

Challenges of Web services in Indian languages

Swaran Lata

Country Manager , W3C India
& Head , TDIL Programme , Dept of Information Technology ,
Govt of India

E-mail : swaran@w3.org, slata@mit.gov.in

Modalities for Government on the Web

- **Provide:** public services on the web, either transactional or information services or both.
- **Engage:** with citizens and businesses, on government terms or on the citizens terms.
- **Enable:** public sector information re-use .

Need For Internationalization in E –governance :

- Adoption of open standards
- Adoption of W3C Internationalization Standards
- Internationalization and localization requirements for schemas
- Implementation of international resource identifiers ,
Example: Language Tags, Internationalization XML,URL.
- Adoption of local terminology and practices
- Script issues in text-to-speech implementations
- International typographic requirements
- Compliance to best practices for seamless web-accessibility

Requirement for Implementation of Indian Language :

- Encoding Standard
- Input Mechanism(Key Board)
- Lexical Resources
 - Glossaries
 - Lexicon
- Rendering/Display
- Transliteration
- Database related issues
 - Database which support the character encoding and locale
 - Indexing and sorting depends on the locale details
- Sorting Order in Database
- Interoperability between vendors, platforms and technologies

Best Practices to be Adopted

❖ Storage :

- UNICODE is the universal character encoding scheme for 22 Indic languages.
- Storage application should be in Unicode.

❖ Keyboard :

- Use Inscript keyboard layout
- The Inscript overlay contains characters required for all the Indian scripts

Best Practices to be Adopted (Contd..)

❖ Fonts :

ISO/IEC 14496-22:2009 specifies the Open Font Format (OFF) specification.

The Open Font Format(OFF) addresses the following goals:

- broader multi-platform support;
- excellent support for international character sets;
- excellent protection for font data;
- smaller file sizes to make font distribution more efficient;
- excellent support for advanced typographic control.

“Localization and Adoption of Open Standards”

- ❖ Need to adopt and follow standards on the language technology components for successful localization and wider access of Information and knowledge
- ❖ Adoption of World Wide Web Consortium standard (W3C) in respect of Localization
- ❖ Web based applications need to be developed in such a way that applications should be interoperable for seamless access of knowledge. W3C develops such technologies with specifications, guidelines, software and tools.
- ❖ Indian Language Web-Browsers need to be W3C compliant
- ❖ W3C guideline may easily ensure interoperability between storage, display and access to service even in the multilingual paradigm.

Semantic Web Implementation in E-governance

- Improving access of resources.
- Enable people to create data stores on the Web, build vocabularies, and write rules for handling data.
- Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.
- E-Government agencies can share and exchange information through semantic web

Common Locale Data Repository

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

Need of Locale Data

Locale data is used for:

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

Issues

The issues that primarily affect the e-governance applications are:

- interoperability between the heterogeneous systems and data repositories,
- absence of standard taxonomy
- absence of compliance to best practices for seamless web-accessibility
- lack of internationalization & localization of the multimodal application interface

Best practices for E-governance applications

- Designing and developing in a way that removes barriers to localization or international deployment. This includes such things as enabling the use of Unicode
- Providing support for features that may not be used until localization occurs
- Enabling code to support local, regional, language, or culturally related preferences.
- Separating localizable elements from source code or content, such that localized alternatives can be loaded or selected based on the user's international preferences as needed.

W3C activity in India

Education and Outreach to all stake holders

Bi-directional communication between Stake holders and W3C Consortium.

Objective of W3C India Office

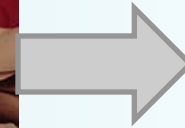
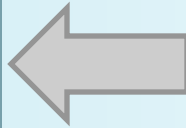
Promotion and proliferation of W3C Standards

feedback to W3C for the existing and future standards of W3C

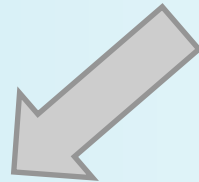
**Autonomous Societies/ Attached
Offices of DIT (STQC, NIC, CDAC,
ERNET)**



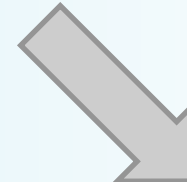
**BIS and other
International bodies
(Unicode, IETF & ISO)**



