ICANN CALL FOR EXPRESSIONS OF INTEREST (EOIs)

For New gTLD String Similarity Examiners
25 February 2009

1 Introduction

Generic top-level domains (gTLDs) are an important part of the structure of the DNS. Examples of existing gTLDs include .BIZ, .COM, .INFO and .JOBS. A complete listing of all gTLDs is available at http://www.iana.org/gtld/gtld.htm. The responsibility for operating each gTLD (including maintaining the authoritative registry of all domain names registered within that gTLD) is delegated to a particular organization. These organizations are referred to as "registry operators" or "sponsors," depending upon the type of agreement they have with ICANN.

Following years of community-driven policy development that recommended the introduction of new gTLDs, ICANN is preparing a process to receive applications to operate new generic top-level domain (gTLD) registries. This new program is described in detail at http://www.icann.org/en/topics/new-gtld-program.htm. ICANN has published a draft Applicant Guidebook at http://www.icann.org/en/topics/new-gtlds/comments-2-en.htm that provides detailed information about the process for applying to operate a new gTLD. The Applicant Guidebook will constitute the request for proposals (RFP) for new gTLDs.

The development of the Applicant Guidebook is an iterative process, which includes seeking public comment on draft versions. The comment resulting from the publication of the first draft Applicant Guidebook led to the identification of several overarching issues that will require additional examination and discussion to resolve. Although ICANN has prepared a revised Applicant Guidebook, the information in the Guidebook is not yet fixed and the new gTLD process is not yet launched. While that work goes forward, steps will also be taken to assure there will be a robust, effective and timely evaluation process in place to review applications once the round is launched. Retaining competent evaluation panels with sufficient expertise, resources and geographic diversity is expected to take many months. Some preliminary steps, such as the publication of this call for expressions of interest, are being taken now, even as important decisions regarding the overall implementation process are still being considered.

ICANN is now seeking expertise to enable the formation of panels to evaluate applications against the criteria published in the Applicant Guidebook. Expressions of Interest (EOIs) in providing management and evaluation services are sought in the following five areas of assessment:

1. Has the applicant demonstrated their technical capability to run a registry for the purpose specified in the application as defined by the criteria in the Applicant Guidebook?

2. Has the applicant demonstrated their financial and organizational capability as defined by the criteria in the Applicant Guidebook?

3. In the context of the criteria specified in the Applicant Guidebook, does the gTLD represent a geographical name, and if so, have authenticated support from the relevant government?

4. Will the introduction of the proposed gTLD string likely result in user confusion with (i.e., due to similarity with) (i) a reserved name; (ii) an existing TLD; or (iii) other proposed gTLDs?

5. In the context of resolving contention among two or more applicants for the same or similar gTLD string, does an applicant claim to represent a community and if so, satisfy the criteria for prevailing in a comparative evaluation?

ICANN also seeks information from potential providers regarding estimation of reasonable timeframes for each type of evaluation (e.g., per string or per application) and anticipated costs associated with conducting the evaluation. The cost and time to process an application are critical factors that must be carefully considered in the information provided by the interested parties.

This EOI describes the criteria and requirements for providers that seeking to perform the inquiry of question 4, the string similarity examination role. Providers should respond by 13 April 2009 23:59 UTC with the required information that is described below. From the information provided, ICANN will invite respondents to exchange additional information.

Contracts will not be awarded from this EOI, but ICANN expects to use the responses to identify entities capable of providing the various evaluation roles and better refine the costs and time frames for conducting evaluation as part of the new gTLD process.

2 Background

The Internet Corporation for Assigned Names and Numbers (ICANN) is a not-for-profit, multi-stakeholder, international organization that has responsibility for Internet Protocol (IP) address space allocation, protocol identifier assignment, generic (gTLD) and country code (ccTLD) top-level domain name system management, and root server system management functions. ICANN's mission is to coordinate, at the overall level, the global Internet's systems of unique identifiers, and in particular to ensure the stable and secure operation of theseI systems. It coordinates policy development reasonably and appropriately related to these technical functions, consistent with ICANN's core values. Among these values are:

- Preserving and enhancing the operational stability, reliability, security, and global interoperability of the Internet;

- Where feasible and appropriate, depending on market mechanisms to promote and sustain a competitive environment;

- Introducing and promoting competition in the registration of domain names where practicable and beneficial in the public interest; and

- Seeking and supporting broad, informed participation reflecting the functional, geographic, and cultural diversity of the Internet at all levels of policy development and decision-making.

