ABSTRACT
This paper reveals the fact about CLDR. The Common Locale Data Repository is a volunteer-developed and maintained resource coordinated and administered by the Unicode Consortium. CLDR is associated with languages and country names, date formats, time zone names, and so forth. It is based on BCP 47.

INTRODUCTION
Not long ago, computer systems were like separate worlds, isolated from one another. The internet and related events have changed all that. A single system can be built of many different components, hardware and software, all needing to work together. Many different technologies have been important in bridging the gaps; in the internationalization arena, Unicode has provided a lingua franca for communicating textual data. However, there remain differences in the locale data used by different systems. Its goal is to gather basic linguistic information for various "locales," essentially combinations of a language and a location.

LOCALITY
In computing, locale is a set of parameters that defines the user's language, country, and any special variant preferences that the user wants to see in their user interface. Usually a locale identifier consists of at least a language identifier and a region identifier. A locale is an identifier (id) that refers to a set of user preferences that tend to be shared across significant swaths of the world. Traditionally, the data associated with this id provides support for formatting and parsing of dates, times, numbers, and currencies; for measurement units, for sort-order (collation), plus translated names for time zones, languages, countries, and scripts. The data can also include support for text boundaries (character, word, line, and sentence), text transformations (including transliterations), and other services. There are many tools available on various technologies like java.util.Locale, .Net Culture, LANG (setlocale in C, C++), NLS_LANG in Oracle.

SCOPE AND NEED OF CLDR
There are many things that locale data can cover. It could cover industry specific topics. CLDR limits its scope to a few specific topics. It is depend upon user’s choice which can be change over time due to some political or cultural reasons like Introduction of new currencies, like the Euro, Standard sorting of Old Marathi or Manipuri and many of these preferences have varying degrees of standardization like in United States most of the people uses 12 hour time, however some of the people uses 24 hour time format. So, CLDR is very broad topic in itself. Localization is about Linguistic and Cultural adaption of digital content to the requirements of foreign markets and the management of multilingualism across the global and digital information flow. Text processing depends on languages spelling, stemming, tokenization, word/line/sentence boundaries, thesauri, terminology, morphological analysis, font and stylistic traditions, collation and IT systems are depend on language negotiation localization, message selection, user interface, presentation, number/date/time/etc. formatting, list presentation.

CLDR FORMAT
The Unicode Locale Data Markup Language (LDML) is the XML format for the communication of locale data. This provides a common format for systems to interchange locale data so that they can get the same results in the services provided by internationalization libraries. It also provides a standard format that can allow users to customize the behavior of a system. Unicode LDML can used to let a user encapsulate specialized sorting behavior for a specific domain, or create a customized locale for a minority language. It is also used in the Unicode Common Locale Data Repository (CLDR). CLDR uses an open process for reconciling differences between the locale data used on different systems and validating the data, to produce with a useful, common, consistent base of locale data. There are two kinds of data that can be expressed in LDML: language-dependent data and supplementary data. In either case, data can be split across multiple files, which can be in multiple directory trees. Types of Locale Data are as follows: Dates/time formats, Number/Currency formats, Measurement System, Translated names for language, territory, script, timezones, currencies, Script and characters used by a language and Collation Specification as Sorting, Searching and Matching. The CLDR is totally based on BCP 47. Structure of CLDR controlled by Locale Data Markup Language (LDML) specification http://unicode.org/reports/tr35. That element contains the following elements:

<identity>: The identity element contains information identifying the target locale for this data, and general information about the version of this data.

<version number="Revision: 1.227 ">
The version element provides, in an attribute, the version of this file. The contents of the element can contain textual notes about the changes between this version and the last. For example:

<version number="1.1">Various notes and changes in version 1.1</version>
<fallback>: Sometimes, it happens that there may not be full data for a particular locale. It provides a mechanism to indicate what the best available fallback locales would be. For example, for a Marathi speaker, the best fallback if data is unavailable might be Marathi. That is, suppose we have found a Marathi bundle, but it does not contain translation for the key "IN" (for the country India). It is best to return "Indiae", rather than falling back to the value default for the key "IN" (for the country India). It is best to return "Indiae", rather than falling back to the value default for the key "IN" (for the country India). It is best to return "Indiae", rather than falling back to the value default for the key "IN" (for the country India). It is best to return "Indiae", rather than falling back to the value default for the key "IN" (for the country India). It is best to return "Indiae", rather than falling back to the value default for the key "IN" (for the country India). It is best to return "Indiae", rather than falling back to the value default for the key "IN" (for the country India).

