# Developing a locale file for OpenOffice.org

Javier SOLA javier@khmeros.info - Alberto Escudero-Pascual aep@it46.se [1]

Version 2.0 - 1st May 2006

## Table of Contents

1. Introduction....1
2. Creating a locale file for OpenOffice.org 2.x....3
2.1 Introduction to LocaleGEN....3
2.2 Language and Country ISO Codes....4
2.3 Checking for your locale in OpenOffice.org....4
3. Creating a new locale....5
3.1 XML Header Section....5
3.2 LC_INFO Section....6
3.3 LC_CTYPE Section....6
3.4 LC_FORMAT Section....8
3.4.1 Placeholders....10
3.4.2 Usage DATE....11
3.4.3 Usage TIME....12
3.4.4 Usage DATE_TIME....13
3.4.5 Usage FIXED_NUMBER....14
3.4.6 Usage CURRENCY....14
3.4.7 Usage PERCENT_NUMBER....15
3.4.8 Usage SCIENTIFIC_NUMBER....15
3.5 LC_COLLATION Section....16
3.6 LC_SEARCH Section....16
3.7 LC_INDEX Section....16
3.8 LC_CALENDAR Section....18
3.9 LC_CURRENCY Section....19
3.10 LC_TRANSLITERATION Section....20
3.11 LC_MISC Section....20
3.12 LC_NumberingLevel....21
3.13 LC_OutlineNumberingStyles....22
4. Final XML Validation....23

## 1. Introduction

The written culture of a country is not only defined by its language. Many other conventions are applied on written texts.

As an example, in the US they write *long dates* placing the name or number of month before the day of the

1 The first version of this document (prior to LocaleGen) was written as part of the work on an "Open Source Localization Toolkit" under a grant provided to the Open Forum of Cambodia by the ICT R&D Grants Programme for Asia Pacific. The donors were APDIP (United Nations Development Program), the Internet Society, APNIC, Pan Asia Networking (IDRC), and Asia Media Information and Communication Centre. The present version has been improved and adapted to LocaleGen in the framework of the WordForge project, funded by the Open Society Institute. 

month, capitalize the month and make the day of the month an ordinal (May 2nd, 1960), while in Spain, the day is written first, the month is not capitalized and the word " de " has to be placed between the day and the month and between the month and the year (2 de mayo de 1960).

In other countries like Japan the words "day, month and year" has to be used in combination with the numbers in the date (1960 年 5 月 2 日). Other languages like Thai, will rather use the Buddhist calendar and dates will be completely different. In Igboland, for example months has 7 weeks, one week has four days and one hence one year is 91 weeks.

The same applies to numbers (the period is used as a decimal separator in some countries and the comma is used in others). The number one thousand can be written as: 1 000; 1,000 or 1.000. Different convention also are applied to currency writing formats (number of decimals, currency identifier placed before or after, etc.), measurement units and some other pieces of data.

Sorting order, word order, alphabetic order, collation order, or whatever you prefer to call it (all equivalent terms), is also script and culture dependent. Cultures that use the same script not necessary follow the same rules. In English the sorting order starts with "a b c d f…" while in Spanish it starts with "a b c ch d e f…". Swedish will index "w" as a "vv".

Accents and diacritics are also classified differently in different languages. In Icelandic, for example, certain accented letters are considered to be separate base characters: a á b d ð e é f g h i í j k l m n o ó p r s t u ú v x y ý þ æ ö.

The same language can be written with more than one written system, that is the case of Serbian that can be written with Latin or Cyrillic alphabets or Hausa that can be written in Latin and Arabic.

All this data and its representation is usually referred to in the computer world as LOCALE data. Locale data either refers to the general use of a language (independently of which country or region it is used in) or to the specific use of a language in a given country or regions (the conventions for the use of the Spanish language is Spain are different to the ones used for the Spanish language as used in Chile).

Programs that are localized to many languages tend to place all the data related to a language or to a country (region) in a file called a LOCALE for that culture.

OpenOffice.org requires that you place all the cultural data for your language/region in a file that has a format specific to OpenOffice.org. This file is an XML file encoded in UTF-8.

# 2. Creating a locale file for OpenOffice.org 2.x

In this document we described two ways to create a locale file for OpenOffice.org. The

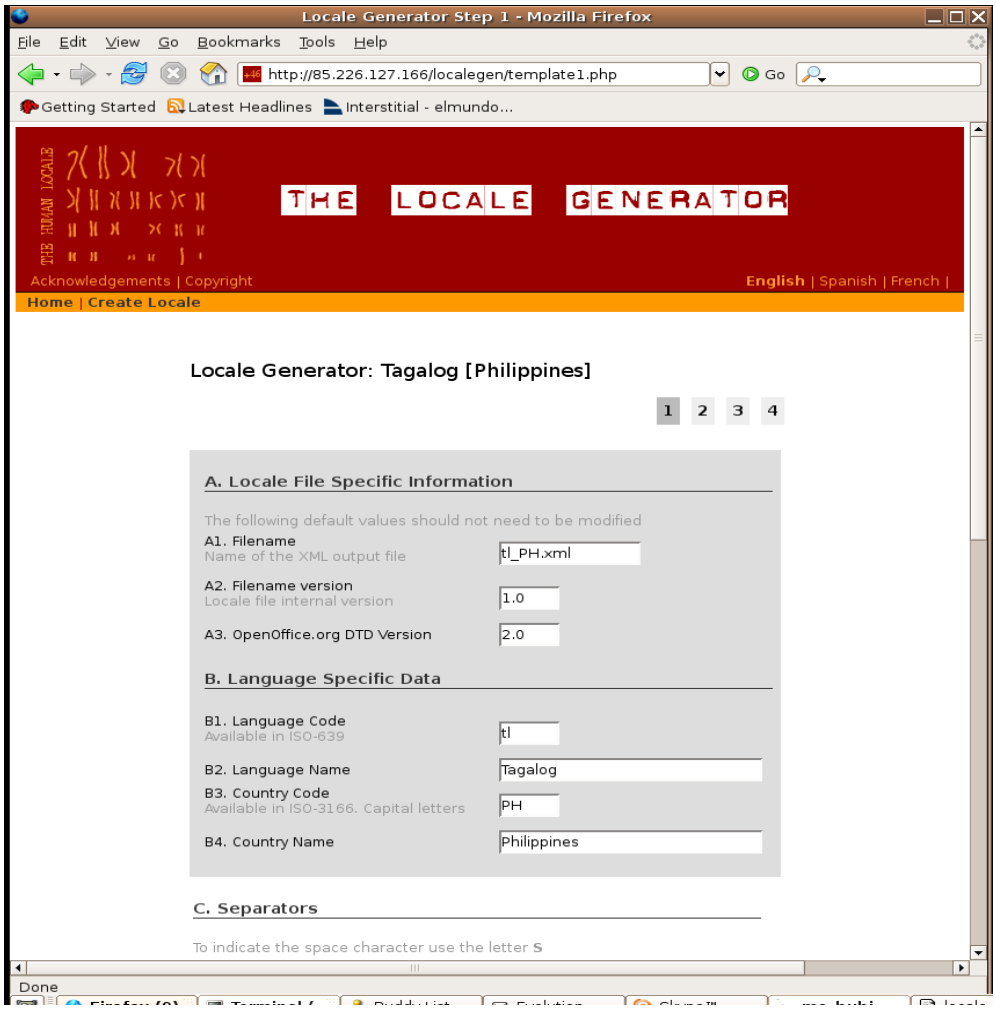
1) The first and more traditional method implies to modify an existing template http://www.khmeros.info/tools/template.xml
2) The second method is to use an online tool called: LocaleGEN

This document describes in detail the thirteen sections that a locale file has. In each section we include a reference to LocaleGEN by means of a "box with grey background color".

