Proposal for
Generation Panel for Arabic Script
Label Generation Ruleset for the Root Zone

Task Force on Arabic Script Internationalized Domain Names (TF-AIDN)
Middle East Strategy Working Group (MESWG)

1. General Information

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 considered mandatory (meaning omitting them is considered a spelling error). Further, some diacritical marks are also used to specify consonantal sounds, e.g. use of SHADDA diacritic for gemination of consonants (e.g. see [1] for relationship between script and phonetics of Urdu).

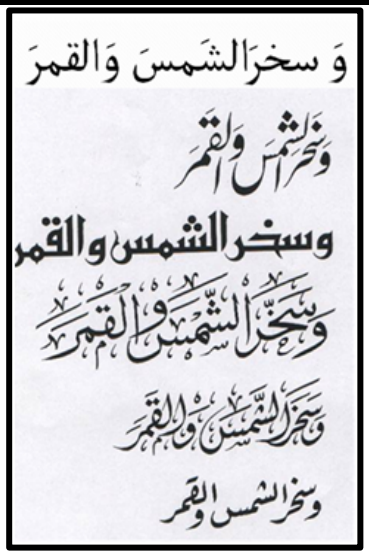


Figure 1: Arabic Writing Styles

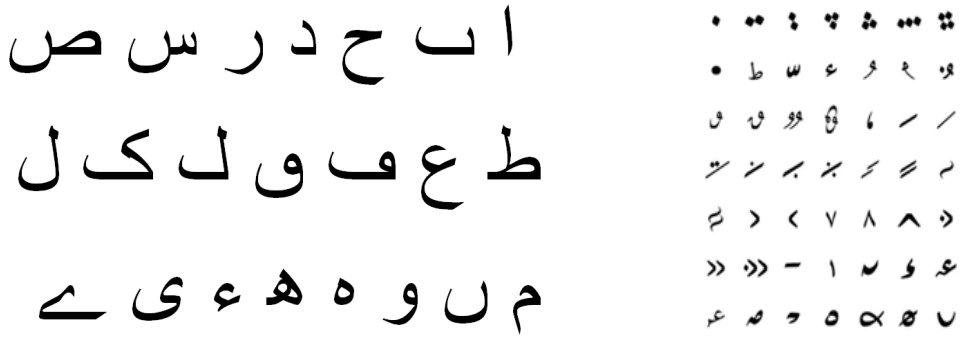
In addition to the basic letters and combining marks, the script also has a set of digits, punctuation marks, and other symbols (both combining marks and letters). Unicode standard encodes two sets of digits for Arabic script, one set for Arabic language and another set for other languages. In addition to encoding, languages also differ in shape of some digits (e.g. digits 4, 6, 7). This latter difference is generally handled in 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 generally considered calligraphic traditions, with different glyphs being considered equivalent for Arabic language, but may be distinctive in other languages. In the printing industry and computer systems, Arabic and Persian are mainly written in Naskh calligraphic style; but Urdu, sometimes Persian, and many other languages in South Asia are written in Nastalique calligraphic style.

Arabic script is cursive and therefore its letters are generally written connected with one another. It has two types of letters: one sub-set joining with only the letters before them, and another sub-set joining to both the adjoining letters. The cursive joining feature is a basic requirement of the script, with joined letters forming a ligature. One or more ligatures are juxtaposed to form words. Space is not put

between words to separate them in calligraphic writing and is also not regularly used in typing text using 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 [8]. The cursive behavior can vary with style of writing (e.g. see [2] for Nastalique) as well as, in few cases, with the writing tradition for particular languages.

Arabic language was initially written with base letters called *Rasm*. These letters under-represented the sounds, where the exact pronunciation had to be deciphered through context by the reader. To facilitate reading, especially by those who did not speak Arabic as their first language, the script was extended by adding two kinds of marks: mandatory *Ijam* (marks composed with *Rasm* to fully specify the consonants) and optional *Tashkil* (marks for other purposes, including *Harakat* for specifying vowels). The process has been used since the 7th century to extend the 17 letter shapes in Arabic language to the 28 letters today, now considered as its basic set (see also a longer list and corresponding discussion in [3]). Over past centuries, Arabic script has been adopted by many other languages with different phonetic requirements. Therefore, the script has been extended to include additional letters. The letter formation process in the script has been mostly by using the same base forms (*Rasm*), composing one or more *Ijam* and/or changing their orientation. For example, Persian adds some characters by using THREE DOTS with existing shapes; Urdu uses the extension by Persian language and further adds letters with SMALL HIGH TAH combining mark; Seraiki uses a combination of SMALL HIGH TAH and ONE DOT ABOVE; etc. In some languages, the base form (*Rasm*) is also modified, e.g. Pashto adds a small RING to some base forms and a small TAIL to a base form in one case. Some base letter shapes (*Rasm*), consonant letter forming marks (*Ijam*) and other diacritics (*Tashkil*; which include the vowel diacritics or *Harakat*) are illustrated in Figure 2 below. The composed *Rasm* and *Ijam* are now perceived as a single consonantal letters by native users of the script (i.e. the two components are not considered independently) and Unicode Standard **[4]** also encodes this composed form as a character.



(a) (b)

Figure 2. Examples of Arabic Script: (a) Base Forms (*Rasm*), and (b) Diacritical Marks (*Ijam* and *Tashkil* (including *Harakat*))

Arabic script is written horizontally, from right to left, but digits are written from left to right, making the script bi-directional. Therefore, based on the Unicode standard, implementations of the Arabic script formatting should conform to the Unicode Bidirectional Algorithm [4].

Beyond the horizontal linearity in writing the letters, there is also super- and sub-scripting of Tashkil in Arabic script. There is also variation in this super- and sub-scripting as well as in basic letter shapes (e.g. FEH, GAF, and NOON due to regional differences between *Mashriqi/Ḥafṣ* or *Maghribi/Warš* varieties).