<localeDisplayNames>: It provides translated display names for languages, territories, scripts, variants and keywords used in CLDR and most of this information is at the language level, since it typically does not vary by territory, only language. An example

```xml
<localeDisplayNames>
  <language type="aa"><exemplarCharacters type="proposed-x1001">अफ़ार</exemplarCharacters><draft="unconfirmed"></language>
  <language type="ab"><exemplarCharacters type="proposed-x1001">अंधाध्याजित</exemplarCharacters><draft="unconfirmed"></language>
  <language type="ach"><exemplarCharacters type="proposed-x1001">अकोली</exemplarCharacters><draft="unconfirmed"></language>
  <language type="ae"><exemplarCharacters type="proposed-x1001">अब्ाा�ज़यन</exemplarCharacters><draft="unconfirmed"></language>
  <!-- Other language examples -->
</localeDisplayNames>
```

<layout>: This top-level element specifies general layout features. It currently only has one possible element (other than <special>), which is always permitted.

```xml
<layout>
  <orientation lines="top-to-bottom" characters="left-to-right" />
</layout>
```

The lines and characters attributes specify the default general ordering of lines within a page, and characters within a line. The values are:

### Orientation Attributes

<table>
<thead>
<tr>
<th>Vertical</th>
<th>Horizontal</th>
</tr>
</thead>
<tbody>
<tr>
<td>top-to-bottom</td>
<td>left-to-right</td>
</tr>
<tr>
<td>bottom-to-top</td>
<td>right-to-left</td>
</tr>
</tbody>
</table>

If the lines value is one of the vertical attributes, then the characters value must be one of the horizontal attributes, and vice versa.

<characters>: Characters allows for creation of exemplar character sets. An exemplar set specifies the set of characters that must be present in order to properly render the language. Auxiliary exemplar set defines additional characters that may appear in foreign words or phrases but in Lower case only.

```xml
<characters>
  <!-- Character sets go here -->
</characters>
```
<numbers>: This tag specifies proper localized formatting of numeric quantities in Decimal, Scientific, Currency and Percentages. It also includes localized decimal, thousands separators, currency symbols, etc.

<symbols>
  <decimal/>. <decimal>
  <group/>. <group>
  <list/>. <list>
  <percentSign>%</percentSign>
  <nativeZeroDigit>0</nativeZeroDigit>
  <patternDigit>#</patternDigit>
  <plusSign>+</plusSign>
  <minusSign>-</minusSign>
  <exponential>E</exponential>
  <perMille>‰</perMille>
  <infinity>∞</infinity>
  <nan>☹</nan>
</symbols>

<units>: These elements specify the localized way of formatting quantities of units such as years, months, days, hours, minutes and seconds— for example, in English, "1 day" or "3 days". The English rules that produce this example are as follows ({} indicates the position of the formatted numeric value):

<unit type="day">
  <unitPattern count="one">{0} day</unitName>
  <unitPattern count="other">{0} days</unitName>
</unit>

<collations>: It contains one or more collation elements, distinguished by type. It contains Information in collation directory, not main. It is the XML version of Java/ICU collation syntax. Its Unicode collation algorithm has base http://unicode.org/reports/tr10

<posix>: A colon-separated list of strings that would normally be recognized as "yes" and "no" responses. For cased languages, this shall include only the lower case version. POSIX locale generation tools must generate the upper case equivalents, and the abbreviated versions, and add the English words wherever they do not conflict.

