

# Romanization systems and their implementation

Peeter Päll

20th Session of ECSEED

Zagreb, 9 February 2011

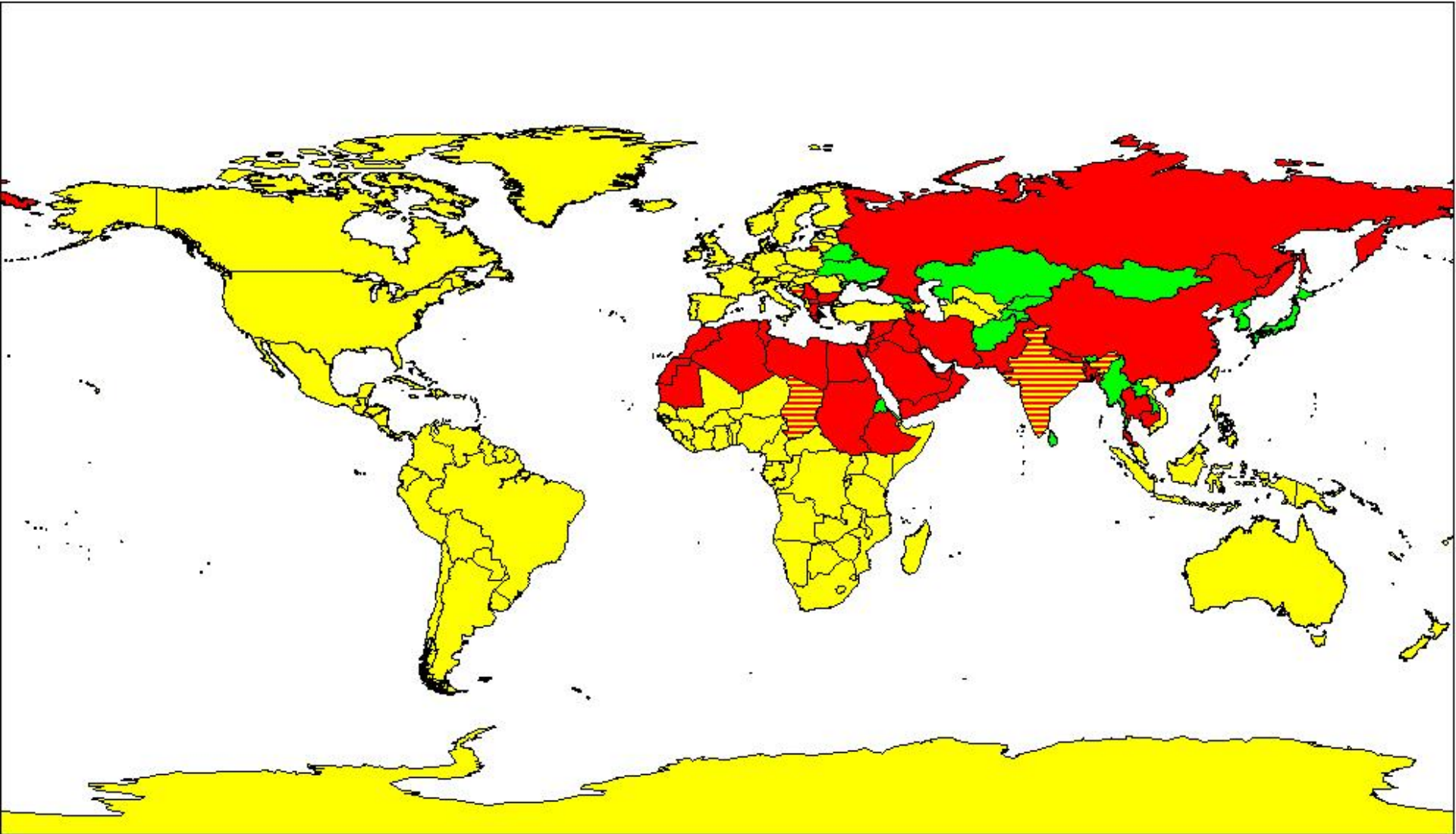
# Mandate of the WG on Romanization Systems

- The basic mandate of the WG consists of considering and reaching agreement on a romanization system for geographic names proposed by a (donor) country. The process of adoption of such a system by the United Nations involves the following prerequisites:
  - Sufficient time for appropriate consultations and an expression of all views on technical matters between the sponsoring country and the working group members.
  - The romanization system should be based on sound scientific principles, generally reversible, and implemented on cartographic products (maps and charts) by the proposing country.
  - The new system is referred to the next U.N. Conference on the Standardization of Geographical Names in form of a resolution which, in turn, is voted by the Conference. When the resolution is adopted, the subject romanization system becomes the United Nations standard.

