

Package ‘ISOCodes’

January 26, 2010

Version 0.3-0

Date 2010-01-26

Title Selected ISO codes

Author Christian Buchta and Kurt Hornik

Maintainer Kurt Hornik <Kurt.Hornik@R-project.org>

Description ISO language, territory, currency, script and character codes. Provides ISO 639 language codes, ISO 3166 territory codes, ISO 4217 currency codes, ISO 15924 script codes, and the ISO 8859 character codes as well as the UN M.49 area codes.

Depends R (>= 2.5.0)

Encoding UTF-8

License GPL-2

Repository CRAN

Date/Publication 2010-01-26 10:45:45

R topics documented:

ISO_15924	2
ISO_3166	3
ISO_4217	4
ISO_639	4
ISO_8859	6
UN_M.49	7
Index	9

`ISO_15924`*ISO 15924 Script Codes*

Description

International Organization for Standardization (ISO) codes for the representation of names of scripts (writing systems, “a set of graphic characters used for the written form of one or more languages”).

Usage

```
data("ISO_15924")
```

Format

`ISO_15924` is a data frame with variables

`Alpha_4` a character vector with the 4-letter (alpha-4) ISO 15924 script codes.

`Numeric` a character vector with numeric script codes providing some measure of mnemonicity.

`Name` a character vector with the (English) script names.

`PVA` a character vector with the Property Value Alias defined by [Unicode](#) (ISO 10646) (if available).

`Date` a [Date](#) object with the date the script was registered.

Details

The following number ranges are used for the numeric codes:

000-099	Hieroglyphic and cuneiform scripts
100-199	Right-to-left alphabetic scripts
200-299	Left-to-right alphabetic scripts
300-399	Alphasyllabic scripts
400-499	Syllabic scripts
500-599	Ideographic scripts
600-699	Undeciphered scripts
700-899	(unassigned)
900-999	Private use, aliases, special codes

Source

<http://www.unicode.org/iso15924/>

References

http://en.wikipedia.org/wiki/ISO_15924

ISO_3166

ISO 3166 Country Codes

Description

International Organization for Standardization (ISO) codes for the representation of names of countries and their subdivisions. Consists of three parts. Part 1: Country codes, defines codes for country and dependent area names. Part 2: Country subdivision code, defines codes for the principal subdivisions of a country or dependent area. Part 3: Code for formerly used names of countries, defines codes for superseded ISO 3166-1 codes.

Usage

```
data ("ISO_3166_1")
data ("ISO_3166_2")
data ("ISO_3166_3")
```

Format

ISO_3166_1 is a character frame with variables `Alpha_2`, `Alpha_3`, and `Numeric` (giving the two-letter, three-letter and three-digit numeric country codes) and `Name`, `Official_name`, and `Common_name` (giving the respective names).

ISO_3166_2 is a character frame with variables `Code`, `Type`, `Name`, `Country`, and `Parent`, giving the code, type and name of the subdivision, the country it belongs to, and a parent subdivision in case this is different from the country.

ISO_3166_3 is a character frame with variables `Alpha_4` (the 4-letter code of the retired country), `Alpha_3`, `Numeric`, and `Name` (the original 3166-1 code elements of the country), and `Date_withdrawn` and `Comment`.

Source

Converted from XML files provided by Debian's **iso-codes** package (<http://pkg-isocodes.alioth.debian.org/>).

References

http://en.wikipedia.org/wiki/ISO_3166

ISO_4217

ISO 4217 Currency Codes

Description

International Organization for Standardization (ISO) codes for the representation of currencies.

Usage

```
data("ISO_4217")
data("ISO_4217_Historic")
```

Format

ISO_4217 is a character frame with variables `Letter`, `Numeric` and `Currency`, giving the 3-letter and 3-digit codes and the names of the respective currency.

ISO_4217_Historic is a character frame with the currency codes retired from ISO 4217, containing variable `Date_withdrawn` in addition to the variables in ISO_4217.

Source

Converted from XML files provided by Debian's **iso-codes** package (<http://pkg-isocodes.alioth.debian.org/>).

