प्रगत संगणन विकास केंद्र
CENTRE FOR DEVELOPMENT OF ADVANCED COMPUTING

I.   Internationalization (i18n) & Localization (l10n)

II. Overview of proposed Localization Guidelines

III. Minimal set of localisation guidelines

IV. Meta Data and Data standards

Internationalization (i18n)

&

localization (l10n)

Definitions

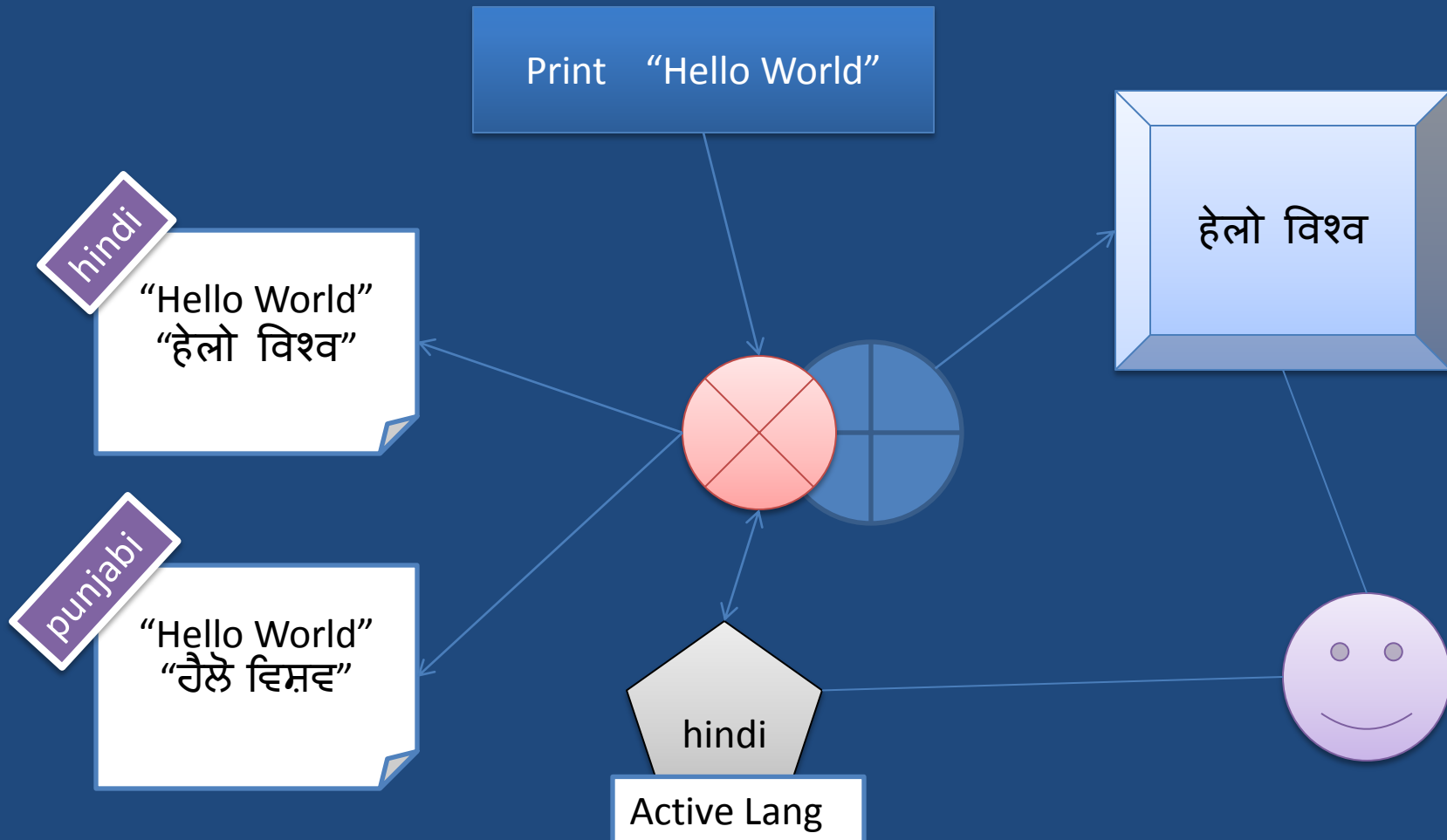| Translation	L10N	I18N	G11N
• *"**Translation** is the text transfer from a source to a target language".* • Translation is not just about text conversion but also deals with conversion of spoken words from source to target language, without losing its meaning.	• *Refers to the actual adaptation of the product for a specific market. It includes translation, adaptation of graphics, adoption of local currencies, use of proper forms for dates, addresses, and phone numbers, and many other details, including physical structures of products in some cases".*	• Design and development of a product, application or document that enables **easy localisation** for target audiences that differ in region, culture, or language. design and development of a product, application or document that enables **easy localisation** for target audiences that differ in region, culture, or language.	• *"Expansion of marketing strategies to address regional requirements of all kinds"*

