ICANN Org Report on Languages and Scripts in the Trademark Clearing House (TMCH)

Executive Summary

ICANN org has prepared this report for discussion within the IDN EPDP Working Group and is prepared to answer any questions regarding information contained in this report. Should the Working Group have additional questions and/or research requests, ICANN org will review those requests in terms of available budget and resources.

On 20 May 2021, the Generic Names Supporting Organization (GNSO) Council passed a resolution to initiate the Expedited Policy Development Process (EPDP) on Internationalized Domain Names (IDNs) and adopted its charter.[1] The IDN EPDP charter includes several data and metric requirements, including one requirement for data related to the Trademark Clearinghouse (TMCH)[2]:

> *2. Determine whether certain data is required to help understand a specific issue or answer a charter question (charter question numbers are indicated next to the data points):*
> - *Breakdown of the scripts/languages represented in a validated and active trademark in the TMCH (f1)*

Based on this requirement and discussions with the IDN EPDP Working Group Leadership, ICANN org looked at the following research questions:

- **What languages and scripts exist in the TMCH?[3]**
- **How are variant labels managed and calculated within the TMCH?**
- **If variant analysis is conducted, how many cases of variant analysis are there?**

To answer these questions, ICANN org consulted with ICANN org subject matter experts (SMEs) and the TMCH validator (Deloitte),[4] and collected data from the TMCH, specifically the Domain Name Label (DNL) List[5] and Sunrise Label (SURL) List.[6] The results of ICANN org's research are summarized below.

Languages and Scripts in the TMCH

---

[1] See: https://gnso.icann.org/en/group-activities/active/idn-epdp.
[2] See: https://gnso.icann.org/sites/default/files/file/field-file-attach/idns-epdp-charter-10may21-en.pdf, p. 19.
[3] ICANN org notes that this report focuses on scripts as it is difficult to determine the full set of languages in the TMCH because language information is not supplied by verified trademark holders. Please see the Languages and Scripts in the TMCH section for additional information.
[4] See: https://trademark-clearinghouse.com/content/trademark-clearinghouse-team.
[5] The DNL list is composed of domain name labels that correspond with trademarks that have been verified by and added to the TMCH. This list is used by registries and registrars during the Trademark Claims period and may be accessed by certified organizations, such as ICANN Org. Please see the Domain Name Label (DNL) List section of this report for additional information.
[6] The SURL list consists of domain name labels corresponding with trademarks eligible for the Sunrise Period and have been verified by and added to the TMCH. This list is used by registries and registrars during the Sunrise period. Please see the SURL List section of this report for additional information.

According to the TMCH validator, in November 2021, the TMCH verified 47,058 trademark records from 135 jurisdictions. In the New gTLD TMCH Frequently Asked Questions ([FAQ](#)), in response to the question "What languages will be serviced in Trademark Clearinghouse?", it states that **the TMCH "will verify trademark data from multiple global regions" and "accept trademark data from all over the world, supporting users in multiple languages."[7]** Additionally, the New gTLD Applicant Guidebook states that the TMCH service provider should **"use globally accessible and scalable systems so that multiple marks from multiple sources in multiple languages can be accommodated and sufficiently cataloged"** and **"allow for multiple languages."**[8] The TMCH accepts submissions of trademarks in both Latin and non-Latin scripts[9] and can also accept trademark labels containing two different scripts (e.g., a domain name label that includes both Chinese and Latin script characters).[10] ICANN org notes that it is difficult to determine the full set of languages in the TMCH because language information is not supplied by verified trademark holders.

With each trademark that is authenticated, the TMCH supports both an opportunity for verified trademark holders to register domain names in new gTLDs prior to registration for the general public ("Sunrise Service") and a notification service ("Trademark Claims") at the start of general registration. Additionally, when a trademark is recorded in the TMCH, the record includes domain name labels that correspond to the trademark. **As noted in the Trademark Clearinghouse Frequently Asked Questions (FAQ), "[w]hen you submit your information to the Clearinghouse, part of your file will consist of the different domain names corresponding to that record" and "[u]p to 10 domain names corresponding to one trademark record are included in the initial cost of verification by the Clearinghouse."[11]**

These domain name labels, if the holder of a trademark verified by the TMCH decides to participate in the Sunrise Service or Trademark Claims Period, then appear in the Domain Name Label (DNL) list; for those trademark holders eligible for Sunrise Service period, their labels will also appear in the Sunrise Label (SURL) list.[12] The Sunrise Service period allows holders of verified trademarks, using their signed mark data (SMD) file, to register domain names corresponding with their trademark before General Registration. The Trademark Claims period generally follows the Sunrise Service period. During the Trademark Claims period, any potential domain name registration matching any labels on the DNL list will trigger notification to potential registrants and the trademark holder if the registration proceeds.[13]

---

[7] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/faqs.
[8] See: https://newgtlds.icann.org/en/applicants/agb/trademark-clearinghouse-04jun12-en.pdf, p. 3
[9] See: https://trademark-clearinghouse.com/help/faq/are-trademarks-non-latin-script-accepted.
[10] See: https://trademark-clearinghouse.com/help/faq/can-i-submit-trademark-composed-two-different-scripts.
[11] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/faqs.
[12] The Sunrise and Trademark Claims periods are the minimum rights protection mechanisms requirements that registry operators must execute when launching their New gTLD. Registry operators may also include other Qualified Launch Programs. For additional information regarding Qualified Launch Programs, see: https://newgtlds.icann.org/en/announcements-and-media/announcement-10apr14-en.
[13] See Trademark Clearinghouse, Results & Discussion, and Appendix B for more information on the TMCH and matching rules.

Based on this, when looking at scripts in the TMCH, it is important to distinguish between trademarks and domain name labels. For each trademark, there may be a certain number of labels that are generated (or none), and which will appear in the DNL. Additionally, verified trademark holders may opt to participate in Sunrise Services (though are not obligated to and may choose not to), and those respective labels will appear in the SURL. This means that the number of trademarks recorded in the TMCH may not match the number of labels in the DNL, and the number of labels in the DNL will not necessarily match the number of labels in the SURL.

The table below shows the count of verified trademarks in the TMCH by script, as well as the count of labels in the DNL and SURL list. **As shown below in the [Results & Discussion](#) section, the number of non-Latin trademarks and labels represent a small number of all verified trademarks and labels in the TMCH.** For example, the number of Han labels in the DNL is 377, which represents around 1 percent of all labels in the DNL.

Executive Summary Table 1: Number of Trademarks and Labels by Script

Script	Number of Trademarks (Out of 47,058 TMCH Records)	Number of Labels in DNL (Out of 34,143 Labels)	Number of Labels in SURL (Out of 27,337 Labels)
Arabic	102	39	32
Bengali	3	0	0
Common[14]	92	0	0
Cyrillic	158	75	50
Devanagari	10	3	1
Georgian	1	2	0
Greek	5	1	1
Han	901	377	309
Hangul	75	50	46
Hebrew	33	8	4
Hiragana	11	1	0
Katakana	137	64	40

---

[14] According to the TMCH validator, "common" trademark scripts are trademarks that consist of numbers.