1.1 Target Script for the Proposed Generation Panel

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 [4]:

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

These do not include the Unicode ranges for the Arabic Presentation Forms, which are DISALLOWED by IDNA 2008 protocol, except Arabic Tail Fragment (U+FE73) that is PVALID and therefore will be considered. In addition, two code points Zero Width Non-Joiner (U+200C) and Zero Width Joiner (U+200D), which are used to manage the shaping of Arabic letters (and therefore word boundaries in case of multiple words written together in a label), will also be considered because they have been discussed in the Arabic Script Case Study.

There may be other characters in Arabic script which may lack encoding in the current version of Unicode (and which may be encoded in later versions), e.g. see [10]. Such characters, which are known and documented, will remain relevant for the analysis.

For Internationalized Domain Names (IDNs) in Arabic script, these ranges will be constrained by many factors. These include first and foremost the rules defined in RFCs 5892, limiting the ranges to code points that are PVALID and CONTEXT O/J and those included (or excluded) based on the exception list provided in the RFC. Further constraints on the use of the script within IDNs are given in RFC 5893. Additionally, the character set may be further limited to letters in contemporary use by languages. The letters or marks which are not in current use but are encoded to document historical texts or those not used in normal writing but used to document religious texts could be excluded. There are further restrictions for the root zone suggested in the IAB Statement: The interpretation of the rules in ICANN gTLD Applicant Guidebook[1] and the follow-up discussion[2]. Based on these and other factors there may be restrictions imposed by the Maximal Starting Repertoire (MSR) to be issued by the Integration Panel (IP). The final script set for discussion towards inclusion in the LGR for the root zone will be devised in the context of these restrictions.

[1] See http://www.iab.org/documents/correspondence-reports-documents/2012-2/iab-statement-the-interpretation-of-rules-in-the-icann-gtld-applicant-guidebook/

[2] See https://www.iab.org/documents/correspondence-reports-documents/2012-2/response-to-icann-questions-concerning-the-interpretation-of-rules-in-the-icann-gtld-applicant-guidebook/

1.2 Principal Languages using the Script

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 and North Africa:** Modern Standard Arabic and the local variants of the Arabic language, used mainly by the Arab countries.
- **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:** Some other languages of North and sub-Saharan Africa use Arabic script (along with Latin script). The extended Arabic script used by some of these languages is commonly known as the Ajami script. Examples include Swahili, Fula and Mandinka and others.
- **South-East Asia:** In South-East Asia, Jawi is the Arabic alphabet for writing the Malay, Acehnese, Banjarese, Minangkabau, Tausug and several other regional languages. Jawi is one of the two official scripts in Brunei, and is used as an alternate script in Malaysia. Usage wise, Jawi is used 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.

Arabic script has had significant historical use. The historical use of the script is not directly relevant for devising the LGR for the Root Zone as the LGR will focus on significant contemporary use of the script. However, it may still be important to identify characters encoded in the Unicode standard for historical reasons and not in contemporary use.

Arabic script has been used in East Asia, such as China, as well as numerous countries of Central Asia, such as Kyrgyzstan, Turkmenistan, Uzbekistan, and Azerbaijan. It is difficult to estimate the extent of the current use of the script due to language policies of current states in these areas.

In Europe, Arabic script has historically been used in countries with significant Muslim populations, such as Albania, Bosnia, Belarus, and Turkey. Most historic traditions have fallen into disuse. We have incomplete knowledge about Eastern European writing traditions, particularly due to language policies in these countries which were part or closely associated with the former Soviet Union. It is also unclear how much revival there could be of Arabic script writing traditions for non-European languages in Europe by recent immigrants, particularly from Africa.

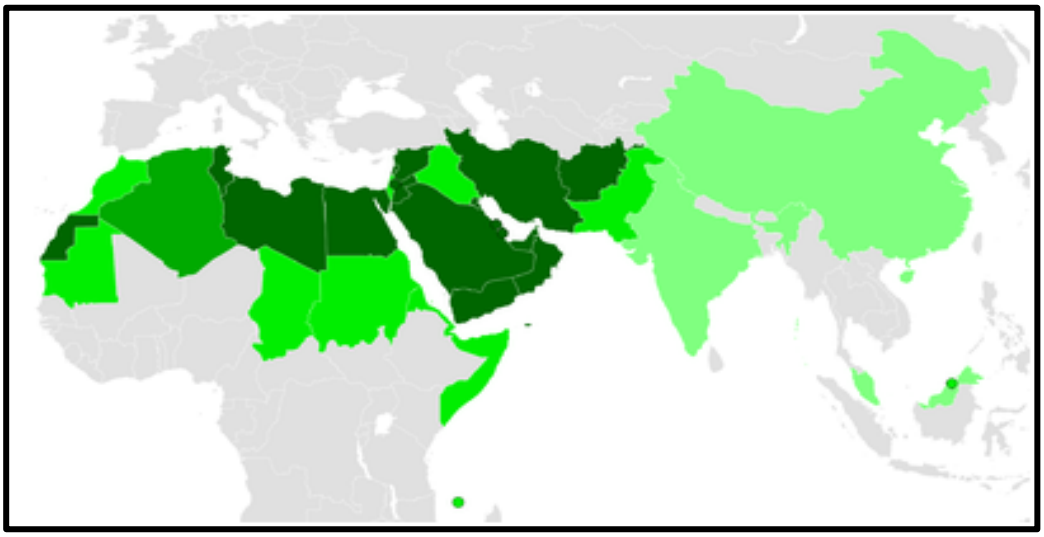
In Africa, there have been numerous writing traditions, usually relating to dominant or influential Muslim strata of societies, before European colonialism. The most recent estimates identify 80 to 95 historical or contemporary Arabic scriptwriting traditions of African languages in addition to Arabic

language [8]. With the recent increasing Latin script based language policies and literacy campaigns, the use of Arabic script in most African countries has fallen into disuse, particularly where Arabic is not generally the first language. Where it has remained, use is usually restricted to personal, religious, and cultural domains. Estimates of current use are difficult due to language policies as well as the 'insider character' of what remains of writing traditions. There is a certain degree of Arabic script usage revival, particularly due to the impetus from the Bible translation communities (with Arabic script being perceived as a generally prestigious script for such documents), as well as due to increasing funding by Arab countries of non-western primary educational systems.

