



# 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/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

# 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

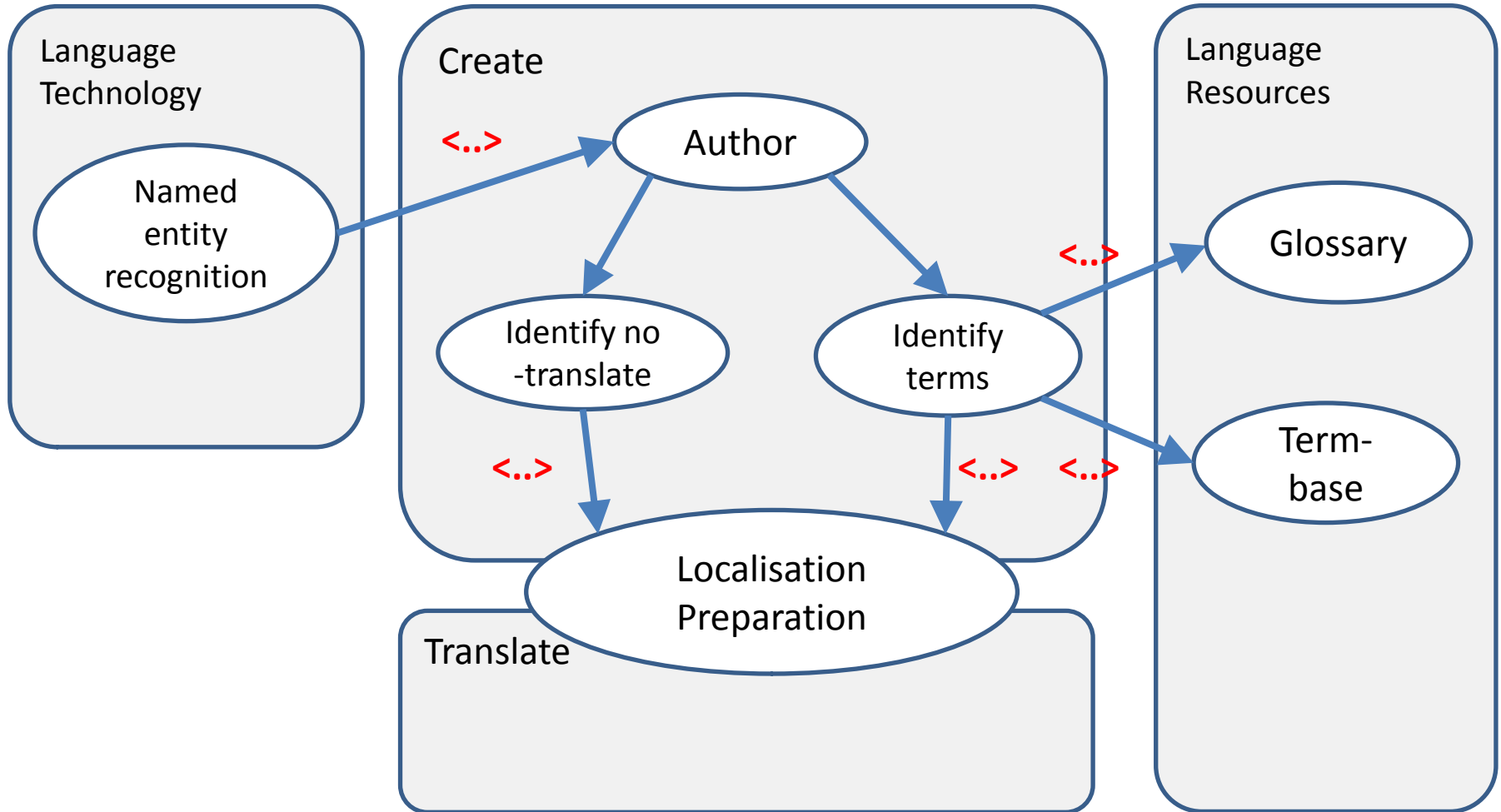
Create Content

Translate Content

Consume Content

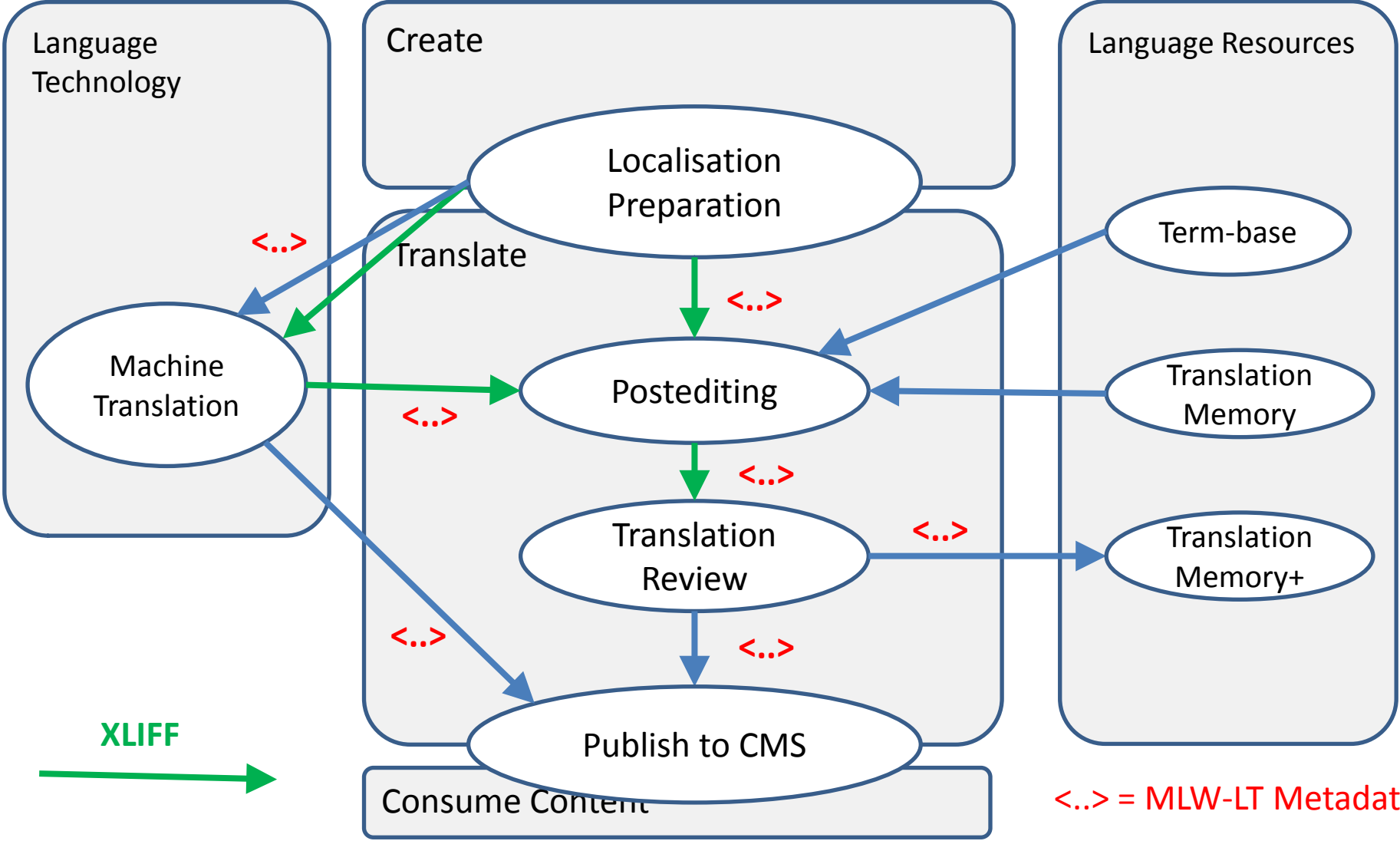
Language  
Resources

# Source Content Processing

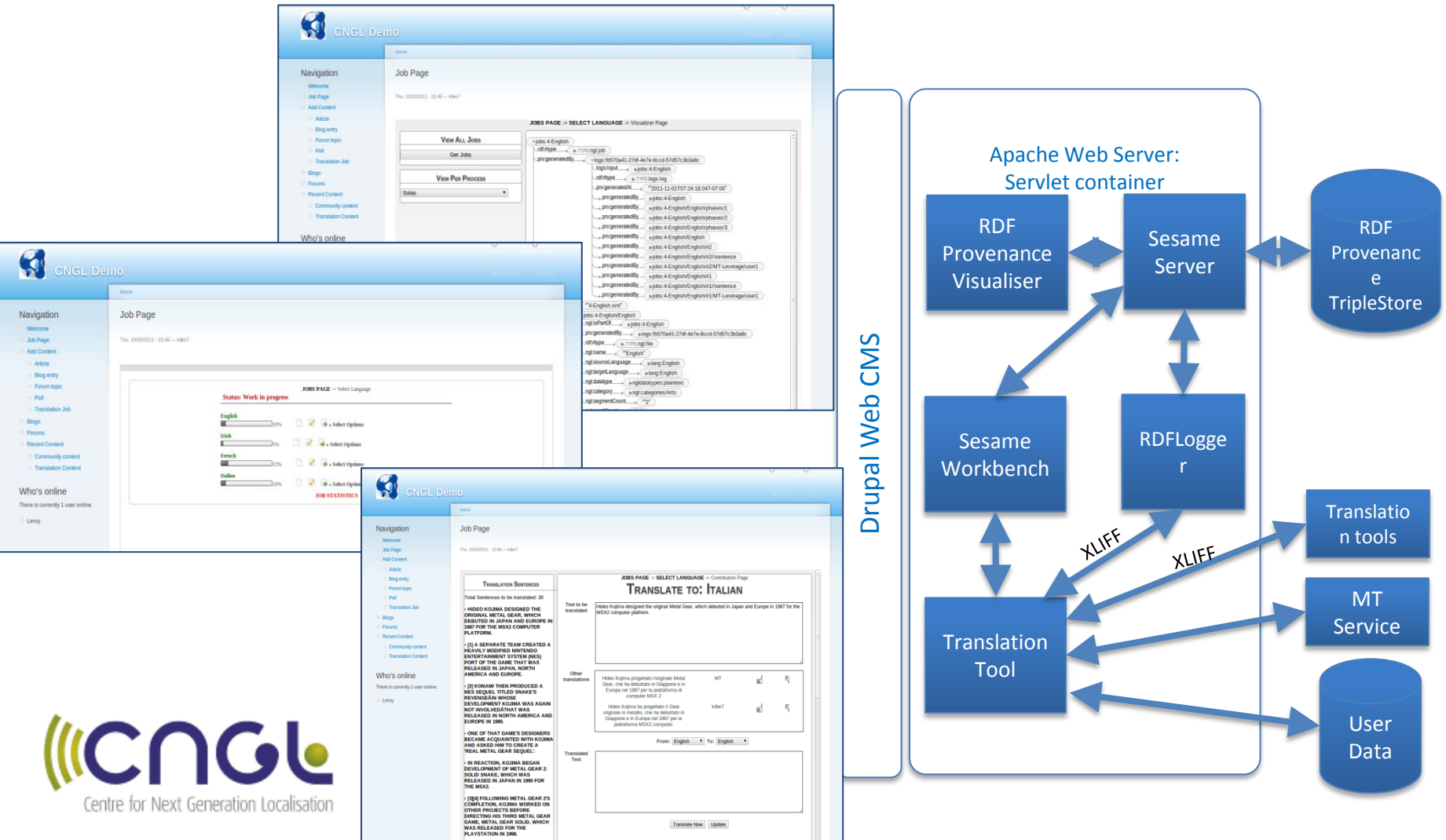


**<..> = Possible MLW-LT Metadata**

# Localisation Quality Assurance

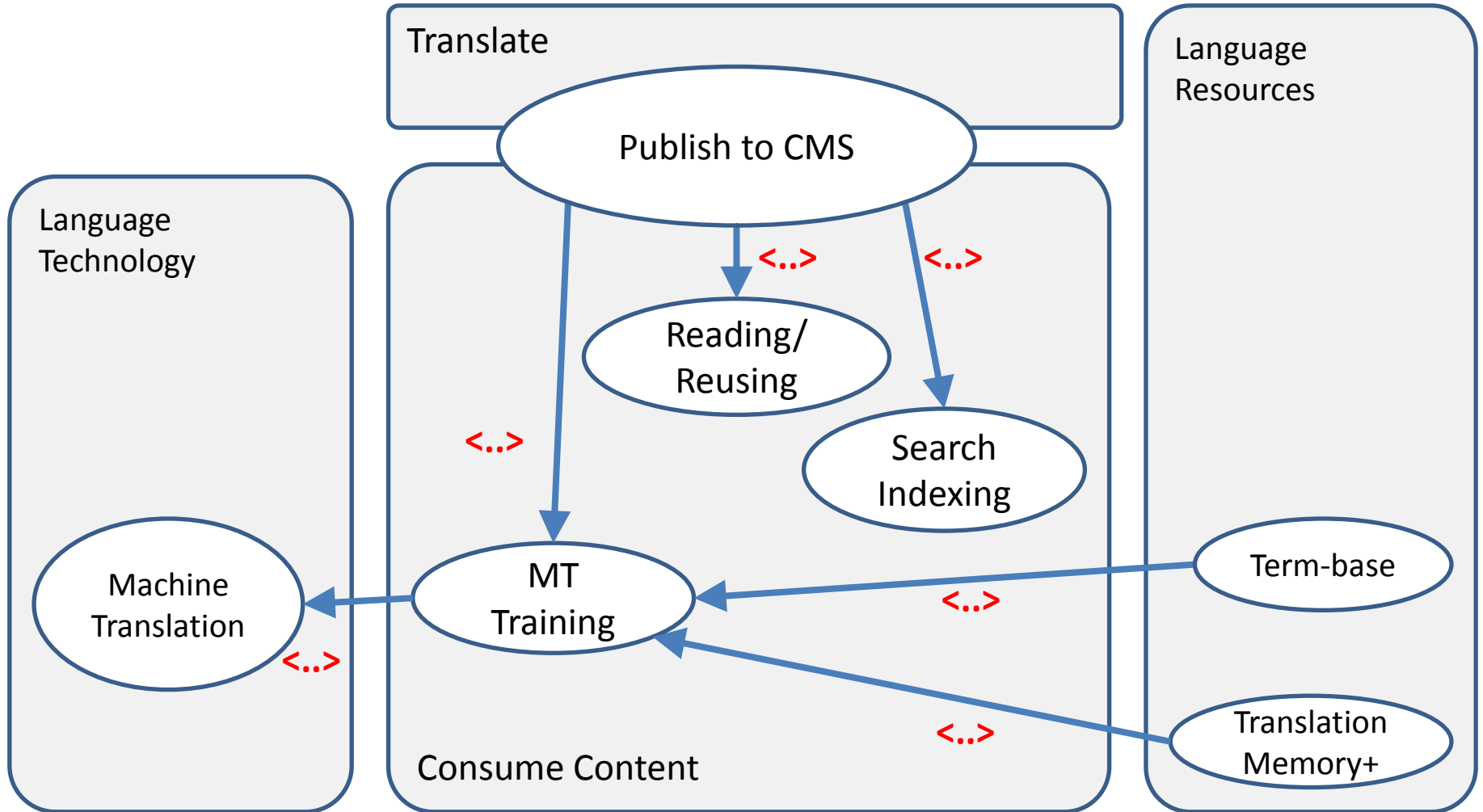