Latin[15]	45470	800	591
Mixed Script[16]	46	80	34
Thai	14	5	0
Total	47058	1505	1108

Variants in the TMCH

The handling of variant labels of trademarks is dependent upon the policies and approved Registry Services of the gTLD registry operator that may, but is not required to, implement a Registry Service for the allocation and registration of IDNs and potential IDN variants. The Rights Protection Mechanism (RPM) requirements related to the TMCH, specified in Section 1 of Specification 7 to the Registry Agreement make references to how a registry operator may handle the registration or allocation of IDN variants. Registries control their own policies in regard to whether they offer IDN registrations, which scripts or languages they support, and how variant code points are handled. For example, Section 2.4.2 of the RPM requirements, states **when a registry has implemented IDN variant registration policies, the registry operator can allocate IDN variants if registration policies are based on published IDN tables and are imposed consistently across the differing registration periods** (e.g., Sunrise to General Registration).[17] Additionally, the RPM requirements state that, **in cases where a registry operator has established variant policies, the registry operator must check all labels in a variant set against the DNL list before any domain names in the set are registered.**[18]

There is no mechanism for variant analysis in the TMCH. While the matching rules do note that they allow for multiple languages–"[t]he implementation approach is to support any official language used in a jurisdiction that grants trademark rights"[19]**–the matching rules and the TMCH do not account for IDN variant labels.**

---

[15] The TMCH validator includes both IDN and non-IDN trademarks in the "Latin" category. Therefore, Latin in this table accounts for all other trademarks that are not in a non-Latin script. In the columns, and subsequent tables, referring to the DNL and SURL, Latin refers only to those Latin labels that include an IDN character (e.g., 'ä').

[16] More detail on "mixed scripts" is provided in the Results & Discussion section. See also Footnote 40.

[17] Section 2 of the TMCH Rights Protection Mechanism Requirements states that General Registration will begin "on the first day following the Sunrise Period in which domain names are generally made available to all registrants that are qualified to register domain names within the TLD." For additional information, please see: https://newgtlds.icann.org/en/about/trademark-clearinghouse/rpm-requirements-14may14-en.pdf, p. 2, 7-8.

[18] *Ibid.*, p.12.

[19] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/matching-rules-14jul16-en.pdf, p. 6.

Background

Data and Metric Requirements for the EPDP on IDNs

The IDN EPDP charter includes several data and metric requirements, including one requirement for data related to the Trademark Clearinghouse (TMCH)[20]:

> *2. Determine whether certain data is required to help understand a specific issue or answer a charter question (charter question numbers are indicated next to the data points):*
> - *Breakdown of the scripts/languages represented in a validated and active trademark in the TMCH (f1)*

This requirement forms the basis of the research questions examined in this report, which are discussed [below](#).

Internationalized Domain Names (IDNs) and Variants

Internationalized Domain Names (IDNs) are domain names that use characters other than the traditional American Standard Code for Information Interchange (ASCII) characters (a-z, A-Z, 0-9, and the hyphen [-]), including those names containing Latin letters with diacritical marks or characters from non-Latin scripts (e.g., Arabic or Chinese). As detailed in ICANN org's report "Internationalized Domain Names (IDNs): Where are we now?"[21] from June 2021, there are 62 country-code top-level domain (ccTLD)[22] IDNs (61 have been delegated and 1 pending delegation) and 92[23] generic top-level domain (gTLD) IDNs. These TLDs represent 23 scripts; the TLD administrators self-identified 37 languages for those TLDs. Regarding second-level domain (SLD) IDNs, as of 2019, there were nearly 9 million SLD IDN registrations. Of these, only about 1.6 million registrations were under gTLDs, meaning most were under ccTLDs. The IDN registrations under gTLDs were spread across 449 gTLDs, with Chinese having the most registrations.[24]

IDN variants, as defined in the *Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels*, are an:

---

[20] See: https://gnso.icann.org/sites/default/files/file/field-file-attach/idns-epdp-charter-10may21-en.pdf, p. 19.

[21] See: https://www.icann.org/en/system/files/files/idns-where-are-we-now-16jun21-en.pdf, p. 3-4.

[22] The TMCH is not required to accept nor is it (at this time) available to ccTLDs.

[23] Though the June 2021 report states that there are 93 gTLD IDNs, there are, as of January 2022, 92 gTLD IDNs due to contract terminations. For an up-to-date list of delegated gTLD IDNs, please visit: https://www.iana.org/domains/root/db.

[24] *Ibid.*, p. 5-8.

alternate code point (or sequence of code points) that could be substituted for a code point (or sequence of code points) in a candidate label to create a variant label that is considered the "same" in some measure by a given community of Internet users.[25]

IDN variants can also be defined as "an alternative domain name that can be perceived as the same thing."[26] The registration of such variants at the second level is dependent upon any applicable policies, including registry policy, contractual requirements such as IDN implementation guidelines,[27] and the approval of a Registry Service allowing such registrations; a registry's policy may require, for example, that a variant is registered to the same registrant or that it must be blocked.

However, as stated in *IDN Variant TLD Implementation: Motivation, Premises and Framework*, variants are difficult to define as "the definition of variant label is arbitrary and based on what is defined by the user-community, so cannot be derived algorithmically or technically."[28] This is demonstrated in the following examples:

> according to the Arabic script community, شبكہ would be a variant of the label شبكة even though the two are visually distinct, while the Khmer script community has suggested that strings like ស្ត្ប(= ស្ប� ñ) and ស្ត្ប(= ស្ប� ជ) should be considered variant labels because they are visually identical, even though they are inherently different code point sequences.[29]

The Trademark Clearinghouse

As part of implementation of the New gTLD Program, ICANN sought to develop a mechanism that would enable trademark holders to protect their rights during the expansion of the Domain Name System (DNS) via the Program.[30] Together with the ICANN community, ICANN developed the ICANN Trademark Clearinghouse (TMCH), which is "a central repository for information to be authenticated, stored, and disseminated, pertaining to the rights of Trademark Holders"[31] and serves "the goal of providing protection for verified legal rights."[32] The Trademark Clearinghouse's basic function is to authenticate information from rights holders and provide this information to registries and registrars. The major benefits of recording a trademark with the

---

[25] See: https://www.icann.org/en/system/files/files/draft-lgr-procedure-20mar13-en.pdf, p. 8.

[26] See: https://www.icann.org/resources/pages/glossary-2014-02-04-en#i.

[27] For the current IDN Implementation Guidelines, please see: https://www.icann.org/resources/pages/implementation-guidelines-2012-02-25-en.

