

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

Table of Contents

1.	General Information	2
1.1	Use of Latin Script characters in domain names	3
1.2	Target Script for the Proposed Generation Panel	4
1.2.1	Diacritics	5
1.3	Countries with significant user communities using Latin script	6
2.	Proposed Initial Composition of the Panel and Relationship with Past Work or Working Groups	7
3.	Work Plan	13
3.1	Suggested Timeline with Significant Milestones	13
3.2	Sources for funding travel and logistics	16
3.3	Need for ICANN provided advisors	17
4.	References	17

1. General Information

The Latin script¹ or Roman script is a major writing system of the world today, and the most widely used in terms of number of languages and number of speakers, with circa 70% of the world's readers and writers making use of this script² ([Wikipedia](#)).

Historically, it is derived from the Greek alphabet, as is the Cyrillic script. The Greek alphabet is in turn derived from the Phoenician alphabet which dates to the mid-11th century BC and is itself based on older scripts. This explains why Latin, Cyrillic and Greek share some letters, which may become relevant to the ruleset in the form of cross-script variants.

The Latin alphabet itself originated in Italy in the 7th Century BC. The original alphabet contained 21 upper case only letters: A, B, C, D, E, F, Z, H, I, K, L, M, N, O, P, Q, R, S, T, V and X. Soon after, this repertoire was extended. For example, the letter G developed from C and J from I, while letter V and U split and a ligature³ of VV became W. For our purposes, the basic set of the historic repertoire of Latin script is considered: A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y and Z. The Latin script is written left-to-right.

With the spread of the Roman Empire and the Catholic Church, the script became increasingly used to represent other languages apart from Latin. As a result, new letters were added to the script, using different mechanisms for the purposes of representing speech sounds which were unknown to - at first - Latin or - much later - European languages: For example, some letters were formed by combining two glyphs into one ligature, such as 'æ' from 'a' and 'e', as used in Danish and Norwegian. Also, new letters were borrowed through contact in between languages and scripts, such as 'þ' (thorn) used for Scandinavian languages, which was borrowed from the Runic alphabet, or 'ʿ/ع' borrowed from the Arabic letter 'ع' ([Wikipedia](#)). Borrowing was an ongoing process, which continued throughout the history of the use of Latin Script. Even repeated processes of borrowing occurred. For example, Greek 'Υ' (gamma), which was the original source for both C and G, was borrowed again in its original shape for the representation of the voiced velar fricative for the writing of languages such as Ewe (ISO 639-3⁴ ewe) ([Wikipedia](#)).

In addition to borrowing, new letters were developed using different mechanisms such as the addition of various modifiers to existing letter shapes: bars or strokes, e.g. 'R/ɾ' as used in Kanuri (ISO 639-3 kau) ([Wikipedia](#)), hooks, e.g. 'K/k' as used in Hausa (ISO 639-3 hau) ([Wikipedia](#)), horizontal rotation or mirroring, e.g. 'Ɔ' based on 'E' as used by the Pan-Nigerian alphabet ([Wikipedia](#)), vertical rotation or mirroring, e.g. 'Λ/λ' on the basis of 'v', as used by Ibibio (ISO 639-3 ibb) ([Wikipedia](#)), making (part of) a letter shape more cursive or italic, e.g. 'U/u' on the basis of 'v' as used by Ewe (ISO 639-3 ewe) ([Wikipedia](#)), or the re-use of further signs such as punctuation marks to develop

¹ *Script* is used here to indicate the whole writing system including basic letters, ligatures and diacritics. See also RFC 6365 and ISO 15924.

² However, several orthographies on the basis of different scripts are frequently used simultaneously, both historically and contemporarily.

³ In writing and typography, a ligature occurs where two or more graphemes or letters are joined as a single glyph.

⁴ Due to the variation of language names, ISO 639-3:2007 codes are placed in brackets after such language names in the present document to aid in identifying the correct idiom.

letters, e.g. 'ʔ' (glottal stop) based on '?' (question mark) as used by Chipewyan (ISO 639-3 chp) ([Wikipedia](#)).

Typologically the Latin script is an Alphabet, which means that segments or units of the writing system generally tend to represent consonants and vowels (rather than other linguistic units)⁵. Today, Latin script is a bicameral system, which features upper and lower case forms for most letters. There may be little visual similarity between a letter's upper and lower case forms, for example, 'A' and 'a'. However, some orthographies have reversed the development back towards a unicameral system, without a distinction in between upper and lower case letters, such as the second version of the African Reference Alphabet from 1982 ([Wikipedia](#)).

