

## Localization in e-Governance Perspective

WELCOME  
सुस्वागतम्  
நல்வரவு  
வூவ்வுரவு  
సుసాగతం  
सुस्वागतम्  
সুসাগতম  
ಸುಸ್ವಾಗತೆ  
സുസ്വാഗതം  
সুসাগতম  
ಸುಸ್ವಾಗತಮ  
सुस्वागतम्  
خوش آمدید

Akshat S. Joshi  
C-DAC, Pune  
akshatj@cdac.in

Venue :  
Circuit House,  
Guwahati

28<sup>th</sup> May 2013

# Brief outline

- Importance of localization
- Internationalization / localization
- The game changer – Standards
- Indian standards – GIGW, e-governance localization best practices, (RFP)
- Various standards in context of Indian languages
- Surface level localization – what it takes
- Deeper level localization – what it takes

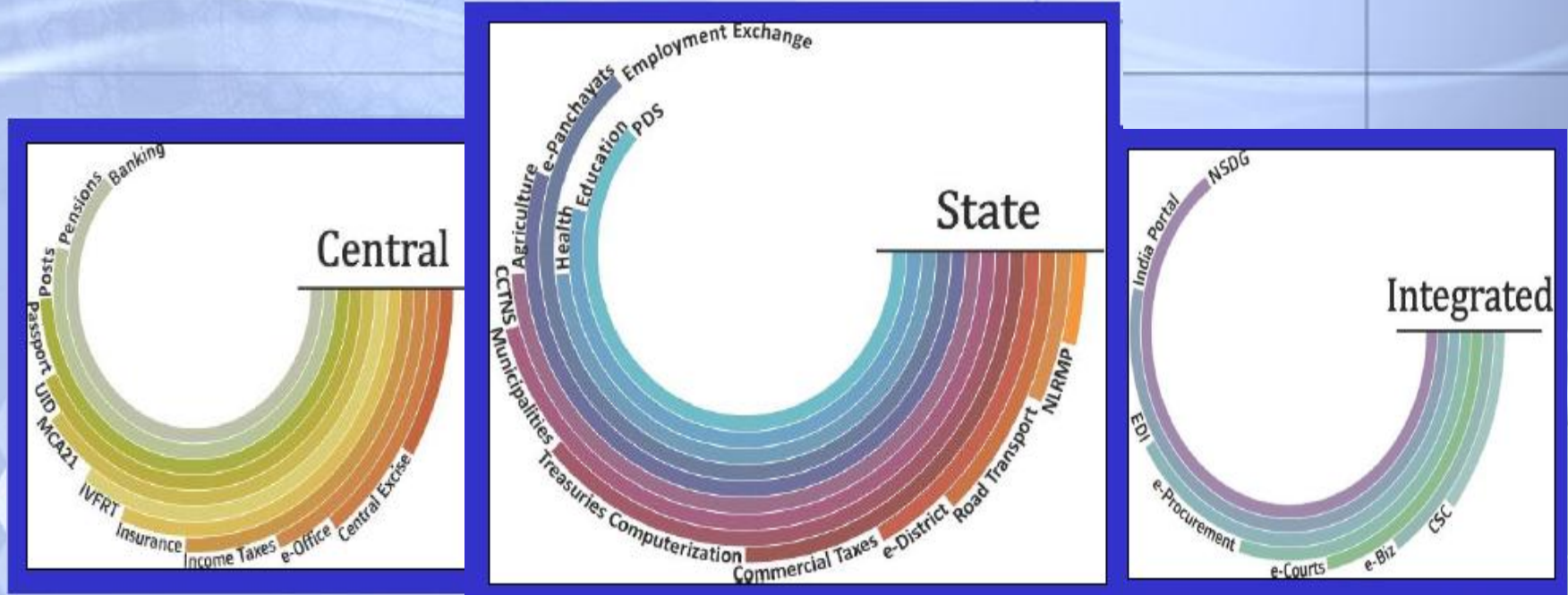
## C-DAC Offering:

- Localization Framework
- Leveraging on prior knowledge – collaborative model



# Mission Mode Projects

**NeGP comprises of 31 Mission Mode Projects (MMPs) encompassing 11 Central MMPs, 13 State MMPs and 7 Integrated MMPs spanning multiple Ministries/ Departments.**





Hence Indian Language Interface becomes very crucial to the easy access to the information



**साहित्य अकादेमी**  
**SAHITYA AKADEMI**  
National Academy of Letters

- About Us
- Fellows
- Authors
- Literacy Activities
- E-Newsletters
- Awards
- Library
- New Arrivals
- Projects & Schemes
- Policies
- Cultural Exchange
- Publications

**Awards**

Akademi Awards

Even from its inception in 1954, the Sahitya Akademi awards prizes to the most outstanding literary.

more...



**साहित्य अकादेमी**  
**SAHITYA AKADEMI**  
नेशनल एकेडेमी ऑफ लेटर्स

- हमारे बारे में
- फेलो
- लेखक
- साहित्यिक क्रियाएँ
- ई-समाचारपत्रिकाएँ
- पुरस्कार
- नीतियाँ
- पुस्तकालय
- नए आगमन
- परियोजना और योजना
- सांस्कृतिक विविधता कार्यक्रम
- प्रकाशन

**पुरस्कार**

अकादेमी पुरस्कार

1954 में अपना स्थापना के समय से ही साहित्य अकादेमी प्रतिवर्ष अपने द्वारा मान्यता प्रदान भारत की प्रमुख भाषाओं

अधिक जानकारी...

**English**

अकादेमी पुरस्कार

जिनमें संबंधित भाषाओं के प्रचार, आधुनिकीकरण या उसके संघर्ष के लिए यथार्थ योगदान दिया हो, इस सम्मान से विभूषित किया जा सके

अधिक जानकारी...

