Proposal for Arabic Script Root Zone LGR

LGR Version: 1

Authors: Task Force on Arabic Script IDN (TF-AIDN)

                    https://community.icann.org/display/MES/TF-AIDN+Work+Space

Date: 18 November 2015

Document Version: 3.4

Contents

General Information

The purpose of this document is to give an overview of the proposed LGR in the XML format and the rationale behind the design decisions taken. It includes a discussion of relevant features of the script, the communities or languages using it, the process and methodology used and information on the contributors. The formal specification of the LGR can be found in the accompanying XML document:

• Proposed-LGR-ArabicScript-20151118.xml

Labels for testing can be found in the accompanying text document:

• Labels-ArabicScript-20151118.txt

1 Script and Languages Covered

Arabic script ancestry includes Aramaic, Syriac and Nabataean scripts with Nabataean being its parent script. Typologically, Arabic script has been classified as an Abjad writing system, as the consonants are obligatorily written using letters and the short vowels are represented optionally using diacritical marks. However, there are language specific variations to this general rule, with ALEF, WAW and YEH letters representing either long vowels or consonants (based on context), and, in some languages, a few vowel diacritics are considered mandatory and omitting them is considered a spelling mistake. Furthermore, some diacritical marks are also used to specify consonantal sounds, e.g. use of *Shadda* diacritic for gemination of consonants.

Figure 1: Arabic Writing Styles

In addition to basic letters and combining marks, the script also features digits, punctuation marks and other symbols (both combining marks and characters).Unicode standard encodes two sets of digits for Arabic script, one set for Arabic language and another set for other languages. Also, the languages differ in the shaping of some digits (e.g. digits 4, 6, 7). However, the latter difference is generally handled by the font.

Arabic script is written in many different writing styles including Naskh, Nastalique, Kufi, Thuluth, Diwani, Riqa, etc., as illustrated in Figure 1.These styles are considered calligraphic traditions, with different glyphs for the same letters. While these glyphs are generally considered equivalent for Arabic language, they may be distinctive in orthographies of other languages (and consequently represent different letters).In the printing industry and computer systems, Arabic,
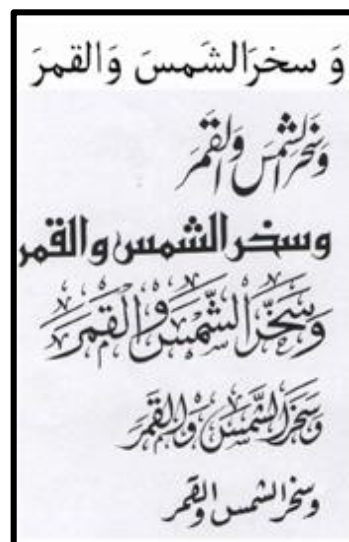
Persian and Sindhi are mainly written in Naskh calligraphic style, whereas many languages of South Asia such as Urdu and occasionally Persian are written in Nastalique calligraphic style.

Arabic script is cursive and therefore its letters are generally written from right-to-left and connected with one another. It has two types of letters: one sub-set joining with only the previous letter[1], and another sub-set joining with adjacent letters to both sides. The cursive joining feature is a basic requirement of the script, with joined letters forming ligatures. One or more ligatures are juxtaposed to form words. Space is neither put between words to separate them in calligraphic handwriting nor regularly used in typing text on computers. Readers use letter shaping (joined or non-joined forms) and other linguistic cues to parse the ligature sequence into words. Unicode describes the general cursive joining behavior of Arabic script characters in Chapter 8 of the standard.

Arabic script has the following specifications:

ISO 15924 code: **Arab**
ISO 15924 no.: 160
English Name: Arabic

The complete set of characters in the Arabic script fall in the following Unicode ranges:

Arabic: U+0600 – U+06FF
Arabic Supplement: U+0750 – U+075F
Arabic Extended A: U+08A0 – U+08FF

For the current analysis, these are limited to the code points shortlisted by the Maximal Starting Repertoire ver. 2(MSR-2), released by ICANN and based on Unicode 6.3.0.

The languages that currently use Arabic script can be divided into four major groups based on their language branch and/or geographical distribution, as follows:

- **Middle East, North Africa and the (African) Sahel area:** Modern Standard Arabic and the local variants of the Arabic language, used mainly by Arab countries[2] and African countries of the Sahel, such as Chad.

- **South and West Asia:** Many Indo-Iranian and Turkic languages of the southern and western Asia use the Arabic script as their main writing system. The extension of the script used by this family of languages is commonly known as the Perso-Arabic script. Examples include Persian, Pashto, Sindhi, Urdu and many more.

- **North and Sub-Saharan African Languages:** Languages of North and sub-Saharan Africa use Arabic script (often much along with Roman/Latin or other scripts).The extended

---

[1]In the right-to-left direction, letter before occurs on the right of the target letter.

[2] Defined here as the member states of the Arab League.

Arabic script used by some of these languages is referred to as Ajami in some countries, and Ajami script is a popular term denoting the use of Arabic script for the writing of African languages in the relevant literature.. Examples include Wolof, Hausa, Tamazight, Comorian and others.

- **South-East Asia:** In South-East Asia, the use of Arabic script for writing the Malay, Acehnese, Banjarese, Minangkabau, Tausug and for several other regional languages is referred to as Jawi. Furthermore, Jawi is one of the two official scripts in Brunei, and is used in Malaysia as an alternate to Latin script. Jawi is employed mainly for religious and cultural purposes of the Malay language. Day-to-day usage of Jawi is maintained in more conservative Malay-populated areas such as Pattani in Thailand, Kelantan in Malaysia, Singapore and used by some of the locals in the Muslim population regions of the southern Philippines. Arabic script is also used by the Cham language in Cambodia.

A non-exhaustive list of languages using Arabic script is provided in Appendix A.

2 Process Undertaken for Developing the Proposal

There has already been a detailed analysis of Arabic script done by the community in an earlier phase of the LGR program, which resulted in an Arabic Issues Report. This report was preceded by work done through a self-formed and community-led group called Arabic Script IDN Working Group (ASIWG). These analyses by the Arabic script community have been used as a basis for the current work, especially the Arabic Issues Report and RFC 5564, in addition to other literature and the expertise available in the current task force.

2.1 Team diversity and process

The current work has been undertaken by the Task Force on Arabic Script IDN (TF-AIDN) acting as the Arabic script Generation Panel (GP). TF-AIDN is a community-based effort to address all relevant technical solutions for promotion and the use of Arabic script IDNs.TF-AIDN gathers experts from a variety of backgrounds (bringing varied linguistic and technical perspectives) and covering a wide variety of languages. Geographically, the GP for Arabic script has members from across the relevant regions, including South East Asia, South Asia, the Middle East, North Africa and Sub-Saharan Africa. There are thirty-three members belonging to eighteen different countries from these regions, as listed in Appendix H.

The task force consists of members from a diverse set of disciplines and with very different perspectives. The members represent national and regional policy makers, technical community in general, technical community directly working with the DNS (e.g. registries), security and law enforcement community, academia (technical and linguistic), members of Community Based Organizations (CBOs), and members with experience with local language computing using Unicode and specifically IDNs.

It is not possible to include one expert per language as Arabic script is used for the writing of a large number of languages. However, care has been taken to gather expertise for the coverage of linguistic use of Arabic script. Panel members directly cover many of the major languages, which use Arabic script, including Arabic, Malay, Pashto, Persian, Punjabi, Sindhi, Seraiki, Torwali, and Urdu. In addition, the group members also have further expertise in the use of Arabic script for languages of Africa and South Asia. While all members are familiar with general use of Latin script, in addition, some members of the groups also have expert knowledge of further scripts, such as Latin, Ethiopic, Tifinagh, or Bamun.

In the cases where the team did not have necessary internal expertise for particular code points and their use in a given language, the task force members reached out to relevant experts in the communities and got their feedback. For example, experts were consulted for some code points used in the Arabic language in Maghreb countries, Malay, Saraiki, some languages of Africa, Kurdish and some languages of North Pakistan. To further ensure necessary coverage and representative solutions, TF-AIDN released its work on character repertoire and variants by an open call for public comments, which was widely circulated by its members. Feedback from the community was received on the use of Arabic language in North Africa (Maghreb), Kashmiri and Kurdish through this process, which was discussed among the members for inclusion in the present work. This three-tier process (wide internal expertise, direct outreach and consultation with experts and openly and widely circulated public comments) allowed the task force to address the issues to compile a representative solution.

The work has been carried out since November of 2013, when the group formed to put forward a "Proposal for Generation Panel for Arabic Script Label Generation Ruleset for the Root Zone". Since, the group has had fortnightly conference calls (weekly towards the end of 2014 and early 2015), as well as three face-to-face meetings, in March 2014, June 2014 and February 2015.In addition, the group has been actively engaged over email, through the public mailing list of the Task Force. The group has frequently undertaken work in sub-groups or working groups, focusing on specific sub-tasks and then reporting back to the task force for final confirmation. Including the sub-group meetings, the group has had more than 45 calls during its work. Beyond the technical work, the group has been actively reporting progress to the community at ICANN meetings, Arab IGFs, MENOGs, and local events, to get community feedback and to attract more participation.

The work has been structured to handle the following steps:

1. Defining the code point principles

    - Public Comment

2. Finalizing the code point repertoire

    - Public Comment

3. Defining the code point variant principles

4. Finalizing the code point variant sets

   - Public Comment

5. Analyzing allocatable vs. blocked variants

6. Finalizing code point variant types

7. Analyzing Whole Label Evaluation rules

8. Finalizing WLE rules

9. Documenting the process

10. Creating XML LGR for Arabic script LGR proposal

    - Public Comment (to be undertaken by ICANN)

11. Finalizing the Documentation and XML LGR proposal for submission to ICANN

2.2 Analysis of code point repertoire

Before the actual analysis of the use of individual code points, TF-AIDN had initial discussions to devise principles to include, exclude or defer from inclusion relevant code points. Subsequently, and based on data from various languages, there have been discussions on every single Arabic script code point contained in Unicode during the audio calls and face-to-face meetings. Despite the exclusion of 48 code points by the Maximal Starting Repertoire and 64 code points by IDN 2008, TF-AIDN nonetheless included these characters in its considerations for inclusion within the code point repertoire and based on implications for the user community. 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, TF-AIDN had to consider ratings according to the Expanded Graded Intergenerational Disruption Scale(EGIDS).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[3] as per the

---

[3]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

"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[4].

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[5].

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.

For most of the languages considered, the use of Arabic script is evident and widely established. However, for some languages of Africa and Central Asia, Arabic orthographies, which had been in use for centuries, have since been officially replaced with other orthographies (mostly Roman/Latin or Cyrillic script-based ones). In very few cases, language policy was enforced to such a degree that any previous writing practices have effectively been replaced., However, in numerous other cases, there is still active use of Arabic script, for example, for many languages spoken in Africa called Ajami. For details see Appendix B: Use of Ajami Writing System.

2.3   Analysis of code point variants

TF-AIDN considered fonts, which follow general Arabic calligraphy principles, and font sizes, which are normally used. One can always make fonts stylistically ambiguous (by not following calligraphy conventions) and font size very small to make text difficult to read. In latter cases, any arbitrary letters can become confusing in any script. The following principles for defining variants were agreed upon by the members of GP:

1. Two code points are variants if they are visually identical in any one of the four positions/forms (isolated, initial, medial or final form).

(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".

[4]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).

[5]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.

2. Two code points are variants if they are visually similar and if at least a sub-set of the Arabic script-using community employs those two code-points as stylistic variations of one another within the calligraphic tradition.

3. Two code points are variants if they are visually similar to the degree that at least a sub-set of the Arabic script using community may consider the two as stylistic variations of each other, even where such consideration is not part of the established calligraphic tradition.

4. Two code points are variants if they only differ in their dot orientation, i.e. where those dots, which are used as part of the letter[6], are rotated but their count and placement (above or below) does not change. Dot orientation is used in calligraphic tradition to fit dots in tighter spaces.

5. Code points representing the same letter in Western (African) Arabic orthography[7] and Conventional Arabic orthography are considered variants, as Arabic script readers from the two orthographic traditions may confuse such letters. The two sets are given below.

Conventional Arabic Orthography	Western (African) Orthography
ف	ٮ
ق	ف
ن	ں
ك	ک
ى	ے

6. Variant sets are symmetric (if A is a variant of B, then vice versa is also true) and transitive (if A is a variant of B and B is a variant of C, then A is also a variant of C).

In addition to that, the task force agreed upon the following principles for variant type:

a) By default, all code point variant pairs have a blocked type.
b) The type of any pair of code points, which are interchangeably used within a community, is allocatable.

---

[6]This intends *nuqta* or *ʾiʿǧām* placed on *rasm* to form (new) letters, not *taškīl* or *ḥarakāt*, such as U+065C, which are used rather as modifiers.

[7]There is established variation in between two to three sets of letters employed in in different Arabic script writing traditions, whereby one set is considered the standard in the majority of Arab and Asian countries, while the other set is employed for the most part throughout the Maghreb, West Africa and the Sudanic belt. The latter is referred to here as Western (African) Arabic orthography.
While the differences between the sets themselves are stable, there is a terminological gap and confusion. Maghrebian (Morocco, Algeria, Tunisia, Libya and Mauritania) users seem to consider those letter Maghrebian variants, while certain Arab scholars consider them Andalusian Arabic calligraphy (from Al-Andalus, a medieval Islamic state). The Maghrebian calligraphy is a variation of the Andalusian one. Technical and (Western) Scholarly documents today make mostly reference to Warsh (orthography), e.g. http://www.unicode.org/L2/L2014/14207-warsh-model.pdf or http://scriptsource.org/entry/yq85dn9q24, or Maghrebi script (e.g. van den Boogert, Nico (1989). "Some notes on Maghrebi Script". In Manuscripts of the Middle East: a journal devoted to the study of handwritten materials of the Middle East 4, pp. 30–43). Both terms might however be confusing to some users of Arabic language. It is worth noticing that the term "Warsh" is essentially and exclusively used for recitation and the reading method of the Holy Quran and it does not refer to the writing style.

c) The type of any set of code points which are visually identical or similar, and which are actively used by two different communities (and inherently considered the same character) is allocatable.

3   Code Point Repertoire

The current section lists the code points included or excluded for LGR for the Root Zone based on the MSR-2. Specific usage is documented for the code points which have been included. If usage is identified for code points which have been excluded, it has also been documented for future reference.

It should be noted that the current work is limited to the analysis of the repertoire short-listed in the MSR-2 as per the Procedure. TF-AIDN realizes that the current MSR is limited to version 6.3 and will keep following later versions of MSR updated to include later versions of the Unicode standard. In these cases, if the revised MSR will include additional Arabic script code points, TF-AIDN will analyze them and make separate additional proposal(s) for adding the relevant subset to the current repertoire being proposed, also considering any security and stability implications.

TF-AIDN is also aware of the recent IAB statement. As TF-AIDN has not included any combining marks in its proposal for the Root Zone LGR and only included the code points which represent the composed forms, the task force does not consider that the issue raised by IAB remains relevant in context of this proposal.

3.1   Summary of code point repertoire included and excluded

Summary of analysis is presented below, as per the following representation:

- Code points excluded by IDNA2008 protocol: white
- Code points excluded by MSR-2: pink
- Code points excluded by TF-AIDN: blue
- Code points included by TF-AIDN: yellow

Unicode Arabic script code chart.

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0	؀ 0600	ؐ 0610	ؠ 0620	ذ 0630	ـ 0640	ؐ 0650	ؠ 0660	؀ 0670	پ 0680	ۂ 0690	غ 06A0	گ 06B0	ۀ 06C0	ي 06D0	۠ 06E0	۰ 06F0
1	؁ 0601	ؑ 0611	ء 0621	ر 0631	ف 0641	آ 0651	١ 0661	خ 0671	ځ 0681	ڑ 0691	ڡ 06A1	گ 06B1	ہ 06C1	ڑ 06D1	ۡ 06E1	۱ 06F1
2	؂ 0602	ؒ 0612	آ 0622	ز 0632	ق 0642	أ 0652	٢ 0662	؂ 0672	ڂ 0682	گ 0692	ڢ 06A2	ڲ 06B2	ۂ 06C2	ئ 06D2	ۢ 06E2	۲ 06F2
3	؃ 0603	ؓ 0613	أ 0623	س 0633	ك 0643	ٓ 0653	٣ 0663	إ 0673	ج 0683	ړ 0693	ڣ 06A3	گ 06B3	ۃ 06C3	؍ 06D3	ۣ 06E3	٣ 06F3
4	؄ 0604	ؔ 0614	ؤ 0624	ش 0634	ل 0644	ٔ 0654	٤ 0664	؄ 0674	چ 0684	ف 0694	ڤ 06A4	ف 06B4	و 06C4	؎ 06D4	ۤ 06E4	٤ 06F4
5	؅ 0605	ؕ 0615	إ 0625	ص 0635	م 0645	ٕ 0655	٥ 0665	؅ 0675	ڇ 0685	پ 0695	ڥ 06A5	ڵ 06B5	و 06C5	ۥ 06D5	ر 06E5	٥ 06F5
6	؆ 0606	ؖ 0616	ئ 0626	ض 0636	ن 0646	ٖ 0656	٦ 0666	ؤ 0676	چ 0686	ق 0696	ڦ 06A6	ل 06B6	ۆ 06C6	ۦ 06D6	ۦ 06E6	٦ 06F6
7	؇ 0607	ؗ 0617	ا 0627	ط 0637	ه 0647	ٗ 0657	٧ 0667	ٷ 0677	چ 0687	ف 0697	ڧ 06A7	ل 06B7	ۇ 06C7	ۗ 06D7	ۧ 06E7	٧ 06F7
8	؈ 0608	ؘ 0618	ب 0628	ظ 0638	و 0648	٘ 0658	٨ 0668	ئ 0678	ڈ 0688	ژ 0698	ڨ 06A8	ۀ 06B8	ۈ 06C8	ۘ 06D8	ۨ 06E8	٨ 06F8
9	؉ 0609	ؙ 0619	ة 0629	ع 0639	ى 0649	ٙ 0659	٩ 0669	ٹ 0679	د 0689	ڙ 0699	ک 06A9	ۀ 06B9	ۉ 06C9	ۙ 06D9	۩ 06E9	٩ 06F9
A	؊ 060A	ؚ 061A	ت 062A	غ 063A	ي 064A	٪ 065A	ٮ 066A	ث 067A	ڊ 068A	ښ 069A	ڪ 06AA	ۀ 06BA	ۊ 06CA	ۚ 06DA	۪ 06EA	ۺ 06FA
B	؋ 060B	؛ 061B	ث 062B	ۻ 063B	٫ 064B	٫ 065B	ٯ 066B	ڻ 067B	ڋ 068B	ۋ 069B	گ 06AB	ٹ 06BB	ۋ 06CB	ۛ 06DB	۫ 06EB	ۻ 06FB
C	، 060C	ALM 061C	ج 062C	کپ 063C	٬ 064C	٬ 065C	، 066C	ة 067C	ڌ 068C	ۍ 069C	ڬ 06AC	ۋ 06BC	ی 06CC	ۜ 06DC	۬ 06EC	ۼ 06FC
D	؍ 060D		ح 062D	ؽ 063D	�ٍ 064D	ٍ 065D	★ 066D	ٽ 067D	ۮ 068D	ص 069D	ڭ 06AD	ٹ 06BD	ٿ 06CD	؞ 06DD	ۭ 06ED	ﭝ 06FD
E	؎ 060E	؞ 061E	خ 062E	ۍ 063E	ۄ 064E	ۀ 065E	پ 066E	ڽ 067E	ڎ 068E	ۻ 069E	ۀ 06AE	ﮠ 06BE	ۍ 06CE	۞ 06DE	ۮ 06EE	ﮢ 06FE
F	؏ 060F	؟ 061F	د 062F	ؿ 063F	ۈ 064F	ۀ 065F	ۇ 066F	ٺ 067F	ۋ 068F	گ 069F	گ 06AF	چ 06BF	ۏ 06CF	۟ 06DF	ژ 06EF	ھ 06FF