<posix>
  <messages>
    <yesstr draft="contributed">हाँ</yesstr>
  </messages>
</posix>

<references>: The references section supplies a central location for specifying references and standards. The uri should be supplied if at all possible. If not online, then a ISBN number should be supplied, such as in the following example:

<reference type="R2" uri="http://www.ur.se/nyhetssjurnalistik/3lan.html">Landskoder på Internet</reference>

BCP 47

A Best Current Practice (BCP) means that a certain manner of proceeding is in general the most logical choice a de facto standard of sorts. BCP is the name published by the Internet Engineering Task Force. A Best Current Practice is only a noncommittal suggestion, as is to be proceeded in a certain case. It is thus more flexible than a standard: As requirements and conditions change, from time to time, another manner of proceeding can be more promising, and subsequently implemented. Should the requirements permanently change, a BCP should be revised. It describes the structure, content, construction, and semantics of language tags for use in cases where it is desirable to indicate the language used in an information object. It also describes how to register values for use in language tags and the creation of user-defined extensions for private interchange. The language tag is composed of one or more parts, known as "subtags". Each subtag consists of a sequence of alphanumeric characters. BCP 47 can be used in no. of modern standards such as http, html, xml, png. Each language tag is composed of one or more "subtags" separated by hyphens. With the exception of private use language tags and grandfathered language tags, the subtags occur in the following order: a language subtag (potentially followed by up to three extended language subtags)- an optional script subtag, an optional region subtag, optional variant subtags, optional extension subtags, optional private use subtags. Language subtags are mainly derived from ISO 639-1 and ISO 639-2, script subtags from ISO 15924, and region subtags from ISO 3166-1 alpha-2 and UN M.49. Variant subtags are not derived from any standard. No extension subtags have yet been defined. The Language Subtag Registry, maintained by IANA, lists the current valid public subtags. Language tags consist of only language subtag, or a language subtag and a region subtag. For example, bn represents Bengali, and consists of the language subtag (from ISO 639-1), while bn-IN represents Indian Bengali, and consists of the language subtag bn followed by the region subtag IN (from ISO 3166-1). Subtags are not case sensitive, but the specification recommends using the same case as in the Language Subtag Registry, where region subtags are uppercase, script
subtags are titlecase and all other subtags are lowercase. Syntax will be:
Language-Tag = langtag
   / privateuse ; private use tag
   / grandfathered ; grandfathered registrations
   langtag = (language [* script] [* region] [* variant]
    (* extension) [*-privateuse])

For Example, bn language code is used for Bengali language; Beng Script code is used for Bengali Script and IN is used as Region Code for India. So, to represent the whole Language tag we can declare it as bn-Beng-IN (Bengali language uses Bengali Script in India.)

CLDR WITH ISO 8601, 15924, 3166, 639, 4217

ISO 639 (Language Codes): ISO 639 is the set of international standards that lists short codes for language names. It provides two sets of three-letter alphabetic codes for the representation of names of languages. ISO 639 consists of different parts, of which two parts have been approved and a third part that is in the final approval (FDIS) stage. The other parts are works in progress. The ISO 639 code is divided into two parts i.e. Two-character code (639-1) Example: hi for Hindi, kn for kannada and Three character code (639-2, 639-3,639-5) Example: mar for Marathi, san for sanskrit (ISO 639-1)
- ISO 639-1 (alpha2 codes676) (136 codes)
- ISO 639-2 (alpha3 codes1751) (about 500)
- ISO 639-3 (alpha3 codes) (about 7000)
- ISO 639-4 (principles for encoding)
- ISO 639-5 (language families)
- ISO 639-6 (alpha4 codes) (under development)

There's no definitive list associating scripts with languages (because most languages could be written with most scripts, and this actually happens when people are transliterating foreign languages). This is a N-to-N relation, which only includes the standardized orthographies, and even lots of standard languages are missing the most basic information about the well-known scripts they are written with. For example: * language "br" (Breton)->script "Latn" (Latin) is missing, despite it is the only standardized script for the language. There are many indec scripts are not associated with the most language using it.

ISO 15924 (Script Code): There is a list in ISO 15924 used in Unicode 4.0 in which some Indian Scripts are included. Some more Scripts are added in Unicode 5.0 as well.

