

**Study of the issues present in the registration of IDN TLDs in GREEK characters**

**By the  
Greek Case Study Team**

**For the  
ICANN Variant Issues Project**

# **Study of the issues present in the registration of IDN TLDs in GREEK characters**

**By the  
Greek Case Study Team**

**For the  
ICANN Variant Issues Project**

## **Greek Case Study Team Members:**

1. Vaggelis Segredakis (Coordinator)
2. Asimina Giannopoulou
3. Catherine Tsapikidou
4. Evangelos Melagrakis
5. Fotia Panayiotou
6. George Papapavlou
7. Giannis Papaioannou
8. Panagiotis Papaspiliopoulos

## Table of Contents

<b>1. Introduction - Disclaimer .....</b>	<b>1</b>
<b>2. Definitions .....</b>	<b>2</b>
<b>3. Useful key points regarding the Greek Language.....</b>	<b>4</b>
<b>3.1. The Greek Language Question .....</b>	<b>4</b>
<b>3.2. The Greek Orthography.....</b>	<b>5</b>
<b>4. Proposed Characters for Registrations.....</b>	<b>5</b>
<b>5. The sigma and final sigma .....</b>	<b>6</b>
<b>6. The IDNA approach of translating Greek words into domain names .....</b>	<b>6</b>
<b>7. Homographs.....</b>	<b>7</b>
<b>8. Homophones.....</b>	<b>8</b>
<b>9. Polytonic Domain Name Variants .....</b>	<b>8</b>
<b>10. Greeklish.....</b>	<b>9</b>
<b>11. Registrations containing the tonos accent .....</b>	<b>9</b>
<b>12. Registrations containing the final sigma.....</b>	<b>9</b>
<b>13. Proposed solutions .....</b>	<b>10</b>
<b>13.1. The variants proposal.....</b>	<b>10</b>
<b>13.2. The small characters proposal .....</b>	<b>11</b>
<b>14. Team Recommendations .....</b>	<b>12</b>
<b>APPENDIX A: Proposed allowed characters for Greek TLD registrations.....</b>	<b>14</b>

# Study of the issues present in the registration of IDN TLDs in GREEK characters

## 1. Introduction - Disclaimer

This report is the work of the Greek Study Team for IDN Variant TLDs. The team was created by ICANN to address the issues presented by the registration of IDN TLDs in Greek.

This document is considered by the team as a step to identify various issues that are present in the Greek IDN registrations in the Greek script but not between scripts. The Greek, Cyrillic and Latin variant teams must interact to identify cross-script issues that might be present between Latin, Greek and Cyrillic characters. These issues might be difficult to be identified by each team alone. Unicode provides a test to check whether whole strings are confusable across strings; the team did not examine that test and cannot say whether it is effective. The team recommends that no IDNs in the Latin, Greek or Cyrillic script be delegated until this issue is discussed thoroughly, unless the string submitted presents no confusability issues. The reader of this report could benefit from reading the article titled "IDN homograph attack" ([http://en.wikipedia.org/wiki/IDN\\_homograph\\_attack](http://en.wikipedia.org/wiki/IDN_homograph_attack)) in understanding some of the issues associated with the similarities and the confusability issues of these character sets.

The reader of this report is considered to be knowledgeable of the ICANN IDN Variant Issues Project (<http://www.icann.org/en/topics/new-gtlds/idn-variant-tlds-delegation-20apr11-en.pdf>) where the concept of variant is roughly defined when it comes to domain labels. Especially in Greek, a variant of a label representing a word (words in small letters are usually accented by the tonos mark) could be considered to be the same word in capital letters, since the words in capital letters are not accented. This leads to different Unicode representation of the words and different Punycode representation, although they consist of the same characters, even if the IDNA2003 character case mapping was to be used.

The term **variant** will be used in the rest of this document to represent only partially what could be considered a variant in the Greek language. Members of the team suggested that the meaning of a variant in Greek should extend to include the variations of a word in Katharevousa and Dimotiki (different forms of the Greek language through history), the monotonic and polytonic differences e.t.c. To keep this report simple for the reader, the report uses the term variant for Greek words or strings, including their accented and not accented forms, retaining the final sigma at the position the applicant requested for it or replacing it by the medial sigma. Furthermore, the variant in Greek takes into account the variations created by changing a string or word from small letters to capital letters, using the Greek grammar as a guide for this transformation.