	075	076	077
0	ݐ 0750	ݠ 0760	ݰ 0770
1	ݑ 0751	ݡ 0761	ݱ 0771
2	ݒ 0752	ݢ 0762	ݲ 0772
3	ݓ 0753	ݣ 0763	ݳ 0773
4	ݔ 0754	ݤ 0764	ݴ 0774
5	ݕ 0755	ݥ 0765	ݵ 0775
6	ݖ 0756	ݦ 0766	ݶ 0776
7	ݗ 0757	ݧ 0767	ݷ 0777
8	ݘ 0758	ݨ 0768	ݸ 0778
9	ݙ 0759	ݩ 0769	ݹ 0779
A	ݚ 075A	ݪ 076A	ݺ 077A
B	ݛ 075B	ݫ 076B	ݻ 077B
C	ݜ 075C	ݬ 076C	ݼ 077C
D	ݝ 075D	ݭ 076D	ݽ 077D
E	ݞ 075E	ݮ 076E	ݾ 077E
F	ݟ 075F	ݯ 076F	ݿ 077F

	08A	08B	08C	08D	08E	08F
0	ࢠ 08A0	ࢰ 08B0				ࣰ 08F0
1	ࢡ 08A1	ࢱ 08B1				ࣱ 08F1
2	ࢢ 08A2	ࢲ 08B2				ࣲ 08F2
3	ࢣ 08A3					ࣳ 08F3
4	ࢤ 08A4				ࣄ 08E4	ࣴ 08F4
5	ࢥ 08A5				ࣅ 08E5	ࣵ 08F5
6	ࢦ 08A6				ࣆ 08E6	ࣶ 08F6
7	ࢧ 08A7				ࣇ 08E7	ࣷ 08F7
8	ࢨ 08A8				ࣈ 08E8	ࣸ 08F8
9	ࢩ 08A9				ࣉ 08E9	ࣹ 08F9
A	ࢪ 08AA				࣊ 08EA	ࣺ 08FA
B	ࢫ 08AB				࣋ 08EB	ࣻ 08FB
C	ࢬ 08AC				࣌ 08EC	ࣼ 08FC
D	ࢭ 08AD				࣍ 08ED	ࣽ 08FD
E	ࢮ 08AE				࣎ 08EE	ࣾ 08FE
F	ࢯ 08AF				࣏ 08EF	ࣿ 08FF

3.2 Code point repertoire included

The following code points are included as they are being actively used by different language communities. Applicable references are given to the languages, applicable EGIDS scale and additional documentation. The list of languages given is representative but not complete, as comprehensive list may require significantly more analysis, which is not needed because even a single example is a sufficient criterion for inclusion.

Details of the code points excluded from the repertoire are given in Appendix F: Code Point Repertoire Excluded.

S. No.	Unicode	Glyph	Code Point Name	Example Languages using the Code Point (note: not an exhaustive list of languages using the code point)	EGIDS Reference	Reference/Comments	Reference Number
1.	0620	ي	ARABIC LETTER KASHMIRI YEH	Kashmiri	4 Kashmiri	http://www.unicode.org/L2/L2009/09215-kashmiri.pdf	11, 115
2.	0621	ء	ARABIC LETTER HAMZA	Arabic, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
3.	0622	آ	ARABIC LETTER ALEF WITH MADDA ABOVE	Arabic, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
4.	0623	أ	ARABIC LETTER ALEF WITH HAMZA ABOVE	Arabic, Malay, Persian, Torwali	1 Arabic	RFC 5564	0, 100
5.	0624	ؤ	ARABIC LETTER WAW WITH HAMZA ABOVE	Arabic, Malay, Persian, Urdu	1 Arabic	RFC 5564	0, 100

6.	0625	إِ	ARABIC LETTER ALEF WITH HAMZA BELOW	Arabic, Malay	1 Arabic	RFC 5564	0, 100
7.	0626	ئ	ARABIC LETTER YEH WITH HAMZA ABOVE	Arabic, Malay, Pashto, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
8.	0627	ا	ARABIC LETTER ALEF	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
9.	0628	ب	ARABIC LETTER BEH	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
10.	0629	ة	ARABIC LETTER TEH MARBUTA	Arabic, Malay, Persian, Urdu	1 Arabic	RFC 5564	0, 100
11.	062A	ت	ARABIC LETTER TEH	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
12.	062B	ث	ARABIC LETTER THEH	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
13.	062C	ج	ARABIC LETTER JEEM	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
14.	062D	ح	ARABIC LETTER HAH	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
15.	062E	خ	ARABIC LETTER KHAH	Arabic, Baluchi, Kashmiri, Malay,	1 Arabic	RFC 5564	0, 100

				Persian, Punjabi, Sindhi, Urdu			
16.	062F	د	ARABIC LETTER DAL	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
17.	0630	ذ	ARABIC LETTER THAL	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
18.	0631	ر	ARABIC LETTER REH	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
19.	0632	ز	ARABIC LETTER ZAIN	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
20.	0633	س	ARABIC LETTER SEEN	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
21.	0634	ش	ARABIC LETTER SHEEN	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
22.	0635	ص	ARABIC LETTER SAD	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
23.	0636	ض	ARABIC LETTER DAD	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100

24.	0637	ط	ARABIC LETTER TAH	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
25.	0638	ظ	ARABIC LETTER ZAH	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
26.	0639	ع	ARABIC LETTER AIN	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
27.	063A	غ	ARABIC LETTER GHAIN	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
28.	0641	ف	ARABIC LETTER FEH	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
29.	0642	ق	ARABIC LETTER QAF	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
30.	0643	ك	ARABIC LETTER KAF	Arabic	1 Arabic	RFC 5564	0, 100
31.	0644	ل	ARABIC LETTER LAM	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
32.	0645	م	ARABIC LETTER MEEM	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100

33.	0646	ن	ARABIC LETTER NOON	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
34.	0647	ه	ARABIC LETTER HEH	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
35.	0648	و	ARABIC LETTER WAW	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
36.	0649	ى	ARABIC LETTER ALEF MAKSURA	Arabic, Baluchi, Kashmiri, Malay, Persian, Punjabi, Sindhi, Urdu	1 Arabic	RFC 5564	0, 100
37.	064A	ي	ARABIC LETTER YEH	Arabic, Urdu, Malay, Punjabi, Kashmiri, Baluchi, Sindhi, Pashto	1 Arabic	RFC 5564	0, 100
38.	0672	أ	ARABIC LETTER ALEF WITH WAVY HAMZA ABOVE	Baluchi, Kashmiri	4 Kashmiri	http://omniglot.com/writing/kashmiri.htm	0, 102
39.	0679	ٹ	ARABIC LETTER TTEH	Urdu, Punjabi, Kashmiri, Baluchi	1 Urdu	http://omniglot.com/writing/urdu.htm	0, 112
40.	067A	ٺ	ARABIC LETTER TTEHEH	Sindhi	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111
41.	067B	ٻ	ARABIC LETTER BEEH	Sindhi, Saraiki	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111
42.	067C	ټ	ARABIC LETTER TEH WITH RING	Pashto	1 Pashto	http://omniglot.com/writing/pashto.htm	0, 108
43.	067D	ٽ	ARABIC LETTER TEH WITH THREE DOTS ABOVE DOWNWARDS	Sindhi	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111

44.	067E	پ	ARABIC LETTER PEH	Baluchi, Persian, Punjabi, Kashmiri, Sindhi, Urdu	1 Persian	http://omniglot.com/writing/persian.htm	0, 109
45.	067F	ٹ	ARABIC LETTER TEHEH	Sindhi	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111
46.	0680	ڀ	ARABIC LETTER BEHEH	Sindhi	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111
47.	0681	ځ	ARABIC LETTER HAH WITH HAMZA ABOVE	Pashto, Khowar	1 Pashto	Pashto Academy Peshawar University; http://omniglot.com/writing/pashto.htm	0, 108, 138
48.	0683	ڃ	ARABIC LETTER NYEH	Sindhi, ANT, Wolof	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111
49.	0684	ڄ	ARABIC LETTER DYEH	ANT, Sindhi,Gawri, Mandinka, wolof	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111
50.	0685	څ	ARABIC LETTER HAH WITH THREE DOTS ABOVE	Burushaski, Gawri, Khowar, Kohistani, Pashto, Torwali, Wolof	1 Pashto	Burushaski Research Academy, Karachi; http://omniglot.com/writing/pashto.htm	0, 108, 138
51.	0686	چ	ARABIC LETTER TCHEH	Urdu, Persian, Punjabi, Kashmiri, Baluchi, Sindhi	1 Persian	http://omniglot.com/writing/persian.htm	0, 109
52.	0687	ڇ	ARABIC LETTER TCHEHEH	Torwali, Shina, Sindhi	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111
53.	0688	ڈ	ARABIC LETTER DDAL	Urdu, Pashto, Punjabi, Kashmiri, Baluchi	1 Urdu	http://omniglot.com/writing/urdu.htm	0, 112
54.	0689	ډ	ARABIC LETTER DAL WITH RING	Pashto	1 Pashto	Pashto Academy Peshawar University; http://omniglot.com/writing/pashto.htm	0, 108, 138
55.	068A	ڊ	ARABIC LETTER DAL WITH DOT BELOW	Sindhi, ANT	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111

#	Code	Glyph	Name	Languages	Usage	Source	Refs
56.	068B	ڋ	ARABIC LETTER DAL WITH DOT BELOW AND SMALL TAH	Saraiki	5 Saraiki	http://omniglot.com/writing/saraiki.htm; Saraiki actively used in print; large population; Newspapers, television and radio. E.g. https://id-id.facebook.com/jhoke.saraiki	0, 110, 139
57.	068C	ڌ	ARABIC LETTER DAHAL	Sindhi, Mandinka	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111
58.	068D	ڍ	ARABIC LETTER DDAHAL	Sindhi, Arabic Tamil	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111
59.	068E	ڎ	ARABIC LETTER DUL	Malay	3 Malay	Information technology - Jawi coded character set for information interchange MS 2443:2012, Department of Standards, Malaysia. http://www.standardsmalaysia.gov.my	0, 137
60.	068F	ڏ	ARABIC LETTER DAL WITH THREE DOTS ABOVE DOWNWARDS	Sindhi	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111
61.	0691	ڑ	ARABIC LETTER RREH	Baluchi, Kashmiri, Urdu, Punjabi	1 Urdu	http://omniglot.com/writing/urdu.htm	0, 112
62.	0693	ړ	ARABIC LETTER REH WITH RING	Pashto	1 Pashto	Pashto Academy Peshawar University; http://omniglot.com/writing/pashto.htm	0, 108, 138
63.	0695	ڕ	ARABIC LETTER REH WITH SMALL V BELOW	Kurdish	3 Kurdish	http://omniglot.com/writing/kurdish.htm See use in Kurdish at http://www.kurdpress.com/	0, 106, 140
64.	0696	ږ	ARABIC LETTER REH WITH DOT BELOW AND DOT ABOVE	Pashto	1 Pashto	Pashto Academy Peshawar University; http://omniglot.com/writing/pashto.htm	0, 108, 138

65.	0697	ڗ	ARABIC LETTER REH WITH TWO DOTS ABOVE	ANT (Alphabet National du Tchad)	1 ANT	ANT (Alphabet National du Tchad) is the national standard for Chad/Tchad; See Figures in L2/10-288R (used for "tr" sound as given in the table in Section 6.1.2): http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf and Appendix E ANTis the official standard for Wolof, Pulaar, Soninke, Mandinka, Seereer, Joola, Balan and others. See http://www.ethnologue.com/country/TD/languages for a more complete list of languages spoken in Chad.	0, 119, 146
66.	0698	ژ	ARABIC LETTER JEH	Baluchi, Kashmiri, Pashto, Punjabi, Saraiki, Urdu	1 Urdu	http://omniglot.com/writing/urdu.htm	0, 112
67.	0699	ڙ	ARABIC LETTER REH WITH FOUR DOTS ABOVE	Sindhi, Torwali	2 Sindhi	Torwali online and printed dictionaries; http://omniglot.com/writing/sindhi.htm	0, 111, 143
68.	069A	ښ	ARABIC LETTER SEEN WITH DOT BELOW AND DOT ABOVE	Pashto	1 Pashto	Pashto Academy Peshawar University; http://omniglot.com/writing/pashto.htm	0, 108, 138
69.	069F	ظ	ARABIC LETTER TAH WITH THREE DOTS ABOVE	Jawi, Hausa, Various languages of Ethiopia (Afaan Oromo, Silt'e,		Ethiopia Wetter, Andreas (2006): Arabic in Ethiopia. In Kees (C.) H. Versteegh (Ed.): Encyclopedia of Arabic Language and Linguistics.	0, 121, 123, 130

				Harari, Amharic), and others		Volume I. A-Ed, vol. 2. With assistance of Mushira Eid, Alaa Elgibali, Manfred Woidich, Andrzej Zaborksi. Leiden: E. J. Brill, pp. 51–56. Hausa Warren-Rothlin, Andy (2014): West African scripts and Arabic-script orthographies in socio-political context. In Meikal Mumin, Kees (C.) H. Versteegh (Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in Semitic Languages and Linguistics, 71), pp. 261–289. For use of Ajami script for Hausa, see "Language planning in West Africa – who writes the script?" by FREDERIKE LÜPKE; available at http://www.elpublishing.org/docs/1/02/ldd02_08.pdf	
70.	06A0	ﻉ	ARABIC LETTER AIN WITH THREE DOTS ABOVE	Malay, ANT	3 Malay	http://omniglot.com/writing/malay.htm https://en.wikipedia.org/wiki/Jawi_alphabet[8]	0, 107, 129, 144

[8] https://en.wikipedia.org/wiki/Jawi_alphabet

Jawi (Jawi: جاوي Jāwī; Pattani: Yawi; Acehnese: Jawoë) is an Arabic alphabet for writing the Malay language, Acehnese, Banjarese, Minangkabau, Tausūg and several other languages in Southeast Asia.

