

TC 37 Meeting, Madrid 24-29 June 2012

MultilingualWeb – Language Technology

A New W3C Working Group



Who We Are



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

Arle Lommel

Senior Consultant at DFKI

Standards Coordinator at GALA

PhD in Ethnology (Indiana Univ.)

Member of MultilingualWeb-LT

W3C Member

Participant in TC 37



arle.lommel@dfki.de



Funded by:

Administered by:



Standards ...

... are great.



Funded by:



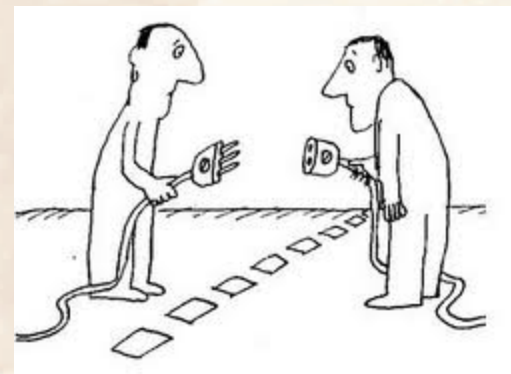
Administered by:



Standards ...

... are great.

Everyone should
have their own.



New Standards for New Needs



Multilinguality on the Web depends on automation of methods and processes.

Standard metadata can help make this happen.



Funded by:



Administered by:



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 in I18n Activity
 - <http://www.w3.org/International/multilingualweb/lt/>
- **Aims:** define meta-data for web content that facilitates its interaction with language technologies and localization processes.



Funded by:



Administered by:



EU Project FP7-ICT

MultilingualWeb-LT



- 28 participants from 20 organisations
 - Chairs: *Felix Sasaki, David Filip, Dave Lewis*
- Timeline:
 - Feature Freeze Nov 2012
 - **Recommendation complete Dec 2013**



Funded by:



Administered by:



EU Project FP7-ICT

Standardizing Data Categories



- ITS 1.0 has:
 - Translate, Loc note, Terminology, Directionality, Ruby, Language Info, Element Within Text
- ITS 2.0 may add:
 - MT-specific instructions, workflow/process, quality, provenance, domain, legal, etc.



Funded by:



Administered by:



Adding Intelligence to the Web



- ITS 1.0 focused on XML
 - useful for XHTML, DITA, DocBook
- MLW-LT/ITS 2.0 adds HTML5 and CMS-based ‘deep web’
- Support for microdata and RDFa
- Development of Uses Cases/Guidelines



Funded by:



Administered by:



MLS-LT Main Tasks



- Develop metadata standards through broad consensus across communities, involving
 - content producers
 - localisation workers,
 - language technology experts
 - Web browser vendors and users.
- Establish a network between stakeholders in these communities to ensure
 - long-term adoption of the standard
 - Future integration of LT and emerging Web technologies.



Funded by:



Administered by:



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.



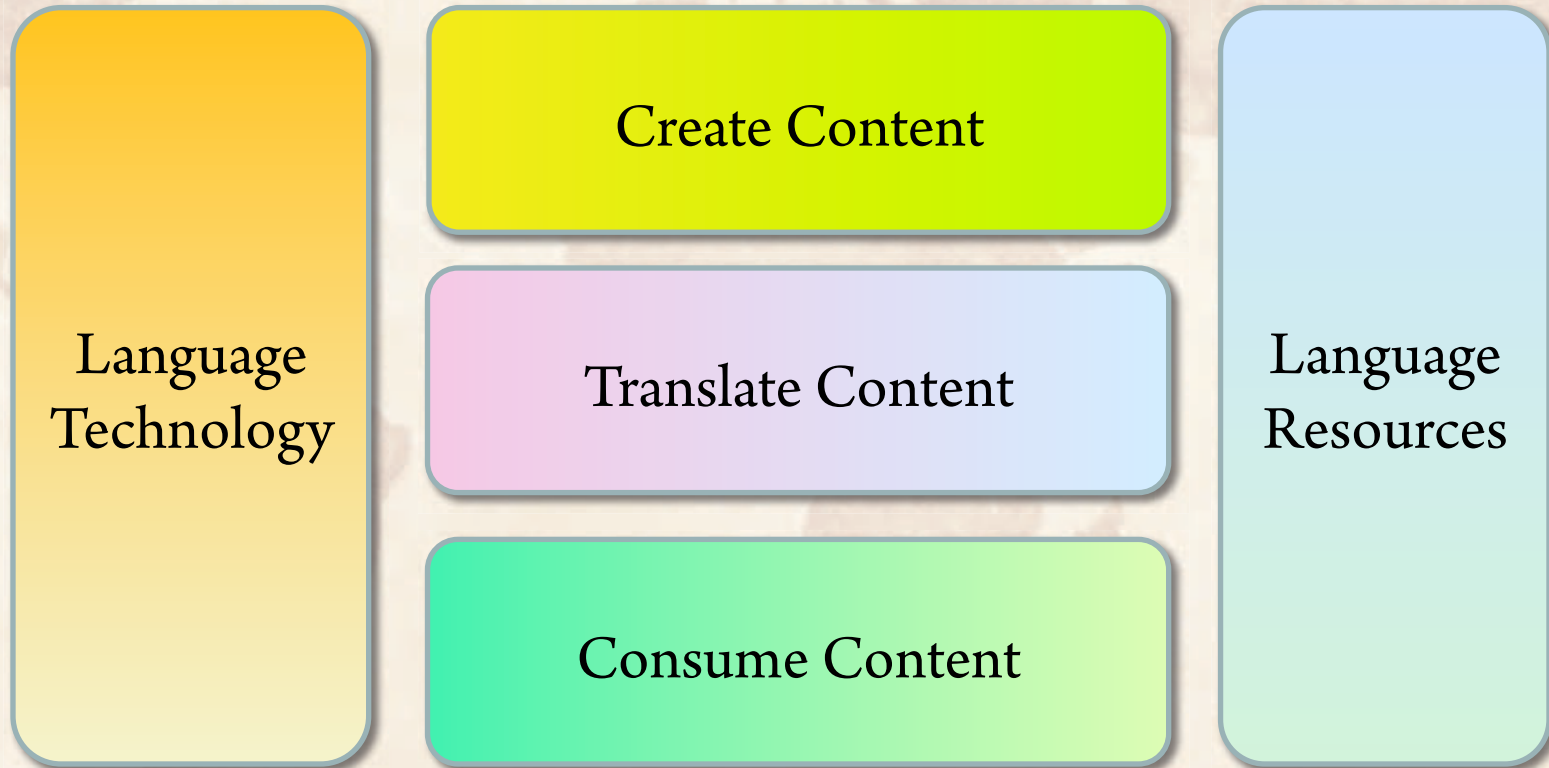
Funded by:



