**Radix Registry Comment on New gTLD Name Collision Risk Mitigation**

Dear ICANN Board,

On behalf of Radix Registry, we would like to take this opportunity to voice our concern over the recent name collision report put together by the Interisle Consulting Group and the ICANN staff recommendations that are being proposed in response to the former.

Radix Registry is a part of the New gTLD Applicant Group (“NTAG”), and has worked with members from their Working Group to formulate the NTAG response on this issue as well. As rightly explained in that letter, the findings contained in Interisle’s report do not give sufficient cause to delay the new gTLD program in the manner proposed by ICANN staff.

While we do support and endorse the NTAG letter sent to the Board, we would like to bring out some additional issues that are of relevance. In the absence of data used for the study, we have minimal ability to provide ICANN with quantitative reasoning for substantiating our requests, but there are some glaring loopholes in the study and the resulting staff recommendations, which have not been highlighted by the NTAG. The most prominent ones are listed below:

1. **Quality of queries is MUCH more important than quantity of queries**

The Interisle study simply counts the absolute number of queries received per string. Clearly, this should not be the only metric used by ICANN staff to make the far-reaching recommendations that it is proposing. The study makes no mention of the following more important statistics:

1. The number and type of unique domain names queried in each string
2. The number and distribution of unique sources / IP addresses of queries in each string

The reason we believe that the above statistics would be more important to analyze is logical. Imagine a hypothetical, but possible situation, in which there are a handful of large organizations which use the .corp extension in their internal networks, and would therefore be responsible and account for the bulk of the 1.4 million queries seen in .corp (Rank 2). If this is true, there may be a finite number of unique domain name queries issued in .corp, and originating from a handful of easily identifiable sources / IP addresses in such a case. It is also likely that the same would apply to .brand strings like .hsbc (Rank 14), where a small number of unique domain name queries may have been issued from a few easily identifiable sources. If this is found to be true, mitigating such a risk in a targeted manner would actually be fairly straightforward. However, if not true, then we believe that the recommendation we are proposing will serve to mitigate the risk anyway.

Conversely, imagine a large number of smaller stock broking businesses that possibly use the .trade (Rank 303) extension (44,000 queries) on their internal network for trading or storage of sensitive and confidential information. Naturally, if this were true, there would be a relatively larger number of unique queries in .trade originating from a widely distributed number of unique sources / IP addresses. This risk could potentially be a lot more challenging and complicated to deal with. But ICANN staff currently proposes to qualify it as “low risk”, without analyzing the data more thoroughly and therefore allow it to proceed on to delegation.

Thus, we believe that it would be grossly incorrect to classify some strings as being less or more risky than other strings simply by summing the number of queries, and without analyzing the unique domain names queries and unique sources of these queries.

1. **Any new data used for further studies can easily be compromised / gamed**

In the situation that ICANN does choose to carry out further studies in relation to the “uncategorized risk” strings, we would like to caution ICANN against the usage of any data that has been / will be collected after the commissioning of the Interisle study. There is an almost undeniable likelihood of such “new” data being gamed or compromised. Some important points to consider:

1. The Internet community and any interested members of the general public are now aware that “query counts” have been used to determine the level of risk that any proposed gTLD poses
2. They are also aware of the exact list of 281 proposed gTLDs for which further studies of data will be carried out in order to ascertain refined risk levels
3. It is extremely easy to purchase a software or a service that can send a gigantic amount of targeted queries from multiple sources for non-existing strings of a particular kind/s for resolution
4. There are several parties that have vested interests, or stand to benefit from the delay or potential non-delegation of the so-called “riskier” strings
5. Thus it would be / is incredibly easy for any newer data to be manipulated for personal gain

In light of the above, we assert that ICANN should make it mandatory for all future studies to use only pre-dated data (before commissioning of the Interisle study), which is more likely to be more representative of actual usage of proposed gTLDs in internal networks.