In South and Middle America, there have been historical cases of the use of Arabic script, which are mostly associated to slavery. It is unclear but improbable that any modern-day writing traditions remain.

1.3 Countries with Significant User Communities for the Script

Arabic script is used for writing a diverse set of languages spoken across the Middle East, North Africa, Sub-Saharan Africa, South Asia, West Asia and East Asia. Some major countries with use of Arabic script are depicted in the map [5]. See Appendix A for a more detailed overview of usage of the script based on languages.



	is the only official orthography
	is the only official orthography, but others are recognized for national or regional languages
	is official alongside other orthographies
	is official at a sub-national level (China, India) or is a recognized alternative orthography (Malaysia)

1.4 Related Scripts

There are different ways Arabic script may be related to other scripts. As discussed, ancestry of the Arabic script includes Aramaic, Syriac and Nabataean scripts. Nko script has been claimed to be based on Arabic script. Thaana or Dhivehi script of the Maldives is derived from Arabic script. Further African scripts, such as Vai, might have borrowed from Arabic script, but are not based on it in such a sense.

Arabic script has been written in multiple calligraphic styles, which show significant variety in mutual intelligibility: a simple Kufi style evolved into a more complex Naskh style for the Arabic language. Subsequent evolution of Nastalique and its use in Persian and South Asian languages is significant, to the extent that sometimes the styles are not mutually intelligible. Examples of these writing styles are given earlier in this document. For the purposes of the Arabic Script Label Generation Ruleset for the Root Zone, these may not be considered as different scripts but as a variation in writing style within the script.

2. Proposed Initial Composition of the Panel

The Middle East Strategy Working Group (MESWG) is a community-based working group coordinated by ICANN to promote the DNS Ecosystem in the Middle East, North Africa as well as Afghanistan, Iran and Pakistan. In this regard, MESWG is in the process of implementing an action plan based on its Middle East Strategy Document in order to perform technical, training and dissemination activities. Details are available at https://community.icann.org/display/MES/Middle+East+Working+Group. As part of its activities, MESWG has formed the Task Force on Arabic Script Internationalized Domain Names (TF-AIDN). TF-AIDN will focus on technical issues and solutions to promote Arabic IDNs' definition, secure deployment and ease of use for the community. Taking a holistic approach, the current technical matters include the following:

- Arabic Script Label Generation Ruleset (LGR) for the Root Zone
- Second level LGRs for the Arabic script
- Arabic script Internationalized Registration Data Protocol and Practice
- Universal acceptability of Arabic script IDNs and variants
- Technical challenges around registration of Arabic IDNs and variants
- Operational software for Arabic script IDN registry and registrar operations
- DNS security matters specifically related to Arabic IDNs and variants
- Technical training material around Arabic script IDNs

One of the first tasks being undertaken by the TF-AIDN is to act as a Generation Panel for the Arabic script to define the Arabic Script Label Generation Ruleset (LGR) for the Root Zone. All members of the TF-AIDN will participate as members of the Generation Panel for Arabic Script LGR.

MESWG, the parent group of TF-AIDN, is responsible for the process of forming the TF-AIDN and its oversight. This includes defining the processes of its operations. However, TF-AIDN is independent of MESWG for the technical decision-making. The TF-AIDN is open in its membership. The membership process is as follows:

Anybody who is interested to join the task force on a voluntary basis can apply at any time by emailing a CV and Statement of Interest (SOI) to tf-aidn-admin@meswg.org. The CV and SOI are forwarded to the MESWG. MESWG members discuss the merits of each applicant, on a case by case basis, to decide if the applicant is to be invited as a member of TF-AIDN. The primary criteria used in these discussions include (but are not limited to) the current membership and expertise of TF-AIDN and what additional expertise the applicant is adding to the current pool. Linguistic and geographical diversity is also taken into account.

2.1 Panel Chair and Members (with Expertise)

The current membership of TF-AIDN, as recommended by MESWG, includes the following (in alphabetical order). All the members of TF-AIDN will be members of the Generation Panel for Arabic script LGR for the Root zone.

No.	Name	Designation	Organization	Country	Language Expertise
1	Abdelaziz Hilali	Professor (and Chair of ISOC Morocco)	Moroccan Telecommunications graduate institution (INPT)	Morocco	Arabic
2	AbdulRahman I. Al-Ghadir	Researcher	Saudi NIC	Saudi Arabia	Arabic
3	Adel M. Riyad	Senior System Administrator	NTRA	Egypt	Arabic
4	Alireza Saleh	Director of IPM/IRNIC	.IR Internet Registry	Iran	Persian
5	Allan Ghazi Salahedin	Director of Projects and Technical Development	Ministry of Telecom and IT	Palestine	Arabic
6	Ammar Hussain Jaffri	Chief Executive	Pakistan Information Security Association	Pakistan	Urdu
7	Behnam Esfahbod	Research Assistant	Stevens Institute of Technology	Iran / USA	Persian
8	Hago Elteraifi Mohamed Dafalla	System Administrator and Network Manager	Faculty of Engineering and Technology, University of Gezira	Sudan	Arabic
9	Hania Sabbidin-Dimassi	Research Assistant	ICT Division, UN-ESCWA	Lebanon	Arabic
10	Hazem Hezzah	IT Expert for ICT development	League of Arab States	Egypt	Arabic