- This section is covered by Section # of the LocaleGEN

## *2.1 Introduction to LocaleGEN*

*Locale*GEN http://www.it46.se/localegen is an online tool that allows in four (4) steps to create a locale file for OpenOffice.org 2.0.x. The tool focuses on "information gathering" and usability, leaving the encoding of that information and the tricky internals to the tool. Users will not have to know what XML or a DTD is to create a locale file. Furthermore, the locale file produced by *Locale*GEN includes several checks that ensures its validation.



Back in September 2004, I created my first locale for OpenOffice.org. During the preparation of the Swahili (Tanzania) locale (sw-TZ.xml) for OpenOffice.org 1.1.3 i noticed that there were two main challenges in the task: the first one was to know which information was needed and second was to try to understand how to write the file itself. The creating of the "XML file" was a tedious work as most of the information to create a locale file was scattered in mailing lists.

After working with locale files in OpenOffice.org, we understood that the main perceived difficulty was not in finding out the "relevant data" for a certain culture rather the complexity of getting all that data in the OpenOffice.org XML format.

Back in 2004, Javier Sola started to write a guide to the creation of locale files in OpenOffice.org based on all that dispersed information: http://www.khmeros.info/tools/openoffice_locale_.htm

So what we did was to use the guide as a specification and create an automatic tool: LocaleGEN. The idea was to enable, via a simple step-by-step user interface, the collection of all necessary data to create "automatically" locale files for OpenOffice.org. The goal was to allow contributors in the Internet to create locales for their languages without bothering about this "document" at all.

## 2.2 Language and Country ISO Codes

The first step is to know the standardized ISO codes for language and the country you want to create the locale for. The following ISO standards contain the information: languages names in ISO639-2 and country codes in ISO-3166-1.

http://www.loc.gov/standards/iso639-2/englangn.html
http://www.iso.org/iso/en/prods-services/popstds/languagecodes.html
http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/index.html

For example "sw" is the ISO639-2 code for Swahili and TZ the ISO3166-2 country code for Tanzania.

The name of the locale file is composed of the language and country codes (language code in small letters and country code in capital letters), separated by an underscore (not a hyphen), and with the .xml extension.

Some examples of names are: km_KH.xml, es_CL.xml, en_US.xml, sw_TZ.xml etc.

## 2.3 Checking for your locale in OpenOffice.org

The first think that you might consider doing, it is to check if your locale already exist in OpenOffice.org. The directory that contains all the locale files is:

http://l10n.openoffice.org/source/browse/l10n/i18npool/source/localedata/data/

If there is a file for your culture already, you might want to check it to see if any corrections are needed. If they are, you should file an issue with OpenOffice.org related to mistakes in the file. If there is no file, that means that you need to create a new one.

- If you need to modify the locale for your language, consider inserting the existing data into LocaleGEN first.
- LocaleGEN includes several sanity checks that will ensure that the file generated is ready to be submitted to OpenOffice.org

# 3. Creating a new locale

If you decided to create the file manually, the easier way is to modify an existing one. To try to make it easy for you, we have created a template file that originates in the present en-US.xml file.

Download the template file from:
http://www.khmeros.info/tools/template.xml

The template includes most of the information that will be needed. A manual creating of the locale file implies to "localize the template.xml file" to your own culture.

Reminder: Sections in "grey" refer to the automatic creation of a locale using *Locale*GEN

- Instead of manually editing a template, *Locale*GEN will create the file based on the answers that you give to a set of questions.
- Questions are clustered in Sections and Sections are clustered in 4 Steps.

We will now go through this *template file* section by section to see what needs to be changed:

## *3.1 XML Header Section*

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE Locale SYSTEM 'locale.dtd'>
<Locale versionDTD="2.0" allowUpdateFromCLDR="no" version="1.0">
```

This first block gives general information about the locale itself.

The first line defines the version of XML that the file is written "1.0" and the encoding "UTF-8".
The second line defines the name of the DTD associated with the XML file. The DTD or Document Type Definition specifies which elements (the markup tags) and attributes (values associated with specific tags) needs to be present in the XML locale file. In simple words, the DTD helps to validate that the file is structurally correct.

The third line contains three values:

**version**: This value is for your own internal control. It is the version of the locale file for your culture. You can start with version="1.0" (as it is now) and change it when needed (later, if you do big upgrades of the locale file).

**allowUpdateFromCLDR**: This value indicates if you accept that changes in CLDR[2] locale file are migrated automatically into your OpenOffice.org locale file. If your OpenOffice.org locale and the CLDR locale contain the same general data (such as separators, day and month names and so on), as it is already the case for many locales, then you might be interested on using the value "yes", not to have to think about upgrading your locale every time something new comes out in the CLDR locale. If you rather prefer to keep your data different from the CLDR (there are disagreements), and you do not want the values to be changed to those of the CLDR locale, then you should use the value "no".

**versionDTD**: This attribute refers to the version of the DTD locale.dtd. The DTD is available in the

2 The Common Locale Data Repository (CLDR) http://www.unicode.org/cldr/ is a central location for locale data managed by the Unicode consortium that is becoming a central reference location for all locale data. You can browse the CLDR locale repository here: http://unicode.org/cldr/data/common/main/.

OpenOffice.org source code /i18npool/source/localedata/data/locale.dtd. This value does not change very often as it can imply changes in the locale files, the default value is verstionDTD="2.0”

As a general comment for the Header Section, you should not change anything, unless you know what you are doing, specially in the first two lines.

- The Header Section of the locale is addressed in input boxes of Section A and B of LocaleGEN
- After choosing the language and the country is spoken, it should automatically find ISO codes.
- Field A1 indicates the locale file name and, it is created automatically based on an internal database of ISO codes.
- Field A2 indicates the version number of “your” locale, so you can keep track of different versions. The default value is 1.0.
- Field A3 indicates the OpenOffice.org DTD version that LocaleGEN is working with. The default value is 2.0. This value should not be changed.

**Exceptions**:

- In most of the cases (if not all) the default values of Sections A and B should not be modified.
- If your language or country is not available, pick any other language and country and edit the A1, B1, B2, B3 and B4 of Sections A nd B manually

## 3.2 LC_INFO Section

```
<LC_INFO>
<Language>
<LangID>wa</LangID>
<DefaultName>Walloon</DefaultName>
</Language>
<Country>
<CountryID>BE</CountryID>
<DefaultName>Belgium</DefaultName>
</Country>
</LC_INFO>
```

In the LC_INFO section, you need to provide the language and region (country) codes, as well as language and the region full names in **English** (not in your own language and script).

- The LC_INFO of the locale is addressed in input boxes of Section A and B of LocaleGEN
- After choosing the language and the country is spoken, it should automatically find ISO codes.

## 3.3 LC_CTYPE Section

```
<LC_CTYPE unoid="generic">
<Separators>
<DateSeparator>/</DateSeparator>
<ThousandSeparator>.</ThousandSeparator>
<DecimalSeparator>,</DecimalSeparator>
<TimeSeparator>:</TimeSeparator>
<Time100SecSeparator>,</Time100SecSeparator>
<ListSeparator>;</ListSeparator>