Usually the domain names are not treated as words; however, many of them are, since a known word is easy to memorize. That is the reason this report mainly deals with issues that involve words. However, this does not mean that the proposed solutions are not applicable to domain names which are not words. Furthermore, since this is a report about TLDs only, rules have been suggested about the actual meaning of the word – in case of a word – which could not be applied on a larger number of registrations and they might seem strange to the reader. A more protective approach has been applied because of the rare nature of TLDs and these rules are not suggested for second or lower level registrations.

This report is considered by this team to be the Greek part of a bigger report in the issues presented by the use of Variants in IDN TLDs registrations. Some of the concepts discussed, especially in definitions, are not extensively presented because of this.

## 2. Definitions

**2.1. Homograph:** A word that shares the same written form with a word of the same or different script but may have different meanings. The same could apply for any string of letters, even if not a word. The characters that make this possible between scripts are mentioned as homograph characters in the document (e.g. the Latin A and the Greek Α).

**2.2. Homophone:** Two different words that are pronounced the same. The same could apply for any string of letters, even if not a word.

**2.3. Greeklish** (Source: Wikipedia, <http://en.wikipedia.org/wiki/Greeklish>, text in Italics): a portmanteau of the words Greek and English, also known as Grenglish, Latinoellinika/Λατινοελληνικά or ASCII Greek, is the Greek language written using the Latin alphabet. Unlike standardized systems of Romanization of Greek, as used internationally for purposes such as rendering Greek proper names or place names, or for bibliographic purposes, the term Greeklish mainly refers to informal, ad-hoc practices of writing Greek text in environments where the use of the Greek alphabet is technically impossible or cumbersome, especially in electronic media. Greeklish was commonly used on the Internet when Greek people communicate by forum, e-mail, IRC, instant messaging and occasionally on SMS, mainly because older operating systems didn't have the ability to write in Greek, or in a Unicode form like UTF-8. Nowadays most Greek-related content appears in native Greek.

**2.4. Aliased name:** A domain name that has been aliased with one or more names under the concept of Name Aliasing. The technical solution that is currently available for aliasing a domain name to another is the use of a CNAME or a DNAME record in a zone file, essentially mapping a domain name or a DNS tree to another. Voices for other technical solutions have been raised over the last few years but without any results.

**2.5.Name Aliasing:** The abstract concept of two or more domain names “behaving as one” by Policy or technical means. This concept has still unresolved issues, definitional, technical and political. Currently this concept is technically served by the use of CNAME and DNAME records in a zone file, allowing for the aliasing of a domain name or a DNS tree to another. With the introduction of the concept of variants, arguments have been raised for the necessity of the presentation of a mechanism that would allow the users to keep in sync domain name trees but without the limitations of DNAME and CNAME. Experts have discussed the use of a CNAME+DNAME kind of record but without concluding on the advantages and disadvantages such a record would present. Issues with the use of DNSSEC, and the inability to address what “behaving as one domain name” stands for in the real world have stalled the discussions on this issue.

**2.6.Bundling Domain Names:** The registry policy of registering certain domain names as a set, depending on select characteristics (e.g. homograph domain names). In some cases, bundling is used for name aliasing purposes.

**2.7.Tonos:** Greek accent mark, acute accent (Greek Tonos, U+0384)

**2.8.Dialytika** (diaeresis): Greek accent mark (appears on the letters “ι̇” (e.g. Greek small letter iota with dialytika, U+03CA) and “υ̇” (e.g. Greek small letter upsilon with dialytika, U+03CB) to show that a pair of vowel letters is pronounced separately, rather than as a diphthong – see <http://en.wikipedia.org/wiki/Diphthong>). It can also be combined with tonos over the same letters, Greek small letter iota with dialytika and tonos, U+0390 and Greek small letter upsilon with dialytika and tonos, U+03B0.

