INTRODUCTION

This publication contains all of the romanization systems, tables of correspondences, and Roman-script spelling conventions that are currently approved by the U.S. Board on Geographic Names (BGN) and its British counterpart, the Permanent Committee on Geographical Names for British Official Use (PCGN).   It therefore supersedes the *Transliteration Guide* of 1961; the *Romanization Guide* of 1964, 1967, and 1972; and the publication *Romanization Systems and Roman Script Spelling Conventions* of 1994.   Each romanization system and spelling convention presented is identified as being a BGN/PCGN system or a BGN/PCGN agreement, with the date of its joint adoption by the BGN and PCGN indicated in most cases. (See Table 1.)

Within the U.S. Government, BGN/PCGN romanization systems and agreements are used primarily for the purpose of establishing standardized Roman-script spellings of those foreign geographic names that are written in non-Roman scripts or in Roman alphabets that contain special letters.   Geographic names that have been romanized and names originally written in Roman script are made available for general use on the GEOnet Names Server, an on-line service of the National Geospatial-Intelligence Agency.   This database, which covers virtually every foreign country in the world, provides information as to the name, type, and location of every geographic feature listed, as well as variant spellings of names for finding purposes.

In most cases, familiarity with the writing system of a given language is all that is needed in order to apply the appropriate BGN/PCGN romanization system or agreement correctly.   In some cases, however, a thorough knowledge of both the language and its writing system is an absolute requirement. The latter category includes the systems for Arabic, Hebrew, Persian, and Pashto, i.e., systems for languages in which vowels are not ordinarily represented in the script.   The BGN/PCGN Romanization systems for those languages and for the other languages represented in this publication generally contain elements of transliteration – the process of recording the graphic symbols of one writing system in terms of the corresponding graphic symbols of a second writing system – and of transcription – the process of recording the phonological and/or morphological elements of a language in terms of a specific writing system.*

There are a number of principles that BGN and PCGN agreed should be considered when devising a romanization system; it should be as reversible and parsimonious as possible, should include diacritical signs to a minimal degree, and should neither be a guide to pronunciation nor a language treatise.   Some BGN/PCGN romanization systems, e.g., those for Bulgarian and Georgian, exhibit a high degree of reversibility; i.e., the Roman letters that serve as the equivalents of the non-Roman characters of the source script may be converted to the original characters almost unambiguously. Other BGN/PCGN systems and agreements, e.g., those for Amharic and Korean, are not easily reversible.   The BGN/PCGN romanization system for Thai provides an extreme example of a

non-reversible system; in that system, the Roman letter **t** is used to represent a total of nineteen different Thai characters in syllable-final position. The Thai romanization system, therefore, can be said to be undifferentiated, since it contains several instances of a single Roman letter or letter combination serving as the equivalent of more than one Thai character.

_____

* These definitions were agreed upon in 1971 by the U.N. Working Group on a Single Romanization System for Each Non-Roman Writing System and were included in that group's report in the U.N. document, <u>Second United Nations Conference on the Standardization of Geographical Names</u>, London, 10-13 May 1972, vol. II, p. 115.

The Roman letters and letter combinations that are shown as equivalents of the non-Roman characters in the BGN/PCGN romanization systems generally reflect the letters and letter combinations that are used in English orthography. In many cases, however, the number of Roman-letter equivalents needed for a particular system exceeds the number of appropriate letters and letter combinations available in English orthography. As a result, several Roman letters may be shown with diacritical marks in order to provide the necessary differentiation of graphic symbols and insure proper reversibility. In the Persian alphabet, for example, there are four different characters that are pronounced like the letter z in English. In order to differentiate the romanizations of those four characters, the BGN/PCGN system for Persian utilizes the ordinary letter z and three z's with diacritical marks, i.e., ẕ, z̄, and ẓ. In addition to their use in Roman-letter equivalents in this publication, diacritical marks are used with Roman letters and with non-Roman characters in many languages of the world: For example, the cedilla (¸) is used with the letter c to form ç in French, and the breve (˘) is used with the letter И to form Й in Russian. Diacritical marks are just as important as the basic letters and characters of any orthography or romanization system, and, therefore, should not be omitted. However, in cases where the user is unable to reproduce correctly the appropriate letter-diacritic combination, omission of the diacritical marks is permissible. Modifying marks that occur internally in both Roman letters and non-Roman characters, e.g., the horizontal bar in the Croatian letter đ and the horizontal bar in the Kazakh Cyrillic character Ұ, are not generally considered to be diacritical marks but are just as significant and, therefore, should always be retained.

It should be noted that for clarity of presentation and ease of reference the terms *character* and *letter* have been used in a mutually exclusive way throughout this publication. The term *character* has been used to refer to a graphic symbol used only in a non-Roman-script writing system, thereby restricting the term *letter* to a graphic symbol used only in a Roman-script writing system or in a set of romanization equivalents.

Finally, it may well be pointed out that although the romanization systems and agreements contained in this publication have been approved by the BGN and the PCGN for application to geographic names, some or all of the systems may prove similarly applicable to personal names and to text and indeed have frequently been used for such purposes by organizations both within the U.S.

and the U.K.

Please refer to the appendices for Unicode values for letters used in BGN/PCGN romanization systems, as well as for hints on optimizing computer software and operating systems to allow display of Unicode characters and letters. Please refer to the BGN website for updates to this guide and additional information: http://geonames.usgs.gov, as well as to the PCGN Website: http://www.pcgn.org.uk. Information on other transliteration systems for toponyms is also maintained on the Website of the United Nations Group of Experts on Geographical Names Working Group on Romanization Systems: http://www.eki.ee/wgrs/.