<LongDateDayOfWeekSeparator> </LongDateDayOfWeekSeparator>
```

<LongDateDaySeparator> </LongDateDaySeparator>
<LongDateMonthSeparator> di </LongDateMonthSeparator>
<LongDateYearSeparator> </LongDateYearSeparator>
</Separators>
<Markers>
<QuotationStart>«</QuotationStart>
<QuotationEnd>»</QuotationEnd>
<DoubleQuotationStart>“</DoubleQuotationStart>
<DoubleQuotationEnd>”</DoubleQuotationEnd>
</Markers>
<TimeAM>AM</TimeAM>
<TimePM>PM</TimePM>
<MeasurementSystem>metric</MeasurementSystem>
</LC_CTYPE>

The LC_CTYPE section contains four different subsections: Separators, Markers, Time and Measurement System.

**Short Date, Time, Thousands and List Separators**

Pay attention to the <DecimalSeparator> and the <ThousandSeparator>. If in your country you write “one thousand three hundred fifty five and half” as (1,355.5), you do not need to change the proposed template value, but if you usually write (1.355,5), then you have to invert them.

The list separator is presently not used in OpenOffice.org, but it is defined to separate objects in enumerated list, such as: socks; shoes; Swiss army knife; toothbrush and clean underware. If you are planning to use something in your own script as a list separator, consult first in the mailing list, as it might not be possible.
In the languages that <space> is used as thousands (aka group) separator (1 355,5) the space should be type as U+00A0 that corresponds to the non-breaking space.

- This section is covered by Section C, Questions C1 to C6
- To input space use the capital letter “S”
- When S is indicated in the thousands separator, LocaleGEN will use the non-breakable space

**Long Date Separators**

The next four items are related to “Long Date” separators. You should think in long dates, as dates where you spell out some of the “numbers”. If you create the locale file manually and not assisted by LocaleGEN, you need to ensure that the separators that you use in this section are in agreement with the separators used in the *Date formats section.*

For consistency, automatic dates should have the same format that the ones you are declaring here. This means that Whatever <DateSeparator>, <TimeSeparator> and <Time100SecSeparator> you declare in the LC_CTYPE Section, they have to agree with the ones you use in dates formats below with FormatIndex numbers 21, 43, 44, 45 and 47. <DateSeparator> and <TimeSeparator> should contain only one code-point (one character).

The declarations in this section are independent of the *Date formats section,* with only one exception: <LongDateDayOfWeekSeparator>. Date formats declared in the *Date formats section* can include the “placeholder” letters NNNN (complete name of the day of the week). The <LongDateDayOfWeekSeparator> declared in this section will be automatically inserted after the occurrence of NNNN[3].

3 This is the reason why in some date formats you will find patterns like (NNNNMMMM DD, YYYY) in which MMMM is placed right after NNNN.
In “run time”, OpenOffice.org will display (NNNN<LongDateDayOfWeekSeparator>MMMM DD, YYYY. For example, if we declare <LongDateDayOfWeekSeparator> as “, “ the date 3/12/23 will be formated as Wednesday, March 12, 2023

As the automatic separator is inserted, and if we consider the <LongDateDayOfWeekSeparator> ",", it will be interpreted in this template locale as something like:

- This section is covered by Section C, Questions C7 to C10
- To input space use the capital letter “S”
- LocaleGEN will handle all the consistency requirements between this section and the *Date formats section*

**Markers:**
In general they are two groups/types of markers. The advanced directional quotation marks where the opening mark is different from the closing mark or the old-style straight marks ( ' and ") where the same mark is used for Start and End.

- This section is covered by Section D, Questions D1 to D4
- LocaleGEN offers a *drop-down menu* with different proposed alternatives
- If your markers are not in the *drop-down menu* choose any marker and modify manually the locale file

**TimeAM, TimePM:**
The words or abbreviations do your culture use to indicate morning (0:00 to 12:00) or afternoon (12:00 to 24:00) in a 12 hour clock. Translation of AM and PM. This concept might not make sense for many cultures. In that case, leave AM and PM.

- This section is covered by Section G, Questions G5 and G5
- If this concept does not make sense in your culture, keep AM and PM and report it to aep@it46.se

**Measurement System:**
Three values can be used in this section metric(default), US or UK (if your country uses the US or UK measurement systems).

- This section is covered by Section L, Question L1
- LocaleGEN offers a *drop-down menu* with the three different proposed alternatives

## 3.4 LC_FORMAT Section

<LC_FORMAT replaceFrom="[CURRENCY]" replaceTo="[$₦-470]">

**[$$-####]:**
This is the first of the section that sets the rule that [CURRENCY] will be replaced for [$$-##] inside this section. If you are using the template, you have to replace the second $ sign with the sign(s) used to indicate your currency (up to three characters). In the example we have replaced the “second” $ by ₦. The number is the Microsoft language ID for your language in hexadecimal format. You can find it here:

http://www.microsoft.com/globaldev/reference/lcid-all.mspx

If your language does not appear in this list, then say so in the OpenOffice.org Localization list and they will assign a number for you. You can use the unassigned default number FFFF meanwhile.

- This section is covered by Section I, Questions I2 and I4
- LocaleGEN will try to automatically find the correct values for the currency symbol and Microsoft

language ID
- If Language ID is not found, the default value is FFFF

```
<FormatElement msgid="FixedFormatskey1" default="true" type="medium" usage="FIXED_NUMBER" formatindex="0">
<FormatCode>General</FormatCode>
</FormatElement>
<FormatElement msgid="FixedFormatskey2" default="true" type="short" usage="FIXED_NUMBER" formatindex="1">
<FormatCode>0</FormatCode>
</FormatElement>
<FormatElement msgid="FixedFormatskey3" default="false" type="medium" usage="FIXED_NUMBER" formatindex="2">
<FormatCode>0.00</FormatCode>
</FormatElement>
[...]
</LC_FORMAT>
```

What comes next, it is the longest and more “cryptic” section of the locale file. This section is related to Automatic Formatting of numbers, dates, currencies etc. An intuitive way to know what this section is doing is to think in the formatting of a *Cell* in a Spreadsheet (Calc). We can input a number and later apply a “format” to the *Cell* so a date a time or even an associated currency is displayed instead.

Every element of the LC_FORMAT has the following structure:

```
<FormatElement msgid="msgid" default="true | false" type="short | medium | large" usage="FIXED_NUMBER | DATE | TIME | ..."
formatindex="##">
<FormatCode>Placeholder</FormatCode>
</FormatElement>
```

The <FormatElement> item contains:

1. msgid: Human readable text associated with the <FormatElement>
2. default (true|false): if the <FormatElement> is the default for certain *type* and *usage*. Only one can be true per type/usage combination
3. type (short|medium|long): a way to classify <FormatElement>'s of same *usage*.
4. usage DATE | FIXED NUMBER | TIME | DATE_TIME | CURRENCY | SCIENTIFIC_NUMBER | PERCENT_NUMBER): usage of the <FormatElement>.
5. formatindex: Unique identifier/index of a <FormatElement> . e.g: formatindex="21"
6. <FormatCode>: The placeholder/expression for the <FormatElement>.

The DTD allows you to re-order the <FormatElement> in any position inside of the LC_FORMAT section. The following conditions are required by the DTD:

This is the list of rules that you need to follow (check the comments in the *locale.dtd* for more info):

- formatindex 0-49 should ALL be used with the exception of 10,11, 48 and 49.
- No order is required in the way you list the <FormatElement>.
- formatindex 10 and 11 are not necessary as correspond to fractions formats "# ?/?" and "# ??/??" and are identical in every locale generated by the number formatter.
- formatindex 21 (YYYY/MM/DD) and 47 (YYYY/MM/DD HH:MM:SS AM/PM) are used to edit existing data. This are the formats that are applied in the *Edit box* and not in the Cell itself for DATE and DATE_TIME usage. There are necessary not to loose information.
- formatindex 32 and 33 are always for ISO 8601 (YY-MM-DD) and (YYYY-MM-DD) respectively
- formatindex 43 contains [HH] but should be understood as HH > 24 h. The common HH is always a value < 24 h (module 24)
- formatindex 44 does not use the hour code but the MM:SS.00 that includes the <Time100SecSeparator>
- formatindex 45 uses [HH]:MM:SS.00. It is used to *Edit* the full time without loosing the full time information

- formatindex 48 and 49 are for internal use
- If the replaceFrom is not used in the header of the section, formatindex 12, 13, 14, 15 and 17 should use the currency symbol instead of [CURRENCY]
- formatindex 16 that is using CCC is deprecated.

- This section is covered in Sections I and K, Questions I5 and K1
- LocaleGEN will try to automatically create all this section by asking what is the order of your culture to write Year Month and Day (YMD), the position of the currency symbol, the number of decimals used etc.

### 3.4.1 Placeholders

If you decide to edit the file manually it is a good idea that you know what the each placeholder means:

| | |
|---|---|
| G | Era |
| Y | Year |
| M | Month |
| D | Day (Number) |
| N | Day of Week (Name) |
| Q | Quarter |
| H | Hour |
| M | Minutes |
| S | Seconds |
| 00 | 1/100 Sec |
| C | Currency |

The number of times a placeholder letter is repeated is an indication of the number of characters to be used, but the number of characters does not always correspond exactly with the number of times the placeholder letter is repeated in the format. For example: ‘D’ means day of the month, with one or two characters (as needed: 2 for day 2, 12 for day 12), ‘DD’ means that two digits must always be used (day 5 must be 05). M is a one or two digit month, MM a two digit month, MMM a three letter month (short month name) and MMMM a long month name (long and short month names for your language are defined further down in the locale).

When you change the order of the placeholders, it is a good idea to maintain the separators and the amount of “information” in each format (i.e. keep the general structure of each *formatindex* and just reorder and include the correct separators needed).

By default, numbers in these formats will be Arab (English) numbers. If you want the placeholders to be replaced by a number in your own script, you need to precede the format with [NatNum1].

The following example adds a new format that uses native numbers:

<FormatElement msgid="DateFormatskey22" default="false" type="medium" usage="DATE_TIME" formatindex="50">
<FormatCode>[NatNum1]NNNN, "Day" D "of" MMMM "of the year" YYYY</FormatCode>

<DefaultName></DefaultName>
</FormatElement>

This will not translate the expressions "Day" or "of the year", but all the placeholders that represent the day of the month or the year will be in the numbers used in the script defined in your locale by the <UnicodeScript> number. For example, if this was part of a Khmer locale, the numbers would appear using Khmer script digits.

Now that you know what each placeholder means and the general structure of the the <FormatElement> edit the following section accordingly:

### 3.4.2 Usage DATE

<FormatElement msgid="DateFormatskey1" default="true" type="short" usage="DATE" formatindex="18">
<FormatCode>M/D/YY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey2" default="false" type="medium" usage="DATE" formatindex="28">
<FormatCode>NN DD/MMM YY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey3" default="false" type="medium" usage="DATE" formatindex="34">
<FormatCode>MM/YY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey4" default="false" type="medium" usage="DATE" formatindex="35">
<FormatCode>MMM DD</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey5" default="false" type="medium" usage="DATE" formatindex="36">
<FormatCode>MMMM</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey6" default="false" type="medium" usage="DATE" formatindex="37">
<FormatCode>QQ YY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey7" default="false" type="medium" usage="DATE" formatindex="21">
<FormatCode>MM/DD/YYYY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey8" default="true" type="medium" usage="DATE" formatindex="20">
<FormatCode>MM/DD/YY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey9" default="true" type="long" usage="DATE" formatindex="19">
<FormatCode>NNNNMMMM DD, YYYY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey10" default="false" type="long" usage="DATE" formatindex="22">
<FormatCode>MMM D, YY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey11" default="false" type="long" usage="DATE" formatindex="23">
<FormatCode>MMM D, YYYY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey12" default="false" type="long" usage="DATE" formatindex="25">
<FormatCode>MMMM D, YYYY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey13" default="false" type="long" usage="DATE" formatindex="27">
<FormatCode>NN, MMM D, YY</FormatCode>

<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey14" default="false" type="long" usage="DATE" formatindex="29">
<FormatCode>NN, MMMM D, YYYY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey15" default="false" type="long" usage="DATE" formatindex="30">
<FormatCode>NNNNMMMM D, YYYY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey16" default="false" type="long" usage="DATE" formatindex="24">
<FormatCode>D. MMM. YYYY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey17" default="false" type="long" usage="DATE" formatindex="26">
<FormatCode>D. MMMM YYYY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey18" default="false" type="short" usage="DATE" formatindex="31">
<FormatCode>MM-DD</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey19" default="false" type="medium" usage="DATE" formatindex="32">
<FormatCode>YY-MM-DD</FormatCode>
<DefaultName>ISO 8601</DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey20" default="false" type="medium" usage="DATE" formatindex="33">
<FormatCode>YYYY-MM-DD</FormatCode>
<DefaultName>ISO 8601</DefaultName>
</FormatElement>
<FormatElement msgid="DateFormatskey21" default="false" type="medium" usage="DATE" formatindex="38">
<FormatCode>WW</FormatCode>
<DefaultName></DefaultName>
</FormatElement>

### 3.4.3 Usage TIME

<FormatElement msgid="TimeFormatskey1" default="false" type="short" usage="TIME" formatindex="39">
<FormatCode>HH:MM</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="TimeFormatskey2" default="false" type="medium" usage="TIME" formatindex="40">
<FormatCode>HH:MM:SS</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="TimeFormatskey3" default="true" type="short" usage="TIME" formatindex="41">
<FormatCode>HH:MM AM/PM</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="TimeFormatskey4" default="true" type="medium" usage="TIME" formatindex="42">
<FormatCode>HH:MM:SS AM/PM</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="TimeFormatskey5" default="false" type="medium" usage="TIME" formatindex="43">
<FormatCode>[HH]:MM:SS</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="TimeFormatskey6" default="false" type="short" usage="TIME" formatindex="44">
<FormatCode>MM:SS.00</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="TimeFormatskey7" default="false" type="medium" usage="TIME" formatindex="45">
<FormatCode>[HH]:MM:SS.00</FormatCode>
<DefaultName></DefaultName>

</FormatElement>

### 3.4.4 Usage DATE_TIME

<FormatElement msgid="DateTimeFormatskey1" default="true" type="medium" usage="DATE_TIME" formatindex="46">
<FormatCode>MM/DD/YY HH:MM AM/PM</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="DateTimeFormatskey2" default="false" type="medium" usage="DATE_TIME" formatindex="47">
<FormatCode>MM/DD/YYYY HH:MM:SS</FormatCode>
<DefaultName></DefaultName>
</FormatElement>

Here is some rules of the thumb:

- Localize what is in the <FormatCode> lines, don't touch the <FormatElement> lines.
- Do not change the format for entries with formatindex 32 and 33, they correspond to data in a specific ISO formats (year first).
- In entries with usage TIME and formatindex 43, 44 and 45 change only the timeseparator if <TimesSeparator> is not ":" or if <Time100SecSeparator> is not "." (remember that they have to be consistent with the definitions in LC_CTYPE).
- Keep the same number of letters for each piece of data in dates with formatindex 20 (or whichever format has default=true type=medium, if you change this), 21 and 47; change only the order of the elements if needed, and use the separator defined in LC_CTYPE <DateSeparator>.
- Although you might see that some "placeholders" have been localized, do not localize them. As support for localized placeholders for new languages is deprecated.

- This section is covered in Sections I, K and L, Questions I5, I6, I7, K1, L2, L3
- LocaleGEN will try to automatically create all this section by asking what is the order of your culture to write Year Month and Day (YMD), the position of the currency symbol, the number of decimals used etc.

Spaces inside the dates are significant. If you include a space, it will be included in the printed date. The only exception is NNNNMMMM, where the day of the week is attached to the name of the month with no spaces between. OpenOffice.org automatically includes between them the <LongDateDayOfWeekSeparator> that you defined above (", " comma + space for US English).

Inside the date format, you can include strings with text before, between or after the placeholder letters, such as in D "de" MMMM "de" YYYY for Spanish. This is because dates in Spanish should be printed with these words in the middle: 2 de mayo de 1960 (note that there are spaces outside the quotes, which are significant. If there had also been spaces inside the quotes, they would also be taken into account. What you should never do is to put spaces inside and outside, then they would be duplicated in the final date.
Inside some of the date format, you will see "AM/PM". Do not translate these. They are placeholders for the words for AM and PM that you have defined in <TimeAM> and <TimePM>
The placeholder M is used in two different situations. When used within a date, it means "month", but when used within a time structure, it means "minute". Note that it might be used with both meanings within the same format.

You can include additional date formats or any other type of formats starting by formatindex="50". For example, you might want different predetermined date formats to appear in the places in which date formatting is done automatically (such as in *Calc Cell formatting*).

By default, numbers in these formats will be Arab (English) numbers. If you want date formats to use the

numbers of your script, you need to precede the format with [NatNum1].

The following example adds a new format that uses native numbers:

```
<FormatElement msgid="DateFormatskey22" default="false" type="medium" usage="DATE_TIME" formatindex="50">
<FormatCode>[NatNum1]NNNN, "Day" D "of" MMMM "of the year" YYYY</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
```

This will not translate the expressions "Day" or "of the year", but all the numbers that represent the day of the month or the year will be in the script defined in your locale by the <UnicodeScript> number. For example, if this was part of a Khmer locale, the numbers would appear using Khmer script digits.

When adding new date or other styles you should consider:

- Give them new consecutive DateFormatskey numbers. If the last one is DateFormatskey21, you should start with DateFormatskey22
- Use formatindex="##" number in which ## is over 49, as all numbers under 49 are reserved by the system.
- Notice that the the formatindex is a unique identifier for all types of formats (dates, numbers, currency, etc...). If you add dates with formatindex numbers 50 and 51 (for example), and then add number formats, they must have different formatindex such as 52, 53, etc.
- There is no relation between the number that you use in DateFormatskey and the number that you use in formatindex.

### 3.4.5 Usage FIXED_NUMBER

```
<FormatElement msgid="FixedFormatskey1" default="true" type="medium" usage="FIXED_NUMBER" formatindex="0">
<FormatCode>General</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="FixedFormatskey2" default="true" type="short" usage="FIXED_NUMBER" formatindex="1">
<FormatCode>0</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="FixedFormatskey3" default="false" type="medium" usage="FIXED_NUMBER" formatindex="2">
<FormatCode>0.00</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="FixedFormatskey4" default="false" type="short" usage="FIXED_NUMBER" formatindex="3">
<FormatCode>#,##0</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="FixedFormatskey5" default="false" type="medium" usage="FIXED_NUMBER" formatindex="4">
<FormatCode>#,##0.00</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="FixedFormatskey6" default="false" type="medium" usage="FIXED_NUMBER" formatindex="5">
<FormatCode>#,###.00</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
```

### 3.4.6 Usage CURRENCY

```
<FormatElement msgid="CurrencyFormatskey1" default="true" type="short" usage="CURRENCY" formatindex="12">
<FormatCode>[CURRENCY]#,##0;-[CURRENCY]#,##0</FormatCode>
<DefaultName></DefaultName>
```

```
</FormatElement>
<FormatElement msgid="CurrencyFormatskey2" default="false" type="medium" usage="CURRENCY" formatindex="13">
<FormatCode>[CURRENCY]#,##0.00;-[CURRENCY]#,##0.00</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="CurrencyFormatskey3" default="false" type="medium" usage="CURRENCY" formatindex="14">
<FormatCode>[CURRENCY]#,##0;[RED]-[CURRENCY]#,##0</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="CurrencyFormatskey4" default="true" type="medium" usage="CURRENCY" formatindex="15">
<FormatCode>[CURRENCY]#,##0.00;[RED]-[CURRENCY]#,##0.00</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="CurrencyFormatskey5" default="false" type="medium" usage="CURRENCY" formatindex="16">
<FormatCode>#,##0.00 CCC</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="CurrencyFormatskey6" default="false" type="medium" usage="CURRENCY" formatindex="17">
<FormatCode>[CURRENCY]#,##0.--;[RED]-[CURRENCY]#,##0.--</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
```

### 3.4.7 Usage PERCENT_NUMBER

```
<FormatElement msgid="PercentFormatskey1" default="true" type="short" usage="PERCENT_NUMBER" formatindex="8">
<FormatCode>0%</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="PercentFormatskey2" default="true" type="long" usage="PERCENT_NUMBER" formatindex="9">
<FormatCode>0.00%</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
```

#### 3.4.8 Usage SCIENTIFIC_NUMBER