[28] See: https://www.icann.org/en/system/files/files/idn-variant-tld-motivation-premises-framework-25jan19-en.pdf, p. 6.

[29] *Ibid.*

[30] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse. As of November 2021, ICANN has added 1,239 new top-level domains to the DNS (see: https://newgtlds.icann.org/en/program-status/statistics).

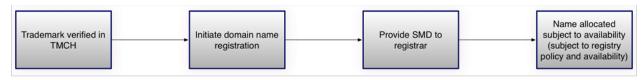[31] See: https://datatracker.ietf.org/doc/html/draft-ietf-regext-tmch-func-spec, p.6.

[32] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/rpm-requirements-14may14-en.pdf, p.1.

Clearinghouse include access to a priority registration period (during "Sunrise Services" period) and notification from the Clearinghouse when a domain matching a trademark has been registered (during "Trademark Claims Service" period), both of which are pictured below.[33]
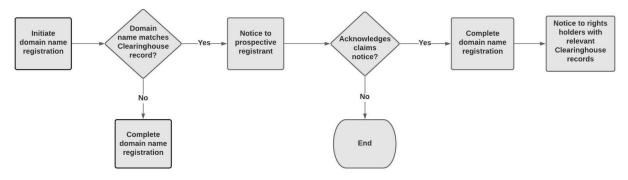
Sunrise Services: Priority access for rights holders to request domain names associated with trademark(s).

Image 1: Sunrise Services

Trademark Claims Service: Notification to a registrant of matching trademark records; notification to Rights Holder after registration if registrant proceeds

Image 2: Trademark Claims Service

When a trademark is recorded in the TMCH, the record can include domain name labels that correspond to the trademark. As noted in the Trademark Clearinghouse Frequently Asked Questions (FAQ), "[w]hen you submit your information to the Clearinghouse, part of your file will consist of the different domain names corresponding to that record" and "[u]p to 10 domain names corresponding to one trademark record are included in the initial cost of verification by the Clearinghouse."[34] This means that any registration matching any of those labels will trigger notification to potential registrants prior to completing the registration, and notification to the trademark holder if the registration proceeds.

Together with the ICANN community, ICANN developed matching rules to guide how domain names (labels) associated with a recorded trademark are created and how to determine

---

[33] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse.
[34] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/faqs.

whether a domain name being registered is a match to a verified trademark in the TMCH.[35] In general, the matching rules are based on determining "identical matches". For example, ICANN's trademark (which is "ICANN") and the domain label being used in a domain name, such as "icann.example", would be an identical match. The rules further state:

> 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. [36]

However, because trademarks can contain elements that cannot be represented in the DNS (e.g., punctuation), the matching rules also allow for "transformations" such as changing a space to a hyphen–to determine "identical" matches.[37] For example, "ICANN TMCH" might be transformed to "icann-tmch" as a domain name label.

The labels generated from these matching rules then appear in the Domain Name Label (DNL) list, which is a list of all labels covered by a trademark in the TMCH with support for Trademark Claims; for those trademark holders eligible for Sunrise Services, their labels will also appear in the Sunrise Label (SURL) list, which is the list of labels covered by a trademark in the TMCH and eligible for Sunrise.[38]

The Appendix to this report contains additional information regarding the TMCH, including: basic functions of the TMCH, eligibility, sunrise and claims services, and criteria for matches.

Research Questions

Based on the IDN EPDP Data and metric requirement described above, ICANN org looked at the following core research questions:

- **What languages and scripts exist in the TMCH?[39]** Specifically, how many trademarks exist in the TMCH for each language and/or script ?
- **How are variant labels managed and calculated within the TMCH?** Is any kind of IDN variant analysis conducted on trademarks entering the TMCH? If so, what is the mechanism for doing so? Additionally, do incoming trademarks already contain variants? Or is variant analysis conducted after the trademark has been entered into the TMCH?
- **If variant analysis is conducted, how many cases of variant analysis are there?** Finally, based on the findings related to the above questions, how many trademarks in

---

[35] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/matching-rules-14jul16-en.pdf.
[36] *Ibid*., p. 3.
[37] *Ibid.,* p. 8.
[38] See: https://datatracker.ietf.org/doc/html/draft-ietf-regext-tmch-func-spec#section-3, p. 5-6.
[39] As previously noted, ICANN org notes that it is difficult to determine the full set of languages in the TMCH because language information is not supplied by verified trademark holders.

the TMCH have variants, what are those variants, and how many instances of trademarks coming in are matched with a variant label?

Methodology

ICANN org gathered statistics and data from the TMCH provider, internal reports, and the December 2021 Domain Name Label (DNL) List and Sunrise List (SURL) to answer the research questions above.

As part of its research, ICANN org also surveyed historical websites and documents pertaining to IDNs, variant labels, and the TMCH. These websites and documents are used to provide a foundation and understanding of the TMCH and IDNs. The information gathered from these websites and documents has been used to provide explanations and examples of IDNs, variant labels, and TMCH functions and periods. Additionally, these websites and documents are used to provide some statistical information regarding scripts in the TMCH and also provide informative data regarding variant labels in the TMCH. The websites and documents referenced throughout this report include SAC060, the New gTLD webpage on the TMCH, the TMCH website, and IETF documents.

Results and Discussion

This section contains the results of ICANN org's research related to the questions above, including accompanying analysis and discussion.

Languages and Scripts in the TMCH

Languages in the TMCH are difficult to account for without a careful analysis that would require an inquiry with verified trademark owners on the intended language of the trademark. Such an analysis would also be faulty due to the fluctuating nature and understanding of languages. For example, the trademark holder of the "Nescafé" name could say that it has no single language or is part of multiple languages. However, ICANN notes that in some documentation related to the TMCH, the terms "language" and "script" might be used interchangeably. For the discussion below, ICANN org would like to note that the primary focus of this report is on scripts rather than languages.

The TMCH was designed to handle trademarks from multiple jurisdictions. The New gTLD Applicant Guidebook states that the designated TMCH service provider should "use globally accessible and scalable systems so that multiple marks from multiple sources in multiple languages can be accommodated and sufficiently cataloged" and "allow for multiple languages."[40] According to the New gTLD Program webpage, the TMCH "will verify trademark data from multiple global regions" and "accept trademark data from all over the world,

---

[40] See: https://newgtlds.icann.org/en/applicants/agb/trademark-clearinghouse-04jun12-en.pdf, p. 3.

supporting users in multiple languages."[41] Additionally, the TMCH accepts trademarks in both Latin and non-Latin scripts[42] and can even, in some cases, accept trademark labels in two different scripts (e.g., a domain name label that includes both Han and Katakana scripts).[43]

As of November 2021, the TMCH validator reportedly received and verified 47,058 trademark records from 135 jurisdictions. Additionally, according to the TMCH website, as many as 291,736 claims notifications have been provided. The TMCH has accepted trademarks globally, including trademarks in Arabic, Bengali, Cyrillic, Devanagari, Georgian, Greek, Han, Hangul, Hebrew, Hiragana, Katakana, Thai, and mixed scripts.[44]

The numbers in the tables below are divided into the following categories: Trademarks; Labels in the Domain Name Label (DNL) list; and, Labels in the Sunrise Label (SURL) list.

