Proposal for an Armenian Script Root Zone LGR

LGR Version: 1

Date: 2015-11-05

Document version: 7

Authors: Armenian Script Generation Panel

1 General Information

The purpose of this document is to give an overview of the proposed LGR 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 (https://www.icann.org/en/system/files/files/proposed-armenian-lgr-05nov15-en.xml). Labels for testing can be found in the accompanying text document (https://www.icann.org/en/system/files/files/armenian-test-labels-05nov15-en.txt).

2 Script for which the LGR is proposed

ISO 15924 Code: Armn

ISO 15924 Key N°: 230

ISO 15924 English Name: Armenian

Latin transliteration of native script name: Hye

Native name of the script: հայ

This proposal is based on Maximal Starting Repertoire (MSR) version: MSR-2 [MSR]

The document was developed based on the Guidelines [Guidelines]

3 Background on Script and Principal Languages Using It

The **Armenian language** is an Indo-European language spoken by the Armenians. It is the official language of the Republic of Armenia and the self-proclaimed Nagorno-Karabakh Republic. It has historically been spoken throughout the Armenian Highlands and today is widely spoken in the Armenian Diaspora. Armenian has its own unique script, the Armenian alphabet, invented in 405 AD by Mesrop Mashtots.

Linguists classify Armenian as an independent branch of the Indo-European language family.

There are two standardized modern literary forms, Eastern Armenian and Western Armenian, with which most contemporary dialects are mutually intelligible.

3.1 Principal Languages Using the Script.

The following lists the Language code and name for each (from ISO 639-3)
ISO 639-1 hy
ISO 639-2 arm (B) hye (T)
ISO 639-3

Variously: hye – Modern Armenian xcl – Classical Armenian axm – Middle Armenian
Glottolog: arme1241
Linguasphere: 57-AAA-a

3.2   Geographic Territories or Countries with Significant User Communities

Geographic territories or countries with significant user communities for the script

Official language in Armenia and Nagorno-Karabakh Republic

Diasporas using the Armenian language in

1	Russia	38	Venezuela
2	United States	39	Cyprus
3	France	40	Estonia
4	Georgia	41	Italy
5	Iran	42	Denmark
6	Ukraine	43	United Arab Emirates
7	Turkey	44	Tajikistan
8	Lebanon	45	Jordan
9	Argentina	46	Moldova
10	Syria	47	Lithuania
11	Canada	48	Israel
12	Greece	49	Kyrgyzstan
13	Abkhazia	50	Chile
14	Bulgaria	51	Norway
15	Uzbekistan	52	Finland
16	Spain	53	Malta
17	Germany	54	Slovakia
18	Poland	55	Slovenia
19	Australia	56	Albania
20	Brazil	57	Mexico
21	Belarus	58	Serbia
22	Turkmenistan	59	Macedonia
23	Kazakhstan	60	South Africa
24	United Kingdom	61	Peru
25	Hungary	62	New Zealand
26	Uruguay	63	India
27	Iraq	64	Ireland
28	Netherlands	65	Portugal
29	Belgium	66	Ethiopia
30	Kuwait	67	Cuba
31	Egypt	68	Singapore
32	Czech Republic	69	China
33	Sweden	70	Japan
34	Austria	71	Thailand
35	Romania	72	Morocco
36	Latvia	73	Luxembourg
37	Switzerland	74	Bangladesh

3.3   Related Scripts

There is some commonality with Latin, Greek and Cyrillic.

4   Overall Development Process and Methodology

See discussion in the following sections for process and methodology for Repertoire and Variants.

5 Repertoire

The proposal is based on MSR-2 version. Armenian MSR-2 table is shown in the Appendix B [MSR]. There are two standardized mutually intelligible modern literary forms, Eastern Armenian and Western Armenian with some difference in orthographies. But as both are written with the same script it doesn't influence the LGR document.

The following web site contains a full description of the Armenian repertoire for use in the Armenian writing system: http://www.omniglot.com/writing/armenian.htm

This defines a formal reference for the justification of the repertoire for this proposal.