```
<FormatElement msgid="ScientificFormatskey1" default="true" type="medium" usage="SCIENTIFIC_NUMBER"
formatindex="6">
<FormatCode>0.00E+000</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
<FormatElement msgid="ScientificFormatskey2" default="false" type="medium" usage="SCIENTIFIC_NUMBER"
formatindex="7">
<FormatCode>0.00E+00</FormatCode>
<DefaultName></DefaultName>
</FormatElement>
</LC_FORMAT>
```

This block is about number formats. The most important issues here are the consistency in the use of <DecimalSeparator>, the <ThousandSeparator> and how to represent currency: should the currency symbol be placed before the amount or after? How many decimals should currency representation have? Look at each format carefully. Remember to change only the lines that include <FormatCode>, and make sure that the representation for numbers that you include here is the same as the <DecimalSeparator>, the <ThousandSeparator> that you did above. This block is based on the US template, if your form is different, you must change it.

- This section is covered in Sections I, K and L, Questions I6, I7 and L2, L3

As described above, if you want different predetermined formats to appear in the places where number

formatting is done automatically, you can include additional FIXED_NUMBER, CURRENCY, PERCENT_NUMBER formats etc. By default, numbers in these formats will be Latin numbers. If you want the new number formats to use the numbers of your script, you need to precede the format with [NatNum1].

Similarly, you should give them new FixedFormatskey, CurrencyFormatskey or PercentFormatskey1 numbers that are consecutive to the already existing ones. For example, if the last FIXED_NUMBER is FixedFormatskey6, you should start with FixedFormatskey7 and so on.

## 3.5 LC_COLLATION Section

<LC_COLLATION>
<Collator default="true" unoid="alphanumeric"/>
<CollationOptions>
<TransliterationModules>IGNORE_CASE</TransliterationModules>
</CollationOptions>
</LC_COLLATION>

Collation is defined as the process of ordering units of textual information. Collation is usually specific to a particular language. Collation is also known as alphabetizing or alphabetic sorting. The Unicode Technical Report #10, "Unicode Collation Algorithm," defines a complete, unambiguous, specified ordering for all characters in the Unicode Standard.

The default collation algorithm *default=true* is the Unicode Collation Algorithm (alphanumeric). It is possible to define your own collation algorithm type “charset”. More information:

http://www.khmeros.info/tools/localization_of_openoffice_2.0.html#Collation

- LocaleGEN does NOT handle the creation of new collation algorithms
- The default algorithm used by the tool is alphanumeric with default=”true”

## 3.6 LC_SEARCH Section

<LC_SEARCH>
<SearchOptions>
<TransliterationModules>IGNORE_CASE</TransliterationModules>
</SearchOptions>
</LC_SEARCH>

Support the Find/Change feature, see source code: i18npool/source/search

- LocaleGEN always includes this section as default

## 3.7 LC_INDEX Section

<LC_INDEX>
<IndexKey phonetic="false" default="true" unoid="alphanumeric">ʾ A Ā B Ɓ C D Ɗ E-K Ƙ L-P R-W Y ʾY Z</IndexKey>
<UnicodeScript>0</UnicodeScript>
<UnicodeScript>1</UnicodeScript>
<UnicodeScript>3</UnicodeScript>
<FollowPageWord>s</FollowPageWord>
<FollowPageWord>w</FollowPageWord>
</LC_INDEX>

Before describing this section, it is important to understand how OpenOffice.org is handling this type of information. The <IndexKey> is a a list the letters of your script that is used to create an "Index". When talking about an Index think in the index that you normally find at the end of a book, listing and groups "words" and pointing to the page or pages that you can find information related to them. Do not confuse an Index with a Table of Contents where the content is ordered sequentially.

By specifying a <IndexKey> (keys for an Index), you are specifying which letters and the order you want to see them in the Index. If you are familiar with hashing tables and mathematical functions... you can think in the IndexKey as the space of letters where you want to "map" words to. Simply talking, think in the letters that you want to see in your Index, as the letters that are common in your dictionary to map words to.

When you write an <IndexKey> you need to consider the following aspects:

- Only composed characters are accepted
- You can name sets of characters by mentioning the first one and the last one separated by a hyphen For example: (A-Z) or (A-Я), corresponding to all latin or cyrillic letters between A and Z, or A and Я respectively.
- If the "letters" are made of more than one unicode codepoint as in the case of digraphs, you need to include them between {}. For example the <IndexKey> for spanish could look like this:
  A-C {CH} D-N Ñ O-Z
- "letters" require more than one unicode codepoint as a result of base character + 1 or more combining diacritics (decomposed) are not supported.

The <UnicodeScript> elemement(s) contains the Unicode Scripts ranges for your language. This information will be used to create a table dynamically to map the letters of those ranges to the (Index) keys.

A list of the Unicode Scripts and their reference number follows:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | BasicLatin | 23 | Sinhala | 46 | MathOperator | 69 | _CJKUnifiedIdeographsExtensionA |
| 1 | Latin1Supplement | 24 | Thai | 47 | MiscTechnical | 70 | CJKUnifiedIdeograph |
| 2 | LatinExtendedA | 25 | Lao | 48 | ControlPicture | 71 | YiSyllables |
| 3 | LatinExtendedB | 26 | Tibetan | 49 | OpticalCharacter | 72 | YiRadicals |
| 4 | IPAExtension | 27 | Myanmar | 50 | EnclosedAlphanumeric | 73 | HangulSyllable |
| 5 | SpacingModifier | 28 | Georgian | 51 | BoxDrawing | 74 | HighSurrogate |
| 6 | CombiningDiacritical | 29 | HangulJamo | 52 | Bloc | 75 | HighPrivateUseSurrogate |
| 7 | Gree | 30 | Ethiopic | 53 | GeometricShape | 76 | LowSurrogate |
| 8 | Cyrillic | 31 | Chero | 54 | MiscSymbol | 77 | PrivateUse |
| 9 | Armenian | 32 | UnifiedCanadianAborig | 55 | Dingbat | 78 | CJKCompatibilityIdeograph |
| 10 | Hebrew | 33 | Ogham | 56 | BraillePatterns | 79 | AlphabeticPresentation |
| 11 | Arabic | 34 | Runic | 57 | CJKRadicalsSupplement | 80 | ArabicPresentationA |
| 12 | Syriac | 35 | Khmer | 58 | KangxiRadicals | 81 | CombiningHalfMar |
| 13 | Thaana | 36 | Mongolian | 59 | IdeographicDescriptionChars | 82 | CJKCompatibilityForm |
| 14 | Devanagari | 37 | LatinExtendedAdditional | 60 | CJKSymbolPunctuation | 83 | SmallFormVariant |
| 15 | Bengali | 38 | Gree | 61 | Hiragana | 84 | ArabicPresentationB |
| 16 | Gurmu | 39 | GeneralPunctuation | 62 | Kata | 85 | NoScript |
| 17 | Gujarati | 40 | SuperSubScript | 63 | Bopomofo | 86 | HalfwidthFullwidthForm |
| 18 | Oriya | 41 | CurrencySymbolScript | 64 | HangulCompatibilityJamo | 87 | ScriptCount |
| 19 | Tamil | 42 | SymbolCombiningMar | 65 | Kanbun | | |
| 20 | Telugu | 43 | Letterli | 66 | BopomofoExtended | | |
| 21 | Kannada | 44 | NumberForm | 67 | EnclosedCJKLetterMonth | | |
| 22 | Malayalam | 45 | Arrow | 68 | CJKCompatibility | | |

The latest version of the list is available here:
http://api.openoffice.org/source/browse/api/offapi/com/sun/star/i18n/UnicodeScript.idl

This will show you a file that has the last version of this Unicode Script list… but unfortunately, the entries are not numbered. If you find the script for your language there, you have to start counting from the

beginning (BasicLatin is number 0) and figure out what is the correct number of your <UnicodeScript>. If your script is in Unicode, but you still do not find the script listed here, you should write to the dev@l10n.openoffice.org list mentioning this and asking what should be done.

Finally, The <FollowPageWord> refers to the abbreviation used in your language to talk about the following page or pages (such as s. and ss. in Spanish or sv. in both cases in French). Think that this are the words that are going to be used in the Index created automatically.

- This section is covered in Sections H, Questions H1 to H4

## 3.8 LC_CALENDAR Section

<LC_CALENDAR>
<Calendar unoid="gregorian" default="true">
<DaysOfWeek>
<Day>
<DayID>sun</DayID>
<DefaultAbbrvName>እሑድ</DefaultAbbrvName>
<DefaultFullName>እሑ< /DefaultFullName>
</Day>
<Day>
<DayID>mon</DayID>
<DefaultAbbrvName>ሰኞ</DefaultAbbrvName>
<DefaultFullName> ኞ</DefaultFullName>
</Day>

[...]
</DaysOfWeek>

<MonthsOfYear>
<Month>
<MonthID>jan</MonthID>
<DefaultAbbrvName>ጃንዩ</DefaultAbbrvName>
<DefaultFullName>ጃንዩወ< /DefaultFullName>
</Month>
<Month>
<MonthID>feb</MonthID>
<DefaultAbbrvName>ፌብሩ</DefaultAbbrvName>
<DefaultFullName>ፌብሩወሪ</DefaultFullName>
</Month>
[...]
</MonthsOfYear>
<Eras>
<Era>
<EraID>bc</EraID>
<DefaultAbbrvName>አ.አለም</DefaultAbbrvName>
<DefaultFullName>ከክርስቶስ ልደት በፊ< /DefaultFullName>
</Era>
<Era>
<EraID>ad</EraID>
<DefaultAbbrvName>አ.ም</DefaultAbbrvName>

<DefaultFullName> ሙሉ ምህረት</DefaultFullName>
</Era>
</Eras>
<StartDayOfWeek>
<DayID>sun</DayID>
</StartDayOfWeek>
<MinimalDaysInFirstWeek>1</MinimalDaysInFirstWeek>
</Calendar>
</LC_CALENDAR>

The next one is easy, if your country uses the Gregorian 12 month, 7 day week Gregorian (western) calendar. For each day of the week and for each month, you have to defined a long (full) name and a short or abbreviated name. You should not translate the ID fields (<DayID> or <MonthID>), they are used for reference. Translate only the abbreviated and Full names (<DefaultAbbrvName> and <DefaultFullName>) using the characters of your language. If you want to define a different type of calendar, you will have to ask about it in the list.

Use the right capitalization. In English the Months are written capitalizing the first letter, in other languages months are written all in small letters.

If your language uses different names for the eras <EraID> of the Gregorian calendar (before and after Christ). Localize those too.

In the US, the week starts on Sunday, but in Spain and other countries, the week is considered to start on Monday. Select the correct one for your language. Indicate it by the corresponding three **english** letter of the "DayID" . For example: sun (sunday), sat (saturday), mon (monday) etc.

The last item of the LC_CALENDAR it relates to cultures where (working) time is measure in weeks. This is very common in Sweden but it is not the case in many other cultures. At the beginning of the year how many days until the first day of the week need to be to be accounted as the "first" working week of the year?

Assume that the first day of the week in your culture is Monday and the first of January is a Friday. Will Friday be in the first week of your year or accounted in week 52 of the previous year?. Common values for <MinimalDaysInFirstWeek> are 1 or 2.

- Days of the week are covered in Sections E, Questions E1 to E7
- Months of the year are covered in Sections F, Questions F1 to F12
- Gregorian Calendar Specific Information is covered in Section G, Questions G1 to G4

## 3.9 LC_CURRENCY Section

<LC_CURRENCY>
<Currency default="true" usedInCompatibleFormatCodes="true">
<CurrencyID>NGN</CurrencyID>
<CurrencySymbol>₦</CurrencySymbol>
<BankSymbol>NGN</BankSymbol>
<CurrencyName>Neira</CurrencyName>
<DecimalPlaces>2</DecimalPlaces>
</Currency>
</LC_CURRENCY>

http://www.khmeros.info/tools/Currency%20codes.pdf
http://www.bsi-global.com/

In recent versions of OpenOffice.org the <Currency ID> =<BankSymbol>. Bank Symbol is normally a three

letter code. Check the following references:

The <Currency Symbol> is normally a “special letter” in your own script. Think in symbols like $, £, ₦ etc. Currency symbol should be in your own script. The items <BankSymbol> and <DecimalPlaces> come from the ISO4217.

<CurrencyName> is the long name of the currency in your local language. If your currency is a new one and is not in here, you should try to find it by yourself in your country, because if you go the standards body maintainer (BSI Global) , they will make you PAY for the data.

http://www.khmeros.info/tools/CurrencyIDs.gif
http://www.evertype.com/standards/iso4217/iso4217-en.html

- Currencies are covered by Section I, Questions I1, I2, I3 and I5
- LocaleGEN will try to find the Bank Symbol and Currency Name using ISO4217

## 3.10 LC_TRANSLITERATION Section

