Based on the requirements of “everyday, general purpose [use ...] in a stable and widespread manner” put forward in the [Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR](https://community.icann.org/download/attachments/43989034/Guidelines%20for%20LGR.pdf?version=2&modificationDate=1430174479000&api=v2), TF-AIDN had to consider ratings according to the Expanded Graded Intergenerational Disruption Scale([EGIDS](https://www.ethnologue.com/about/language-status)).The group decided to only include code points used by languages that are actively written in Arabic script and to exclude code points for which it was unable to find sufficient evidence for their use.

Despite the Integration Panels (IP) guidelines to exclude any languages with an EGIDS rating higher than four, TF-AIDN nonetheless allowed exceptions for code points used by languages for which strong evidence of wider community use was found by the group.

For example, Saraiki language, which has level five on the scale, is included because it is actively used by its community in television programs, printing of newspapers and books, and teaching in universities, etc. Such information has been gathered based on personal and first-hand experience of members of the group and from working with the language communities in language development.

Based on the discussions, the following principles for code point repertoire were finalized. Though initially a wider set of code points were considered, which included required marks, letter combinations, eventually the discussions narrowed the inclusion principle to the following single letter principle:

1. Any code point, which is a letter in established contemporary use in a language.

In addition, the following rules were used to exclude code points.

1. Any code point DISALLOWED by IDNA 2008 protocol.
2. Any code point representing a security or stability issue, which cannot be resolved at any other stage of the analysis (i.e., stage of determining code points, variants or whole label rules).
3. Any code point not listed in the Arabic GP proposal.
4. Any code point either deprecated or not recommended for use in Unicode Standard, with the exception that it meets one of the inclusion criteria with no alternative code point(s).
5. Any code point specifically encoded for historic use without established contemporary use.
6. Any code point representing technical signs only, as encountered, for example, in religious texts.
7. Any code point which does not meet the inclusion criterion, and which is only used for other purposes, for example as a mark, a formatting character or mark, a numerical digit, a punctuation mark, an honorific mark or symbol, or a mathematical symbol.

Additionally, such code points have been excluded to which one or more of the following were applicable:

1. The generation panel lacked sufficient information on the usage.
2. The generation panel could only ascertain the use for such languages that had an EGIDS rating higher than five, as per the “Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR” (see also above).
3. The generation panel had data on the use of code points, but such a code-point was used in free variation with another code point and therefore (possibly) out of scope[[1]](#footnote-1)as per the “Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR”.
4. The generation panel had data on the use of code points, which were excluded fromMSR-1 and therefore out of scope[[2]](#footnote-2).
5. The generation panel had data on the use of code points, but where Integration Panel explicitly expressed disagreement on the validity and relevance of such data in separate communications[[3]](#footnote-3).

In cases where the GP identified data demonstrating the use of code-points but could not include them because of insufficient clarity of the same data within the rules and scope defined by the “Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR”, the code-points were deferred (and therefore excluded).

Such information for the code points is included in the present proposal for possible inclusion in future revisions of the Arabic Script Root Zone LGR. However, such additions would have to be evaluated carefully on a case-by-case basis to ensure that inclusion of such (additional) code-points may not have adverse stability impact (or worse) on the current inclusions.

1. Such was the case e.g. for U+06AC as used by Malay. TF-AIDN gathered data establishing the use of that code-point in the orthography, from internet sources as well as personal communication in between the language community and a member of TF-AIDN (which happens to be part of that community). However further data suggested that such a code-point was in free variation with U+0762, and the “Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR” stipulate that a code-point must be not be an alternate for another code point in current use. IP considered the given case in a response dated March 14, 2014 (<http://forum.icann.org/lists/comments-msr-03mar14/msg00011.html>, - cf. the following footnote) to a public comment on MSR-1 submitted by TF-AIDN, and expressed it’s disposal “towards keeping a less‐preferred variant out of the LGR in the first place”. [↑](#footnote-ref-1)
2. Such was the case e.g. for the U+06AC as employed by Harari [ISO 639-3: har]. A public comment requesting a change of MSR-1 was submitted by TF-AIDN on April 30, 2014 (<http://forum.icann.org/lists/comments-msr-03mar14/msg00006.html>), with a follow-up on May 21, 2014 (<http://forum.icann.org/lists/comments-msr-03mar14/msg00012.html>) and responses from IP on March 14, 2014 (<http://forum.icann.org/lists/comments-msr-03mar14/msg00011.html>). [↑](#footnote-ref-2)
3. Such was the case e.g. for U+06E8 as used by Chimiini [ISO 639-3: swh]. TF-AIDN presented data on the use of that code point in a public comment to MSR-1 (see the previous footnote), since the same code-point was excluded by MSR-1. In the response “The Integration Panel [expressed it’s] disposition [...] to reject the requested addition”, since “the case [...] is rather weak, [and Chimiini] is a mostly oral dialect of Swahili”. Since that comment, the relevant language community has managed to launch a proposal for encoding in Unicode of further code points employed by Chimiini, which was accepted by the Consortium on October 30, 2014. [↑](#footnote-ref-3)