



New gTLD Program: Trademark Clearinghouse

Explanatory Memorandum: Implementing the Matching Rules

Date of publication: 24 September 2012

Updated: 14 July 2016

Since the approval of the New gTLD Program, ICANN has undertaken an open, inclusive and transparent implementation process to address stakeholder concerns, including the protection of intellectual property and community interests, consumer protection, and DNS stability.

When the community raised a set of new gTLD "over-arching issues," ICANN in each case convened teams of experts in the relevant fields to develop sets of solutions. In a manner unique to the ICANN model, solutions were recommended, discussed by the broader Internet community, and approved by the Board for additional review as part of the draft Applicant Guidebook.

Teams of recognized experts were convened in the areas of intellectual property, consumer protection, DNS market economics, registry operations, linguistics and internationalized domain names, and root server stability. This multi-year public participation process included consultations with governments, businesses, NGOs, law enforcement, and the at-large Internet community.

The Board formed the Implementation Recommendation Team (IRT) comprised of 18 intellectual property experts to develop specific rights protection mechanisms for new gTLDs. The IRT final recommendations were reviewed by a cross-constituency based (Special Trademark Issues or STI) team to provide a multi-stakeholder consensus view. These combined efforts produced an enhanced set of trademark protections for new gTLDs that have been further improved through the participation of many in the broader Internet community, including a number of national governments via participation in ICANN's Governmental Advisory Committee.

The implementation model introduces significant new rights protection mechanisms that are not available in the current space. These include the Trademark Clearinghouse, a single database of authenticated registered trademarks will provide data to support trademark claims and sunrise services. The Clearinghouse will be operated by a third party under license or agreement with ICANN and replaces the need for trademark holders to register in many databases as new gTLDs are launched. ICANN will require that every new gTLD operator utilize the Clearinghouse and conduct both a Trademark Claims and a Sunrise Process.

Note: This memorandum has been updated since original publication to provide additional detail on the currently implemented matching rules.

1 Key Points

- Some trademarks entitled for inclusion in the Trademark Clearinghouse include characters that are impermissible in the domain name system (DNS) as domain names.
- The Clearinghouse will change certain DNS impermissible characters in a trademark into DNS permissible equivalent characters, as described in the Applicant Guidebook, for the mandatory Trademark Claims and Sunrise services.
 - The Commercial At ("@", U+0040) symbol can be omitted, replaced by a hyphen, or replaced by the canonical translation of the word 'at' in any official language of the jurisdiction from which the trademark right is protected and verified.
 - The Ampersand ("&", U+0026) symbol can be omitted, replaced by a hyphen, or replaced by the canonical translation of the word 'and' in any official language of the jurisdiction from which the trademark right is protected and verified.
 - Any other impermissible character can be omitted or replaced by a hyphen.
- The transformation rules interact; as the number of impermissible characters increase linearly, the number of permissible variations increases exponentially.
- Rights holders will have the ability to select which of the permissible variations are to be used by the Clearinghouse as matching domain name labels for sunrise and trademark claims services.

2 Introduction and Background

The Trademark Clearinghouse ("Clearinghouse") is a system that facilitates certain rights protection mechanisms in the new gTLD namespace, specifically relating to the registration of domain names. Trademarks can include elements that are not able to be represented in the DNS at the present time. For example, trademarks in many jurisdictions around the world can include colors, punctuation, and design elements such as graphical images.

In contrast, the DNS does not represent design elements. Domain name labels (the components separated by dots in a fully qualified domain name, such as "example" in the domain name "example.test") can contain only letters, digits and hyphens ("LDH" restrictions). Labels are not permitted to begin or end with a hyphen and are restricted in maximum length.¹

In the case of domain names that contain non-US-ASCII characters, or Internationalized Domain Names (IDNs), the underlying technology in the domain name system continues to function based on LDH-restricted labels. These labels are derived from the non-ASCII string via

¹ For example, see RFC952, RFC1035 and RFC1123.

Punycode transcription. They are interpreted by computer programs such as web browsers and e-mail programs and translated back to the original name when displayed to a user.