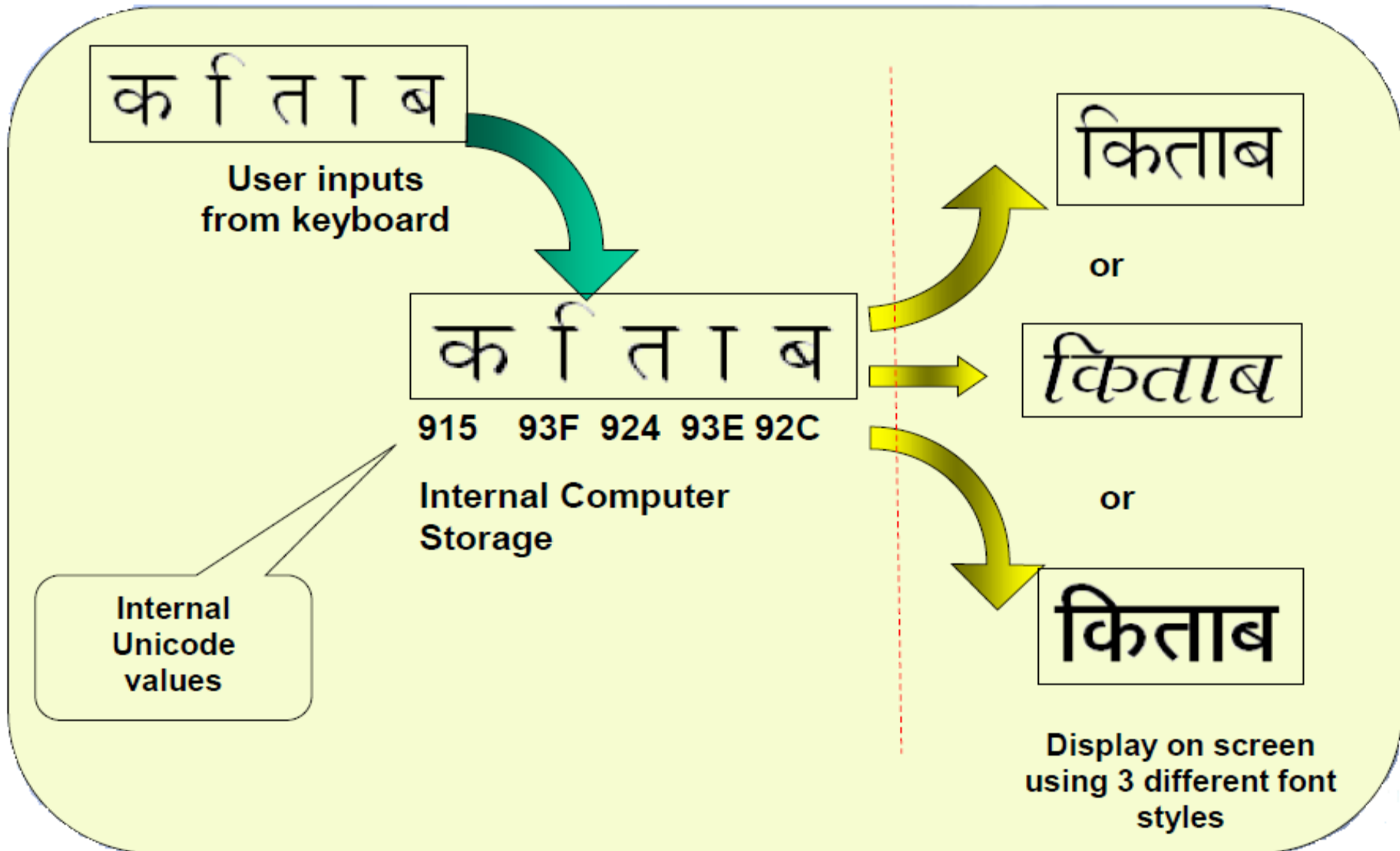
**Hindi**

अकादेमी पुरस्कार

जिनमें संबंधित भाषाओं के प्रचार, आधुनिकीकरण या उसके संघर्ष के लिए यथार्थ योगदान दिया हो, इस सम्मान से विभूषित किया जा सके

अधिक जानकारी...

# Akshara Formation: A glance



## One script :: many languages

- Devanagari - Hindi, Marathi, Konkani, Rajasthani, Nepali, etc.
- Thus the code page Devanagari can support all languages using that particular script.

## One language :: many scripts

- Konkani is written in Roman, Devanagari, Malayalam and Kannada.
- Sindhi is written in Gurmukhi (Punjabi), Arabi (Perso-Arabic), Devanagari, Gujarati and also Roman.
- Sindhi has adopted the Perso-Arabic script for representing their language. In case of Konkani, Devanagari is used as official script.



# Indian language complexities at syllable formation level

In Indian language scripts, a group of characters together form a syllable or a cluster

Syllable = Multiple character Sequence

eg. "brahmaa" (ब्रह्मा) has 2 syllables - 'bra' (ब्र) and 'hmaa' (ह्मा)

'bra' is a sequence of 3 character codes

ब्र = ब + ा + र.

'hmaa' is a sequence of 4 character codes

ह्मा = ह + ा + म + ा.



# Indian language complexities at rendering order level

Many to many correspondences between character code and font code.  
The character code sequence and the display sequence can differ and hence repositioning may be required at font level.

क ि त ा ब

(input from keyboard)

क ि त ा ब

(internal representation)

(179+219+194+218+202)

किताब

(ि+क+त+ा+ब)

(Out put on screen/printer)

(Display composition using font)



## Indian language complexities – alternate spellings

Alternate spellings in Indian languages may pose problems while searching.

हिंदी : ह + ि + ं + द + ी

हिन्दी : ह + ि + न + ् + द + ी

तसवीर : त + स + व + ी + र

तस्वीर : त + स + ् + व + ी + र



# What needs to be done

## **Follow internationalization best practices which will help**

- To build globally acceptable software
- Isolate translatable and non-translatable resources
- Universal encoding
- Interoperability
- Easy localisation

## **At MMP Level**

- Include best practices in RFP
- Ensure adherence of these standards at every stage of MMP development
- Train stakeholders / developers in “localisation”

## Best practices to be included in the RFPs

- **GIGW – Government of India guidelines for Web**
- **e-Governance localization guidelines**
- **W3C standards**
- **WCAG**
- **Translation style guides**



# Internationalization & Localization

## Internationalization

- Internationalization is the process of designing a software application so that it can be adapted to various languages and regions by reducing engineering efforts

## Localization

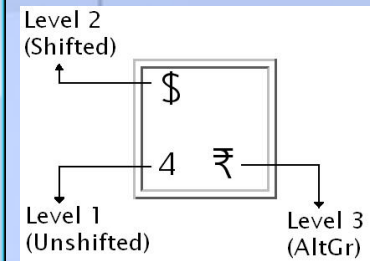
- Adaptation of a product, application or document content to meet the language, cultural and other requirements of a specific region (a "locale").
- Apart from translation of the UI and documentation, it covers
  - Numbers, date, time formats, currency, keyboard, Collation, Symbols, icons, colors and many more.