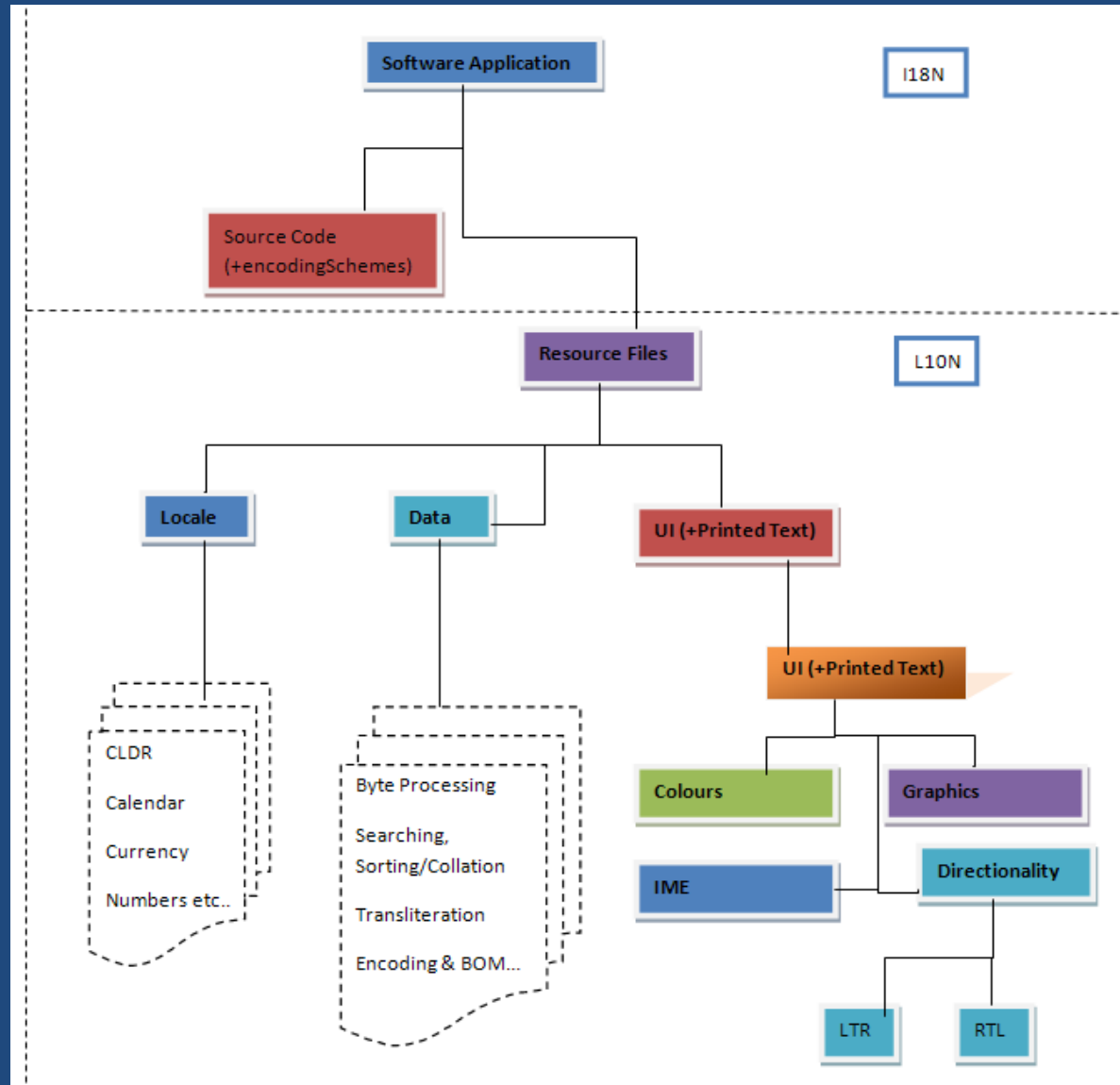# How they are related?
