2022-08-11 IDNs EPDP - Meeting #47

The call for the IDNs EPDP team will take place on **Thursday, 11 August 2022 at 13:30 UTC for 90 minutes.**

For other places see: https://tinyurl.com/4hedmcmv

PROPOSED AGENDA

1. Roll Call & SOI (2 minutes)
2. Welcome & Chair Updates (5 mins)
3. Presentation of small group outputs on String Similarity Review (80 mins)
4. AOB (3 mins)

BACKGROUND DOCUMENTS

Package_ Report String Similarity Small Group Outcome.pdf

PARTICIPATION

Attendance

Apologies: Michael Bauland, Nigel Hickson, Lianna Galstyan

RECORDINGS

Audio Recording

Zoom Recording (including audio, visual, rough transcript and chat)

GNSO transcripts are located on the GNSO Calendar

Notes/ Action Items

Action Item

Action Item 1: EPDP Team members to provide input on whether it is acceptable to move the 25 August EPDP Team meeting 24 hours later to 26 August at 13:30 UTC to avoid a scheduling conflict with the August GNSO Council meeting.

Notes

Welcome and Chair Updates

- Last week the EPDP Team discussed that it would continue to discuss strings ineligible for delegation, but the leadership team determined that it would be helpful to develop examples before discussing further, so the group will come back to this topic when examples are ready.
- In terms of the glossary, and regarding the discussion around “primary” TLD, when th EPDP Team talks about primary it is talking about the source label. A source label is needed to calculate variants. There may not be value in being strict in saying that the EPDP Team can’t use the term “primary.” We’ve reached a common understanding that when we talk about primary, we are talking about the source or applied-for label. It might be helpful to move towards “source” or “applied-for label” when we move towards drafting.
- What we mean by set is more challenging. We will need to be deliberate when we use this term in being clear what we mean. In the strict sense, it means the source string and all variants, but we may use the term differently in different contexts, as long as we are clear about the intent in context.
- The EPDP Team will continue to work on glossary elements and terms as the work progresses.
- Admin item: Meeting on 25 August will overlap with the GNSO Council meeting. Can we move the meeting 24 hours later to Friday 26 August at 13:30 UTC? Responses welcome in chat and on the mailing list.

Action Item 1: EPDP Team members to provide input on whether it is acceptable to move the 25 August EPDP Team meeting 24 hours later to 26 August at 13:30 UTC to avoid a scheduling conflict with the August GNSO Council meeting.

Small group outputs on String Similarity Review

- Slide 4 – Background
- Slide 5 – Three Levels of Comparison
- Slide 6 – Problem Statement
- Slide 7 – Small Group Tasks – focus for today is task 1 and task 2
- Slide 8 – Small Group Composition
- Slide 9 – Task 1
- Slide 10 – Example Strings
    - Question: Are all 8 examples standalone strings i.e. no variant relationship, except for row 2?
    - Response: Confirmed.
    - Further clarification: Looking at the columns, the strings may not be variants of one another, but there may be risk of confusion when variants of these strings are introduced.
- Slide 11 – Task 2
- Slide 12 – Selected Examples for Comparison
- Slide 13 – Example 6: Two Applied-For Arabic TLDs
- Slide 14 – Example 6: String Similarity Review
- Slide 15 – Example 6: String Similarity Review consolidated view
    - Note: This slide demonstrates what could happen but not necessarily what will happen. The panel will ultimately make the determination of what is confusingly similar.
    - Question: What is the value of comparing any string to a blocked variant, knowing that those will not be applied for and will not be delegated?
    - Response: The small group was looking from the point of view of an end user who may not know if a variant is blocked. There is still a misconnection risk even if the blocked variant will never be a TLD. There is still a risk of user confusion.
    - Follow up to question: Misconnections happen every day when users type a different extension to the second level domain name or type a second level domain name that was not the right one.
    - Additional question: Wouldn't the user quickly establish that the TLD is not what they expected by the content on the site?
    - Response: The small group was also considering end user expectation in terms of experience of "usage".
    - Question: What is the next step if the applied for string is found to be confusingly similar to a blocked variant?
    - Response: Depending on whether the comparison is between two applied for strings or an existing TLD and an applied for string. If it's two applied for strings, this results in a contention set. If the comparison is between an existing string and an applied for TLD, the applied for TLD can not proceed.
- Additional pages -- Example 7
- Slide 17 – Recommendation: Hybrid Model
- Slide 18 – Rationale for Hybrid Model
- Slide 35 – String Similarity Review Recommendation
- Thanks to the members of the small group as well as staff for their work on these tasks.
- Comment: The inclusion of the blocked variants in the analysis can significantly increase the complexity of the analysis and the number of strings taken into account. The working group should consider the potential impact in its evaluation of the small group's recommendations.
- Additional clarification sought on the misconnection risk.
- Response: Looking at the Arabic example in the deck (slide 13) – Looking at B1 and its variants, assuming someone applies for B1, the blocked variants are considered somehow equivalent to B1. Someone from a given language community who is looking at B1 could consider B3 the same as B1. If you are not comparing block variants in the evaluation, A1 and B1 will proceed to delegation. Someone from the language community that considered B1 and B3 to be the same will look at A1 and think it is B1, because B1 and B3 are the same. The user could be put in harm's way because they could click on a link that directs them to somewhere other than B1, which is not what they expected.
- Question: Who is "somebody"?
- Response: This refers to the end user of one or more language community. Whenever two different code points are variants, they are either visually identical or they are considered the same by one of the language communities using the script.
- Comment: That risk could already be happening and is not unique to IDN labels. A1 and B3 are visually similar, which could cause confusion, but it is not clear that the situation with IDN labels is unique and needs to be mitigated in a different way.
- Misconnections happen, but we don't know if it's due to phishing or mistyping or something else. This group can't address that across all TLDs. But it can do something for the case of IDN variants. The small group took into account both the idea of the consistent user experience and potential risk of harm.
- Additional point to highlight – on slide 18 – in practice, language experts on the panel may be able to look at strings in question as part of the String Similarity Review on a case-by-case basis to simplify the problem space.
- RySG will review the small group recommendations and come back with feedback.
- In reference in example 6, in previous discussions, one member had expressed support for comparing blocked to blocked variants to future proof the evaluation in case a blocked variant became allocatable.
- Question – why would the end-user type B3 and end up at A1? Why wouldn't they type B1? If they type B3, wouldn't they get a message that it's an invalid domain?
- Response: The example focuses on case where someone sends a link to someone else to click. Someone sends example.a1 that is malicious and people click on it thinking they are going somewhere else.
- Summary: EPDP Team will come back to this topic in two weeks. In the meantime, members can send questions to the list or reach out to staff for further explanation of the proposal.
- Comment: At some point, we need to shift into mindset of what is implementable. We need to think about what would happen if many IDNs end up in contention. SubPro recommended that for brands there could be an opportunity to change the applied for string in the case of contention. Could this opportunity be extended to those in contention for IDNs?