# CMS-L10N integration via RDF & XLIFF





# Leverage Target Quality Meta-data



<..> = MLW-LT Metadata

# Rich Meta-data for TM Leverage

The screenshot displays the Swordfish translation software interface. The main window is titled "Swordfish - C:\Users\Lucia\Desktop\FilesWebinar\Test3.xlf". The interface is divided into several panes:

- Source Text: en:** Contains the source text in English. The first segment is "2010 1st XLIFF Inernational Symposium - Limerick, **Test**". The second segment is "The first XLIFF Symposium 2010 is taking place in Limerick, Ireland on 22nd September Test. It will be part of the LRC preconference Test. 1st XLIFF International Test 22 September 2010, Limerick, Test".
- Translations: es:** Contains the translated text in Spanish. The first segment is empty. The second segment is empty.
- TM Matches:** Shows a 94.0% match rate. Two matches are listed:
  - 2010 1st XLIFF Inernational Symposium - Limerick, **Ireland**
  - 2010 Primer Simposio Internacional sobre XLIFF - Limerick, Irlanda
- Match Properties:** A pop-up window showing detailed metadata for the selected match:

|                 |                      |
|-----------------|----------------------|
| 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               |

The bottom status bar shows "Words: 7 Characters: 56" and "Context: 8".

# Next Steps

- Contribute to MLW-LT requirements gathering
  - Breakout session Friday
  - Feedback on Requirements
    - New ones? Priorities?
    - <http://www.w3.org/International/multilingualweb/lt/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 ....