

Report of Public Comments

Title: Label Generation Ruleset for the Root Zone Version 1 (LGR-1)															
Publication Date:	16 February 2016														
Prepared By:	Integration Panel (IP)														
<table border="1"> <tr> <td colspan="2">Comment Period:</td> <td colspan="2" rowspan="3" style="text-align: center;">Important Information Links</td> </tr> <tr> <td>Comment Open Date:</td> <td>04 December 2015</td> <td style="text-align: center;">Announcement</td> </tr> <tr> <td>Comment Close Date:</td> <td>27 January 2016</td> <td style="text-align: center;">Public Comment Box</td> </tr> <tr> <td colspan="2"></td> <td colspan="2" style="text-align: center;">View Comments Submitted</td> </tr> </table>		Comment Period:		Important Information Links		Comment Open Date:	04 December 2015	Announcement	Comment Close Date:	27 January 2016	Public Comment Box			View Comments Submitted	
Comment Period:		Important Information Links													
Comment Open Date:	04 December 2015					Announcement									
Comment Close Date:	27 January 2016			Public Comment Box											
		View Comments Submitted													
Staff Contact:	Sarmad Hussain	Email:	sarmad.hussain@icann.org												
Section I: General Overview and Next Steps															
<p>The Label Generation Ruleset for the Root Zone defines the mechanism to mechanically determine valid IDN TLDs and their variants. LGR-1 [PDF, 786 KB] is developed following the Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels [PDF, 1.39 MB]. LGR-1 is designed to be the first installment of a Root Zone LGR that meets the requirement for a conservative set of label generation rules for stable and secure operation of the Internet's Root Zone. This first version of the LGR contains code points and variants for the Arabic script, to be followed by future versions that will add additional scripts as they become available.</p> <p>Once each version the Root Zone LGR is finalized, it is intended to be applied mechanically as part of the registration process for TLDs to determine the validity of the applied-for label and the maximal set of allocable variant labels. Further mechanisms need to be developed to determine which of the allocable variants from this maximal set will actually be allocated to the applicants.</p>															
Section II: Contributors															
<p><i>At the time this report was prepared, a total of one community submission had been posted to the Forum. The contributor is listed below. To the extent that quotations are used in the foregoing narrative (Section III), such citations will reference the contributor's initials.</i></p>															
<u>Organizations and Groups:</u>															
Name		Submitted by	Initials												
<u>Individuals:</u>															
Name		Affiliation (if provided)	Initials												
John C Klensin			JCK												
Section III: Summary of Comments															
<p><i>General Disclaimer: This section is intended to broadly and comprehensively summarize the comments submitted to this Forum, but not to address every specific position stated by each contributor. Staff</i></p>															

recommends that readers interested in specific aspects of any of the summarized comments, or the full context of others, refer directly to the specific contributions at the link referenced above (View Comments Submitted).

JCK (1) The user community for a string

The relevant population for the root zone represents the entire global user community.

JCK (2) What is being "integrated"?

The report should not be adopted or integrated LGR published until a variety of script proposal have been received and considered by IP. "Integration" of an LGR is not plausible without at least two scripts.

JCK (3) Language integration and Language-dependent Characters Within a Script

It is also not clear how the decomposition problem may impact its work in the future. It is not clear if Arabic Generation Panel or IP have complete understanding of various languages that use Arabic script and how they use it.

JCK (4) Languages using the Arabic Script

Though Arabic Generation Panel has broad linguistic expertise, it still does not cover the variety of languages and the regions where Arabic script is used. It is not clear whether issues which may arise due to the languages not represented may challenge the solution being proposed.

Section IV: Analysis of Comments

General Disclaimer: This section is intended to provide an analysis and evaluation of the comments received along with explanations regarding the basis for any recommendations provided within the analysis.

Integration Panel Response to comment JCK-(1)

The cited statement of the "Least Astonishment Principle" including the text "should not present recognition difficulties to the zone's intended user population" is in fact a verbatim quote from the [Procedure]. Therefore, it is not clear that the comment contains anything that is actionable.

Integration Panel Response to JCK-(2)

The Integration Panel agrees that there exists a tension between making progress and an ideal of completeness and that any partial design bears some degree of risk, even if small. Because it is not practical to wait for a full set of script proposals, all realistic scenarios for the LGR involve a partial set, to be augmented later. The question is then how to mitigate the risk that stems from starting with a single script version of the LGR. To that end the Integration Panel performed internal tests of the integration process, using the available draft LGRs for other scripts, as well as the two completed scripts available at the time, Arabic and Armenian. These test yielded valuable insights into the types of dependencies that can and cannot be expected between LGRs for different scripts.