Trademarks

Eligible trademark holders may submit their trademarks to the TMCH to be verified and added to the Trademark Database (TMDB).[45] As noted above, the trademark could be in any number of scripts. It is important to note that the "language" of the trademark is not submitted, and cannot be reliably assumed from the script or scripts used. The table below shows the breakdown of trademarks by script, according to the TMCH validator.[46] The TMCH currently has verified trademark records in 15 scripts, including trademarks with mixed scripts.[47] Not including Latin[48], the Han script (which is used in the Chinese, Japanese, and Korean languages) is represented in the largest number of non-Latin script trademarks. Overall, non-Latin script trademarks make up a small portion (3.179 percent) of total verified trademarks in the TMCH.

---

[41] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/faqs.

[42] See: https://trademark-clearinghouse.com/help/faq/are-trademarks-non-latin-script-accepted.

[43] See: https://trademark-clearinghouse.com/help/faq/can-i-submit-trademark-composed-two-different-scripts. See also Footnote 40.

[44] See: https://trademark-clearinghouse.com/content/stats-0.

[45] For additional information regarding eligibility for the Trademark Claims Period, please see the TMCH Eligibility and Rights Protection Mechanisms section under Appendix B.

[46] See: https://trademark-clearinghouse.com/content/stats-0.

[47] As noted in Guideline 5 of the *Guidelines for the Implementation of Internationalized Domain Names Version 3.0*: "All code points in a single label will be taken from the same script as determined by the Unicode Standard Annex #24: Script Names <http://www.unicode.org/reports/tr24>. Exceptions to this guideline are permissible for languages [scripts] with established orthographies and conventions that require the commingled use of multiple scripts. Even in the case of this exception, visually confusable characters from different scripts will not be allowed to co-exist in a single set of permissible code points unless a corresponding policy and character table is clearly defined." See: https://www.icann.org/resources/pages/idn-guidelines-2011-09-02-en.

[48] The TMCH validator includes both IDN and non-IDN trademarks in the "Latin" category. Therefore, Latin in this table accounts for all other trademarks that are not in a non-Latin script. In subsequent tables regarding the DNL and SURL, Latin refers only to those Latin labels that include an IDN character (e.g., 'ä').

Table 1: Number of Trademarks per Script[49]

Trademark Script	Number of Submitted Trademarks	Total Percentage of non-Latin script Trademarks[50]	% of all Trademarks *(Out of 47,058 marks)*
Arabic	102	6.82	0.217
Bengali	3	0.20	0.006
Common[51]	*92*	*n/a*	0.196
Cyrillic	158	10.56	0.336
Devanagari	10	0.67	0.021
Georgian	1	0.07	0.002
Greek	5	0.33	0.011
Han	901	60.32	1.915
Hangul	75	5.01	0.159
Hebrew	33	2.21	0.070
Hiragana	11	0.74	0.023
Katakana	137	9.16	0.291
Latin[52]	*45470*	*n/a*	96.625
Mixed Script[53]	46	3.07	0.098
Thai	14	0.94	0.030
Total	47058	100	100

---

[49] See: https://trademark-clearinghouse.com/content/tmch-stats.

[50] Not including "Latin" or "Common". See Footnote 41 and Footnote 44, respectively.

[51] According to the TMCH validator, "Common" trademark scripts are trademarks that consist of numbers. The "Common" category is accounted for in this table due to its inclusion in the report provided by the TMCH validator. This category is excluded from the non-Latin script trademarks category because the numbers represented use ASCII code. For example, the TMCH validator used the hypothetically verified trademark for the Spanish liquor "Licor cuarenta y tres (43)". This trademark would be categorized as "Common" but represented in ASCII code.

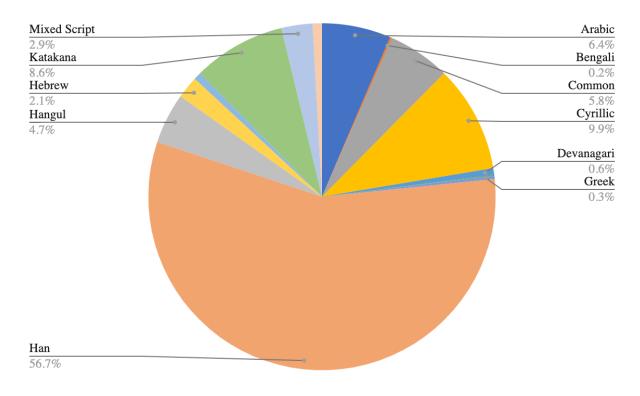[52] See Footnote 41.

[53] See Footnote 40.

Chart 1: Percentage of Trademarks Submitted by Script (excluding Latin)

Label	Value
Arabic	6.4%
Bengali	0.2%
Common	5.8%
Cyrillic	9.9%
Devanagari	0.6%
Greek	0.3%
Han	56.7%
Hangul	4.7%
Hebrew	2.1%
Katakana	8.6%
Mixed Script	2.9%

Domain Name Label (DNL) List

As explained in the Background section, with each trademark record, domain name labels (DNL) are generated based on matching rules. The corresponding labels are placed on the DNL list and are used by registries during the Trademark Claims Period to determine whether a requested domain name matches a record in the Clearinghouse. The Trademark Claims Period occurs during the first 90 days of general registration for a new gTLD and may be extended depending on the registry operator's policies.[54] During the Trademark Claims Period, the prospective registrar provides a notice to prospective registrants of the trademark record, and a notice to trademark holders if the registration proceeds. Due to label matching rules, variations, and the trademark holders' preference, the numbers on the DNL list do not necessarily match the number of trademarks that have been submitted and verified. For example, a trademark holder may opt to withhold the use of their verified trademark from the Claims service, resulting in no label additions to the DNL list for that particular trademark record. In contrast, for another trademark record, several labels could be generated, all of which would appear on the DNL list.

---

[54] The Trademark Claims Period must last at least 90 days. Depending on the registry operator's policies, the trademark claims period may exceed the 90 day minimum. For information regarding New gTLD launch periods, see: https://newgtlds.icann.org/en/program-status/sunrise-claims-periods.