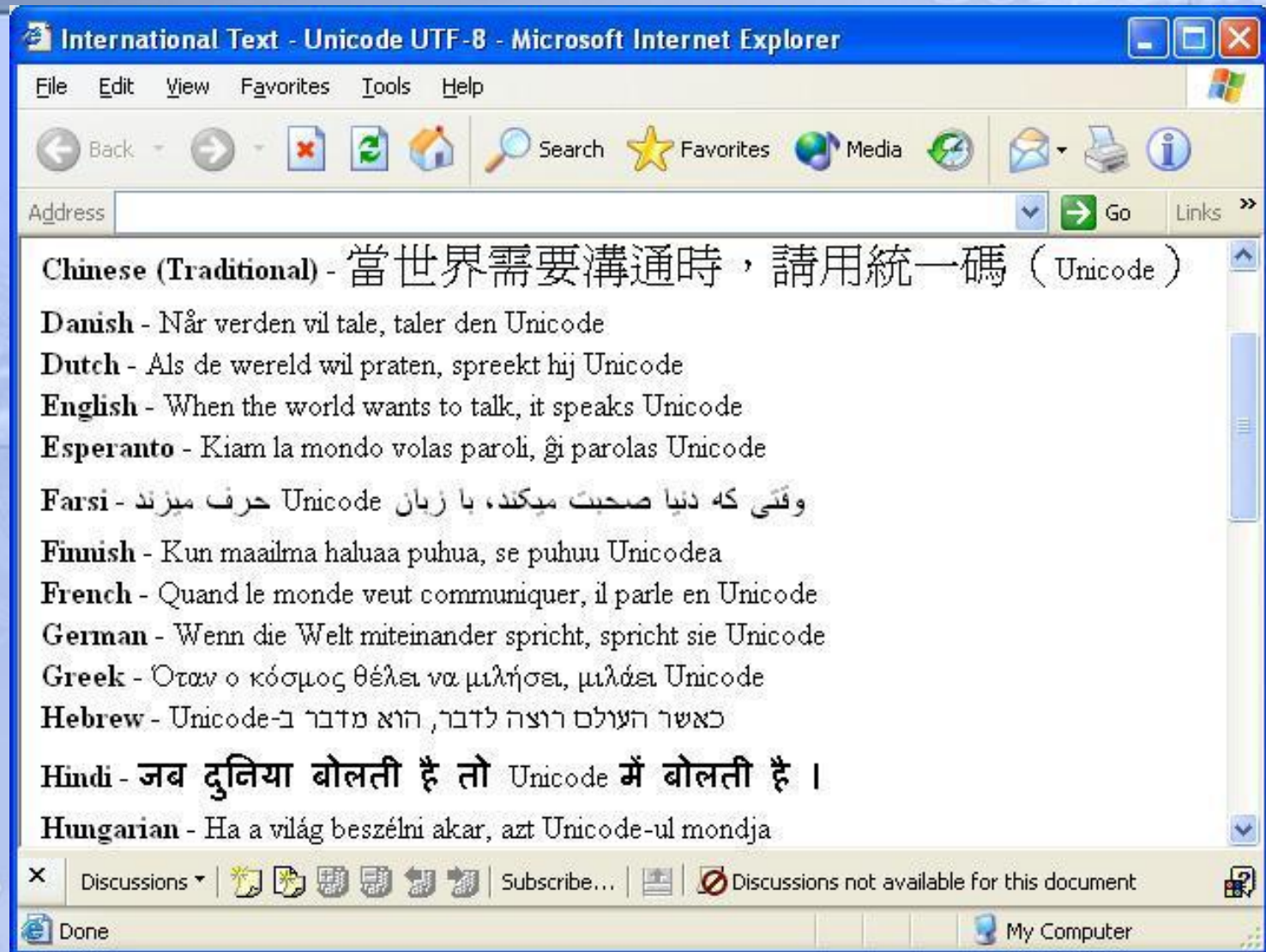
## Standards to be followed

- Inputting standards – Enhanced INSCRIPT Keyboard. (Onscreen Keyboard for apps)
  - Stylus based devices (tablets / mobiles) – either INSCRIPT keyboard and or 12 keys Virtual Keyboard
  - Non-stylus based devices (tablets / mobiles) – physical INSCRIPT or 12 keys keyboard.
- Storage (Universal encoding – UTF8 / Unicode)
- Display - Unicode compliant “Open Font Format” – on lines of SakalBharati.
- FUEL – Frequently Used Entry List
- Term banks, Domain specific terms
- CLDR – Common Locale Data Repository
- Roman phonetic equivalents.
- W3C / WCAG(Web Content Accessibility Guidelines)
- Desirable to have Content Management System at backend for all the MMPs – for ease of localization.