**2.9.Katharevousa:** (Greek: Καθαρεύουσα, [kaθaɾevusa], lit. "puristic [language]"), is a form of the Greek language conceived in the early 19th century as a compromise between Ancient Greek and the Modern Greek of the time, with a vocabulary largely based on ancient forms, but a much-simplified grammar. Originally, it was widely used both for literary and official purposes, though seldom in daily language. In the 20th century, it was increasingly used for official and formal purposes, until Dimotiki became the official language of Greece in 1976 (Source: Wikipedia, <http://en.wikipedia.org/wiki/Katharevousa>).

**2.10.Dimotiki:** (Greek: δημοτική [γλώσσα] [ðimotiɾci], "[language] of the people") is the modern vernacular form of the Greek language. The term has been in use since 1818. Demotiki refers particularly to the form of the language that evolved naturally from ancient Greek, in opposition to the artificially archaic Katharevousa, which was the official standard until 1976. The two complemented each other in a typical example of diglossia until the resolution of the Greek language question in favour of Demotiki (Source: Wikipedia, <http://en.wikipedia.org/wiki/Dimotiki>).

### 3. Useful key points regarding the Greek Language

#### 3.1. The Greek Language Question

(Source: Wikipedia, [http://en.wikipedia.org/wiki/Greek\\_language\\_question](http://en.wikipedia.org/wiki/Greek_language_question), text in Italics)

*The Greek language question was a dispute discussing the question whether the language of the Greek people (Dimotiki) or a cultivated imitation of Ancient Greek (Katharevousa) should be the official language of the Greek nation. It was a highly controversial topic in the 19th and 20th centuries and was finally resolved in 1976, when Dimotiki was made the official language. The language phenomenon in question—which occurs elsewhere in the world—is called “diglossia”. This term was coined at the turn of the 20th century by Ioannis Psycharis, a leading participant in the Greek controversy.*

*While Dimotiki was the mother tongue of the Greeks, Katharevousa was an archaic and formal variant that was pronounced like modern Greek, but adopted both lexical and morphological features of ancient Greek that the spoken language had lost over time, like: morphological features, phonological features, syntactical features and lexical features. These differences meant that Katharevousa was only partly intelligible to a Greek without higher education. There was no single Katharevousa. Instead, proponents of the formal language utilized ever-changing variants that never were standardized. These variants were nearly Attic in extreme cases, but they could also be closer to spoken Greek and could be understood by the majority of the people.*

***Diglossia:** For a person who does not speak Greek and whose mother tongue (e.g. English) exhibits no comparable form of diglossia, it is hard to understand the motivation of the Greek language question, as it concerns the coexistence of two—in extreme cases—completely different forms of Greek that greatly exceeds the usual stylistic difference between written and spoken language.*

*Katharevousa was made the official language of Greece after its independence (1830), since the “unpolished” language of the people was not thought of as able to fit the needs of a modern state.*

*By 1900, the discussion about the form of the Greek language had become a matter of public interest. Proponents of Katharevousa denounced proponents of Dimotiki as “μαλλιαροί” (hairy, furry), “αγελαίοι” (gregarious, social, vulgar) and “χυδαϊσταί” (speakers of slang, plebeians, vulgarians), while the proponents of Dimotiki called their enemies “γλωσσαμύνητες” (defenders of language, purists), “σκοταδιστές” (obscurantists, or more or less: the ones living in spiritual darkness), “αρχαιοπληκτοι” (ancient-maniacs), “μακαρονισταί” (imitators of archaic languages, macaronic people) or “συντηρητικοί” (conservatives). The educational system was in an alarming state and completely ineffective: The children were completely unable to express themselves in the unfamiliar formal language, which severely harmed their speech acquisition instead of educating them. By 1917 the Dimotiki had been successfully introduced into a limited number of primary schools; but again, it was repeatedly replaced with Katharevousa.*



*Only on April 30, 1976 was the era of linguistic purism ended in Greece when the Greek Government banned Katharevousa from use in schools and, only a few months later, passed a law concerning the use of the Dimotiki in official texts and documents, which effectively terminated the diglossia. Ironically, the law in question was formulated in Katharevousa.*