11	Inam Ullah	Chairman	Mother Tongue and Heritage for Education and Research (MOTHER)	Pakistan	Pashto, Torwali, Urdu, and Various languages in South Asia
12	Meikal Mumin	PhD Candidate	University of Cologne & University of Naples L'Orientale	Germany	Various African languages
13	Mohammad Ismail Nofal	System Administration Team Lead	Zain Jordan	Jordan	Arabic
14	Nadhem Al-Fardan	Senior Security Architect	Cisco Systems	Saudi Arabia	Arabic
15	Noha Fathy Mohamed	Web Content Editor	Internet Governance for MENA Program	Egypt	Arabic
16	Rinalia Abdul Rahim	Managing Director	Compass Rose Sdn Bhd	Malaysia	Malay
17	Sarmad Hussain	Professor	Al-Khawarizmi Institute of Computer Science, UET	Pakistan	Urdu, Punjabi
18	Syed Iftikhar Shah	Director ICT	Ministry of IT	Pakistan	Seraiki, Urdu
19	Tarik Merghani	CTO	.SD Registry	Sudan	Arabic
20	Tariq Rahim Soomro	Assistant Professor	College of Engineering & Information Technology, Al-Ain University of Science & Technology	Pakistan / UAE	Sindhi, Urdu

The members fill the following relevant slots, as designated by the call for Generation Panels.

Name	Role	Relevant Experience

Sarmad Hussain	Generation Panel Chair	• Conducted multiple projects in localization and language processing of languages in Pakistan and developing Asia
		  o Coordinated the development of language tables and variant analysis for multiple languages (scripts) including Bangla (Bangla), Dzongkha (Tibetan), Khmer (Khmer), Lao (Lao), Mongolian (Cyrillic), Nepali (Devanagari), Pashto (Arabic), Sinhala (Sinhala), Urdu (Arabic) through the PAN Localization project for feedback for IDNA 2008 protocol development
		• Member and coordinator of Arabic Issues Case Study Team, Variant Issues Project, ICANN
		• Member of the Integrated Issues Report, Variant Issues Project, ICANN
		• Member and lead author of the report on Examining the User Experience Implications of Active Variant TLDs, ICANN
		• Member of Arabic Script IDN Working Group (ASIWG)
		• Member of SSAC, ICANN
		• Member of IDN ccTLD Technical Committee, Ministry of IT, Government of Pakistan
		  o Coordinator of Language Table sub-committee (coordinated the effort to develop a single language table for the 66+ languages spoken in Pakistan, which was submitted to ICANN with the IDN ccTLD application for Pakistan)
		  o Member of the Technical sub-committee
		  o Member of the Policy sub-committee (coordinated the development of IDN ccTLD policy for Pakistan, including the variant policy)

Abdelaziz Hilali	Community Representative	• Professor at the Moroccan Telecom graduate institution (INPT) (www.inpt.ac.ma) • PhD in Applied Mathematics and Computer Science in 1987, University of Grenoble • President of the Moroccan delegation representing the civil society in the different phases of the World Summit of the Information Society (WSIS) • Deputy Director for training at INPT (2007-2010) • Deputy Director of Corporate Relations at INPT (since 2010) • Chair of de The African Regional At-Large Organization (AFRALO-ICANN) (AGM 2013- 2015), • Chair and one of the principal founding members of the ISOC Morocco and the Moroccan IPv6Task Force • (www.misoc.org) • President of the Mediterranean Federation of Internet Associations (Tunis 2007) (www.fmai.tn.org) • Deputy Director of Studies at INPT (1995-2000) • Director of the Incubator to create innovative companies in Information Technology at INPT (2001-2007) • Lecturer of Internet governance at universities and high schools in Morocco and Mediterranean countries

Ammar Hussain Jaffri	Community Representative	• Over 40 years of experience in Governance, Digital Forensics, Information Technology, Security, Education • Led different law enforcement departments in the Federal Government of Pakistan. Retired as the Additional Director General of the Federal Investigation Agency of Pakistan in 2010 • Served or serving as the Pakistan point of contact for a number of international initiatives on Cyber Security like the Microsoft Law Enforcement Forum, G-8 24/7 High Tech Crime Network, the OIC-CERT, SAARC CERT, APCERT, Council of the European Union, and with INTERPOL • Speaker on a wide range of subjects including Information Security, Cyber Security, Digital Forensics, Electronic Governance, Business Continuity Planning, Electronic Banking, and emerging trends in ICTs
Hazem Hezzah	Community Representative	• Responsible for the preparation and follow ups of application for the new Arab gTLDs (.arab) and its IDN equivalent (.عرب) by League of Arab States
Mohammad Ismail Nofal	Community Representative	• Systems administration team lead at Zain Jordan • Maintaining and tuning the DNS cache and name server environments • Managing the implementation of the DNSSEC and IPv6 transformation
Noha Fathy Mohamed	Community Representative	• Web Content Editor, Internet Governance for MENA program o Research and produce content specific to Internet Governance in the MENA region focusing on thematic Internet policy topics o Support capacity building on Internet Governance for different stakeholders from MENA region o Support advocacy campaign on Internet Rights in MENA region
Rinalia Abdul Rahim	Community Representative	• Vice-Chair, ICANN At-Large IDN Working Group • Observer, Project Team that developed the Procedure for developing and maintaining Label Generation Rules for the Root Zone with respect to IDNA • Member of the ICANN At-Large Advisory Committee (ALAC) 2011-2013
Adel M. Riyad	DNS/IDNA/Unicode Expert	• Administrator Egypt IDN ccTLD registry system • Administrator Egypt IDN ccTLD DNS
Behnam Esfahbod	DNS/IDNA/Unicode Expert	• Unicode Expert at ICANN IDN Variant Issues Project, Arabic Script Team (2011-2012). • Senior Software Engineer at IRNIC, the dot-IR and dot-Iran ccTLD Registry (2006-2010)