# Enhanced Inscript Keyboards



# Unicode



# Unicode Compliant – Open Type Fonts

Sakal Bharat **Assamese** Normal

কঁহুৱা বন মোৰ অশান্ত মন,

Sakal Bharat **Devanagari** Normal

भारतीय भाषाएं और सामाजीक परिवर्तन ।

Sakal Bharat **Gujarati** Normal

ભારતીય ભાષાઓ અને સામાજિક પરિવર્તન ।

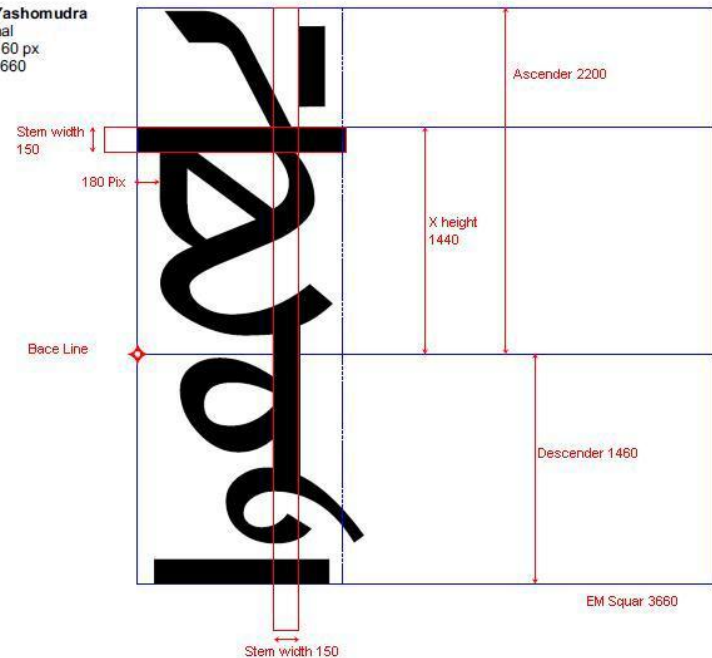
Sakal Bharat **Oriya** Normal

ଇଂଗିଡ଼ ପ୍ରଶ୍ନ ଦୟାକରି ମନୋନୀତ କରନ୍ତୁ

Sakal Bharat **Punjabi** Normal

ਪੰਜਾਬ ਪੰਜ ਦਰਿਆਵਾਂ ਦੀ ਧਰਤੀ ਹੈ।

The Sharir: Yashomudra  
Weight: Normal  
Stem width: 160 px  
Em square: 3660

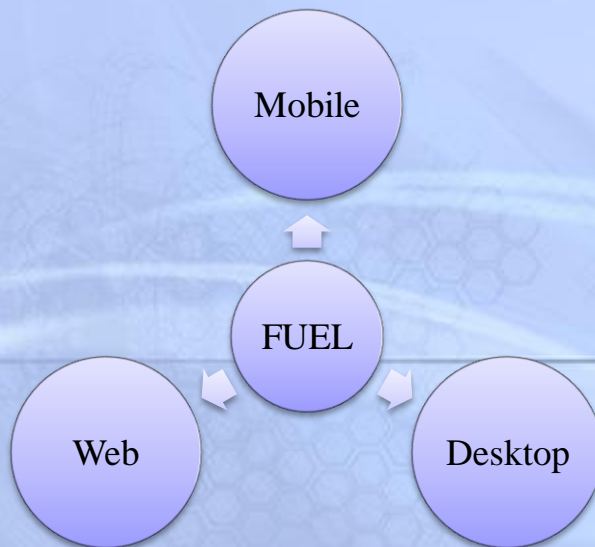


کنپیوٹر کی دوڑ میں ہم کہاں ہیں؟



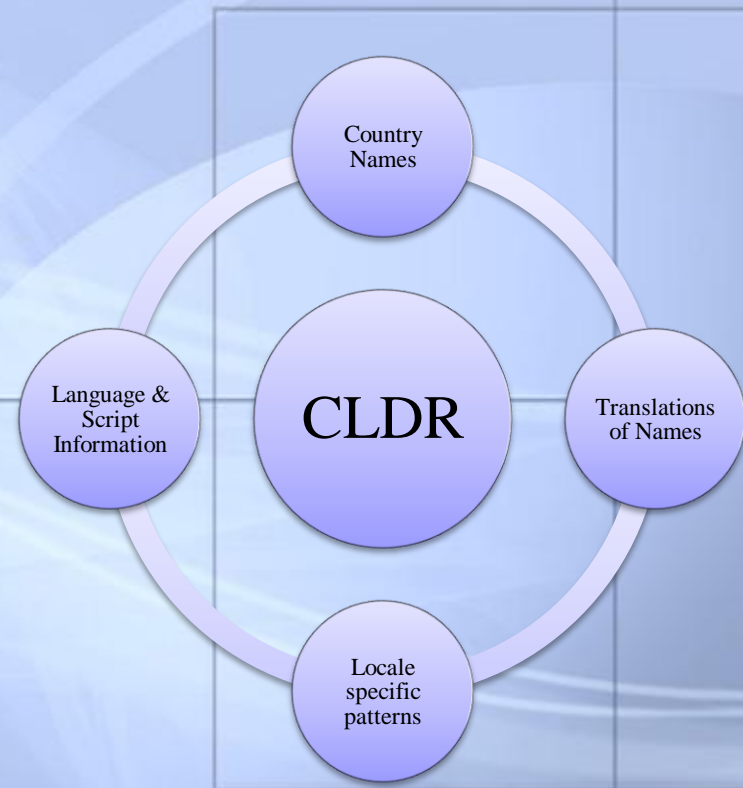
## Some of the important initiatives

- Frequently Used Entry List



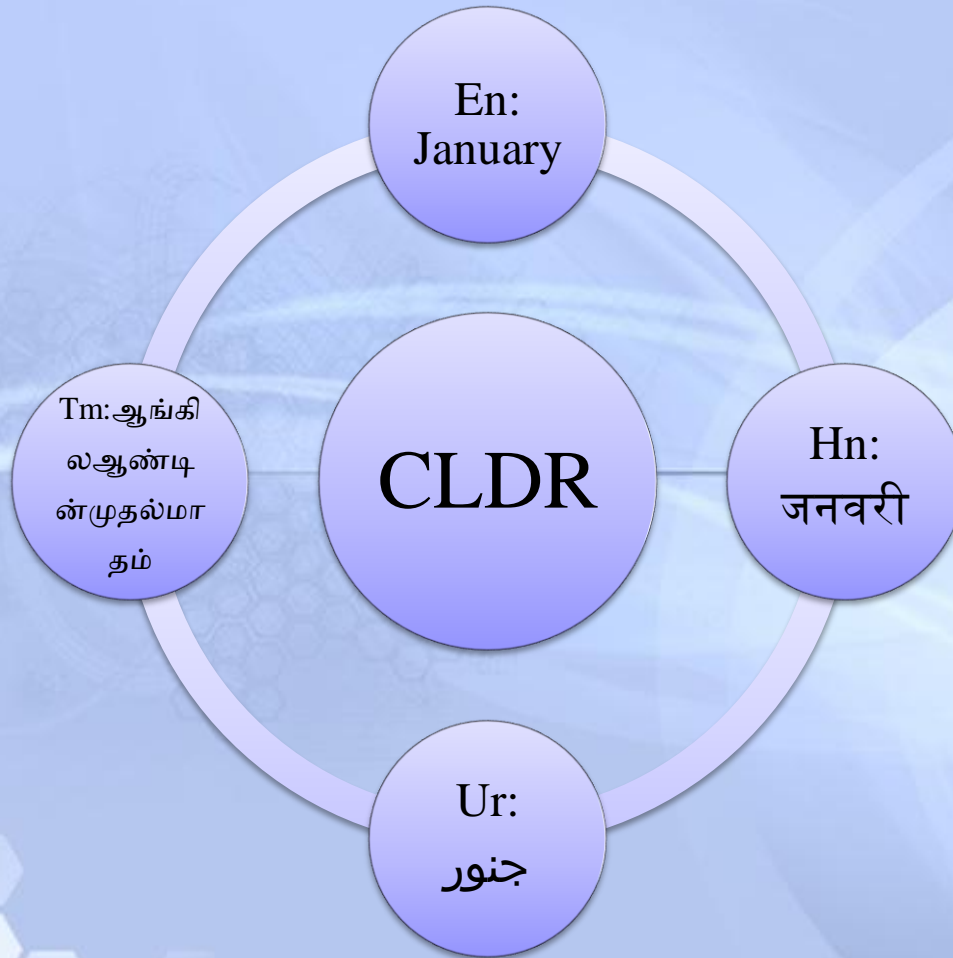
- an open source initiative to standardize terms for open source software programs
- aimed at resolving the problem of term inconsistency and lack of standardization in Computer software translation, across various platforms.