The following subset of code points identified as the Unicode range U+0561..U+0586 will be available for the root zone:

0561	ա	ARMENIAN SMALL LETTER AYB
0562	բ	ARMENIAN SMALL LETTER BEN
0563	գ	ARMENIAN SMALL LETTER GIM
0564	դ	ARMENIAN SMALL LETTER DA
0565	ե	ARMENIAN SMALL LETTER ECH
0566	զ	ARMENIAN SMALL LETTER ZA
0567	է	ARMENIAN SMALL LETTER EH
0568	ը	ARMENIAN SMALL LETTER ET
0569	թ	ARMENIAN SMALL LETTER TO
056A	ժ	ARMENIAN SMALL LETTER ZHE
056B	ի	ARMENIAN SMALL LETTER INI
056C	լ	ARMENIAN SMALL LETTER LIWN
056D	խ	ARMENIAN SMALL LETTER XEH
056E	ծ	ARMENIAN SMALL LETTER CA
056F	կ	ARMENIAN SMALL LETTER KEN
0570	հ	ARMENIAN SMALL LETTER HO
0571	ձ	ARMENIAN SMALL LETTER JA
0572	ղ	ARMENIAN SMALL LETTER GHAD
0573	ճ	ARMENIAN SMALL LETTER CHEH
0574	մ	ARMENIAN SMALL LETTER MEN
0575	յ	ARMENIAN SMALL LETTER YI
0576	ն	ARMENIAN SMALL LETTER NOW
0577	շ	ARMENIAN SMALL LETTER SHA
0578	ո	ARMENIAN SMALL LETTER VO
0579	չ	ARMENIAN SMALL LETTER CHA
057A	պ	ARMENIAN SMALL LETTER PEH
057B	ջ	ARMENIAN SMALL LETTER JHEH
057C	ռ	ARMENIAN SMALL LETTER RA
057D	ս	ARMENIAN SMALL LETTER SEH
057E	վ	ARMENIAN SMALL LETTER VEW
057F	տ	ARMENIAN SMALL LETTER TIWN
0580	ր	ARMENIAN SMALL LETTER REH

0581	ց	ARMENIAN SMALL LETTER CO
0582	ւ	ARMENIAN SMALL LETTER YIWN
0583	փ	ARMENIAN SMALL LETTER PIWR
0584	ք	ARMENIAN SMALL LETTER KEH
0585	օ	ARMENIAN SMALL LETTER OH
0586	ֆ	ARMENIAN SMALL LETTER FEH

6   Variants

There is some similarity between Armenian letters GIM and ZA but we considered that it is not enough to include them into variants. So our conclusion is: **The Armenian repertoire does not contain in-script variants.**

The GP decision is in accordance with a prediction of the [Procedure] cited below:
"We have identified a certain number of scripts, with small repertoires and used by one or by only a few languages, as somewhat less difficult than the cases that have attracted the most attention. At present we judge them to be relatively unproblematic, and they may not need to have very large generation panels or protracted discussion. (This does not mean that they would not require a generation panel to do the processing; it is just recognized that the discussions about their repertoire are unlikely to be complex or problematic.) The scripts identified as most likely belonging to this class are Armenian, Georgian, and Thaana."

6.1   Cross-script Homoglyphs

Cyrillic, Greek and Latin scripts are examples of related scripts where some cross-script homoglyphs exist.

6.1.1   Armenian and Cyrillic homoglyphs

Armenian Letter	**Cyrillic Letter**
ш U+561 Armenian small letter AYB	ш U+448 Cyrillic small letter SHA
һ U+0570 Armenian small letter HO	һ U+04BB Cyrillic small letter SHHA
o U+0585 Armenian small letter OH	o U+043E Cyrillic small letter O

The ARMENIAN LETTER YI and j U+006A LATIN SMALL LETTER J are not considered homoglyphs as the GP considers that the presence of a "dot" changes the visual perception of the letter. That is why this case is not included into the cross-script variants.

6.1.2   Armenian and Greek homoglyphs

Armenian Letter	**Greek Letter**
η U+0572 Armenian small letter GHAD	η U+03B7 Greek small letter ETA
ı U+0582 Armenian small letter YIWN	ı U+03B9 Greek small letter IOTA
o U+0585 Armenian small letter OH	o U+03BF Greek small letter OMICRON

4

6.1.3   Armenian and Latin homoglyphs

Armenian Letter	**Latin Letter**
ɡ U+0581 Armenian small letter CO	g U+0067 Latin small letter G
ɦ U+0570 Armenian small letter HO	h U+0068 Latin small letter H
ո U+0578 Armenian small letter VO	n U+006E Latin small letter N
o U+0585 Armenian small letter OH	o U+006F Latin small letter O
q U+0566 Armenian small letter ZA	q U+0071 Latin small letter Q
ս U+057D Armenian small letter SEH	u U+0075 Latin small letter U
ı U+0582 Armenian small letter YIWN	ɩ U+0269 Latin small letter IOTA

The Armenian letter YI and j U+006A Latin small letter J are not considered homoglyphs as the GP considers that the presence of a "dot" changes the visual perception of the letter. That is why this case is not included into the cross-script variants.

