

Proposal for an Armenian Script Root Zone LGR

Version 3

Date: 2015-06-04

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 (Proposed-LGR-Armn-20150604.xml).

1.1 Script for which the LGR is proposed

Code: Armn

N°: 230

English Name: Armenian

Property Value Alias: Armenian

Date: 2004-05-01

English name of the script: Hye

Native name of the script: հայ

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.

1.2 Principal languages using that script

List Language code and name for each (from ISO 639-3)

ISO 639-1

ISO 639-2

ISO 639-3

Variously:

hye – Modern Armenian

xcl – Classical Armenian

axm – Middle Armenian

Glottolog

Linguasphere

1.3 Geographic territories or countries with significant user communities for the script

Official language in Armenia

 Nagorno-Karabakh Republic

Big Diaspora using the Armenian language in

Argentina
Brazil
Cyprus
France
Georgia (Samtskhe-Javakheti)
Hungary
Iraq
Lebanon
Poland
Romania
Syria
Turkey
Ukraine
United States

Total Armenian population in the world is about 10mln.

1.4 Related scripts

There is some commonality with Latin, Greek and Cyrillic.

2. Composition of the Armenian Generation Panel

2.1. The composition of the Armenian Generation Panel is shown in Appendix 1.

3. Final report

3.1 Repertoire

Armenian MSR2 table is shown in the Appendix 2 [1].

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	ւ	ARMENIAN SMALL LETTER MEN
0575	յ	ARMENIAN SMALL LETTER YI
0576	ի	ARMENIAN SMALL LETTER NOW
0577	շ	ARMENIAN SMALL LETTER SHA
0578	ն	ARMENIAN SMALL LETTER VO
0579	չ	ARMENIAN SMALL LETTER CHA
057A	պ	ARMENIAN SMALL LETTER PEH
057B	զ	ARMENIAN SMALL LETTER JEH
057C	ր	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	լ	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

3.2 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.**

Our GP decision is in accordance with a prediction of the document [2] 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.”

3.3 Cross-script homoglyphs

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

Armenian-Cyrillic homoglyphs

Armenian Letter	Cyrillic Letter
հ U+0570 Armenian small letter HO	Һ U+04BB Cyrillic small letter SHHA
օ U+0585 Armenian small letter OH	о U+043E Cyrillic small letter O

Note:

‘ն’ U+0578 Armenian small letter VO and ‘п’ U+43F Cyrillic small letter PE are usually somewhat distinct, unless both are in italic serif style, like ‘*n*’ vs. ‘*n*’, where they would be homoglyphs. However, they are not included into cross-script variants because italic font styles are insufficient justification for creating a cross-script variant, especially as that is not the most common style for presenting domain names to the user.

Armenian and Latin homoglyphs

Armenian Letter	Latin Letter
զ U+0566 Armenian small letter ZA	q U+0071 Latin small letter Q
հ U+0570 Armenian small letter HO	h U+0068 Latin small letter H
ն U+0578 Armenian small letter VO	n U+006E Latin small letter N
ւ U+057D Armenian small letter SHE	u U+0075 Latin small letter U
ց U+0581 Armenian small letter CO	g U+0067 Latin small letter G
յ U+0582 Armenian small letter YIWN	ı U+0269 Latin small letter IOTA
օ U+0585 Armenian small letter OH	o U+006F Latin small letter O

Notes:

The ARMENIAN SMALL LETTER ZA and the LATIN SMALL LETTER Q are not perfect homoglyphs but the difference may not be perceivable at normal size. A label, such as .quo would be readily accepted by users as “the same” as the label .quo (in Latin).

The ‘g’ homoglyph situation only exists in sans-serif style, which, however, is a very common choice for user interfaces.

Armenian YIWN and Latin IOTA are considered homoglyphs as their visual perception is quite similar.