- Common Locale Data Repository



- Numeric Formats, Date time formats, Currency specifiers

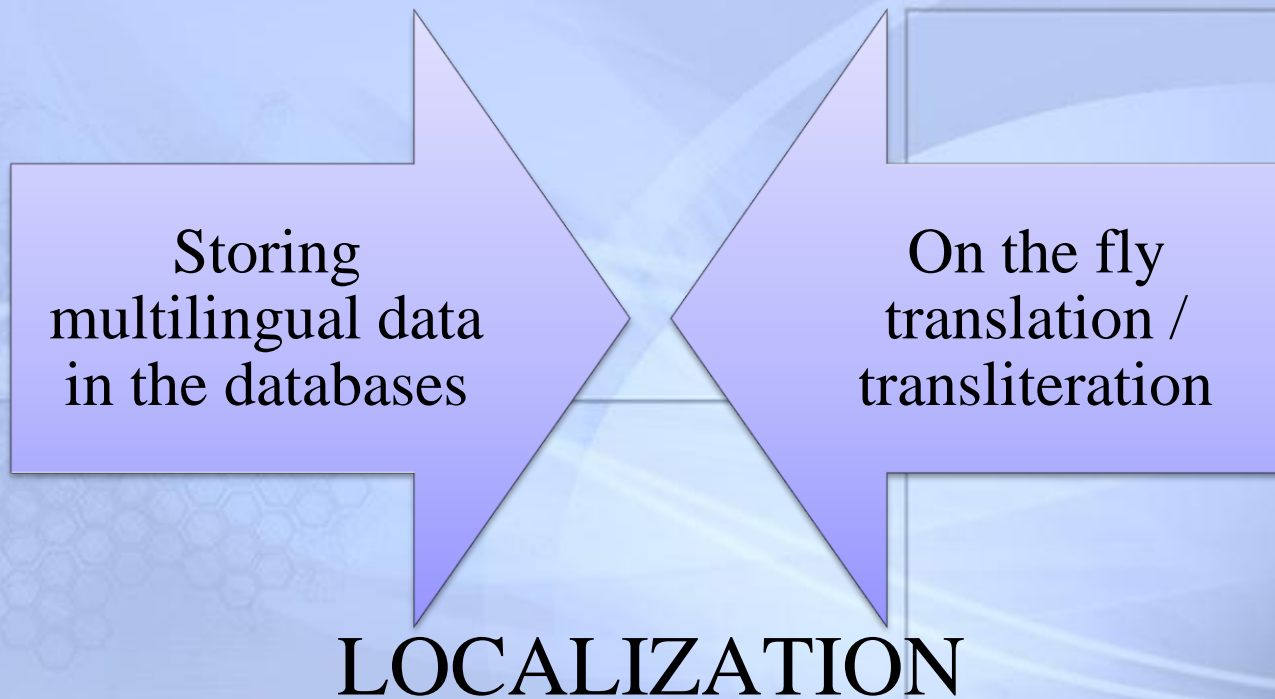
# CLDR – An example



```
<monthWidth type="wide">
<month type="1">जनवरी </month>
<month type="2"> फरवरी </month>
<month type="3">मार्च </month>
<month type="4">एप्रैल </month>
<month type="5">मेई </month>
<month type="6">जून </month>
<month type="7">जूलै </month>
<month type="8">अगस्त </month>
<month type="9">सितंबर </month>
<month type="10">अक्तूबर
</month>
<month type="11">नवंबर </month>
<month type="12">दिसंबर
</month>
```

For further details please refer:  
<http://cldr.unicode.org/>

## Localization – Different techniques



# Searching in Indian Languages

Alternate  
spellings

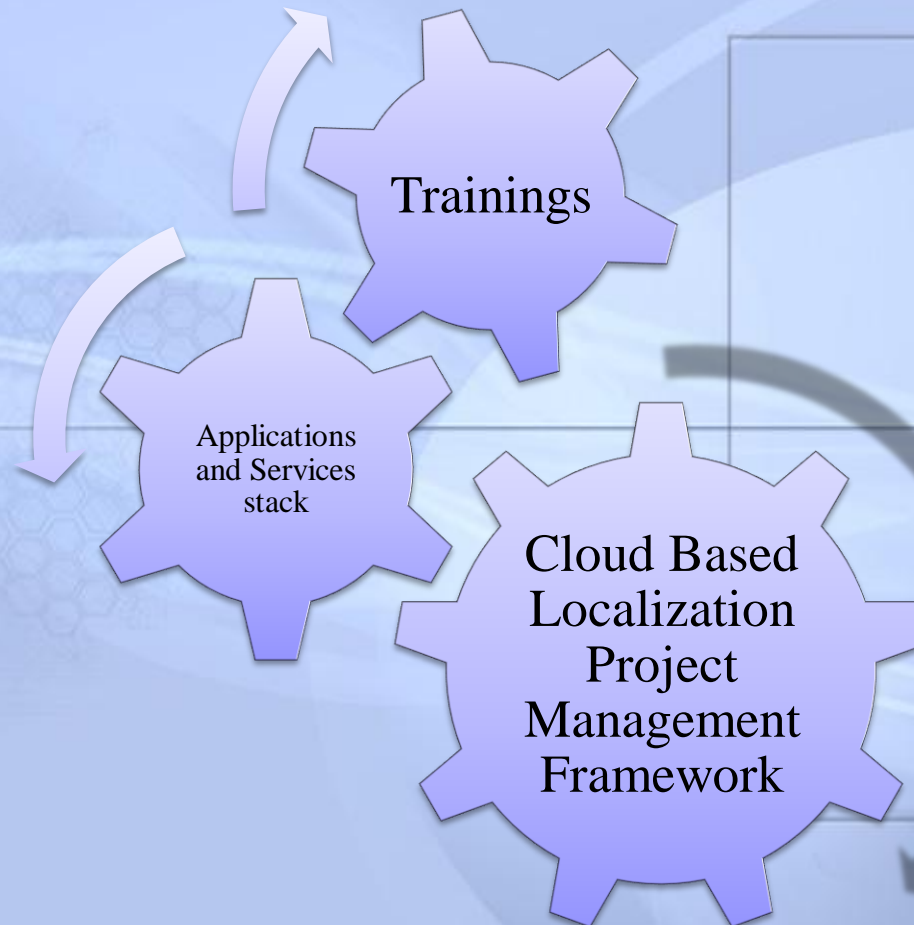
- Cognitively
- Visually
- Culturally
- Historically

Indian Language  
Numerals

- When a numeral in Indian Language is encountered, the software should treat it on its numeric value.