		○ Multilingual web-based domain name registration and management system ○ Extensible Provisioning Protocol (EPP) registrar-registry web service, supporting ○ DNSSEC, IDNs and domain name bundling ○ Real-time Unicode-enabled domain name WHOIS server • Community Manager at Persian Computing Community (since 2002)
Hago Elteraifi Mohamed Dafalla	DNS/IDNA/Unicode Expert	• System Administrator and Network Manager, Faculty of Engineering and Technology, University of Gezira ○ installing and configuring network workstations ○ creating and maintaining user accounts ○ backing up the system ○ distributing software across the networked workstations ○ providing end user support ○ training Staff and Student in the Faculty and in the University
Tariq Rahim Soomro	DNS/IDNA/Unicode Expert	• Researcher of IDNs and Unicode ○ Implementation issues of IDNs ○ Urdu, Sindhi, and Arabic language IDNs • Participated in early development of ASCII based (and later Unicode based) Sindhi Fonts and their applications for Govt. of Sindh, Pakistan
Inam Ullah	Linguistic Expert	• Worked on various projects related to the orthographies of Urdu, Pashto, Torwali and the use of Arabic script for the languages of northern Pakistan, including Khowar, Burushaski, Gawri, Gojri, Hindko etc. • Coordinating communities with emerging orthographies of under-documented languages of Pakistan • Participated in the Language Table sub-committee meetings to develop a single language table for the 66+ languages spoken in Pakistan
Meikal Mumin	Linguistic Expert	• Conducted linguistic scientific research on the use of Arabic script for writing African languages • Co-organizer of two scientific conferences on the use of Arabic script for writing African languages • Co-edited collected volumes and published papers on the use of Arabic script for writing African languages • Initiated TASIA network (http://tasia.org/) as loose association of scientific (mostly linguistic) researchers on the use of Arabic script for writing African languages
Allan Ghazi	Policy Expert	• Implementation and follow-up E-Gov project

Salahedin		• IPV6 deployment Project • Member of (AMAG) • Arab Multi-stakeholder Advisory Group for the Arab Internet Governance forum • Member of PIR (Public interest Registry) Advisory council in Virginia –USA. • Member in PIR workgroups policy and governance, competition and globalization
Hania Sabbidin-Dimassi	Policy Expert	• Supported ESCWA activities to promote the Arabic Domain Names System and assisted in RFC number 5564 entitled "Linguistic Guidelines for the Use of the Arabic Language in Internet Domains"; • Participated in activities to apply for the new Arab gTLDs (.arab) and its IDN equivalent in Arabic; • Takes part in ESCWA activities for regional integration in building the information society and bridging the digital divide in the Arab region; • Key person in ESCWA's activities on digital Arabic content.
Syed Iftikhar Shah	Policy Expert	• Member of GAC from Ministry of IT, Government of Pakistan since 2010 • Member of ICANN Engagement Strategy in the Middle East • Member of IDN VIP ICANN Project Arabic Script IDN Variant Committee • Member of Commonwealth Connects and Point of Contact (PoC) from Pakistan • Member of Project Management Institute (PMI) • Member of National Task Force for IPv6 transition in Pakistan • Member of National Technical Committee regarding the establishment of IDN ccTLD in local languages
AbdulRahman I. Al-Ghadir	Registry/ Registrar	• IDNA Researcher in SaudiNIC working on different topics related to Arabic: o Working in finding a solution for handling variants in IDNs and applying it on IDN ccTLD o Studying the similarity of Arabic script characters and defining variants tables • Participated in multiple IETF mailing groups related to IDNA • Consultation on IDNA related subjects in SaudiNIC
Alireza Saleh	Registry/ Registrar	• CEO IRNIC since 2009; CTO IRNIC since 2003 • Member of Board, .asia, since 2008 • Member andcontributorinIDNA2008IETFworking group • Moderator of wiki page of ICANN .test experiment • ICANNRSTEP Panelist

		• Root LGR panel member of ICANN • Designer and project manager of implementing the IDN registration at IRNIC • Team Member of drafting the IDN registration policy under .ir in Persian • Given many IDN presentations and have participated in various panels at ICANN,APTLD And CENTR • Member of ICANN IDN Variant TLDs Project for Arabic Script (Arabic - vip) • One of the original members of ASIWG (Arabic Script IDN Working Group)
Nadhem Al-Fardan	Registry/ Registrar	• Experience in the following area: o DNS Security o Applied Cryptography o Secure network protocols
Tarik Merghani	Registry/ Registrar	• Internet Project manager (Sudatel 1997 – 1998) • Worked for Sudanet (first ISP in Sudan) as Chief Engineer (1998 – 2000), as Sudanet CTO (2000 – 2002), Sudanet Deputy GM (2003) and Sudanet GM on 2004 • Founder and Board member of Sudan Internet Society 2002 • CTO of Registry system .sd and policy maker for both .sd and .سودان • Responsible for the IDN project (made the variant table to .سودان) • Member of steering committee of Arabic IDN new gTLDs .عرب and .arab • Head of IDN working group at AFTLD • Member of Arab IGF AMAG

Detailed CVs are also attached.

2.2 Panel Diversity

The Generation Panel (GP) for the Arabic Script LGR gathers experts from a variety of backgrounds (bringing varied linguistic and technical perspectives) and covering a wide variety of languages. It is not possible to include one expert per language as Arabic script is used for a significant number of languages. However, care has been taken to have expertise for 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 include expertise in specialized use of Arabic script, including languages of northern Pakistan and Ajami languages in Africa. Beyond the specialized expertise, the group consists of experts who have worked on the use of Arabic script for various languages, including Baluchi, Khowar, Barushaski, Balti, and many other languages. Some of the experts also have knowledge of other scripts.

