



# The Multilingual Web

<http://www.multilingualweb.eu/>

David Filip, David Lewis, Felix Sasaki

dave.lewis@cs.tcd.ie

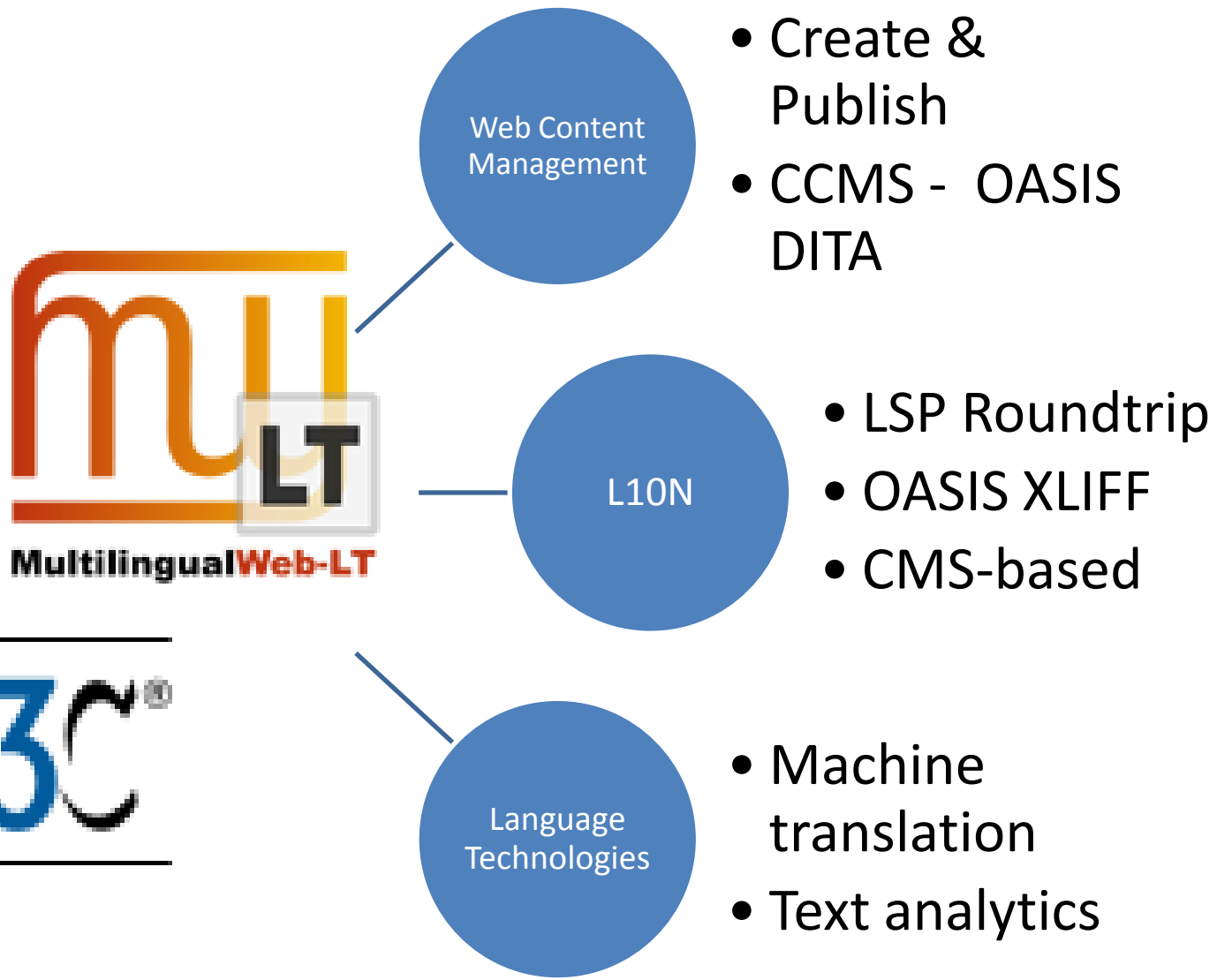
# Why Multilingual Web?