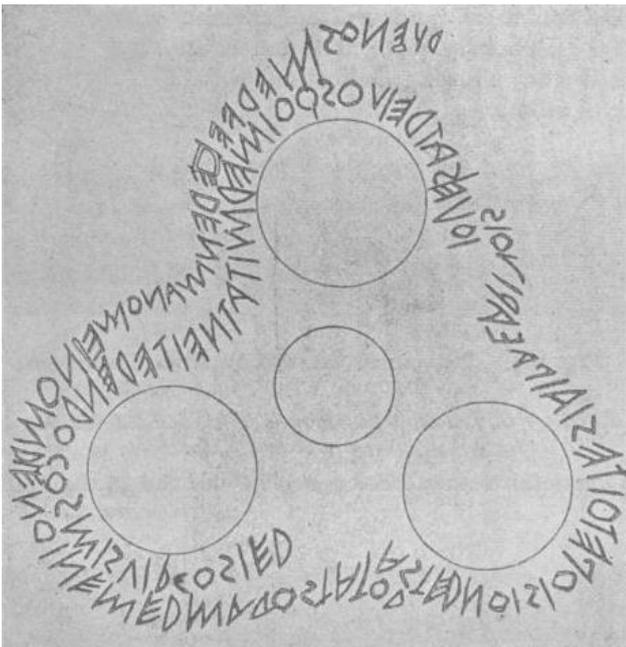


Figure 1: The Duenos Inscription, 6th Century B.C., one of the earliest surviving documents in Latin (taken from [Wikipedia](#))

1.1 Use of Latin Script characters in domain names

Traditional domain names are domain names without the extension that IDN gives. These are the domain names that most of us are familiar with. The characters in traditional domain names are limited to a subset of the ASCII characters set. In Top Level Domains (TLDs) the permitted characters are the characters in the English alphabet:

⁵ The exact representation of linguistic features however is different for every orthography. Some languages, such as Esperanto, use it more phonemically, while other languages, such as English, use it so that other aspects, such as etymology, are represented too. For example, the spelling of 'night' connects it with German 'Nacht', although 'gh' is no longer pronounced. Therefore, the degree to which phonemes are represented can vary from orthography to orthography, and this characteristic has been described by the term orthographic depth in linguistic literature (cf. Katz & Frost 1992).

a b c d e f g h i j k l m n o p q r s t u v w x y z

Those characters, 'a-z', form a subset of the Latin Script.

In traditional domain names, or traditional domain name labels, both upper case 'A-Z' and lower case 'a-z' can be used and mixed Upper and lower case versions of a letter are considered to be equal. For example “com”, “Com” and “COM” are identical from a domain name perspective. We will see below, that in IDN domain name U-labels (but not A-labels), only the lower case letters may be used.

However, in Second Level Domains (SLDs) and below, digits (0 1 2 3 4 5 6 7 8 9) and the hyphen '-' (U+002D) can also be included. As in all domain names, the dot '.' is used to separate the parts (or “labels”) of the domain name, e.g. “www”, “example1” and “com” in “www.example1.com”. In this example “com” is the top-level domain name label, that may only contain characters 'a-z' (except for valid A-labels), whereas the other lower level labels may contain digits and the hyphen.

1.2 Target Script for the Proposed Generation Panel

As per the *Procedure to Develop and Maintain the Label Generation Rules for the DNS Root Zone in Respect of IDNA Labels* (referred to simply as [Procedure] in the following), only code points included in Maximal Starting Repertoire, Second Version (Referred to simply as [MSR-2] in the following) will be considered.

The Latin script has the following specifications:

ISO 15924 code: Latn

ISO 15924 no.: 215

English Name: Latin

The set of code points in the Latin script, as specified by [MSR-2], contains 326 selected code points from the following Unicode ranges, i.e. 305 letters and 21 Combining Diacritical Marks:

Script	Range of Unicode code points
Controls and Basic Latin	U+0061 – U+007A
Controls and Latin-1 Supplement	U+00DF - U+00F6 U+00F8 - U+00FF
Latin Extended-A only lowercase	U+0101 – U+017F
Latin Extended-B	U+0180 – U+024F
IPA Extensions	U+0250 – U+02AF
Combining Diacritical Marks	U+0300 – U+036F
Combining Diacritical Marks Supplement	U+1DC0 – U+1DFF
Latin Extended Additional	U+1E00 – U+1EFF
Latin Extended-C	U+2C60 – U+2C7F

MSR-2 excluded the following Latin script UNICODE ranges:

Latin Extended-D; technical use (phonetic)/obsolete/punctuation	U+A720 – U+A7FF
Latin Ligatures; compatibility characters not PVALID in IDNA 2008	U+FB00 – U+FB0F
Full-width Latin Letters; compatibility characters not PVALID in IDNA 2008	U+FF00 – U+FF5E

When there is a single code point which gives the combined letter and diacritical mark, that will be used in preference to using the combining diacritic mark code point. Furthermore, only lower case letters are considered, as upper case ones may not be used in IDNs following [IDNA 2008].

1.2.1 Diacritics

Diacritics are marks which combine with other letters. In Latin script, many languages make use of diacritics to modify letters. Diacritics may appear anywhere around a letter, most commonly above 'é', below 'ṭ', or through 'ð' the letter. Several diacritics may attach to the same letter, such as in the Vietnamese: 'ê̂'. However, often diacritics are used for specialized purposes, such as phonetic notation/romanization, and therefore may not be part of the general orthographic system.