**NASSCOM, MAIT, TRAI
and other INDUSTRY
BODIES/Associations**



**E-governance, E-
Infrastructure, E-
learning group of DIT**



**ACADEMIA
(IIT & IIIT's etc)**

W3C Indian Members

Current members :

- TCS (Submitted application for membership)
- Department of Information Technology, Government of India ,New Delhi
- Comviva Technologies , Gurgaon
- Akmin Technologies, Chennai
- Setu Software Systems,
- Yantra Software , Hyderabad
- Atal Behari Vajpayee - Indian Institute of Information Technology and Management, Gwalior

W3C E-Government Activity

W3C's e-Government Activity currently offers guidelines focused on helping governments to publish data in open, standards-based formats so that information can be more easily accessed on the Web.

The screenshot shows a Mozilla Firefox browser window displaying the W3C eGovernment website. The browser's address bar shows the URL <http://www.w3.org/egov/>. The website content includes the W3C logo and the text "Technology and Society domain". The main heading is "eGovernment at W3C" with the subtitle "Better Government Through Better Use of the Web." Below this, there is a welcome message and a paragraph about the eGovernment Activity. A "News" section features an article titled "eGov at W3C Technical Plenary Week (TPAC)" dated Wednesday, November 17th 2010. The article text mentions a series of discussions on eGovernment topics in Lyon, France, and refers to "The first day" of the event. On the right side of the page, there are sections for "Groups" (listing "eGovernment Interest Group" and "Government Linked Data (GLD) Working Group") and "Documents" (listing "Improving Access to Government through Better Use of the Web"). A "News Search" section is also present with radio buttons for "All Words", "Some Word", and "Entire phrase", and a "Search" button. The Windows taskbar at the bottom shows the Start button, several open applications (eGovernment at W3C..., elitex-ppt, Microsoft PowerPoint ...), and the system tray with the time 10:36 AM.

W3C[®] Technology and Society domain

eGovernment at W3C

Better Government Through Better Use of the Web.

Welcome! The World Wide Web Consortium invites all who work within governments or those who work or engage with governments to participate in the many opportunities sponsored by W3C and the Web community.

W3C's eGovernment Activity currently offers speakers, discussion forums, educational information, and guidelines focused on helping governments to publish data in open, standards-based formats so that information can be more easily accessed on the Web.

News

eGov at W3C Technical Plenary Week (TPAC)

Wednesday, November 17th 2010

Two weeks ago, we held a series of discussions of eGovernment topics at the W3C [Technical Plenary and Advisory Committee meeting \(TPAC\)](#) in Lyon, France.

[The first day](#) consisted of a broad ranging discussion across eGovernment issues, lent some shape by presentations from W3C staff members. Some of the topics provoking the most discussion were

eGovernment is the use of the Web and other information technologies by governments to interact with the citizenry, between departments and divisions, and with other governments.

Groups

- [eGovernment Interest Group](#)
- Currently being formed: [Government Linked Data \(GLD\) Working Group](#)

Documents

- [Improving Access to Government through Better Use of the Web](#)

News Search

All Words
 Some Word
 Entire phrase

Engaging the Stakeholders through W3C India Website

- **QUESTIONNAIRE**

A set of questionnaire has been developed to evolve the inputs from the industries in the following domains: **Accessibility, Multilingualism, Mobile and Literacy**. The inputs would be crucial for developing the vision and course of action for enabling above domains in Indian Languages

<http://w3cindia.in/que.aspx>

- **EXPRESSION OF INTEREST**

W3C India Invites expression of interest to work in the following areas for detailed study of existing standards Areas of Interest are **Mobile web, Speech , Styling , XML(Extensible Markup language),WAI(Web Accessibility Initiative), XHTML(Extensible Hypertext Markup language) . E-Governance best practices in Indian Languages.**