7   Whole Label Evaluation (WLE) Rules

The GP reviewed the question of Whole Label Evaluation rules and came to a positive decision that they are not needed.

8   Contributors

Composition of the Armenian Generation Panel

1	Igor Mkrtumyan	President .am Administrative Contact	ISOC Armenia AM NIC	Armenia	English, Russian, Armenian
2	Grigori Saghyan	Vice President	ISOC Armenia	Armenia	English, Russian, Armenian
3	Lianna Galstyan	Board member	ISOC Armenia	Armenia	English, Russian, Armenian
4	Vladimir Sahakyan	Manager of ASNET AM Vice President of ISOC AM	Institute for Informatics and Automation Problems of the National Academy of Sciences, Republic of Armenia; ISOC Armenia	Armenia	English, Russian, Armenian
5	Anna Karakhanyan	Board member	ISOC Armenia	Armenia	English, Russian, Armenian
6	Ruben Hakobyan	President	Armenian Association of Type Designers	Armenia	English, Russian, Armenian
7	Kristina Babajanyan	Executive Director	"ABC Domain" Registrar Company	Armenia	English, Russian,

					Armenian
8	Hrant Dadivanyan	Manager, .am technical contact	AM NIC	Armenia	English, Russian, Armenian

1	Igor Mkrtumyan	Chair, Contact person, coordinator, and official representative of the Generation Panel to ICANN and to the Integration Panel	Experience in the management of .am Registry and AM NIC, performing the role of the country domain administrative contact, communications with IANA, WIPO, and solution of UDRP cases.
2	Grigori Saghyan	Expert in local culture, customs, and practices	Experience in the management of CERT AM, Internet security, representing the country in ITU meetings, development of .am Registry and AM NIC policy, development of Registry-Registrar and Registrar-Registrant contracts, Internet security problem solution.
3	Lianna Galstyan	Expert in the specific languages/scripts with good knowledge in Unicode	Experience in the management of ISOC AM as ICANN At-Large Structure, development of policy of .am Registry, management of registrar and webhosting companies. She also has a good knowledge of Armenian language and culture as a linguist
4	Vladimir Sahakyan	Expert in the specific languages/scripts with good knowledge in Unicode	Experience in the management of Academic Scientific Research Network (ASNET), upgrading the academic and research network to IPv6 readiness, and linguistic expertise in the Armenian language/script with good knowledge of Unicode and IDNA.
5	Anna Karakhanyan	Expert in the specific languages/scripts with good knowledge the Armenian script	Experience in the management of Internet Registry and Registrar, development of policy of .am Registry, operations of AMNIC, linguistic expertise in the Armenian language/script with good knowledge of Unicode.
6	Ruben Hakobyan	Expert in the specific languages/scripts with good knowledge in Unicode	Experience in Armenian and Russian Hyphenation macros; ArmSCII-8 to UNICODE transcoder for MS Word and Excel, development of new generation of Armenian TrueType fonts and linguistic expertise in the Armenian language/script with good knowledge of Unicode.

7	Kristina Babajanyan	Expert familiar with registry/ registrar operations and standards in the Armenian script	Experience in the management of a Registrar of .am Registry, webhosting services, management of Registrar's part of .am zone, registry-registrar and registrar-registrant operations and its legal problems, good knowledge of Unicode and standards of Armenian language/script.
8	Hrant Dadivanyan	Expert knowledgeable of ICANN's bottom - up process and local registration policies	Experience in the management of AMNIC, .am zone file and DNS services, communications with IANA on zone file changes, DNSSEC operations, IPv6 readiness of .am DNS servers and IDN, good knowledge of routers, system administration and TCP/IP theory and practice.

9   References

[Procedure] Internet Corporation for Assigned Names and Numbers, "Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels." (Los Angeles, California: ICANN, March, 2013) http://www.icann.org/en/resources/idn/variant-tlds/draft-lgr-procedure-20mar13-en.pdf

[Requirements] Integration Panel "Requirements for LGR Proposals from Generation Panels" available online as https://www.icann.org/en/system/files/files/Requirements-for-LGR-Proposals-20150424.pdf

[MSR] Internet Corporation for Assigned Names and Numbers, "Maximal Starting Repertoire", (Los Angeles, California: ICANN April 2015) –Annotated Reference Tables, Non-CJK: https://www.icann.org/en/system/files/files/msr-2-non-cjk-13apr15-en.pdf

[Guidelines] Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR, https://www.icann.org/en/system/files/files/Guidelines-for-LGR-20150424.pdf

Appendix A - Notes on second level domains

Visual similarities in the second level domains will be addressed by registration rules. There will be restrictions in the policy of domain registration. The problem will be solved by limiting Armenian domain names strictly to the Armenian MSR table, Latin dash (code point '2d') and Latin numbers (code points '30' - '39')".