For matching purposes, alternative language characters used in the creation of IDNs will not experience matching rule transformations. IDNs will be noted at entry to the Clearinghouse and will be translated into IDNA (IDNA2008) compatible labels. See chapter 5 for a detailed description of the conversion.

Characters that cannot be represented in the DNS are expected to occur in trademarks that are recorded in the Clearinghouse. However, the range of textual elements that can be presented in trademarks is dramatically larger than the range of textual elements that can be presented in domain names.

The Applicant Guidebook's Module 5 chapter on the Trademark Clearinghouse², section 6.1.5, defines the framework of how DNS impermissible characters will be treated for the purpose of determining a match between a trademark string and a domain name label. This section is based on the recommendations of the Special Trademark Issues (STI) work team and describes the mechanism of an "identical match" to define at a high level how the comparisons are to occur.

6.1.5 The Trademark Clearinghouse Database will be structured to report to registries when registrants are attempting to register a domain name that is considered an "Identical Match" with the mark in the Clearinghouse. "Identical Match" means that the domain name consists of the complete and identical textual elements of the mark. In this regard: (a) spaces contained within a mark that are either replaced by hyphens (and vice versa) or omitted; (b) only certain special characters contained within a trademark are spelled out with appropriate words describing it (@ and &); (c) punctuation or special characters contained within a mark that are unable to be used in a second-level domain name may either be (i) omitted or (ii) replaced by spaces, hyphens or underscores and still be considered identical matches; and (d) no plurals and no "marks contained" would qualify for inclusion.

This memo describes each of the elements of section 6.1.5, reviews community input on the implementation of this provision, and illustrates the implementation approach.

2.1 Defining an Identical Match

All Clearinghouse trademark comparisons occur by comparing the textual elements of a mark with the second level label of the domain name being registered. When all and only the complete and identical textual elements exist in both the trademark and the label, it is considered an identical match. For example, the trademark <ICANN> and the domain label "ICANN" to be used in a domain name such as "icann.org" would be an identical match.

² <http://newgtlds.icann.org/en/applicants/agb/trademark-clearinghouse-04jun12-en.pdf>

Four additional criteria, any of which could result in additional matches to a trademark record, are also applicable according to the community-developed requirements.

(a) Spaces contained within a mark that are either replaced by hyphens (or vice versa) or omitted

A mark containing one or more spaces cannot be directly represented in a domain name, as spaces are not permissible DNS characters. Applying this rule, those spaces may be dropped or replaced by hyphens to generate domain names that will identically match the trademark. If a trademark in the phrase "ICANN Example" were to exist and we were considering the top-level domain "test," then the domain names "icannexample.test" and "icann- example.test" would both be considered identical matches to the trademark. Note that if the trademark in question were "ICANN- Example," it would only match "icann- example.test," as the hyphen is a DNS-permissible character and thus is not eligible to be translated into some other value for the purposes of an identical match.

(b) Only two special characters contained within a trademark are spelled out with appropriate words describing it (@ and &)

Two special cases have been defined for initial support in the Clearinghouse: the Ampersand ("&", U+0026) and Commercial At ("@", U+0040) characters. The meaning and translation of these characters, however, is language specific.³ Appendix 1 provides a full overview of possible substitutions per Trademark region (Registered trademarks) and per country (other Clearinghouse allowed marks)

The need for language specificity raises a number of implementation decisions on, for example:

- Which languages should be supported;
- Whether more than one translation should be possible; how many are practical and reasonable to provide for a given trademark;
- Which party or parties should determine which translations to apply to any particular trademark;
- What mechanisms are required to apply those translations?

The issues considered in developing an implementation approach for this "rule b" are discussed in section 2 below.

(c) Punctuation or special characters contained within a mark that are unable to be used in a second-level domain may either be (i) omitted or (ii) replaced by spaces, hyphens or underscores and still be considered identical matches

³ For example, "X&Y" can be read "X and Y" in English; "X y Y" in Spanish; "X und Y" in German, "X et Y" in French, and so forth. This appears to be limited only by the number of languages one wishes to consider.