```
<LC_TRANSLITERATION>
<Transliteration unoid="LOWERCASE_UPPERCASE"/>
<Transliteration unoid="UPPERCASE_LOWERCASE"/>
<Transliteration unoid="IGNORE_CASE"/>
</LC_TRANSLITERATION>
```

Transliteration is a type of “string conversion”. It is used for specify character conversion algorithms. Languages that use Latin characters normally use the “upper to lowercase conversion” algorithms, where characters are “converted-transliterated” in a one-to-one mapping

Some languages, such as Japanese, Chinese or Korean (CJK), have complicated transliteration schemes that are supported in OpenOffice.org. Transliteration procedures need to be written before they are included here.

http://l10n.openoffice.org/i18n_framework/HowToAddLocaleInI18n.html

- Currencies does not covered Transliteration and uses the Latin default section

## 3.11 LC_MISC Section

```
<LC_MISC>
<ReservedWords>
<trueWord>true</trueWord>
<falseWord>false</falseWord>
<quarter1Word>1st quarter</quarter1Word>
<quarter2Word>2nd quarter</quarter2Word>
<quarter3Word>3rd quarter</quarter3Word>
<quarter4Word>4th quarter</quarter4Word>
<aboveWord>above</aboveWord>
<belowWord>below</belowWord>
<quarter1Abbreviation>Q1</quarter1Abbreviation>
<quarter2Abbreviation>Q2</quarter2Abbreviation>
<quarter3Abbreviation>Q3</quarter3Abbreviation>
<quarter4Abbreviation>Q4</quarter4Abbreviation>
</ReservedWords>
</LC_MISC>
```

The LC_MISC Section contains four subsections: true/false <trueWord> <falseWord>, above/below <aboveWord> <belowWord> and quarters and their abbreviation <quarter#Word> <quarter#Abbreviation>. Translate the words “true”, “false”, “above” and “below” to your language.

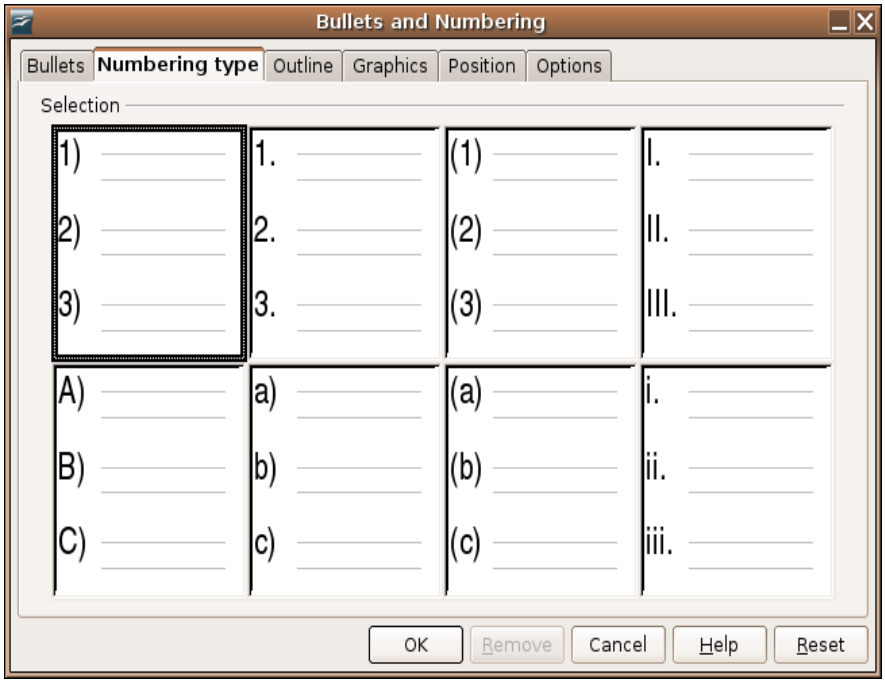
The Quarter's expression(s) in English are used to describe a period of three months. There are related to the way reporting and planning is done in companies and organizations. Translate 1st, 2nd etc. quarters and their abbreviations. Translate the following four expressions, in reference to whatever is more used in your language, quarters, trimesters or any other way that your language uses to call the four three-month groups contained in a year.

- Reserved words are covered by Section J, Questions J1 to J12

## 3.12 LC_NumberingLevel