### **3.2.The Greek Orthography**

(Source: Wikipedia, [http://en.wikipedia.org/wiki/Greek\\_diacritics](http://en.wikipedia.org/wiki/Greek_diacritics), text in Italics)

*Greek orthography has used a variety of diacritics starting in the Hellenistic period. The complex polytonic orthography notates Ancient Greek phonology. The simple monotonic orthography, introduced in 1982, corresponds to Modern Greek phonology and requires only two diacritics.*

*Polytonic orthography (πολύς "much", "many", τόνος "accent") is the standard system for Ancient Greek. The acute accent ( ´ ), the grave accent ( ` ), and the circumflex ( ^ ) indicate different kinds of pitch accent. The rough breathing ( ρ ) indicates the presence of an /h/ sound before a letter, while the smooth breathing ( ϱ ) indicates the absence of /h/.*

*Since in Modern Greek the pitch accent was replaced by a dynamic accent, and the /h/ was lost, most polytonic diacritics have no phonetic significance, and merely reveal the underlying Ancient Greek etymology.*

*Monotonic orthography (μονός "single", τόνος "accent") is the standard system for Modern Greek. It retains only the acute accent (tonos) to indicate stress and the diaeresis ( ¨ ) to indicate a diphthong: compare modern Greek παιδάκια /paǰ̂ðaca/ "lamb chops", with a diphthong, and παιδάκια /pêðaca/ "little children" with a simple vowel. Tonos and diaeresis can be combined on a single vowel.*

## **4.Proposed Characters for Registrations**

Following discussions between the members of the team, the opinion mostly shared between the members is to disallow the ancient special characters and only allow characters that are today used in the spelling of Greek words. Furthermore, the team recommends allowing only the Monotonic characters. An appendix of these characters can be found at the end of this document.

The team reached this decision on the basis that TLD IDN registrations should be useful for as many people as possible. The use of ancient characters or polytonic characters is not easy, even for people using Greek keyboards. Furthermore, the users of the Greek language are accustomed in the contemporary Monotonic way of writing and this choice of characters will create a good use experience for the IDN TLDs in Greek. However, this recommendation stands only for the top level labels and it does not prohibit registrations in polytonic or ancient characters for the lower levels where some users could be expected to be accustomed to the use of these characters.

## 5.The sigma and final sigma

Sigma is the 18th character of the Greek alphabet. There are three types of sigma: Lower case sigma (Unicode U+03C3), Upper case sigma (Unicode U+03A3) and small letter Final sigma (Unicode U+03C2). The use of the first two is obvious. The small letter final sigma is used only at the end of words, instead of the lower case sigma.

In IDNA2003, the three letters were mapped to each other and the small letter final sigma was replaced by the IDNA2003 processor by the letter small sigma. This way, the translation of a word from the Punycode was not satisfactory when a final sigma was used, because the Nameprep translation was converting it to a small letter sigma.

In IDNA2008 a different approach was followed. The following paragraphs explain the rationale behind the decision (Source: IETF, RFC5894, “Internationalized Domain Names for Applications (IDNA) Background, Explanation, and Rationale”, August 2010, text in Italics):

*“In IDNA2003, all characters are case folded and mapped by clients in a standardized step.*

*Even in scripts that generally support case distinctions, some characters do not have uppercase forms. For example, the Unicode case-folding operation maps Greek Final Form Sigma (U+03C2) to the medial form (U+03C3) and maps Eszett (German Sharp S, U+00DF) to "ss". Neither of these mappings is reversible because the uppercase of U+03C3 is the uppercase Sigma (U+03A3) and "ss" is an ASCII string. IDNA2008 permits, at the risk of some incompatibility, slightly more flexibility in this area by avoiding case folding and treating these characters as themselves. Approaches to handling one-way mappings are discussed in Section 7.2.*

*Because IDNA2003 maps Final Sigma and Eszett to other characters, and the reverse mapping is never possible, neither Final Sigma nor Eszett can be represented in the ACE form of IDNA2003 IDN nor in the native character (U-label) form derived from it. With IDNA2008, both characters can be used in an IDN and so the A-label used for lookup for any U-label containing those characters is now different. See Section 7.1 for a discussion of what kinds of changes might require the IDNA prefix to change; after extended discussions, the IDNABIS Working Group came to consensus that the change for these characters did not justify a prefix change.”*

## 6.The IDNA approach of translating Greek words into domain names

The example used for the identification of the issues arising from the use of the IDNA protocols will be the word “Ελλάς” which is the Country name.