						Also, see Jawi Keyboard developed by Technical Committee in Multi-Lingual Computing of Malaysia's Standards body at https://en.wikipedia.org/wiki/Jawi_keyboard	
71.	06A2	ڢ	ARABIC LETTER FEH WITH DOT MOVED BELOW	Hausa,Fulfulde	2 Hausa	http://omniglot.com/writing/hausa.htm ; http://www.silcam.org/documents/AlphabetandOrthographyStatementforFulfuldeFUBAjamiyafortheinternet.pdf Newspapers and books published in Hausa using Arabic script; See http://aflang.humnet.ucla.edu/Hausa/Pronunciation/writing.html : "Ajami continues to be used by many Hausas, especially in the composition of poetry, though there is at least one newspaper in Arabic orthography and a number of books have been published in Arabic orthography. Further evidence of Ajami script use for African languages including Fula and Hausa in "Language planning in West Africa – who writes the script?"	0, 101, 130, 131, 132

Jawi is one of the two official scripts in Brunei, and is used as an alternative script in Malaysia … Today, the script is used for religious and Malay cultural administration inTerengganu, Kelantan, Kedah, Perlis and Johor. … The Malays in Pattani still use Jawi today.

						by FREDERIKE LÜPKE; available at http://www.elpublishing.org/docs/1/02/ldd02_08.pdf	
72.	06A4	ڤ	ARABIC LETTER VEH	Malay, Kurdish	3 Malay, Kurdish	http://omniglot.com/writing/kurdish.htm ; http://omniglot.com/writing/malay.htm http://www.fas.harvard.edu/~iranian/Sorani/sorani_1_grammar.pdf See use in Kurdish at http://www.kurdpress.com/	0, 106, 107, 127, 140
73.	06A6	ڦ	ARABIC LETTER PEHEH	Sindhi	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111
74.	06A7	ڧ	ARABIC LETTER QAF WITH DOT ABOVE	Hausa, Tuareg, Mongofin, Bamana	2 Hausa	http://omniglot.com/writing/hausa.htm ; http://www.silcam.org/documents/AlphabetandOrthographyStatementforFulfuldeFUBAjamiyafortheinternet.pdf Newspapers and books published in Hausa using Arabic script; See http://aflang.humnet.ucla.edu/Hausa/Pronunciation/writing.html : "Ajami continues to be used by many Hausas, especially in the composition of poetry, though there is at least one newspaper in Arabic orthography and a number of books have been published in Arabic orthography. For use of Ajami script for Hausa, see "Language planning in West	0, 101, 130, 131, 132

					Africa – who writes the script?" by FREDERIKE LÜPKE; available at http://www.elpublishing.org/docs/1/02/ldd02_08.pdf	
75.	06A8	ڨ	ARABIC LETTER QAF WITH THREE DOTS ABOVE	Almaghreb Alarabi (Western Arabic)	1 Arabic	This character is used in the local dialects (i.e., الدارجة ad-dārija or العامية al-`āmmiyya) of the western parts of the Arab world (e.g., Morocco, Algeria, Tunisia, …). For example, city of Gabes in Tunisia (قابس); e.g. see its logo http://ar.wikipedia.org/wiki/%D9%88%D9%84%D8%A7%D9%8A%D8%A9_%D9%82%D8%A7%D8%A8%D8%B3#/media/File:Logo_Governorate_Gabes.svg	0, 124
76.	06A9	ک	ARABIC LETTER KEHEH	Urdu, Malay, Sindhi, Punjabi, Baluchi,	1 Urdu	http://omniglot.com/writing/urdu.htm	0, 112
77.	06AA	ڪ	ARABIC LETTER SWASH KAF	Sindhi	2 Sindhi	http://omniglot.com/writing/sindhi.htm	0, 111
78.	06AB	ګ	ARABIC LETTER KAF WITH RING	Pashto	1 Pashto	Pashto Academy Peshawar University; http://omniglot.com/writing/pashto.htm	0, 108, 138
79.	06AD	ڭ	ARABIC LETTER NG	Kyrghyz, Uyghur, Kazakh	2 Uyghur	For Uyghur: http://www.omniglot.com/writing/uyghur.htm. Also see: http://www.omniglot.com/writing/kirghiz.htm	0, 105, 114, 133, 134

					http://www.ethnologue.com/language/kir "WritingArabic script [Arab], used in China." http://www.ethnologue.com/country/CN/languagesKyrgyz[kir] West and southwest Xinjiang Uyghur Autonomous Region, Wuqia, Akqi, Akto, Tekes, Zhaosu, Baicheng, and Wushi counties. 160,000 in China (2000 census). 60,000 Northern Kirghiz, 40,000 Southern Kirghiz (Shearer and Sun Hong Kai 2002). Older adults monolingual. Ethnic population: 161,000 (2000 census). Status: 2 (Provincial). Statutory provincial language in Xinjiang Uyghur Autonomous Region (1984, Ethnic Regional Autonomy Act, Articles 10 and 21). Language of recognized nationality: Kyrgyz.	
80.	06AE	ﭫ	ARABIC LETTER KAF WITH THREE DOTS BELOW	L'Alphabet National du Tchad (ANT)	1 ANT	ANT (Alphabet National du Tchad) is the national standard for Chad/Tchad; See Figures in L2/10-288R (pp. 19-20) - http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf	0, 116

81.	06AF	گ	ARABIC LETTER GAF	Urdu, Persian, Punjabi, Baluchi, Sindhi,	1 Persian	http://www.omniglot.com/writing/persian.htm	0, 109
82.	06B0	ڰ	ARABIC LETTER GAF WITH RING	Saraiki	5 Saraiki	http://www.omniglot.com/writing/saraiki.htm	0, 110
83.	06B1	ڱ	ARABIC LETTER NGOEH	Sindhi	2 Sindhi	http://www.omniglot.com/writing/sindhi.htm	0, 111
84.	06B3	ڳ	ARABIC LETTER GUEH	Sindhi	2 Sindhi	http://www.omniglot.com/writing/sindhi.htm	0, 111
85.	06B5	ڵ	ARABIC LETTER LAM WITH SMALL V	Kurdish, Pashai	3 Kurdish	http://www.omniglot.com/writing/kurdish.htm See use in Kurdish at http://www.kurdpress.com/	0, 106, 140
86.	06BA	ں	ARABIC LETTER NOON GHUNNA	Urdu	1 Urdu	http://www.omniglot.com/writing/urdu.htm	3, 112
87.	06BB	ڻ	ARABIC LETTER RNOON	Baluchi, Hindko, Saraiki, Sindhi	2 Sindhi	http://www.omniglot.com/writing/sindhi.htm	3, 111
88.	06BC	ڼ	ARABIC LETTER NOON WITH RING	Pashto	1 Pashto	Pashto Academy Peshawar University ; http://www.omniglot.com/writing/pashto.htm	3, 108, 138
89.	06BD	ڽ	ARABIC LETTER NOON WITH THREE DOTS ABOVE	Malay Standard	3 Malay	http://www.omniglot.com/writing/malay.htm	0, 107
90.	06BE	ھ	ARABIC LETTER HEH DOACHASHMEE	Sindhi, Saraiki, Punjabi, Urdu	1 Urdu	http://www.omniglot.com/writing/urdu.htm	0, 112
91.	06C0	ۀ	ARABIC LETTER HEH WITH YEH ABOVE	ANT, Kurdish	1 ANT	ANT (Alphabet National du Tchad) is the national standard for Chad/Tchad; See Figures in L2/10-288R (pp. 19-20); http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf See use in Kurdish at http://www.kurdpress.com/	0, 116, 140

92.	06C1	ه	ARABIC LETTER HEH GOAL	Punjabi, Saraiki, Sindhi, Urdu	1 Urdu	http://omniglot.com/writing/urdu.htm	0, 112
93.	06C2	ۂ	ARABIC LETTER HEH GOAL WITH HAMZA ABOVE	Urdu	1 Urdu	As per Section 3 in http://www.columbia.edu/~mk2580/urdu_section/handouts/izafat.pdf: "In the case of silent word-final choṭī he, the iẓāfat again is indicated not by a zer, but in this case by a little hamza(h) piggybacking on top of the final choṭī he." Combining Hamza not recommended for use by IAB statement and combining marks not included by TF-AIDN; So combined form needs to be included. See https://www.iab.org/documents/correspondence-reports-documents/2015-2/iab-statement-on-identifiers-and-unicode-7-0-0/. Usage: the usage is not optional; see http://www.bbc.com/urdu - see links at the top for Home Page (literally "First page"): صفحۂ اول	0, 125, 135, 141
94.	06C3	ة	ARABIC LETTER TEH MARBUTA GOAL	Punjabi, Urdu	1 Urdu	Code point UZT 76 of Urdu Zabta Takhti 1.01, the official code page	0, 126

					standard for Govt. of Pakistan, approved in 2001; see http://cle.org.pk/Publication/papers/2001/uzt1.01.pdf	
95.	06C4	و	ARABIC LETTER WAW WITH RING	Kashmiri	4 Kashmiri	Document shared by CDAC India; http://omniglot.com/writing/kashmiri.htm	0, 102
96.	06C6	وٗ	ARABIC LETTER OE	Beja, Kurdish, Uyghur, Kashmiri, Kazakh	2 Uyghur	An introduction to Latin script Uughur: by Waris Abdukerim Janbaz State Library of Victoria; http://omniglot.com/writing/kashmiri.htm See use in Kurdish at http://www.kurdpress.com/	0, 102, 140, 142
97.	06CB	ۋ	ARABIC LETTER VE	Kazakh, Kyrgyz, Uyghur	2 Uyghur	Uyghur in Arabic script used in China; http://omniglot.com/writing/uyghur.htm Kazakh in Arabic script used in China and Iran; http://omniglot.com/writing/kazakh.htm http://en.wikipedia.org/wiki/Kazakh_alphabets Kazakh - "The Arabic script is officially used in People's Republic of China in the Altay Prefecture and the Ili Kazakh Autonomous Prefecture of the Xinjiang Uyghur Autonomous Region. It is also used in Iran and	0, 103, 114, 136

						Afghanistan. This is a modified script based on the alphabet used for Kazakh before 1927."	
98.	06CC	ى	ARABIC LETTER PERSIAN YEH	Urdu, Persian/Persian	1 Urdu	http://omniglot.com/writing/urdu.htm	0, 112
99.	06CD	ۍ	ARABIC LETTER YEH WITH TAIL	Pashto	1 Pashto	Pashto Academy Peshawar University; http://omniglot.com/writing/pashto.htm	0, 108, 138
100.	06CE	ێ	ARABIC LETTER YEH WITH SMALL V	Kurdish	3 Kurdish	http://www.fas.harvard.edu/~iranian/Sorani/sorani_1_grammar.pdf (pg. 7); See use in Kurdish at http://www.kurdpress.com/	0, 127, 140
101.	06CF	ۏ	ARABIC LETTER WAW WITH DOT ABOVE	Malay	3 Malay	http://omniglot.com/writing/malay.htm	3, 107
102.	06D0	ې	ARABIC LETTER E	Pashto	1 Pashto	Pashto Academy Peshawar University; http://omniglot.com/writing/pashto.htm	0, 108, 138
103.	06D1	ۑ	ARABIC LETTER YEH WITH THREE DOTS BELOW	Bamana, Madinka	4 Bamana	Vydrin, Valentin Feodos'evich; Dumestre, Gérard (2014): Manding Ajami samples. Mandinka and Bamana. In Meikal Mumin, Kees (C.) H. Versteegh (Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in Semitic Languages and Linguistics, 71), pp. 225–260.	0, 122

104.	06D2	ے	ARABIC LETTER YEH BARREE	Urdu	1 Urdu	http://omniglot.com/writing/urdu.htm	0, 112
105.	06D5	ە	ARABIC LETTER AE	Kurdish, Kirghiz, Kazakh, Uyghur	2 Kurdish, Uyghur	For Kurdish: http://omniglot.com/writing/kurdish.htm. For usage see http://www.kurdpress.com/. For Uyghur: http://www.omniglot.com/writing/uyghur.htm.	0, 106, 114, 140
106.	0751	ݑ	ARABIC LETTER BEH WITH DOT BELOW AND THREE DOTS ABOVE	Wolof, Hausa	4 Wolof 2 Hausa	Ngom, Fallou (2010): Ajami scripts in the Senegalese speech community. In Journal of Arabic and Islamic Studies 10, pp. 1–23. Available online at http://www.lancs.ac.uk/jais/volume/docs/vol10/v10_01_Ngom_1-23.pdf, checked on 7/1/2011. Warren-Rothlin, Andy (2014): West African scripts and Arabic-script orthographies in socio-political context. In Meikal Mumin, Kees (C.) H. Versteegh (Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in Semitic Languages and Linguistics, 71)(see pg. 275) Though considered old use by Currah, Ngom, in his article "Murid Ajami sources of knowledge: the myth and the	7, 121, 128, 130, 147

					reality" in *From Dust to Digital* states "It has been argued that the local Wolof Ajami letters (so-called les lettres vieillies) are found in ancient texts. This suggests that Wolof Ajami letters are rarely or not found in current Ajami materials. This is incorrect. The so-called les lettres vieillies remain the most widely used letters in Wolof communities as illustrated by Figs. 11.5, 11.7, 11.8 and 11.12, and by the plethora of Wolof Ajami materials in the digital repositories cited throughout this article" (pg. 357); see http://www.openbookpublishers.com/htmlreader/978-1-78374-062-8/11.Ngom.xhtml#_idTextAnchor144	
107.	0752	ﺒ	ARABIC LETTER BEH WITH THREE DOTS POINTING UPWARDS BELOW	Serer, Wolof	4 Wolof	http://omniglot.com/writing/wolof.htm For usage, see: BENEATH THE SURFACE? CONTEMPORARY AJAMI WRITING IN WEST AFRICA, EXEMPLIFIED THROUGH WOLOFAL by Lüpke and Bao-Diop 2014 (http://eprints.soas.ac.uk/19019/	7, 113, 130

					1/03%20Luepke%20%2526%20Bao%20Diop%20clean.pdf) For use of Ajami/Wolofal script for African languages, see "Language planning in West Africa – who writes the script?" by FREDERIKE LÜPKE; available at http://www.elpublishing.org/docs/1/02/ldd02_08.pdf	
108.	0756	ﺖّ	ARABIC LETTER BEH WITH SMALL V	Wolof, Pular, Sonigke	2 Pular 4 Wolof	http://omniglot.com/writing/wolof.htm http://www.paul-timothy.net/pages/wolof/wolofal/orthography/wolofal_orthography_currah_20-may-2011_a4.pdf http://paul-timothy.net/pages/ajamisenegal/je_sais_le_wolofal_06-jan-2013_a4.pdf For use of Ajami/Wolofal script for African languages, see "Language planning in West Africa – who writes the script?" by FREDERIKE LÜPKE; available at http://www.elpublishing.org/docs/1/02/ldd02_08.pdf	7, 113, 120, 130
109.	0760	ﻴڜ	ARABIC LETTER FEH WITH TWO DOTS BELOW	Fulfulde	3 Fulfulde	Warren-Rothlin, Andy (2014): West African scripts and Arabic-script orthographies in socio-	7, 121, 130

|						
					political context. In Meikal Mumin, Kees (C.) H. Versteegh (Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in Semitic Languages and Linguistics, 71), pp. 261–289. For use of Ajami/Wolofal script for African languages, see "Language planning in West Africa – who writes the script?" by FREDERIKE LÜPKE; available at http://www.elpublishing.org/docs/1/02/ldd02_08.pdf	
110.	0762	کۡ	ARABIC LETTER KEHEH WITH DOT ABOVE	Malay	3 Malay	Jawi Keyboard standard by Department of Standards Malaysia Also see http://www.unicode.org/L2/L2003/03176-gafs.pdf	7, 118, 129
111.	0763	گۡ	ARABIC LETTER KEHEH WITH THREE DOTS ABOVE	Almaghreb Alarabi	1 Arabic	For example city of Agadir in Morocco (أكَّادير) Also see http://www.unicode.org/L2/L2003/03176-gafs.pdf	7, 118
112.	0766	مۡ	ARABIC LETTER MEEM WITH DOT BELOW	ANT	1 ANT	See Figures in L2/10-288R (pp. 19-20) http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf	7, 116, 121

						Warren-Rothlin, Andy (2014): West African scripts and Arabic-script orthographies in socio-political context. In Meikal Mumin, Kees (C.) H. Versteegh (Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in Semitic Languages and Linguistics, 71), pp. 261–289.	
113.	0767	ﻨٟ	ARABIC LETTER NOON WITH TWO DOTS BELOW	Wolof, Serer	4 Wolof	http://omniglot.com/writing/wolof.htm http://www.paul-timothy.net/pages/wolof/wolofal/orthography/wolofal_orthography_currah_20-may-2011_a4.pdf http://paul-timothy.net/pages/ajamisenegal/je_sais_le_wolofal_06-jan-2013_a4.pdf	7, 113, 120
114.	0768	ﻄٗ	ARABIC LETTER NOON WITH SMALL TAH	Saraiki	5 Saraiki	http://en.wikipedia.org/wiki/Saraiki_alphabet	7, 145
115.	076A	ڪ	ARABIC LETTER LAM WITH BAR	ANT, Wolof, Gawri	1 ANT 4 Wolof	http://omniglot.com/writing/wolof.htm See Figures in L2/10-288R http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf (pp. 19-20) Warren-Rothlin, Andy (2014): West African scripts and Arabic-script orthographies in socio-political context. In Meikal Mumin, Kees (C.) H. Versteegh	7, 113, 116, 120

					(Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in Semitic Languages and Linguistics, 71), pp. 261–289. http://www.paul-timothy.net/pages/wolof/wolofal/orthography/wolofal_orthography_currah_20-may-2011_a4.pdf http://paul-timothy.net/pages/ajamisenegal/je_sais_le_wolofal_06-jan-2013_a4.pdf	
116.	076E	ح	ARABIC LETTER HAH WITH SMALL ARABIC LETTER TAH BELOW	Khowar	3 Khowar	Anjuman Taraqi-i-Khowar, Chitral; http://omniglot.com/writing/khowar.htm	9, 104
117.	076F	ح	ARABIC LETTER HAH WITH SMALL ARABIC LETTER TAH AND TWO DOTS	Khowar	3 Khowar	Anjuman Taraqi-i-Khowar, Chitral; http://omniglot.com/writing/khowar.htm	9, 104
118.	0770	ش	ARABIC LETTER SEEN WITH SMALL ARABIC LETTER TAH AND TWO DOTS	Khowar	3 Khowar	Anjuman Taraqi-i-Khowar, Chitral; http://omniglot.com/writing/khowar.htm	9, 104
119.	0771	ر	ARABIC LETTER REH WITH SMALL ARABIC LETTER TAH AND TWO DOTS	Khowar	3 Khowar	Anjuman Taraqi-i-Khowar, Chitral; http://omniglot.com/writing/khowar.htm	9, 104
120.	08A0	ب	ARABIC LETTER BEH WITH SMALL V BELOW	Direction de la Promotion des Langues Nationales (DPLN), Senegal	1 DPLN	L2/10-288R; http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf Figure 2a in (pg. 21).	12, 117

						"The Directíon de la Promotion des Langues Nationales (DPLN) 1 is under the Ministry of Education of Senegal. The DPLN worked to standardize the orthographies for languages of Senegal. This includes languages using the Arabic script. The first version of the harmonized script (HS) was done in 1986; the latest in 1990. The DPLN charts have not been decreed by the Senegal government, because the harmonization was intended to be multi-national, and therefore a decree could not (or should not) be made by one country… This is the official standard for the languages listed (Wolof [wol], Pulaar [fuc], Soninke [snk], Mandinka [mnk], Seereer [srr], Joola [dyo] and Balant [bjt])." (pg. 5)	
121.	08A2	جٚ	ARABIC LETTER JEEM WITH TWO DOTS ABOVE	DPLN	1 DPLN	L2/10-288R; http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf	12, 117
122.	08A3	ظٚ	ARABIC LETTER TAH WITH TWO DOTS ABOVE	DPLN; Wolof	1 DPLN	http://omniglot.com/writing/wolof.htm ;	12,113, 117

					L2/10-288R; http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf	
123.	08A4	ڤ	ARABIC LETTER FEH WITH DOT BELOW AND THREE DOTS ABOVE	ANT	1 ANT	See Figures in L2/10-288R http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf (pp. 19-20)	12, 116
124.	08A5	ڦ	ARABIC LETTER QAF WITH DOT BELOW	ANT	1 ANT	See Figures in L2/10-288R http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf (pp. 19-20)	12, 116
125.	08A6	ڧ	ARABIC LETTER LAM WITH DOUBLE BAR	ANT	1 ANT	See Figures in L2/10-288R http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf (pp. 19-20)	12, 116
126.	08A7	ڨ	ARABIC LETTER MEEM WITH THREE DOTS ABOVE	ANT	1 ANT	See Figures in L2/10-288R http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf (pp. 19-20)	12, 116
127.	08A8	ئ	ARABIC LETTER YEH WITH TWO DOTS BELOW AND HAMZA ABOVE	Fulfulde	3 Fulfulde	Warren-Rothlin, Andy (2014): West African scripts and Arabic-script orthographies in socio-political context. In Meikal Mumin, Kees (C.) H. Versteegh (Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in Semitic Languages and Linguistics, 71), pp. 261–289. For use of Ajami script for Fula languages, see "Language planning in West Africa – who writes the script?" by FREDERIKE	12, 121

					LÜPKE; available at http://www.elpublishing.org/docs/1/02/ldd02_08.pdf	
128.	08A9	ئ	ARABIC LETTER YEH WITH TWO DOTS BELOW AND DOT ABOVE	Fulfulde	3 Fulfulde	Warren-Rothlin, Andy (2014): West African scripts and Arabic-script orthographies in socio-political context. In Meikal Mumin, Kees (C.) H. Versteegh (Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in Semitic Languages and Linguistics, 71), pp. 261–289.	12, 121

4 Final Recommendation of Variants for Top Level Domains (TLDs)

During interactions with the Integration Panel, one of the technical requirements highlighted was that the allocatable variants must be minimized for the Root Zone. The initial proposal developed by TF-AIDN still generated significant allocatable variants (see the details of this analysis in Appendix G). Thus, this initial analysis was further revised by the Task Force to meet these technical requirements for the TLDs. The allocatable variants were further constrained for TLDs by changing the type of additional variant code points to *blocked*, noting that TLDs are not applied-for by the general registrants and that TLD applicants choose the relevant label after substantial effort, considering usability, correctness, manageability and other factors related to the label. A less conservative strategy could be determined for other levels, as per registry policies.