Diacritics may perform different roles depending on the language: Some languages consider letter + diacritic as one letter. An example is Norwegian (both Bokmål and Nynorsk varieties), which lists 'Æ', 'Ø' and 'Å' at the end of its alphabet. In contrast, Italian recognizes diacritics as a further layer of the orthography. For example 'è', which differs from 'e', is not listed as a separate letter of the alphabet, but occurs in dictionaries following unmodified 'e'.

Similarly, diacritics may express different units of language and may be treated differently from other elements of the writing system. In numerous languages, they are an obligatory part of the alphabet and serve to distinguish entirely different phonemic units of a language, such as the combining stroke⁶ on 'l' in Navajo, i.e. 'l' vs. 'ḷ', where the former represents an approximant and the latter a fricative, two different types of consonants. Also, diacritics may be applied to other letters to systematically modify phonemes, such as U+0308 (COMBINING DIAERESIS) which in German is an obligatory part of the alphabet, generally indicating a change in the quality of a vowel as in "Ofen" ['o:fən] "oven" vs. "Öfen" ['ø:fən], the plural form of the same. However in other languages, they may be optional in some contexts, such as in Italian, where homographs, which differ in the position of the word-level stress (the so-called *parole sdrucciole*), can optionally be written with a combining accent mark U+02CA ' (MODIFIER LETTER ACUTE ACCENT) to indicate the position of word-level stress in antepenultimate positions, e.g. "súbito" "immediately" vs. "subito" "suffered" (about.com).

Diacritics may express segmental units of a languages, such as single phonemes, or may express supra-segmental features such as word-level stress, e.g. U+02CA ' (MODIFIER LETTER ACUTE ACCENT) in Spanish "corazón" "heart". Or they may express tone, such as in Thai-word "mái" "wood", where the same diacritic indicates a high tone. Nonetheless diacritics may serve to distinguish minimal pairs in numerous languages, irrespective of whether their omission is

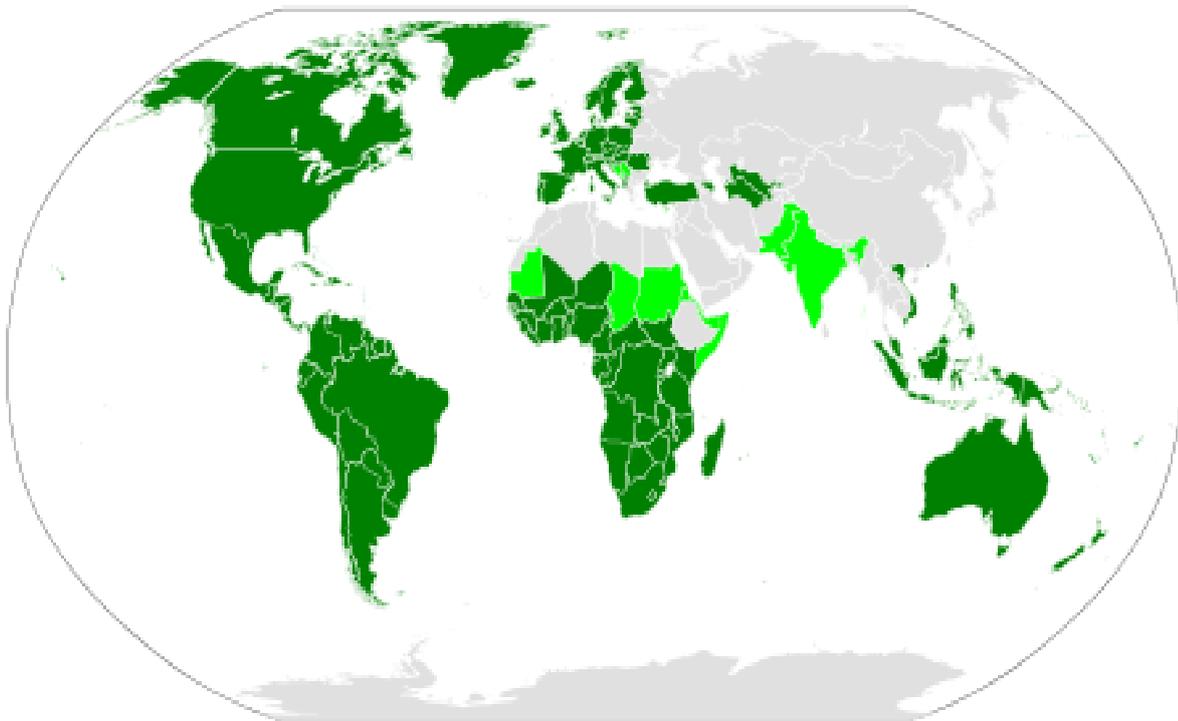
⁶ The Combining stroke is not encoded as a combining mark in Unicode.

considered a spelling mistake (such as "vô" "grandfather" vs. "vó" "grandmother" in Portuguese or "è" "is" vs "e" "and", in Italian), or whether the use of such diacritics is optional (such as in the Italian example "súbito" vs. "subito" presented above).