ISO 3166 (Region Codes): to denote the area in which a term, phrase, or language is used and provides two and three-character codes for representing names of countries. These two standards were developed independently, and there was no attempt to use the same code for a language as that for the country in which it is spoken. One should use codes from each list independently. However, the language code and country code may be used together to indicate a language variation spoken in a particular country as defined by BCP 47. For example, Panjabi as spoken in the India may be indicated with the following: pa-IN.

ISO 4217 (Currency Codes): ISO 4217 is the international standard describing three-letter codes (also known as the currency code) to define the names of currencies established by the International Organization for Standardization (ISO). The ISO 4217 code list is the established norm in banking and business all over the world for defining different currencies, and in many countries the codes for the more common currencies are so well known publicly, that exchange rates published in newspapers or posted in banks use only these to define the different currencies, instead of translated currency names or ambiguous currency symbols. ISO 4217 codes are used on airline tickets and international train tickets to remove any ambiguity about the price.

ISO 8601(Date and Time): ISO 8601 is an international standard for date and time representations issued by the International Organization for Standardization (ISO). Specifically, the standard is titled "Data elements and interchange formats — Information interchange — Representation of dates and times." The signature feature of ISO 8601 date and time representations is the ordering of date and time values from the most to the least significant or, in plain terms, from the largest (the year) to the smallest (the second).

CLDR AND UNICODE

These are the Draft CLDR data available in UNICODE CLDR initiative for Indian languages.

<table>
<thead>
<tr>
<th>S.NO.</th>
<th>LANGUAGE</th>
<th>COUNTRY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Assamese</td>
<td>India</td>
</tr>
<tr>
<td>2.</td>
<td>Bengali</td>
<td>Bangladesh, India</td>
</tr>
<tr>
<td>3.</td>
<td>Gujarati</td>
<td>India</td>
</tr>
<tr>
<td>4.</td>
<td>Hindi</td>
<td>India</td>
</tr>
<tr>
<td>5.</td>
<td>Kannada</td>
<td>India</td>
</tr>
<tr>
<td>6.</td>
<td>Konkani</td>
<td>India</td>
</tr>
<tr>
<td>7.</td>
<td>Malayalam</td>
<td>India</td>
</tr>
<tr>
<td>8.</td>
<td>Marathi</td>
<td>India</td>
</tr>
<tr>
<td>9.</td>
<td>Oriya</td>
<td>India</td>
</tr>
<tr>
<td>10.</td>
<td>Punjabi</td>
<td>India, Pakistan</td>
</tr>
<tr>
<td>11.</td>
<td>Punjabi (Gurmukhi)</td>
<td>India</td>
</tr>
<tr>
<td>12.</td>
<td>Sanskrit</td>
<td>India</td>
</tr>
<tr>
<td>13.</td>
<td>Tamil</td>
<td>India</td>
</tr>
<tr>
<td>14.</td>
<td>Telugu</td>
<td>India</td>
</tr>
<tr>
<td>15.</td>
<td>Urdu</td>
<td>India, Pakistan</td>
</tr>
</tbody>
</table>

CLDR AND LOCALIZATION

Localization is the process of adapting software for a specific region or language by adding locale-specific components and translating text. It is usually abbreviated as l10n (where 10 stand for the number of letters between the l and the n in Localization). Internationalization is the process of designing a software application so that it can be adapted to various languages and regions without engineering changes. It is usually abbreviated as i18n where 18 stand for the number of letters between the i and the n in internationalization. CLDR is by far the largest and most extensive standard repository of locale data. This data is used by a wide spectrum of companies for their software internationalization and localization.

CLDR - A strategic tool for the development of
**E-Governance.**

The Indian government is using IT to facilitate governance. The last couple of years have seen e-governance drop roots in India. IT enables the delivery of government services as it caters to a large base of people across different segments and geographical locations. It also gives citizens easy access to tangible benefits, be it through simple applications such as online form filling, bill sourcing and payments, or complex applications like distance education and tele-medicine.

Nowhere, can the benefit of CLDR and public-private partnerships be seen more clearly than in e-Governance initiatives. E-Governance, by nature, is citizen-centric rather than officer-centric. Therefore, it is necessary to be interactive, rather than just presenting information. For e-Governance to be truly effective and sustainable, it will have to go to the very doorsteps of the people - and CLDR can help a lot in making this possible.

