

Proposal for the Georgian Script Root Zone LGR

LGR Version 2

Date: 2016-09-15

Document version: 1

Authors: GEORGIAN SCRIPT GENERATION PANEL

1 General Information/ Overview/ Abstract

The purpose of this document is to give an overview of the proposed Georgian 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-Georgian-20160915.xml

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

- Labels-GeorgianScript-20160915.txt

2 Script for which the LGR is proposed

ISO 15924 Code: Geor

ISO 15924 Key N°: 240

ISO 15924 English Name: Georgian Mkhedruli

Latin transliteration of native script name: Mkhedruli

Native name of the script: მხედრული

Maximal Starting Repertoire (MSR) version: MSR-2

3 Background on Script and Principal Languages Using It

The Georgian scripts are the three writing systems used to write the Georgian language: Asomtavruli, Nuskhuri and Mkhedruli. Mkhedruli (Georgian: მხედრული) is the current Georgian script and is therefore the standard script for modern Georgian and its related Kartvelian languages, whereas Asomtavruli and Nuskhuri are used only in ceremonial religious texts and iconography. In the following, the term Georgian script is used synonymously with Mkhedruli.

Like the two other scripts, Mkhedruli is purely unicameral. Mkhedruli first appears in the 10th century - the oldest Mkhedruli inscription found is dated back to 982 AD. Mkhedruli is written horizontally following the standard Left-to-Right direction, with spaces between words. The modern Georgian alphabet consists of 33 letters, (5 vowels and 28 consonants) with no upper case characters – the Georgian script comes in only a single case.

The Georgian script was first added to the Unicode Standard in October, 1991 with the release of version 1.0 [0].

The Georgian script is used to write the Georgian language, which is the official language for Georgia. It is a Kartvelian or South Caucasian language spoken by about 4.1 million people mainly in Georgia, and also in Armenia, Azerbaijan, Iran, Kazakhstan, Kyrgyzstan, Russia, Tajikistan, Turkey, Turkmenistan, Ukraine, USA and Uzbekistan.

Additionally, the Georgian script is also used to write the following languages:

- Mingrelian, a South Caucasian language spoken in north-western Georgia by perhaps half a million people.
- Laz, a South Caucasian language closely related to Mingrelian and spoken in Turkey and Georgia by about 33,000 people.

- Svan, a South Caucasian language with about 30,000 speakers mainly in the northwest of Georgia.
- Abkhaz, a Northwest Caucasian language, was also once written with the Mkhedruli alphabet, but is now written with the Cyrillic alphabet.

The Panel has looked at these languages during the analysis. As long as the Mingrelian is written in the Mkhedruli script, it adopts the rules of the Georgian script, with some of its archaic characters are no longer in use in everyday speaking / writing. They are used only for scientific or religious and historical writings. These characters include U+10F7 and U+10F8, designated by Unicode as “Extensions for Mingrelian”, and U+10F2, which is only used in archaic scripts, mainly of religious character. However, these three code points can be completely substituted by other characters of Mkhedruli script, for example use U+10D8 and U+10DC instead of U+10F7, so the Georgian Script GP decided not to allow their use, as Mingrelian is fully supported by the Georgian Script LGR.

4 Overall Development Process and Methodology

In order to come up with the principles that are used to determine valid code variants, labels and points, the Georgian Script GP organized several face to face meetings with its members and defined 33 from 37 code points from MSR Version 2, needed for the root zone.

Due to the unique characters of Mkhedruli, the panel unanimously adopted the decision that there are no confusing code points or label constraints in the use of the selected code points, as there is no confusing similarities with any other scripts.

Panel members also engaged in an online conversation with ICANN in order to ensure the quality and the direction of the decisions to develop a valid Root Zone LGR proposal, including references for every code point included, label rules, variants and provide XML file.

5 Repertoire

The repertoire for the Georgian script can be found at

<http://www.omniglot.com/writing/georgian.htm#mkhedruli> [100]. The generation panel has short-listed the following code points for the root zone LGR, based on the Maximal Starting Repertoire (MSR-2).