As shown below, there are a total of 18 script/mixed-script labels found on the DNL list. Greek, Han-Hiragana-Katakana, Han-Katakana, and Hiragana each have 1 label on the DNL list. There are 5 script labels that are mixed scripts in both the DNL list and SURL list; these script labels are Cyrillic-Latin, Greek-Latin, Han-Hiragana-Katakana, Han-Katakana, and Han-Latin.[55] The average string count per script on the DNL list is 83.4, but, as seen in Table 3, the string count for a script label can range from 1 to 800. As noted in Table 3, the Latin script makes up more than half of the total number of strings on the DNL list with 800 (53.3 percent). Han makes up a quarter of the total number strings with 377 (25.1 percent) strings counted.

Table 2: Total Number of Strings per Label Script in the DNL List

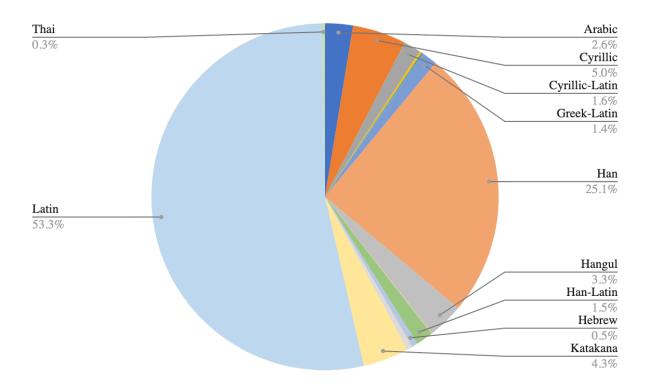| Label Script	Total IDN DNL Count	% of IDN DNL	% of all DNL *(Out of 34,143 labels)*
Arabic	39	2.6	0.114
Cyrillic	75	5.0	0.220
Cyrillic-Latin[56]	24	1.6	0.070
Devanagari	3	0.2	0.009
Georgian	2	0.1	0.006
Greek	1	0.1	0.003
Greek-Latin	21	1.4	0.062
Han	377	25.1	1.104
Hangul	50	3.3	0.146
Han-Hiragana-Katakana[57]	1	0.1	0.003
Han-Katakana	1	0.1	0.003
Han-Latin	23	1.5	0.067
Hebrew	8	0.5	0.023
Hiragana	1	0.1	0.003

---

[55] See Footnote 40.

[56] See Footnote 40.

[57] Related to Footnote 40, in the *Final Proposed Draft Guidelines for the Implementation of Internationalized Domain Names Version 4.0*, it states in the "Additional Notes": "For example, Japanese language normally mixes Han, Hiragana, and Katakana scripts. Also, for Chinese, Japanese and Korean languages, the IDN tables commonly mix "a-z" Latin letters." See: https://www.icann.org/en/system/files/files/idn-guidelines-10may18-en.pdf, p. 5.

IDNA2008_NOTVALID[58]	7	0.5	0.021
Katakana	64	4.3	0.187
Latin[59]	800	53.3	2.343
Thai	5	0.3	0.015
Total	1502	100%	4.399

Chart 2: Percentage of IDN Strings per Label Script in the DNL List

Thai
0.3%

Arabic
2.6%

Cyrillic
5.0%

Cyrillic-Latin
1.6%

Greek-Latin
1.4%

Han
25.1%

Latin
53.3%

Hangul
3.3%

Han-Latin
1.5%

Hebrew
0.5%

Katakana
4.3%

Sunrise Label (SURL) List

The SURL list contains only those labels corresponding to trademark holders that are eligible and have chosen to participate in the Sunrise Period.[60] Like the DNL list, this list contains labels matching trademarks that have been verified by the TMCH. In addition to the submission of

---

[58] IDNA2008_NOTVALID are strings that did not pass the IDN 2008 validation test. For additional information about IDN 2008 character validation guidance, please see:
https://www.icann.org/resources/pages/idna-protocol-2012-02-25-en.

[59] As noted in Footnote 41, "Latin" in this table refers only to those labels that include Latin IDN characters (e.g., ä).

[60] For gTLD Sunrise Period information, see:
https://newgtlds.icann.org/en/program-status/sunrise-claims-periods.

relevant trademark information, the trademark holder must also submit proof of use for eligibility to register a matching domain name in the Sunrise Period.[61] This period allows trademark holders to request their trademark's corresponding domain name prior to its availability to the general public. Not all trademarks qualify for participation in the sunrise period, and some trademark holders choose to opt-out of the Sunrise Period. Depending on varying degrees of eligibility and participation, the count of DNLs on the SURL list will differ from the count of DNLs on the DNL list and from the total number of trademark records that have been verified.

In total, there are 15 script/mixed-script labels found on the SURL list. The list of scripts excludes ASCII-based labels, and does not include labels for Georgian, Hiragana, or Thai. The Devanagari, Greek, Han-Hiragana-Katakana, and Han-Katakana script labels each have 1 string count on the SURL list. The average string count per label script in the SURL list is 74.2, though the string count for a script label can range from 1 to 591. As noted in Table 4, the Latin script makes up slightly more than half of the total number of strings with 591 (53.1 percent) strings counted. Han makes up a little over a quarter of the total number of strings with 309 (27.7 percent) strings counted.
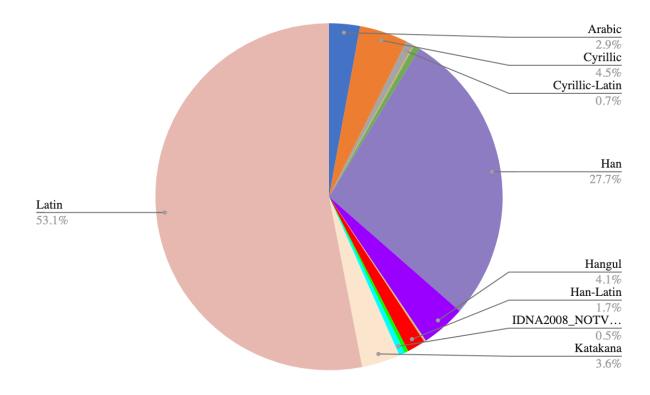
Table 3: Total Number of Strings per Label Script in the SURL list

Label Script	Total IDN SURL Count	% of IDN SURL	% of all SURL *(Out of 27,337 labels)*
Arabic	32	2.9	0.117
Cyrillic	50	4.5	0.183
Cyrillic-Latin	8	0.7	0.029
Devanagari	1	0.1	0.004
Greek	1	0.1	0.004
Greek-Latin	5	0.4	0.018
Han	309	27.7	1.130
Hangul	46	4.1	0.168
Han-Hiragana-Katakana	1	0.1	0.004
Han-Katakana	1	0.1	0.004
Han-Latin	19	1.7	0.070
Hebrew	4	0.4	0.015

---

[61] For additional information regarding eligibility for the Sunrise Period, please see the TMCH Eligibility and Rights Protection Mechanisms section under Appendix B.

