







Advancing the Multilingual Web, Thematic Network







[1st Year Review]

07 June 2011

































I. Overview



Overall objectives for the first period

1

• Run two workshops where participants can share their experiences with standards, guidelines, best practices and initiatives related to authoring content for the Web, and discuss areas needing attention. Produce a set of proceedings and a workshop report for each.

2

• Produce an initial version of an internationalization checker and training curriculum, and get input on these from project partners. Put in place a mechanism for harvesting test results.

3

• Put in place a project website that provides information about the workshops, future and past, and build around it a network of social media and other channels for dissemination of information.

4

• Set up and run the administrative framework for the management of the project.

• • · Contents

- I. Project Overview
- II. Madrid workshop
- III. Pisa workshop
- IV. Practical items
- V. Communication & mobilisation



II. Madrid workshop



1st Workshop: Madrid

"The Multilingual Web – Where are We?"



Universidad Politécnica de Madrid

26-26 October, 2010

Date on Gantt chart fell in August, due to date of project start. Moved to October to ensure better attendance and support EC plans.





Madrid Workshop: Key outcomes

- 101 attendees (25% higher than target of 80). All partners represented.
- 34 presentation authors (12 from consortium) in 29 speaking slots.
- Good mix of speakers from industry, research, standards and industry bodies, browser and application developers, etc.
- Sessions covering a wide range of topic areas: Developers, Creators, Localizers, Machines, Users.
- Evening reception for additional networking.
- Video recordings of the speakers, and live minuting via IRC (enabled, amongst other things, participation by a deaf person in the workshop).
- Overall rating on evaluation form: 58% of respondents gave 5/5.
 Remaining 42% scored 4/5.



Madrid Workshop: Speaker mix

- Welcome: UPM, University of Limerick/LRC
- Developers: W3C, Mozilla, Opera, Microsoft, Unicode Consortium
- **Creators**: BBC World Service, Loquendo, Universidad Politécnica de Madrid, Linguaserve, World Wide Web Foundation
- Localisers: SAP, CNGL, Lucy Software, Institute Jozef Stefan
- **Machines**: DFKI, CNR-ILC, DEADELUS, bioloom group, VU University of Amsterdam, SAP
- **Users**: Facebook, Google, Sidar Foundation, Dept of Information Technology Govt. of India



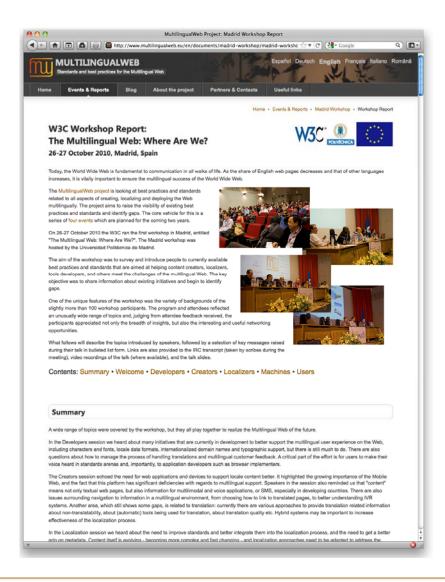
Madrid Workshop: Delegate comments

- "Well organised. Good variety of presentations/topics & participants."
- "Excellent mix pitched at accessible level"
- "Very good networking opportunities"
- "Excellent organization. High level and wide variety of the presentations. Very interesting."
- "Great & stimulating, of lot of work to be done"
- "Good balance overall between technical & process & data. ¡good work!
 Very helpful and interesting. Thank you"



Madrid Workshop: Outputs

- Key output is the Workshop Report.
 Summarises all talks, with bulleted key points, plus an overall summary of the workshop.
 Links to all videos, slides and IRC transcripts.
- Program with links to videos, slides and IRC transcripts, released in the week after the workshop.
- Pointers to social media commentaries:
 - 6 blog posts
 - · Large number of tweets using #mlw hashtag.
 - Photos on Flickr.
- All publicly available on the web at http://www.multilingualweb.eu/documents/ma drid-workshop





Madrid Workshop: Sample messages

Users need to be talking to browser developers to get i18n standards implemented.

Standards development should be speeded up, and use a more 'agile' approach

The Mobile Web is and will become increasingly important, but the platform has significant deficiencies with regard to the multilingual Web.

"Content" means not only textual web pages, but also information for multilimodal and voice applications, or SMS, especially in developing countries

There is a need for standardization of metadata related to the localization process and recognition of the importance of the Semantic Web technologies to support that

A new way of working is needed, founded on cooperation and sharing of standardized resources, rather than hiding data away.

