

The IDN Variant Case Study – Greek Team
June 18 Working Sessions
Actions and Notes

Actions/Next Steps:

1. Update at Public Session 20 June: Panagiotis agreed to give the update on the work of the team from the Saturday sessions in Vaggelis' absence;
2. Statements of Interest: Agreed to revise the template along the lines of the GNSO version and present a draft for the team to review;
3. Definitions Document: Leave as is for now;
4. Questions Document: Use as a framework for the work plan;
5. Project Plan/Deliverables: Agreed on initial draft steps;
6. Schedule: Develop a meeting schedule with Vaggelis and other team members;
and
7. Face-to-Face Meetings: Agreed on at least one to help the team decide on final list of variants (timing TBD).

Notes: Working Session 1

1. Brief introduction of the team:

The Greek team is comprised of the following 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

In addition, the following ICANN staff/consultants will also participate the discussions as support or subject matter experts:

Julie Hedlund - Case study support
Andrew Sullivan - IDNA expert consultant
Nicholas Ostler - Linguistic consultant
Kim Davies - Security expert from staff
Francisco Arias - Registry operations expert from staff
Steve Sheng – Policy expert from

2. Statements of Interest (SOIs):

- Tighten up the language that may be too broad (look at GNSO example);
- Send a revised draft to the full team to review;
- Finalize and ask team members to complete the questions (some information may be consistent across work team members);
- Use as guidance on work team member expertise; and
- Publish the SOIs on the wiki.

3. Draft Definitions for Case Study Teams:

Evangelos:

- Most of the definitions come from well-known documents, such as requests for comment (RFCs).
- For the Greek script it is important to identify whom we are serving. There could be an interest in registering names that may come from scripts earlier than the recent changes to the Greek script, which is less than 30 years old.
- We need to identify which will be the character set of the Greek script that we will allow to be used for registration purposes. Then we can see if there are additional issues in the language variance table.
- If we try to serve the Greek script throughout history, then we can find variations that existed throughout the centuries, but we may not have a lot of issues for Modern Greek and we should not open the possibilities for any kind of Greek script developed through the centuries.
- If we stick to the Modern Greek script than we may be able to accept the definitions in the draft document as they are currently written.

Nicholas:

- The Arabic group is worrying about multiple scripts, so it would be a great luxury if this team could get down to a specific script.
- I was wondering whether the Greek Orthodox Church agrees with restricting the script to only something relatively modern and I don't know whether we are responsible for the Coptic script.

Evangelos:

- I was careful to say that we should not restrict to the monotonic (“monotoniko”) only. Certainly even now there is a need for a polytonal (“polytoniko”) script.
- Centuries of development of the Greek language have used polytonic characters, up to the recent change to the “monotoniko” way of writing Greek, which was introduced less than 30 years ago.
- Polytonic characters have been used in older forms of writing Greek. Going back to history, broadly we can say that we can first recognize the “ancient” form of the Greek language—capitalized form only, no accents, with some Archaic characters like “digamma”, “koppa”, “sigma”, “sampi” etc. Then we can distinguish the “Alexandrian or Hellenistic Common” form of the Greek language, were the accents and spirits begin to appear to assist non-native Greek

- speakers to stress the vowels and correctly pronounce Greek (323 BC to 395 AD). Then accents and spirits become indispensable through the “Byzantine” form of writing Greek (395-1453 AD), where Greek was a dominant and international language. Closing to recent times, a new form of writing Greek was gradually developed, the “Neohellenic form”, from which “katharevousa” was evolved early in the 19th century. “Katharevousa” used polytoniko accents and spirits and grammar elements from the older forms of Greek. Since “katharevousa” was mainly used by scholars and not by common people, a newer form of writing Greek, “Dimotiki”, more popular to the common people gradually gains momentum and it is in use even today. “Dimotiki” still used polytonic characters until 1982, where it was decided to use officially “Dimotiki” with only one accent and no spirits (“Dimotiki with the Monotoniko system”). I am giving this very brief and simplistic history of the development of the Greek language simply to point out that we should try to support only the use of “Dimotiki” perhaps with both “monotonic” and “polytonic” characters, and this could accommodate 95% of all expressed needs, including those of the Greek Orthodox Church.
- My answer to your question is that we will try to accommodate the needs of those who have used the polytonal script but we should try to avoid ancient or older forms of the script. We can just use “Dimotiki” with or without “polytonic” characters.

| Panagiotis: The polytonal letters are supported in the second level domain [under \[.gr\]](#).