New gTLDs have previously been established based on proposals that were submitted to ICANN during two specific application periods. Materials from the 2000 application round, which led to the delegation of .AERO, .BIZ, .COOP, .INFO, .MUSEUM, .NAME and .PRO, are

available at http://www.icann.org/tlds/app-index.htm. Materials from the 2003 round, which led to the delegation of .ASIA, .CAT, .JOBS, .MOBI, .TEL and .TRAVEL, are available at http://www.icann.org/tlds/stld-apps-19mar04. Applications received during both of these rounds were evaluated on the basis of instructions and criteria contained in the respective RFPs published by ICANN. Applicants that were successful went on to negotiate and enter gTLD agreements with ICANN.

ICANN is seeking to establish String Similarity Examiners to assess each application for a new gTLD to determine:

> *…if the applied-for string so nearly resembles another visually that it is likely to deceive or cause confusion. For the likelihood of confusion to exist, it must be probable, not merely possible that confusion will arise in the mind of the average, reasonable Internet user. Mere association, in the sense that the string brings another string to mind, is insufficient to find a likelihood of confusion.*

The standard will be applied in three sets of circumstances, when comparing:

- Applied-for gTLD strings against existing TLDs and reserved names;
- Applied-for gTLD strings against other applied-for gTLD strings; or
- Applied-for gTLD strings against strings requested as IDN ccTLDs.

Policy discussions indicate such a degree of similarity would increase the risk to Internet users of fraud or other harms based on confusion between the applied-for string and existing TLDs, reserved words, or other applied-for strings, in the context of the criteria specified in the Applicant Guidebook. It is recommended that potential providers review all drafts of the Applicant Guidebook and other resources on the new gTLD program available at http://www.icann.org/en/topics/new-gtld-program.htm.

The number of applications that will be received for new gTLDs is unknown; however, it is estimated to be several hundred or more. It is therefore vital that the provider be able to convene—or have the capacity to convene—as many evaluators as is necessary to evaluate all the applications, in a timely and complete manner. For example, the provider may wish to consider the process it will use to evaluate applications, and how that process will scale if 100, 250, 500, 700 or 900 applications are received. There should be a statement describing how 2000 applications would be processed (even though this is thought to be highly unlikely). The provider should also consider how the number of applications may impact evaluation timeframes and costs of evaluations.

It is expected that there will be more than one application round. Therefore, there may be an opportunity for cyclical work in evaluating applications. In the longer term, the work may become continuous with new gTLD applications being submitted and evaluated at any time.

In addition, given the international nature of the ICANN community and the likelihood that applications will be received for both ASCII and non-ASCII new gTLDs, it will be important that the provider can convene—or have the capacity to convene—globally diverse panels familiar with internationalized domain names (IDNs). A non-ASCII domain name, i.e., an IDN, is one that utilizes characters from the full Unicode set rather than just the “letter-digit-hyphen” characters specified in the original DNS standards. Using IDNs, for example, make it possible to add TLDs in Arabic, Hebrew, Cyrillic, and other scripts. For more information on IDNs, please visit http://www.icann.org/en/topics/idn/.

3 String similarity

The strings proposed by applicants for new gTLDs could be either identical or similar within the meaning of the standard in the previous section to (a) existing TLDs, (b) reserved words, or (c) strings proposed by other applicants in the same round. Strings that are identical to existing TLDs or reserved words are easily identified, and will be rejected during the application process. Strings that are identical to other applied-for strings during the same round are also easily identified, and are automatically placed into contention sets to be handled by the contention resolution process. The design of the DNS makes it impossible to insert more than one instance of the same string into the root, so the importance of detecting and rejecting identical strings is both obvious and uncontroversial.