IDNA2008_NOTVALID[62]	6	0.5	0.022
Katakana	40	3.6	0.146
Latin	591	53.1	2.162
Total	1114	100	4.075

Chart 3: Percentage of IDN Strings per Label Script in the SURL List

Variant Labels in the TMCH

The registry operator may, but is not required to, implement policies and approved Registry Services for the allocation and registration of IDN variants.[63] If an approved Registry Service

---

[62] IDNA2008_NOTVALID are strings that did not pass the IDN 2008 validation test. Based on the IDNA2008 standard, these labels should be considered invalid for registration purposes. For additional information about IDN 2008 character validation guidance, please see: https://www.icann.org/resources/pages/idna-protocol-2012-02-25-en.

[63] As of December 2021, 144 gTLDs had implemented no policies pertaining to IDN variants; 298 TLDs had implemented policies that allow for IDN variants; and, 431 TLDs had policies that blocked IDN variants. See gTLD Registry Agreements here: https://www.icann.org/en/registry-agreements?first-letter=a&sort-column=top-level-domain&sort-direction=asc&page=1.

allows the Registry Operator to offer IDN registration, the policies of such Registry Service would not have an effect on what labels are generated as part of the TMCH's matching rules.[64] If registries offer IDN registrations, they must also implement IDN tables, which are tables listing the various characters that a TLD supports. If any of the characters in the registries' IDN tables are considered variants, the IDN tables will indicate which characters are variants and which characters it is a variant of. The requirements and responsibilities for calculating variant labels belong to the registry operator and do not apply to the TMCH itself.

Thus, there is no mechanism for variant analysis in the TMCH and thus no cases of variant analysis exist. While the matching rules do note that they allow for multiple languages– "[t]he implementation approach is to support any official language used in a jurisdiction that grants trademark rights"[65]–the rules and the TMCH do not account for IDN variant labels:

> [i]n 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 [letter-digits-hyphens]-restricted labels. These labels are derived from the non-ASCII string via 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. [Emphasis Added]*** IDNs will be noted at entry to the Clearinghouse and will be translated into IDNA (IDNA2008) compatible labels.

For example, if a trademark in traditional Chinese characters is recorded in the TMCH, the matching rules do not define a process for calculating variant labels in simplified Chinese characters (e.g., 愛 vs. 爱 'love').

The Rights Protection Mechanism (RPM) requirements related to the Trademark Clearinghouse specified in Section 1 of Specification 7 to the Registry Agreement make reference to how a registry (but not the TMCH) may handle the registration or allocation of IDN variants. Section 2.4.2. states that in the situation where a registry has implemented IDN variant registration policies, the registry:

> MAY Allocate or register IDN variant labels generated from a label included in a valid SMD file during the Sunrise Period, provided that (i) such IDN variant registration policies are based on the Registry Operator's published IDN tables for the TLD and (ii)

---

[64] See also Appendix A for information on an ancillary service provided by the TMCH (and not mandated or implemented by ICANN org) related to "ongoing notifications" past the Trademark Claims Period. These notifications can include notification of registrations of domains that are variations of a trademark. For example, the TMCH could notify a trademark holder for "trademark" that there was a registration of "trädemark." However, the TMCH only notifies trademark holders of registrations containing variations of Latin characters; notifications would not include variants of non-ASCII scripts.

[65] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/matching-rules-14jul16-en.pdf.

such policies are imposed consistently in the Sunrise Period, any Limited Registration Period, any Launch Program and during General Registration.[66]

Additionally, Section 4.1.2 states that Registries "MAY implement additional matching rules at the TLD level, provided that the Claims Services are still implemented for any Claims Registration satisfying such additional matching rules"[67] and in Section 4.1.3:

> if Registry Operator has established IDN variant policies for Allocation of domain names in the TLD, Registry Operator must check all labels in a variant set against the Domain Name Label List before any domain names in the set are registered.[68]

Closing

ICANN org has prepared this report for discussion within the IDN EPDP Working Group and is prepared to answer any questions regarding information contained in this report. Should the Working Group have additional questions and/or research requests, ICANN org will review those requests in terms of available budget and resources.

---

[66] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/rpm-requirements-14may14-en.pdf, p. 7-8.
[67] See: Ibid, p. 12.
[68] See: Ibid, p.12.

Appendix

Appendix A: TMCH Eligibility, Matching Rules, and Ancillary Services

The TMCH consists of two components, the Trademark Validator (TMV) and Trademark Database (TMDB).

A TMV is an organization that has been authorized by ICANN org to "authenticate and validate registrations in the TMDB ensuring the marks qualify as registered or are court-validated marks or marks that are protected by statute or treaty."[69] Once a trademark has been verified, the trademark data can be added to the domain name label (DNL) list and those participating in the Sunrise Period can have their trademark data added to the sunrise label (SURL) list and be provided with a signed mark data file (SMD), as explained below.

The TMDB is a database that "concentrates the information about the 'verified' Trademark records from the TMVs" and provides "information to the gTLD Registries and Registrars to support Sunrise or Trademark Claims services."[70]
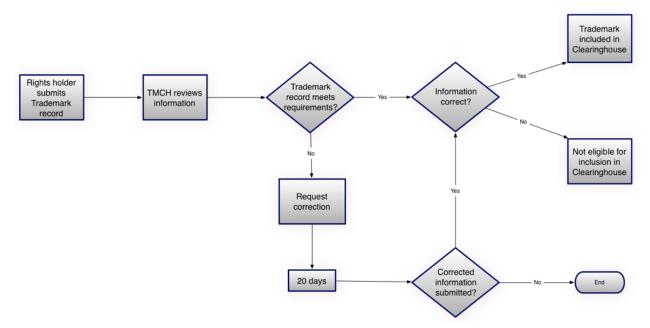
TMCH Eligibility

There are 2 forms of eligibility for participation in the TMCH: Basic Eligibility and Sunrise Eligibility. A trademark holder needs to go through several steps to qualify for Basic Eligibility in the TMCH, which includes the Trademark Claims Period and excludes the Sunrise Period. The trademark holder or their representative needs to submit the trademark record, which includes "trademark information, the class of goods or services, the jurisdiction where the trademark is registered, and other relevant information according to the type of trademark being submitted"[71], to the TMCH where the TMV will review the information. If the information meets the TMV's requirements, the trademark record is added to the TMDB.

If the information does not meet the TMV's requirements or is found to be inaccurate, the TMCH contacts the trademark holder to request corrections. Trademark holders are allotted 20 days to provide their corrections to the TMCH. If the corrected information meets the TMV's requirements the trademark record is added to the TMDB. If the corrected information does not meet the TMV's requirements, the trademark record is not added to the TMDB. Image 1 below provides an overview of Basic Eligibility for the TMCH.