# How C-DAC can help



# Why LPMF ?

- Hurricane task to enable all the MMPs with region specific Indian languages if the architecture is not internationalized.
- Difficult to have seamless access to information without standards and process flow
- For SIM (SIMultaneous) release in multiple languages, application developers, translators, evaluators need to work in tandem.
- Volume, last minute localization
- Non-availability of skilled manpower, especially in the field of language translation

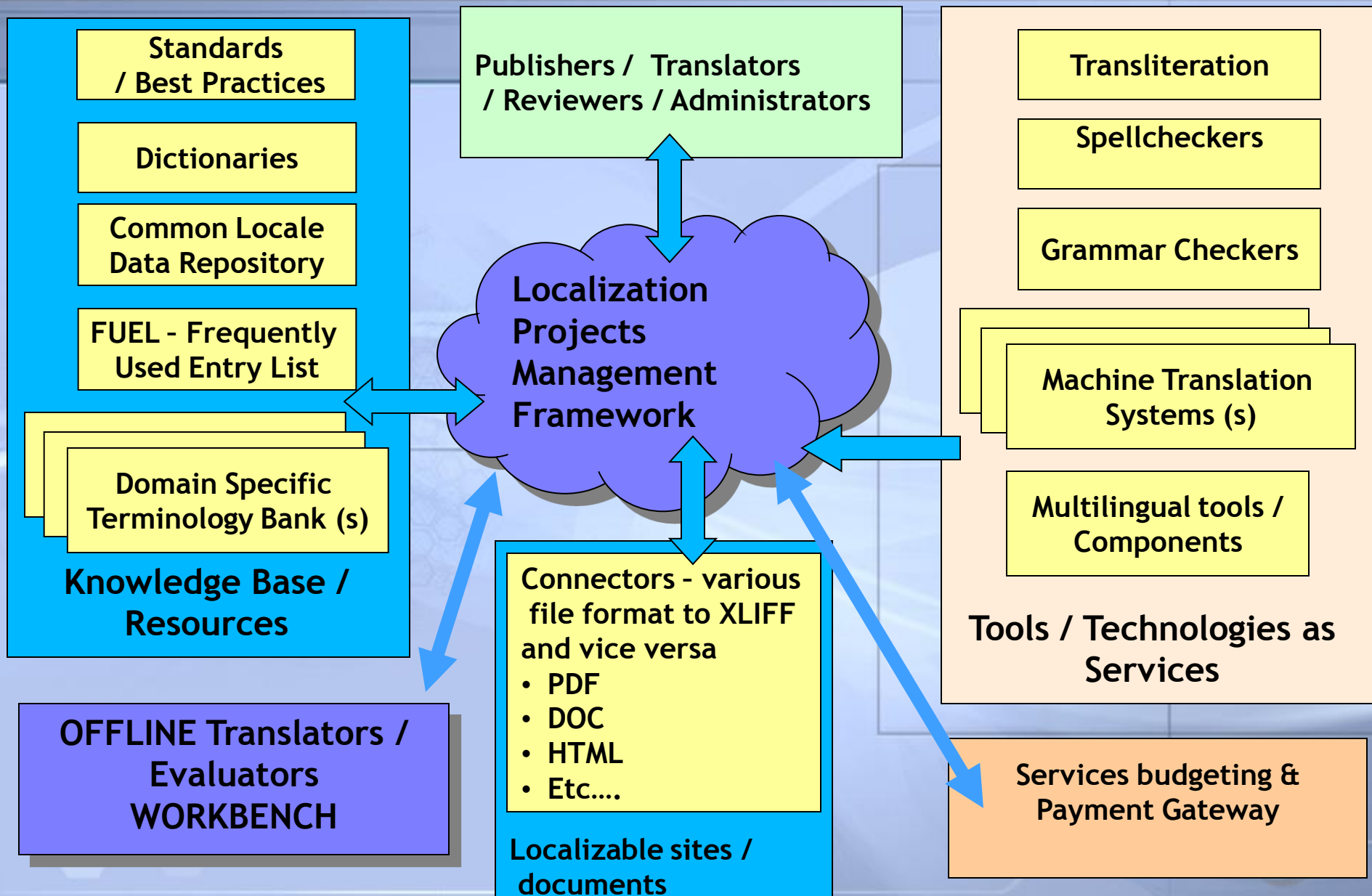
*Workflow software can greatly streamline the localization process by managing the sequence of review/edit tasks, providing the status of tasks and processes, and notifying participants of changes in state, new work, or other information*



# How LPMF ?

- **Unified standards – Unicode, XLIFF, TMX, SBX, CLDR to name a few.**
- **Localization Projects Management Framework offered as CLOUD service.**
- **Linguistic tools and resources**
  - Term Banks, Dictionaries, CLDR, FUEL
  - Spellchecker, grammar Checker, Transliteration tools
  - Search engines, Machine Translation System
- **Entire Work Flow Management**
  - Publisher, Translator, evaluator, administrators modules managements
- **Services offering and budgeting**
  - Document, books, applications localization services
- **Community participation**
  - Crowd sourcing, knowledge sharing, chat, RSS feeds.
- **Payment Gateways**
  - For payments to translators, evaluators and relevant stakeholders
- **Facility to plug-in Third party tools / technologies.**

# Localization in Cloud





## Salient features of LPMF

- Workflow optimization
- Knowledge Repository
- Standards usage - XLIFF
- Machine Translation System (Consortia MT engine)
- Synchronization of various components
- For surface localization no need of source code (on the fly)
- Bringing translators, evaluators, Project Managers, publishers, tools provider on a single platform.
- Workflow saves money, time and infrastructure.
- Believes on leveraging prior / existing knowledge

# Translation / Transliteration through a browser toolbar

iPlugin Highlight Scriptkey Keyboard Floating Keyboard Help Convert

EnglishToHindi  
HindiToUrdu  
UrduToHindi

BBC Home News Sport Radio TV Weather Languages

BBC HINDI.com

हमारे कार्यक्रम सुनें रेडियो प्लेयर चलाएँ

مدد جائی اسے یومید بنائیں

بہار بٹا

بھارت اور بڑوس  
کھیل کی دنیا  
منورجن ایکسپریس  
آپکی رائے  
گچہ اور جائے

Learning English

ریڈیو

ہمارے کاریہ کرم  
ہزاروں سے

سٹوڈنٹس

ہم کون ہیں  
ہمارا بتا

05:48 تک کے سماچار GMT منگلوار 07 اگست 2007 کو

بازہ بیڑوں تک راحت پہنچانے میں  
بھارت سمیت دکن ایشیا کے کئی دیہوں میں بازہ  
سے پر بھاوت لوگوں تک راحت سامگری پہنچانے  
میں راحکرمیوں کو ڈھکوں کا سامنا کرنا پڑ رہا ہے  
ناؤ ڈرگھٹنا میں 15 سے اذہیک موتیں +  
...ہمت میں بی لکھا ہے!

