The Integration Panel (IP) has reviewed "Principles for Inclusion and Exclusion of Code Points in Latin Script for the Root Zone (Latin LGR)e" and has the following comments:

The IP congratulates the Latin GP on the formulation of its "Principles for Inclusion and Exclusion of Code Points in Latin Script for the Root Zone (Latin LGR)". They appear to cover the important considerations and will likely serve the GP well in arriving at at list of proposed candidate code points. The IP would like to caution that the final decision on whether to include or exclude a code point may not be possible by rote application of these (or any other set of) principles, and that additional factors may have to be considered in individual cases.

The IP is looking forward to the next stage of the Latin GP's work and to reviewing actual examples of draft code points.

Additional notes:

The IP would like to note that all entries in an LGR need to be in Unicode Normalization Form C (see RFC 7940) and further that IDNA requires NFC, even if it doesn't agree with the native typing order, or conventions regarding precomposed, decomposed or mixed composed usage. RFC5890 states: "A "U-label" is an IDNA-valid string of Unicode characters, in Normalization Form C (NFC)". Because entries are normalized dual encoding cannot exist.

In creating the repertoire each combining sequence needs to be individually justified and should be separately enumerated; combining marks should not be individually members of the repertoire.

In applying these principles, attention must be paid to the foundational documents for this work as summarized in the "Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR".

Further, the exclusion principles should mention explicitly that the LGR repertoire is constrained by MSR: « A code point not in the latest version of the MSR is excluded. If there is a clear need to add one, the GP will contact the Integration Panel to assess the possibility of adding one to the MSR ».

The IP has reviewed "Analysis of Variants in the Latin Script for the Root Zone" and has the following comments:

The actual guiding principle (contained in the second paragraph of the document) appears to cover the important considerations and will likely serve the GP well in arriving at at list of proposed candidate variants. The IP would like to caution that the final decision on whether to include or exclude a variant may not be possible by rote application of this (or any other) principle, and that additional factors may have to be considered in individual cases.

The IP is looking forward to the next stage of the Latin GP's work and to reviewing actual examples of draft variants.

Additional notes:

The IP has some concerns about the remainder of the document.

The procedure sets a very narrow limit on the kinds of cases that can be considered variants for the Root Zone; this is the basis of the statement by the IP that is quoted in a footnote. It might perhaps be better if this statement were incorporated into the definition of "scope".

In that section, the opening remark about script mixing seem unconnected to the discussion that follows. A straight listing of which related scripts the GP will consider would be more useful.

The IP would like to point out that the example given the document of Latin è (U+00E8) and Cyrillic è (U+0450) may be moot because the final Cyrillic repertoire does not contain U+0450. In general, it is expected that the analysis of cross-script repertoires remain limited to code points that are in the respective scripts' LGRs or draft LGRs.

The general discussion of "classes of variants" may be "of interest to the reader", but it isn't helpful in understanding which principles the Latin GP will follow in deciding whether something is a variant or not -- most of the items discussed are not applicable in the context of the Root Zone LGR.

In the context of the Root Zone, the Procedure is quite clear in that it considers simple similarity of appearance to be outside the scope of the Root Zone LGR. In admitting exact homoglyphs, the IP has been making the argument that ‘e’ in Latin (U+0065) and ‘e’ in Cyrillic(U+0435) are not just visually indistinguishable, but that their distinct code points effectively represent a disunification by script property. (A disunification not unlike that of 01DD and 0259, which are disunified based on case, or the two sets of Arabic digits disunified largely on directional properties).

In the context of other script LGRs for the Root Zone, the IP has argued strongly against embodying rules intended to deal with spelling issues. Therefore, any orthographic variation (spelling differences) would require a very compelling case being made; the examples given may not rise to that level. For instance, ‘ss’ (U+0073 U+0073) and ‘ß’ (U+00DF) are separately available on the second level, in the .de ccTLD (and presumably others). This would strongly argue against the claim that German usage would require them to be variants - in fact the opposite might be concluded.

Consideration of established practice in existing Latin-based IDNs ought to be an important principle. The procedure makes reference to the "Least Astonishment Principle". This principle argues against solutions that produce unexpected or surprising behavior. Having the Root Zone exhibit fundamentally different design decisions with respect to variants than those found on the second level would have to be justified by strong arguments based on factors special to the Root Zone. Absent such factors, the expectation would be that the various levels are more or less compatible in their treatment of IDN labels for a given script.

Finally, the claimed normalization exceptions appear based on a misunderstanding of the normalization algorithm. In normalizing to precomposed form (Normalization Form C), the first step is to fully decompose the input string and then to reorder all combining marks in a canonical order. Because of that, the two examples of e with grave and dot below would become identical at that stage of normalization. In the final stage of the algorithm, as much of the sequence as possible is composed. But because both inputs have the same fully decomposed and reordered form, their final NFC form is identical.

Or, put differently, only one of the two forms is in NFC, the other is unnormalized and as such not admissible in the LGR.