---

[69] See: https://datatracker.ietf.org/doc/html/draft-ietf-regext-tmch-func-spec, p. 7.
[70] See: Ibid.
[71] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/faqs.
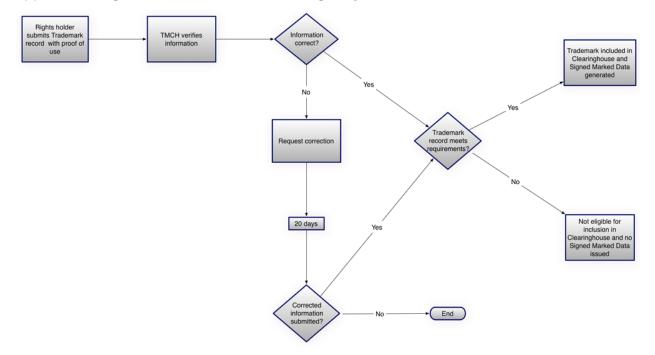
Appendix Image 1: Overview of Basic Eligibility for TMCH[72]

The process to qualify for Sunrise Eligibility is similar to the process for Basic Eligibility. The main difference is the submission of proof of use of the trademark. A trademark holder or their representative submits the trademark record to the TMCH where the TMV reviews and validates the trademark. In addition to the materials that are submitted in the Basic Eligibility process, the trademark holder must provide a "declaration stating that the trademark is indeed being used as [they] say it is" and a "sample, such as an advertisement, a branded product, or some other example showing the trademark in use."[73] If the trademark record information, including the proof of use, is correct and meets the TMV's requirements, the trademark record is added to the TMDB and the trademark holder is provided a SMD that can be used for participation in Sunrise Periods. If the trademark record information, including the proof of use, is not correct or does not meet the TMV's requirements, the rights holder is notified and allotted 20 days to provide any corrections. If the corrected trademark record information, including the proof of use, is verified, then the trademark record is included in the TMCH and the trademark holder is provided with an SMD file. Image 2 below provides an overview of Sunrise Eligibility for the TMCH.

---

[72] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/rights-holders.
[73] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/faqs.

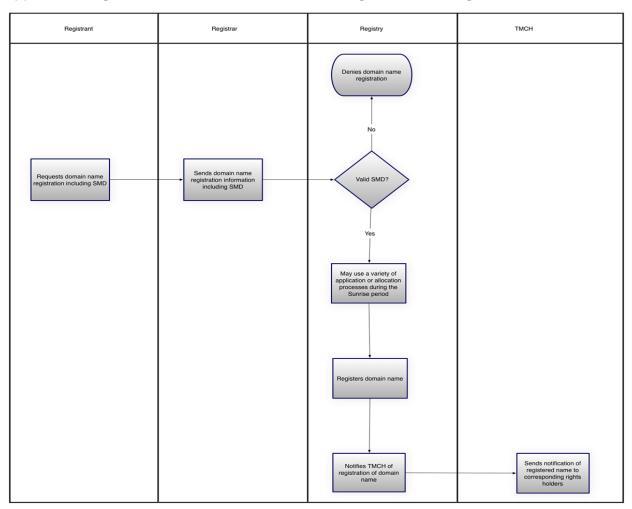Appendix Image 2: Overview of Sunrise Eligibility for TMCH[74]

Additional Information Regarding Sunrise and Claims Periods

There are 2 periods that can be utilized by trademark holders to protect their trademark rights. The first is the Sunrise Period, which "allows trademark holders an advance opportunity to register domain names corresponding to their marks before names are generally available to the public."[75] During the sunrise period, the registrant (trademark holder) submits a request that includes the SMD file for a domain name to the registrar. The registrar submits this information to the registry. The registry checks the validity of the SMD file. If the SMD file is not valid the registration is denied. If the SMD file is valid, the registry registers the domain name and notifies the TMCH of the domain name's registration. The trademark holder subsequently receives a notification from the TMCH that the domain name corresponding with their SMD file has been registered. Image 3 below provides an overview of the sunrise period.

---

[74] See: Ibid.
[75] See: https://datatracker.ietf.org/doc/html/draft-ietf-regext-tmch-func-spec, p 3.

Appendix Image 3: Overview of Domain Name Registration During the Sunrise Period[76]

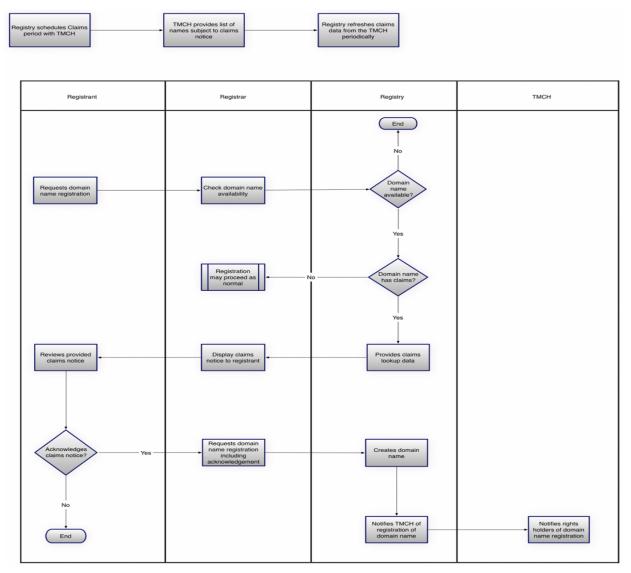| Registrant	Registrar	Registry	TMCH

The second period of the TMCH is the Trademark Claims Period, which is a minimum 90 day period following the sunrise period. During this period, "anyone attempting to register a domain name matching a trademark that is recorded in the ICANN Trademark Clearinghouse (TMCH) will receive a notification displaying the relevant trademark information."[77] This process begins with the registry scheduling a claims period with the TMCH, followed by the TMCH providing a list of names subject to claims notice to the registry. When a registrant tries to register a domain name, the registrar checks the domain name's availability with the registry. The registry determines whether a domain name has trademark claims. If it does not, the registration may proceed as usual. If the domain name has trademark claims, the registrar retrieves the Trademark Claims Notice information and displays the claims notice to the prospective registrant. The registrant reviews the claims notice and determines whether they would like to proceed with domain name registration or discontinue the registration of the domain name. If the registrant proceeds with domain name registration, they must acknowledge the claims notice.

---

[76] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/rights-holders.
[77] See: https://datatracker.ietf.org/doc/html/draft-ietf-regext-tmch-func-spec, p 3.

Once the notice has been acknowledged, the registrar registers the domain name with the registry. When domain name registration is complete, the registry must notify the TMCH within 26 hours of the registration and the TMCH notifies the trademark owner of the domain name registration. Image 4 below provides an overview of domain name registrations during the Trademark Claims Period.