The Clearinghouse will always allow a DNS-impermissible character in a trademark to be omitted or replaced by a hyphen for identical match purposes. Spaces and underscores are themselves not DNS permissible characters and thus are only being supported by being omitted or changed to a hyphen when the Clearinghouse performs identical matching tests. If a trademark in the phrase "ICANN_Example" (using the low line/underscore, U+005F) were to exist and we were registering in the top level domain "test," then the domain names "icannexample.test" and "icann-example.test" would both be considered identical matches to the trademark.

An optimization became apparent in the detailed implementation analysis. Reconsidering the phrase "ICANN Example" (using the space, U+0020), when applying the rule that any DNS-impermissible character (such as a space) can be omitted or replaced with a hyphen, the results are identical to the application of rule (a). As a result, in implementation, rule (a) will be implemented by performing a rule (c) transformation, which provides the exact same results.

(d) No plural and no "marks contained" would qualify for inclusion.

The definition of "identical match" in the Clearinghouse does not stretch to include plurals or "marks contained." For example, the trademark "ICANN Example" will not identically match the domain names "icann-examples.test" (note the plural) nor would the actual trademark "ICANN" identically match "icann-example.test." Such expansions could easily require the Clearinghouse to perform legal analysis in order to decide whether or not they match, which is outside the desired scope of Clearinghouse activities.

2.2 Critical principles

Reviewing the community discussions leading to the proposed implementation, ICANN identified a set of principles to guide the implementation of the matching rules.

2.2.1 The Clearinghouse must be able to apply the matching rules consistently.

First, the Clearinghouse is designed to verify factual information. It is explicitly beyond the scope and remit of the Clearinghouse to perform legal analysis, provide legal advice or apply discretionary assessments as to the range of domain names that a party owning given a certain trademark may or should be interested in protecting. Secondly, a process that resulted in uneven matching results depending on the different determinations made on appropriate languages or character combinations would undermine the objective of having a single globally reliable system with equivalent treatment across jurisdictions.

2.2.2 Application of the matching rules must be done in a technically feasible and commercially viable manner

The matching process should be automatable to generate predictable and repeatable results.

2.2.3 The approach adopted should provide value for the cost.

The implementation should account for the set of marks that are likely to be most valuable to rights holders, with consideration for special circumstances.

2.2.4 Application of the matching rules must protect trademark rights as agreed during the community development of the Clearinghouse processes, without either unduly expanding or limiting the scope of verified rights.

The IRT and the STI agreed that Clearinghouse processes should protect existing rights, but not seek to expand rights beyond those conferred according to trademark law.

3 Rule B Implementation

3.1 Language support implications

In the initial evaluation of the options, the following criteria⁴ (as agreed upon by the community) were consulted:

- 2.5.3 Use globally accessible and scalable systems so that multiple marks from multiple sources in multiple languages can be accommodated and sufficiently cataloged (database administrator and validator).
- 2.5.4 Accept submissions from all over the world - the entry point for trademark holders to submit their data into the Clearinghouse database could be regional entities or one entity.
- 2.5.5 Allow for multiple languages, with exact implementation details to be determined.

Taken together, these criteria seem to suggest that a broad selection of languages is most desirable. This interpretation was consistent with reactions from rights holders during the Implementation Assistance Group (IAG) about the need for trademark language support. Accordingly, the implementation approach is to support any official language used in a jurisdiction that grants trademark rights.

3.2 Generating the appropriate Rule B translation(s) for a given trademark

The criteria for identical match contained in the Guidebook and how they could be implemented was one of the issues raised for discussion in the IAG. Much of the discussion focused on the special characters in "Rule B" below particularly, since language requirements were not specified in the Guidebook.

To generate consistent and repeatable results, an authoritative translation and spelling for the "&" and "@" characters must be adopted for each language. A process that yields

⁴ <http://newgtlds.icann.org/en/applicants/agb/trademark-clearinghouse-04jun12-en.pdf>

different translations for the same character in the same language would be contrary to the goal of a predictable process.