While generally diacritics take effect only on those letters with which they are combined, there are a few languages in which diacritics have an effect on adjacent letters, particularly where several letters are used to represent single phonemes, i.e. di-, tri- or quadri-graphs. An example is Maltese, where the combining stroke over 'h' affects the preceding letter 'g', since 'gh' is a digraph in that language which contrasts with simple 'g', e.g. "gallettina" "biscuit" and "għasfur" "bird" ([Wikipedia](#)).

1.3 Countries with significant user communities using Latin script

Per [Wikipedia](#) the distribution of the Latin script on the world map is:



Dark green marks countries where the Latin script is the sole main script.

Light green marks countries where Latin co-exists with other scripts.

Grey marks areas, in which supposedly Latin-script is not used or used only unofficially for second language. However there are several widely known cases which may be eligible for consideration for the LGR which run counter to this classification, such as French in Algeria, English in Egypt, or even specialized yet widely used written representations, such as [Arabic chat alphabet](#), or the use of Latin transliteration for Chinese languages, known as [Pinyin](#) or for Japanese romanji.

There are no reliable figures regarding the number of languages using Latin script or the number of readers and writers of the script using community. Per [Worldstandards.eu](#), languages using Latin script are spoken by 2.6 billion people (36% of the world population). However such figures, just as

the visual depictions from Wikipedia, are far from exact. Various linguistic factors play into this, including issues of classifications of spoken languages (e.g. dialect vs. language), socio-linguistic factors such as the use of different languages, scripts, and orthographies by the same users for different contexts, as well as issues of language policy, which may not recognize or may even ban the use of languages, script, and orthographies for political reasons, irrespective of the actual use.

Appendix A is composed using data found on [OMNIGLOT](#) and [ETHNOLOGUE](#) sites. However, this constitutes only a sample of data. The Generation Panel will consider all languages using Latin script, provided they fulfil the criteria of inclusion stipulated by the [Procedure]. In that context, it is the criteria of modern wide-spread use which shall serve as crucial factor for inclusion. The Integration Panel [MSR-2] has made use of the EGIDS (Expanded Graded Intergenerational Disruption Scale) as documented in [SIL-Ethnologue] to evaluate modern use, and in [MSR-2] used a cut-off between EGIDS level 4 (Educational) and level 5 (Developing). This however is only seen as a guideline for inclusion, and where the Panel sees data positively confirming a relevant use of a language of a higher EGIDS rating, the panel shall document such for the final proposal and provide feedback to the Integration Panel within its review of [MSR-2].

2. Proposed Initial Composition of the Panel and Relationship with Past Work or Working Groups

The current working group listing includes first the members of the panel, then the observers, in alphabetical order:

No.	Name	Position	Organization	Country	Language Expertise
1.	Abdeslam Nasri	Member	ATOS	Algeria	Arabic, French
2.	Ahmed Bakhat Masood	Member	Pakistan Telecom Authority	Pakistan	Urdu, English
3.	Bakiau Takentebwebwe	Member	CCK - Kiribati Regulator for Telecom	Kiribati	Gilbertese, English
4.	Bill Jouris	Member	Inside Products	USA	English, German, Japanese
5.	Dennis Tan Tanaka	Member	Verisign	USA	Spanish, English
6.	Elvin Prasad	Member	Government	Fiji	Fijian, English
7.	Fiammetta Caccavale	Member	Student	Denmark/ Italy	Italian, English
8.	Hazem Hezzah	Member	League of Arab States	Egypt	Arabic, German, English
9.	Jean-Jacques Subrenat	Member	NCUC; Individual Users; NMI/CC; ICG	France	French, English
10.	Mats Dufberg	Member	Internet Foundation in Sweden	Sweden	Swedish, English, some Danish

No.	Name	Position	Organization	Country	Language Expertise
11.	Meikal Mumin	Member	Institute for African Studies and Egyptology, University of Cologne & “L’Orientale” University of Naples	Germany	German, English, Italian, French, and various African and Middle Eastern languages
12.	Michael Bauland	Member	Knipp Medien und Kommunikation GmbH	Germany	German, English, Finnish
13.	Mirjana Tasić	Chair	Register of National Internet Domain Names of Serbia (RNIDS)	Serbia	Serbian, Croatian, Bosnian, Montenegrin, English
14.	Nebiye Petek Kurtböke	Member	IJDLC EURALEX	Turkey	Turkish, Italian, English
15.	Ousmane Moussa Tessa	Member	Université A. Moumouni, Niamey	Niger	Zarma, French, English, Hausa
16.	Seun Ojedeji	Member	AFRINIC BOD member	Nigeria	Yoruba, English
17.	Tran Canh Toan	Member	VNNIC	Vietnam	Vietnamese, English
18.	Violet Rose Ningakun	Member	University of Papua New Guinea	Papua New Guinea	Melanesian Pidgin/ Tok pisin, English
19.	Danko Jevtovic	Observer	Register of National Internet Domain Names of Serbia (RNIDS)	Serbia	Serbian, English
20.	Jiankang Yao	Observer	Computer Network Information Center (CNIC, CAS)	China	Mandarin Chinese, Pinyin and English
21.	Matthias Brenzinger	Observer	University of Cape Town	South Africa	
22.	Tarik Merghani	Observer	AFTLD	Sudan	

The panel currently includes members from several areas of the script using community. However, as the Latin script is used by hundreds of languages, it is not possible to have direct representation from community members or speakers from all of them. At the same time, there are numerous languages, for which resources are scarce or difficult to access. The panel finds it lacks sufficient expertise, and sees a need for access to qualified advisers. It will work together with ICANN and request support in obtaining help from external experts, who are not formally members of the panel, to act as advisors to the panel following the [Procedure], which the GP considers as critical to the success of its work. At the same time the GP will be working to invite more members.