Currently there are various approaches to providing translation related information about non-translatability, about (automatic) tools being used for translation, about translation quality etc

Hybrid systems may be important to increase effectiveness of the localization process

A language neutral representation may provide a way of dealing more effectively with multilingual searches

There are issues surrounding navigation to information in a multilingual environment, from choosing how to link to translated pages, to better understanding IVR systems

There is an appetite for multilingual content around the world, but there are significant organizational and technical challenges in the way of reaching people in continents such as Africa and Asia

Speech technologies enable multilinguality to be addressed in a wide variety of sectors and applications

Fonts for mobile devices need OpenType rendering support for complex scripts, not just font glyphs.



Madrid Workshop: Relevant success indicators

Objective	Expected	Actual
Workshop attendees	80	101
No. of hits on announcements of workshop on W3C home page	2000 (full year)	 As of 2 June 2011: "W3C Workshop, Call for Participation: The Multilingual Web - Where Are We?" 22 jul 2010, 75,801 views "MultilingualWeb Workshop, Madrid. Speaker proposals due 17th September" 13 sep 2010, 65,036 views "MultilingualWeb Workshop, Madrid. Initial program published." 1 oct 2010, 45,415 views "MultilingualWeb Madrid slides, video recordings, and IRC notes now available" 23 nov 2010, 46,902 views "MultilingualWeb Madrid report now available" 15 dec 2010, 43,970 views
Views of workshop report on project website	200 (full year)	As of 21 April 2011: 96,740



Q&A

II. Madrid workshop



III. Pisa workshop



2nd Workshop: Pisa

"Content on the Multilingual Web"

Consiglio Nazionale delle Ricerche

4-5 April, 2011







Pisa Workshop: Project changes

Original plans in DoW

- Hold workshop in February (due to offset from project start).
- Make it a small workshop (around 40 people).
- Focus on discussion.
- Use the discussions to generate a set of prioritised recommendations.
- Focus on content development only.

Revised plans

- Date moved to April to allow for better attendance and fit preparations around Xmas break.
- Aim for around 80 people to attend.
- Reuse the successful format used in Madrid, ie. more speaker time than discussion time.
- Produce a set of proceedings as for Madrid with a summary workshop report.
- Use the same wide ranging sessions as in Madrid, to continue networking across disciplines.



Pisa Workshop: Key outcomes

- 95 attendees (237% higher than target of 40). All partners represented.
- 37 presentation authors (10 from consortium) in 34 speaking slots.
- Another good mix of speakers from industry, research, standards and industry bodies, browser and application developers, etc. Speakers from several localization vendors (eg. SDL, Lionbridge, GTS).
- Sessions covering same wide range of topic areas as Madrid, plus a new Policy session.
- Evening reception for additional networking highly rated.
- In addition to video recordings of the speakers, and live minuting via IRC, there was live video-cast and external participation via IRC.
- Very positive feedback from attendees (verbally and in social media).
- Implemented sponsorship program: 2 sponsors. FLaReNet endorsed.



Pisa Workshop: Speaker mix

- **Welcome**: Institute for Informatics and Telematics (IIT), Italian National Research Council (CNR), W3C Italy Office, European Commission Joint Research Centre (JRC)
- Developers: CWI/W3C, Opera Software, W3C, VZ Netwerke, Thompson Reuters
- **Creators:** Microsoft European Devt. Centre, University of Economics Prague, INAF, European Commission/Opera Software, SDL
- Localisers: SAP, DFKI, Enlaso, LRC, CNGL, University of Limerick, Andrä AG, XTM International, Opera Software, PangeaMT, GTS Translation
- **Machines**: CNGL, Trinity College Dublin, Software AG, Tilde, Josef Stefan Institute, CRS4 Sardinia, University of Sassari
- Users: Lionbridge, University of Amsterdam, Trusted Translations Inc, Facebook, SDL
- Policy: TAUS, FAO of UN, ILSP "Athena" RC



