

COALITION FOR ONLINE ACCOUNTABILITY

WWW.ONLINEACCOUNTABILITY.NET

1818 N STREET N.W., 8TH FLOOR • WASHINGTON, D.C. 20036-2406 • TEL: (202) 355-7906 • FAX: (202) 355-7899 • E-MAIL: INFO@ONLINEACCOUNTABILITY.NET

COMMENTS OF THE COALITION FOR ONLINE ACCOUNTABILITY

April 15, 2010

The Coalition for Online Accountability (COA) offers the following comments on the draft version of the NORC Study on Whois Data Accuracy. See <http://www.icann.org/en/public-comment/#whois-accuracy-study>.

COA consists of eight leading copyright industry companies, trade associations and member organizations of copyright owners. These are the American Society of Composers, Authors and Publishers (ASCAP); Broadcast Music, Inc. (BMI); the Entertainment Software Association (ESA); the Motion Picture Association of America (MPAA); the Recording Industry Association of America (RIAA); the Software and Information Industry Association (SIIA); Time Warner Inc.; and the Walt Disney Company. All these entities depend upon reliable access to accurate Whois data in order to safeguard their intellectual property rights, and those of their members, in the online environment. Consequently, COA has been deeply involved in the debates within ICANN over Whois policy for more than a decade.

The NORC study marks an important contribution to this debate, and provides some welcome empirical evidence that should enhance the quality of ICANN's future decision-making in this sphere. The following main conclusions can be drawn from the NORC study:

1. *The problem may be worse than we thought*

The NORC study provides a more systematic, more comprehensive, and more detailed demonstration of what COA has been stressing for a decade: the domain name registrant contact data available through Whois is riddled with false information. This pervasive inaccuracy undermines the ability of the Whois service to fulfill its mission of enabling contact with the party responsible for a registered domain name. In fact, the NORC study confirms that the problem of inaccurate Whois data is even worse than indicated by earlier studies (notably the GAO study of 2005).

Both the GAO study and the more comprehensive NORC study found that the Whois data for over 7% of gTLD domain name registrations was either patently false, incomplete, or the Whois data simply could not be accessed. As the NORC study puts it, 7.8% of the registrations "fully failed" to be associated with accurate registrant contact data. But the NORC

American Society of Composers
Authors & Publishers (ASCAP)

Entertainment Software Association (ESA)

Software & Information Industry Association (SIIA)

Broadcast Music Inc. (BMI)

Motion Picture Association of America (MPAA)

Time Warner Inc.

Recording Industry Association of America (RIAA)

The Walt Disney Company

Counsel: Steven J. Metalitz (met@msk.com)

study drilled deeper, concluding that another 20.9% of the registrations “substantially failed” the accuracy test. In the great majority of these cases, the NORC researchers were “unable to link or even locate the registrant.”

With nearly 30% of the sample of more than 1400 registrations in the 5 largest gTLDs classified as fully or substantially failing an accuracy test, the scope of the problem is clearly far greater than some of the previous studies suggested. These failure rates extrapolate to “full failure” of Whois accuracy with respect to some 9 million gTLD registrations, and “substantial failure” with regard to an additional 24.2 million registrations.¹

2. The prevalence of privacy/proxy registration is masking a broader problem

Even though the NORC study finds a more serious problem than the earlier investigations, it actually understates the scope of the problem. NORC classifies 14.7% of the registrations as associated with Whois data for a privacy or proxy registration service, not the beneficial registrant itself.² (page 9) NORC counts nearly all these as registrations associated with (more or less) accurate data. Not surprisingly, the contact information for the proxy or privacy service proved to be substantially or fully inaccurate in only 8% of the cases. (page 16)

Thus, for about another 13.5% of all gTLD registrations (92% of 14.7%), Whois enables contact, not with the party actually responsible for use of the domain name, but with a third party whose contact details appear in Whois instead. In this universe of some 15.7 million (or more) additional gTLD registrations, the real question is whether it is possible to contact the actual registrant – the customer of the proxy service – or even to learn that party’s identity and contact data.

The experience of COA participants is that proxy and privacy services frequently do not reveal this data, even when presented with what the Registrar Accreditation Agreement (RAA), in section 3.7.7.3, describes as “reasonable evidence of actionable harm” resulting from use of the domain name. Thus, the NORC data show that the proportion of registrations for which accurate contact information on the party responsible for the resource is not available through Whois is far greater than 30%, and perhaps nearly 40%, of the total.³

¹ The NORC study reports a total of 101,225, 988 registrations in the five gTLDs studied, accounting for over 98% of total gTLD registrations. This figure is badly outdated. As of April 14, 2010, the top five gTLDs included about 115,985,550 registrations. See <http://www.domaintools.com/internet-statistics/>. The latter figure is used for the calculations in the text of this comment.

