



180

**Addition of Address Form Data to Unicode CLDR**

2011.04.26

**Status:**

Open

**Description of Issue:**

The Unicode Consortium is considering the addition to CLDR of address form metadata. This metadata is intended for presenting a form for users to fill in with address data. The format and data is being donated by Google. The consortium is soliciting feedback on these changes. Feedback should be submitted as comments to <http://unicode.org/cldr/trac/ticket/3572>.

**Background**

Google's address widget metadata contains information on how address fields should be laid out, and how to format and validate addresses entered by the user. The metadata has been exposed for the open-source community through [an Appengine service](#). However, it is impossible for the open-source community to propose or make any change to the data.

The CLDR project is the appropriate place to host the address metadata for the following reasons:

1. It is by far the largest and most extensive standard repository of locale data, with a well-established process of vetting the data to keep it to its highest quality.
2. It currently lacks detailed information on how address fields should be laid out, and how to format and validate addresses entered by the user.
3. Contributing this data to CLDR ensures wider adoption of the data, which in turn improves the quality of this data.

**Existing Address Metadata support in CLDR**

Currently address metadata exists in several places in CLDR:

1. common/main/[Locale].xml: contains **localized country names** under ldml/localeDisplayNames /territories
2. supplemental/postalCodeData.xml: contains regex of postal code for 158 countries. The information in this file was contributed by Google in 2009, and contains a subset of the data we are proposing to contribute in this document.

**Future Plans**

We plan to later follow up with a separate proposal to contribute translations of different address fields and provinceNameType.

**References:**

1. Address and Phone Number Internationalization: Standards, Technologies and Best Practices, the 34th Internationalization and Unicode Conference. ([presentation slides](#))

**Detailed Proposal****Proposed changes in CLDR**

Deprecate postalCodeData.xml, and add the following file to common/supplemental:

1. **addressformdata.xml** - One file that contains country-level address information for countries/regions in the world.

## Example contents

```
<addressFormData>
<postalCountry iso3166="TW">
  <layout order="LargeToSmall">%Z%n%S%C%n%A%n%O%n%N</layout>
  <layout order="SmallToLarge">%N%n%O%n%A%n%C, %S %Z</layout>
  <requiredFields>ACSZ</requiredFields>
  <postalCodeValidationRule>\d{3}(\d{2})?</postalCodeValidationRule>
  <postalCodeType>postal</postalCodeType>
  <provinceNameType>county</provinceNameType>
  <centralPostOfficeURL>http://www.post.gov.tw</centralPostOfficeURL>
</postalCountry>
<postalCountry iso3166="US">
  <layout order="SmallToLarge">%N%n%O%n%A%n%C %S %Z</layout>
  <uppercaseFields>CS</uppercaseFields>
  <requiredFields>ACSZ</requiredFields>
  <postalCodeValidationRule>\d{5}([ \-]\d{4})?</postalCodeValidationRule>
  <postalCodeType>zip</postalCodeType>
  <provinceNameType>state</provinceNameType>
  <centralPostOfficeURL>http://www.usps.com</centralPostOfficeURL>
</postalCountry>
</addressFormData>
```

## Detailed Breakdown of elements

### 1. <layout order=..>

*Required/Optional*

Optional. Default value: %N%n%O%n%A%n%C

*Meaning*

Layout of address fields in the order specified *order*. It encodes how different fields should be laid out together for a particular country. There are two possible orders: LargeToSmall lays out larger territorial unit before smaller ones, while SmallToLarge does the reverse. The order is language dependent and which order to use is defined in the locale specific files under common/main. Only a few countries have both orders commonly used, and therefore specified here. Most of the countries only have one order specified.