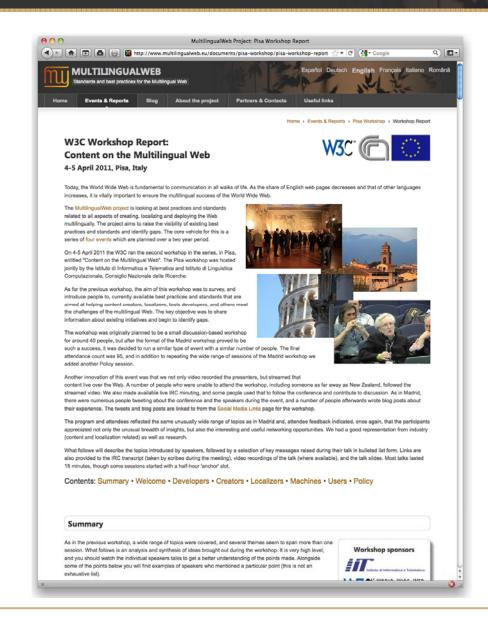
Pisa Workshop: Delegate comments

- "Excellent #mlwpisa conference. Broad range of topics and mix of academic and commercial participants/presentations."
- "Nice end to a good workshop thanks to all at #mlwpisa. We really enjoyed the conference"
- "Again thank you very much for the invitation to participate at the workshop, it was very stimulating and very useful to make contacts to further improve multilinguality on the web"
- "congratulations to a great meeting! It was obvious how much work had gone in the preparations; it will be a hard act to follow."
- "I want to thank you for all your efforts and, of course, for the result -- a very successful Workshop"



Pisa Workshop: Outputs

- Same outputs as for Madrid:
- Workshop Report. Summarizes all talks, with bulleted key points, plus an overall summary of the workshop. Links to all videos, slides and IRC transcripts.
- Program with links to videos, slides and IRC transcripts, released in the week after the workshop.
- Pointers to social media commentaries:
 - 8 blog posts
 - Approx. 100 tweets using #mlwpisa hashtag.
 - · Photos on Flickr.
- All publicly available on the web at http://www.multilingualweb.eu/documents/pis a-workshop





Pisa Workshop: Sample messages

The community needs to participate now in development of standards such as HTML5, CSS, XForms, IDN, etc.

Standards development should be speeded up, and use a more 'agile' approach

The Social Web created an explosion of content that is nowadays directly relevant to corporate and organizational strategies. This is leading to a change, where immediacy trumps quality in many situations

Inconsistencies need to be addressed in user interfaces for users wanting to switch between localized variants of sites. This needs investigation and standardisation. In some cases this is down to differences in browser support

We need better training of authoring tool implementers and authors in internationalisation, to make it easier for content authors to produce well internationalised content

Changes in the way content is generated and technology developments are also expected to shift emphasis further onto the long tail of translation

ITS customizations need to work in authoring tools

Tools need to deal with comparable texts (ie. texts that are not completely parallel, since one page has slightly different content than another)

The industry is at the beginning of radical change, such as it hasn't seen for 25 years (van der Meer), and interoperability and standards will be key to moving forward into the new era

More work is needed on the elaboration and use of metadata to support localization processes, building on the foundation provided by ITS, but also using Semantic Web technologies such as RDF

There is a shift to greater adaptation of content and personalisation of content for local users. The ability to componentise is a key enabler for this, as is some means of helping the user find relevant content, such as geolocation or content negotiation

In order to cope with the increased emphasis on automated approaches to handling data and producing multilingual solutions, there is a need for increased interoperability between initiatives via standard approaches



Pisa Workshop: Relevant success indicators

Objective	Expected	Actual		
Workshop attendees	40	95		
No. of hits on announcements of workshop on W3C home page	2000 (full year)	 As of 2 June, 2011: "MultilingualWeb workshop, Pisa, Call for Participation" 13 feb 2011, 38,588 views "MultilingualWeb Pisa, Program published!" 11 mar 2011, 23,715 views "Don't forget to register for the W3C MultilingualWeb Workshop in Pisa" 23 mar 2011, 20,736 views "MultilingualWeb Pisa slides, video recordings, and IRC notes now available" 4 may 2011, 7,699 views "MultilingualWeb Pisa report now available" 26 may 2011, 2,362 views (Bit.ly click throughs for tweet announcing Pisa CFP, 13 Jan, 12,608 clicks) 		
Views of workshop report on project website	200 (full year)	7,041 For 26 May to 6 June, 2011		



Q&A

III. Pisa workshop



IV. Practical items



Practical items

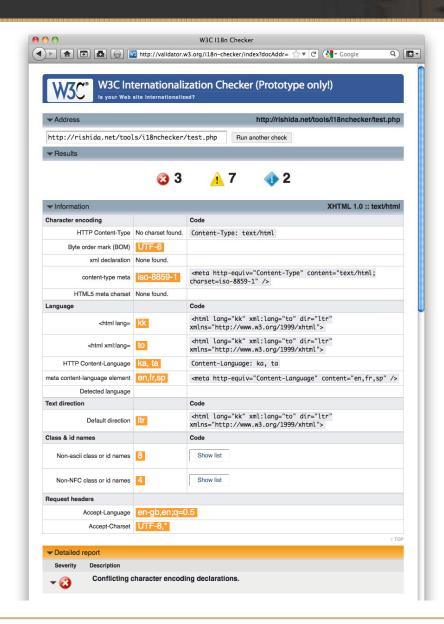
- Internationalization Checker
- Standards curriculum
- Test results

NOTE: Not funded as part of the project! However, opportunities for review and comment by partners are included in the plan.