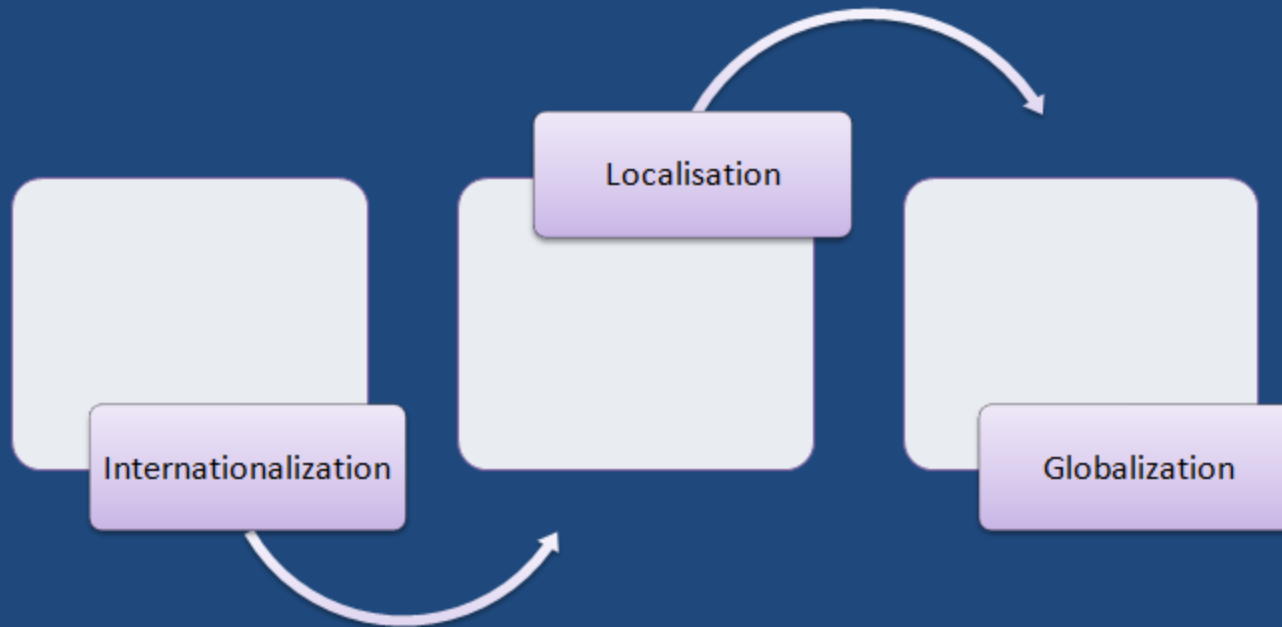
Internationalization frameworks