These individuals, listed without short CVs and currently marked as observers, have at some point expressed their willingness to support the work of the panel, but have not reacted to further communiques during the process leading to the seating of the proposed panel and the submission

of the present proposal. In the interest of openness and because of the panel’s wide remit, the members of the suggested panel have decided not to remove those individuals for the moment. The panel expressly remains open to new members, or observers becoming full members, throughout its work. Relevant expertise of panel members is demonstrated in the following table:

N.	Name	Role	Designation	Relevant experience
1.	Abdeslam Nasri	ICT Architect/ Arabic Generation Panel	ICT Architect and Project Manager / AtoS	<ul style="list-style-type: none"> ● 2014 to present: Member of the Arabic GP ● 2014 to present: Member of the Task Force on Arabic IDN (TF-AIDN) ● Expertise in various IT domains like software development, Internet development and multi-tiered architectures, Enterprise architecture. PSPO I and TOGAF certification ● Panellist at the Internet Governance Forum
2.	Ahmed Bakhat Masood	Regulator/ DNS/ Arabic Generation Panel/Security	Deputy Director (ICT/Network)/ Pakistan Telecom Authority	<ul style="list-style-type: none"> ● 2013 to present: Member of Task Force on Arabic IDN (TF-AIDN) ● 2014- to present: Member of Program Committee Middle East DNS Forum) ● 1998 to present: Pakistan Telecom Authority (PTA) ● Initiation of different ICT projects for community development like IXP for Pakistan ● Coordination for Ipv6 Task Force for Pakistan Network Management, Network Security including DNSSec and Network forensic ● Coordination with APNIC, SANOG, ICANN and academia for trainings on modern technologies like IPV6, DNSSec, IRM ● Network and Security management ● Implementation of ISO 27001 standards in PTA
3.	Bakiau Takentebwebwe	Administration and Finance	Manager at CCK, - Regulator Telecom	<ul style="list-style-type: none"> ● 2012 to present Administration and Finance manager ● 2008 to 2012 Finance Manager ● 2004 to 2008 Personal officer ● administer Ki domain name only for updating client’s account when we receive payment for particular domains
4.	Bill Jouris	Computer performance measurement and analysis expert	COO of Inside Products	<ul style="list-style-type: none"> ● 2016 to present: technical work on project to identify homographic conflicts in domain names ● 2015 to present: work on various IETF RFCs ● 2012 to 2015: Director, Computer Measurement Group
5.	Dennis Tan	Registry	Senior	<ul style="list-style-type: none"> ● Member of the IDN Implementation Guidelines Review Team

N.	Name	Role	Designation	Relevant experience
	Tanaka	Operator/IDN	Platform Manager	<ul style="list-style-type: none"> ● IDN Product Manager
6.	Elvin Prasad	Government of Fiji	Senior Engineer ICT	<ul style="list-style-type: none"> ● GAC representative Fiji
7.	Fiametta Caccavale	Master student in IT and Cognition. Bachelor in Literature with Major in Linguistics	University of Copenhagen	<ul style="list-style-type: none"> ● 2016 to present: student of IT and Cognition at the University of Copenhagen. Particular interested in the field of Natural Language Processing ● 2013 to 2016: Bachelor in Literature and Linguistics at the University of Padova, Italy
8.	Hazem Hezzah	Arabic Generation Panel member/ National and regional policy maker	IT Expert for ICT Development / League of Arab States	<ul style="list-style-type: none"> ● 2013-present: Member of the Task Force for Arabic Script IDNs (TF-AIDN) ● 2012- present: Member of the Multistakeholder advisory group and preparation team for the Arab Internet Governance Forum. ● 2012-present: Participated in preparation, evaluation and contracting for the (.arab) gTLDs, and currently preparing policies for the new gTLD. ● 1991-2011: Performed various IT related roles as support, consultant and technical project manager. ● Languages: English, German, use of Latin script for Arabic chat language
9.	Jean-Jacques Subrenat	Policy Expert	Président, IndividualUsers.org (European Individual Users' Association)	<ul style="list-style-type: none"> ● Member of the NTIA IANA Functions' Stewardship Transition Coordination Group (ICG) ● Member of the NETMundial Coordination Council ● President of the Steering Committee, IndividualUsers.org (elected in October 2015) Member of the ICANN Board of Directors 2007-10 during which: ● Member of President's Strategy Committee (where he was a co-author of the "Implementation Plan for Improving Institutional Confidence") ● Structural Improvements Committee; Public Participation Committee (as its first Chair) ● Member of Board Working Groups: ALAC Review, Board Review, ccNSO Review (as its Chair)

