars would then be given a chance to supply secondary input about their published policies and practices and how sampled requests were actually handled.

The key challenge appears to be obtaining a sufficient data sample. Researchers could not assess the feasibility of finding willing and able volunteers to supply this sample or effort to do so because:

- Finding diverse participants appears likely to require a very large outreach campaign.
- Some essential parties (e.g., Privacy/Proxy providers, Registrars) may object to the study.
- Law enforcement agencies are unlikely to supply data pertaining to on-going investigations.
- Many others parties may be unable to supply data due to business sensitivities or privacy laws.
- Even willing participants are likely to have limitations on the data elements they can disclose.
- Incentives and data collection aids may be needed to encourage sufficient participation.

Pre-Study Feasibility Survey Goals

Given sampling uncertainties and extensive third-party dependencies, respondents could not reliably estimate study cost or duration. To resolve these uncertainties while establishing the foundation needed to run the full study anyway, we propose to conduct a pre-study feasibility survey of limited duration.

If this survey can identify a pool of potentially willing and able participants (including request originators, Privacy/Proxy providers, and Registrars), along with measures and tools required to enable accurate timely data collection, bids can be solicited to conduct the full-blown exploratory study. However, if the feasibility study determines that obtaining a sufficient set of willing and able participants is unlikely, this would also help the community decide whether to pursue, revise, or abandon the full-blown exploratory study.

Pre-Study Feasibility Survey Approach

The pre-study feasibility survey shall be conducted by a team of senior researchers with a solid grasp on the problem, study goals, and contacts needed for community outreach. This team will attempt to:

- Assess community willingness and ability to participate in a full Relay & Reveal study, including request originators, Privacy/Proxy providers, and Registrars
- Identify potentially-willing participants and limitations on their participation, including their business sensitivities and privacy laws
- Identify availability of requested data elements and conditions for sharing it, including measures needed to protect relay and reveal requests and responses
- Explore the impact of incentives and data collection tools on participation, including viable methods for timely accurate reporting and follow-up
- Solicit actual request examples for use in formulating a full Relay & Reveal study, and to assess Privacy/Proxy provider and Registrar ability to supply secondary input

To accomplish this, the following project is proposed:

1. Conduct an online survey, broadly publicized to potentially interested parties;
2. Follow up with in-depth interviews with a small but representative set of survey respondents;
3. Perform both statistical and qualitative analyses of survey and interview results;
4. Document findings in a draft report, to be presented to the GNSO; and
5. Deliver all tools, data, and intermediate results (including a list of potentially-willing participants) to serve as a launch point for the full exploratory study

The survey will be carefully designed, tested, and validated so as to get the best possible answers to the questions raised above. An iterative approach will be used to refine the survey to maximize the chance of gathering good data. A standard on-line survey tool will be used to conduct the survey in English and a few additional languages to enable international participation. Community outreach efforts will be used to solicit participation from all participant categories. Follow-up calls will be conducted with randomly-selected respondents in each participant category to obtain further detail.

Pre-Study Feasibility Survey Duration and Cost

The cost to perform this pre-study feasibility survey is estimated at \$80K, to be completed over a 4 month period. This estimate is based largely but not exclusively upon a proposal supplied by a researcher that declined to bid on the full study but willing to undertake this pre-study survey. If GNSO Council approves this pre-study feasibility survey, that initial proposal would be expanded and refined prior to project start. In particular, language translation is required to conduct this study; the above-estimated project cost and duration reflects this extension.