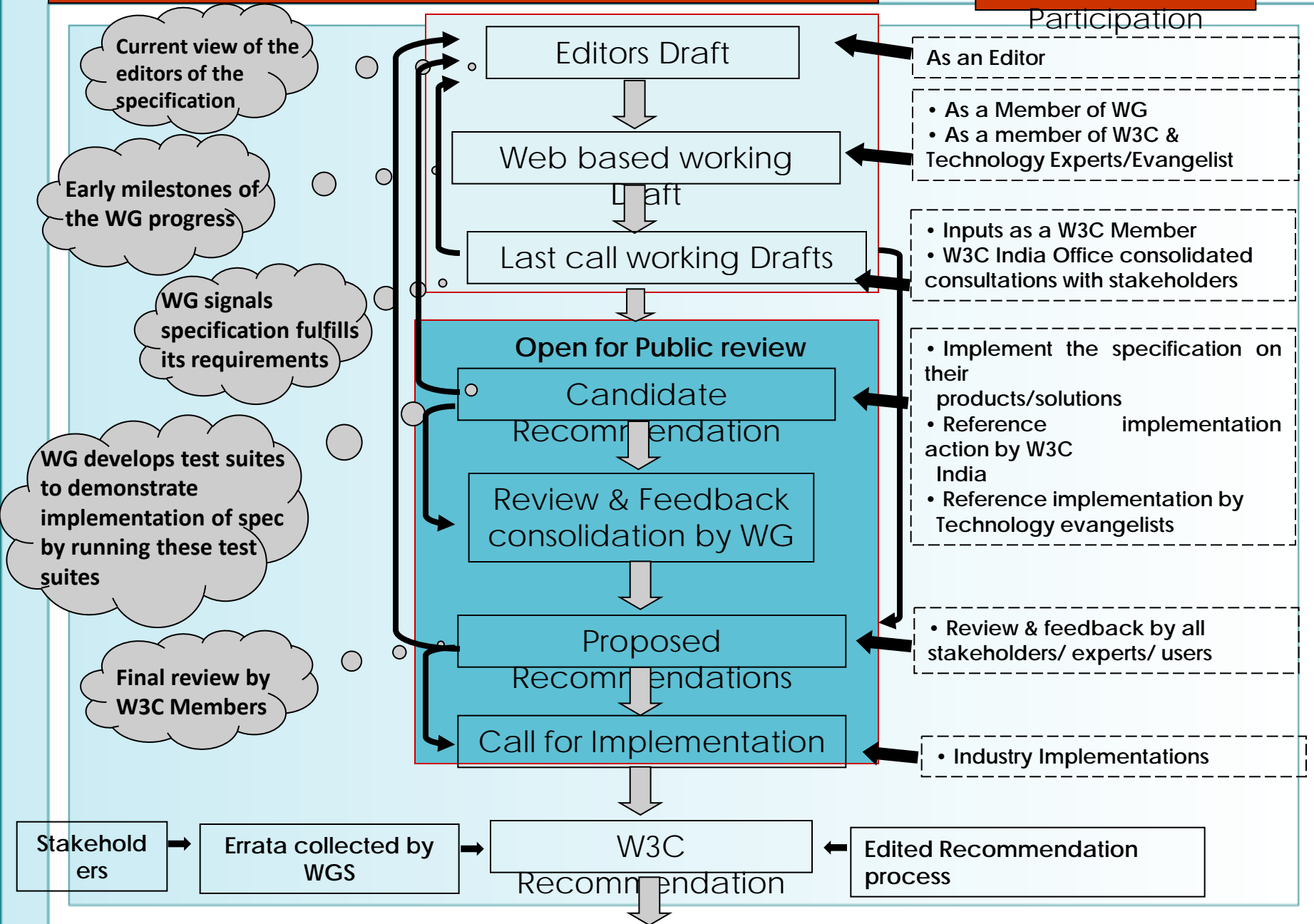
http://w3cindia.in/word_pdf/exp_of_int.doc

Drafts Available on W3C India website

- [Indian language requirements in CSS \(Cascading Style sheets\)](#)
- [WCAG Guidelines in Hindi Language](#)
- [Accessible Guidelines by STQC](#)
- [WCAG 2.0 Guidelines for Indian website by NIC](#)
- [CLDR for Indian languages](#)
- [Language Tags](#)

Other Activities

- W3C Speech Technology and Standardization in India
- Internationalization and localization best practices guidelines for e-governance applications
- Keyboard Standardization
- Xml Internationalization best practices
- Base document for Mobile Ok Checker
- XML Schema for POS