For reading the type of variants in the tables below, note that:

- Where variant type is same in both directions, 'a/a' is written as 'a' and 'b/b' is written as 'b'
- Where variant type is not same in both directions, the type is given as 'a/b' or 'b/a' and further specified in the final column.

Further, in the tables for variant types, alternate sets of rows for a code point in the first column have been shaded for better readability (and therefore should be treated as a formatting artifact only).

Note of table numbering: Tables in this section are labeled as "Table #a" (e.g. Table 1a) because the original version of these tables, "Table #" (e.g. Table 1), are given in Appendix G, as discussed above.

Table 1a: YEH Class

Unicode Code Point	Isolated Form	Initial Form	Final Form	Medial Form	Applicable Principle No.
0649	ى	ـب	بى	ببب	1
06CC	ى	يـب	بى	بيب	1
064A	ي	يـب	بي	بيب	1
06D0	ې	ٻـب	بې	ببب	4
067B	ٻ	ٻـب	بٻ	ببب	4
06CD	ؽ		بؽ		3
06D2	ے		بے		2 (06CC), 5 (064A)
0626	ئ	ئـب	بئ	بئـب	3

Code Point	Code Point	Type (b=blocked/ a=allocatable)	Notes
0649	06CC	b	
0649	064A	b	
0649	06D0	b	
0649	067B	b	
0649	06CD	b	
0649	06D2	b	
0649	0626	b	
06CC	064A	a	The two are visually identical and the same label could be typed using one or the other based on the set language settings and keyboard layout of a user. E.g. ccTLD labels ایران (Iran) and ملیسیا (Malaysia) will be written with 06CCor 064A depending on the country – both versions are identical need to be available to the entire community.
06CC	06D0	b	
06CC	067B	b	
06CC	06CD	b	
06CC	06D2	b	
06CC	0626	b	
064A	06D0	b	
064A	067B	b	
064A	06CD	b	
064A	06D2	b	
064A	0626	b	

06D0	0626	b	
06D0	067B	b	
06D0	06CD	b	
06D0	06D2	b	
067B	06CD	b	
067B	06D2	b	
067B	0626	b	
06CD	06D2	b	
06CD	0626	b	
0626	06D2	b	

Table 2a: HEH Class

06C1	ہ	بـہ	ـہـ	بـہـب	1
06D5	ہ		ـہ		1
0647	ه	هـب	ـه	بـهـب	1
06BE	ھ	ھـب	ـھـ	بـھـب	1
0629	ة		ـة		3
06C3	ة		ـة		3 (with Heh) 1 (with 0629)
06C0	ۀ		ـۀ		3 1 (06C2)
06C2	ۂ		ـۂ		3 1 (06C0)

06C1	06D5	b	
06C1	0647	a	Labels in Arabic language using 0647 in the final and isolated positions will be typed in other languages using 06C1 (Urdu, Pashto, Saraiki, etc.) which is identical in isolated and has a variant glyph or identical glyph form in final position.
06C1	06BE	b	
06C1	0629	b	
06C1	06C3	b	This is not allocatable in either direction because, unlike the allocatable relationship suggested between 0647 and 0629 due to variation in the Arabic language, Urdu and other languages using 06C1 do not exhibit such variation
06C1	06C0	b	

06C1	06C2	b/a	06C1 ->b 06C2 (blocked), 06C2 ->a 06C1 (allocatable); Heh with Hamza is sometimes confused with the simple Heh in some languages, e.g. Urdu
06D5	0647	b	
06D5	06BE	b	
06D5	0629	b	
06D5	06C3	b	
06D5	06C0	b/a	06D5 ->b 06C0 (blocked), 06C0 ->a 06D5 (allocatable); Heh with Hamza may be confused with the simple Heh in some languages, e.g. Kurdish
06D5	06C2	b	
0647	06BE	b	
0647	0629	b/a	0647 ->b 0629[9](blocked), 0629 ->a 0647 (allocatable). In Arabic language, Heh may be substituted by Teh Marbuta; RFC6365
0647	06C3	b	
0647	06C0	b	
0647	06C2	b	
06BE	0629	b	
06BE	06C3	b	
06BE	06C0	b	
06BE	06C2	b	
06C3	0629	a	The two are visually identical and the same label could be typed using one or the other based on the set language settings and keyboard layout of a user. E.g. ccTLD labels السعودية (Saudi Arabia) will be written with 0629 or 06C3 depending on the country. Labels in Arabic language using 0629 in the final and isolated positions will be typed in other languages using 06C3 (Urdu, etc.) which is identical in isolated and has a variant glyph or identical glyph form in final position.
06C3	06C0	b	
06C3	06C2	b	
0629	06C0	b	
0629	06C2	b	
06C0	06C2	b	

---

[9]Two sets of data are collected for second level IDNs to determine the need by SaudiNIC under its IDN ccTLD: From about 2000 registered names at the second level which should have ended in Teh Marbutah, 67.18% were registered with Heh instead (e.g. for حديقة the user registered حديقه); From about 2000 DNS queries for IDN domain names, 17.87% of the queries related to domains which actually contain words that end with Heh (like فقه), the user mistyped with Teh Marbutah (فقة); Thus, there is evidence that this direction may also be allocated. Nevertheless, TF-AIDN notes that root level is different from 2nd level and it should be more restricted and hence it can be blocked in the root. At second level, it may be desirable to keep this as an allocatable variant for better usability experience for users who are casual about adhering to correct spelling rules.

Table 3a: KAF Class

0643	ك	كب	بك	بكب	1
06A9	ک	کب	بک	بکب	1
06AA	ڪ	ڪب	بڪ	بڪب	2, 5 (0643)

0643	06A9	a	The two have identical shapes in initial and medial positions and are used by different language communities to refer to the same letter e.g. common words like كتاب (book) will be written with 0643 in Arabic language but with 06A9 in Urdu, Persian, etc.
0643	06AA	a	The two have similar (interchangeable) shapes in initial and medial positions and are used by different language communities to refer to the same letter e.g. common words like ڪتاب (book) will be written with 0643 in Arabic language but with 06AA in Sindhi. Similarly, 06AA is used in lieu of 0643 in Western African orthographies.
06A9	06AA	a	The two have similar (interchangeable) shapes in initial and medial positions and are used by different language communities to refer to the same letter e.g. common words like ڪتاب (book) will be written with 06A9 in Urdu, Persian, etc. languages but with 06AA in Sindhi. These are used as distinct characters in Sindhi.

Table 4a: PEH Class

067E	پ	پب	بپ	بپب	1
0752	ݒ	ݒب	بݒ	بݒب	4
06D1	ۑ	پب	بۑ	بپب	1
06BD	ڽ	پب	بڽ	بپب	1

067E	0752	b	
067E	06D1	b	
067E	06BD	b	
0752	06D1	b	
0752	06BD	b	
06D1	06BD	b	

Table 5a: NOON WITH TWO DOTS BELOW Class

0767	ݧ	ݧب	بݧ	بݧب	1
08A9	ࢩ	ب࢏	࢏ب	ب࢏ب	1

0767	08A9	b	

Table 6a: FEH Class

0641	ف	فب	بف	بفب	1 (06A7)
06A2	ڢ	ڢب	بڢ	بڢب	5 (0641)
06A7	ڧ	ڧب	بڧ	بڧب	5 (0642)
0642	ق	قب	بق	بقب	6 (06A7)

0641	06A2	a	Used interchangeably in Africa for languages using Western (African)orthography
0641	06A7	b	
0641	0642	b	
06A2	06A7	b	
06A2	0642	b	
06A7	0642	a	Used interchangeably in Africa for languages using Western (African)orthography

Table 7a: TTHEH Class

0679	ٹ	ٹب	بٹ	بٹب	1
06BB	ٻ	ٻب	بٻ	بٻب	1

0679	06BB	b	

Table 8a: NOON Class

0646	ن	نب	بن	بنب	1
06BA	ں	ںب	بں	بںب	1, 5 (0646) Used in Urdu along with 0646 as two distinct letters

0646	06BA	a	Used interchangeably in Africa for languages using Western (African)orthography

Table 9a: TEH Class

062A	ت	تب	بت	بتب	4

067A	ﭺ	ﻨﺐ	ﺑﺚ	ﺑﻨﺐ	4

062A	067A	b	

Table 10a: NYEH Class

0683	ﭺ	ﭻﺐ	ﺑﺞ	ﺑﭻﺐ	4
0684	ﭺ	ﭻﺐ	ﺑﺞ	ﺑﭻﺐ	4

0683	0684	b	

Table 11a: VEH Class

06A4	ﭪ	ﻗﺐ	ﺑﻒ	ﺑﻘﺐ	1
06A8	ﭢ	ﻗﺐ	ﺑﻖ	ﺑﻘﺐ	1

06A4	06A8	b	

Table 12a: THEH Class

062B	ﺙ	ﺛﺐ	ﺑﺚ	ﺑﺜﺐ	4
067D	ﺙ	ﻨﺐ	ﺑﺚ	ﺑﻨﺐ	4

062B	067D	b	

Table 13a: GAF Class

06AF	ﮒ	ﺑﮒ	ﮔﺐ	ﺑﮕﺐ	3
06AD	ﯕ	ﮐﺐ	ﺑﯕ	ﺑﮑﺐ	1
06AB	ﮎ	ﺑﮒ	ﮔﺐ	ﺑﮕﺐ	3; interchangeably used in Pashto with 06AF
0763	ﮒ	ﮐﺐ	ﺑﮒ	ﺑﮕﺐ	1; with 06AD; 3; with 06AF; Iraqi Arabic uses 06AF, whereas Moroccan Arabic uses 0763

06AB	06AD	b	
06AB	06AF	b	
06AB	0763	b	
06AD	06AF	b	
06AD	0763	b	
06AF	0763	b	

Table 14a: DUL Class

068E	دَ		بدَ		4
068F	دَ		بدَ		4

068E	068F	b	

Table 15a: ALEF with HAMZA Class

0623	أ		بأ		3
0622	آ		بآ		3
0625	إ		بإ		3
0627	ا		با		3
0672	أ		بأ		3

0623	0627	a/b	0623 ->a 0627 (allocatable), 0627 ->b 0623. Alef with Hamza above is simplified to Alef in Arabic language; الأردنforالاردن(the current IDN ccTLD for Jordan)
0623	0622	b	
0623	0625	b	
0623	0672	b	
0627	0622	b/a	0622 ->a 0627 (allocatable), 0627 ->b 0622. Alef with Madda above is simplified to Alef in Arabic language
0627	0625	b/a	0625 ->a 0627 (allocatable), 0627 ->b 0625. Alef with Hamza below is simplified to Alef in Arabic language; إماراتforامارات(the current IDN ccTLD for United Arab Emirates)
0627	0672	b/a	0672 ->a 0627 (allocatable), 0627 ->b 0672. Simplified to Alef in Kashmiri language
0622	0625	b	
0622	0672	b	
0625	0672	b	

Table 16a: Waw Class

0648	و		بو		3
0624	ؤ		بؤ		3

0624	0648	a/b	0624 ->a 0648 (allocatable), 0648 ->b 0624; Waw with Hamza is simplified to Waw in some languages, e.g. Arabic and Urdu (blocked)

5 Whole Label Evaluation (WLE) Rules

TF-AIDN considers two motivations behind adding whole label evaluation rules. The first one is to reduce the confusability within a label while the second one is to maintain the conservative approach and minimize the number of allocatable variants.

The confusability is introduced due to same or similar shape of characters, as well as language specific behavior encoded separately in the Unicode.

1. Restrict occurrences of variant code points within a label to reduce confusability, with the exception of the following cases:
    a. Variants do not have identical shape in any position, therefore do not cause any confusion
    b. Variants are distinct characters used within a language and therefore distinguishable by the community

Based on the analysis, the following pairs of code points should not be mixed in a label. These constraints do not limit label generation for any particular language and are aimed at reducing the allocation of redundant labels formed by code point variants. Labels which contain both code points should be blocked.

Table 17: WLE Rules for Code Points

No.	Code points which cannot co-occur within a label	Reference Variant Mapping Table	Notes
1	0643 and 06A9	3a	ک and ك
2	0643 and 06AA	3a	ڪ and ك
3	06CC and 0649	1a	ی and ى
4	06C1 and 0647	2a	ہ and ﻫ
5	06C1 and 06D5	2a	ہ and ە
6	0647 and 06D5	2a	ﻫ and ە
7	0647 and 06BE	2a	ﻫ and ھ
8	0629 and 06C3	2a	ة and ۃ
9	06D1 and 06BD	4a	ی and ڽ
10	067E and 06BD	4a	پ and ڽ

11	0641 and 06A2	6a	ب and ف
12	0642 and 06A7	6a	ف and ق
13	0641 and 06A7	6a	ف and ف
14	06AB and 06AF	13a	گ and گ
15	06AB and 0763	13a	ڭ and گ
16	06AFand 0763	13a	ڭ and گ

References

0 . The Unicode Standard 1.1

3 . The Unicode Standard 3.0

7 . The Unicode Standard 4.1

9 . The Unicode Standard 5.1

11 . The Unicode Standard 6.0

12 . The Unicode Standard 6.1

100 . RFC 5564 Linguistic Guidelines for the Use of the Arabic Language in Internet Domains

101 . Omniglot Hausa http://omniglot.com/writing/hausa.htm

102 . Omniglot Kashmiri http://omniglot.com/writing/kashmiri.htm

103 . Omniglot Kazakh http://omniglot.com/writing/kazakh.htm

104 . Omniglot Khowar http://omniglot.com/writing/khowar.htm

105 . Omniglot Kirghiz http://omniglot.com/writing/kirghiz.htm

106 . Omniglot Kurdish http://omniglot.com/writing/kurdish.htm

107 . Omniglot Malay http://omniglot.com/writing/malay.htm

108 . Omniglot Pashto http://omniglot.com/writing/pashto.htm

109 . Omniglot Persian(Farsi) http://omniglot.com/writing/persian.htm

110 . Omniglot Saraiki http://omniglot.com/writing/saraiki.htm

111 . Omniglot Sindhi http://omniglot.com/writing/sindhi.htm

112 . Omniglot Urdu http://omniglot.com/writing/urdu.htm

113 . Omniglot Wolof http://omniglot.com/writing/wolof.htm

114 . Omniglot Uyghur http://omniglot.com/writing/Uyghur.htm

115 . Unicode, Kashmiri, Yeh http://www.unicode.org/L2/L2009/09215-kashmiri.pdf

116 . Unicode, Chad ANT, pp. 19-20 http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf

117 . Unicode, DPLN, p.21 http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf

118 . Unicode, Jawi and Moroccan Arabic GAF, http://www.unicode.org/L2/L2003/03176-gafs.pdf

119 . Unicode, Chadian, p.5 http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf

120 . Wolof, Paul Timothy http://paul-timothy.net/pages/ajamisenegal/primers/je_sais_le_wolofal_harmattan_20-oct-2015_a4.pdf

121 . Hausa, pp. 261-289 Warren-Rothlin, Andy (2014): West African scripts and Arabic-script orthographies in socio-political context. Meikal Mumin, Kees (C.) H. Versteegh (Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in Semitic Languages and Linguistics, 71)

122 . Mandika, Bamana, pp. 225-260 Vydrin, Valentin Feodos'evich; Dumestre, Gérard (2014): Manding Ajami samples. In Meikal Mumin, Kees (C.) H. Versteegh (Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in Semitic Languages and Linguistics, 71)

123 . Ethiopian, Wetter, Andreas (2006): Arabic in Ethiopia. In Kees (C.) H. Versteegh (Ed.): Encyclopedia of Arabic Language and Linguistics. Volume I. A-Ed, vol. 2. With assistance of Mushira Eid, Alaa Elgibali, Manfred Woidich, Andrzej Zaborksi. Leiden: E. J. Brill, pp. 51-56.

