

# **Unicode Common Locale Data Repository**

**By :**

**Swaran Lata**

**Country Manager, W3C India Office  
6, CGO Complex, Electronics Niketan  
New Delhi**

**E-mail : [slata@mit.gov.in](mailto:slata@mit.gov.in)**

# What is Locale?

A locale is the part of a user's environment that brings together information about how to handle data that is specific to the end user's particular country, language, or territory.

The data associated with locale provides support for

- formatting and parsing of
  - dates,
  - times,
  - numbers, and
  - currencies;
- measurement units,
- sort-order (collation),
- translated names for
  - time zones,
  - languages,
  - countries, and
  - scripts.

# Locale data is used for:

- **Software internationalization and localization**
- **Adapting software to the conventions of different languages for such common software tasks as:**
  - **formatting of dates, times, and time zones,**
  - **formatting numbers and currency values**
  - **sorting text**
  - **choosing languages or countries by name**

# **Localization (L10N)**

**Localization refers to the adaptation of a product, application or document content to meet the language, cultural and other requirements of a specific target market (a "locale").**

# **Internationalization (I18N)**

**Internationalization is the design and development of a product, application or document content that enables easy localization for target audiences that vary in culture, region, or language.**

# Common Locale Data Repository

- **CLDR is by far the largest and most extensive standard repository of locale data**
- **Provide locale data in the XML format for use in computer applications**
- **It facilitate information sharing among various industries**
- **Its goal is to gather basic linguistic information for various "locales"**
- **This data will be used for software I18N and L10N**
- **Basic purpose is the development of standards and technical reports**

## **The basic lists that CLDR gathers are:**

- 1. Dates formats**
- 2. Time Zones**
- 3. Number formats**
- 4. Currency formats**
- 5. Measurement System**
- 6. Collation Specification**
  - Sorting**
  - Searching**
  - Matching**
- 7. Translation of names for language, territory, script, time zones, currencies**
- 8. Script and characters used by a language.**

# XML in CLDR

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE Idml SYSTEM "../common/dtd/ldml.dtd">
<ldml>
<identity>
<version number="$Revision: 4753 $"/>
<generation date="$Date: 2010-02-26 00:32:00 -0600 (Fri, 26 Feb 2010) $"/>
<language type="hi"/>
</identity>
<days>
  <dayContext type="format">
  <dayWidth type="abbreviated">
  <day type="sun">रवि.</day>
  <day type="mon">सोम.</day>
  <day type="tue">मंगल.</day>
  <day type="wed">बुध.</day>
  <day type="thu">बृह.</day>
  <day type="fri">शुक्र.</day>
  <day type="sat">शनि.</day>
  </dayWidth>
```

```
<dayWidth type="wide">
<day type="sun">रविवार</day>
<day type="mon">सोमवार</day>
<day type="tue">मंगलवार</day>
<day type="wed">बुधवार</day>
<day type="thu">बृहस्पतिवार</day>
<day type="fri">शुक्रवार</day>
<day type="sat">शनिवार</day>
</dayWidth>
</dayContext>
<dayContext type="stand-alone">
<dayWidth type="narrow">
<day type="sun">र</day>
<day type="mon">सो</day>
<day type="tue">मं</day>
<day type="wed">बु</day>
<day type="thu">गु</day>
<day type="fri">शु</day>
<day type="sat">श</day>
</dayWidth>
</dayContext>
</days>
```

Contd.