1. **The threshold for dividing strings into “low risk” and “uncategorized risk” is arbitrary**

The Interisle report section 8.3.1 suggests that one way for setting the threshold for dividing strings into “low risk” and “uncategorized risk” could be by reference to the number of queries for existing TLDs that have empty zone files. The report mentions two such existing TLDs - .sj and .bv, both ccTLDs associated with mostly uninhabited Norwegian colonies. ICANN simply picked .sj, and decided that all the proposed strings that appeared in the data stream more frequently than .sj (49,842 queries) should be classified as “Uncategorized risk”.

The result of this arbitrary selection is that .bio (Rank 281) with 50,000 queries (rounded to the nearest thousand) is part of the “uncategorized risk” list, and is delayed by 3 to 6 months, whereas .engineering (Rank 282) with 49,000 queries (rounded to the nearest thousand) is part of the “low risk” list, and can proceed without any significant delays. What this means is that there may have been a handful of more queries for .bio (as little as 344 more queries) than for .engineering, which somehow made .bio appear more “risky” to ICANN than did .engineering.

Another validation for the fact that this division is unjustified comes from Digicert, one of the world’s largest CAs, which is also a founding member of the CA/Browser Forum (<http://en.wikipedia.org/wiki/DigiCert>). In a presentation conducted specifically to discuss the Name Collision issue on the 22nd of August 2013, Digicert’s associate general counsel, Jeremy Rowley (<https://artemis.net/tld-security-forum/speakers/#Rowley>) stated that Digicert’s opinion was that including 20% of all proposed strings in the “uncategorized risk” segment was unnecessary (See Slide 9 of the presentation: <https://www.artemis.net/wp-content/uploads/2013/08/Artemis-Presentation.pptx>). In fact, it was verbally stated that only the top 14 of the proposed gTLDs from the Interisle report are substantially risky.

**Recommendation for ICANN Board**

**Allow all strings to be delegated subject to certain mitigation measures**

We firmly believe that there is a relatively simple series of steps that can be taken to mitigate the real threat that ICANN staff foresees with respect to name collision. Evaluating and affecting the following can allow all strings to proceed on to delegation while successfully avoiding unnecessary delays:

**Phase 1: 3 months after delegation**

1. Registries should not sell / register / activate any domain names in their gTLD for a 3 month (120 day) period following delegation
2. Registries should use these first 3 months (Phase 1) after delegation to simply gather data about every single query in their gTLD (Processes such as Sunrise notification should run simultaneously during Phase 1)
3. Needless to say, all such queries would receive NX Domain responses during Phase 1
4. At the end of Phase 1, the Registry should use the list of queries received during the last 3 months to build a set of all unique SLDs that were queried in their gTLD
5. ALL these unique SLDs should be blocked, and not be made available for sale

**Phase 2: Next 3 months**

1. During the next 3 months (Phase 2), the Registry should proceed with their business plans for all SLDs, barring those contained in the blocked list
2. Any queries for these blocked names will continue to receive NX domain responses during Phase 2 as well
3. During the first 30 days of Phase 2, the Registry should make a set of prescribed efforts to effectively communicate with the unique sources / IP addresses from where queries on the blocked list (during Phase 1) originated
4. A pre-set time frame of 60 days should be granted to allow the concerned organizations / authorities to fix the issues or make the necessary alterations to their internal networks
5. At the end of Phase 2, the SLDs on the blocked list should be made available for sale

We genuinely believe that a mitigation strategy based on the above recommendations or some variation thereof, if applied to all TLDs will serve to alleviate even unforeseen threats in TLDs that ICANN staff has inadvertently qualified as “low risk”.

We appreciate ICANN’s proactive response to the Interisle study, and look forward to our recommendation being addressed as part of the risk mitigation exercise. We would also like to express our willingness to work with ICANN staff in order to further the new gTLD program in a manner that is mutually beneficial.

Sincerely,

Brijesh Joshi

Radix Registry