Assuming that for each supported language, a translation of these characters can be specified, the following are possible methods that have been considered for implementing this rule.

1. Applying every language translation to any trademark containing an “&” or “@” character to yield a set of identical matches. Given the number of languages expected to be supported, this would result in an extraordinarily high number of matches, which could exceed the actual scope of rights pertaining to a trademark, as well as going beyond the set of strings that rights holders are interested in protecting.
2. Applying a standard set of languages (for example, the 6 UN languages) in every case, according to a mapping like the below.

Special Character	Arabic	Chinese	English	French	Russian	Spanish
& (Ampersand)	و	和	And	et	и	y
@ (Commercial)	ع	在	At	A	в	en

This approach would be produce consistency in the matching operations, but could result in under- or over-inclusive results. For example, in the case of a mark not meaningful to speakers of any of the six UN languages, this protection would likely be unsatisfactory to the rights holder. At the same time, this approach would extend protection in every case to languages where the mark might not be meaningful. Essentially, this could offer protection that is not needed and eliminate protection for the most desired domain labels associated with a given mark.

3. Applying the translation of “@” and “&” in the official language(s) of the jurisdiction where the trademark is registered or otherwise protected. Assuming, as above, that an authoritative reference can be created for this to identify the official languages for each jurisdiction, the mapping of these characters to the appropriate language could be effectuated.
4. Enabling the rights holder to choose the most appropriate language(s) for translation of the “@” and “&” characters on a per-record basis. Some initial feedback on the matching rules issues suggested that flexibility is desired, as the rights holder would be in the best position to determine which translation(s) would be most appropriate, and may not value the variation(s) that would be automatically assigned. However, one possible outcome of this approach would be selection of the maximum available set in every case, causing cost and scaling issues for Clearinghouse services.

A combination of options 3 and 4 discussed above seems to provide the best range of protection. That is, a mapping can be made based on an extended table for all official languages in the relevant jurisdictions, indicating which languages should be applied. A

rights holder can select a different language basis for the character translation as considered appropriate. This allows flexibility for cases that do not fall inside the rules (like a case where a trademark is in a different language than those used in the jurisdiction).

Every mark containing a recognized special character is entitled to use the canonical mappings into each of the national language(s) of the jurisdiction in which the trademark rights were recognized. Consistent with Rule "C," dropping the special character or replacing it with a hyphen (like any other DNS impermissible character) is also acceptable.

This will allow rights holders to make the permissible and appropriate translation decisions as part of their entry into the Clearinghouse. Rights holders will have the opportunity to select which translations for each impermissible character they are concerned about. A canonical translation list of jurisdictions and available translations will be developed and posted at the Trademark Clearinghouse website. Note that this translation framework is capable of supporting expansion to support additional special characters, should the need arise, and can readily handle revisions to map additional languages into the canonical translation for any jurisdiction if needed.

4 Implications of Rule Interactions

Rules (b) and (c) both map a single character to two or more options. As a result, they create the potential that a single trademark could result in a substantial number of identical matches, depending on how rules (b) and (c) are interpreted.

In extreme cases, the interaction of the matching rules can result in combinations that create operational complexity. If each character in a domain name is considered independently, then, as the number of non-LDH (letter digit or hyphen) characters in the trademark increases in a linear fashion, the number of Clearinghouse exact matches increases exponentially. Behaviors such as attempting to obtain a sunrise registration by creating a trademark containing a substantial number of special characters and then automatically generating a dramatic number of identical matches as a result, potentially drives changes to the operational cost model for the Clearinghouse.

Consider the fictional trademark "E * X * A * M * P * L * E". This example has 18 DNS impermissible characters (there are 12 U+0020 [space] and 6 U+002A [asterisk] characters). These particular impermissible characters have the possibilities of mapping to a hyphen or being omitted from the label; other specially mapped characters (currently, '@' U+0040 and '&' U+0026) have at least three possible mappings and could potentially have more than 20 possible translation options each, depending on the jurisdiction involved.

The design considered several principled ways in which one could apply the Guidebook matching rules in such a case.