S.No.	Unicode Code Point	Glyph	Name
1	10D0	ა	GEORGIAN LETTER AN
2	10D1	ბ	GEORGIAN LETTER BAN
3	10D2	გ	GEORGIAN LETTER GAN
4	10D3	დ	GEORGIAN LETTER DON
5	10D4	ე	GEORGIAN LETTER EN
6	10D5	ვ	GEORGIAN LETTER VIN
7	10D6	ზ	GEORGIAN LETTER ZEN
8	10D7	თ	GEORGIAN LETTER TAN
9	10D8	ი	GEORGIAN LETTER IN
10	10D9	კ	GEORGIAN LETTER KAN
11	10DA	ლ	GEORGIAN LETTER LAS
12	10DB	მ	GEORGIAN LETTER MAN
13	10DC	ნ	GEORGIAN LETTER NAR
14	10DD	ო	GEORGIAN LETTER ON
15	10DE	პ	GEORGIAN LETTER PAR
16	10DF	ჟ	GEORGIAN LETTER ZHAR
17	10E0	რ	GEORGIAN LETTER RAE
18	10E1	ს	GEORGIAN LETTER SAN
19	10E2	ტ	GEORGIAN LETTER TAR
20	10E3	უ	GEORGIAN LETTER UN
21	10E4	ფ	GEORGIAN LETTER PHAR
22	10E5	ქ	GEORGIAN LETTER KHAR
23	10E6	ყ	GEORGIAN LETTER GHAN
24	10E7	ღ	GEORGIAN LETTER QAR
25	10E8	შ	GEORGIAN LETTER SHIN
26	10E9	ჩ	GEORGIAN LETTER CHIN
27	10EA	ც	GEORGIAN LETTER CAN
28	10EB	ძ	GEORGIAN LETTER JIL
29	10EC	წ	GEORGIAN LETTER CIL
30	10ED	ჭ	GEORGIAN LETTER CHAR
31	10EE	ხ	GEORGIAN LETTER XAN

32	10EF	ჯ	GEORGIAN LETTER JHAN
33	10F0	ჰ	GEORGIAN LETTER HAE

ა	ბ	გ	დ	ე	ვ	ზ	თ	ი
ანი	ბანი	განი	დონი	ენი	ვინი	ზენი	თანი	ინი
ani	bani	gani	doni	eni	vini	zeni	tani	ini
1	2	3	4	5	6	7	9	10
კ	ლ	მ	ნ	ო	პ	ჟ	რ	ს
კანი	ლასი	მანი	ნარი	ონი	პარი	ჟანი	რაე	სანი
k'ani	lasi	mani	nari	oni	p'ari	zhani	rae	sani
20	30	40	50	70	80	90	100	200
ტ	უ	ფ	ქ	ღ	ყ	შ	ჩ	ც
ტარი	უნი	ფარი	ქანი	ღანი	ყარი	შინი	ჩინი	ცანი
t'ari	uni	pari	kani	ghani	q'ari	shini	chini	tsani
300		500	600	700	800	900	1000	2000
ძ	წ	ჭ	ხ	ჯ	ჰ			
ძილი	წილი	ჭარი	ხანი	ჯანი	ჰაე			
dzili	ts'ili	ch'ari	khani	jani	hae			
3000	4000	5000	6000	8000	9000			

*Source: <http://www.omniglot.com/writing/georgian2.htm>

6 Variants

There are neither in-script or cross-script variants for Georgian.

7 Whole Label Evaluation (WLE) Rules

There are no WLE rules needed for Georgian script.

8 Contributors

Name	Role	Designation	Organization
Konstantine Karosanidze	DNS / IDNA / Unicode Expert	UNIX System Administrator	Information Technologies Development Centre
Mzia Gogilashvili	Policy Expert	Chief Specialist/International Affairs Coordinator Strategy Development Department	Georgian National Communications Commission
David Birman	Community Representative	Chief Digital Officer	Bank of Georgia
Sophio Elizbarashvili	Generation Panel Chair	PR & social media manager	Information Technologies Development Centre
Malcolm Taylor	XML code Developer	Chief Specialist of System Administration	Information Technologies Development Centre
Marine Beridze	Linguistic Expert	Chief executive of linguistic data digital processing department	Arnold Chikobava State Institute Of Linguistics
Ia Feradze	Registry Expert	.GE Domain Administrator	Caucasus Online

9 Proposal References

[0] The Unicode Standard, Version 1.0 [<http://unicode.org/charts/PDF/U10A0.pdf>]

[100] Georgian alphabet (Mkhedruli) [<http://www.omniglot.com/writing/georgian2.htm>]

[Proposal] Proposal for the Generation Panel for the Georgian Script Label Generation Ruleset for the Root Zone

[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>