ہندو فرائٹ ڈاؤنلوڈ  
یاد آپ اس پےج کو ہندی میں نہیں دیکھ  
پا رہے ہیں تو کپڑا یہاں کلک کرے

مدد کے لیے کلک کرے

مہبئی ٹرین دھاکے میں 13 کے خلاف اروپ  
"حطروں سے بیٹھے کے لئے ٹار بھارت"  
بانچ اور سٹی منکرہوں نے ہشکار کیا

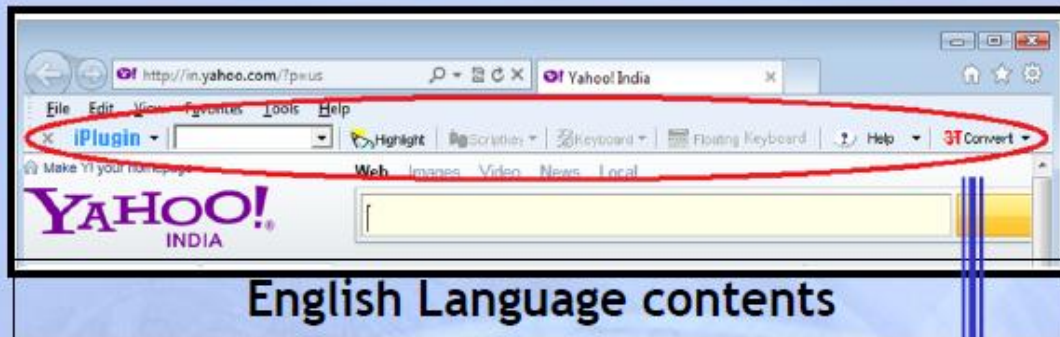
سلاخوں کے پیچھے سجنے

پروہما ہم پہلے کی 62  
برسی  
دیکھیں

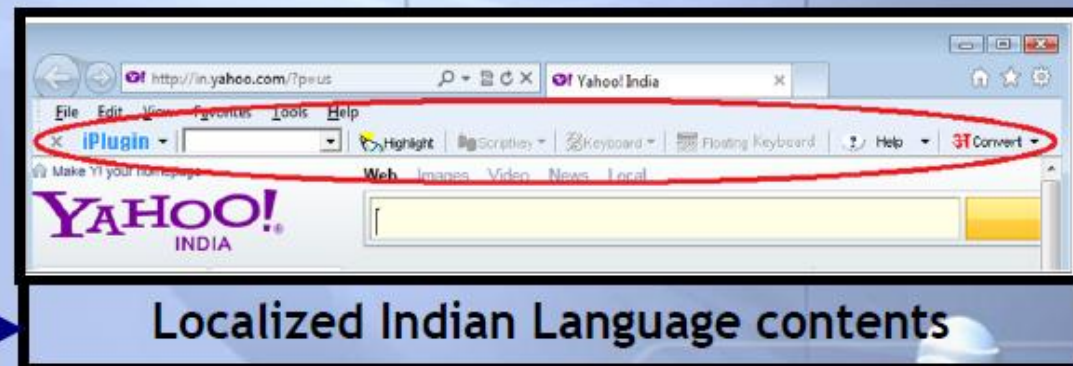
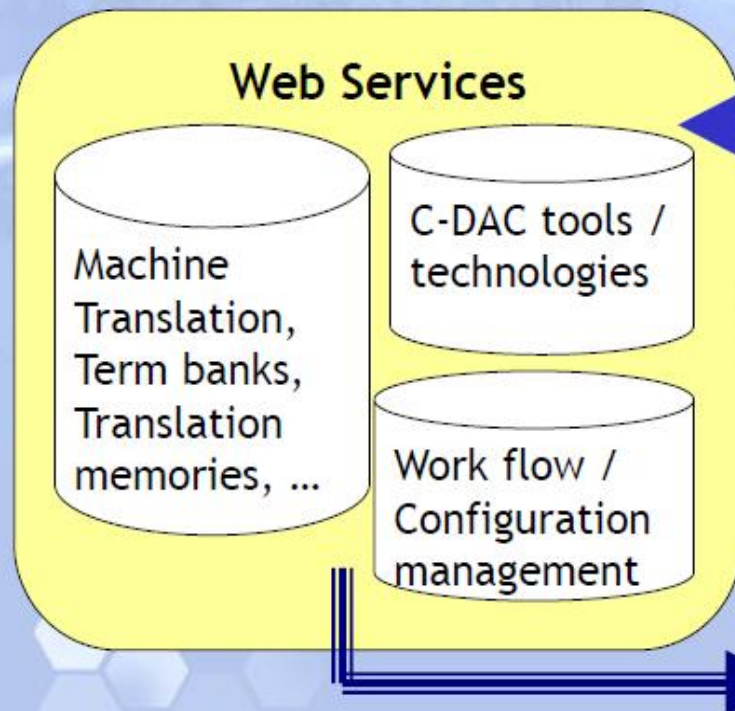
تائیان کو سمات کرنے پر  
سہمنی