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

Armenian and Greek homoglyphs

Armenian Letter	Greek Letter
ղ U+0572 Armenian small letter GHAD	η U+03B7 Greek small letter ETA
լ U+0582 Armenian small letter YIWN	ι U+03B9 Greek small letter IOTA
օ U+0585 Armenian small letter OH	ο U+03BF Greek small letter OMICRON

Note:

Armenian YIWN and Greek IOTA as well as Armenian GHAD and Greek ETA are considered homoglyphs as their visual perception is quite similar. They are included in the cross-script variants.

The case of OH/OMICRON/Cyrillic O/Latin O is considered cross-script homoglyphs because of strong similarity.

As result of integration, blocked variants can exist between script-specific repertoires.

Even if a label is applied for in one script, the existence of certain other labels in another script may block it from being delegated [3].

For example, both .uno and .ւոո would be well-formed TLDs in Latin and Armenian,

If one of these two labels is applied for and allocated, the other would be blocked. Otherwise, both could be independently allocated to different applicants causing user confusion. [3]

Armenian GP is hardly expecting any other application for an Armenian TLD but in cases when there is homoglyph similarity in related scripts it should be blocked.

The Armenian GP has added out of repertoire blocked variants for all homoglyphs investigated above, except for cases identified as merely "similar".

3.4 Second level domains

b. Visual similarities in the second level domains will be solved 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 domains names with the visually similar code points of related scripts.

Reference:

1. Maximal Starting Repertoire - MSR-2 (MSR2-non-cjk-14nov14-en.pdf)
2. Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels
3. Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR

Appendix 1. Composition of the Armenian Generation Panel

	Name	Designation	Organization	Country	Languages
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

	Name	Role	Experience
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	Experience in the management of ISOC AM as ICANN At-Large Structure, development of policy of .am Registry, management

		languages/scripts with good knowledge in Unicode	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.

Appendix 2. 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.

	053	054	055	056	057	058
0		Հ 0540	Ր 0550		Տ 0570	Ր 0580
1	Ա 0531	Զ 0541	Յ 0551	Մ 0561	Ճ 0571	Գ 0581
2	Բ 0532	Ղ 0542	Ի 0552	Բ 0562	Ղ 0572	Լ 0582
3	Գ 0533	Ճ 0543	Փ 0553	Գ 0563	Ճ 0573	Վ 0583
4	Դ 0534	Մ 0544	Բ 0554	Գ 0564	Ճ 0574	Բ 0584
5	Ե 0535	Յ 0545	Օ 0555	Ե 0565	Յ 0575	Օ 0585
6	Զ 0536	Ն 0546	Ֆ 0556	Զ 0566	Ն 0576	Փ 0586
7	Է 0537	Շ 0547		Է 0567	Շ 0577	Է 0587
8	Ը 0538	Ո 0548		Ը 0568	Ո 0578	
9	Թ 0539	Չ 0549	Ս 0559	Թ 0569	Չ 0579	Ս 0589
A	Ժ 053A	Պ 054A	Վ 055A	Ժ 056A	Պ 057A	Վ 058A
B	Ի 053B	Ղ 054B	Ճ 055B	Ի 056B	Ղ 057B	
C	Լ 053C	Ռ 054C	Ն 055C	Լ 056C	Ռ 057C	
D	Խ 053D	Ս 054D	Վ 055D	Խ 056D	Ս 057D	
E	Ծ 053E	Վ 054E	Ր 055E	Ծ 056E	Վ 057E	
F	Կ 053F	Տ 054F	Ն 055F	Կ 056F	Տ 057F	Գ 058F