Roadmap

AREA	PROPOSED ACTIVITIES	
EDUCATION AWARENESS AND OUTREACH		
FOCUSED CONFERENCES/WORKSHOPS IN PRIORITY AREAS OF NATIONAL INTEREST	<ol style="list-style-type: none"> 1. CONFERENCE ON MOBILE WEB INITIATIVE IN INDIA 2. FOCUSED WORKSHOP ON W3C SPEECH STANDARD(PLS) 	
RECRUITMENT OF NEW W3C MEMBERS FROM INDIA	<ol style="list-style-type: none"> 1. DIALOG WITH POTENTIAL INDIAN INDUSTRIES LIKE YAHOO R & D , BANGALORE ,MICROSOFT , BANGALORE, BARRIER BREAK TECHNOLOGIES, MUMBAI, BRAAHMAM NET SOLUTIONS P. LTD., NOIDA , WIPRO , BANGALORE, INFOSYS , 2. INVITING MEMBERSHIP FROM PREMIER ACADEMIC INSTITUTIONS LIKE IITS / IIITS, IISC BANGALORE, LNMIIT JAIPUR. 	
PROMOTION AND PROLIFERATION OF W3C STANDARDS		
MOBILE WEB INITIATIVE	<ol style="list-style-type: none"> 1. REVIEW AND ENHANCEMENT OF W3C MOBILE WEB BEST PRACTICES IF REQUIRED 	TARGET LANGUAGE HINDI
SPEECH TECHNOLOGY AND STANDARDIZATION IN INDIA	<ol style="list-style-type: none"> 1. REQUIREMENT, ANALYSIS INITIATED FOR PRONUNCIATION LEXICON SPECIFICATION(PLS 1.0) IN INDIC LANGUAGES 	TARGET LANGUAGE HINDI , BANGLA

AREA	PROPOSED ACTIVITIES	
CSS IMPLEMENTATION FOR INDIC LANGUAGES	<ol style="list-style-type: none"> 1. CREATING DRAFT THAT PRESENTS VARIOUS CSS STYLING ISSUES IN INDIC SCRIPT IN 8 INDIAN LANGUAGES(HINDI, BENGALI, GUJARATI, PUNJABI, MARATHI, TAMIL, TELUGU, MALAYALAM) 2. FINALIZATION OF AKSHARA DEFINITION 3. NEW LISTING OF HINDI CHARACTERS 4. MEETINGS WITH LANGUAGE EXPERTS AND PUBLISHING COMPANIES IN ORDER TO DEFINE THE VARIOUS NEW RULES REQUIRE IN THE IMPLEMENTATION OF CSS FOR INDIAN LANGUAGES. 	
E-GOVERNANCE	<ol style="list-style-type: none"> 1. FINALIZATION OF INTERNATIONALIZATION AND LOCALIZATION BEST PRACTICES GUIDELINES FOR E-GOVERNANCE APPLICATIONS 2. GUIDELINES FOR WEB APPLICATION PLATFORMS 3. GUIDELINES FOR DATABASE 	
REFERENCE IMPLEMENTATION OF CSS AN HTML W3C STANDARDS	1. BUILDING THE REFERENCE IMPLEMENTATION FOR TWO W3C STANDARDS(CSS & HTML)	TARGET LANGUAGE
		HINDI

AREA	PROPOSED ACTIVITIES
BI-DIRECTIONAL COMMUNICATION BETWEEN STAKE HOLDERS AND W3C CONSORTIUM	
SENSITIZING INDUSTRY ASSOCIATIONS FOR ENHANCED PARTICIPATION FROM INDUSTRY IN W3C INDIA ACTIVITY	<ol style="list-style-type: none"> 1. MEETINGS INVOLVING INDUSTRY ASSOCIATIONS LIKE NASSCOM, MAIT, TRAI, COAI, IMAI, CCAOI, ISPAI, TEMA. 2. FOLLOW-UP MEETINGS & PARTICIPATION IN THEIR EVENTS/CONFERENCES.
FEEDBACK TO W3C	
PARTICIPATION IN W3C GLOBAL AND VISION TASK FORCE	<ol style="list-style-type: none"> 1. FEEDBACK FROM INDIA IN THE FOUR THEMATIC DOMAINS NAMELY ACCESSIBILITY, MULTILINGUALISM, MOBILE AND LITERACY AS FOCUSED INITIATIVES TO CHALK OUT ITS MEDIUM TERM AND LONG TERM STRATEGY.
DRAFT RECOMMENDATIONS	<ol style="list-style-type: none"> 1. DEVELOPMENT OF INITIAL DRAFT RECOMMENDATIONS OF CSS & PLS 1.0

Thank You