References

http://en.wikipedia.org/wiki/ISO_4217

ISO_639

ISO 639 Language Codes

Description

International Organization for Standardization (ISO) codes for the representation of languages. Consists of four parts, with more parts work in progress. ISO 639-1 consists of 185 two-letter (alpha-2) codes used to identify the world's major languages. ISO 639-2 has three-letter (alpha-3) codes for 485 languages. ISO 639-3 extends the ISO 639-2 alpha-3 codes with an aim to cover all known natural languages. ISO 639-5 defines alpha-3 codes for language families.

Usage

```
data("ISO_639_2")
data("ISO_639_3")
data("ISO_639_3_Retirements")
data("ISO_639_5")
```

Format

`ISO_639_2` is a character data frame with variables `Alpha_3_B` and `Alpha_3_T` (the ISO 639-2 bibliographic and terminological codes), `Alpha_2` (the corresponding ISO 639-1 alpha-2 code if available), and `Name`.

`ISO_639_3` is a data frame with the following variables:

`Id`: a character vector with the ISO 639-3 3-letter (alpha-3) identifiers.

`Part2B`: a character vector with the equivalent ISO 639-2 B-code identifiers of the bibliographic applications code set (if existent).

`Part2T`: a character vector with the equivalent ISO 639-2 T-code identifiers of the terminology applications code set (if existent).

`Part1`: a character vector with the equivalent ISO 639-1 2-letter (alpha-2) identifiers (if existent).

`Scope`: a factor with levels "I" (Individual), "M" (Macrolanguage) and "S" (Special).

`Type`: a factor with levels "L" (Living languages), "E" (Extinct languages), "A" (Ancient languages), "H" (Historic languages), "C" (Constructed languages), and "S" (Special).

`Name`: a character vector with the reference language names.

`Comment`: a character vector with a comment relating to one or more of the other variables.

`ISO_639_3_Retirements` is a data frame giving the languages retired from ISO 639-3, with variables:

`Id`: a character vector with the retired codes

`Ret_Reason`: a factor with levels "C" (change), "D" (duplicate), "N" (non-existent), "S" (split), and "M" (merge).

`Change_To`: a character vector which in the cases of C, D, and M gives the identifier to which all instances of the `Id` should be changed.

`Ret_Remedy`: a character vector with instructions for updating an instance of the retired (split) identifier.

`Effective`: a `Date` object giving the date the retirement became effective.

`ISO_639_5` is a data frame with the following variables:

`Id` a character vector with the 3-letter (alpha-3) ISO 639-5 identifiers.

`English_Name` the family names in English.

`French_Name` the family names in French.

`Part2` a factor indicating how the family relates to 639-2, with levels "g" (group: consists of several related languages), "r" (rest group: a group of several related languages, from which some specific languages have been excluded), or "" (no 639-2 code).

`Hierarchy` an indication of which other language families or groups the current language family or group is a member of (given as 639-5 ids separated by ' : ').

Details

While most languages are given one code by the ISO 639-2 standard, twenty-two of the languages described have two three-letter codes, a “bibliographic” code (ISO 639-2/B, B-code), which is derived from the English name for the language and was a necessary legacy feature, and a “terminological” code (ISO 639-2/T, T-code), which is derived from the native name for the language. The range ‘qaa’ to ‘qtz’ is reserved for local use.

ISO 639-3 is a superset of ISO 639-1 and of the individual languages in ISO 639-2. ISO 639-1 and ISO 639-2 focused on major languages, most frequently represented in the total body of the world’s literature. Since ISO 639-2 also includes language collections, whereas Part 3 does not, ISO 639-3 is not a superset of ISO 639-2. Where B and T codes exist in ISO 639-2, ISO 639-3 uses the T-codes.

ISO 639-2 contains codes for some individual and group languages and so any code in it is either in 639-3 or 639-5; 639-5 families may be missing from 639-2.

Source

<http://www.loc.gov/standards/iso639-2/> for ISO 639-2;
<http://www.sil.org/iso639-3/download.asp> for ISO 639-3;
<http://www.loc.gov:8081/standards/iso639-5/> for ISO 639-5.