Administered by:



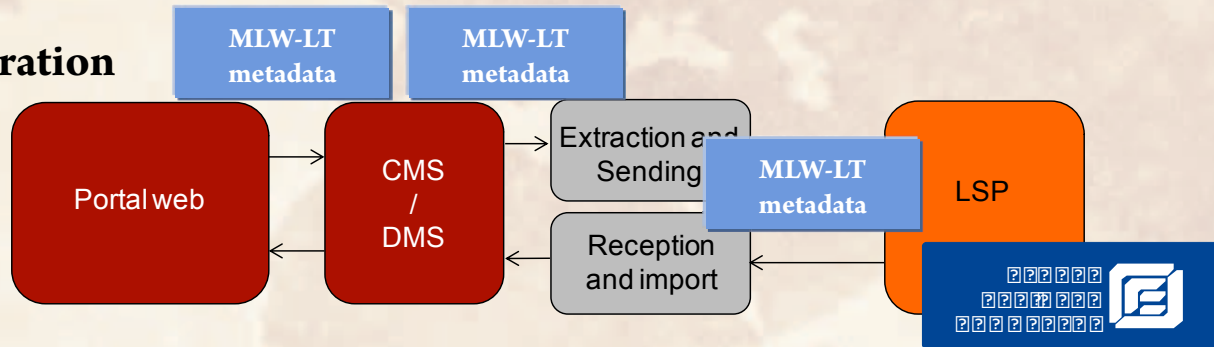
MLW-LT Scope of Use Cases



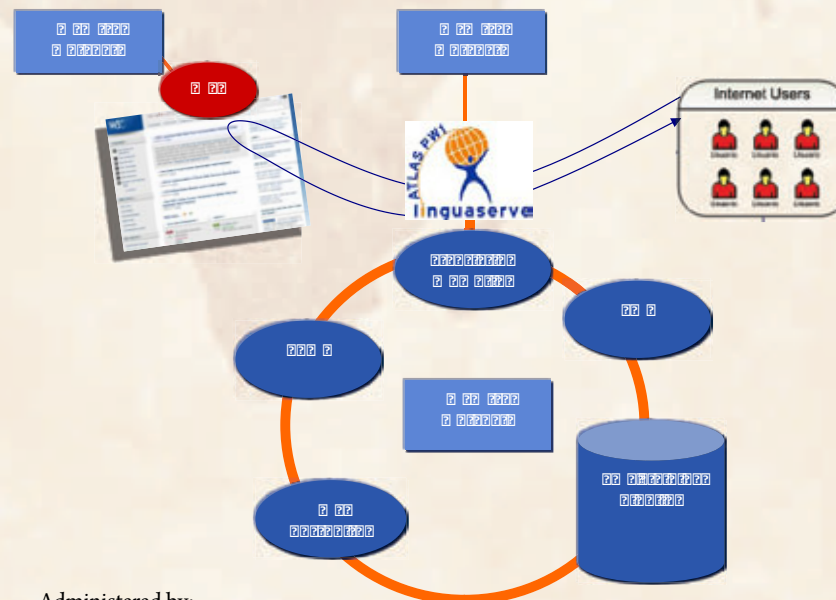
Implementations (e.g. Linguaserve)



CMS-L10N integration



Real Time Translation System



Funded by:



Administered by:



MLW-LT will work with...



- Stakeholders from W3C MultilingualWeb and META-NET.
- 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
- *The presence of GALA and its sector is vital to ensure applicability*



Funded by:



Administered by:



Candidate Stakeholders

- End users/content consumers
- Web owners
- IT Companies
- Content authors
- CMS-based
 - Localisation Management
 - Translator/Posteditor/ Reviewer
- LSP-based (CAT/TMS users)
 - Translator/Posteditor/ Reviewer
 - Translation/Review Process Manager
- Providers/developers of:
 - MT services
 - Text analytics services
 - Localisation tools
 - Linguistic technologies
 - Search systems
 - Content management systems
- Systems integrators
- Communication companies
- Public bodies (gov., NGO)
- Minority Languages
- Language specialists

ISO/TC37 in MLW-LT



- Please visit and review:

<http://www.w3.org/TR/its20>

- Get involved in MLW-LT Working Group
 - Participate as W3C members
 - Feedback via public list and WG site
 - Participation in Workshops
 - Implementations



Funded by:



Administered by:



[www.w3.org/International/
multilingualweb/lt/](http://www.w3.org/International/multilingualweb/lt/)

www.multilingualweb.eu



Enjoy the reception!