

*Introduction to W3C India
Internationalisation Programme*

November 2017

W3C India Internationalisation (i18n) Programme

- W3C has launched an aggressive Internationalisation Program designed to identify and obtain the resources to significantly enhance and improve support for non-Latin languages on the Web.
 - *Support for the major Indic Languages is a part of this program*
- *The merger of IDPF with W3C has resulted in a set of activities called **PUBLISHING@W3C***
 - *The Indian Publishing community needs to bring their unique requirements to that conversation.*

W3C and Internationalisation



- W3C's mission is a Web for All, regardless of language, script or culture
- W3C has provided tools to help with I18N for many years.
- However, we need higher bandwidth discussions with leading website developers and language experts.

These conversations can inform us what today's major barriers are so we can conceive of new tools and language enablement.

PUBLISHING@W3C

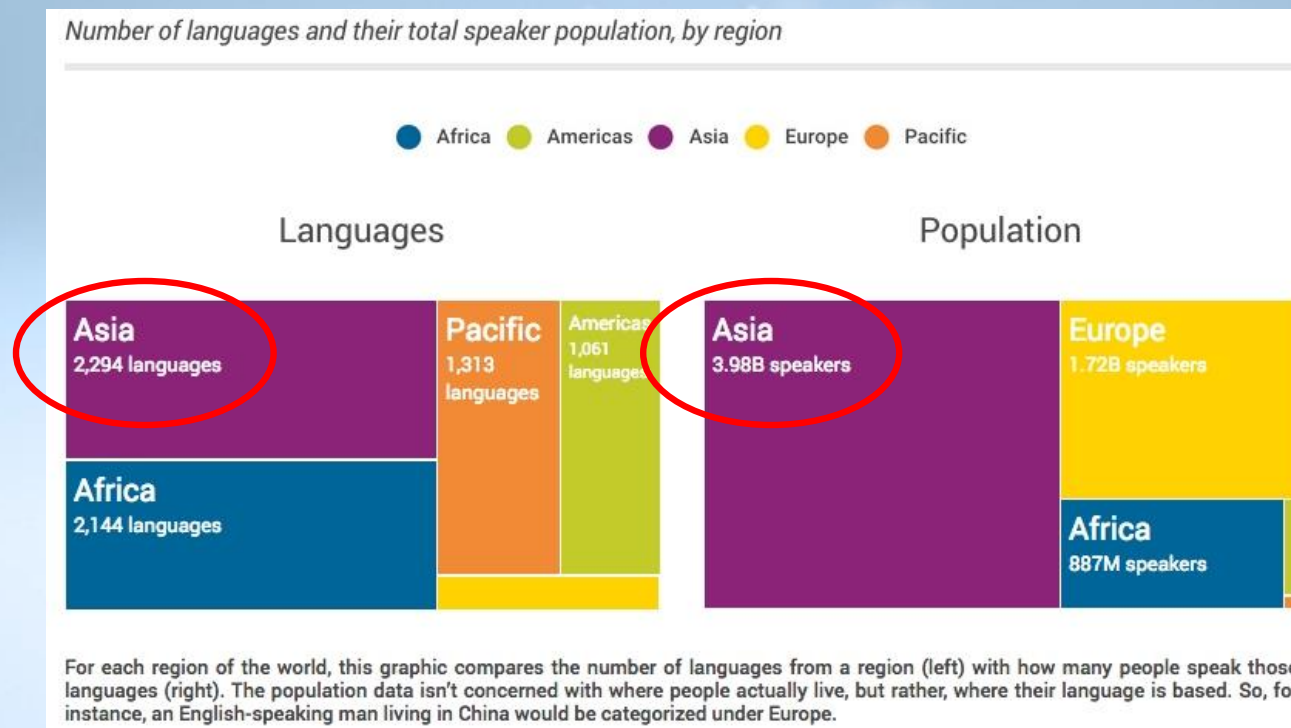
We welcome global participation in Publishing@W3C activities from any organization interested in joining the conversation about the future of publishing.

The Publishing Business Group fosters ongoing participation by members of the publishing industry and overall publishing ecosystem in the development of the Web for publishing, and serves as a conduit for feedback between the publishing ecosystem and W3C.

The mission of the Publishing Working Group is to enable all publications—with all their specificities and traditions—to become first-class entities on the Web.

What is the I18N Current Landscape?

- Ethnologue reports 7,100 languages currently spoken
- Wikipedia reports 100 languages represent 85% of the world population
- 90+ languages are each spoken by 10+ million people



What does that look like in India?

- Ethnologue reports 454 languages currently spoken in India
- Wikipedia reports The 2001 census recorded 29 individual languages as having more than 1 million native speakers (0.1% of total population).

