

L2/09-030

Title: Proposal to publish *kHanyuPinyin* data in Unihan**Authors:** Richard Cook, John Jenkins**Date:** 23 January 2009

This is a proposal to publish an important new field of pīnyīn readings in future Unihan releases. This important data was originally input by 井作恆 Jǐng Zuòhéng, proofed by 聃媽歌 Magda Danish (using software donated by 文林 Wénlín Institute, Inc. and tables prepared by 曲理查 Qū Lǐchá), and proofed again and prepared for the Unicode Consortium by 曲理查 Qū Lǐchá. Suffice it to say, this data is important, and has really been proofed! A draft of the property metadata is as follows:

Tag ⇒ kHanyuPinyin

Description ⇒ The 漢語拼音 Hànyǔ Pīnyīn reading(s) appearing in the edition of 《漢語大字典》 Hànyǔ Dà Zìdiǎn (HDZ) specified in the “kHanYu” property description (q.v.). Each mapping “ABCDE.XYZ” (given in the same form as in “kHanYu”) is followed by a colon, followed by a space-separated list of one or more pīnyīn readings. Where multiple pīnyīn readings are associated with a given mapping, these are ordered as in HDZ (for the most part reflecting relative commonality). Each Unihan Hànzì may have multiple “mapping: pīnyīn” sets, also space-separated: there may be variation among the lists of pīnyīn readings in each such set.

For example, the “kHanyuPinyin” value for 鹵 U+5364 is “10093.130: xī lǚ 74609.020: lǚ xī”. This means that 鹵 U+5364 is found in “kHanYu” at entries 10093.130 and 74609.020. The former entry has the two pīnyīn readings xī and lǚ (in that order), whereas the latter entry has the readings lǚ and xī (reversing the order).

-- Release Notes --

This data set includes readings for 34,131 distinct HDZ Hànzì, 34,302 HDZ references, and 1,457 distinct pīnyīn syllables.

Category ⇒ Readings**CompletionLevel** ⇒ Complete**Delimiter** ⇒ space**Syntax** ⇒ /^(\\d{5}\\d{2}0)*(\\d{5}\\d{2}0:)[^\\d\\.]+\$/**Status** ⇒ Provisional**Records** ⇒ 34131**Distinct** ⇒ 34131