The first option considered was to universally apply each translation: all DNS impermissible characters that are not specially mapped are either dropped or transformed into hyphens under rule (c). This would result in two domain name strings, "e---x---a---m---p---l---e" and

"example" respectively. While this option was appealing as posing the fewest implementation and operational problems, it seems to unduly limit the protection to the rights holder's trademark. A less restrictive option was considered desirable.

The next option was to allow a more relaxed approach for DNS impermissible characters, where any given character value can be transformed uniformly across the string. For example, all U+0020 characters must be dropped, or all U+0020 characters must be transformed to hyphens, but U+002A characters would be treated independently. This resulted in a manageable implementation and operational set of issues (complexity became a function of the number of unique DNS impermissible characters in the trademark), but still seemed to potentially limit the string in undesirable ways. With this approach, the trademark would generate four strings: example, e-x-a-m-p-l-e, e--x--a--m--p--l--e, and e---x---a---m---p---l---e.

Further variations were considered, attempting to identify less restrictive results to ensure that the maximum levels of protection were available to the rights holder that were both technically feasible and congruent with the legal rights verified by the Clearinghouse. The least restrictive system permits each impermissible character to be managed independently, so any given U+0020 (space) could be dropped or transformed into a hyphen, independently of any other character in the string. This has the unfortunate technical side effect of creating exponential growth in the number of possible strings that can be matched. In the example above with 18 impermissible characters, it creates 262.144 possible exact matches.

As those variations are recognized as identical matches by the Clearinghouse, they must be able to be protected upon the Further variations were considered, attempting to identify less restrictive results to ensure that the maximum levels of protection were available to the rights holder that were both technically feasible and congruent with the legal rights verified by the Clearinghouse.

Seen purely from a practical perspective, many of the permutations are not expected to be of particular interest or value to a rights holder. On its face, the domain name "exa---mp-le.tld" may not raise a concern of confusion with the fictional mark "example." Generating substantial numbers of matches when only a handful are of particular interest to protecting the rights holder's interests, would be a massive cost driver that would increase registration costs in the Clearinghouse to a much higher level. At the same time, there may be particular strategies or marks for which such protections are both desired and appropriate. The Clearinghouse is not in a position to make those determinations for a particular case.

Using the broadest interpretation, the number of identical matches to a trademark could be very large. If the underlying system requires that the strings be enumerated at any time (and such an enumeration seems likely to be necessary), one consequence is that it might be necessary to exclude certain matches. There are technical limits to how much data can be stored and represented at a commercially viable price point for the system. Such a decision would impact rights holders and would be best informed by the rights holders' brand protection and expansion strategy; information the Clearinghouse does not have and should

not request. Community input suggested several ways to put a rights holder in some degree of control.

- (a) Developing automation to apply any rules consistently. Based on asking the rights holder to provide the jurisdiction or even the language of the mark, some community input suggested that appropriate translations could be generated automatically by the Clearinghouse software. This option creates a degree of discretion on the part of the Clearinghouse, which could affect the legal rights of a mark holder.
- (b) Asking Clearinghouse personnel to perform the translation as part of the verification process. While this is practical to implement from a technical standpoint, and while Clearinghouse personnel will be interacting with the records as part of the verification process, this approach also creates too large a degree of discretion on the part of the Clearinghouse. Even if it did not create an unacceptable degree of liability, it would also increase the overall cost and complexity of performing verifications.
- (c) Having rights holders suggest appropriate translations for their marks containing special characters as part of their entry into the system. If rights holders are able to select which translations or permutations for identical matches are important to them, this seems to address almost all concerns.

However, the impact of volume is significant. If all marks can generate large numbers of identical matches without limitation, it will become an operational drain on the system and could threaten the commercial viability of the Clearinghouse. As such, the solution provider will have the freedom to price the service in accordance with the number of exact matches selected by the rights holder. In keeping with the principle that parties will bear their own costs, a rights holder will have the option to select those strings that are interesting for protection purposes and to ignore those strings that are not, and should be expected to be charged according to how many exact matches are requested. This treats the marks equitably and fairly, while giving the rights holder the necessary flexibility to protect those domain name strings required by his or her brand protection strategy.