First, Second, and Third languages by number of speakers in India (2001 Census)						
Language	First language speakers ^{91[10]}	First language speakers as a percentage of total population ¹¹¹	Second language speakers ¹¹⁰	Third language speakers ¹¹⁰	Total speakers ^{121[10]}	Total speakers as a percentage of total population ¹¹¹
Hindi ^[b]	422,048,642	41.03	98,207,180	31,160,696	551,416,518	53.60
Bengali	83,369,769	8.10	6,637,222	1,108,088	91,115,079	8.86
Telugu	74,002,856	7.19	9,723,626	1,266,019	84,992,501	8.26
Marathi	71,936,894	6.99	9,546,414	2,701,498	84,184,806	8.18
Tamil	60,793,814	5.91	4,992,253	956,335	66,742,402	6.49
Urdu	51,536,111	5.01	6,535,489	1,007,912	59,079,512	5.74
Kannada	37,924,011	3.69	11,455,287	1,396,428	50,775,726	4.94
Gujarati	46,091,617	4.48	3,476,355	703,989	50,271,961	4.89
Odia	33,017,446	3.21	3,272,151	319,525	36,609,122	3.56
Malayalam	33,066,392	3.21	499,188	195,885	33,761,465	3.28
English	226,449	0.02	86,125,221	38,993,066	125,344,736	12.18
Sanskrit	14,135	<0.01	1,234,931	3,742,223	4,991,289	0.49

What is this Project About?

- W3C intends to marshal the resources of organisations that care about these problems to strengthen internationalisation support on the web



Photo by [Climate KIC](#) on [Unsplash](#)



Shutterstock Image ID: 72752062

- Set 3-10 year stretch goals of progress
Measureable progress on language support, author support, developer support
Goals set independently so that we can move faster in areas where we get greater support

Where are We Starting From?

Definition of How Well the Web Currently Supports a Particular Language

- We have 27 Criteria that we are Measuring
- For the top 72 Languages (initial targets), of which 18 are Indic, we will measure these elements and give the language one of four categorizations:
 - Advanced – Web Support of high-quality rendering
 - Basic – Web Supports commonly required features
 - Minimal – Web readable but missing basic features for the community
 - Broken – fundamental requirement prevents Web Usage

Where do Indic Languages Stand?

Key:

- 3 All needs covered (ie. OK), or not applicable
 - 2 Basic needs covered, but work needed for advance publishing
 - 1 Can create interoperable web pages, but work still needed for basic features
 - 0 Something prevents interoperable or effective use of the language in webpages
- * next to a language name indicates a tentative score, pending validation by experts

Results:

Number of languages listed: **72**

- **20** may be Ok, based on what information we have for them
- **32** need work for advanced publishing
- **18** need work for basic features
- **2** don't work on the Web
- **?** **52%** of cells still need investigation.

Results:

Number of languages listed: **18**

- **8** may be Ok, based on what information we have for them
- **1** need work for advanced publishing
- **9** need work for basic features
- **0** don't work on the Web
- **?** **49%** of cells still need investigation.

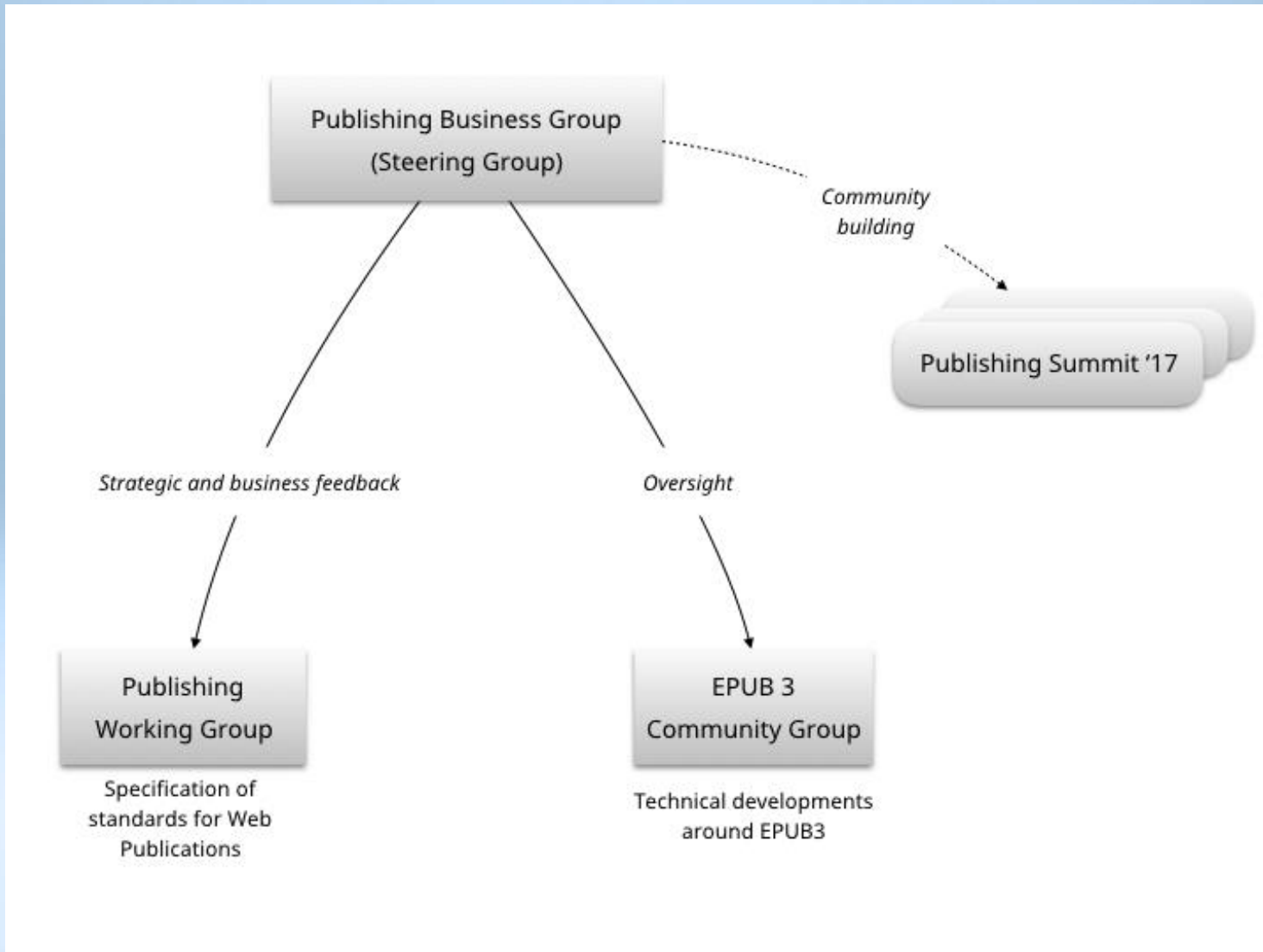
Where do Indic Languages Stand?