**CLDR in W3C**

The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential.

**Goals of W3C**

- Web for Everyone
- Web on Everything
- Knowledge Base
- Trust and Confidence

CLDR data can be used for many W3C activities like internationalization, Language Tags and Locale Identifiers for the World Wide Web. To support users in different languages, software systems must not only use translated text, but must also be adapted to local conventions. An identifier for such a set of default conventions for a given user community (identified by language, script, region, variant, etc.) is commonly referred to as a locale.

Locale Specific data is the name used for denoting default conventions of a language, script or a region community. Those data which are specific to a locale is called locale specific data. A collection of common, necessary locale data for all world languages is provided by Common Locale Data Repository (CLDR) and this locale data can contribute very much to error free and fast translation and avoid confusion.

**CLDR AND INDIAN LANGUAGE TAGS**

Identification of Language and its locale is a challenge by itself. IETF has developed BCP 47 for both Language Tags and Locale. Since Indian languages have varieties. Indian Language hierarchy is much unmanaged. It has dialect, sub dialect, regional and stylistic differences and patois, also different communities have different requirements in terms of terminology, librarians, computer systems, translators. Language needs to identify in written form and locale provides the bases how to write data in particular language correctly. To make data accessible on internet through locale there should be harmony between operating environment, Application server, Web servers and acceptance of CLDR/LDML of the particular language. A variety of problems were associated with language tags, despite their success. The main problem associated with Language tags and CLDR is script variation. Most languages are customarily written in a single script. They may be transcribed in another script, but most native speakers and most content in that language use a single script. A few languages are written equally—or at least “commonly”—in more than one script. Some of the languages are undergoing transitions (Devanagari script written in more than one script. For example, Kashmiri can be written in either Arabic or Devanagari script. Both traditions are historical to the language, not artificially imposed. The most notable problem was the relative ease of registration for language tags compared to the action of the various ISO maintenance and registration bodies. Many of the registered tags were later deprecated due to standards action and another problem is instability in ISO 3166 (the region codes). Codes in ISO 3166 are changing all the time, which is not a surprise, given that countries are changing name, boundaries, and organization with some regularity. ISO 3166 not only remove old codes but they sometimes introduce new country or region.

**SPECIFIC ISSUES RELATED TO INDIAN LANGUAGES**

**INDIAN (Missing Date format)**

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Code</th>
<th>English</th>
<th>Proposed</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Indian/date/ full/standard</td>
<td>EEEE,</td>
<td>EEEE,d</td>
<td>EEEE,</td>
</tr>
<tr>
<td></td>
<td>standard/standard</td>
<td>MMMM d,</td>
<td>MMMM,</td>
<td>MMMM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y</td>
<td>y</td>
<td>ddd</td>
</tr>
<tr>
<td></td>
<td><strong>Example</strong></td>
<td>Tuesday,</td>
<td>मंगलवार,</td>
<td>मंगलवार,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bhadra 3</td>
<td>२३  अठढ़</td>
<td>१९२१  अठढ़</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1921</td>
<td>२३</td>
<td>१९२१</td>
</tr>
<tr>
<td>2.</td>
<td>Indian/date/ long/standard</td>
<td>MMMM d,</td>
<td>d MMMM</td>
<td>yyyy</td>
</tr>
<tr>
<td></td>
<td>standard/standard</td>
<td>y</td>
<td>y</td>
<td>MMMM d</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Example</strong></td>
<td>Bhdara</td>
<td>२३ अठढ़</td>
<td>१९२१ अठढ़</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1921</td>
<td>२३</td>
<td>१९२१</td>
</tr>
<tr>
<td>3.</td>
<td>Indian/date/ medium/</td>
<td>MMMM d,</td>
<td>dd-MM-yyy</td>
<td>y MMMM</td>
</tr>
<tr>
<td></td>
<td>standard/standard</td>
<td>y</td>
<td>y</td>
<td>ddd</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Example</strong></td>
<td>Bhdara</td>
<td>२३-०६-</td>
<td>१९२१ अठढ़</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1921</td>
<td>२३-०६</td>
<td>१९२१</td>
</tr>
<tr>
<td>4.</td>
<td>Indian/date/ short/</td>
<td>M/d/yy</td>
<td>d-M-yy</td>
<td>yyyy-MM</td>
</tr>
<tr>
<td></td>
<td>standard/standard</td>
<td></td>
<td></td>
<td>ddd</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Example</strong></td>
<td>6/23/21</td>
<td>२३-७-२१</td>
<td>१९२१-०६-२१</td>
</tr>
</tbody>
</table>

