Why Localisation Standardisation Activities Matter to You

MultilingualWeb – Language Technology
A New W3C Working Group

Pedro L. Díez Orzas
A few words about who is talking:

Pedro L. Díez-Orzas

CEO at Linguaserve I.S. S.A
Professor at Univ. Complutense de Madrid
PhD in Computational Linguistics
Member of MultilingualWeb-LT
W3C Member
CTN 191 – AENOR – Terminology

pedro.diez@linguaserve.com
Standards

They are great.

Everyone should have their own.
New Standards for Old Needs

How standards proliferate:
(see: a/c chargers, character encodings, instant messaging, etc)

Situation: There are 14 competing standards.

14?! Ridiculous! We need to develop one universal standard that covers everyone’s use cases. Yeah!

Soon:

Situation: There are 15 competing standards.

http://xkcd.com/927/
New Standards for New Needs

Viability of Multilinguality on the Web depends on the level of mechanisation of methods and processes.

A certain level of standard metadata can decisively help to make this happen.
But it cannot always take that long...

The Space Shuttle and the Horse's Rear End

http://www.astrodigital.org/space/stshorse.html
MultilingualWeb-LT

• New W3C Working Group under I18n Activity
  – [http://www.w3.org/International/multilingualweb/lt/](http://www.w3.org/International/multilingualweb/lt/)

• **Aims:** define meta-data for web content that facilitates its interaction with language technologies and localization processes.

• Already have 28 participants from 20 organisations
  – Chairs: Felix Sasaki, David Filip, Dave Lewis

• **Timeline:**
  – Feature Freeze Nov 2012
  – Recommendation complete Dec 2013
MLW-LT Approach

• Standardise Data Categories
  – ITS (1.0) has: Translate, Loc note, Terminology, Directionality, Ruby, Language Info, Element Within Text
  – MLW-LT could add: MT-specific instructions, workflow, quality-related provenance, legal?

• Map to formats
  – ITS focussed on XML
    • useful for XHTML, DITA, DocBook
  – MLW-LT also targets HTML5 and CMS-based ‘deep web’
  – Use of microdata and RDFa

• Uses Cases
MLS-LT Main Tasks

• Develop this metadata standard through broad consensus across communities, involving content producers, localisation workers, language technology experts, browser vendors and users.

• Establish a network between stakeholders in these communities to ensure the long-term adoption of the standard and future opportunities in integrating LT and emerging Web technologies.
MLW-LT Success Criteria

• The MultilingualWeb-LT Working Group is expected to demonstrate interoperable implementations for different Use Cases.

• ‘Interoperable’ here means that at least the metadata must be available in various parts of Web-related technologies like CMS systems, localisation chains, etc., or in profiles of related formats like XLIFF.
MLW-LT Scope of Use Cases

Language Technology

Create Content

Translate Content

Consume Content

Language Resources
Source Content Processing

Language Technology
- Named entity recognition

Create
- Author
  - Identify no-translate
  - Identify terms
  - Localisation Preparation

Translate

Language Resources
- Glossary
- Term-base

<..> = Possible MLW-LT Metadata
Localisation Quality Assurance

Language Technology

Create

Language Resources

Translate

Localisation Preparation

Term-base

Postediting

Translation Memory

Translation Review

Translation Memory+

Publish to CMS

<XLIFF>

<..> = MLW-LT Metadata

Consume Content

GALA 2012 MONACO

THE LANGUAGE OF BUSINESS
THE BUSINESS OF LANGUAGE
Monaco
26-28 March 2012
Leverage Target Quality Meta-data

Translate

Publish to CMS

Language Resources

Language Technology

Machine Translation

Reading/Reusing

Search Indexing

Term-base

Translation Memory+

Publish to CMS

MT Training

Consume Content

<..> = MLW-LT Metadata
MLW-LT Implementations (Linguaserve)

CMS-L10N integration

Real Time Translation System
MLW-LT will try to work with...

- Beginning with the network of stakeholders built via the W3C MultilingualWeb and META-NET projects.
- ITS Interest Group
- HTML Working Group
- Internationalization Core Working Group
- RDF Web Applications Working Group
- OASIS XML Localisation Interchange File Format (XLIFF) TC
- Unicode Localization Interoperability (ULI) TC

- ETSI Localisation Industry Standards (LIS) ISG
- ISO TC 37
- OASIS Open Architecture for XML Authoring and Localization Reference Model (OAXAL) TC
- OASIS Darwin Information Typing Architecture (DITA) TC
- OASIS Content Management Interoperability Services (CMIS) TC

- It is essential the presence of GALA and the sector that represents in this initiative.
Candidate Stakeholders

- Final users
- Web owners
- Content Consumer
- CMS Developer
- Content Authors
- CMS-based
  - Localisation Management
  - Translator/Posteditor/Reviewer
- LSP-based (CAT/TMS users)
  - Translator/Posteditor/Reviewer
  - Translation/Review Process Manager
- MT Service Provider
- Text Analytics Service Provider
- Localisation Tool developer
- Linguistic Technology developers
- Systems Integrator
- IT Companies
- Search engine crawler
- Communication companies
- Public bodies
- Minority Languages
- Language specialists
Where next?

- Mapping the future of the MLW
  MLW-MultiModal Interaction ....
  MLW-Audio-Visual Content ....
  MLW-JavaScript ....

- Active participation in the near future relationship between MultilingualWeb and Linked Open Data.

Open Linked Data, defined as: «Linked Data is about using the Web to connect related data that wasn’t previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF."»
GALA in MLW-LT

- Contribute to MLW-LT requirements gathering
- [http://www.w3.org/International/multilingual-web/lt/wiki/Requirements](http://www.w3.org/International/multilingual-web/lt/wiki/Requirements)
- Get involved in MLW-LT Working Group
  - Participate as W3C members
  - Feedback via public list and WG site
  - Participation in Workshops
  - Implementations
www.w3.org/International/multilingualweb/lt/

www.multilingualweb.eu