



# MultilingualWeb – Language Technology

A New W3C Working Group

Felix Sasaki, David Filip, David Lewis

# MultilingualWeb-LT

- New W3C Working Group under I18n Activity
  - <http://www.w3.org/International/multilingualweb/l10n/>
- **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

# 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, 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

# Candidate Stakeholders

- Content Author
- 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
- CMS Developer
- Localisation Tool  
developer
- Systems Integrator
- Search engine crawler
- Content Consumer

# Scope of Use Cases

Language  
Technology

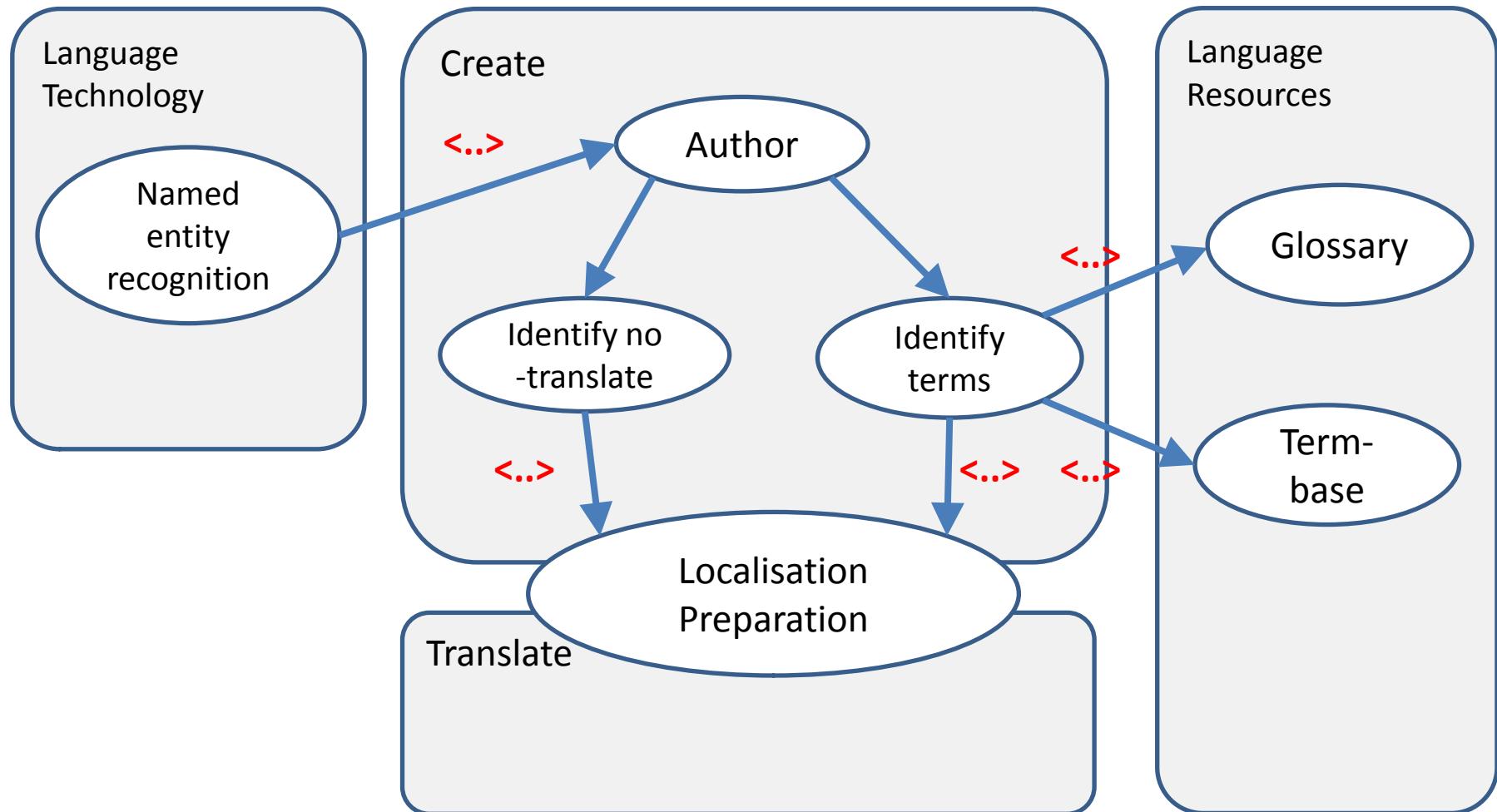
Create Content

Translate Content

Consume Content

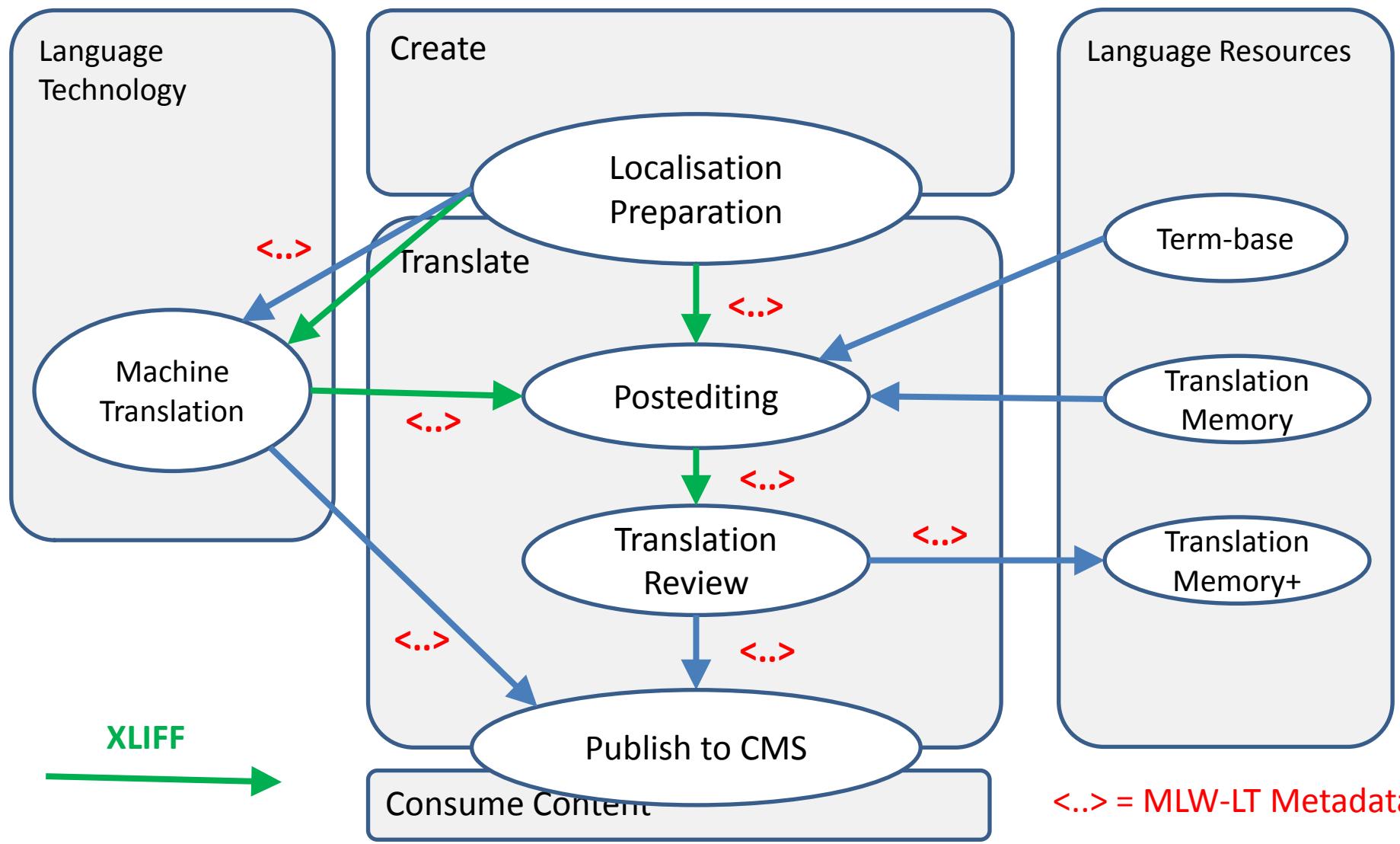
Language  
Resources

# Source Content Processing



**<..>** = Possible MLW-LT Metadata

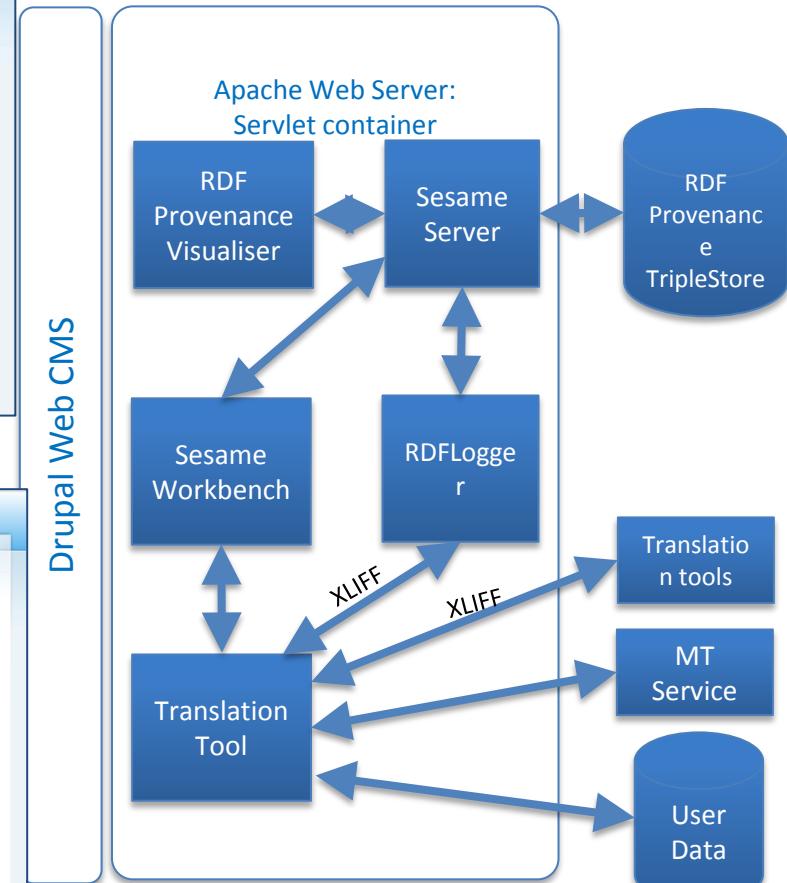
# Localisation Quality Assurance



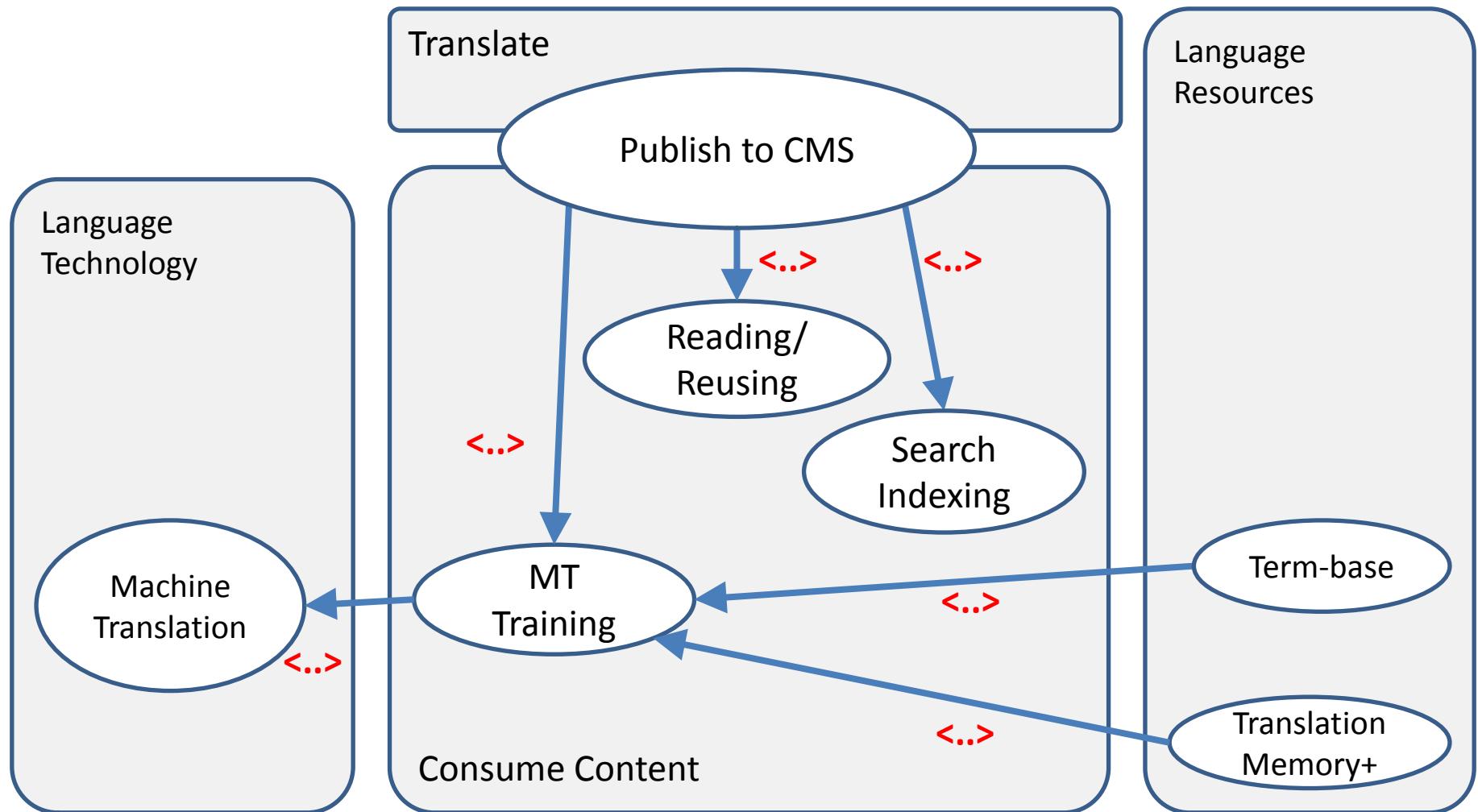