Less easily identified, and potentially more controversial, are strings that are similar[1], but not identical, to (a), (b), or (c). Both the GNSO (Recommendation 2: "Strings must not be confusingly similar to an existing top-level domain or a Reserved Name") and the GAC (Principle 2.4: "In the interests of consumer confidence and security, new gTLDs should not be confusingly similar to existing TLDs") have identified "confusing similarity" as a reason to reject an applied-for new gTLD string. Policy discussions indicate that the most important reason to disallow similar strings as top-level domain names is to protect Internet users from the increased exposure to fraud and other risks that could ensue from confusion of one string for another. This reasoning must be balanced against unreasonable exclusion of top-level labels and denial of applications where considerable investment has all ready been made. As the top-level grows in number of registrations, drawing too large a circle of "similarity protection" around each existing string will quickly result in the unnecessary depletion of available names. The unnecessary exclusion of names would also tend to stifle the opportunity of community representation at the top-level and innovation.

An algorithm will score visual similarities between strings for each pair of applications, as a partial guidance for determination of the likelihood of string confusion. The String Similarity Algorithm ("algorithm") is available in several character sets. The algorithm is a tool the examiners use to provide one objective measure as part of the process of identifying strings likely to result in confusion. The algorithm is also available to applicants for testing and informational purposes. The algorithm, user guidelines, and additional background information are available at http://icann.sword-group.com/icann-algorithm/.

The algorithm calculates scores for visual similarity between any two strings, using factors such as letters in sequence, number of similar letters, number of dissimilar letters, common prefixes, common suffixes, hyphenation, and string length. The algorithm will provide rank ordered visual similarity scores for each combination (pair) of applied-for and existing TLDs. The algorithm will help the evaluators sort potential cases of similarity and provide some evidence of the degree of similarity between pairs of strings. It is not a determinative tool.

During the Evaluation of applications, String Similarity Examiners will review all applied-for strings and decide whether the strings proposed in any pair of applications are so similar to another applied-for string (based on the standard above) that they should be placed into direct string contention as part of a contention set. (Applied for strings that are found to be so similar to an existing TLD or Reserved Word that the standard is met will be rejected.) Such a determination, based on human judgment assisted by criteria and algorithm outcomes, is performed for each pair of applications. When all applications have been checked in this way, the outcome is a matrix of direct string contentions between pairs of applications. Applications

[1] Although two strings might be "similar" in appearance, sound, or meaning, for the purposes of this RFI, only *visual* similarity (appearance) is relevant.

without any string contention can proceed without further action, but contention must be resolved for all others.

4 Criteria

ICANN anticipates expressions of interest (i.e., answers to questions posed in section 5 below) from providers to conduct the string similarity examination of applications must meet the following criteria:

1. The provider must be an internationally recognized firm or organization with significant demonstrated expertise in the evaluation and assessment of this type of similarity. The provider must demonstrate experience in making decisions of this level of importance: balancing the potential for exposing Internet users to increased risks of fraud and other harms versus unnecessary exclusion of Internet labels and denial of applications where considerable investment has been made.

2. The provider must be competent to assess (or be able to call upon experts who are competent to assess) the way in which changes to the top level (root zone) of the DNS might increase the exposure of Internet users to fraud or other harms.

3. The provider must be able to convene (either in advance or rapidly on-demand) a linguistically and culturally diverse panel capable, in the aggregate, of evaluating Applications from a wide variety of different communities, which may:
   - be local or global in scope;
   - be based on geography, political affiliation, common interests, or other factors;
   - involve either commercial or non-commercial interests (or both); and
   - be either objectively defined or self-defining.[2]

4. The provider must propose a structure and plan for the string similarity panel that is viable for a varying number of Applications, and how that process will scale if 100, 250, 500, 700 or 900 applications are received. There should be a statement describing how 2000 applications would be processed (even though this is thought to be highly unlikely). The number of comparisons will increase exponentially with increase in the number of applications and, as the number of Applications increases, the number of those that will present string similarity issues, may also increase but this number will not be known in advance.