Practical items: I18n Checker

- Provides information about i18n features of a page, and provides guidance to authors via error, warning and info messages, with links to further reading.
- First version produced prior to project kick-off meeting in Bucharest.
- Reviewed during kick-off meeting.
- Moved to http://validator.w3.org and various fixes and improvements made.
- Major overhaul under way to integrate to other W3C validation tools, improve robustness, and extend capabilities.

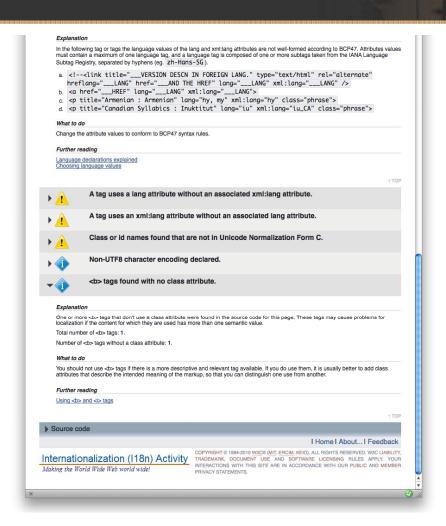




Practical items: I18n Checker

Improvements currently in progress:

- Allow file upload.
- Increase range of formats supported to include HTML5 and XHTML 1.1.
- Localize the checker target, 40 languages.
- Say something about linguistic quality, eg. spellchecking 'global English'
- Provide results in XML format for incorporation into other applications.
- Integrate with W3C Unicorn validation tool reports include HTML, CSS, MobileOK, etc.
- Auto-detection of actual language of page, to compare with declarations.
- Improvements to robustness and user interface.





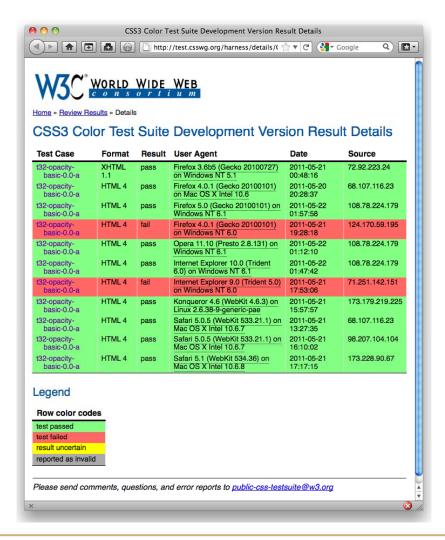
Practical items: Curriculum & testing

Educational curriculum design

- Discussed during kick-off meeting.
- Partners decided to create a sub-group to look at the curriculum.
- Not much movement since the initial wiki setup.
 To be discussed in this meeting.

Test results harvesting

- Early discussion at kick-off meeting, but plan was to implement a method for submitting test results by this meeting and start work after.
- Plans delayed, due to major shift in test suite formats planned by W3C later this year. To be discussed later in this meeting.





Practical items: Relevant success indicators

Objective	Expected	Actual
I18n checker in use	1 st version produced for kick-off FTF, and available for use by public.	Done. http://www.w3.org/validator.w3.org/i18n-checker/



Q&A

IV. Practical items



V. Communication & mobilisation



C&M: Project web site and logo

- Web site address: http://multilingualweb.eu/
- Available since the beginning of the project. Based on Joomla, to test i18n limitations.
- Translated into 5 languages besides English.
- Project logo represents conjoined M and W of MultilingualWeb, but in forms inspired by Syriac letter SHEI.
- Logo has been used for an Opera browser extension icon.