124 . Western Arabic, Qaf with three dots above, city of Gabes, Tunisia

125 . Urdu, Heh goal with hamza above, Section 3 in http://www.columbia.edu/~mk2580/urdu_section/handouts/izafat.pdf

126 . Urdu, Teh marbuta goal, Code point UZT 76 of Urdu Zabta Takhti 1.01, the official code page standard for Govt. of Pakistan, approved in 2001; see http://cle.org.pk/Publication/papers/2001/uzt1.01.pdf

127 . Kurdish-Sorani, p.7 in http://www.fas.harvard.edu/~iranian/Sorani/sorani_1_grammar.pdf

128 . Wolof, Beh with dot below and three dots above, WOLOFAL ORTHOGRAPHY by Galen Currah, revised 20 May 2011

129 . Malay, Jawi Keyboard standard by Department of Standards Malaysia https://en.wikipedia.org/wiki/Jawi_keyboard, accessed on 13 November 2015

130 . Ajami usage, Language planning in West Africa - who writes the script?" by FREDERIKE LÜPKE; http://www.elpublishing.org/docs/1/02/ldd02_08.pdf

131 . Fulfulde,
http://www.silcam.org/documents/AlphabetandOrthographyStatementforFulfuldeFUBAj
amiyafortheinternet.pdf

132 . Hausa, Newspapers and books published in Hausa using Arabic script,
http://aflang.humnet.ucla.edu/Hausa/Pronunciation/writing.html

133 . Kyrgyz, http://www.ethnologue.com/language/kir

134 . Kyrgyz, http://www.ethnologue.com/country/CN/languages

135 . Urdu, http://www.bbc.com/urdu

136 . Kyrgyz, http://en.wikipedia.org/wiki/Kazakh_alphabets, accessed on 13 November 2015

137 . Malay, Information technology - Jawi coded character set for information interchange MS
2443:2012, Department of Standards, Malaysia. http://www.standardsmalaysia.gov.my

138 . Pashto Academy Peshawar University

139 . Saraiki, https://id-id.facebook.com/jhoke.saraiki

140 . Kurdish, http://www.kurdpress.com/

141 . Combining Hamza not recommended for use by IAB statement and combining marks not
included by TF-AIDN, so combined forms need to be included; also see
https://www.iab.org/documents/correspondence-reports-documents/2015-2/iab-
statement-on-identifiers-and-unicode-7-0-0/

142 . An introduction to Latin script Uughur: by Waris Abdukerim Janbaz State Library of
Victoria

143 . Torwali online and printed dictionaries

144 . https://en.wikipedia.org/wiki/Jawi_keyboard, accessed on 13 November 2015

145 . http://en.wikipedia.org/wiki/Saraiki_alphabet, accessed on 13 November 2015

146 . ANT (Alphabet National du Tchad) is the national standard for Chad/Tchad; See Figures in
L2/10-288R (used for "tr" sound as given in the table in Section 6.1.2):
http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf and Appendix E

147 . Wolof, http://www.openbookpublishers.com/htmlreader/978-1-78374-062-
8/11.Ngom.xhtml#_idTextAnchor144

Appendix A: Some of the Languages using Arabic Script

There are many languages spoken around the world, which are written using Arabic script. This appendix lists only those languages that have larger populations writing the language in Arabic script today. The short-listing is only to keep the list tractable and does not imply that unlisted languages are less significant. The languages not listed here were also be considered for analysis of script by the Generation Panel. There are cases where a language may have a large population, but only a small part of it writes it in Arabic script. Such languages are also excluded from this list. Languages defined as living by ISO 639-3 (http://www-01.sil.org/ISO639-3/codes.asp) are included, however codes may refer to either groups of languoids (e.g. macro languages) or individual languages. Furthermore, a non-exhaustive list of countries in which the language is written using Arabic script is provided. The languages may be used in many more countries by expatriate communities as well as for academic, religious, or other specialized purposes. Languages have official status in the countries listed, unless they are marked with an * mark.

Language	ISO 639-3 Code(s)	Countries	Local Names of the Script
Acehnese	ace	Indonesia*	جاوي
Amharic	amh	Ethiopia	*Ajäm*
Arabic	Aao, abh, abv, acm, acq, acw, acx, acy, adf, aeb, aec, afb, ajp, ajt, aju, apc, apd, ara, arb, ary, arz, auz, avl, ayh, ayl, ayn, ayp	Algeria, Bahrain, Chad, Comoros, Djibouti, Egypt, Eritrea, Iraq, Israel*, Jordan, Kuwait, Lebanon, Libya, Mauritania, Morocco, Oman, Palestine, Qatar, Saudi Arabia, Somalia, Sudan, Syria, Tunisia, United Arab Emirates, Yemen, Guinea*, Mauritius*, Uganda*	العربية
Aynu		China*	
Azerbaijani	azb, aze, azj	Azerbaijan, Iran*, Iraq*, Georgia*, Armenia*, Turkey*, Syria* and Russia*	
Bakhtiari	bqi	Iran*	
Baluchi	bcc, bgn, bgp	Pakistan*, Iran*, Afghanistan*	
Balti	bft	Pakistan*	
Banjarese	bjn	Indonesia*, Malaysia*	جاوي
Beja	bej	Egypt*, Sudan* and Eritrea*	
Bosnian		Bosnia and Herzegovina*	

Brahui	brh	Pakistan*, Afghanistan*, Iran*	
Burushaski	bsk	Pakistan*	
Cham	cja	Cambodia	
Crimean Tatar		Uzbekistan*	
Gilaki	glk	Iran*	
Hausa	hau	Nigeria, Niger	عجمي
Hindko	hnd, hno	Pakistan*	
Karakalpak		Uzbekistan*	
Kanuri	kau, kby, knc	Niger ,Nigeria*	*Tarjumo/Turjiman*
Kashkay	qxq	Iran, Iraq	
Kashmiri	kas	Pakistan*, UK*, India*	
Kazakh	kaz	Kazakhstan, Russia*, China*	
Khowar	khw	Pakistan*	
Kohistani/Shina	scl	Pakistan*	
Kurdish	kur, kmr, ckb, sdh, lki	Iraq, Turkey*, Iran*, Syria*, Lebanon*, Armenia*, Georgia*, Kyrgyzstan*, Azerbaijan*, Kazakhstan*,Afghanistan*	
Kyrgyz	kir	Kyrgyzstan, China*, Afghanistan*, Tajikistan*	
Luri	ldd	Iran*	
Malay	zlm, zsm, meo, mfa	Malaysia, Brunei, Indonesia*, Singapore*,Thailand*	جاوي
Marwari	mve	India* and Pakistan*	
Mazandarani	mzn, srz	Iran*	
Minangkabau	min	Indonesia*	جاوي
Pashto	pbt, pbu, pst	Afghanistan, Pakistan*	

Persian and Dari	far, pes, prs	Iran, Afghanistan, Tajikistan*	الفبای فارسی دري
Potwari	phr	Pakistan*	
Punjabi	pan, pmu, pnb	Pakistan*	
Salar		China*	
Seraiki	skr	Pakistan*, India*	
Shabaki		Iraq*	
Sindhi	snd	Pakistan*	
Somali	som	Somalia, Ethiopia, Kenya, Djibouti*, and Yemen*	(likely that Arabic script is not in contemporary use) Wadaad's (وَداد) writing
Swahili	swh, swc, swa	Kenya, Tanzania, Uganda, DR Congo*	Kiarabu
Tajiki	tgk	Tajikistan, Kyrgyzstan*, Uzbekistan*, Turkmenistan*	
Tausug	tsg	Philippines*, Malaysia* and Indonesia*	جاوي
Turkmen	tuk	Afghanistan*, Iran*	
Urdu	urd	India, Pakistan, Bangladesh*, Nepal*, UAE*	اردو،نستعليق
Uyghur	uig	China*, Afghanistan*	
Uzbek	uzb, uzn, uzs	Uzbekistan, Afghanistan*, China*	
Wolof	wol, wof	Senegal, Mauritania, Gambia*	*Wolofal* وَلَفْ

Appendix B: Use of Ajami Writing System

For most of the languages referenced in the table below, the use of Arabic script is evident and widely established. However, for some languages of Africa and Central Asia, Arabic orthographies, which had been in use for centuries, have since been officially replaced with other orthographies (mostly Roman/Latin or Cyrillic script-based ones). In very few cases, such language policy was enforced to such a degree that any previous writing practices have effectively been replaced., However, in numerous other cases, there is still active use of Arabic script, as described by, for example, Lüpke and Bao-Diop in their article "Beneath the Surface? Contemporary Ajami Writing in West Africa, Exemplified through Wolofal" published in 2014. The authors note that Ajami scripts (which, as discussed above, are modified Arabic alphabet used to write languages in Africa) "are used throughout West Africa; Figure 1 shows the distribution of those of their uses that have been attested in the literature. [...] The earliest document showing an African language written in Arabic characters, featuring Songhai, dates from the 13th century [...] documents testifying a conventionalized Ajami writing system are attested from the 16th century onwards. From then on, a wealth of religious and profane literature in Hausa, Fula and Wolof [...] is visible proof of this literary tradition. As shown in Figure 1, in addition to these *linguae francae*, Ajami writing is reported for an important number of other languages, and it is to be expected that more occurrences of Ajami writing in smaller languages will emerge once more research interest in this domain is ignited" (pp. 3-4).
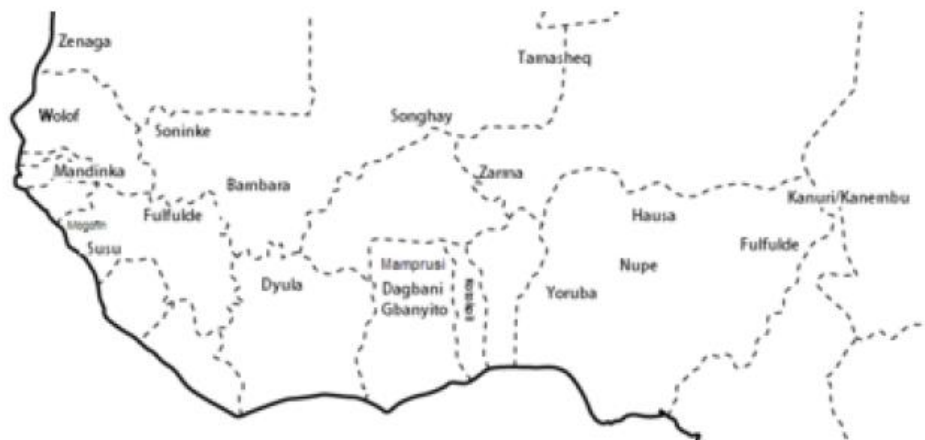

Figure 1: Map of languages for which Ajami use is attested (Souag 2010:1)

Another view is given by Lüpke (2004)[10].

---

[10]Frederike Lüpke (2004). Language planning in West Africa - who writes the script?. In Peter K. Austin (ed.) Language Documentation and Description, vol 2. London: SOAS. pp. 90-107 (at http://www.elpublishing.org/docs/1/02/ldd02_08.pdf)
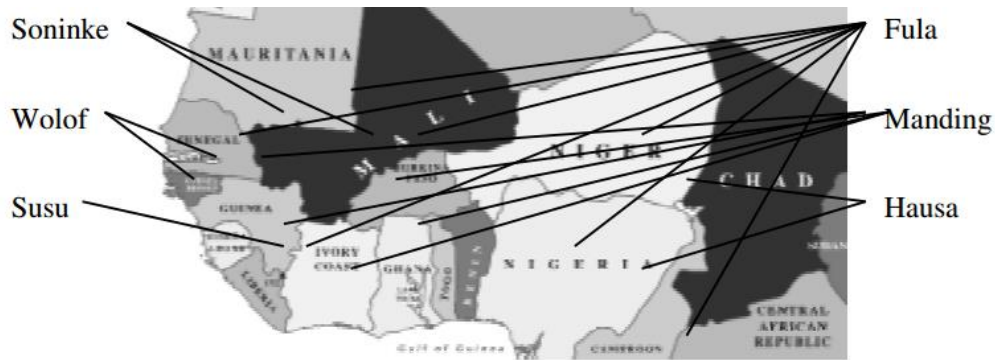
Map 1: West Africa and countries with languages using Arabic-based scripts

Lüpke and Bao-Diop further note that "Despite being (near to) invisible to educators, language planners and development activists, a pre-colonial literacy tradition continues to be practiced throughout those areas of Africa that are in the sphere of influence of Islam. This writing tradition uses Arabic-based scripts (also called Ajami) for the writing of African languages" (p. 1). They reiterate (on pp. 2-3) the observation by Souag (2010) that "This practice is attested in practically all Muslim areas of West Africa, including at least Senegal, Gambia, Guinea-Bissau, Guinea, Mauretania, Mali, Niger, Ghana, Nigeria and Cameroon. It continues to the present despite being propagated almost exclusively through traditional religious instruction, usually without government funding or recognition; in this sense, it might be called a non-governmental literacy, as opposed to literacy whose norms are passed on through a government-organized school system."

Lüpke and Bao-Diop provide an overview of the extensive use of Ajami scripts use in present-day West Africa from their fieldwork in Guinea, Cameroon and Senegal, illustrating "how Ajami writing becomes visible as soon as the Eurocentric perspective on reading and writing is abandoned" (p. 1). Though they also note that for some of the languages, like Pulaar, the use of Ajami writing is declining, others, like Wolofal (Ajami writing of Wolof) "continue to thrive. Although Arabic-based scripts have been officially harmonized and recognized in Senegal in 2002 […] in practice they still remain in informal contexts and defy standardization, which does not hamper their success" (p. 2).

Interestingly, Lüpke and Bao-Diop note that "different policies were adopted in French and British colonies regarding not only the role of African languages but also the interference with the existing writing traditions. So in contrast to the British policy regarding Hausa, whose Arabic script was actively discouraged and replaced, the French colonizers of Guinea for instance ignored indigenous Fula writing traditions, since their goals was to create a population literate in French. This divergence resulted in different contemporary weights of the Ajami traditions for Hausa in Nigeria and Fula in Guinea respectively. In Nigeria, where Hausa Ajami survives, it is flanked by very visible use of Romanized Hausa, particularly in formal context and

in publishing, although "[t]the informal use of the Ajami in manuscript by scholars, merchants and others continues today wherever there are Hausa speakers (Philips 2000:27)"".

Lüpke and Bao-Diop go on to write that "there are still books and newspapers produced in it. In contrast for the Futa Jalon region of Guinea, the facts point to a continuing popularity of Ajami; Despite a standardized Roman orthography created for the Fulfulde variety of Fula in Guinea and used in adult literacy campaigns, the Ajami tradition persists until today, seeing the birth of new genres, and resulting in a flourishing writing environment. Salvaing and Hunwick (2003) remark "Today, even slightly educated fold are capable of reading and writing Fulfulde in ajami script, at least for matters of everyday life and private correspondence. The great spread of written Fulfulde does not seem to have been hindered by the abandonment of teaching Fulfulde in public schools fifteen years ago, when the government, based on the work of the Military Committee for National Recovery, give preference to French." This assessment is matched by fieldwork experience of the first author in the Futa Jalon; there, speakers of the minority language Jalonke, who are all bilingual in Jalonke and Fula, use an exographic […] writing practice and write almost exclusively in Fula Ajami. Yet this literacy is nowhere officially documented or instrumentalized in formal education or adult literacy in Guinea – almost unbelievable since it seems to be the form of literacy that is closest to UNESCO's definition of literacy as "the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts". Remarkably Ajami writing in Fula is persistent and dominant in the Futa Jalon in Guinea and in the Adamawa region of Cameroon […] Admittedly, because of its particular social context, discovering Ajami writing is not an easy task for outsiders, particularly in those of its uses that are of the foremost interest of this paper: the writing of personal letters, of grocery lists, of financial accounts, of moral, religious and social pamphlets. These domains are outside those contexts of Ajami writing that are in the scope of literary studies, and in consequence most accounts of Ajami systems focus on the historically rooted literature traditions, whose works, be they religious or poetic (two closely interlinked domains), are still widely circulated today."

The figure below was taken from Lüpke and Bao-Diop 2014, p. 13 and provides examples from Cameroon of the use of Arabic script for writing of Fula and Hausa.
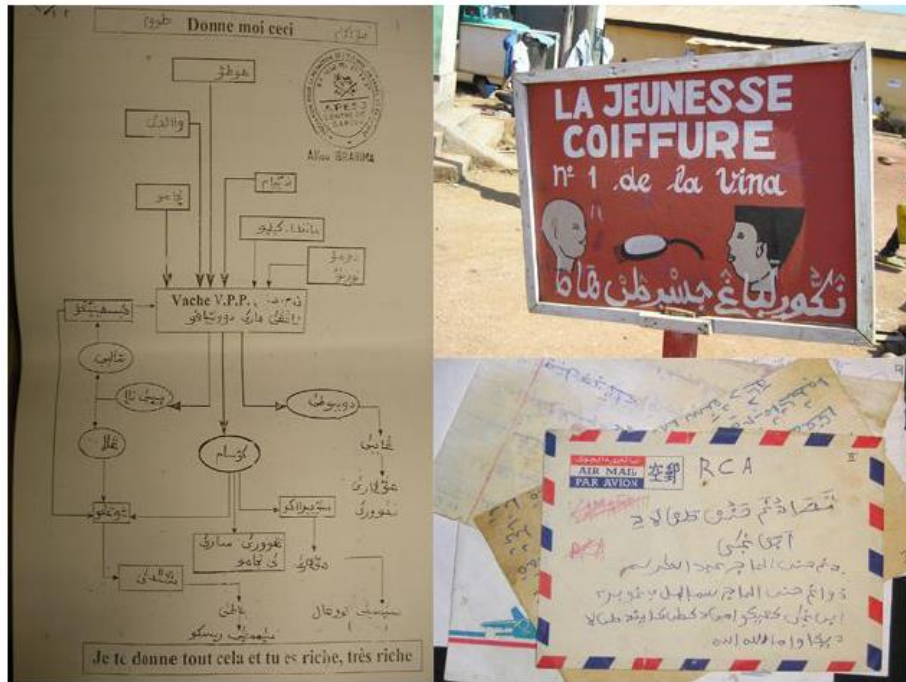
Figure 4: Use of Ajami in Cameroon © F. Lüpke 2004

The figure above shows a bilingual and biscriptual advertisement in Ngaoundéré in the top right, a diagram illustrating the many benefits of cattle in Fula Ajami (APES Garoua) on the left, and finally a collection of personal letters in Fula and Hausa Ajami from Garoua in the bottom right.