As we can see, the example word contains the tonos accent mark above the letter small case alpha (Unicode U+03AC) and as a word that finishes on a sigma, this sigma is a small letter final sigma (Unicode U+03C2).

The same word in upper case is spelled “ΕΛΛΑΣ”. No tonos accent mark is used, nor a final sigma, since there is no uppercase letter final sigma. The letters used in these positions are the uppercase letter alpha (Unicode U+0391) and uppercase letter sigma (Unicode U+03A3). These letters according to the IDNA2003 are mapped to the small letters alpha and sigma, respectively U+03B1 and U+03C3. In IDNA2008 uppercase letters are disallowed and the application may map them to allowed characters.

If the domain name “.ελλάς” was to be typed, it would be common for a user to type it as:

.ελλάς. xn--hxarsa0b, small letters form  
.ελλας. xn--mxahsa0b, misspelled without the tonos  
.ΕΛΛΑΣ (.ελλασ). xn--mxahsa5b, capital letters form

One would argue that .ελλας without the tonos is not a correct type. However, due to the use of smartphones and other devices lacking a proper keyboard, typing without the tonos accent is sometimes used because it is not always easy for the user to input the accented letter. This variant should be considered a way to address possible shortcomings of the current technology of handheld and keyboard-less devices.

If we try to equalize the user experience of a user using ASCII domain names and a user using Greek domain names, they might both expect to type the domain name in uppercase and lowercase and end up at the same result. However, this is not the case for the Greek user or any non-ASCII user. This is happening because in IDNA2008 the capital characters are not mapped to small characters. This task is left to the application, without any specifications from the IETF or any other standardization organization. One could expect that .ΕΛΛΑΣ should be transformed to .ελλασ. On another scenario, the variant .ελλας could be considered the transformation from capital letters to small letters for a final sigma aware application.

## 7.Homographs

The homographs are two or more words in one script or different scripts that look-alike to the user.

Examples:

**ΕΕΤΤ** (Greek script) and **EETT** (Latin script). In small letters these words differentiate adequately: **εεττ** (Greek script) and **eett** (Latin script). Similar issues might appear between Greek script and Cyrillic script.

**Πότε** (Greek Script, tonos accent over alpha) and **Ποτέ** (Greek script, tonos accent over epsilon, different meaning). These two words have both the same uppercase typing, **ΠΟΤΕ**.

Would it be possible for two different users to have these homographs registered? One could argue that it is possible, since the punycodes are different. However, this

team's opinion is that these domain names are probably confusing for the average user and restrictions on these registrations should be applied to avoid such confusion. Measures to avoid confusion are used by the Greek registry, involving the introduction of Homograph characters, the requirement of same registrant for homograph domain names and the use of bundling. For further information see <https://grweb.ics.forth.gr/homographs.jsp?lang=en>. This team suggests a similar table should be created, in cooperation with the Latin and Cyrillic teams to address this issue between the three scripts.

## **8.Homophones**

These are domain names that differ in one or more characters but are pronounced the same. As examples we use the words “λοττο” and “λοτο”, “λοττο” and “lotto”, or even “com” and “κομ”. Although these words could be treated as different domain names since they consist of different characters, it is apparent that the registration of such a sound variant could lead to confusion.

However, this is not an issue that is specific for the Greek script since it is a common issue in domain name registrations. The Greek team will not address this issue.

## **9.Polytonic Domain Name Variants**

The Greek words that are spelled using the old, Polytonic way of writing, have special accent marks, usually at the first letter of the word if the letter is a vowel. These special accent characters were used initially to make easier the pronunciation of each specific word since they were used to change the tone and duration of the letter. These accent marks remained in the vocabulary until the 1970s, although they had no effect in the pronunciation of the word. They were eliminated by the adoption of the monotonic writing by the Greek government.