- Commercial translation market \$21-26bn annually
- Enables order of magnitude more in cross border trade
- 51% European retailers sell via Web, but only 21% cross border
- 30% Europeans buy online, only 7% cross border
- Only a third of Web users now claim English as first language
- Biggest growth in Chinese, Russian and Arabic
- Interoperability problems estimated to impose 20% overhead on language services industry

# Multilingual Web Initiative

- W3C is active in ensuring the multilingual viability of the World Wide Web
  - Unicode, IRI, language tags
- Internationalisation Activity
  - Best Practice, Internationalisation checker
  - Internationalisation Tag Set (ITS)
- Multilingual Web Initiative
  - EC-funding for W3C workshop series
  - Assemble Stakeholders who create, localise, publish and use multilingual web content, applications and technologies

# Multilingual Web – LT (Language Technology)



---

**W3C**<sup>®</sup>

---

# MultilingualWeb-LT Approach

- Standardise Data Categories
  - Document header rules, element attributes, precedence rules
  - ITS (1.0) has: Translate, Loc note, Terminology, Directionality, Ruby, Language Info, Element Within Text
  - MLW-LT could add: MT-specific instructions, localisation 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

# Scope of Use Cases

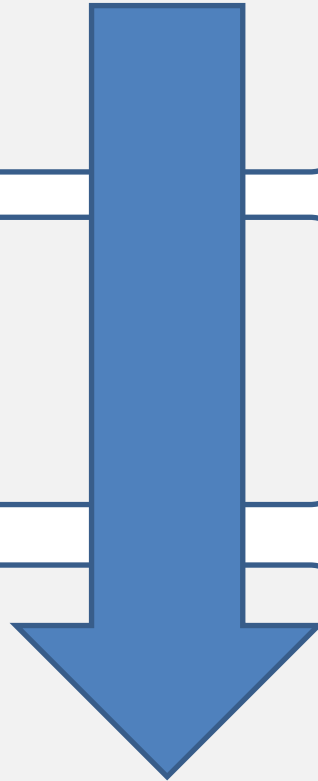
Language  
Technology

Create Content

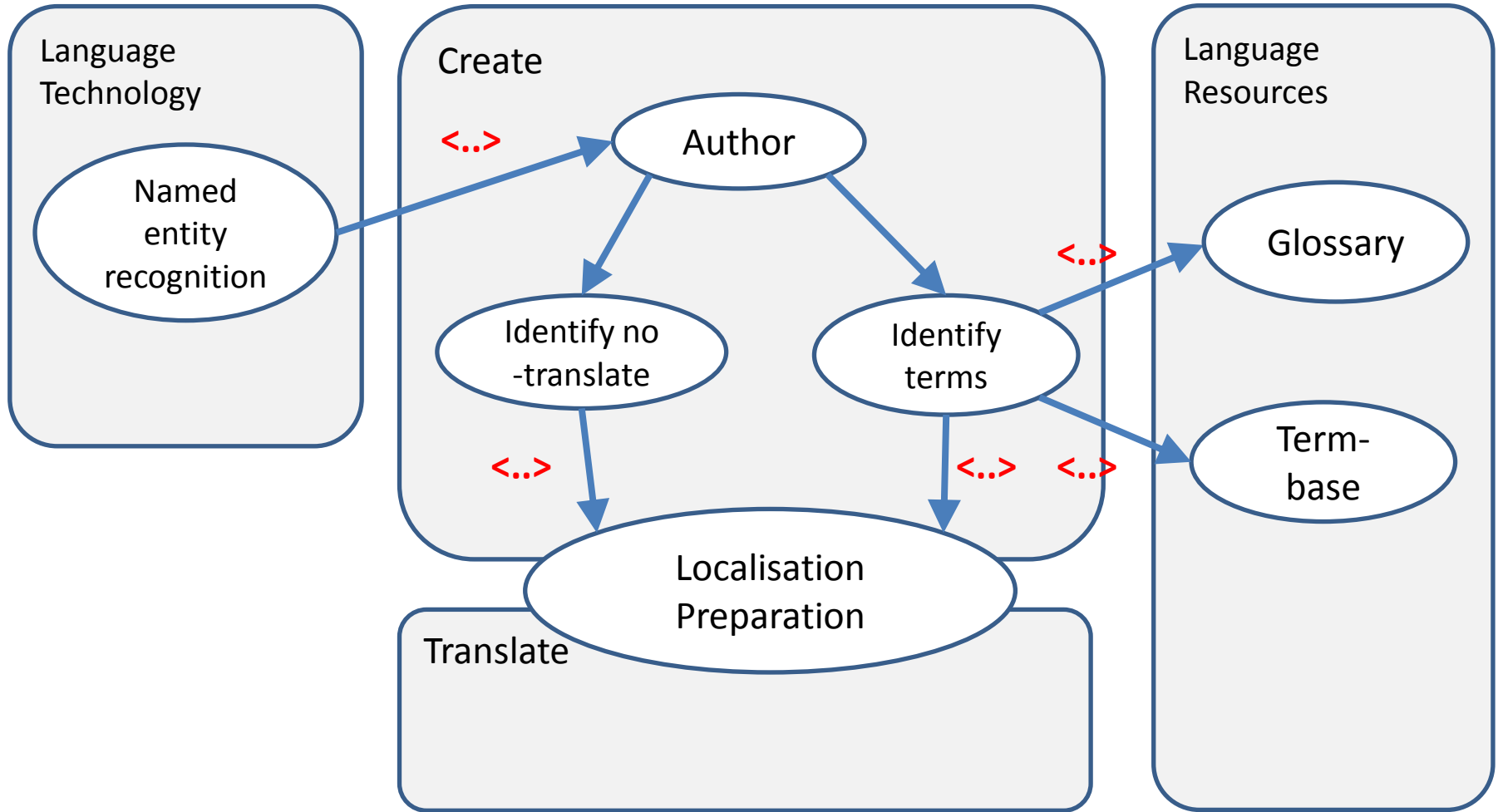
Translate Content