Appendix Image 4: Overview of Domain Registration During the Trademark Claims Period[78]

Trademark holders are able to submit their trademark data to the TMCH prior to and during the launch of a new gTLD. If eligible, the trademark data is verified by the TMV and the trademark holder (or their representative) receives an authentication key (SMD file) that allows them to

---

[78] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/rights-holders.

participate in every new gTLD's sunrise period. Trademark holders may also participate in the trademark claims period, in which the TMCH will notify the trademark holder when domain names corresponding with their verified trademark record have been registered, if the domain name registration is completed, in addition to notifying anyone attempting to register a domain name matching a trademark as mentioned above.

Additional Information Regarding TMCH Matching Rules

The Background section of this report contains information regarding the matching rules–which are used to generate labels corresponding to a trademark record–are compared against potential domain registrations, and have the capability of triggering claims notifications. The rules are largely based on identifying "identical matches", but since some trademark characters cannot be used in domain names, the rules include four additional criteria for identifying corresponding domain label matches[79]:

1. Spaces contained within a trademark that are either replaced by hyphens (or vice versa) or omitted
2. Only two special characters contained within a trademark are spelled out with appropriate words describing it (@ and &)
3. Punctuation or special characters contained within a trademark 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
4. No plural and no "marks contained" [a mark that appears amidst other words] would qualify for inclusion

Ongoing Notifications: An Ancillary Service of the TMCH

A trademark holder also has the option of adding on an Ongoing Notification service, which is one of several ancillary services provided by the TMCH validator - Deloitte - (and may also be provided by other private organizations) and is not required by ICANN org. The trademark holder must manually activate this service in their TMCH account. With this service, and following the Trademark Claims Period, the TMCH will notify a trademark holder of "potential intellectual property infringement indefinitely, beyond the original 90 day period."[80] Similar to the Trademark Claims Period, when an individual has registered a new gTLD domain name that matches the Ongoing Notifications Labels linked to a verified trademark record in the TMCH, the trademark holder will receive an email informing them of this domain name registration.[81]

The Ongoing Notification service provides notifications for domain names that are an exact match to the trademark along with the domain names that contain the trademark, partially contain the trademark, or are similar to the trademark. The TMCH validator provides the following examples on their website:

---

[79] See: https://newgtlds.icann.org/en/about/trademark-clearinghouse/matching-rules-14jul16-en.pdf, p. 3-5.
[80] See: https://trademark-clearinghouse.com/content/ongoing-notifications.
[81] Ibid.

_EXACT: trademarkclearinghouse & trademark-clearinghouse
_CONTAINING: thetrademarkclearinghouse-db (text prior or after the exact match label)
_PARTIAL: clearinghouse & mark-clearing (part of the exact match label)
_SIMILAR: trademarkc1earin6house & tràdémark-çlearinghoüse (a variation of the exact match label using a variant)[82]

This service is not applicable to all variant labels as this only applies to ASCII-derived trademarks and their corresponding script variants, meaning there would be no notifications for variants related to non-ASCII scripts, such as Cyrillic, Greek, or Chinese.[83]

---

[82] *Ibid.*

[83] For additional information regarding types of variants and accepted variants for the TMCH Ongoing Notification service, please see: https://trademark-clearinghouse.com/content/variants.

Appendix C: SAC060 and the TMCH

In 2013, the Security and Stability Advisory Committee (SSAC) released SAC060, the *SSAC Comment on Examining the User Experience Implications of Active Variant TLDs Report*,[84] which was issued in response to a request by the ICANN Board for comments on ICANN's report E*xamining the User Experience Implications of Active Variant TLDs*.[85] The SSAC's comment contained 14 recommendations, three of which referred directly to the TMCH:

- **Recommendation 10**: The current rights protection regime associated with the Trademark Clearinghouse (TMCH) process is susceptible to homographic attacks. The roles of the involved parties, specifically registrars, registries, and TMCH, related to matching must be made clear.
- **Recommendation 12**: The matching algorithm for TMCH must be improved.
- **Recommendation 13**: The TMCH must add support for IDN variant TLDs. Particularly during the TM Claims service, a name registered under a TLD that has allocated variant TLDs should trigger trademark holder notifications for the registration of the name in all of its allocated variant TLDs.

Regarding Recommendation 10, the SSAC notes, as explained in the previous section, that the method currently being used to calculate variants in the TMCH is done with "[v]ariant calculation at the registry level, and checking TMCH for the existences [sic] of marks for variants in the calculated variant set."[86] The SSAC discusses the advantages and disadvantages of this, stating that an advantage of this approach is that the role of the TMCH is to "record existing rights, and not make determinations concerning the scope of particular rights and whether certain (variant) strings qualify for the same right." However, the downside of such an approach, as described by SSAC, is that

> Registries could have different IDN tables, even for the same script. Thus it is possible that miscreants can use the different rules to generate and register variants in other TLDs and cause security, stability, or resiliency concerns or result in squatting and other related issues.

The SSAC notes that a different approach–made explicit in Recommendation 13–would be to have variant calculation conducted via the TMCH, meaning the matching rules would also define "if a variant of the registered string matches" and that "[a] name registered that has variants will trigger trademark holder notifications for the registration of the name or its variants."[87] The SSAC states that such an approach would help to mitigate the downside of the Registry Operator approach.

---

[84] See: https://www.icann.org/en/system/files/files/sac-060-en.pdf.
[85] See: https://www.icann.org/en/system/files/files/active-ux-21mar13-en.pdf.
[86] *Ibid.*, p. 16
[87] *Ibid.*, p. 18.

In the absence of an approach that includes variant calculation in the TMCH, the SSAC recommends in Recommendation 11 that all variant calculations for second-level domains should be done using "all the implemented [Label Generation Rules (LGRs)]" covering a particular script.[88] The SSAC states that without consistent use of the LGRs "it is possible that non-overlapping sets of variants are generated (e.g. one registry generates for the second level domain A variants A1, A2, A3 in tldA and another registry generates variants A2, A3, A4 in tldB)."[89] Bad actors could potentially exploit this situation.

It should be noted, however, that since the SSAC issued this report, the ICANN community has conducted considerable work related to LGRs as well as IDN tables. A Registry Operator can now refer to 46 different LGRs[90] in creating their IDN tables, which are submitted to ICANN for review.[91]

Finally, Recommendation 12 calls for the matching rules in the TMCH to be improved in the way it handles matching, noting that the rules do not consider non-ASCII based scripts:

> "identical match" as defined by TMCH is not really an identical match as in "bit-by-bit" or "character-by-character comparison" as a transformation stage is included before the actual matching . . . the transformation stage currently as specified from is unclear and does not take non-ASCII based scripts into account.[92]

---

[88] *Ibid.*, p. 19.
[89] *Ibid.*
[90] See: https://www.icann.org/resources/pages/second-level-lgr-2015-06-21-en.
[91] See: https://www.icann.org/en/blogs/details/recent-improvements-to-icanns-idn-table-review-process-18-8-2021-en.
[92] *Ibid.*, p. 19.