# CMS-L10N integration via RDF & XLIFF

The figure displays three screenshots of the CNGL Demo application, illustrating its localization features:

- Screenshot 1:** Shows a "Job Page" with a sidebar navigation menu. The menu includes sections like "Welcome", "Job Page", "Add Content", "Article", "Blog entry", "Forum topic", "Poll", "Translation Job", "Blogs", "Forums", "Recent Content", "Community content", and "Translator Content". A "Who's online" section indicates there is currently 1 user online.
- Screenshot 2:** Shows a "Job Page" with a "Status: Work in progress" message. It features a progress bar for "English" (0%) and a "Select Options" button. Below this is a "JOB STATISTICS" section with four language options: English, Irish, French, and Italian, each with a "Select Options" button.
- Screenshot 3:** Shows a "Job Page" with a sidebar navigation menu. The menu includes sections like "Welcome", "Job Page", "Add Content", "Article", "Blog entry", "Forum topic", "Poll", "Translation Job", "Blogs", "Forums", "Recent Content", "Community content", and "Translator Content". A "Who's online" section indicates there is currently 1 user online. The main content area shows a "TRANSLATION SENTENCES" section with a list of sentences to be translated, and a "JOBS PAGE -> SELECT LANGUAGE -> Contribution Page" section where text can be translated from English to Italian.



# Leverage Target Quality Meta-data



<..> = MLW-LT Metadata

# Rich Meta-data for TM Leverage

The screenshot shows the Swordfish application interface. The main window has two main panes: 'Source Text: en' on the left and 'Translations: es' on the right. In the Source Text pane, there is a text area containing the following content:

text/html; charset=UTF-8  
2010 1st XLIFF International Symposium -  
Limerick, Test

Below this, there are several lines of text that are partially visible or redacted.

In the Translations pane, there is a large empty text area.

To the right of the main panes, there is a vertical panel titled 'TM Matches' which displays a list of matches with a confidence score of 94.0%:

- 2010 1st XLIFF International Symposium - Limerick, Ireland
- 2010 Primer Simposio Internacional sobre XLIFF - Limerick, Irlanda

Below the TM Matches panel is a 'Match Properties' dialog box. It contains the following key-value pairs:

company-name	Irc
contact-email	Lucia.Morado@ul.ie
contact-name	Lucia
date	2010-05-01T12:00:00Z
job-id	1
original	Symposium2.htm
source-language	en
state	translated
target-language	es
tool-company	SDL Trados
tool-id	Trados

At the bottom of the interface, there is a toolbar with various icons and a status bar displaying 'Words: 7 Characters: 56' and 'Context: 8'. A dropdown menu shows the value 'needs-review-translation'.

# Next Steps

- Contribute to MLW-LT requirements gathering
  - Breakout session Friday
  - Feedback on Requirements
    - New ones? Priorities?
    - <http://www.w3.org/International/multilingualweb/l1t/wiki/Requirements>
- Get involved in WG
  - Participate as W3C members
  - Feedback via public list and WG site
  - Requirements Workshop in Dublin in 11-12 June
  - Implementations
- Where next ?– mapping the future of the MLW
  - MLW-MultiModal Interaction ....
  - MLW-Audio-Visual Content ....
  - MLW-JavaScript ....