5. Considering the string similarity criteria defined in Module 4 of the Applicant Guidebook and described in Section 3 of this document, the provider must propose a panel that is capable of:
   - exercising subjective judgment in making its evaluations,
   - reaching conclusions that are compelling and defensible, and
   - documenting the way in which it has done so in each case.

[2] An example of an objectively defined community is “the registered voters in the city of Perth, Australia”; an example of a self-defining community is “people who are interested in dogs.”

6. The provider must convene and operate the string similarity panel so as to prevent communication between the panel (or any of its members) and any party with an interest in the Applications being evaluated, except as may be explicitly permitted by the process as defined in the Applicant Guidebook and to avoid conflicts of interest. The provider selected and each of its evaluators (including any additional experts) will execute a confidentiality agreement with regard to material contained in the applications under review.

7. The provider should be comfortable that the Applicant Guidebook is comprehensive and satisfactorily expresses all selection criteria, but understand that it is not finalized. It is possible, that the provider will be selected before the Applicant Guidebook is finalized, it will have the opportunity to review the text to ensure that the basis for the evaluation is clear. The criteria must be objective, measurable, publicly available at the outset of the evaluation process, and described fully in the Applicant Guidebook. All applications will be evaluated against these criteria.

8. The evaluation process for selection of new gTLDs will respect the principles of fairness, transparency, avoiding potential conflicts of interest, and non-discrimination.

5 Response to EOI Requirements

Interested parties should respond to each of the eight subject areas below. Responses will be gauged on the basis of the criteria defined in this document and Applicant Guidebook. Candidates desiring to express their interest to ICANN in the comparative evaluation role in the new gTLD program should provide the following:

1. A Statement of Suitability that includes a detailed description of the candidate's ability to perform the work described in the previous section which demonstrates knowledge, experience and expertise, including but not limited to projects, consulting work, research, publications and other relevant information.

2. Evidence of the candidate's knowledge of and familiarity with ICANN, its role, structure and processes, including the Internet's Domain Name System (DNS) and past gTLD application and evaluation rounds.

3. The *curriculum vitae* for each person proposed by the candidate to manage or lead work on this project, the candidate's selection process for persons being proposed to ICANN, and explanation of the role that each named person would play. Also indicate the experience and availability of proposed evaluators. ICANN will consider the professional background of available and proposed panelists prior to selecting a provider in order to assess their areas and level of expertise and to identify any conflicts that would prevent them from making an objective evaluation of any application.

4. A warrant that the candidate, if selected, will operate under ICANN's non-disclosure agreement and standard consulting agreement, and that neither the candidate nor any individual who might be engaged to work on this project (whether or not declared pursuant to (4) above) has a known conflict of interest.

5. A statement of the candidate's plan for ensuring fairness, nondiscrimination and transparency.

6. Considering the nature of the expertise necessary for evaluating strings for user confusion on a global scale, a statement of the candidate's plan for ensuring that the examiners will consist of qualified individuals and that the candidate will make every effort to ensure a consistently diverse and international panel.

7. Project and operational timelines.

    a. A proposed work schedule for planning and starting panel operations including key milestone dates, consistent with but more detailed than those specified in this document.
    b. Projected targets for the time frame necessary for it to complete a thorough and careful evaluation of all applications. Identification of volumes of applications that can be processed in those timeframes and at what volumes batch processing of applications might be necessary.

8. Costs. The candidate should provide a detailed statement of the proposed fee structure, including any variable provisions that may be based on the number of string similarity examinations conducted, the number of examinations that involve IDNs, or other factors.

6 Deadline

Interested providers must submit expressions of interest by email to string-sim-eoi@icann.org by 13 April 2009 23:59 UTC. A confirmation email will be sent for each response received.

Also send queries regarding this request to string-sim-eoi@icann.org will be accepted until 3 April 2009, 23:59 UTC. Queries and answers will be posted to a page on the ICANN web site dedicated to this purpose.

If selected, the successful candidate is expected to be ready to assist ICANN with finalization of the Applicant Guidebook, prepare for the evaluation phase, and be ready to start the evaluation within four months after release of the final Applicant Guidebook.

Thank you for your interest.