Uppercase letters

0531	Ա	ARMENIAN CAPITAL LETTER AYB
0532	Բ	ARMENIAN CAPITAL LETTER BEN
0533	Գ	ARMENIAN CAPITAL LETTER GIM
0534	Դ	ARMENIAN CAPITAL LETTER DA
0535	Ե	ARMENIAN CAPITAL LETTER ECH
0536	Զ	ARMENIAN CAPITAL LETTER ZA
0537	Է	ARMENIAN CAPITAL LETTER EH
0538	Ը	ARMENIAN CAPITAL LETTER ET
0539	Թ	ARMENIAN CAPITAL LETTER TO
053A	Ճ	ARMENIAN CAPITAL LETTER ZHE
053B	Ի	ARMENIAN CAPITAL LETTER INI
053C	Լ	ARMENIAN CAPITAL LETTER LIWN
053D	Խ	ARMENIAN CAPITAL LETTER XEH
053E	Վ	ARMENIAN CAPITAL LETTER CA
053F	Կ	ARMENIAN CAPITAL LETTER KEN
0540	Հ	ARMENIAN CAPITAL LETTER HO
0541	Ջ	ARMENIAN CAPITAL LETTER JA
0542	Ղ	ARMENIAN CAPITAL LETTER GHAD
0543	Ճ	ARMENIAN CAPITAL LETTER CHEH
0544	Մ	ARMENIAN CAPITAL LETTER MEN
0545	Ի	ARMENIAN CAPITAL LETTER YI
0546	Ն	ARMENIAN CAPITAL LETTER NOW
0547	Շ	ARMENIAN CAPITAL LETTER SHA
0548	Ո	ARMENIAN CAPITAL LETTER VO
0549	Չ	ARMENIAN CAPITAL LETTER CHA
054A	Պ	ARMENIAN CAPITAL LETTER PEH
054B	Ջ	ARMENIAN CAPITAL LETTER JHEH
054C	Ռ	ARMENIAN CAPITAL LETTER RA
054D	Ս	ARMENIAN CAPITAL LETTER SEH
054E	Վ	ARMENIAN CAPITAL LETTER VEV
054F	Տ	ARMENIAN CAPITAL LETTER TIWN
0550	Ր	ARMENIAN CAPITAL LETTER REH
0551	Յ	ARMENIAN CAPITAL LETTER CO
0552	Ի	ARMENIAN CAPITAL LETTER YIWN
0553	Փ	ARMENIAN CAPITAL LETTER PIWR
0554	Թ	ARMENIAN CAPITAL LETTER KEH
0555	Օ	ARMENIAN CAPITAL LETTER OH
0556	Ֆ	ARMENIAN CAPITAL LETTER FEH

Modifier letters

0559	◌̈	ARMENIAN MODIFIER LETTER LEFT HALF RING • technical use
055A	՝	ARMENIAN APOSTROPHE
055B	՜	ARMENIAN EMPHASIS MARK
055C	՚	ARMENIAN EXCLAMATION MARK
055D	՛	ARMENIAN COMMA
055E	՛	ARMENIAN QUESTION MARK
055F	՝	ARMENIAN ABBREVIATION MARK

Lowercase letters

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	ճ	ARMENIAN SMALL LETTER ZHE
056B	ի	ARMENIAN SMALL LETTER INI

056C	լ	ARMENIAN SMALL LETTER LIWN
056D	խ	ARMENIAN SMALL LETTER XEH
056E	վ	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	զ	ARMENIAN SMALL LETTER JHEH
057C	ր	ARMENIAN SMALL LETTER RA
057D	ս	ARMENIAN SMALL LETTER SEH
057E	վ	ARMENIAN SMALL LETTER VEV
057F	տ	ARMENIAN SMALL LETTER TIWN
0580	ր	ARMENIAN SMALL LETTER REH
0581	հ	ARMENIAN SMALL LETTER CO
0582	լ	ARMENIAN SMALL LETTER YIWN
0583	փ	ARMENIAN SMALL LETTER PIWR
0584	թ	ARMENIAN SMALL LETTER KEH
0585	օ	ARMENIAN SMALL LETTER OH
0586	ֆ	ARMENIAN SMALL LETTER FEH
0587	և	ARMENIAN SMALL LIGATURE ECH YIWN

Punctuation

0589	։	ARMENIAN FULL STOP
058A	֊	ARMENIAN HYPHEN

Religious symbols

058D	Ⲁ	RIGHT-FACING ARMENIAN ETERNITY SIGN
058E	ⲁ	LEFT-FACING ARMENIAN ETERNITY SIGN

Currency symbol

058F	֏	ARMENIAN DRAM SIGN
------	---	--------------------