**Gregorian (Missing Date Interval Format)**
The Committee process is converted into five processes. These processes are as follows:
1. When gathering data for a region and language, it is important to have multiple sources for that data to produce the most widely acceptable data.
2. Initial versions of data were based on the best available sources, but CLDR data will be modified and improved, in successive versions, by more input from the contributors inside and outside of the Unicode Consortium.
3. CLDR is a Repository, not a Registration. It will be vetted against the best available information.
4. Inputs are open, and gathered via the CLDR Survey Tool or recorded in a bug/feature request database (CLDR Bug Reports).
5. Changes in response to requests in the database may be entered into the repository snapshot over time by the maintainers of the repository, but the final approval of the release of any version of CLDR is up to the decision of the CLDR Technical Committee.

**VETTING PROCESS FOR DATA**

Once data for a country and language has been received, the data from the different sources will be compared to show agreements and differences. Initial data contributions are normally marked as "draft=provisional"; this may be changed once the data is vetted.

Note that there are two types of data in the repository:

**Comparison data:** The contributor can be an individual or an organization. Data is normally gathered by calling public APIs, to ensure that the data matches what is actually in use. The data is only for comparison, and will not be changed except where necessary to update the data to match the external source. The only requirement is that all changed data be versioned, and the Version Numbering Scheme be used.

**Common Data:** The contents are decided upon by the CLDR Technical Committee, following its procedures and this process.

**COMMITTEE PROCESS**

The Committee process is converted into five processes. These processes are as follows:

1. When gathering data for a region and language, it is important to have multiple sources for that data to produce the most widely acceptable data.
2. Initial versions of data were based on the best available sources, but CLDR data will be modified and improved, in successive versions, by more input from the contributors inside and outside of the Unicode Consortium.
3. CLDR is a Repository, not a Registration. It will be vetted against the best available information.
4. Inputs are open, and gathered via the CLDR Survey Tool or recorded in a bug/feature request database (CLDR Bug Reports).
5. Changes in response to requests in the database may be entered into the repository snapshot over time by the maintainers of the repository, but the final approval of the release of any version of CLDR is up to the decision of the CLDR Technical Committee.

**VETTING PROCESS FOR DATA**

Once data for a country and language has been received, the data from the different sources will be compared to show agreements and differences. Initial data contributions are normally marked as "draft=provisional"; this may be changed once the data is vetted.

Note that there are two types of data in the repository:

**Comparison data:** The contributor can be an individual or an organization. Data is normally gathered by calling public APIs, to ensure that the data matches what is actually in use. The data is only for comparison, and will not be changed except where necessary to update the data to match the external source. The only requirement is that all changed data be versioned, and the Version Numbering Scheme be used.

**Common Data:** The contents are decided upon by the CLDR Technical Committee, following its procedures and this process.
SUGGESTIONS ON HINDI CLDR

Date Formats: We need to include (i) Saka शक (National Calendar of Govt. of India) (ii) Vikrama Sambat (शक सम) and (iii) Gregorian Calendar in the CLDR. In both the विक्रमी संवत (present year + 57) – चंद्र (ि. सं. (2055) the counting of month starts from March. The Number of days in the month is either 28, 31 or 32. For this we need to define an extra tag “Calendar Type in the XML”. We need to include variations of the names of the days as given below:

Thursday - गुरवार, गुरवार, गुरवार
Saturday - शनिवार, शनिवार, शनिवार
Sunday - रविवार, रविवार, रविवार

There is need to correct the spellings of month as given below:

अक्ू्र (Preferred) – अक्ू्र
िस्ं्र (Preferred)
यवं्र (Preferred)
�दसं्र (Preferred)

The abbreviated forms of the months in an year needs to be Included as given below:

<table>
<thead>
<tr>
<th>तिथि</th>
<th>ससंह / हक्ता</th>
<th>समय अवधि</th>
<th>युग</th>
</tr>
</thead>
<tbody>
<tr>
<td>प्रतिपदा</td>
<td>रविवार / इतवार</td>
<td>आज</td>
<td>सत युग</td>
</tr>
<tr>
<td>द्वितीया</td>
<td>सोमवार</td>
<td>कल</td>
<td>द्रष्टा युग</td>
</tr>
<tr>
<td>तृतीया</td>
<td>मंगलवार</td>
<td>परसी</td>
<td>द्रापर युग</td>
</tr>
</tbody>
</table>

The quarters of the year may be represented as

<table>
<thead>
<tr>
<th>भाग</th>
<th>तिथि</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Quarter</td>
<td>पहली तिमाही</td>
</tr>
<tr>
<td>Second Quarter</td>
<td>दूसरी तिमाही</td>
</tr>
<tr>
<td>Third Quarter</td>
<td>तीसरी तिमाही</td>
</tr>
<tr>
<td>Fourth Quarter</td>
<td>चौथी तिमाही</td>
</tr>
</tbody>
</table>

Various Time-periods may be included as:

<table>
<thead>
<tr>
<th>तिथि</th>
<th>ससंह / हक्ता</th>
<th>समय अवधि</th>
<th>युग</th>
</tr>
</thead>
<tbody>
<tr>
<td>चनुमी</td>
<td>गुरुवार / गुरुवार</td>
<td>दिव</td>
<td>कलियुग</td>
</tr>
<tr>
<td>पंचमी</td>
<td>गुरुवार / गुरुवार</td>
<td>दिव</td>
<td>कलियुग</td>
</tr>
<tr>
<td>षट्टी</td>
<td>शुक्ल</td>
<td>प्रात: / सुबह</td>
<td>ठापत</td>
</tr>
<tr>
<td>सप्तमी</td>
<td>शशिक</td>
<td>प्रात: / सुबह</td>
<td>प्रभात</td>
</tr>
<tr>
<td>अष्टमी</td>
<td>अष्टमी</td>
<td>प्रात: / सुबह</td>
<td>ठापत</td>
</tr>
</tbody>
</table>
नवमी | मध्यावाह / दोपहर  
---|---  
दशमी | संध्या / शाम / सांझ  
एकादशी | रात / रात्रि  
द्वादशी | तिमाही  
त्रयोदशी | दूसरी तिमाही / छुमाही  
चतुर्दशी | तीसरी तिमाही  
अष्टमास्त्वा / धूमंगिमा | वर्ष / साल  
---|---  
दशशादी |  
शतशादी |  
सहस्रशादी |  
निषेध | पलक को झुकाने में लगा समय  
इत्य | चिँडियों की आयाज़ ची ची  
प्लः | कोई का कोई - कोई  
दौर्य |  
अक्षर | 24/60 / 2/5 सेकंड  
पल | 60 अक्षर – 24 सेकंड  
छड़ | 60 पल – 24 मिनट  
पहर | 7.5 छड़ – 3 घंटे  
दिन + रात | 60 / 24 घंटे / 8 पहर  

**Measurement System:**
The Indian traditional measuring systems for distance, weight etc. also may be included in the Hindi CLDR