```
<fields>
  <field type="era">
    <displayName>काल</displayName>
  </field>
  <field type="year">
    <displayName>साल</displayName>
  </field>
  <field type="month">
    <displayName>महीना</displayName>
  </field>
  <field type="week">
    <displayName>सप्ताह</displayName>
  </field>
</ldml>
```

Contd.

# Steps To be Followed for accessing XML data in CLDR

**Step 1 :** Visit <http://cldr.unicode.org/index/downloads/cldr-1-8-1>

**Step 2:** Following image will be shown

Release No.	Date	Rel. Note	Data	Spec	Delta	CVS Tag
1.8.1	2010-04-29	<a href="#">Version1.8.1</a>	<a href="#">CLDR1.8.1</a>	<a href="#">LDML1.8.1</a>	<a href="#">Changes1.8.1</a>	<a href="#">release-1-8-1</a>

**Step 3 :** Click CLDR 1.8.1 under Data column

**Step 4 :** Page with following downloads will be open



	<a href="#">Parent Directory</a>				-
	<a href="#">core.zip</a>	29-Apr-2010	11:26	3.2M	
	<a href="#">posix.zip</a>	29-Apr-2010	11:26	16M	
	<a href="#">tests.zip</a>	29-Apr-2010	11:26	645K	
	<a href="#">tools.zip</a>	29-Apr-2010	11:26	6.5M	

**Step 5 :** Download core.zip and save it on your system

**Step 6 :** Open **hi.xml** under `..\core\common\main`

# Our Efforts

- Locale data for six languages viz. **Hindi , Nepali, Bengali , Assamese, Malayalam and Gujarati** are uploaded in CLDR 1.8.1
- UNICODÉ agreed to add new fields required, specific for Indian scenario
- Development of CLDR data for remaining languages is under process

## Sample Feedback data

First Quarter	पहली तिमाही
Second Quarter	दूसरी तिमाही
Third Quarter	तीसरी तिमाही
Fourth Quarter	चौथी तिमाही

विक्रम संवत	शक संवत	ग्रिगोरियन
महीना/मास	महीना/मास	महीना/मास
चैत्र	चैत्र	जनवरी
वैशाख	वैशाख	फरवरी
ज्येष्ठ	ज्येष्ठ	मार्च
आषाढ	आषाढ	अप्रैल
श्रावण	श्रावण	मई
भाद्र	भाद्र	जून
आश्विन	आश्विन	जुलाई
कार्तिक	कार्तिक	अगस्त
अग्रहायन	अग्रहायन	सितम्बर
पौष	पौष	अक्टूबर
माघ	माघ	नवम्बर
फाल्गुन	फाल्गुन	दिसम्बर

Contd.

तिथि	सप्ताह / हफ़्ता	समय अवधि	युग
प्रतिपदा	रविवार / इतवार	आज	सत युग
द्वितीया	सोमवार	कल	त्रेता युग
तृतीया	मंगलवार	परसों	द्वापर युग
चतुर्थी	बुधवार	नरसों	कलियुग
पंचमी	गुरुवार / बृहस्पतिवार	दिन	
षष्ठी	शुक्रवार	प्रातः / सुबह / प्रभात	
सप्तमी	शनिवार / शनीश्वर	पूर्वाह्न	
अष्टमी		अपराह्न	
नवमी		मध्याह्न / दोपहर	
दशमी		संध्या / शाम / सांझ	
एकादशी		रात / रात्रि	
द्वादशी		तिमाही	

Contd.

# Modification Required in UNICODE CLDR data for Indian Languages:

1.

New fields to be added in CLDR	Examples
➤ Seasons	वसन्तः , ग्रीष्मः, वर्षा , शरद् , हेमन्तः , शिशिरः
➤ Local Measurement System for Area	1 हेक्टेयर – 2 एकड़ 100 हेक्टेयर – 1 वर्ग
➤ Local Measurement System Distance	220 गज – 1 फर्लांग 100 सेंटीमीटर – 1 मीटर
➤ Local Measurement System Weights	250 ग्राम – 1 पाव 1000 ग्राम – 1 किलोग्राम
➤ Fortnight	पक्ष / पखवारा

**2. No provision to submit data for 7 Indian languages namely**

- **Manipuri,**
- **Bodo,**
- **Sindhi,**
- **Santali,**
- **Maithili,**
- **Santali and**
- **Kashmiri.**

**3. New locales with all fields need to be added in CLDR.**

# Conclusion

- **Standardization of Locale Data Repository is a very important task from the Localization of Indian Languages perspective**
- **Locale data for all languages is available at one place**
- **XML format facilitates data sharing**
- **It makes resources available to anyone, at no cost**
- **For details please visit : <http://cldr.unicode.org/>**



**Thank you**