- GNU Gettext (PO, MO)
- Microsoft Internationalization API (RESX)
- Java Internationalization (Resource)
- Ruby On Rails i18n API (YML)
- Mozilla Applications (DTD, Properties)
- PHP Scripts (INI, PHP)

Localization in Action

Overview of Proposed

Localization Guidelines

Standard

"a document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context".

Localization Standard: XLIFF

An XML based intermediate format which is used to store, carry and interchange localizable data. According to XLIFF specification *13+: "XLIFF is the XML Localization Interchange File Format designed by a group of software providers, Localization service providers, and Localization tools providers. It is intended to give any software provider a single interchange file format that can be understood by any Localization provider."

```xml
<xliff version="1.2">
 <file original="Graphic Example.psd"
  source-language="en-US" target-language="ja-JP"
  tool="Rainbow" datatype="photoshop">
  <header>
   <skl>
    <external-file uid="3BB236513BB24732" href="Graphic Example.psd.skl"/>
   </skl>
   <phase-group>
    <phase phase-name="extract" process-name="extraction"
     tool="Rainbow" date="20010926T152258Z"
     company-name="NeverLand Inc." job-id="123"
     contact-name="Peter Pan" contact-email="ppan@example.com">
     <note>Make sure to use the glossary I sent you yesterday.
      Thanks.</note>
    </phase>
   </phase-group>
  </header>
  <body>
   <trans-unit id="1" maxbytes="14">
    <source xml:lang="en-US">Quetzal</source>
    <target xml:lang="ja-JP">Quetzal</target>
   </trans-unit>
   <trans-unit id="3" maxbytes="114">
    <source xml:lang="en-US">An application to manipulate and
     process XLIFF documents</source>
    <target xml:lang="ja-JP">XLIFF 文書を編集、または処理
     するアプリケーションです。</target>
   </trans-unit>
   <trans-unit id="4" maxbytes="36">
    <source xml:lang="en-US">XLIFF Data Manager</source>
    <target xml:lang="ja-JP">XLIFF データ・マネージャ</target>
   </trans-unit>
  </body>
 </file>
</xliff>
```

Localization Standard: SRX

Can we segment following on the basis of a dot (.) or comma (,) ?

Mr. John, who manages our library, is excellent at teaching mathematics.

Dr. Ray, Block No. 34, J.L.P. Engineering Complex, New Delhi

Localization Standard: SRX

Rules based on XML vocabulary was developed for breaking the text into translatable segments/ smaller fragments. SRX is defined in two parts: <languagerules>: specification about rules applicable for each language. <maprules>: specification about how rules are applied for each language.

Localization Standard: TMX

The translation memory data exchange standard between applications. It is divided into two parts: Translation Unit <tu> and Segment of translation memory text <seg>.

Locale Data

- A locale represents information on the rules and preferences that is specific to the end user's particular country, language and territory.

- The data associated with locale provides support for: presenting, formatting, parsing of local data elements like dates, timestamps, numbers, currencies, measurement units, sort-order (collation), holidays, calendars, translated names for time zones, languages, countries and Scripts.

Locale Data


```
dev@localhost:~

File  Edit  View  Search  Terminal  Help

[dev@localhost ~]$ locale
LANG=en_US.UTF-8
LC_CTYPE="en_US.UTF-8"
LC_NUMERIC="en_US.UTF-8"
LC_TIME="en_US.UTF-8"
LC_COLLATE="en_US.UTF-8"
LC_MONETARY="en_US.UTF-8"
LC_MESSAGES="en_US.UTF-8"
LC_PAPER="en_US.UTF-8"
LC_NAME="en_US.UTF-8"
LC_ADDRESS="en_US.UTF-8"
LC_TELEPHONE="en_US.UTF-8"
LC_MEASUREMENT="en_US.UTF-8"
LC_IDENTIFICATION="en_US.UTF-8"
LC_ALL=
[dev@localhost ~]$ 
```

Minimal set of

localisation guidelines

Point 1: Default Homepage in Marathi

All Contents should be in Marathi on Home Page.

Point 2: All Subsequent Web-Pages in Marathi.

The subsequent linked pages should be in Marathi. Many of the pages, pdf files linked are in English.

Point 3: All Menu titles of the web pages are in Marathi.

The Menu Titles must be in Marathi. We found some of the sites these are in English.

Point 4: All Web-pages developed using UTF-8 encoding.

On many WebPages the character encoding information was not found within the document of view source page. The charset attribute specifies the character encoding for the HTML document. We can declare the UTF-8 encoding in our HTML files using meta charset.