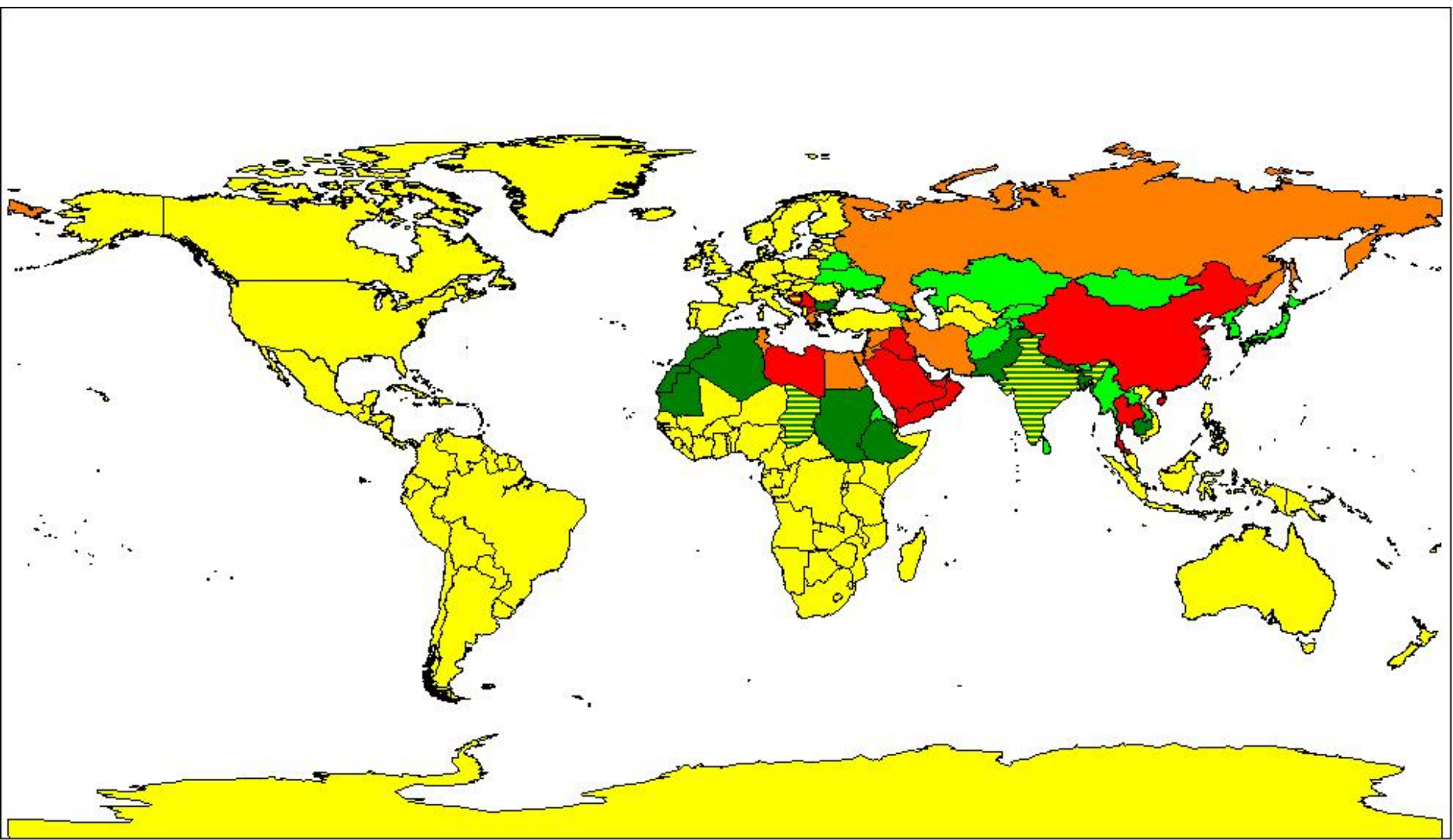
# Current standing: approved (28)

- **Implemented** (at least to some extent): Chinese, Greek\*, Hebrew\*, Macedonian Cyrillic\*, Mongolian (in China), Persian\*, Russian\*, Serbian, Thai, Tibetan, Uighur (11)
- **Currently under revision:** Amharic, Arabic, Bulgarian, Khmer (4)
- **Never implemented:** Assamese, Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Nepali, Oriya, Punjabi, Tamil, Telugu, Urdu (13)

# Romanization systems by status



# Romanization systems by implementation



# Current standing: on the agenda (17)

- **National romanization systems:**
  - Byelorussian, Dzongkha, Georgian, Japanese, Korean, Maldivian, Ukrainian (7)
- **Other romanization systems available** (mainly BGN/PCGN):
  - Armenian, Burmese, Kazakh, Kirghiz, Lao, Mongolian (Cyrillic), Pashto, Tajik, Tigrinya (9)
- **No romanization systems available:**
  - Sinhalese (1)

# Implementation

- Romanization as applied in
  - maps
  - listings (gazetteers)
  - texts
- Automatic conversion
  - Unicode Transliteration Guidelines
    - <http://cldr.unicode.org/index/cldr-spec/transliteration-guidelines>
  - Application in Google Maps

# Unicode Transliteration Guidelines

## Introduction

*This document describes guidelines for the creation and use of CLDR transliterations. Preliminary [charts](#) are available for the available transliterations -- be sure to read the known issues there. Please file any feedback on this document or those charts at [Locale Bugs](#).*

Transliteration is the general process of converting characters from one script to another, where the result is roughly phonetic for languages in the target script. For example, "Phobos" and "Deimos" are transliterations of Greek mythological "Φόβος" and "Δεῖμος" into Latin letters, used to name the moons of Mars.