In a similar fashion Warren-Rothlin discusses in his 2014 publication "West African Scripts and Arabic-Script Orthographies in Socio-Political Context" the use of Ajami (Arabic) script for Hausa (and other languages including Fulfulde). Digraphia is also a common for these languages, as stated, which does not preclude the use of Arabic script: "A stretch of the Sahel between the Mossi and Jula in the west and the Sango and Maba in the east is home to the highest concentration of the distinct linguistic communities in Africa, dominated however by three major regional languages – Hausa, Fulfulde and Chadian Arabic. These three languages share geography, having overlapping spheres of influence; but they differ in genetics, coming as they do from three different linguistic families […] They share certain features of distribution, all having a high proportion of second-language to first-language speakers (Hausa18:25 million centered in Nigeria, Fulfulde 2:2 million in Cameroon, Chadian Arabic 2:3 million centered in Chad), and national status, being spoken by large proportions of their host countries' populations (Hausa 25% in Nigeria and 80% in Niger, Fulfulde 20% in Cameroon and 8% in Nigeria, Chadian Arabic 50% in Chad) […] In addition to the extensive regional use of Arabic itself, Arabic-script or عجمى'a'jami literary traditions existed in the 19th century for languages in all three phyla of the region […] Digraphia is a common feature of African Ajami situations, usually involving Arabic and Roman scripts, as in the three cases under discussion here. Government and NGO publications,

commercial signboards and adverts, and religious literature … may exist in both Arabic and Roman scripts […]" (pp. 261-264).

The author further notes that, "As for religious movements, Kano is home to the Arabic-script Hausa الفجر *Al-Fijir* newspaper and the vast production, especially by the Tijaniyya of small religious pamphlets in handwritten *Warš*-style, *maġribī*-script. Hausa (also Fulfulde) containing the 99 names of Allah, collections of *ḥadīṯ*, theology, history, songs, and instruction on the use of amulets. This remains a strong movement (also produced in Zaria, Sokoto and doubtless elsewhere too) can be found for sale outside almost any mosque throughout the region […]" (p. 267). The author adds that "Today, of the languages treated here, Hausa is the one most immersed in a digraphic situation, though it is perhaps surprising that even Roman-script Hausa, far predominant over *ajamin Hausa*, does not yet have an adequate standard orthography. Hausa is also the language most liable to appear in both standard Arabic-script orthographies – Hafs in the printed *Al-Fijir* newspaper (with adaptations marked by hand on the front page and only irregularly inside), and Warš in the handwritten Tijaniyya booklets, for which the orthography presented here is typical" (pp. 274-275). The following table taken from Warren-Rothlin 2014, p. 275, demonstrates the Hausa Arabic script orthography.

Table 1: Hausa consonant graphemes: Warš-based Arabic-script and Anglophone-context[24] Roman-script orthographies (multiple sources including Yahaya 1988: 31–42; Philips 2000: 19–27; Newman 2000: 729–740)[25]

Consonants	Labial	Alveolar	Palatal[26]	Velar	Glottal
Plosive					
voiceless		ت /t/ t	+ث /tʃ/ c	ک /k/ k	+ع /ʔ/ '
voiced	ب /b/ b	د /d/ d	ج /dʒ/ j	+غ /g/ g	
glottalized[27]	+ب /ɓ/ ɓ	+ط /ɗ/ ɗ	+ي /ƙ/ 'y	ف /k'/ ƙ	
Fricative					
voiceless	ڤ /f/ f	س /s/ s	ش /ʃ/ sh		ح /h/ h
voiced		ز /z/ z			
glottalized		+ظ /s'/ ts			
Nasal	م /m/ m	ن /n/ n			
Flap		ر /ɾ/ r			
Approximant		ل /l/ l	ي /j/ y	و /w/ w	

In their 2014 publication "Beneath the Surface? Contemporary Ajami Writing in West Africa, Exemplified through Wolofal", Lüpke and Bao-Diop provide insight into the use of Wolofol/Ajami (Arabic) script for Wolof: "Wolof, the de facto national language of Senegal, is one of the West African languages with a deeply rotted Ajami writing tradition. Wolof is spoken by about 80% of country's population as a language of their mostly multilingual repertoire. A formal Ajami tradition for this language is attested from the 17th century onwards […] although an earlier emergence of this script is very likely. Today, Wolofal is tightly linked to the Islamic brotherhood of Mourides. The Mourides are very influential in the Senegalese religious landscape, and the use of Wolofal for religious and poetic writing in their realm has resulted in an important body of literature, for the most part preserved in private libraries and compiled by hand" (p. 14).

Quoting McLaughlin, they stress the present-day use of Wolofal: "Wolofal originated within a religious context, but it is also fairly widely used on the contemporary scene by those who are familiar with the Arabic alphabet but not the Roman, to keep records and notes and especially to write letters. The use of Wolofal for writing Wolof appears to be much more widespread than the use of the Roman alphabet for the same purpose, a fact that is due to almost universal attendance by Muslim children on Qur'anic school, where they master the rudiments of the Arabic writing system. Public school education is conducted in French, and thus students who attend those schools learn to write in the Roman script – but attendance at such schools is not as attendance at Qur'anic schools, which was even lower in the past" (p. 14).

Lüpke and Bao-Diop continue to point out that " Texts of these two genres are in wide circulation in the entire country. While there is no publishing house producing Wolofal publications, several printers/book stores […] sell photocopies produced in large numbers of these popular texts […] Two television channels *Touba TV* and *RDV (Radio Dunyaa Vision)* have regular programmes featuring Wolofol. To these broadcasts, poets or authors of regligious and moral texts are invited to read from their work, and provide a commentary. These different media results in a great visibility of Wolofal, not just within the Mouride movement, but in the entire Senegal" (p. 19).

While discussing the non-religious contexts for use of Wolofol in Touba and beyond, the authors note that, "Given that the local government of Touba is religious, Wolofal has gained currencey as a literacy used in formal contexts there as well – a context normally occupied by French and the Latin script in the rest of Senegal. […] Official and private letters are […] an important domain for Wolofal literacy, not just in Touba but in the entire Wolof-speaking and Arabic script-writing population of the country. […] Wolofal is omnipresent in the linguistic landscape of Touba, and very visible in the entire Senegal. Figures 9 exemplify the prominence of Wolofal".

Figure 9: Ajami advertisments in Touba, Senegal © S. Bao-Diop 2009

Regarding DPLN Standard, Priest and Hosken in their 2010 document for "Proposal to add Arabic script characters for African and Asian languages" for DPLN Standard write that "The *Direction de la Promotion des Langues Nationales* (DPLN) is under the Ministry of Education of the Senegal. The DPLN worked to standardize the orthographies for the languages of Senegal. This includes languages using the Arabic script. The first version of the harmonized script (HS) was done in 1986; the latest in 1990. The DPLN charts have not been decreed by the Senegal government, because the harmonization was intended to he multi-national, and therefore a decree could not (or should not) be made by one country […] This is the official standard for the languages listed (Wolof [wol], Pulaar [fuc], Soninke [snk], Mandinka [mnk], Seereer [srr]', Joola [dyo] and Balant [bjt]). Not all of these languages are currently using the Arabic script […] There are are few *other* languages, not listed in the DPLN charts, using the Arabic script" (p. 5). It stressed again that, as pointed out by the authors, the DPLN has intentionally not been decreed, since languages affected by this standard are spoken across other countries, and therefore this should be ratified at a multi-national level. Thus, the reason for the lack of formal acceptance of the DPLN standard is not necessarily a lack of acceptance by local communities, but a need for acceptance at the political level across national boundaries. The authors also note that "Many (or most) of the characters in the Senegal standard were derived from the Islamic Educational, Scientific and Cultural Organization (ISESCO) recommendations" (p. 10). Figure 4a taken from page 23 of Priest and Hosken's 2010 proposal demonstrates those recommendations:

الأبجدية العربية الإفريقية
الحروف المفردة

EcritLAf.Ar

Column headers (left to right, handwritten language labels): Somali, Comores, Nuba, Dinka, Swahili, Yoruba, Wolof, Songhai/Zarma, Kanuri, Soso, Mandinke, Soninke, Hausa, Fulfulde, Tamachek, Arab

الرمز اللاتيني	الرمز العربي	... (African language columns) ...	العربية
a	ا		ا
b	ب		ب
p	پ		
ɓ			
t	ت		ت
t	ث		ث
c	چ		
j	ج		ج
y			
c			
h	ح		ح
kh	خ		خ
d	د		د
ɗ			
d	ذ		ذ
r	ر		ر
ɍ			
z	ز		ز
ż	ژ		
s	س		س
š	ش		ش
ṣ	ص		ص
d	ض		ض
ṭ	ط		ط
ɗ			
ẓ	ظ		ظ
c	ع		ع
gh	غ		غ
ŋ	ڠ		
f	ف		ف
v	ڤ		
ɓ			
q	ق		ق
k	ك		ك
g	گ		
ƙ			
l	ل		ل
m	م		م
n	ن		ن
ñ	ݧ		
h	ه		ه
w	و		و
y	ي		ي

Source: ISESCO à N'Djamén 2003

Ajami script as used in Chad has been standardized through the National Alphabet of Chad (ANT), and written into law by the Chadian government. As pointed out by Priest and Hosken, this is a national standard applicable to all recognized languages of Chad: "Each Chadian language is to be written using either the Latin orthography or the Arabic orthography, or possibly using both orthographies in different contexts" (p. 4). The illustration below gives the calendar from 2012, indicating current use of the script for the region (provided by Lorna Evans[11]).

The following chart given on page 283 of Warren-Rothlin's 2014 "West African Scripts and Arabic-Script Orthographies in Socio-Political Context" summarizes ANT's consonant graphemes. ANT is given in Appendix E.

---

[11] Lorna Evans is the current name of Lorna Priest.

Table 7. Alphabet national du Tchad consonant graphemes: MSA-based Arabic-script and Francophone-context Roman-script orthographies (primary source: DAPLAN Tchad 2009)

Consonants	Labial	Dental	Alveolar	Emphatic	Palatal	Velar	Pharyngeal	Glottal
Plosive								
voiceless	+ب /p/ p		ت /t/ t	(ط) /ŧ/ t	+چ /tʃ/ c	ك /k/ k		ء ئ،ؤ /ʔ/ '
voiced	ب /b/ b		د /d/ d	(ض) /ɖ/ d	ج /dʒ/ j	ؼ /g/ g		
Glottalized	+ٮ /ɓ/ +ɓ		ط /ɗ/ +ɗ		+ڃ /ʄ/ +y			
Prenasalized	+ٮ /mb/ mb		+ٮ /ⁿd/ nd		+ڃ /ⁿj/ nj	+ڠ /ⁿg/ ng		
Double Articulated								
voiceless	+ڮ /kp/ kp							
voiced	+ڲ /gb/ gb							
Fricative								
voiceless	ف /f/ f	(ث) /θ/ s	س /s/ s	(ص) /ʂ/ s	ش /ʃ/ ch	خ /x/ kh	(ح) /ħ/ h	ه /h/ h
voiced	+ۋ /v/ v	(ذ) /ð/ z	ز /z/ z	(ظ) /ðˤ/ z		(غ) /ɣ/ kh	(ع) /ʕ/ '	+ۿ /ɦ/ ɦ
Prenasalized	+ۋ /mv/ mv							
Lateral								
voiceless			+ڸ /ɬ/ sl					
voiced			+ۻ /ɮ/ zl					
Nasal	م /m/ m		ن /n/ n		+ني /ɲ/ +ny	+ڠ /ŋ/ +ng		
Flap	+ۋ /v/ vb		ر /r/ r					
Approximant			ل /l/ l		ي /j/ y	و /w/ w		

Appendix C: Example Labels from Various Languages Used for Analysis

Urdu	Arabic	Arabic	Persian	Pashto	Sindhi	Sindhi
اب	الاتحاد	أبو بکر	ابراہیم	ارزگان	اچ	ظهیر
ابراہیم	الإنسان	أحمد	احمد	استازی	اچ	عابده
اپنے	الأمر	أکتوبر	اداره	اشرف	الله ڏنو	عالیشان
احسان	الأمن	أم	آبمیوه	اعظمی	آءُ	عمر
احسن	الأول	أم کلثوم	آشپز	افغانی	بخشڻ	غلام
اختر	الأولی	أمام	آبدا	انځورگر	بصري	غلط
اردو	البلاد	أمس	آینه	اندبنمن	بهشت	غم
اعظم	التاریخ	أمینة	باران	او	بیمار	فانوس
افضل	الثانی	أیوب	بیمارستان	اوس	بدل	فرق
اقبال	الثوار	بنت	پیاز	ایمل	بک	فقیر بخش
امریکہ	الثورة	جامعة	چای	بارکزی	بڪري	قِي
امیر	الجدید	جدید	حمام	پرانسته	یت	قِرڻ
انسان	الجدیدة	جدیدة	خانه	پرون	پلو	قُوڪ
ایمان	الجزائر	حج	خبر	پسرلی	پلی ڏنو	قاصد
ایوب	الجمعة	حفصة	خداحافظ	پښتو	پریل	قافلو
آمنہ	الحکومة	حقوق	خرید	تاسی	پڳ	قطب
باتیں	الحیاة	حل	خیابان	تایید	پلو	ڪاٽو
باغ	الخارجیة	خاصة	درخت	تل	تارا	ڪٽ
بین الاقوامی	الخاصة	خالد	درود	تنی	ترڻ	ڪٽومل
پاکستان	الدکتور	خبر	دیوانه	ټپي	تپي	ڪارو
تعلیم	الدوری	خدیجة	روزنامه	ثناءالله تسل	تنویر	ڪتو
جگہ	الدول	دولة	زمین	جوړه	نڪڙ	ڪلیم
جواب	الدولة	رئیس	زہرا	چمتو	ڻونٽ	گابو
حال	الدولی	زکاة	زیان	حبیب‌الله تری	ٿابو	گڏھ
حسین	الدولیة	سبتمبر	سارا	حاخی	ٽیلهو	گھائو
حضرت	الرئیس	سنة	سامان	خُدران	ٿاندا	گھر
حقیقت	السابق	سنوات	ساناز	خُواني	ٽوپي	ڳیرو
خط	الساعة	شرکة	سبز	ځارنوالي	ثمر	ڳڻون
خوشبو	السبت	صلاة	سروش	خپل	ثواب	لال
دور	السعودیة	صیام	سلام	ختیځ	جامي	لال بخش
دیا	السوری	طرابلس	سہیل	روښانتیا	جتي	لوٽ
ذریعۂتعلیم	السیاسی	طریق	شلوغ	ریره	جهاز	ماڙي
راہ	السیاسیة	عام	شیرینی	زغرده	جهرڪي	مثل
رمضان المبارک	السید	عبد	عباس	سبا	جهیڙو	ملڪ
روز	الشباب	عبدالله	علیرضا	سختدریځي	جِي	نصیر
زبان	الشعب	عثمان	فرش	سعدالدین شپون	جڄ	نماز
زکوۃ	العالم	علی	فریدون	سلطانزوی	چریو	نند
زمین	العام	عمر	فلان	سوله	چنیسر	هاٿي
زندگی	العامة	عمل	قند	بنحُه	چِٽ	هارون

ساته	العديد	عملية	كار	بنرنه	حجم	هجرت
سب	العراق	فاعل	كاوه	بنكلى	حليم	ولړ
سعد	العراقي	فريق	كتاب	بنكول	حوالو	ولي
سفيان	العراقية	قريبة	كون	عبدالشكورررشاد	خجر	وٽ
سبيل	العربي	قوات	كيانا	عصمت قانع	خلق	ياقوت
شاعرى	العربية	كبير	گوشت	غر	خيرل	يعقوب
شخص	العمل	كوثر	گياه	غزنوي	دادو	يكو
طابره	الفريق	لبنان	لباس	غوبنتي	دال	
عائشه	القانون	لجنة	مجتبى	غونډه	دودو	
عبدالرحمن	القدم	لقاء	محمد	فراه	ډك	
عبدالرزاق	القذافي	ليبيا	محمود	گډون	ډيل	
عبدالقادر	القسم	مبتدأ	مدرسه	كابل	ډيد	
عبدالله	الكثير	متحرك	مرتضى	كله	ذاچي	
عرب	الكويت	مجلس	مرضيه	كونر	ڈيڈر	
على	اللاعب	مجموعة	مريم	لومريتوب	ذرو	
عمران	اللاعبين	محرك	مسافر	مشاعره	ذوبي	
عمل	اللجنة	محمد	مسجد	مصطفى	رابيل	
عورت	الله	مدير	مطيعالله	مقالي	راحت	
غالب	الليبي	مدينة	نان	مورب	رستو	
غلام	الماضي	مرفق	نهار	ميلمستون	زرينه	
فارسى	المباراة	مريم	همسايه	ميرويس	زمين	
قادر	المتحدة	مستوى	يحيى	نژدى	زينت	
كاركردگى	المجتمع	مسجد		نن	سك	
كام	المجلس	مصر		ننګيالى	سګ	
كتاب	المغرب	مضاف إليه		نوم	سلسلو	
لابور	المقبل	معضلة		هرات	سليم	
لوگ	المملكة	مفعول		هلالي	شاهي	
محبت	المنتخب	مفعول به		هلمند	شهباز	
محمد	الموسم	مقود		هيخ	شهر	
مسعود	النادي	مليون		وردگ	صالح	
نظر	الناس	منطقة		وكړي	صورت	
نفيس	النصر	موعد		ولسمشر	صوف	
باته	النظام	موقع			ضعيف	
بشام	الوزراء	مؤازرة			ضمير	
بم	الوطن	مؤتمر			طارق	
وقت	الوطني	نادي			طريقو	
يحيى	الوطنية	ندى			طوطو	
يہاں	الوقت	نظام			ظلم	
يونيورسٹى	اليوم	نفسه				
	يوم	نهاية				
	وزير	وزارات				
		وزارة				