For HTML it is possible to include this information inside the head element near the top of the document:<**meta** http-equiv="Content-Type" content="text/html; charset=utf-8"> HTML5 also allows the following syntax to mean exactly the same:
<**meta** charset="utf-8"> XHTML documents have a third option: to express the character encoding via XML declaration:
<**?xml** version="1.0" encoding="utf-8"**?>**

Point 5: Lang attributes lang="mr" specified.

    lang="mr" can be specified in the <head> tag of view source page.

Point 6: Meta tags defined in Marathi.

    Meta elements are typically used to specify page description, keywords, author of the document, last modified and other metadata.

Code example of meta tags:

```
<head>
<title>Not a Meta Tag, but required anyway</title>
<meta name="description" content="मराठी">
<meta http-equiv="content-type" content="text/html; charset=UTF-8">
</head>
```

Point 7: Are Image ALT/Captions, titles and text in Marathi.

- There should be caption in Maharashtra Emblem.

- ALT/Caption Specifies an alternate text for an image.

- The alt text should describe the image if the image contains information. The alt attribute provides alternative information for an image if a user for some reason cannot view it (because of slow connection, an error in the src attribute, or if the user uses a screen reader). A visually impaired reader using a screen reader will hear the alt text in place of the image.

- *Reference Website*: There should be proper Caption in Emblem.
  http://www.maharashtra.gov.in/en/home

Point 8: Font for the website has been provided through latest HTML5 Font SRC provisions.

With the @font-face rule, web designers do no longer have to use one of the "web-safe" fonts.

Example:

```
@font-face
{
        font-family: fontName
        src: url('fontFile.svg'),//Chrome and Safari
        url('fontFile.eot');//for IE9 browsers
         src: url('fontFile.ttf'),//Firefox and Opera
}
div
{
        font-family:fontName;
}
```

Point 9: All Downloads (PDF, DOC, Excel, etc.) are in Marathi language with Unicode compliance.

All the downloads should be in Marathi Language.

Point 10: All page titles in Marathi.

All Page titles should be Marathi. We found many page titles are English.

Point 11: Numbers should be in Marathi on all pages, Documents, PDFs, Images, etc.

Point 12: Contact us information given in Marathi.

Point 13: All user defined alert/error/pop-up messages are in Marathi.

Point 14: Feedback form is available should be in Marathi.

Point 15: Typing facility in Marathi is given for Interactive Website.

Point 16: Typing should be INSCRIPT layout supported.

Point 17: Onscreen Floating keyboard is available.

Point 18: Provision for increasing font size is available.

Example of provision of increasing font size of web can be seen in

Maharashtra Govt. Website. http://maharashtra.gov.in/

Point 19: In-site search support is available for Marathi language.

An in-site search is a site-specific search field. This is an internal search, which searches your website for content that matches the visitors query.

Point 20: **Website works on Hand held devices.**

We require this Information from the Website Information Manager, Since this depends on the Dept. to Dept. whether they have enabled for Handheld devices or not.

Point 21: Site map of website is in Marathi.

Sitemaps provide a way for Web sites to specify what pages within the site should be indexed and what new content has been added. Basically, it provides a communication channel between the search engine and the site. A sitemap is an XML file that contains a list of site URLs and related attributes detailing what should be indexed within a specific site. It must be UTF-8 encoded.

Reference site for Website map:

http://www.maharashtra.gov.in/web/guest/site-map

Meta Data and

Data Standards

Metadata

Metadata

Metadata is key to ensuring that resources will survive and continue to be accessible into the future.