N.	Name	Role	Designation	Relevant experience
10.	Mats Dufberg	IDN/DNS/ Linguist	Internet Foundation In Sweden	<ul style="list-style-type: none"> ● DNS specialist ● New GTLD Pre-Delegation testing ● BA Phonetics/Linguistics
11.	Meikal Mumin	Linguist	Institute for African Studies and Egyptology, University of Cologne & “L’Orientale” University of Naples	<ul style="list-style-type: none"> ● Member of Arabic Generation Panel ● Member of Task Force on Arabic Script IDNs (TF-AIDN) ● PhD candidate and researcher with an expertise in Writing Systems and orthographies, in particular such used in Africa and the Middle East. ● User of several Modern European Languages
12.	Michael Bauland	DNS/Registry/ Registrar / IDN	Knipp Medien und Kommunikati on GmbH	<ul style="list-style-type: none"> ● Development of IDN table for < بازار > (bazaar.) ● 2007-present: Senior Software Engineer at Knipp ● 2003-2007: Senior Research Assistant at Leibniz University of Hanover
13.	Mirjana Tasić	Registry / DNS/Unicode Expert / IDN	Executive Advisor, RNIDS (Register of National Internet Domain Names of Serbia)	<ul style="list-style-type: none"> ● 08/2012–12/2012 ICANN IDN variant TLD Program: Project (P2.1) – Procedure to Develop and Maintain the Label Generation Rules for the DNS Root Zone in Respect of IDNA Labels - ICANN volunteer ● Introduction and implementation of IDN ccTLD Fast Track Process for ccTLD < cp6 > < xn—90a3ac >: string evaluation, domain delegation, sunrise and open registration. ● 07/2006–03/2009 Acting Director of RNIDS (volunteer work). Preparation and implementation of .rs landrush procedures; organization and implementation of the transition process from .yu to .rs domain. ● 04/2006–07/2006 Founder of RNIDS (volunteer work). ● 04/1994–09/2008 YU TLD (YU Top Level Domain) Administrator (volunteer work). Managed operation of .yu DNS; Maintained database of .yu domains. ● 1992–1994 Chairwoman, Technical Committee, Academic Network of Yugoslavia. Actively participated in the introduction of internet in Serbia. (volunteer work)

N.	Name	Role	Designation	Relevant experience
				<ul style="list-style-type: none"> 1991–10/2010 Administrator of Class B IP address (147.91) assigned to the University of Belgrade, Serbia. (volunteer work)
14.	Nebiyé Petek Kurtböke	Linguist/ Lexicography	Member of the Review Editorial Board	<ul style="list-style-type: none"> PhD Linguistic Member of EURALEX
15.	Ousmane Moussa Tessa	Education/ Mathematics/ Scientific text processing	Associate Professor	<ul style="list-style-type: none"> PhD in Educational Sciences PhD in Mathematics Master in Educational Science Master in Mathematics
16.	Seun Ojedeji	DNS/ Policy / Native speaker	Chief Network Engineer at Federal University of Oye-Ekiti	<ul style="list-style-type: none"> AFRINIC Member Board of Directors Open Source Foundation for Nigeria Vice President FOSSFA Council Chair AFRINIC Policy Development Working Group co-chair Principal Network Engineer System analyst/Network engineer
17.	Tran Canh Toan	DNS/ IDN	Deputy Director of Technical Department, VNNIC	<ul style="list-style-type: none"> Technical operations of ccTLD registry for Vietnam
18.	Violet Rose Ningakun	Community Member	Instructional Designer	<ul style="list-style-type: none"> 2016 - Member of the Asia Pacific School of Internet Governance
19.	Danko Jevtovic	Observer		
20.	Jiankang Yao	Observer		
21.	Matthias Brenzinger	Observer		
22.	Tarik Merghani	Observer		

The Generation Panel intends to take into consideration and investigate recommendations from other relevant Panels such as the Integration Panel as well as from other Generation Panels which have contributed prior relevant work, particularly when making choices about the repertoires and variant relationships within, such as the “feasibility and risks of supporting the sharp s in the LGR” [MSR-2: 19] and, if it should consider the inclusion of this code point in the LGR, to investigate the case for or against making it a blocked variant of 'ss', or the view of Armenian GP regarding relevant cross-script variants.

ICANN’s Variant Issues Project Study Group for the Latin Script produced *Considerations in the use of the Latin script in variant internationalized top-level domains* in 2011. These considerations are

relevant in this context.

3. Work Plan

The role of the Latin Generation Panel is to establish the repertoire and Label Generation Rules for top level internationalized domain names in Latin script. In establishing the repertoire, the Generation Panel will strictly adhere to the inclusion principle, building their proposed repertoire ‘from the ground up’ and positively affirming each and every code point in their LGR proposals, giving justification for the inclusion of every single code point in their proposed repertoire (cf. pp. 30-31, of [MSR-2]). Meanwhile code points that are not part of the MSR, but which are an integral part of an important language, will be suggested to be added to the MSR.