References

http://en.wikipedia.org/wiki/ISO_639

ISO_8859

ISO 8859 Character Codes

Description

International Organization for Standardization (ISO) codes for 8-bit character encodings for use by computers. The data set gives the maps of the characters to Unicode (i.e., the respective ISO 10646 codes).

Usage

```
data ("ISO_8859")
```

Format

A character array of dimension $256 \times 15 \times 3$, with the first dimension corresponding to the character codes from 0 to 255 (0x00 to 0xff), the second to the parts of the ISO 8859 standard, and the third to the [Unicode](#) (ISO 10646) code and name, and the respective character.

Details

The ISO 8859, more formally ISO/IEC 8859, standard is divided into numbered, separately published parts, such as as ISO/IEC 8859-1, ISO/IEC 8859-2, etc., each of which may be informally referred to as a standard in itself. There are currently 15 parts as of 2006 excluding the abandoned ISO/IEC 8859-12 standard:

Part 1	Latin-1 Western European
Part 2	Latin-2 Central European
Part 3	Latin-3 South European
Part 4	Latin-4 North European
Part 5	Latin/Cyrillic
Part 6	Latin/Arabic
Part 7	Latin/Greek
Part 8	Latin/Hebrew
Part 9	Latin-5 Turkish
Part 10	Latin-6 Nordic
Part 11	Latin/Thai
Part 13	Latin-7 Baltic Rim
Part 14	Latin-8 Celtic
Part 15	Latin-9
Part 16	Latin-10

Source

<ftp://ftp.unicode.org/Public/MAPPINGS/ISO8859>.

References

http://en.wikipedia.org/wiki/ISO_8859

Examples

```
## ISO 8859 characters at position 200 (number 199).
data("ISO_8859")
ISO_8859[200, , ]
```

UN_M.49

UN M.49 Area Codes

Description

Country and area code classifications (M49) from the United Nations Statistics Division.

Usage

```
data("UN_M.49_Countries")
data("UN_M.49_Regions")
```

Details

UN M.49 is a standard for area codes used by the United Nations for statistical purposes. Each area code is a 3-digit number which can refer to a wide variety of geographical, political, or economic regions, like a continent, a country, or a specific group of developing countries.

UN_M.49_Countries contains the codes for countries and areas as a character frame with variables Code, Name and ISO_Alpha_3 giving the 3-letter UN M.49 code and name and the respective alpha-3 ISO 3166 code.

UN_M.49_Regions contains the codes for the composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings as a data frame with the character variables Code, Name, Parent and Children giving the 3-letter UN M.49 code, name, the code of the parent area and the codes of children areas separated by ', ', respectively, and variable Type, a factor with levels "Region" or "Grouping".

Source

<http://unstats.un.org/unsd/methods/m49/m49.htm>

Examples

```
## Name and codes of countries in Southern Europe:
data("UN_M.49_Regions")
data("UN_M.49_Countries")
region <- subset(UN_M.49_Regions, Name == "Southern Europe")
codes <- unlist(strsplit(region$Children, ", "))
subset(UN_M.49_Countries, Code %in% codes)
```

Index

*Topic **datasets**

ISO_15924, [1](#)

ISO_3166, [2](#)

ISO_4217, [3](#)

ISO_639, [4](#)

ISO_8859, [6](#)

UN_M.49, [7](#)

Date, [2](#), [5](#)

ISO_3166, [7](#)

ISO_15924, [1](#)

ISO_3166, [2](#)

ISO_3166_1 (*ISO_3166*), [2](#)

ISO_3166_2 (*ISO_3166*), [2](#)

ISO_3166_3 (*ISO_3166*), [2](#)

ISO_4217, [3](#)

ISO_4217_Historic (*ISO_4217*), [3](#)

ISO_639, [4](#)

ISO_639_2 (*ISO_639*), [4](#)

ISO_639_3 (*ISO_639*), [4](#)

ISO_639_3_Retirements (*ISO_639*), [4](#)

ISO_639_5 (*ISO_639*), [4](#)

ISO_8859, [6](#)

UN_M.49, [7](#)

UN_M.49_Countries (*UN_M.49*), [7](#)

UN_M.49_Regions (*UN_M.49*), [7](#)

Unicode, [2](#), [6](#)