The GP consists of members from very different perspectives, representing 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.

Geographically, the GP for Arabic script has members from across the relevant regions, including South East Asia, South Asia, the Middle East and North Africa. The members belong to fourteen different countries from these regions.

2.3 Relationship with Past Work or Working Groups

Work on Arabic script IDNs started in early 2000s [6]. Some of the members of TF-AIDN have been involved with Arabic script IDNs since then. Many have been members of Arabic Script IDN Working Group (ASIWG) formed in 2008, and have been active in IDNs since then. These members bring the context of early discussions, issues and solutions to the GP. The members have also been active at relevant ICANN forums, including part of the Variant Issues Project and the follow up work on IDNs and variants at ICANN (including the study on LGR Process and IDN usability). MESWG, the oversight group of TF-AIDN, also has many members who have equally long-term experience in Arabic script IDN, clearly understanding the need and requirements of such work. MESWG also continues to actively engage with TF-AIDN for guiding it.

3. Work Plan

3.1 Suggested Timeline with Significant Milestones

The Generation Panel intends to divide the work on LGR for the Root zone in four stages. These are organized as follows:

1. Finalization of Code Points
2. Finalization of Variants

3. Finalization of Whole Label Rules
4. Finalization of LGR Documents for Arabic Script and Submission to ICANN

Each stage will have three sub-phases. In the first phase of each stage, general principles and framework will be decided on which further decisions will be based. GP will interact with IP to get its feedback, as these principles are finalized. After the principles are finalized, in the second phase the data (character repertoire, variants, and labels) will be analyzed to make concrete decisions. Finally, at each stage, the output will be released to the wider community (including ICANN and IP) for feedback. Each of the four tasks above will be finalized after one or more public comment periods.

It is anticipated that the work will take at least one calendar year. The tentative work plan is depicted in the chart below. Though the chart depicts a linear progression, it is understood that in reality at each stage discussion and feedback will incrementally and cyclically mature the previous portion(s) of the work as well, including principles and data. Further, though it is anticipated that the work will finish in a calendar year, the actual time may vary based on the feedback received by the community and the IP at various stages and phases of the work and may go beyond 2014.

3.2 Proposed Schedule of Meeting and Teleconferences

Most of the work will be accomplished through the TF-AIDN email list. The task force will be holding regular teleconferences every two weeks, for the period of the work. In addition, the task force will also organize face-to-face meetings, tentatively during the discussions on variants (in May 2014) and then towards the end of the work (in November 2014).

3.3 Sources of Funding for Travel and Logistics

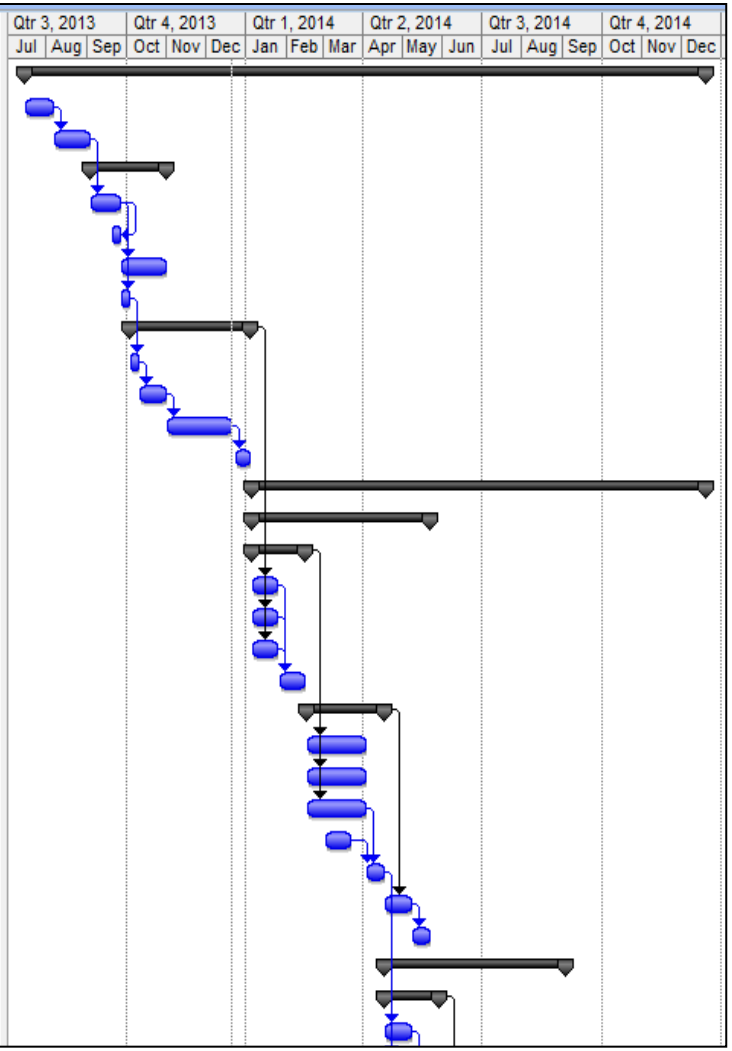
Foremost, all the members of the task force will be volunteering their time for this purpose.

The work has travel and logistic support requirements. ICANN, through its support for MESWG, will support the logistics of the group (e.g. conference calls, assisting in coordination, wiki page for posting information, etc.) to a limited extent. Members of the working group are encouraged to find sources of funding to attend face-to-face meeting(s) related to the TF-AIDN. ICANN, through its support for MESWG, might be able to fund a limited number of those members who are active in the proceedings of the working group, and will also provide support for remote participation for the members not able to attend the meeting(s).