MDDS: Metadata and Data Standards Template
Based on eGIF (e-Governance Interoperability Framework) Standard of UK

#	Item	Description
1	Name	Name / Number of the Generic or Custom Data Element
2	Description	A simple and ambiguous definition of Generic or Custom Data Element.
3	Type	Generic or Custom Generic : commonly used data element across different e-Governance applications. Custom: Used in a particular application only
4	Is Part of	
5	Parts if any	
6	Data Format	Varchar/Character/Decimal(for real/ floating number) / Integer(Whole number)/Date etc Recommended style of printing / display, if required so
7	Max Size	Maximum Size of the data element
8	Validations	Generic Validations for Generic Data and Specific Validations for Custom Data to be applied for acceptance of data.
9	Values	List of Acceptable Values
10	Default Value	For any list of values, the default value to be used unless otherwise

MDDS: Metadata and Data Standards

Name of Data Element : Gender Identification Code (G01.03)

Description of Data Element	**Gender Identification Code of a Person**
Data Element Type (Generic / Custom)	Generic
Is part of any	
Parts if any	
Data Format	Char
Max Size	1
Validation	
Values	**M** - Male **F** - Female **T** - Transgender
Default value	
Owner	Office of RGI
Based on	-New Zealand- e Gov Standard, (http://www.e.govt.nz/Standards/e-gif/authentication/data-formats-v1.1/chapter11.html) (broken Link. Why not use ISO/IEC 5218:2004) -Census of INDIA

MDDS: Metadata and Data Standards

Name of Data Element : Marital Status (G01.04)	
Description of Data Element	Code for **Marital Status** of the Person
Data Element Type (Generic / Custom)	Generic
Is part of any	
Parts if any	
Data Format	Integer
Max Size	1
Validation	
Values	1 - Never Married 2 - Currently Married 3- Widow / Widower 4- Divorced 5- Separated
Default value	1- Never Married
Owner	Office of RGI
Based on	-Australian Govt Institute of Health & Welfare http://meteor.aihw.gov.au/content/index.phtml/itemId/291045

MDDS: iso 639-3 language codes

Recognized Official Language Code	Values	As per ISO 639-3
1	Assamese	asm
2	Bengali	ben
3	Bodo	brx
4	Dogri	doi
5	Gujarati	guj
6	Hindi	hin
7	Kannada	kan
8	Kashmiri	kas
9	Konkani	kok
10	Maithili	mai
11	Malayalam	mal
12	Manipuri	mni
13	Marathi	mar
14	Nepali	nep
15	Oriya	ori
16	Punjabi	pan
17	Sanskrit	san
18	Santali	sat
19	Sindhi	snd
20	Tamil	tam
21	Telugu	tel
22	Urdu	urd
99	Other language (English)	eng

MDDS: Metadata and Data Standards

Religion Code	Values
1	Buddhism
2	Christianity
3	Hinduism
4	Islam
5	Jainism
6	Sikhism
99	Other

Appellation Code	Values in English
1	Mr.
2	Mrs.
3	Ms.
4	Shri
11	Dr.
12	CA
13	Er.
14	Prof.

MDDS: Metadata and Data Standards

Suffix Code	Values
1	IAS
2	IPS
3	IFS
4	MBBS
5	BDS
6	MD
7	MS
8	MDS

Relationship Code	Values
1	Self
2	Spouse
3	Father
4	Mother
5	Son
6	Daughter
7	Brother
8	Sister
9	Father- In- Law
10	Mother- In- Law
11	Brother-In-Law
12	Sister-In-Law
13	Nephew
14	Niece
15	Grandson
16	Granddaughter
17	Grandfather

Ref no. of Generic data element for its Metadata	Name of Data element	Description of Data element	Data format	Maximum Size
G01.03	Gender Identification Code	M - Male F - Female T - Transgender	Char	1
G01.04	Marital Status	1 - Never married 2 - Currently married 3 - Widow / Widower 4 - Divorced 5 - Separated	Integer	1
G01.05-01	Appellation Code	An Appellation is a title for a Person like Mr., Dr. etc. to be prefixed with the name to indicate person's gender, marital status, Professional status etc. Values as per code directory (CD01.04) **Note: Maximum of two Appellations allowed for a person**	Integer	2
G01.06-01	Suffix Code	**Suffix** to the name of the Person to indicate person's positional status like IAS, IPS etc. Values as per code directory (CD01.05)	Integer	2
G01.07-01	Relation Type	H- Head of house hold N- Not head of household (Default value "N")	Char	1
G01.08-01	Relationship Code	Relationship of the Person, with head of the family like self, sister, brother etc.	Integer	2