script		Level	Encoding	Fonts	Font styles	Glyph control	Cursive text	Char transforms	Numbers & digits	Boundaries & selection	Quotation	Inter-char spacing	Ruby	Text decoration	Emphasis & highlight	Bidi text	Other inline	Line breaking	Hyphenation	Justification & alignment	Lists, counters	Initial letter style	Baselines, inline align	Other paragraph	Bidi layout	Vertical text	Notes, footnotes	Page no.s, running heads	Other pagination
language	language																												
Arab	Urdu*	0.15	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Beng	Bengali*	0.46	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
	Chittagonian*	0.46	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Deva	Hindi*	0.42	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
	Marathi*	0.42	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
	Bhojpuri*	0.42	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
	Maithili*	0.42	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
	Awadhi*	0.42	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
	Nepali*	0.42	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Guju	Gujarati*	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Guru	Punjabi*	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Knda	Kannada*	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Latn	English*	0.86	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Mlym	Malayalam*	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Orya	Odia*	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Sinh	Sinhalese*	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
TamI	Tamil*	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Telu	Telugu*	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?

Total number of languages in the table: 18

PUBLISHING@W3C

11



PUBLISHING@W3C

It is the goal of the Publishing Working Group **to provide**, in concert with other W3C Groups, **the necessary technologies on the Open Web Platform** to make **the combination of traditional publishing and the Web** complete in terms of the readers' needs, portability, distribution, archiving, offline access, or reliable cross referencing.

PUBLISHING@W3C

In short, **all publications**—with all their specificities and traditions—**should become first-class entities on the Web.**

PUBLISHING@W3C

Areas for Potential New Work

Accessible Advertising

Advanced features for scholarly/academic/STM

Archival

Color management

Fonts

High quality printing from Web/EPUB directly

Full Support of Indic Languages – let's add this!

What Does Success Look like?

- Completed Gap Analysis on top 10 Languages to Verify They meet Advanced Criteria
- Move 6 Indic languages from Basic to Advanced in 3 years
- Have 5 organisations join the **PUBLISHING@W3C** community as active W3C Members in CY 2018
- These objectives require significant W3C Staff increase and **plus much increased participation from the Indian Language community**

Our Team for This Program

- Alolita Sharma – Overall Program Chair
- Abhijit Dutta – Co-Chair for In-Country Activities
- C-DAC Pune – W3C India Office led by Mahesh Kulkarni, Country Manager
- Richard Ishida – W3C Internationalisation Leader
- Bill McCoy – W3C Publishing Champion
- J. Alan Bird – W3C Global Business Development Ldr

I18N Program Sponsorship Levels



- **Platinum** sponsorship is a commitment of \$100 K per year.
- **Gold** sponsorship is a commitment of \$50 K per year.
- **Silver** sponsorship is a commitment of \$25K per year.
- **Bronze** sponsorship is a commitment of \$10K per year.
- Each Level has unique benefits from participation on the Leadership Board and earmarking of funds to website branding

PUBLISHING@W3C

18

EPUB and the overall **Open Web Platform** are foundations for next-generation publications that are rich, accessible, and cross-device

EPUB 3.1 is well-aligned with overall Web Accessibility initiatives and will continue to evolve and grow in adoption under W3C stewardship

Future **Web Publications** functionality will further converge online/offline, packaged/distributed paradigms

We want to know more about the needs of everyone in the publishing community – academic and scholarly, educational, libraries, trade, magazines, journals, graphic designers and what's missing from the Web

Encourage conversations through many avenues: industry associations like BISG, W3C Publishing Summit, 9-10 November 2017, San Francisco

Will You Support Us in this Effort?



The users of the Web need your help!

W3C



**THANK YOU
FOR YOUR TIME!**