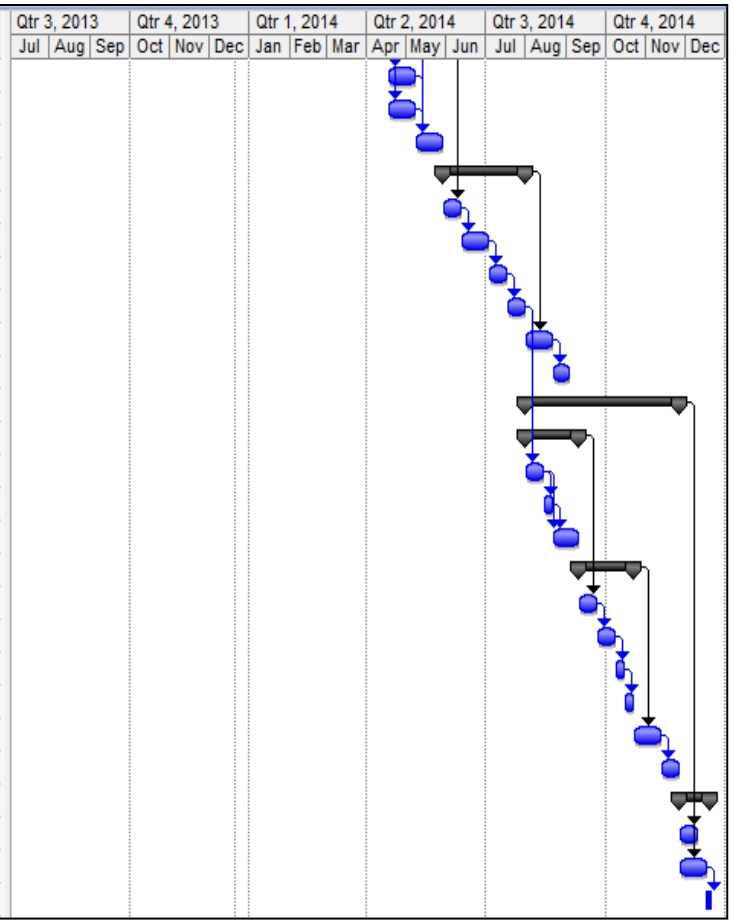
3.4 Need for ICANN Provided Advisors

No advisors are needed at this time, based on earlier discussions and experience during the case study on Arabic Script Issues [9]. ICANN will be requested for advisors, if need arises during the discussions for the development of the LGR for the Root zone.

	Task Name	Duration	Start	Finish	Predecessors
1	**LGR for Arabic Script**	**375 days**	**Mon 15-07-13**	**Fri 19-12-14**	
2	Develop Call for Participation within MESWG	16 days	Mon 15-07-13	Mon 05-08-13	
3	Publicly Release Call for Participation	20 days	Tue 06-08-13	Mon 02-09-13	2
4	**Finalization of Membersip**	**43 days**	**Tue 03-09-13**	**Thu 31-10-13**	
5	Evaluation of Applicants and Finalziation	18 days	Tue 03-09-13	Thu 26-09-13	3
6	Selection/Appoiontment of Chair	5 days	Fri 20-09-13	Thu 26-09-13	5FF
7	Invitaiton to Experts to Ensure Diversity	25 days	Fri 27-09-13	Thu 31-10-13	5
8	Announcement of TF-AIDN	5 days	Fri 27-09-13	Thu 03-10-13	5,6
9	**Formation of Generation Panel**	**66 days**	**Fri 04-10-13**	**Fri 03-01-14**	
10	Introduce Members	5 days	Fri 04-10-13	Thu 10-10-13	8
11	Background Reading	15 days	Fri 11-10-13	Thu 31-10-13	10
12	Proposal Finalization	36 days	Fri 01-11-13	Fri 20-12-13	11
13	Application to ICANN for formation of LGR	10 days	Mon 23-12-13	Fri 03-01-14	12
14	**Work on Arabic Script LGR for Root Zone**	**250 days**	**Mon 06-01-14**	**Fri 19-12-14**	
15	**Character Set**	**98 days**	**Mon 06-01-14**	**Wed 21-05-14**	
16	**Definition of General Principles**	**30 days**	**Mon 06-01-14**	**Fri 14-02-14**	
17	For inclusion	15 days	Mon 06-01-14	Fri 24-01-14	9
18	For Exclusion	15 days	Mon 06-01-14	Fri 24-01-14	9
19	For Deferral	15 days	Mon 06-01-14	Fri 24-01-14	9
20	Interaction with IG for Feedback	15 days	Mon 27-01-14	Fri 14-02-14	17,18,19
21	**Analysis of Data**	**43 days**	**Mon 17-02-14**	**Wed 16-04-14**	
22	Included in Arabic Script IDNs	33 days	Mon 17-02-14	Wed 02-04-14	16
23	Excluded from Arabic Scritp IDNs	33 days	Mon 17-02-14	Wed 02-04-14	16
24	Not Determined at this Time	33 days	Mon 17-02-14	Wed 02-04-14	16
25	Interaction with IG on MSR	15 days	Mon 03-03-14	Fri 21-03-14	
26	Documenting Character Set	10 days	Thu 03-04-14	Wed 16-04-14	25,24
27	Release for Public Comments: Character Set for LGR	15 days	Thu 17-04-14	Wed 07-05-14	21
28	Incorporation of Comments by Public and IG	10 days	Thu 08-05-14	Wed 21-05-14	27
29	**Variants**	**100 days**	**Thu 17-04-14**	**Wed 03-09-14**	
30	**Definition of General Principles**	**30 days**	**Thu 17-04-14**	**Wed 28-05-14**	
31	What is a variant	15 days	Thu 17-04-14	Wed 07-05-14	26



