Comments of the Intellectual Property Constituency ("IPC") to the Draft Report for the Study of the Accuracy of WHOIS Registrant Contact Information by National Research Opinion Center (NORC) at the University of Chicago for ICANN

The IPC welcomes the opportunity to comment on the Draft Report for the Study of the Accuracy of WHOIS Registrant Contact Information by NORC – 17 January 2010. WHOIS accuracy is an important issue for not only Intellectual Property owners, but for law enforcement, business owners and individual stakeholders. Additionally, the mandated requirements for becoming and remaining an accredited ICANN Registrar under the Registrar Accreditation Agreement (RAA) include responsibilities for maintaining accurate and current WHOIS data.

The IPC very much appreciates the effort and expense taken to get a baseline measurement of what proportion of WHOIS records are accurate and applauds the effort of ICANN and NORC to expand the scope of this effort beyond missing and patently false information in the records of the top gTLDs (.com, .net and .org) as was done in 2005 by the GAO. Examining an internationally representative sample of the top five gTLDs (.com, .net, .org, .info and .biz) and contacting registrants directly to learn if information which appeared to be complete was in fact inaccurate as NORC has done, is a necessary step in assuring WHOIS accuracy.

It is noteworthy that .edu, .mil, and .gov were deemed "out of scope" in terms of this survey "due to the higher level of control (and thus accuracy) in registration of registration of domains within those three gTLDs."

The IPC has chosen not to comment on the survey design agreed to by NORC and ICANN and accepts that the sampling of records and extrapolations of what the finding percentages means in terms of domain registrations and accuracy percentages falls within the scope of acceptable survey industry standards.

Therefore, this comment addresses the results of the study, and their policy implications. The results of the study are disheartening for several reasons but most importantly that the scope of the problem, although not knowable exactly, is bigger than suspected, particularly when compared to the 2005 study. The NORC study also suggests that making WHOIS more accurate will result in escalating costs; yet, in IPC's view, the study itself shows that significantly increased accuracy could be achieved with relatively minimal additional costs. Finally, the study is discouraging to the notion that adding perhaps hundreds of new gTLDs into the domain name system in the near future will do anything other than exacerbate the problem.

Baseline measurement of what proportion of WHOIS records are accurate

Using the three criteria agreed to by NORC and ICANN that needed to be met for a WHOIS record to be considered accurate, namely:

1. Was the address of the registrant a valid mailing address?

2. Was the registrant named associated in some way with the given address?
3. When contacted, would the named registrant acknowledge that they were indeed the registrant of the domain name, and confirm all details given as correct and current?

NORC was only able to confirm that 23% of all WHOIS records were fully accurate. The accuracy percentage increases if "minimal" and "limited" failures are permitted which include minor faults in data input by the registrant or if NORC was unable to contact the registrant to confirm ownership, but never exceeds 71% of accurate WHOIS data. That means that 29% of WHOIS data is either substantially or fully inaccurate. This represents over 29 million records of the approximately 101, 226,000-registered domain names (at the time of the study).

Given that one of ICANN's core commitments is the accuracy of WHOIS data, a 29% failure rate is not acceptable in our view.

Nor should the community take too much comfort in the conclusion that 71% of WHOIS data is more or less accurate. This figure includes nearly all the 15% of registrations that NORC determined were made through proxy or privacy services. While the contact information of the service, as listed in WHOIS, could readily be verified in the vast majority of cases, this leaves unanswered whether the WHOIS data could be used for the purpose for which it has always been collected, which is to enable contact with the registrant. This depends on whether the proxy or privacy service fulfilled its responsibilities to relay inquiries from third parties, or to reveal the contact information of the "true" registrant when presented with reasonable evidence of actionable harm flowing from how the domain name is being used (See RAA Section 3.7.7.3). Of course this issue was outside the scope of the NORC study, but should be addressed by ICANN in its future research agenda on WHOIS.

The study states that most of the barriers to accuracy (concerns about privacy, confusion about information needed, clarity in the standard to which information should be entered, etc.,) "can be addressed by the Internet community" but with escalating costs associated "with the level of accuracy sought." The study also suggests that there might be a commercial disadvantage for some ICANN constituents from enforcing greater WHOIS accuracy. This statement of escalating costs is contradicted by the study's own findings.

Criteria 1: Deliverability of the mailing address

By checking addresses in the WHOIS sample set against postal codes and simple mapping data for delivery, NORC easily discerned that 13.3% of those addresses were undeliverable, and that an additional 8.6% were "potentially deliverable".

It stands to reason that if ICANN required registrars to use a basic algorithm that checks both postal codes and deliverability at the point of entry as NORC did on the backend examination, at least 13% of the 29% inaccurate records could have been prevented at registration and perhaps as much as 15-20%.

Criteria 2: Association of name and address

The study states that before it was attempted to associate name and address, the names were coded for type. In dividing the names into personal, organization, proxies, etc., it was determined that over 1.6 million WHOIS records have registrant names that are completely missing or patently false.

Even though this is a small percentage in regard to all registered domain names the IPC feels that this number is completely avoidable with minimal enforcement effort.