Transliteration is *not* translation. Rather, transliteration is the conversion of letters from one script to another without translating the underlying words. The following shows a sample of transliteration systems:

Sample Transliteration Systems

Source	Translation	Transliteration	System
Αλφαβητικός	<i>Alphabetic</i>	Alphabētikós	Classic
		Alfavjtikós	UNGEGN
しんばし	<i>new bridge (district in Tokyo)</i>	shimbashi	Hepburn
		sinbasi	Kunrei
яйца Фаберже	<i>Fabergé eggs</i>	yaytsa Faberzhe	BGN/PCGN
		jajca Faberže	Scholarly
		ajca Faberže	ISO

*Display. Some of the characters in this document may not be visible in your browser, and with some fonts the diacritics will not be correct.*



Devanagari	Romanization	Notes
सेङ्गुप्त	Sēngupta	
सेनगुप्त	Sēnagupta	The final 'a' is not pronounced
मोनिक	Monika	
मोनिच	Monica	The 'c' is pronounced "ch"

## Others

Unicode CLDR provides other transliterations based on the [U.S. Board on Geographic Names](#) (BGN) transliterations. These are currently unidirectional — to Latin only. The goal is to make them bidirectional in future versions of CLDR.

Other transliterations are generally based on the [UNGEGN: Working Group on Romanization Systems](#) transliterations. These systems are in wider actual implementation than most ISO standardized transliterations, and are published freely available on the web (<http://www.eki.ee/wgrs/>) and thus easily accessible to all. The UNGEGN also has good documentation. For example, the [UNGEGN Arabic Tables](#) not only presents the UN system, but compares it with the BGN/PCGN 1956 system, the I.G.N. System 1973, ISO 233:1984, the royal Jordanian Geographic Centre System, and the Survey of Egypt System.

## Submitting Transliterations

If you are interested in providing transliterations for one or more scripts, file an initial bug report at [Locale Bugs](#). The initial bug should contain the scripts and or languages involved, and the system being followed (with a link to a full description of the proposed transliteration system), and a brief example. The proposed data can also be in that bug, or be added in a Reply to that bug. You can also file a bug in [Locale Bugs](#) if you find a problem in an existing transliteration.

For submission to CLDR, the data needs to be supplied in the correct XML format or in the ICU format, and should follow an accepted standard (like UNGEGN, BGN, or others).

- The format for rules is specified in [Transform Rules](#). It is best if the results are tested using the [ICU Transform Demo](#) first, since if the data doesn't validate it would not be accepted into CLDR.
- As mentioned above, even if a transliteration is only used in certain countries or contexts CLDR can provide for them with different variant tags.
- For comparison, you can see what is currently in CLDR in the [transforms](#) folder online. For example, see [Hebrew-Latin.xml](#).
- Script transliterators should cover every character in the exemplar sets for the CLDR locales using that script.



Мьюир Бич  
Muir Beach

Сосалито  
Sausalito

Сан-Франциско  
San Francisco

1

Дейли Сити  
Daly City

Бродмур  
Broadmoor

Колма  
Colma

Брисбен  
Brisbane

Пасифика  
Pacifica

Юг Сан  
Франциско  
South San Francisco

Милбро  
Milbrae

Монтара  
Montara

Мосс Бич  
Moss Beach

Эль  
Гранада  
El Granada

Халф  
Мун Бэй  
Half Moon Bay

Окленд  
Oakland

Аламида  
Alameda

80

580

880

Сан  
Леандро  
San Leandro

Кастро  
Валли  
Castro Valley

Фэрвью  
Fairview

Сан-Франциско  
San Francisco Bay

Хейворд  
Hayward

Данвилл  
Danville

Дьябло  
Diablo

Сан-Рамон  
San Ramon

Блэкхок-Камино  
Тассаджара  
Blackhawk-Camino  
Tassajara

Дублин  
Dublin

Плезантон  
Pleasanton

Ливермор  
Livermore

580

Юнион  
Сити  
Union City

Санол  
Sunol

Государственная  
зона отдыха  
Лейк Дель Валле  
Lake Del Valle State  
Recreation Area

Фремонт  
Fremont

680

84

Ньюарк  
Newark

Санол Риджинал  
Уайлдернесс  
Sunol Regional  
Wilderness

Олон Риджинал  
Уайлдернесс  
Ohlone Regional  
Wilderness

Сан-Матео  
San Mateo

Фостер  
Сити  
Foster City

Сан Карлос  
San Carlos

Редвуд  
Сити  
Redwood City

Эмералд  
Лейк Хилс  
Emerald Lake Hills

Менло Парк  
Menlo Park

Станфорд  
Stanford

Пало-Альто  
Palo Alto

Милпитас  
Milpitas

Маунтин-Вью  
Mountain View

Саннивейл  
Sunnyvale

680

880

Восток  
Футхилс  
East Foothills

Портола  
Валли

Лос Альтос  
Хилс

Лойола