Evangelos:

- On the definitions: we can provide a table from the Unicode (or equivalently the ISO 10646) Standard for the Greek script, which should include monotonic and polytonal and some other characters, but we should use them in the framework of the modern way of writing Greek..
- We have a lot of support in international standards for monotonic, polytonal, and some other characters but there could be problems if we try to introduce the ancient forms of writing Greek. If we stay with the modern “Dimotiki” way of writing then we may not have any variants, which have occurred only in older forms of writing Greek.

Nicholas:

- There may be some variants in the modern script, such as β. I don’t know if you would want to say that one of these is a preferred variant.

Evangelos:

- This is a font variant that we have discussed in Unicode and for us it is the same character.
- In Medieval Greek there was a difference in pronunciation, but now it is used as a unique character.

Nicholas:

- People have been concerned about different resolutions of the characters, for

example in optical character recognition where you would have to resolve them into one.

- These considerations are being discussed in other case study groups.

Evangelos:

- We have a special font for OCR that allows no substitutions. For all the Greek characters there are font variants, but this doesn't constitute a variant – just a way that one is handwriting certain characters. The β issue has been resolved and we aren't using a variant of this character.
- It would be logical to say that we cannot support everything that existed in Greek history, but to support what is actually needed today. If this is the case I don't see any problems with the definitions here.
- It would be helpful for us if Vaggelis (Segredakis) gives consideration to these particular suggestions.

Notes: Working Session 2

1. Draft Questions for Case Study Teams:

Evangelos:

- The questions in the document are a bit generic so it may help to include examples concerning how we will respond to the document.

Nicholas:

- It seems that there are no examples and how they will be dealt with. In the scope of work we don't talk about possible examples and how we will deal with them. Without examples it is hard to define the scope of work. Determining the circumstances could help develop the case study.

Evangelos: Let us run through the questions and see if we have a common understanding of how to address them.

2. Scope of the Work and Issue Reports

Panagiotis: We need as many examples to consider as possible before closing off the scope of work. It would be good for the whole project that even if something is not in the scope of work if we have any experience for handling things to transfer to ICANN.

Section 3.1 Definitions

Evangelos:

- First of all we are asked to see if it is possible to determine a language set from the script. From the morning working session we suggested that maybe we should limit ourselves. So this is the first question that we have to respond to. Then we should see if there are variant characters in the selected set. We need to consider whatever could fall into the definition of a variant character.

- I think we should look out for cases where we have homophones (words that sound the same). If we have concrete examples for variant characters it will facilitate our work very much. So we should:

1) Identify the set and 2) identify the variants.

- I am optimistic that we can identify the language set. We will not cope with the Coptic case under this script.
- We will welcome any input from the community that will point us to any variant cases that we will consider. We will consider examples from other groups.

Panagiotis: Through our Team I think we can produce some problematic test cases of our own and we can think of some examples and then share with the other teams (Latin and Cyrillic) and we have experience of the second level domain and we can use this to produce some cases.

Nicholas: The Greek alphabet is quite small even considering upper and lower case – considering all of the possible cases it is relatively small. Much of the problem is in the possibility is in the characters being confused with other scripts.

Panagiotis: ~~Are w~~We are considering only lower case letters in internationalized domain names (IDNs)?_

Evangelos: I think that is correct.