	Task Name	Duration	Start	Finish	Predecessors
32	What is not a variant	15 days	Thu 17-04-14	Wed 07-05-14	26
33	Typology of variants	15 days	Thu 17-04-14	Wed 07-05-14	26
34	Interaction with IG for Feedback	15 days	Thu 08-05-14	Wed 28-05-14	31,32,33
35	**Analysis of Data**	**45 days**	**Thu 29-05-14**	**Wed 30-07-14**	
36	Initial Analysis based on Arabic Script Issues repo	10 days	Thu 29-05-14	Wed 11-06-14	30
37	Second Review of Variants	15 days	Thu 12-06-14	Wed 02-07-14	36
38	Finalization of Variants, with disposition	10 days	Thu 03-07-14	Wed 16-07-14	37
39	Documenting Variants	10 days	Thu 17-07-14	Wed 30-07-14	38
40	Release for Public Comment: Variant Rules	15 days	Thu 31-07-14	Wed 20-08-14	35
41	Incorporation of Comments by Public and IG	10 days	Thu 21-08-14	Wed 03-09-14	40
42	**Whole label rules**	**85 days**	**Thu 31-07-14**	**Wed 26-11-14**	
43	**Definition of General Principles**	**30 days**	**Thu 31-07-14**	**Wed 10-09-14**	
44	Scope of whole label rules	10 days	Thu 31-07-14	Wed 13-08-14	39
45	Typology of variants	5 days	Thu 14-08-14	Wed 20-08-14	44
46	Interaction with IG for Feedback	15 days	Thu 21-08-14	Wed 10-09-14	44,45
47	**Analysis of Data**	**30 days**	**Thu 11-09-14**	**Wed 22-10-14**	
48	Initial Analysis based on Arabic Script Issues repo	10 days	Thu 11-09-14	Wed 24-09-14	43
49	Second Review	10 days	Thu 25-09-14	Wed 08-10-14	48
50	Finalization of WLRs	5 days	Thu 09-10-14	Wed 15-10-14	49
51	Documenting WLRs	5 days	Thu 16-10-14	Wed 22-10-14	50
52	Release for Public Comment: Whole Label Variant Rule	15 days	Thu 23-10-14	Wed 12-11-14	47
53	Incorporation of Comments by Public and IG	10 days	Thu 13-11-14	Wed 26-11-14	52
54	**Finalizing LGR Documents**	**17 days**	**Thu 27-11-14**	**Fri 19-12-14**	
55	Finalizing document	10 days	Thu 27-11-14	Wed 10-12-14	42
56	Finalizing LGR XML structure	15 days	Thu 27-11-14	Wed 17-12-14	42
57	Submission to ICANN	2 days	Thu 18-12-14	Fri 19-12-14	56



The Generation Panel will remain active after the finalization of the LGR documents to continue to address comments from community, ICANN, and the Integration Panel.

References

[1] Hussain, S. (2004). "Letter to Sound Rules for Urdu Text to Speech System," in the *Proceedings of Workshop on Computational Approaches to Arabic Script-based Languages,* COLING 2004, Geneva, Switzerland. Accessed on 15th Nov. 2013 from http://www.cle.org.pk/Publication/papers/2004/LTS_for_urdu_TTS.pdf.

[2] Hussain, S., Rahman, S., Wali, A., Gulzar, A., Rahman, J. S. (2002).*Grammatical Analysis of Nastalique Writing Style of Urdu,* Center for Research in Urdu Language Processing, NUCES, Lahore, Pakistan. Accessed on 15th Nov. 2013 from http://www.cle.org.pk/research/rep/Nastalique%20Writing%20Style.pdf.

[3] El-Sherbiny, A., Farah, M., Oueichek, I. and Al-Zoman, A. (2010). "Linguistic Guidelines for the Use of the Arabic Language in Internet Domains," RFC 5564, IETF. Accessed on 15th Nov. 2013 from http://tools.ietf.org/html/rfc5564.

[4] The Unicode Consortium. The Unicode Standard, Version 6.3.0, (Mountain View, CA: The Unicode Consortium, 2013. ISBN 978-1-936213-08-5). http://www.unicode.org/versions/Unicode6.3.0/.

[5] Wikipedia. "Arabic Script." Accessed on 15th Nov. 2013 from http://en.wikipedia.org/wiki/Arabic_script.

[6] Al-Zoman, A., (2003). "Supporting the Arabic Language in Domain Names." Accessed on 15th Nov. 2013 from http://www.arabic-domains.org/docs/NIC-docs/SupportingArabicDomainNmaes.pdf.

[7] Wikipedia. "List of languages by number of native speakers", Accessed on 15th Nov. 2013 from https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers.

[8] Mumin, M. (2013) "The Arabic Script in Africa: Understudied Literacy". In *The Arabic Script in Africa – Studies in the Use of a Writing System,* eds. M. Mumin and K. Versteegh. Leiden, Boston: Brill.

[9] ICANN (2012). "Arabic Case Study Team Issues Report." Accessed on 24th Nov. 2013 fromhttp://archive.icann.org/en/topics/new-gtlds/arabic-vip-issues-report-07oct11-en.pdf.

[10] Mumin, M. and Versteegh, K. (2013) "Introduction". In *The Arabic Script in Africa – Studies in the Use of a Writing System,* eds. M. Mumin and K. Versteegh. Leiden, Boston: Brill.

Appendix A: Some of the Languages using Arabic Script

There are many languages spoken around the world, which are written with 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 will also be considered for analysis of script by the proposed 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. For these languages all ISO 639-3 available as "living" are included from http://www-01.sil.org/ISO639-3/codes.asp, which may refer to a macro or an individual language. Further, it is intended to provide a list of countries (which may still be incomplete) in which the language is written in Arabic script. The languages may be used in many more countries by expatriate communities and for academic and religious 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*	
Balochi	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*	فارسی الفبای 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)
Swahili	swh, swc, swa	Kenya, Tanzania, Uganda, DR Congo*	

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