² In a previously released report, analyzing precisely the same data, ICANN staff concluded that a 14.6 % NORC figure (evidently later adjusted to 14.7%) for proxy/privacy registrations is at the bottom edge of a range that could reach as high as 25% of all gTLD registrations. See <http://www.icann.org/en/compliance/reports/privacy-proxy-registration-services-study-28sep09-en.pdf>.

³ For instance, if, using a conservative estimate, the “true registrant” information could not be obtained for half of the registrations associated with a proxy or privacy service (where the service’s more or less accurate contact data appears in Whois), then the “full or substantial failure” accuracy category would include an additional 6.75% of all registrations (1/2 of 13.5). Thus, this category would grow from 28.5 % (7.8 full failure + 20.7 substantial failure) to 35.25% (28.5 + 6.75).

COA believes these findings underscore the need for ICANN to ramp up its efforts to address the abuses inherent in the current system of proxy and privacy registrations, where anarchy reigns, and adult supervision is long overdue. Among many other steps that ICANN should take, a high priority should be given to research projects already in the pipeline that would quantify the extent to which these services provide shelter for illegal, abusive and malicious activities, and the extent to which providers of these services are carrying out any relay or reveal functions.

3. Registrars are at the epicenter of the problem

One shortcoming of the NORC draft report is that, with one exception, it does not report data broken out by gTLD. The one exception is revealing. In Appendix 4, entitled “GAO study replication,” the NORC study presents data that shows a clear differentiation between the “thin Whois” gTLDs (.com and .net), where registrant contact data is maintained only at the registrar level, and those registries with “thick Whois” (.org, .biz and .info). For example, in 2.0% of a larger sample of 2400 domain names, the NORC researchers were unable to access any Whois data at all. This occurred only in the “thin” registries, never in the “thick” gTLDs.⁴ Furthermore, the prevalence of patently false or incomplete Whois data was much higher in .com and .net (5.9% in both cases) than in the “thick Whois” registries (ranging from 2.4 to 4.4 %).

This data indicates that the worst of the problem of Whois data inaccuracy – and the entirety of the problem of Whois data accessibility – is to be found at the registrar level, not among registries. It also underscores that the trend toward thick Whois architecture should be encouraged in order to improve the accuracy as well as the accessibility of Whois data. Release of other data broken out by gTLD may provide other insights, and NORC should be encouraged to do so in the final report.

4. The costs of improving accuracy are manageable

The NORC report also shows that, while there will be costs associated with increasing Whois accuracy, some simple and inexpensive steps could make a difference. These include matching name and address with the credit card used to pay for registration, and even “basic edit checks” to reject registrations with blank contact data fields. COA agrees with NORC that “only the registrars themselves are in a position to use efficient electronic checks of the data, from basic field completion checks through to crosschecks against address deliverability databases and other databases by which identity might be confirmed.” (page 19) However, currently they have little economic incentive to do so, since they would incur some additional costs that their less scrupulous competitors would avoid. ICANN is the entity best positioned to change this calculus. By imposing greater responsibility on accredited registrars to improve the accuracy of Whois data they collect and display, and by rigorously enforcing this obligation on all accredited

⁴ No Whois data whatsoever could be obtained for 2.4% of the registrations in .com. This extrapolates to 2.1 million of the 86.5 million active .com registrations.

registrars, ICANN can transform Whois data accuracy from an uneconomic choice to a simple cost of doing business for all registrars.

Conclusion

In summary, the NORC study documents and quantifies, to an unprecedented degree, that Whois inaccuracy is a serious problem. What the NORC study does not need to spell out, because it is already well established, is that this problem falls squarely within ICANN's remit as the technical coordinator of the domain name system, and that its continuing failure to take effective steps to address the problem amounts to neglect of an ongoing threat to the security and stability of the DNS. ICANN must step up to this problem, in order to fulfill its commitment to manage the domain name system in the public interest. While the impending launch of many new gTLDs provides adds urgency to this need, the NORC study amply demonstrates the scope and breadth of the problem in the existing gTLD space.

Respectfully submitted,

Steven J. Metalitz, counsel to COA

Mitchell Silberberg & Knupp LLP | 1818 N Street, N.W., 8th Floor, Washington, D.C. 20036
USA | tel: (+1) 202 355-7902 | fax: (+1) 202 355-7899 | met@msk.com