<table>
<thead>
<tr>
<th>मीटर</th>
<th>सेर</th>
<th>फूट</th>
<th>गज</th>
<th>माईल</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 खबर</td>
<td>1 चावल</td>
<td>1 माशा</td>
<td>1 मील</td>
<td>1 आधा हाथ</td>
</tr>
<tr>
<td>8 चावल</td>
<td>1 रसी</td>
<td>1 माशा</td>
<td>1 आधा हाथ</td>
<td></td>
</tr>
<tr>
<td>8 रसी</td>
<td>1 माशा</td>
<td>1 टोला</td>
<td>1 गज</td>
<td></td>
</tr>
<tr>
<td>12 माशा</td>
<td>1 टोला</td>
<td>1 छुट्टक</td>
<td>1 मील</td>
<td></td>
</tr>
<tr>
<td>5 टोला</td>
<td>1 छुट्टक</td>
<td>1 पाव</td>
<td>1 गज</td>
<td></td>
</tr>
<tr>
<td>4 छुट्टक</td>
<td>1 पाव</td>
<td>आधा सेर</td>
<td>1 चावल</td>
<td></td>
</tr>
<tr>
<td>8 छुट्टक</td>
<td>आधा सेर</td>
<td>1 सेर</td>
<td>8 माशा</td>
<td></td>
</tr>
<tr>
<td>16 छुट्टक</td>
<td>1 सेर</td>
<td>100 मिलीमीटर</td>
<td>1000 मीटर</td>
<td></td>
</tr>
<tr>
<td>40 सेर</td>
<td>1 मल</td>
<td>250 ग्राम</td>
<td>1 किलोग्राम</td>
<td></td>
</tr>
<tr>
<td>1 यार्म</td>
<td>1 माशा / 15.48 ग्रेन</td>
<td>250 ग्राम</td>
<td>1 किलोग्राम</td>
<td></td>
</tr>
<tr>
<td>1 यार्म</td>
<td>1 माशा / 15.48 ग्रेन</td>
<td>250 ग्राम</td>
<td>1 किलोग्राम</td>
<td></td>
</tr>
<tr>
<td>1 किलोग्राम</td>
<td>1 सेर</td>
<td>5 लीटर</td>
<td>1 किवल</td>
<td></td>
</tr>
<tr>
<td>1 किवल</td>
<td>2.5 मल</td>
<td>1 लिटर</td>
<td>1 लीटर</td>
<td></td>
</tr>
</tbody>
</table>

**Time Formats:** In the time format we should follow the Indian Standard Time. The Full and abbreviated forms of units of time may be represented as

<table>
<thead>
<tr>
<th>घंटा – मिनट – सेकंड</th>
<th>सेर</th>
<th>फूट</th>
<th>गज</th>
<th>मील</th>
</tr>
</thead>
<tbody>
<tr>
<td>घंटा – मिनट – सेकंड</td>
<td>सेर</td>
<td>फूट</td>
<td>गज</td>
<td>मील</td>
</tr>
</tbody>
</table>

**समय मापांक तालिका**

| समय मापांक तालिका | पलक को झुलाने में लगा समय |
It has been suggested that there is no change required in the present CLDR 1.6 format on the following tags:
(1) Character Map (As it has been adopted from UNICODE standard)
(2) Variants
(3) Types
(4) Locale Display

Collation Sequence and Sorting Order
After deliberations It has been decided that, the sorting order notified by Central Hindi Directorate (CHD) would be adopted for the collation sequence and sorting order as given. A cluster tag (वगर) may be included as sub-tag. This may be done in the following way:

CONCLUSION
The standardization of Locale Data Repository is a very important task from the Localization of Indian Languages perspective, Locale data for Indian languages are available in the present version of UNICODE CLDR (Ver 1.6.1), and the data needs to be critically examined, corrected and vetted before the next version of CLDR comes up. Since, Hindi in National Language and being spoken in 11 States in India, we need to examine the Hindi Locale data first. Elaborating the clear distinction between the UNICODE encoding standardization activity and the objectives of Standardization of Locale Data in the standardization process of Locale Data, we need to accommodate variations of a particular language in the XML Locale repository, so as to enable effective localization.

REFERENCES
Localization/Internationalization Standards
http://www.w3cindia.in/2006/08/Talks/Pat_Hall.pdf (Last Access on 06.01.09)
Internationalization and localization
http://en.wikipedia.org/wiki/Internationalization_and_localization (Last Access on 06.01.09)
LDML http://www.unicode.org/reports/tr35/ (Last Access on 06.01.09)
CLDR http://www.unicode.org/cldr/ (Last Access on 06.01.09)
Localization Intelligence: Unicode's CLDR Project
http://gilbane.com/globalization/2008/04/global_intelligence_for_free_u.html (Last Access on 06.01.09)
Language Tags and Locale Identifiers for the World Wide Web http://www.w3.org/TR/ltli/(Last Access on 06.01.09)