Additionally, the study showed that 20% (or over 20 million WHOIS records) did not meet the second criterion of associating the registrant name with the address listed. This included another 1.8 million domain name registrations with partial names for registrants or information inputted that NORC was unable to classify, which the study noted could be "borderline patently false."

An additional 30% of the domain names surveyed included privacy/proxy registrations, organizational names, multiple domain name holders, etc., showed no association, but NORC was able to locate the registrant, in most instances only after applying the third criteria of contacting the registrant. (As noted above, the ability to contact the privacy or proxy registrant does not equate to the ability to contact the party actually responsible for the domain name.)

Criteria 3: Registrant acknowledgement

This third criteria was admittedly more difficult to meet, with varying factors depending on if the registrant was an individual, the WHOIS record was listed to a company or organization, etc., in addition to interviewees not willing to talk the person conducting the survey (independent of topic of compliance and more related to reluctance to spend time participating in a survey).

We agree with NORC that in this instance verifying WHOIS accuracy by contacting registrants would escalate costs considerably and don't deem this a practical measure except perhaps as a last resort before deletion.

Applying all three criteria

As was previously mentioned by the study, only 23% of WHOIS records can be considered fully accurate. Applying 2 out of the 3 criteria results in 42% accuracy. Applying one out of three results in 27% accuracy. And 8% of the records do not satisfy any of the criteria for establishing accurate WHOIS data.

While there are many instances reported in the study that are understandable reasons why a domain name registration does not meet all three criteria for fully accurate WHOIS data (transposition errors, confusion of what information to put where, etc.) there are many millions of registrations that make it into the system that shouldn't. And preventing many of these registrations in the first place need not be difficult or expensive.

Barriers to accuracy

The study suggests that the main barriers to accuracy in WHOIS data are privacy concerns and carelessness and/or little perceived value in domain ownership. While these are understandable as reasons, they are not acceptable excuses for inaccurate WHOIS data.

NORC notes that no proof of identity or address is required when registering a domain name (which would remove many barriers to entering inaccurate information). The study suggests that "requiring that the registrant and address at least match that of the credit card which was used to pay for the account would go some way towards addressing this, given that reasonably stringent proof of identity and address is usually required to obtain a credit card."

We agree that at least matching registrant name and address with credit card information is a good solution in cutting down WHOIS inaccuracy. All registrars are set up to take payments by credit card and comparing that data with WHOIS data would be simple and not expensive. We note that some protocol would be needed to address the reality that many organizations do not issue employees "company" credit cards, meaning that many entities will register domains using a credit card whose billing address is the private home address of an individual employee.

The study notes the registrant is given a broad scope within which they can choose to interpret the requirements for name and address.

Again, matching data with credit card information would help solve this problem.

The study notes that basic checks, if used consistently by all registrars could eliminate some missing data issues.

We agree with the study, that if all registrars made these efforts, it would at least sharply reduce those registrations with addresses that are not valid.

Other issues noted by the study regarding confusion (name and address required 4 times when registering: registrant, administrative, technical and billing contacts) and late 2009 changes to support the entry and maintenance of non-ASCII character sets we feel are understandable and nominal issues.

NORC notes that maintenance of accurate WHOIS data is an issue for two main reasons: 1) registrant motivation for keeping data accurate and current needs to be improved; and 2) since only registrars can use efficient electronic checks of data and registrars all check this data differently, putting together a robust and centralized database would increase costs.

We agree that accuracy of WHOIS data needs to be an important issue for both registrants and registrars. And having a centralized database enabling pattern based checking, as the study suggests, would not only improve maintenance checks but could be used to possibly flag potential fraudulent activity.

We feel that ICANN has not historically enforced compliance of the registrars maintaining WHOIS data accuracy of their registrants, which would help in motivating both to keep records up to date. This effort does not in our view necessarily escalate costs.

We acknowledge that at centralized database or common data format would be associated with some additional costs. We do note, however, that every accredited registrar has made the commitment to provide its data to a centralized Whois service if and when directed to do so by ICANN pursuant to a consensus policy. See RAA Section 3.3.4. Furthermore, the mere proposition of a common data format—highlighted by the study in noting that, "since different registrars used different fields in different ways, … mapping everyone successfully into a consistent set of fields ultimately required a large degree of manual work"—could be addressed at minimal cost by requiring a "thick" WHOIS system in all remaining "thin" registries.

Conclusion

The IPC feels that the NORC survey has clearly identified that the problem of WHOIS accuracy is greater than had been reported in the past and that 29% of the records in the domain name system being substantially or patently inaccurate is an unacceptable percentage.

The IPC feels that improving WHOIS accuracy does not necessarily come with escalating costs as the survey identified sensible and simple solutions in remedying the blatantly inaccurate or absence registration information, including cross checks with postal codes, mapping and address validity software and matching registrant data with data processed for registration payments. While these simple and inexpensive steps would not eliminate WHOIS inaccuracy, they would sharply reduce it.

It is the view of the IPC that ICANN needs to examine if it has the tools in place to live up to a core commitment of accurate WHOIS data and if not, find measures to improve those tools in short order, or find a remedy outside the organization to maintain and enforce accurate WHOIS data. The plan to introduce additional gTLDS into the domain name system makes this an urgent task.