In adopting its decision to move forward with an LGR-1 based on the Arabic script, the Integration Panel was guided by a careful consideration of the nature of the script involved, as well as an understanding of how this script differs from other scripts for which LGRs are in actual or future development. Contrary to appearances derived solely from the LGR-1 document, the Integration

Panel has been active in monitoring the development of LGR proposals for a number of different scripts, as well as reviewed the final LGR proposal for Armenian.

Based on this deeper understanding of the issues, it became clear that the issues faced (or not faced) by the Arabic script were largely unique, or, where similar to issues faced by other scripts, were different enough in the details so as to not pose the expectation of future dependencies between it and other scripts. In particular, the repertoire of the Arabic LGR being entirely disjoint from other scripts, the feared advantage to “early adopters” is not relevant.

In case of the Armenian script, the Integration Panel found the opposite: that is, while the script LGR, taken by itself, appears perfectly acceptable, it does pose issues of interaction with a number of other scripts. The Integration Panel therefore decided to hold back the integration of the Armenian script until such later time by which enough information is available to conclude that it is possible to progress the script without undue risk to the DNS.

The Integration Panel normally integrates submitted script LGRs as soon as they are technically ready and can be integrated safely. The Integration Panel thus releases the result of integration as a series of successive updates. In this particular case, the Integration Panel concludes that there is no actionable technical cause to hold back on the release of LGR-1 as constituted. On the contrary, deferring the integration of a script LGR unless definitely required because of known and unaddressed interaction with another script brings risks of its own: while the process critically relies on their expertise to review the integration work, Generation Panels are volunteer organizations that can be expected to disband soon after they consider their work done.

Integration Panel Response to JCK-(3)

The Integration Panel restricted the Maximal Starting Repertoire (MSR) to exclude a number of homoglyphs or homographs from consideration. This was possible because most of the affected code points are not in widespread ordinary use, having been encoded for specialized purposes, and thus not eligible for the root.

For the case of Arabic, the Integration Panel deferred to the Arabic GP to come up with a recommendation suitable for the issue in the context of the root. The solution proposed, which is to exclude the combining marks, well reflects the nature of the Arabic script and the encoding principles for it followed by the Unicode Standard. To wit, where a mark is necessary for the orthography of a language, Unicode will encode the combination, but where it is an optional mark, it is left as combining mark.

Contrary to the concerns expressed in the comment, the solution is, among many possible options, the one most likely to be robust going forward.

Integration Panel Response to JCK-(4)

The comments made here are not properly addressed to the LGR-1 at this stage, because under the Procedure, the Integration Panel must take the proposal as delivered and review it up or down. Had these comments been raised during public comments on the Arabic LGR proposal, the Arabic GP could have taken them into account in detail, and, where possible, responded by consulting more experts, doing more research and otherwise attempted to buttress their case. Absent such questions during the public comments on the Arabic LGR proposal, the Integration Panel had to rely on its own

evaluation.

The Integration Panel reviewed the proposed evidence for inclusion of any code point into the repertoire of the LGR in a careful, and time consuming process, starting with early drafts of the Arabic LGR proposal. Special attention was paid to ensure that each member of the repertoire could be tracked to the repertoire of at least one writing system using the Arabic script, and that such writing system could be documented as meeting the criteria set out for acceptance into the root zone. See for example [MSR-2] for a discussion of this issue.

The Arabic GP in turn did not rely solely on its members, but, as envisaged in the Procedure, consulted noted experts on languages in the region. A useful source of corroborative input was found in the document archive of the Unicode Consortium, which contained the evidence adduced at the time the characters in question were first proposed for encoding.

It is unquestionable that the state of usage of the Arabic script is in flux. The best that any effort can achieve is to produce a snapshot that matches the best information available on current usage. While the rate of change in the case of Arabic is perhaps the most dramatic, many other scripts are similarly affected. Major modern writing systems are undergoing periodic reforms, which may affect the usage of the existing repertoire. It is not a feasible proposition to wait until “the dust has settled” because, on the global scale, it never does.

Nevertheless, the Integration Panel did pay special attention to getting an understanding of particular risks to future compatibility. These include the introduction of new code points that violate any assumptions made in the creation of the LGR but even more in the risk of needing additional variant relations among code points.

The Integration Panel is well positioned to monitor pending additions to the Unicode Standard; it also paid careful attention to possible variant relations between code points included and not included in the LGR. These were discussed in depth with representatives of the Arabic GP. As a result, the Integration Panel feels satisfied, that wrapping up the LGR for Arabic at this point does not present any risks that are avoidable with reasonable means.

As a final note: the issues expected for other scripts that support a large number of languages are not expected to inform the process followed for the Arabic script – their issues are, while generically similar – too different in their details.