3.2 Basic Character Questions:

Evangelos:

- As an alphabet, the Greek script has a limited number of characters, but when we consider accents and punctuations there are more to consider and this is where variants may arise.
- Also, how will we deal with the same word that is meant to be the same if it is in monotonal and polytonal?
- There may also be some problems with words that have exactly the same appearance in the Greek and Latin script.
- We don't see a problem if we take only the recent development of the Greek script, but we may have problems with accents and punctuation in older forms of writing Greek.
- When we provide examples of variant characters we will invite a response from the ICANN community.

3.3 General Questions About Domain Labels

Evangelos: First we have to determine what are the variants and then we can propose solutions to each particular problem.

Panagiotis: Agree.

3.4 Scripts with Identical Characters

Evangelos: There are many possible issues that may come up, particular with Latin and Cyrillic. It also may come up with IDNs of similar optical appearance so we may want to suggest rules for registration.

3.5 Other Questions

Evangelos:

- In the case of the Greek script we don't have different languages using the Greek script but we have dialects that give us homophones.
- This gives us the opportunity to proceed on a script basis.

Panagiotis:

- With respect to the similarity with Cyrillic and Latin and confusability, i.e. looking the same but not encoded the same way: ICANN should take advantage of the progress that has already been made with respect to the similarity of the scripts.

Evangelos:

- In Unicode the Greek Script comprises two tables.
- As we are working on the variant character issue I can see that there may be issues with other scripts that are using Greek script for other purposes – for example mathematical symbols. All of these characters have different Unicode codes; we can identify them by their values. There is a clear distinction by their code position.
- Concerning confusion: It exists but it is not the main problem. It is variants used for the same purpose.

Panagiotis: While we may consider only TLDs and not mixed TLDs, our experiences can be used for cases at other levels and ccTLDs.

Dennis:

- The primary concern is at the top level, but we are asking the questions for the second level as well – they are tied together through the user experience.
- It helps us to explore what the issues are so this is deliberately vague.

Evangelos: This will help us to come up with some consistent rules.

Dennis:

- I think the user's experience should be consistent, that is, what if the rules are inconsistent?
- You could decide to limit yourselves to the top level but you might be missing an opportunity.

Panagiotis: The last question in section 3.5 is: “What should be the difference in consideration of IDN ccTLDs and IDN gTLDs, if any?” What are we seeking with this question?

Dennis: ccTLDs may have a narrower of character sets than gTLDs; there may be a difference and if you think there is then we are asking you to consider it.

Panagiotis: There are three logical registries that may want to give Greek names: .gr, .cy, and .eu and they may have different procedures.

Dennis:

- In a bundled set of variant strings in a script, you could imagine where one of those strings is a trademark in one language, but isn't in another. So what are the objection procedures either at the top level or secondary level?
- This is more likely to happen in a gTLD world more than in a ccTLD world. Reduction in complexity is a key way to move forward.

Evangelos:

- Our intent is to reduce complexity.
- First we will decide the choice of the character set to simplify things.
- It will be more difficult to determine what are the variants. For example, homophones are possible ways to have problems. We may need rules through all the working groups that some things may need to be out of our scope of work.
- We will need the opinion of others once we come up with our examples of variants.
- Once we have some concrete rules then we can identify solutions to these problems. The third item we need to address is the similarity of appearance. Vaggelis and Panagiotis have worked on this issue before.

2. Initial Draft Project Plan and Deliverables (NOTE: the deadline for the Greek issues report is due on 9/30/2011):

1. Identify the character set;
2. Identify the variants with as many examples as possible (decide if sub teams are necessary for this task and, if so, establish them);
3. Seek comment from the community and other work teams on variant examples; and
4. Look at possible solutions, which may or may not include changes in policy.

3. Schedule:

Determine frequency and timing of meetings after consultation with Vaggelis and the other team members who could not attend the Singapore meetings.

4. Face to face meetings:

We will need at least one face-to-face meeting to determine the final list of variants.

5. Actions and Next Steps (see above)