If domain names using the Polytonic accents were to be allowed, it should be expected that many of the users would try to type the monotonic variation of the word instead of the polytonic. This would happen because the users are used to type monotonic words and secondly because it is not easy to input these accents, even when using a Greek keyboard.

After discussion, the conclusion of this team is that the use of Polytonic characters for TLD registration presents no significant advantage for the user and the use of only monotonic characters is preferable. However, this might not be the case for second level registrations and the registries should be encouraged to decide upon the use of polytonic characters in their policy. Certain user groups (e.g. the Greek Orthodox Church, scholars e.t.c.) might have interest in registering and using domain names in polytonic characters in the second and lower levels.

## **10. Greeklish**

The use of Latin characters for the spelling of Greek words in mobile devices was quite common in the 90s when the support of the Greek characters in such devices was limited. Because of the extended use of the SMS (Short Message Service), this way of texting is not alien to the new generations of users, both of internet and mobile services.

Although as a phenomenon Greeklish could deserve some study, the use of them for registrations presents an issue that is beyond the scope of this team, since these potential registrations are not in Greek characters.

## **11. Registrations containing the tonos accent**

The majority of words in small letters in the Greek language (excluding only monosyllabic words) have to have the tonos accent over one of their vowels to be spelled correctly. This way, the person using this word is able to identify the correct pronunciation of the word.

As we have already mentioned in paragraph 6 of this document, a domain name with the tonos sign over a vowel is a different domain name than the domain name without the tonos. Furthermore, the domain without the tonos is consisting of the letters of the Upper case form of the domain name if a final sigma is not present.

For a satisfactory user experience, this team came to the conclusion that the tonos accent is necessary and should be allowed for domain names in Greek, especially since many labels represent words and the user will be expecting them to have the tonos accent. The exclusion of the tonos sign from the allowed characters would only lead to reduced use of the IDN domain names and a less satisfactory user experience.

Furthermore, it is considered by the team as essential that along the domain with tonos, the domain name without tonos should be registered to the same entity/registrant, since the majority of users will expect the IDN domains to behave as ASCII domains do. Under this scope, it would be even more beneficial if these domain names were “behaving the same”, aliased in some way, by technical means or policy, resulting in the same service (e.g. web pages). This is mainly to cover the case of a user typing words in capital letters (mapped by the client to small letters) or to cover the inability of the user to type through his interface letters with tonos.

The same approach applies for domain names with multiple tonos accents, like two-word domains. A single domain name without the tonos accents has to be allocated to the registrant.

## **12. Registrations containing the final sigma**

The use of the final sigma character in Greek is essential for the correct spelling of the Greek words and male names.

However, as we have already reported in paragraph 6, the use of the final sigma along with the tonos accent in the same domain name could complicate the registration procedure. Both solutions that were proposed by this team had to consider the proper use of final sigma.

### **13. Proposed solutions**

Two proposals were formulated by the members of the team as preferable solutions for the issue of registering TLDs in Greek characters:

#### **13.1. The variants proposal**

This proposal is derived from the intention of allowing a richer user experience by using variants to contemplate with the different character cases and typing of a chosen TLD, especially if it is a representation of a known word.

The main points of the variants proposal:

- The accepted for registration characters are the small monotonic characters, including the characters with tonos, dialytika and their combination.
- The domain name will be accepted at the requested form submitted by the registrant, including tonos, dialytika, final sigmas, at the exact positions they were submitted. The applicant has to submit the domain name in small characters only.
- The same domain will be allocated to the registrant, stripped of the accent marks and final sigma, if any of these characteristics were present at the original domain name application. (Variant 1)
- The same domain will be allocated to the registrant, stripped of the accent marks, retaining the final sigmas at their original positions, if any of these characteristics were present at the original domain name application. (Variant 2)
- Alternative positioning of the tonos accent mark is not to be allowed.
- Alternative positioning of the final sigma is not to be allowed.
- Alternative, same meaning words, if the domain consists of words are to be disallowed. This rule is especially put in place to deal with words in “Katharevousa” and “Dimotiki”. Dimotiki is the contemporary Greek language, where Katharevousa was in use before 1976. However, many word types from Katharevousa are still in use in Dimotiki and the user should be protected of confusability issues between these same meaning words. For example “Πειραιάς” and “Πειραιεύς” are two names of the same city in Greece, respectively in Dimotiki and Katharevousa – The applicant will only choose to register one of these and the other one should be excluded of registration. The team recognizes the difficulties this procedure presents for automated systems, however, since TLD registrations are not expected to be of

great volume, this rule could be easily implemented with relatively low cost for ICANN.