Appendix D: Sample Invalid Labels for Testing WLE and Other Rules

Invalid Label	Applicable Reason	Notes
كلكتہ	WLE	1
كلكتہ	WLE	2
بكلك	WLE	1
بكلاك	WLE	2
یحیی	WLE	3
بہھحہ	WLE	4
بہھحه	WLE	5
بھحه	WLE	6
بھھح	WLE	7
بھحہب	WLE	4
بھحہب	WLE	5
ھحہب	WLE	6
ھھح	WLE	7
مكتہمكرمة	WLE	8
كیبیبا	WLE	9
لیبیبا	WLE	10
لفافہ	WLE	11
فلافِل	WLE	11
محقف	WLE	12
محقفب	WLE	12
آبنفطعجفچ	out of repertoire	آ
بنفٔطعجفچ	out of repertoire	ٔ
بنفطعڈجفچ	out of repertoire	ڈ
بنفرطعجفچ	out of repertoire	ر
بنفضطعجفچ	out of repertoire	ض
بنفطعلچجفچ	out of repertoire	لؓ
بنفطعجفچٷ	out of repertoire	ٷ
بنفطعجفچئے	out of repertoire	ۓ
بنفطعضجفچ	out of repertoire	ض

REPUBLIQUE DU TCHAD	UNITE – TRAVAIL – PROGRES	جمهورية تشاد
PRESIDENCE DE LA REPUBLIQUE	وحدة – عمل – تقدم	رئاسة الجمهورية
PRIMATURE		رئاسة الوزراء
MINISTERE DE L'ENSEIGNEMENT FONDAMENTAL ET DE L'ALPHABETISATION		وزارة التعليم الأساسي ومحو الأمية

Visa : **SGG**   **DECRET N° _____/PR/PM/MEFA/2013**
Fixant l'Alphabet National du Tchad

**LE PRESIDENT DE LA REPUBLIQUE,
CHEF DE L'ETAT,
PRESIDENT DU CONSEIL DES MINISTRES,**

Vu   la Constitution ;

Vu   la Loi N°16/PR/2006 du 13 mars 2006, portant Orientation du Système Educatif Tchadien ;

Vu   le Décret N°066/PR/2013 du 21 janvier 2013, portant nomination d'un Premier Ministre, Chef du Gouvernement ;

Vu   le Décret N°629/PR/PM/2013 du 23 juillet 2013, portant remaniement du Gouvernement ;

Vu   le Décret N°194/PR/PM/2013 du 12 mars 2013, portant structure générale du Gouvernement et attributions de ses membres ;

Vu   le Décret N°838/PR/PM/MEFA/2013 du 05 septembre 2013, portant organigramme du Ministère de l'Enseignement Fondamental et de l'Alphabétisation ;

Sur proposition du Ministre de l'Enseignement Fondamental et de l'Alphabétisation ;

Le Conseil des Ministres, entendu en sa séance du

DECRETE

Article 1 : Il est fixé en République du Tchad un Alphabet National devant servir à l'écriture des langues nationales.

Article 2 : L'Alphabet National du Tchad est en deux versions, la première comprenant les caractères latins et la seconde, les caractères sémitiques (arabes).

La version latine est composée de cinquante et une (51) lettres dont quarante-deux (42) consonnes et neuf (9) voyelles.

La version sémitique est composée de soixante (60) lettres dont cinquante et une (51) consonnes et neuf (9) voyelles.

1

Article 3 : L'Alphabet National du Tchad se présente comme suit :

A) LETTRES MINUSCULES LATINES

a   b   ɓ   c   ch   d   dr   ɗ   e   ɛ   ə   f   g
gb   h   ɦ   i   ɨ   j   k   kh   kp   l   m   mb
mv   n   ṇ   nd   ndr   ng   nj   ŋ   o   ɔ   p   r   ɽ
s   sl   t   tr   u   v   vb   w   y   ỿ   z   zl   '

B) LETTRES MAJUSCULES LATINES

A   B   Ɓ   C   CH   D   DR   Ɗ   E   Ɛ   Ə   F   G
GB   H   Ħ   I   Ɨ   J   K   KH   KP   L   M   MB
MV   N   Ṇ   ND   NDR   NG   NJ   Ŋ   O   Ɔ   P   R   Ṛ
S   SL   T   TR   U   V   VB   W   Y   Ỿ   Z   ZL   '

C) LETTRES SEMITIQUES (ARABES)

خ   ح   ج   ث   ت   پ   ب   ا

ڋ   ج   چ   ج   چ   د   ڊ   ذ   ر   ز   ڔ   ر   ز   ژ

س   ش   ص   ض   ط   ظ   ع   غ   غ   ۼ

ف   ڨ   ڤ   ك   ڮ   ق   ق   ل   ل   ڸ

م   م   مٚ   ن   ن   ه   ۂ   و   ي   ء

ٗ   ٜ   ٙ   ٛ   ٘   ۫   ۪   ٖ   َ
                        ٜ   ٟ

Article 4 : L'Alphabet National du Tchad est illustré comme suit :

A) LES CONSONNES

Nº	Caractères			Exemples			
	API	Alphabet national latin	Alphabet national arabe	Mot en caractères latins	Mot en caractères arabes	Sens	Langue
01	p	p	پَ	pa	پَ	chant	Ngambay
02	b	b	ب	baala	بَالَ	ruisseau	Dangaléat
03	t	t	ت	taara	تَارَ	impôt	Masana
04	d	d	د	dɔ	دَ	tête	Ngambay
05	tʃ	c	چ	corga	چِرْقَ	souris	Kotoko
06	dʒ	j	ج	joo	جْو	deux	Ngam
07	k	k	ك	kuma	كُمَ	Dieu	Nancere
08	g	g	ق	gaaga	قَاقَ	arbre	Bilala
09	f	f	ف	foodə	فُودّ	feu	Bilala
10	v	v	ڤ	vaŋno	ڤَعْنْ	un	Mundang
11	s	s	س	səw	سّوْ	enlever le son du mil	Gor
12	z	z	ز	zana	زَنَ	personne	Mundang
13	h	h	ه	hay	هَيْ	pagaie	Laka
14	m	m	م	manda	مَنْدَ	sel	Kanembu
15	n	n	ن	nokka	نْكَّ	lourd	Musey
16	l	l	ل	lal	لَلْ	vent	Ngambay
17	j	y	ي	yoona	يْونَ	eau	Musey
18	w	w	و	wasə	وَسّ	melon	Laka
19	r	r	ر	kərgə	كّرْقّ	route	Nancere

3

Nº	Caractères			Exemples			
	API	Alphabet national latin	Alphabet national arabe	Mot en caractères latins	Mot en caractères arabes	Sens	Langue
20	ɓ	ɓ	ب	ɓəl	بْلْ	lion	Sar
21	ɗ	ɗ	ط	ɗikna	طِكْنَ	peau (de vache)	Musey
22	ŋ	ŋ	غ	ŋora	غْرَ	dos	Masana
23	ʄ	ƴ	ج	ƴeelo	جِيلْ	haricot	Migaama
24	ɦ	ɦ	ﺋ	ɦalla	ﺋَلَّ	donner	Musey
25	ɲ	n̰	ن	n̰am	نَيْم	lait	Kanembu
26	ʔ	'	ؤ ئ إ أ ء	su'aal	سَأَلْ	question	Arabe local
27	ʃ	ch	ش	chadar	شَدَرْ	arbre	Arabe local
28	mb	mb	م	mbang	مَغْ	soleil, chef	Sar
29	mv	mv	مْ	mvɛrɛ	مّْرَ	garçon	Mundang
30	nd	nd	د	ndar	دَرْ	peau	Sar
31	ɲdʒ	nj	ج	njidi	جِدِ	arbre	Kanembu
32	ŋg	ng	غ	ngaba	غَبَ	lutte	Kanembu
33	ɬ	sl	لْ	slena	لْنَ	choses	Masana
34	ɮ	zl	لْ	zla'na	لَّأَنَ	clôture	Masana
35	kp	kp	اﭖ	kpu	اﭖُ	bois	Mundang
36	gb	gb	ق	gbajole	قَبْجُل	aider	Mundang
37	x	kh	خ	khaatim	خَاتِمْ	bague	Arabe local
38	ʋ	vb	ڤ	puvbup	يُڤُﭖ	manière de se laver	Migaama
39	ɾ	r̰	ر	ar̰um	أَرُمْ	lion	Migaama

Nº	Caractères			Exemples			
	API	Alphabet national latin	Alphabet national arabe	Mot en caractères latins	Mot en caractères arabes	Sens	Langue
40	tɽ	tr	ڗ	treŋdag	رّڠْدَق	piqûre	Maba
41	dɽ	dr	ر	drabag	رَبَق	hyène	Maba
42	ŋdɽ	ndr	ژ	ndrekee	ژّكّي	habit	Maba

La version sémitique de l'Alphabet National distingue des sons qui ne sont pas différenciés dans la version latine, à l'exemple des caractères arabes ص ث س qui sont représentés par le symbole « s » dans l'alphabet latin.

Le tableau suivant présente les consonnes propres à l'alphabet arabe du Tchad :

Nº	Caractères			Exemples			
	API	Alphabet national latin	Alphabet national arabe	Mot en caractères latins	Mot en caractères arabes	Sens	Langue
43	θ	s	ث	saabit	ثَابِتْ	stable	Arabe local
44	ħ	h	ح	humaar	حُمَارْ	âne	Arabe local
45	ð	z	ذ	zanib	ذَنِبْ	péché	Arabe local
46	sˤ	s	ص	saabuun	صَابُونْ	savon	Arabe local
47	dˤ	d	ض	darraaba	ضَرَّابَةْ	gombo	Arabe local
48	tˤ	t	ط	tayyara	طَيَّرَةْ	avion	Arabe local
49	ðˤ	z	ظ	zarf	ظَرَفْ	enveloppe	Arabe local
50	ʕ	'	ع	na'aam	نَعَامْ	autruche	Arabe local
51	ɣ	kh	غ	khanam	غَنَمْ	chèvres	Arabe local

Le chadda ّ indique une consonne longue (ou géminée).

Le sukuun ْ indique que la consonne n'est pas suivie par une voyelle.

5

B) LES VOYELLES

Nº	Caractères				Exemples			
	API	latin (brève)	latin (longue)	arabe (brève) avec ب	arabe (longue) avec ب	Mot en caractères latins	Mot en caractères arabes	Sens	Langue
01	i	i	ii	بِ	بِي	kinti	كِنْتِ	singe	Kera
02	e	e	ee	بْ	بْي	eti	إِتْ	scorpion	Gorane
03	o	o	oo	بُ	بُو	oslna	أُلْنَ	tubercule, pointes	Masana
04	a	a	aa	بَ	بَا	baa	بَا	Dieu	Tupuri
05	u	u	uu	بُ	بُو	usna	أُسْنَ	herbes	Masana
06	ɨ	ɨ	ɨɨ	بٖ	بٖي	mangɨ	مَنْق	bœuf	Bediond
07	ɛ	ɛ	ɛɛ	بٜ	بٜي	kɛrɛ	كّرّ	perdrix	Laka
08	ɔ	ɔ	ɔɔ	بۛ	بۛو	rɔme	رۛم	pardonner	Mundang
09	ə	ə	əə	بۨ	بۨا	əə	أۨ	haricot	Mundang

Article 5 : L'Alphabet National du Tchad s'applique à toutes les langues nationales et reste ouvert aux sons qui seront découverts ultérieurement.

Article 6 : Le Ministère de l'Enseignement Fondamental et de l'Alphabétisation est chargé de l'application du présent Décret qui abroge toutes dispositions antérieures contraires, sera enregistré et publié au Journal Officiel de la République.

Fait à N'Djaména, le

Par le Président de la République

premier Ministre, Chef du Gouvernement

IDRISS DEBY ITNO

Joseph DJIMRANGAR DADNADJI

Ministre de l'Enseignement Fondamental
et de l'Alphabétisation

Mme ALBATOUL ZAKARIA

Appendix F: Code Point Repertoire Excluded

All combining marks have been excluded for multiple reasons. First, they can significantly overproduce and would require additional rules to contain them effectively, complicating the design. Second, even where they are required for some languages, they are optional for others. This also circumvents the issue raised by the IAB statement, as already discussed earlier in this section.

The following code points were excluded from the code-point repertoire of LGR the reasons indicated in the rightmost column.

S.No.	Unicode	Glyph	Code Point Name	Example Languages using the Code Point (note: not an exhaustive list of languages using the code point)	EGIDS	Remarks
1.	0671	ٱ	ARABIC LETTER ALEF WASLA			Quranic Arabic
2.	0674	ء	ARABIC LETTER HIGH HAMZA	Kazakh	1 Kazakh	No evidence found for active use; visually similar to Hamza combining mark (0654)
3.	0682	݂خ	ARABIC LETTER HAH WITH TWO DOTS VERTICAL ABOVE	Pashto	1 Pashto	No evidence found for active use
4.	0690	ڐ	ARABIC LETTER DAL WITH FOUR DOTS ABOVE	Urdu	1 Urdu	Historic use only
5.	0692	ڒ	ARABIC LETTER REH WITH SMALL V	Kurdish	3 Kurdish	No evidence found for active use
6.	0694	ڔ	ARABIC LETTER REH WITH DOT BELOW	Kurdish	3 Kurdish	No evidence found for active use
7.	069B	ڛ	ARABIC LETTER SEEN WITH THREE DOTS BELOW	Harari	6a Harari	No evidence found for active use
8.	069C	ڜ	ARABIC LETTER SEEN WITH THREE DOTS BELOW AND THREE DOTS ABOVE	Moroccan Arabic		No evidence found for active use

9.	069D	صٟ	ARABIC LETTER SAD WITH TWO DOTS BELOW	Turkic		No evidence found for active use
10.	069E	ضۛ	ARABIC LETTER SAD WITH THREE DOTS ABOVE	Berber, Burushaski	6a Burushaski	No evidence found for active use
11.	06A1	ڡ	ARABIC LETTER DOTLESS FEH	Adighe		No evidence found for active use
12.	06A3	ڣ	ARABIC LETTER FEH WITH DOT BELOW	Fulfulde	3 Fulfulde	Warren-Rothlin 2014 - West African scripts and Arabic-script, pg. 285
13.	06A5	ڥ	ARABIC LETTER FEH WITH THREE DOTS BELOW			No evidence found for active use
14.	06B2	ڲ	ARABIC LETTER GAF WITH TWO DOTS BELOW			No evidence found for active use
15.	06B4	ڴ	ARABIC LETTER GAF WITH THREE DOTS ABOVE			No evidence found for active use
16.	06B6	ڶ	ARABIC LETTER LAM WITH DOT ABOVE	Kurdish	3 Kurdish	No evidence found for active use
17.	06B7	ڷ	ARABIC LETTER LAM WITH THREE DOTS ABOVE	Brahui, Chimiini , Pashai	5 Brahui	No evidence found for active use
18.	06B8	ڸ	ARABIC LETTER LAM WITH THREE DOTS BELOW			No evidence found for active use
19.	06B9	ڹ	ARABIC LETTER NOON WITH DOT BELOW	Harari, Kotokoli (Tem)	3 Kotokoli (Tem)	Although Tem is widely spoken, the data on the use of the orthography was voted by ArabicGP to not represent significant community-wide use at this time. Weathers, Andy (April 27, 2013): The rebirth of an Ajami script for the Tem in Togo. Talk held at *The*

						Arabic script in Africa: Synergies resulting from the study of a writing system. Université libre de Bruxelles (ULB) & Universität zu Köln. Bruxelles.
20.	06BF	چٜ	ARABIC LETTER TCHEH WITH DOT ABOVE		Wolof	4 Wolof	Old use, as in: Nelson, Clark & Currah 2013 : "Je sais le wolofal": http://paul-timothy.net/pages/ajamis enegal/je_sais_le_wolofal _06-jan-2013_a4.pdf In a presentation by Lameen Souag, which was published in Afrikanistik 22Online (note however the incorrect encoding in the online publication), the author gives examples of the use of this codepoint in Wolof (Examples are undated or from 1962) and in Fulfulde (Examples date from 1885 and 1935). The citation from 1962 recent enough, but in the light of conflicting evidence it is omitted

21.	06C5	و	ARABIC LETTER KIRGHIZ OE	Kirghiz		No evidence found for active use
22.	06C7	ۇ	ARABIC LETTER U	Kirghiz, Azerbaijani		No evidence found for active use
23.	06C8	ۈ	ARABIC LETTER YU	Uighur		No evidence found for active use
24.	06C9	ۉ	ARABIC LETTER KIRGHIZ YU	Kazakh, Kirghiz		No evidence found for active use
25.	06CA	ۊ	ARABIC LETTER WAW WITH TWO DOTS ABOVE	Kurdish	3 Kurdish	Sorani Kurdish https://en.wikipedia.org/wiki/Kurdish_alphabets; widespread use could not be determined
26.	06D3	ۓ	ARABIC LETTER YEH BARREE WITH HAMZA ABOVE	Urdu	1 Urdu	No evidence found for active use
27.	06EE	ۮ	ARABIC LETTER DAL WITH INVERTED V			No evidence found for active use
28.	06EF	ۯ	ARABIC LETTER REH WITH INVERTED V			No evidence found for active use
29.	06FA	ۺ	ARABIC LETTER SHEEN WITH DOT BELOW	Hausa	2 Hausa	Historic use; p. 276, Warren-Rothlin, Andy (2014): West African scripts and Arabic-script orthographies in socio-political context. In Meikal Mumin, Kees (C.) H. Versteegh (Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in