```
<LC_NumberingLevel>
<NumberingLevel Prefix=" " NumType="4" Suffix=")" />
<NumberingLevel Prefix=" " NumType="4" Suffix="." />
<NumberingLevel Prefix="(" NumType="4" Suffix=")" />
<NumberingLevel Prefix=" " NumType="2" Suffix="." />
<NumberingLevel Prefix=" " NumType="0" Suffix=")" />
<NumberingLevel Prefix=" " NumType="1" Suffix=")" />
<NumberingLevel Prefix="(" NumType="1" Suffix=")" />
<NumberingLevel Prefix=" " NumType="3" Suffix="." />
</LC_NumberingLevel>
```

The next section relates to numbering styles for paragraphs. In the locale file we will define what styles will be included in OpenOffice.org. The options in Ooo Writer are reachable in *Format-->Bullets* and *Numbering-->Numbering Type Tab*. Each line in the locale file defines one of the 8 squares in this page.




For each one of them it defines three things: the style of numbering, what character should be placed before the number (prefix) and what character should be placed after the numbering (suffix). The <NumType> value refers to a list of types of numbers that is defined in:

*offapi/com/sun/star/style/NumberingType.idl*

The list includes some traditional western types of numbering (Latin letters, Arab numbers, Latin Numbers), the possibility of using numbers in the script of the locale (number 12) or creating specific numbering series (such as for example Thai letters). In our reference file number 4 refers to Arab number (as used in English), 2 refers to capital Latin numbers (not the ones used by English language), 0 to capital Latin letters, 1 to lower case Latin letters and 3 to lower case Latin numbers. You should define here the method that would best fit your culture. Some of the locales replace number 4 by number 12, making the first two or three in local numbering, and then use other methods. You should evaluate, for example, if Latin (Arabic) numbers are known to your culture.

- Numbering is covered by Section

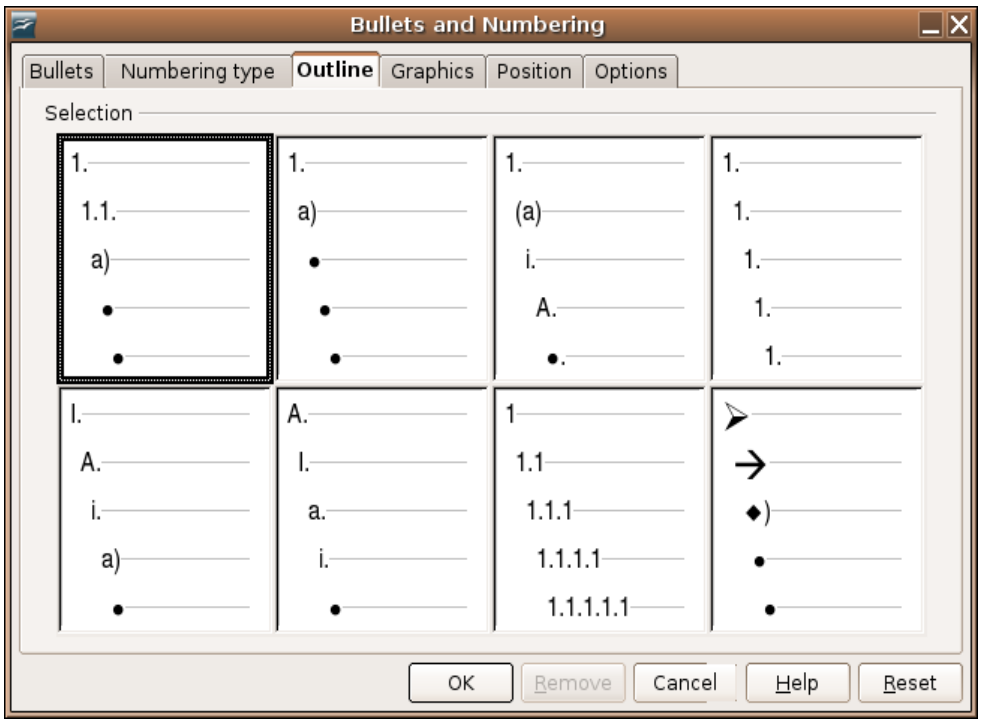
## 3.13 LC_OutlineNumberingStyles

```
<LC_OutLineNumberingLevel>
<OutlineStyle>
<OutLineNumberingLevel Prefix=" " NumType="4" Suffix="." BulletChar="0020" BulletFontName="" ParentNumbering="0"
LeftMargin="0" SymbolTextDistance="50" FirstLineOffset="0" />
<OutLineNumberingLevel Prefix=" " NumType="4" Suffix="." BulletChar="0020" BulletFontName="" ParentNumbering="1"
LeftMargin="50" SymbolTextDistance="50" FirstLineOffset="0" />
<OutLineNumberingLevel Prefix=" " NumType="1" Suffix=")" BulletChar="0020" BulletFontName="" ParentNumbering="0"
LeftMargin="100" SymbolTextDistance="50" FirstLineOffset="0" />
<OutLineNumberingLevel Prefix=" " NumType="6" Suffix=" " BulletChar="2022" BulletFontName="StarSymbol"
ParentNumbering="0" LeftMargin="150" SymbolTextDistance="50" FirstLineOffset="0" />
<OutLineNumberingLevel Prefix=" " NumType="6" Suffix=" " BulletChar="2022" BulletFontName="StarSymbol"
ParentNumbering="0" LeftMargin="200" SymbolTextDistance="50" FirstLineOffset="0" />
</OutlineStyle>
<OutlineStyle>
<OutLineNumberingLevel Prefix=" " NumType="4" Suffix="." BulletChar="0020" BulletFontName="" ParentNumbering="0"
LeftMargin="0" SymbolTextDistance="50" FirstLineOffset="0" />
<OutLineNumberingLevel Prefix=" " NumType="1" Suffix=")" BulletChar="0020" BulletFontName="" ParentNumbering="0"
LeftMargin="50" SymbolTextDistance="50" FirstLineOffset="0" />
<OutLineNumberingLevel Prefix=" " NumType="6" Suffix=" " BulletChar="2022" BulletFontName="StarSymbol"
ParentNumbering="0" LeftMargin="100" SymbolTextDistance="50" FirstLineOffset="0" />
<OutLineNumberingLevel Prefix=" " NumType="6" Suffix=" " BulletChar="2022" BulletFontName="StarSymbol"
ParentNumbering="0" LeftMargin="150" SymbolTextDistance="50" FirstLineOffset="0" />
<OutLineNumberingLevel Prefix=" " NumType="6" Suffix=" " BulletChar="2022" BulletFontName="StarSymbol"
ParentNumbering="0" LeftMargin="200" SymbolTextDistance="50" FirstLineOffset="0" />
</OutlineStyle>
[...]
</LC_OutLineNumberingLevel>
</Locale>
```

The final section is the Outline Numbering Styles. They affect the result of *Format-->Bullets and Numbering-->Numbering* [Outline Tab]. Each <OutlineStyle> block in the locale file defines one of the 8 squares in this page.

Each block has five lines defining the first five “levels of heading”. The first line of a block (for example) will define how paragraphs with style 'Heading 1' will be numbered (including number style, and characters to be placed before and after the number). Similarly, we can define if some bullet characters should be used, the left margin of Heading 1 numbered paragraphs, etc...

In subsequent levels (other lines), it is important to say how many levels of heading will be named in this specific number... for example, if we are defining the numbering of level 3 (Heading 3), the parenting number could be 0 (in which case only one number will be showed) could be 1 (two numbers shown, as in 1.1) or 2, in which case we will have numbers of the style 1.1.1.




If you want to use numbers in the script of the locale, you need to use NumType="12". Information about the different number styles is available in :

*offapi/com/sun/star/style/NumberingType.idl*

- The currenty version of LocaleGEN falls back to en-US

## 4. Final XML Validation

You must check your locale before you send it. Eike gives several possibilities (look into the locale.dtd file if you do not like the one we include below):

1) Temporarily change the DOCTYPE of your file to read (all on one line)
<!DOCTYPE Locale SYSTEM
"http://l10n.openoffice.org/unbranded-source/browse/*checkout*/l10n/i18npool/source/localedata/data/locale.dtd">

2) Upload it to the form available at http://www.validome.org/

- LocaleGEN does automatic validation of the file
- Download the file or check other locales here http://www.it46.se/localegen/locale

When you have checked your file submit it as an ENHANCEMENT issue against the Localization (L10n) project and submit the file.

To submit an issue you first need to:

- Login into the OpenOffice website,
- Hit File Issue on the left hand men
- Proceed in the next page and click in the component l10n in the next on
- Select version current, subcomponent code, type ENHANCEMENT,
- Include a Summary: Locale file for XXX language
- Hit Submit.
- The system will ask you if you want to attach a file and what type. Attach the file that you have been

working on, submit it and you are done.

If you would - nevertheless - like to prepare a more developed locale, please look at the following documents:

http://l10n.openoffice.org/i18n_framework/HowToAddLocaleInI18n.html
http://l10n.openoffice.org/source/browse/l10n/i18npool/source/localedata/data/locale.dtd
http://l10n.openoffice.org/i18n_framework/LocaleData.html
http://api.openoffice.org/docs/common/ref/com/sun/star/i18n/NumberFormatIndex.html