- To serve the applicant better in the use of the allocated variants, two options could be available:
  - The variants enter the root zone file as DNAMEs.
  - The variants are treated as FQDNs.

It is apparent that the second option presents the problem of out-of-sync zone files; however this is a technical issue that cannot be addressed by this report.

### **13.2.The small characters proposal**

This proposal is derived from the intention of creating a consistent and simple user experience, excluding variants. The advantage of this solution is that it straightforward and simple, although in some cases it does not take into account the different representations of a Greek spelled word.

- The accepted for registration characters are the small monotonic characters, including the characters with tonos, dialytika and their combination.
- The only accepted form of the domain name is the form that is originally submitted by the applicant, with the specific tonos positions and the possible final sigmas.
- The same form, stripped of the accents will be disallowed for registration.
- The same form, stripped of the accents and with any combination of sigma/final sigma, will be disallowed for registration.
- Alternative positioning of the tonos accent mark is not to be allowed.
- Alternative, same meaning words, if the domain consists of words are to be disallowed. This rule is especially put in place to deal with words in “Katharevousa” and “Dimotiki”. Dimotiki is the contemporary Greek language, where Katharevousa was in use before 1976. However, many word types from Katharevousa are still in use in Dimotiki and the user should be protected of confusability issues between these same meaning words. For example “Πειραιάς” and “Πειραιεύς” are two names of the same city in Greece, respectively in Dimotiki and Katharevousa – The applicant will only choose to register one of these and the other one should be excluded of registration. The team recognizes the difficulties this procedure presents for automated systems, however, since TLD registrations are not expected to be of great volume this rule could be implemented with relatively low cost for ICANN.

## 14. Team Recommendations

Taking into account the various advantages and disadvantages presented in the proposed solutions of paragraph 13, the team's recommendation is the Variants proposal, par. 13.1.

This proposal is trying to take into account the user's possible needs. It is an effort to make the experience of using Greek IDN TLDs equivalent to the use of ASCII TLDs where the user expects no difference in using upper or lower case characters. Furthermore, this solution is not excluding the use of any future technology advances, especially if a definition of Name Aliasing –whenever and if this happens- would require the use of more than one domain names. At present, the use of DNAME is a partial and not entirely satisfactory technology of making names “behave the same” under Name Aliasing.

At the current phase, a transitional phase between IDNA2003 and IDNA2008 and until the mapping is standardized and devices become easier to use when it comes to typing without a keyboard, this team considers as an appropriate way of dealing with the possible technology shortcomings to suggest the registration of all three types of variants presented in paragraph 6 for each registered word or text according to Variants proposal (par. 13.1). In the case of a domain name that is submitted for registration without a tonos or a final sigma, or upon a registrant's request, the associated variants could be omitted from delegation but always reserved from registration to another entity.

This recommendation was not unanimous. The small characters solution, par. 13.2 has puzzled the members of the team and it is not lightheartedly rejected. However, since this proposal is actually contained in the Variants proposal, one could imagine that if the use of Aliased names does not mature enough in the future to allow for a rich user experience or if technology addresses the mapping and character entry issues with the help of other means, this solution could be revisited and used.

The reason the small characters solution, par. 13.2 was not lightheartedly rejected lies in the current limitations of the technical solutions for addressing the variants issue. Today, only the use of DNAME is a valid option for mapping one domain tree to another. Although this mapping is possible in the DNS, several other protocols need special configurations or even multiple configurations to allow this mapping to serve as a tool. Web servers, email servers, ftp servers, electronic certificates, DNSSEC and other services and applications might not work as expected because of this DNS mapping, creating problems. In the light of this fact, the use of one TLD instead of several variants presents some advantages which are clearly visible and understood.