The panel plans to review code-points, not by sequence in MSR, but based on their use in individual languages. Due to the size of the script using community, the panel intends to build up a repository by focusing first on languages based on their expected relevance, and the EGIDS scale is taken as a proxy of such use. Therefore the panel intends to review Latin-script using languages of EGIDS level 1, before such of levels 2, 3, and 4 to ensure that time will be spent with a focus on code-points of practical importance to the LGR. To this end, the panel will make review of available resources such as IDN tables or LGRs (including reference tables for the second level) for languages that are already well supported in domain names. However, the panel will also seek to avoid systemic bias by considering some additional cases, which may be marginal in practical value to the LGR, to ensure linguistic rights of under-resourced and under-represented language communities.

When analyzing variants, the panel will only consider homoglyphs, which are characters with essentially identical appearance by design, instead of merely similar appearance. However, such distinctions, while commonly applied in the context of DNS, may not be applicable for some script-using communities as demonstrated already by the variant analyses conducted by other GPs such as e.g. the case of *heh* and *tah marbutah* in the Arabic LGR.

3.1 Suggested Timeline with Significant Milestones

The Generation Panel intends to divide the work on the LGR for the Root Zone into eight stages:

1. Initialization of the work
2. Development of Principles
3. Development of the Code Point Repertoire
4. Development of the Code Point Variants
5. Integration of different groups work products
6. Development of the Whole Label Evaluation Rules
7. Preparation of the Latin script LGR for public comment
8. Finalization of the LGR for Latin Script and submission to ICANN

At all stages, there will be consultation with the Integration Panel and the public via periodic public comment phases.

	Activity	Completion
1	Initiate work	2 Weeks
1.1	Review work plan and associate working groups to activities at least as follows: <ol style="list-style-type: none"> 1. Group 1 <ol style="list-style-type: none"> a. Development Code point inclusion criteria b. Development of Code Point Repertoire 2. Group 2 <ol style="list-style-type: none"> a. Development of in-script and cross-script variants criteria b. Development of Code Point Variant Sets c. Coordination with Armenian, Cyrillic and Greek GP 	2 weeks
2.	Develop Principles	4 weeks
2.1	Determine criteria for including code points - Group1	4 weeks
2.2	Determine criteria for within-script and cross-script variants - Group2	4 weeks in parallel with 2.1
2.3	Informal public comment on principles released by GP	4 weeks in parallel with 3
3.	Develop Code Point Repertoire - Group1	24 weeks
3.1	Review EGIDS level 1 languages (60) Review EGIDS level 2 languages (30) Review EGIDS level 3 languages (40) Review EGIDS level 4 languages (50)	18 weeks (approx. 10 languages per week)
3.2	Finalize included and excluded code points	2 weeks
3.4	Consider any additional languages for inclusion	4 weeks
3.5	Release Repertoire by Latin GP for informal public comment	4 weeks in parallel with 5

	Activity	Completion
4.	Develop code point variants based on MSR2 - Group 2	20 weeks in parallel with 3.
4.1	Analyze each MSR2 code point to determine cross-script variant sets and define table with cross-script variants	6 weeks
4.2	Coordinate with other relevant GPs for finalization of cross-script variants	4 weeks
4.3	Analyze each MSR2 code point to determine within-script variant sets	4 weeks
4.4	Review the impact of variant analysis on current delegations of TLDs	4 weeks
4.5	Finalize variant sets, reviewing the symmetry, transitivity, security and stability of the system	2 weeks
4.6	Release Variants by Latin GP for informal public comment	4 weeks in parallel with 5
5.	Integration of different groups work products	4 weeks
5.1	Integration of Group 1 and 2 work products	2 weeks
5.2	Check the status of Cyrillic and Greek GP and make necessary changes	2 weeks
5.3	Release Repertoire and Variants by Latin GP for informal public comment	4 weeks in parallel with 6
6.	Discuss WLE rules needed for Latin script LGR	6 weeks
6.1	Determine relevant WLE rules	6 weeks
6.2	Release WLE rules by Latin GP for informal public comment	4 weeks in Parallel with 7
7.	Prepare Latin script LGR proposal for public comment	8 weeks
7.1	Finalize the documentation for Latin script LGR	4 weeks

	Activity	Completion
7.2	Finalize the XML formulation of the proposal	4 weeks in parallel to 7.1
7.3	Collect labels to test the proposed LGR	2 weeks in parallel to 7.1
7.4	Review and finalize LGR proposal documents	4 weeks
8.	Finalize Latin LGR proposal	10 weeks
8.1	Release by ICANN for formal public comment	6 weeks
8.2	Finalize Latin LGR proposal based on feedback from the public comment phase	4 weeks

Work will be divided among different working subgroups so, that parallel processing of activities could be achieved.

Start week	End week	Task
1	2	Initiate work
3	6	Develop Principles
7	26	Develop Code Point Variants
7	30	Develop Code Point Repertoire
31	34	Integration of different groups work products
35	40	Discuss WLE Rules needed for Latin Script LGR
41	48	Prepare Latin Script LGR Proposal for Public Comment
49	58	Finalize Latin LGR Proposal

Note that this schedule will be updated in a timely manner according to the increasing/decreasing number of action items, and the coordination situation.