Each address field is denoted by a "%" character following by a character to identify a field:

- N: Name (The formatting of names for this field is outside of the scope of the address elements.)
- O: Organization
- A: Address Lines (2 or 3 lines address)
- D: District (Sub-locality): smaller than a city, and could be a neighbourhood, suburb or dependent locality in the UK.
- C: City (Locality)
- S: State (Administrative Area)
- Z: ZIP Code / Postal Code
- X: Sorting code, for example, CEDEX as used in France
- n: newline

Note the fields may mean slightly different things in different countries. **This element is useful when you need to layout address fields for users to enter their address. However, it might not be possible to use this directly to format the address the user entered, because some of the address fields are optional.** In this case, an address formatter is needed to carefully remove formatting characters surrounding an address field when it is empty. Specifying rules to implement such an address formatter is beyond the scope of this document.

Note some of the fields specified may be optional when an address is laid out for in-country use, but required for international use. In such cases, the fields are always specified in the value of the "format" attribute, because it won't lead to any misunderstanding to our best knowledge. Also note the country field is not defined here. The reason is that a country has to be specified before the

value in the *layout* could be used to layout the rest of the address fields in the correct order for that country.

Examples:

Eric Schmidt	Name(N)
Google Inc.	Organization(O)
1600 Amphitheatre Parkway	Address Lines(A)
Mountain View, CA	City(C), State(S)
94043-1351	ZIP Code(Z)

Google Beijing	Organization(O)
Tsinghua Science Park Bldg 6	Address Lines(A)
No. 1 Zhongguancun East Road	Address Lines(A)
Haidian District	District(D)
Beijing 100084	City(C) Postal Code(Z)

Institut National d'Horticulture	Organization(O)
2 rue Lenôtre	Address Lines(A)
49045 Angers Cedex 01	Postal Code(Z) City(C) Sorting code(X)

## 2. <uppercaseFields>

*Required/Optional*

Optional. Default value: C

*Meaning*

Encodes which fields should be written in upper case. The attribute is a set of character that denote the fields, as described in the "format" attribute.

## 3. <requiredFields>

*Required/Optional*

Optional. Default value: AC

*Meaning*

Encodes which fields are required for a postal address. The attribute is a set of character that denote the fields, as described in the "format" attribute.

## 4. <postalCodePrefix>

*Required/Optional*

Optional.

*Meaning*

Contains the postal code prefix that might be used in some countries. E.g. "CH-" is sometimes used in Switzerland to prefix postal code. The prefix could be inserted in front of the "ZIP Code / Postal Code" field if it is present in the "format" attribute.

## 5. <postalCodeValidationRule>

*Required/Optional*

Optional.

*Meaning*

Contains a regular expression that specifies valid postal code.

## 6. <postalCodeType>

*Required/Optional*

Optional. Default value: postal

*Meaning*

Contains an enum that denotes the type of label for the postal code field. Currently, the valid values include:

1. postal
2. zip

## 7. <provinceNameType>

*Required/Optional*

Optional. Default value: ""

*Meaning*

Contains an enum that denotes the type of label for the "state" field. Currently, the valid values include:

1. **state** (Administrative Area for certain countries (e.g., US' California))
2. **province** (Administrative Area for certain countries (e.g., France's Champagne))
3. **prefecture** (Administrative Area for Japan (e.g., Hokkaido))
4. **parish** (Administrative Area for certain countries (e.g., Andorra's Canillo))
5. **island** (Administrative Area for certain countries (e.g., Bahama's Cat Island))
6. **emirate** (Administrative Area for United Arab Emirates (e.g., Abu Dhabi))
7. **department** (Administrative Area, as used for countries like Nicaragua (e.g., Boaco))
8. **county** (Administrative Area for the United Kingdom (e.g., Yorkshire))
9. **area** (Administrative Area for Hong Kong (e.g., Kowloon))
10. **do\_si** (Administrative Area for Korea (e.g., Gyeonggi-do or Busan-si))
11. **district** (Administrative Area of a country such as Nauru)

Note these values are enums, and no translation is included in this field.

## 8. <centralPostOfficeURL>

*Required/Optional*

Optional.

*Meaning*

A URL pointing to the postal office of the country that contains this element.

[Copyright © 1991-2011 Unicode, Inc.](#)

[All Rights Reserved](#)

[Terms of Use](#)

Last updated: - Tuesday, April 12, 2011 5:16:25 PM