If this easy path was the path to be followed by default, this team would expect there would not be a discussion about a Variant IDN Process in ICANN. One could say that variants are difficult to implement, they cannot be fully supported and operative and they do not add to the user experience but instead could lead to user confusion. However, since variants are being studied, ICANN has to identify the requirements they present for the DNS and other protocols and work to evolve the technical environment



to address these issues. Of course this action could take a significant time and the registrants should be aware of all the technical limitations of the use of DNAME in the root before requesting a variant to be added to their choice of domain name.

In any case, the recommendation of the team has to be considered a collective opinion of the members of this team, without reaching unanimity. Taking into account the ICANN intention of submitting the reports for public comment the team considers this recommendation a starting point for further discussion.

Recognizing that a satisfactory technical solution is not currently available for addressing the variants issue within the present technical environment, the team suggests that a discussion on this issue should be initiated between ICANN, technical experts and the registries and users to clearly identify what could be a solution to this issue, which did not exist in the early days of DNS but should be addressed today.

What should be considered as a red line for the members of this team is the absolute belief that both the accent marks and the final sigma have to be included in the allowed for registrations characters since they are paramount for the proper typing of so many words in the Greek language. Although a “no-accent-no-final-sigma” approach was discussed, it was not welcomed by any of the members of the team and it was rejected, because this would lead to a not real representation of the Greek language and its grammar.

Additionally, a second red line for the team is the belief that any variations of the characters of a specific string submitted for registration should be reserved only for the same registrant. This is because of the significant confusability risk present in small variations of a tonos in an alternative place or a different position of the final sigma. The team acknowledges that this decision introduces a restriction for some words (e.g. words like *πότε* and *ποτέ*) but the risk mitigation is by far more significant than the impact of this restriction to the registrations, especially if only TLDs are studied.

Finally, the team recommends the examination and review of issues that exist between scripts, especially between the Greek, Cyrillic and Latin, which was not in the scope of this report, before the introduction of the new IDN TLD registrations procedure by ICANN.

## APPENDIX A: Proposed allowed characters for Greek TLD registrations

Script Description: Greek Unicode (Basic and Extended)

HEXADECIMAL CODE (UNICODE V5.2) ENGLISH NAME

0390 GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS

03AC GREEK SMALL LETTER ALPHA WITH TONOS

03AD GREEK SMALL LETTER EPSILON WITH TONOS

03AE GREEK SMALL LETTER ETA WITH TONOS

03AF GREEK SMALL LETTER IOTA WITH TONOS

03B0 GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND TONOS

03B1 GREEK SMALL LETTER ALPHA

03B2 GREEK SMALL LETTER BETA

03B3 GREEK SMALL LETTER GAMMA

03B4 GREEK SMALL LETTER DELTA

03B5 GREEK SMALL LETTER EPSILON

03B6 GREEK SMALL LETTER ZETA

03B7 GREEK SMALL LETTER ETA

03B8 GREEK SMALL LETTER THETA

03B9 GREEK SMALL LETTER IOTA

03BA GREEK SMALL LETTER KAPPA

03BB GREEK SMALL LETTER LAMDA

03BC GREEK SMALL LETTER MU

03BD GREEK SMALL LETTER NU

03BE GREEK SMALL LETTER XI

03BF GREEK SMALL LETTER OMICRON

03C0 GREEK SMALL LETTER PI

03C1 GREEK SMALL LETTER RHO

03C2 GREEK SMALL LETTER FINAL SIGMA

03C3GREEK SMALL LETTER SIGMA

03C4GREEK SMALL LETTER TAU

03C5GREEK SMALL LETTER UPSILON

03C6GREEK SMALL LETTER PHI

03C7GREEK SMALL LETTER CHI

03C8GREEK SMALL LETTER PSI

03C9GREEK SMALL LETTER OMEGA

03CAGREEK SMALL LETTER IOTA WITH DIALYTIKA

03CBGREEK SMALL LETTER UPSILON WITH DIALYTIKA

03CCGREEK SMALL LETTER OMICRON WITH TONOS

03CDGREEK SMALL LETTER UPSILON WITH TONOS

03CEGREEK SMALL LETTER OMEGA WITH TONOS