Consume Content

Language  
Resources

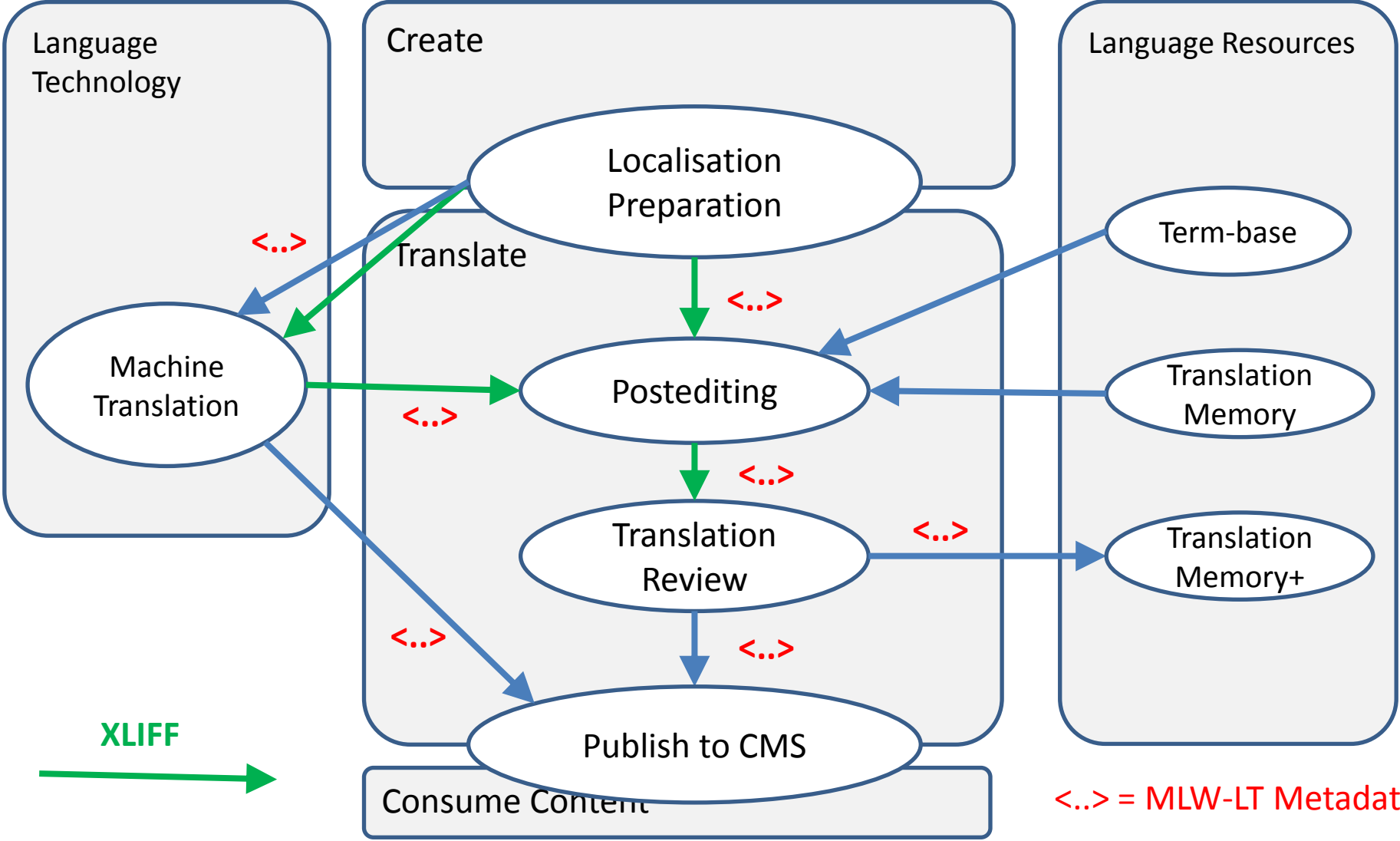


# Source Content Processing



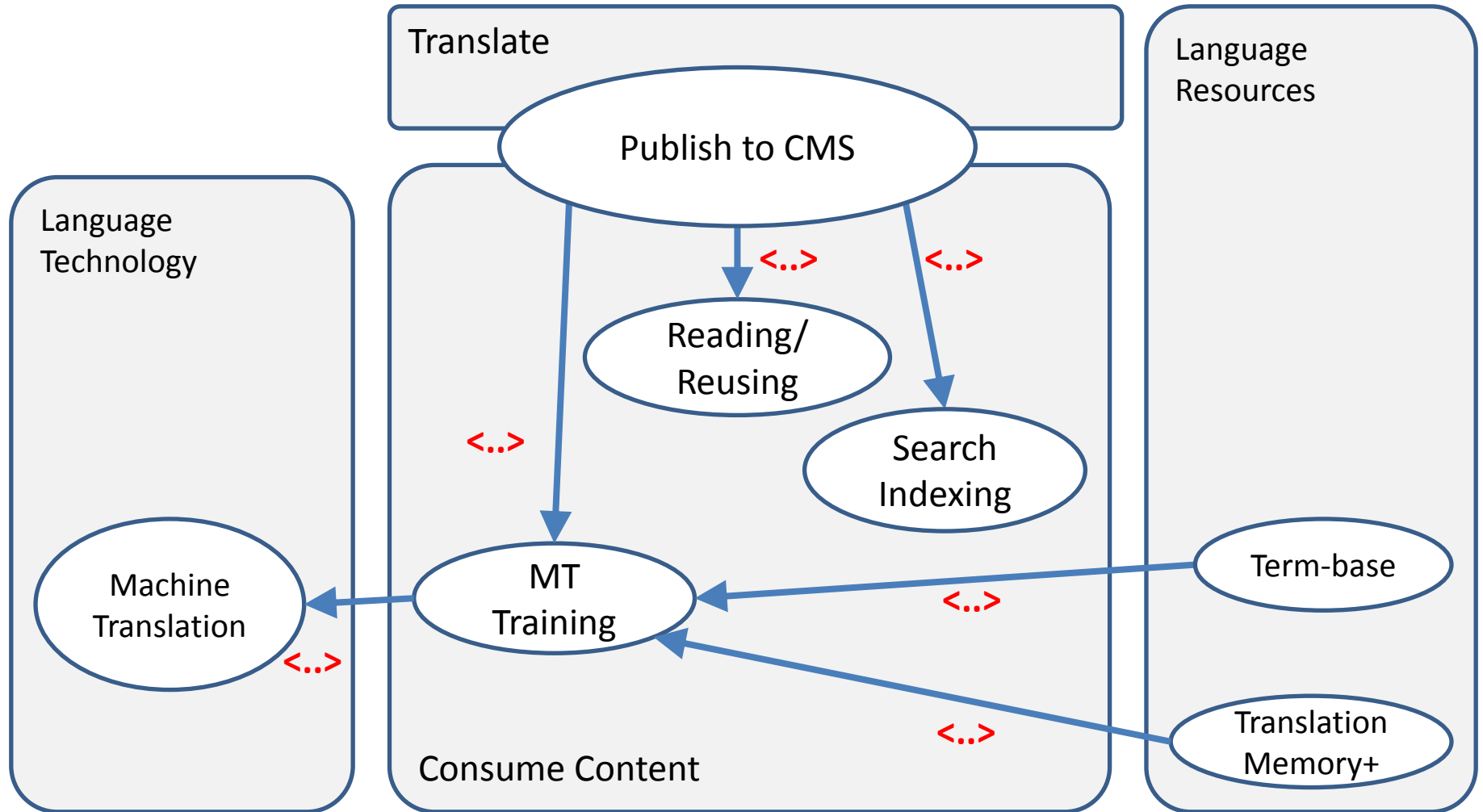
<..> = Possible MLW-LT Metadata

# Localisation Quality Assurance





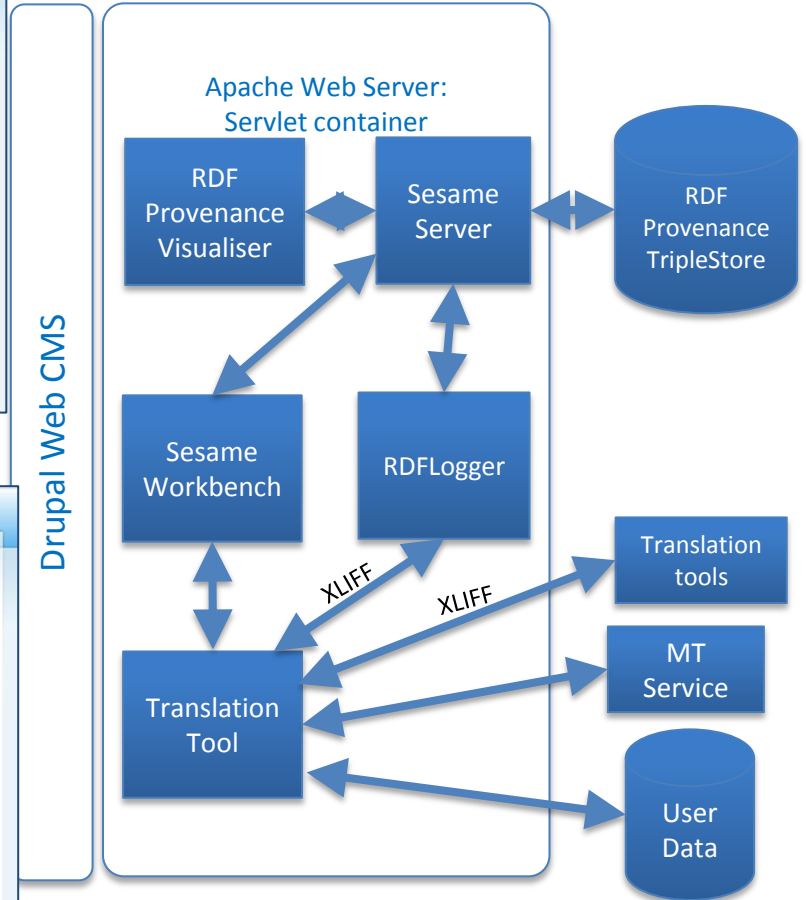
# Leverage Target Quality Meta-data



<..> = MLW-LT Metadata

# CMS-L10N integration via RDF & XLIFF

The image displays three overlapping screenshots of the CNGL Demo CMS interface. The top screenshot shows a 'Job Page' with a 'JOBS PAGE - SELECT LANGUAGE - Visualizer Page' section, displaying a list of jobs with columns for language, progress, and status. The middle screenshot shows a 'Job Page' with a 'Status: Work in progress' section, displaying progress bars for English, French, and Italian. The bottom screenshot shows a 'Job Page' with a 'TRANSLATION SENTENCES' section, displaying a list of sentences to be translated, and a 'TRANSLATE TO: ITALIAN' section with a text input field and a 'Translate' button.



# Its Your Multilingual Web

- Get involved in MultilingualWeb-LT WG
  - Participate as W3C members
  - Follow via public list and WG site
    - <http://www.w3.org/International/multilingualweb/lt/>
  - Requirements Workshop in Dublin in 12-13 June
  - Implementations
- Where next ?– mapping the future of the MLW
  - MLW-Linked Open Data ..... Dublin 11 June
  - MLW-MultiModal Interaction ....
  - MLW-Audio-Visual Content ....
  - MLW-JavaScript ....

<http://www.multilingualweb.eu/>