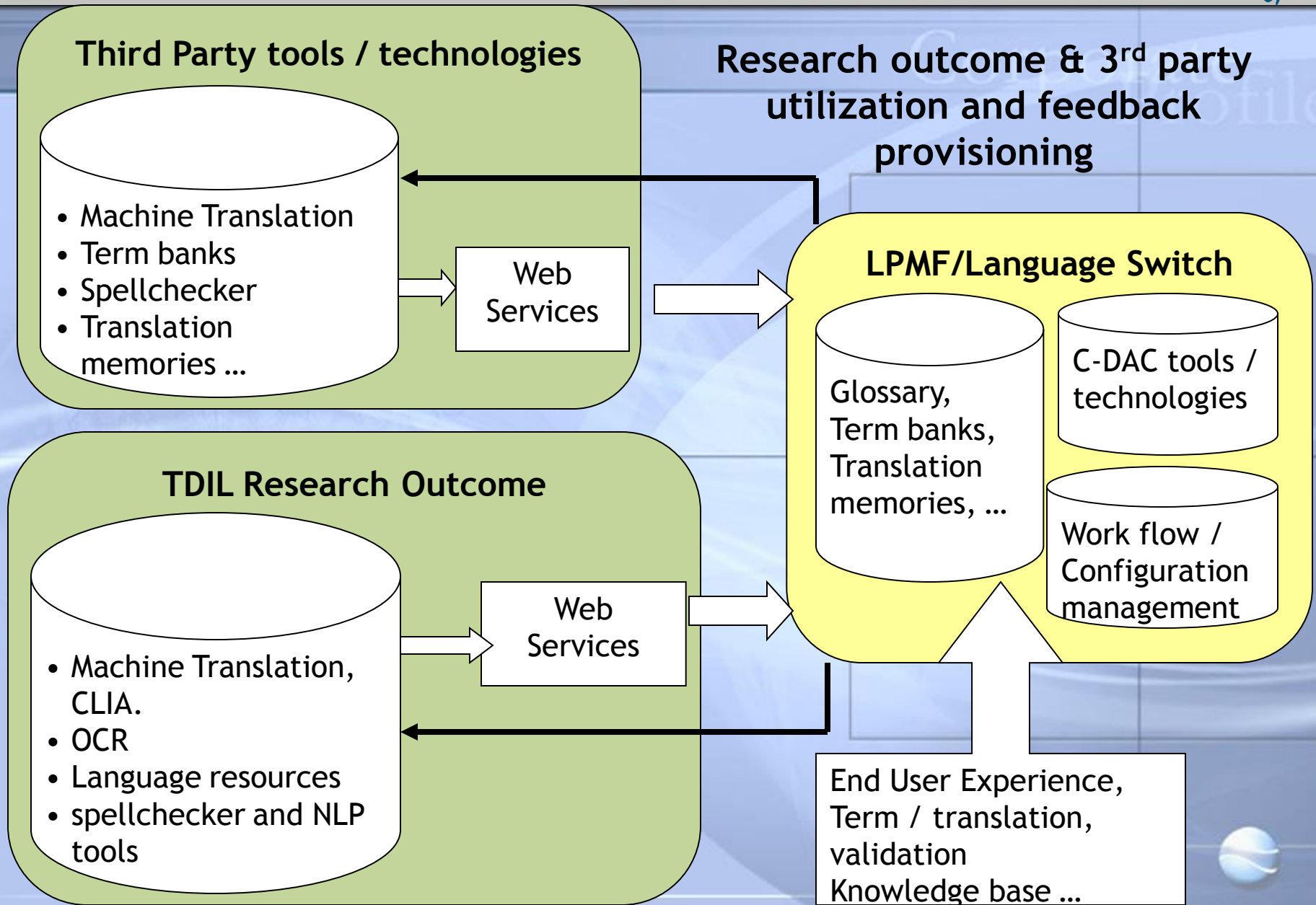
نوبٹس 20 ورڈ کب کے لئے  
بھارنہ ٹیم

# Surface localization Plug-ins for various browsers



The plug-in invokes various web services from within, such as MAT, Spell check, Dictionaries, URL specific Term banks, Translation memories







# Community Participation – Improving the translations

## Validating and improving the translations of NROP I using public voting system

Home | About System | Services | Contact Us

Select Source Language: English  
Select Target Language: Hindi  
Word to Update: bookmark

Search The Word

Source Language: English  
Target Language: Hindi

Search Word: bookmark

Search The Word

Dict_Tab	EXPORT	ENGLISH	Hindi Meaning
Edit	<input type="checkbox"/>	Organize Bookmarks...	पुस्तकचिह्न
Edit	<input type="checkbox"/>	Search Bookmarks and History	पुस्तकचिह्न [100 %] सभी टैब पुस्तकचिह्न [0 %]
Edit	<input type="checkbox"/>	Bookmark This &Link	सभी टैब पुस्तकचिह्न करें (&T)...
Edit	<input type="checkbox"/>	Edit this bookmark	सभी टैब पुस्तकचिह्न बनाएं...
Edit	<input type="checkbox"/>	Page Bookmarked	
Edit	<input type="checkbox"/>	Bookmark Removed	
Edit	<input type="checkbox"/>	You have chosen a keyword that is bookmark. Please select another	
Edit	<input type="checkbox"/>	<bookmark_value>JDBC; definition	सभी टैब पुस्तकचिह्न करें...
Edit	<input type="checkbox"/>	msgbox FileExists(file:///d:/bookma	पुस्तकचिह्न फाइल
Edit	<input type="checkbox"/>	Drop a link to bookmark it	पुस्तकचिह्न फोल्डर
Next	Last	Bookmark folder	पुस्तकचिह्न प्रबन्धक
Check All	UnCheck All	Bookmark Manager	
	Add To Export	Bookmark Menu	

1 2 3 4

Submit

## Summing Up...

- Standards adherence is crucial (especially Unicode, INSCRIPT keyboard, Open Font Format, Script Grammar-Sort order, GIGW, Localization Guidelines, CLDR, FUEL, W3C and WCAG, Term banks )
- Compliance testing / Validation as per the guidelines / best practices.
- Localization Project management framework for SIM release and updates of e-governance applications.
- Use of content management system at backend is crucial.



# धन्यवाद

*Nurturing living languages*

जीवित भाषाओं का पोषण

# Requirement for Internationalization / Localization





## Need for localization

- Earlier software were designed to support only one language.
- With global markets opening up, there is a requirement for supporting multiple languages and diverse cultures.
- Each country or region needed its own version of software.
- In such scenario no option but to
  - Re-develop the entire software / product for different languages (multiple versions, versioning, tracking, issues, global changes)
  - Translate the software in desired languages.
- The above has fallouts : in first case it is repetition of the coding work & in second case since all the menus, help files, dialog boxes, labels which were hard coded required efforts.
- Also since the architecture was not initially meant for supporting multiple languages couple architectural changes were required.



# Localization

- Localization sometimes referred as "l10n", where 10 is the number of letters between 'l' and 'n'.
- Localization refers to the adaptation of a product, application or document content to meet the language, cultural and other requirements of a specific region (a "locale").
- Though localization is synonymous to translation of the UI and documentation, but is more complex covering
  - Numeric, date and time formats
  - Use of currency
  - Keyboard usage
  - Collation and sorting
  - Symbols, icons and colors
  - Text and graphics containing references to objects, actions or ideas which, in a given culture (cultural issues are more serious as it may be subject to misinterpretation)



# Internationalization and globalization

- Internationalization & globalization refers to same concept.
- Internationalization is referred as "i18n", where 18 is the number of letters between 'i' and 'n'.
- Internationalization includes :
  - enabling the code to use Unicode
  - Enabling the code to support local, regional, language, or culturally related preferences, which typically includes date and time formats, local calendars, number formats and numeral systems, sorting and presentation of lists, handling of personal names and forms of address
  - Proper handling of the concatenation of strings, avoiding hard coded / dependence in code of UI string values, etc.
  - Adding markup in your DTD to support bidirectional text, identify, language, or adding to CSS support for vertical text or other non-Latin typographic features.
  - Separating localizable elements from source code or content, so that user can select the same based on his preferences



Corporate  
Profile

- Globalization is a process of developing, manufacturing and marketing software products that are intended for worldwide distribution
- These products are named as 'globally enabled' as they support multiple languages called as multi-locales.
- An important feature of these internationalized products is that the User Interface is kept separate from the core instruction code allowing the software to be translated without requiring re-compilation.
- Globalization is achieved through Internationalization and Localization.