C&M: Internal communications

- Email: multilingualweb-partners@w3.org multilingualweb-pc@w3.org mlw-partners-archive@w3.org
- Wiki
- Teleconferences
 - W3C conferencing bridge
 - IRC with zakim and rrsagent bots for managing meetings and producing minutes

- 1 MultilingualWeb Where are we?
 - 1.1 Ride share/hotel information for partners
- 2 Program Committee
 - 2.1 Next program committee meeting
- 3 Approach
- 4 Dissemination
- 5 Contact pipeline
 - 5.1 Can't come
 - 5.2 Possible contacts
 - 5.3 Awaiting response
 - 5.4 May come
 - 5.5 Expecting to come, but not yet registered
 - 5.6 Registered
- 6 Speaker proposals
- 7 Session planning
 - 8 Registrants 7.1 Program
 - 7.2 Speaker info
 - 7.3 Facilities 7.4 Not fitted in
- - 8.1 Officially nominated partner representatives
 - 8.2 Partner speakers
 - 8.3 Other speakers
 - 8.4 Others
 - 8.5 Unable to make it
- 9 Logistics tracker
- 10 Template for inviting people to the workshop
 - 10.1 People who attended the previous workshop
 - 10.2 People who haven't yet attended
 - 10.3 Another alternative
- 11 Brainstorming ideas
 - 11.1 Workshop title
 - 11.2 From Christian



C&M: External communications





C&M: External communications

multilingualweb.eu

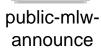


www.org/International













@w3c 25,160 followers

multilingualweb

249 members



@multilingweb





494 visits week of 30 Jun



C&M: Relevant success indicators

Objective	Expected	Actual	
Hit counts for project website	15,000	As of 21 april 2011: 1,323,584	
Subscribers to the public mailing list set up for discussion of standards & best practices	50	<pre>public-multilingualweb@w3.org 53 subscribers public-mlw-announce@w3.org 89 subscribers</pre>	
Number of tweets sent out on W3C twitter channel	24	Not possible to ascertain due to difficulty of viewing past tweets on @w3c channel. 62 tweets were sent on the @multilingweb twitter channel.	



C&M: Consortium partner participation

- Madrid facilities donated by UPM, and several local organizers recruited to handle workshop logistics.
- Free recording and hosting of Madrid video by UPM.
- Pisa facilities donated by CNR, and several local organizers recruited to handle workshop logistics.
- Free video recording from CNR. Free hosting of video by VideoLectures (associated with JSI).
- Program committee of around 10 active participants per workshop, but additional contributions from other partners.
- Partners assisted energetically in contacting potential speakers, disseminating information about workshops, and contributing talks themselves.
- Facebook implemented FB page and donated 200 euros for advertising the 1st workshop. Lionbridge set up LinkedIn page.



C&M: Consortium partner participation

- Facebook managed translations of multilingualweb.eu, and they and other partners contributed the actual translations themselves.
- ICIA hosted the kick-off meeting in Bucharest and took care of local logistics for the meeting.
- CNR-IIT sponsored the Pisa workshop through the sponsorship program, and CNR partners arranged for FLaReNet to endorse the Pisa workshop.
- UAS Potsdam established a Collaboration Agreement between the MLW project and META-NET.
- At least one representative and often more participated in each Workshop.
- All but one partner was represented at the kick-off meeting.
- Partners forward announcements about the workshops to fora and lists,eg. upcoming.yahoo.com, geekagenda, BITKOM Knowledge Management WG, xliff mailing list, META-NET mailing list, TEI-L, DocBook TC, EUATC, EST websit, etc. Announcements are also added to the Multilingual Computing magazine calendar, and broadcast by the LISA and LRC partners to their mailing lists.



C&M: Relevant success indicators

Objective	Expected	Actual
External citations/mentions of MLW project in papers, conference, talks, etc.	10 (full year)	 As of 2 June, 2011: 16 Presentation at 1st International XLIFF Symposium, Ireland Keynote at Business Information Systems 2010, Germany Presentation at Webtech Conference, Germany Presentation at "Markupforum 2010", Germany Talk at Berlin Libraries Colloquium, Germany Keynote at KIM-DINI Kickoff Workshop, Germany Article in Handbook for Technical Communication Presentation at Grand Opening of W3C Office, Germany Talk at tcWorld Conference 2010, Germany Talk at W3C Day of W3C German-Austrian Office, Germany Article in technische kommunikation, Germany Presentation at Localization World, Germany Talk at Technology, Standards & Communication Conference, India Seminar, Aspekte eines multilingualen Web, Germany Workshops, Widgets and more: Cross-platform application development using web standards, Mexico, Khazakhstan, Russia Poster at META-NET meeting, Belgium



Q&A

VI. Communication & mobilisation