The panel will hold fortnightly the conference calls with agenda and materials ready for discussion and all working materials will be shared online.

It is expected that the panel will have a few face-to-face meetings, which will take place probably during phases 2 and 3.

3.2 Sources for funding travel and logistics

Although the members of the Latin panel will be volunteers and provide their time and expertise on a purely voluntary basis, issues of logistics such as travel and stay necessitated by members in a

face-to-face meeting as well as support for conference calls, and postings on the website will require support.

3.3 Need for ICANN provided advisors

The panel might need ICANN support for stage 1 **Developing Code Point Repertoire**. Support might be necessary in analyzing the code point repertoire for all languages of the script using community, a sample of which is presented in Appendix A.

Some coordination might also be needed with Greek and Cyrillic GP.

4. References

1. Frakes, J., et al., "Considerations in the use of the Latin script in variant internationalized top-level domains: Final report of the ICANN VIP Study Group for the Latin script". Los Angeles, Calif.: ICANN, October 2011). <http://archive.icann.org/en/topics/new-gtlds/latin-vip-issues-report-07oct11-en.pdf>
2. Blanchet, M., et al. "Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR". Los Angeles, Calif.: ICANN, April 2015. <https://community.icann.org/download/attachments/43989034/Guidelines%20for%20LGR.pdf>
3. "Considerations for Designing a Label Generation Ruleset for the Root Zone". Los Angeles, Calif.: ICANN, April 2015. <https://community.icann.org/download/attachments/43989034/Considerations%20for%20LGR.pdf>
4. "Requirements for LGR Proposals". Los Angeles, Calif.: ICANN, April 2015. <https://community.icann.org/download/attachments/43989034/Requirements%20for%20LGR%20Proposals.pdf>
5. Integration Panel "Root Zone Label Generation Rules – LGR-1 Overview and Summary". Los Angeles, Calif.: ICANN, February 2016.
6. Common Locale Data Repository. www.unicode.org/cldr/charts/28/summary/root.html
7. www.ethnologue.com
8. www.omniglot.com
9. www.scriptsource.org
10. https://en.wikipedia.org/wiki/History_of_the_Latin_alphabet

11. https://en.wikipedia.org/wiki/Latin_script
12. Maximal Starting Repertoire [MSR-2]. <https://www.icann.org/resources/pages/reports-2013-04-03-en>
13. Katz, Leonard & Ram Frost. 1992. The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis. *Haskins Laboratories Status Report on Speech Research* 111/112. 147–160.
14. Klensin, J., “Internationalized Domain Names in Applications (IDNA): Definitions and Document Framework” = RFC 5890 (2010). <http://tools.ietf.org/html/rfc5890>
15. Fältström, P., ed., “The Unicode Code Points and Internationalized Domain Names for Applications (IDNA)” = RFC 5892 (2010). <http://tools.ietf.org/html/rfc5892>
16. Hoffman, P., et al., “Terminology Used in Internationalization in the IETF” (2011). = RFC 6365 <http://tools.ietf.org/html/rfc6365>
17. [Procedure] Sullivan, A., et al., “Procedure to develop and maintain Label Generation Rules for the Root Zone in respect of IDNA labels” (Marina del Rey, California: ICANN, March 2013). <https://www.icann.org/en/system/files/files/lgr-procedure-20mar13-en.pdf>.
18. Bendor-Samuel, John. 1996. African languages. In Peter T. Daniels & William Bright (eds.), *The World's Writing Systems*, 689–691. New York, Oxford: Oxford University Press.
19. Hartell, Rhonda L. (ed.). 1993. *Alphabets de Langues Africaines*. Dakar: UNESCO, Bureau Régional de Dakar; Societe Internationale de Linguistique.
20. IDNA 2008. See RFCs 5890, 5891, 5892, 5893, 5894 and 5895. <https://tools.ietf.org/html/rfc5890> <https://tools.ietf.org/html/rfc5891> <https://tools.ietf.org/html/rfc5892> <https://tools.ietf.org/html/rfc5893> <https://tools.ietf.org/html/rfc5894> <https://tools.ietf.org/html/rfc5895>
21. ISO 15924 “Codes for the representation of names of scripts”. <http://unicode.org/iso15924/iso15924-codes.html>
22. [RFC1034] Mockapetris, P., "Domain names - concepts and facilities", STD 13, RFC 1034, November 1987.
23. [RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, November 1987.
24. [Armenian Proposal] Armenian Generation Panel (GP). Proposal for an Armenian Script Root Zone Label Generation Rules”, 22 Jul 2015. <https://www.icann.org/public-comments/proposal-armenian-lgr-2015-07-22-en>
25. [SIL-Ethnologue] Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.). 2014. *Ethnologue: Languages of the World*, Seventeenth edition. Dallas, Texas: SIL International.

Online version available as <http://www.ethnologue.com>.

26. [Unicode Standard Annex #15]: <http://unicode.org/reports/tr15/>
27. [Punycode]: Costello, A., "Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)", RFC 3492, March 2003, <https://tools.ietf.org/html/rfc3492>