						Semitic Languages and Linguistics, 71), pp. 261–289.
30.	06FB	ضْ	ARABIC LETTER DAD WITH DOT BELOW			No evidence found for active use
31.	06FC	غْ	ARABIC LETTER GHAIN WITH DOT BELOW	Tem (Kotokoli), Togo		Weathers, Andy (April 27, 2013): The rebirth of an Ajami script for the Tem in Togo. Talk held at *The Arabic script in Africa: Synergies resulting from the study of a writing system*. Université libre de Bruxelles (ULB) & Universität zu Köln. Bruxelles.
32.	06FF	ۿ	ARABIC LETTER HEH WITH INVERTED V			No evidence found for active use
33.	0750	�	ARABIC LETTER BEH WITH THREE DOTS HORIZONTALLY BELOW			No evidence found for active use
34.	0753	ﺕ	ARABIC LETTER BEH WITH THREE DOTS POINTING UPWARDS BELOW AND TWO DOTS ABOVE			No evidence found for active use
35.	0754	ﻦ	ARABIC LETTER BEH WITH TWO DOTS BELOW AND DOT ABOVE	Bamana	4 Bamana	Vydrin, Valentin Feodos'evich; Dumestre, Gérard (2014): Manding Ajami samples. Mandinka and Bamana. In Meikal

						Mumin, Kees (C.) H. Versteegh (Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in Semitic Languages and Linguistics, 71), pp. 225–260.
36.	0755	بٚ	ARABIC LETTER BEH WITH INVERTED SMALL V BELOW	Serer, Senegal	5 Serer	No evidence found for active use
37.	0757	ختּ	ARABIC LETTER HAH WITH TWO DOTS ABOVE			No evidence found for active use
38.	0758	چ	ARABIC LETTER HAH WITH THREE DOTS POINTING UPWARDS BELOW	Harari	6a Harari	No evidence found for active use
39.	0759	ڙ	ARABIC LETTER DAL WITH TWO DOTS VERTICALLY BELOW AND SMALL TAH	Saraiki	5 Saraiki	No evidence found for active use
40.	075A	دٚ	ARABIC LETTER DAL WITH INVERTED SMALL V BELOW			No evidence found for active use
41.	075B	ڔ	ARABIC LETTER REH WITH STROKE			No evidence found for active use
42.	075C	ڜ	ARABIC LETTER SEEN WITH FOUR DOTS ABOVE	Shina, Torwali	6a Shina	No evidence found for active use
43.	075D	ڝ	ARABIC LETTER AIN WITH TWO DOTS ABOVE			No evidence found for active use

44.	075E	غ	ARABIC LETTER AIN WITH THREE DOTS POINTING DOWNWARDS ABOVE			No evidence found for active use
45.	075F	غ	ARABIC LETTER AIN WITH TWO DOTS VERTICALLY ABOVE			No evidence found for active use
46.	0761	ڡ	ARABIC LETTER FEH WITH THREE DOTS POINTING UPWARDS BELOW			No evidence found for active use
47.	0764	ڤ	ARABIC LETTER KEHEH WITH THREE DOTS POINTING UPWARDS BELOW			No evidence found for active use
48.	0765	م	ARABIC LETTER MEEM WITH DOT ABOVE			No evidence found for active use
49.	0769	ن	ARABIC LETTER NOON WITH SMALL V	Gujari	5 Gujari	No evidence found for active use
50.	076B	ژ	ARABIC LETTER REH WITH TWO DOTS VERTICALLY ABOVE	Ormuri	6a Ormuri	No evidence found for active use
51.	076C	ؤ	ARABIC LETTER REH WITH HAMZA ABOVE	Ormuri	6a Ormuri	No evidence found for active use
52.	076D	ش	ARABIC LETTER SEEN WITH TWO DOTS VERTICALLY ABOVE	Ormuri, Gawri	6a Ormuri	No evidence found for active use
53.	0772	ح	ARABIC LETTER HAH WITH SMALL ARABIC LETTER TAH ABOVE	Torwali	6a Torwali	No evidence found for active use
54.	0773	ا	ARABIC LETTER ALEF WITH EXTENDED ARABIC-INDIC DIGIT TWO ABOVE	Burushaski	6a Burushaski	No evidence found for active use
55.	0774	ا	ARABIC LETTER ALEF WITH EXTENDED ARABIC-INDIC DIGIT THREE ABOVE	Burushaski	6a Burushaski	No evidence found for active use
56.	0775	ى	ARABIC LETTER FARSI YEH WITH EXTENDED ARABIC-INDIC DIGIT TWO ABOVE	Burushaski	6a Burushaski	No evidence found for active use

57.	0776	ئ	ARABIC LETTER FARSI YEH WITH EXTENDED ARABIC-INDIC DIGIT THREE ABOVE	Burushaski	6a Burushaski	No evidence found for active use
58.	0777	ي	ARABIC LETTER FARSI YEH WITH EXTENDED ARABIC-INDIC DIGIT FOUR BELOW	Burushaski	6a Burushaski	No evidence found for active use
59.	0778	ۇ	ARABIC LETTER WAW WITH EXTENDED ARABIC-INDIC DIGIT TWO ABOVE	Burushaski	6a Burushaski	No evidence found for active use
60.	0779	ۉ	ARABIC LETTER WAW WITH EXTENDED ARABIC-INDIC DIGIT THREE ABOVE	Burushaski	6a Burushaski	No evidence found for active use
61.	077A	ﮮ	ARABIC LETTER YEH BARREE WITH EXTENDED ARABIC-INDIC DIGIT TWO ABOVE	Burushaski	6a Burushaski	No evidence found for active use
62.	077B	ﮯ	ARABIC LETTER YEH BARREE WITH EXTENDED ARABIC-INDIC DIGIT THREE ABOVE	Burushaski	6a Burushaski	No evidence found for active use
63.	077C	ڇ	ARABIC LETTER HAH WITH EXTENDED ARABIC-INDIC DIGIT FOUR BELOW	Burushaski	6a Burushaski	No evidence found for active use
64.	077D	ﺲ	ARABIC LETTER SEEN WITH EXTENDED ARABIC-INDIC DIGIT FOUR ABOVE	Burushaski	6a Burushaski	No evidence found for active use
65.	08AA	ﺭ	ARABIC LETTER REH WITH LOOP			No evidence found for active use
66.	08AB	و	ARABIC LETTER WAW WITH DOT WITHIN			No evidence found for active use
67.	08AC	ﻴ	ARABIC LETTER ROHINGYA YEH			No evidence found for active use

Appendix G: Initial Analysis of Variant Mappings and Types

The following analysis shows the initial agreement of the TF-AIDN. Each table represents the set of code points from an earlier version of the repertoire, which form a variant set. The tables show the form of each of the characters in all the four forms (initial, medial, final and isolated). The final column states the applicable principle for which the code point is included in the variant set.

All applicable forms of a character against each Unicode code point are given (in combination with letter ب). Further, the principle(s) motivating the variant from Section 2.3 are listed in the final column for each table.

For interpreting the presentation of the variant types in the tables below, note that:

- Where variant type is same in both directions, 'a/a' is written as 'a' and 'b/b' is written as 'b'
- Where variant type is not same in both directions, the type is given as 'a/b' or 'b/a' and further specified in the final column.

Furthermore, in the tables for variant types, alternate sets of rows for a code point in the first column have been shaded for better readability (and therefore should be treated as a formatting artifact only).

Table 1: YEH Class

Unicode Code Point	Isolated Form	Initial Form	Final Form	Medial Form	Applicable Principle No.
0649	ى	بـ	بى	ببـ	1
06CC	ى	يـ	بى	بيـ	1
064A	ي	يـ	بي	بيـ	1
06D0	ې	بـ	بې	ببـ	4
067B	ݕ	بـ	بݕ	ببـ	4
06CD	ؽ		بؽ		3
06D2	ے		بے		2 (06CC), 5 (064A)
0626	ئ	ئـ	بئ	بئـ	3

Code Point	Code Point	Type (blocked/ allocatable)	Notes
0649	06CC	b	
0649	064A	b	
0649	06D0	b	

0649	067B	b	
0649	06CD	b	
0649	06D2	b	
0649	0626	b/a	0649 ->b 0626 (blocked), 0626 ->a 0649 (allocatable)
06CC	064A	a (in both directions)	The two are visually identical and the same label could be typed using one or the other based on the set language settings and keyboard layout of a user. E.g. ccTLD labels ایران (Iran) and ملیسیا (Malaysia) will be written with 06CCor 064A depending on the country – both versions are identical need to be available to the entire community.
06CC	06D0	b	
06CC	067B	b	
06CC	06CD	b	
06CC	06D2	b	
06CC	0626	b/a	06CC ->b 0626 (blocked), 0626 ->a 06CC (allocatable)
064A	06D0	b	
064A	067B	b	
064A	06CD	b	
064A	06D2	b	
064A	0626	b/a	064A ->b 0626 (blocked), 0626 ->a 064A (allocatable)
06D0	0626	b	
06D0	067B	b	
06D0	06CD	b	
06D0	06D2	b	
067B	06CD	b	
067B	06D2	b	
067B	0626	b	
06CD	06D2	b	
06CD	0626	b	
0626	06D2	b	

Table 2: HEH Class

06C1	ه	ـهـب	بـ	بـهـب	1
06D5	ه		به		1
0647	ه	هب	به	بهب	1
06BE	ه	هب	به	بهب	1
0629	ة		بة		3
06C3	ة		بة		3 (with Heh) 1 (with 0629)

06C0	ۀ		بهٔ		3 1 (06C2)
06C2	ۂ		بۂ		3 1 (06C0)

06C1	06D5	b	
06C1	0647	a	Labels in Arabic language using 0647 in the final and isolated positions, 06C1 in other languages (Urdu, Pashto, Saraiki, etc.), the latter being identical with the in final and isolated positions.
06C1	06BE	b	
06C1	0629	b	
06C1	06C3	b	
06C1	06C0	b	
06C1	06C2	b/a	06C1 ->b 06C2 (blocked), 06C2 ->a 06C1 (allocatable)
06D5	0647	a	Labels in Arabic language using 0647 in the final and isolated positions will be typed in other languages using 06D5 (Uyghur, Kazakh, Kyrgyz) which is identical in isolated and has a variant glyph or identical glyph form in final position.
06D5	06BE	b	
06D5	0629	b	
06D5	06C3	b	
06D5	06C0	b/a	06D5 ->b 06C0 (blocked), 06C0 ->a 06D5 (allocatable)
06D5	06C2	b	
0647	06BE	b	
0647	0629	a	In Arabic language, Heh can be substituted by Teh Marbuta; RFC6365
0647	06C3	b	
0647	06C0	b	
0647	06C2	b	
06BE	0629	b	
06BE	06C3	b	
06BE	06C0	b	
06BE	06C2	b	
06C3	0629	a	The two are visually similar or identical and same label could be typed using one or the other based on which country one resides and which keyboard one uses. E.g. ccTLD labels السعودية (Saudi Arabia) will be written with 0629 or 06C3 depending on the country. Labels, which in final and isolated positions employ 0629 in Arabic language will be spelled using 06C3 in other languages (Urdu, etc.) . The latter is identical in isolated positions

			and additionally has a variant glyph or identical glyph form in final positions.
06C3	06C0	b	
06C3	06C2	b	
0629	06C0	b	
0629	06C2	b	
06C0	06C2	b	

Table 3: KAF Class

0643	ك	كب	بك	بكب	1
06A9	ک	کب	بک	بکب	1
06AA	ڪ	ڪب	بڪ	بڪب	2, 5 (0643)

0643	06A9	a	The two have identical shapes in initial and medial positions and are used by different language communities to refer to the same letter e.g. common words like كتاب (book) will be written with 0643 in Arabic language but with 06A9 in Urdu, Persian, etc.
0643	06AA	a	The two have similar (interchangeable) shapes in initial and medial positions and are used by different language communities to refer to the same letter e.g. common words like ڪتاب (book) will be written with 0643 in Arabic language but with 06AA in Sindhi. Similarly, 06AA is used in lieu of 0643 in Western (African) orthographies.
06A9	06AA	a	The two have similar (interchangeable) shapes in initial and medial positions and are used by different language communities to refer to the same letter e.g. common words like ڪتاب (book) will be written with 06A9 in Urdu, Persian, etc. languages but with 06AA in Sindhi. These are used as distinct characters in Sindhi.

Table 4: PEH Class

067E	پ	پب	بپ	بپب	1
0752	ݒ	ݒب	بݒ	بݒب	4
06D1	ۑ	ۑب	بۑ	بۑب	1
06BD	ڽ	ڽب	بڽ	بڽب	1

067E	0752	b	
067E	06D1	b	
067E	06BD	b	
0752	06D1	b	

0752	06BD	b	
06D1	06BD	b	

Table 5: NOON WITH TWO DOTS BELOW Class

0754	ڔ	نب	بڔ	بينب	1
0767	ڧ	نب	بن	بينب	1
08A9	ݩ	ب؟	؟ب	ب؟ب	1

0754	0767	b	
0754	08A9	b	
0767	08A9	b	

Table 6: FEH Class

0641	ف	فب	بف	بفب	1 (06A7)
06A2	ڢ	ڢب	بڢ	بڢب	5 (0641)
06A7	ڧ	ڧب	بڧ	بڧب	5 (0642)
0642	ق	قب	بق	بقب	6 (06A7)

0641	06A2	a	Used interchangeably in Africa for languages using Western (African) orthography
0641	06A7	b	
0641	0642	b	
06A2	06A7	b	
06A2	0642	b	
06A7	0642	a	Used interchangeably in Africa for languages using Western (African)orthography

Table 7: TTHEH Class

0679	ٹ	ٹب	بٹ	بٹب	1
06BB	ڻ	ڻب	بڻ	بڻب	1

0679	06BB	b	

Table 8: NOON Class

0646	ن	نب	بن	بنب	1
06BA	ں	ںب	بں	بنب	1, 5 (0646)

0646	06BA	a	Used interchangeably in Africa for languages using Western (African)orthography. The two have exactly same shape in initial and medial positions.

Table 9: TEH Class

062A	ت	تب	بت	بتب	4
067A	ٿ	ٿب	بٿ	بٿب	4

062A	067A	b	

Table 10: NYEH Class

0683	چ	چب	بچ	بچب	4
0684	ڄ	ڄب	بڄ	بڄب	4

0683	0684	b	

Table 11: VEH Class

06A4	ڤ	ڤب	بڤ	بڤب	1
06A8	ڨ	ڨب	بڨ	بڨب	1

06A4	06A8	b	

Table 12: THEH Class

062B	ث	ثب	بث	بثب	4
067D	ٽ	ٽب	بٽ	بٽب	4

062B	067D	b	

Table 13: NG Class

06AD	ڭ	ڭب	بڭ	بڭب	1
0763	ڣ	ڣب	بڣ	بڣب	1

06AD	0763	b	

Table 14: DUL Class

068E	ڎ		بڎ		4
068F	ڏ		بڏ		4

068E	068F	b	

Table 15: ALEF with HAMZA Class[12]

0623	أ		بأ		3
0622	آ		بآ		3
0625	إ		بإ		3
0627	ا		با		3
0672	ٲ		بٲ		3

0623	0627	a	Alef with Hamza above used interchangeably with Alef in Arabic language; الأردن and الاردن (Jordan)
0623	0622	a	Data from second level registrations by Saudi NIC shows the end-users use the two interchangeably, e.g. registrants registering "Adam" as أدم and آدم
0623	0625	b	
0623	0672	b	
0627	0622	a	Alef with Madda above used interchangeably with Alef in Arabic language
0627	0625	a	Alef with Hamza below used interchangeably with Alef in Arabic language; إمارات and امارات (United Arab Emirates; also the ccTLD label)
0627	0672	a	May be used interchangeably with Alef in Kashmiri language
0622	0625	b	
0622	0672	b	
0625	0672	b	

Table 16: Waw Class

0648	و		بو		3
0624	ؤ		بؤ		3

0624	0648	a/b	0624 ->a 0648 (allocatable), 0648 ->b 0624 (blocked)

Appendix H: Members of the Task Force on Arabic Script IDNs (TF-AIDN)

The table below lists the members of TF-AIDN, associated country and the languages (using Arabic script) which they speak or are actively involved in for research purposes.

S.No.	Name	Country	Language(s)
1	Abdelaziz Hilali	Morocco	Arabic
2	Abdeslam Nasri	Algeria / France	Arabic
3	AbdulRahman I. Al-Ghadir	Saudi Arabia	Arabic
4	Adel M. Riyad	Egypt	Arabic
5	Ahmed Bakhat Masood	Pakistan	Urdu
6	Allan Ghazi Salahedi	Palestine	Arabic
7	Ammar Hussain Jaffri	Pakistan	Urdu
8	Andrea Feltner	USA	
9	Arash Naderpour	Iran	Persian
10	Behnam Esfahbod	Iran / USA	Persian
11	Emily Taylor	UK	
12	Hago Elteraifi Dafalla	Sudan	Arabic
13	Hania Sabbidin-Dimassi	Lebanon	Arabic
14	Hazem Hezzah	Egypt	Arabic
15	Inam Ullah	Pakistan	Pashto, Torwali, Urdu; Various languages in South Asia
16	James Mitchell	Australia	Arabic
17	Kemal Abdulwehab	Ethiopia	Arabic, Amharic, Tigrigna
18	Mahmoud Lattouf	Jordan	Arabic
19	Meikal Mumin	Germany	Various African languages
20	Mohammad Ismail Nofal	Jordan	Arabic
21	Nabil Benamar	Morocco	Arabic
22	Nadhem Al-Fardan	Saudi Arabia	Arabic
23	Nasser Kettani	Morocco	Arabic

24	Noha Fathy Mohamed	Egypt	Arabic
25	Rinalia Abdul Rahim	Malaysia	Malay
26	Said Zazai	Afghanistan	Pashto, Dari
27	Salman Alqais	Bahrain	Arabic
28	Sufyan Aqtam	Palestine	Arabic
29	Syed Iftikhar Shah	Pakistan	Seraiki, Urdu
30	Tarik Merghani	Sudan	Arabic
31	Tariq Rahim Soomro	Pakistan / UAE	Sindhi, Urdu
32	Wael Nasr	Bahrain	Arabic
33	Zain Al-Abdeen Baig	UAE	Urdu

---

[13] Contributed to TF-AIDN as a member of the community before joining the staff.