Visual similarities of related scripts will be blocked by the domain registration program as it will check the scripts for the correspondence to the Armenian MSR table and will not allow domain-names with the visually similar code points of related scripts.

Appendix B - Excerpt from Maximal Starting Repertoire - MSR-2

Annotated Repertoire Tables, Non-CJK
Integration Panel
Date: 2014-11-14
How to read this file:
This file shows all non-CJK characters that are included in the MSR-2 with a yellow background. The set of these code points matches the repertoire specified in the XML format of the MSR. Where present, annotations on individual code points indicate some or all of the languages a code point is used for.
This file lists only those Unicode blocks containing non-CJK code points included in the MSR. Code points listed in this document, which are PVALID in IDNA2008 but excluded from the MSR for various reasons are shown with pinkish annotations indicating the primary rationale for excluding the code points, together with other information about usage background, where present. Code points shown with a white background are not PVALID in IDNA2008, or are ineligible for the root zone (digits hyphen).
Repertoire corresponding to the CJK Unified Ideographs: Main (4E00-9FFF), Extension-A (3400-4DBF), Extension B (20000-2A6DF), and Hangul Syllables (AC00-D7A3) are included in separate files. For links to these files see "Maximal Starting Repertoire - MSR-2: Overview and Rationale".
How the repertoire was chosen:
This file only provides a brief categorization of code points that are PVALID in IDNA2008 but excluded from the MSR. For a complete discussion of the principles and guidelines followed by the Integration Panel in creating the MSR, as well as links to the other files, please see "Maximal Starting Repertoire - MSR-2: Overview and Rationale".
Brief description of exclusion types:
-Obsolete (historic, archaic),
-Limited or declining use (educational, threatened, nearly extinct),
-Symbol (characters classified as letters that are symbolic in nature),
-Numeric (characters used in numerical context),
-Punctuation (characters classified as letters that look like punctuation),
-CONTEXTJ (context - join controls),
-CONTEXTO (context - others),
-Unstable (encoding model changed),
-Deprecated (no longer in use, alternate code preferred),
-Technical use (phonetic, poetry),
-Religious use (annotation, cant illation),

-Homoglyph (digraph of x y),
-Deferred repertoire (7.0 repertoire not yet included in IANA tables). The optional parenthetical annotations provided give further information where appropriate or available.

0530   Armenian   058F

	053	054	055	056	057	058
0		Չ 0540	Ր 0550		Հ 0570	Ր 0580
1	Ա 0531	Ձ 0541	Յ 0551	ա 0561	ձ 0571	ց 0581
2	Բ 0532	Ղ 0542	Ի 0552	բ 0562	ղ 0572	ւ 0582
3	Գ 0533	Ճ 0543	Փ 0553	գ 0563	ճ 0573	փ 0583
4	Դ 0534	Մ 0544	Ք 0554	դ 0564	մ 0574	ք 0584
5	Ե 0535	Յ 0545	Օ 0555	ե 0565	յ 0575	o 0585
6	Զ 0536	Ն 0546	Ֆ 0556	զ 0566	ն 0576	ֆ 0586
7	Է 0537	Շ 0547		է 0567	շ 0577	և 0587
8	Ը 0538	Ո 0548		ը 0568	ո 0578	
9	Թ 0539	Չ 0549	ՙ 0559	թ 0569	չ 0579	։ 0589
A	Ժ 053A	Պ 054A	՚ 055A	ժ 056A	պ 057A	֊ 058A
B	Ի 053B	Ջ 054B	՛ 055B	ի 056B	ջ 057B	
C	Լ 053C	Ռ 054C	՜ 055C	լ 056C	ռ 057C	
D	Խ 053D	Ս 054D	՝ 055D	խ 056D	ս 057D	֍ 058D
E	Ծ 053E	Վ 054E	՞ 055E	ծ 056E	վ 057E	֎ 058E
F	Կ 053F	Տ 054F	՟ 055F	կ 056F	տ 057F	֏ 058F

10