This strategy will result in the most cost-efficient and fair solution for rights holders who will pay for what they ask to receive, and aligns the cost to the complexity of delivering the service, rather than having some marks subsidize the operational cost of other, more complicated marks. Market forces can be expected to moderate both the cost and the complexity of the implementation with this approach.

5 Implementation

Based on this implementation, the Matching Rules can be restated simply as follows:

For purposes of the Trademark Claims and sunrise services, "Identical Match" means that a domain name consists of the complete and identical textual elements of the mark. In this regard:

- (B) Special characters @ and & contained within a trademark may be spelled out with appropriate words; and
- (C) Other special characters contained within a mark that are unable to be used in a second-level domain name may either be: (i) omitted; or (ii) replaced by hyphens.

Plural versions of a mark or domain names containing the mark are not considered an Identical Match for purposes of these baseline services.

The Clearinghouse tries to assist mark holder in specifying their mark name in characters that are valid under IDNA2008. Currently the following references have been used to build the IDNA2008 libraries

- IANA URL <ftp://ftp.iana.org/assignments/idna-tables-6.3.0/idna-tables-6.3.0.txt>
- RFC5892 URL <http://www.ietf.org/rfc/rfc5892.txt>

The mapper function of the Net::IDN::UTS46⁵ library is used to convert non-IDN characters (such as 'Microsoft 'wide' characters which are not allowed by IDN) into a more suitable form. This is to assist end users who do not understand why not all Microsoft word characters are allowed in IDN. This step does not interfere in any way with the standard IDN conversion mechanism.

To verify the compliance of the mapping of the TMCH with IDNA2008 all cases documented in the `ldnaTest6.2.0.txt` (<ftp://unicode.org/Public/idna/6.2.0/ldnaTest.txt>) have been checked and confirmed.

The translation of a mark to its "identical match" labels can be described with following pseudo-code:

```
base := lowercase(Unicode_Normalization_Form_C(mark_name))

reference := empty string
all_combinations := empty list of empty lists
for all characters c in base
    reference := reference + c
    if different direction (IDN2_BIDI)
        end - "Invalid combination of characters - string is bi-
directional"
    if IDN_to_ascii(reference) then
        add list with one element c to all_combinations
    else
        if c is '&' or '@' then
            add list of transliterations (based on type of mark and
jurisdiction), "-" and "" to all_combinations
        else
```

⁵ UTS 46 mapping: see <http://www.unicode.org/reports/tr46>

add list "-" and "" to all_combinations

label_list := all ordered combinations of exact one element from each list from all_combinations that is a valid u-label (and hence can be converted via NFC and IDN2008 to a valid a-label - more specifically the length of the a-label should be less than 64)

Example

```
"abcd" would generate list [ [ 'a' ], [ 'b' ], [ 'c' ], [ 'd' ] ]
    generating one label : abcd
"ab cd" would generate list [ [ 'a' ], [ 'b' ], [ ' ', '-' ], [ 'c' ],
[ 'd' ] ]
    generating two labels : abcd ab-cd
"ab+cd" would generate list [ [ 'a' ], [ 'b' ], [ ' ', '-' ], [ 'c' ],
[ 'd' ] ]
    generating two labels : abcd ab-cd
German mark "ab&cd" would generate list [ [ 'a' ], [ 'b' ], [ 'und', ' ',
'- ' ], [ 'c' ], [ 'd' ] ]
    generating three labels : abcd ab-cd abundcd
```

A list of transliterations for the <@> and <&> characters can be found in the TMCH API documentation (<http://www.trademark-clearinghouse.com/sites/default/files/files/downloads/tmch-api2-2014-08-04.zip> in the transliterations.txt file). It also contains a list of country codes and jurisdictions.

To verify if two labels are identical the LDH form of the label should be used and verified on a character by character basis. Two labels are identical if their LDH representation are a character-by-character exact match.

The entire CNIS service and the Claims service are based on the A-label representation. Accordingly, the normalization of a label that is checked against the TMCH lists is performed by the relevant registry operator.