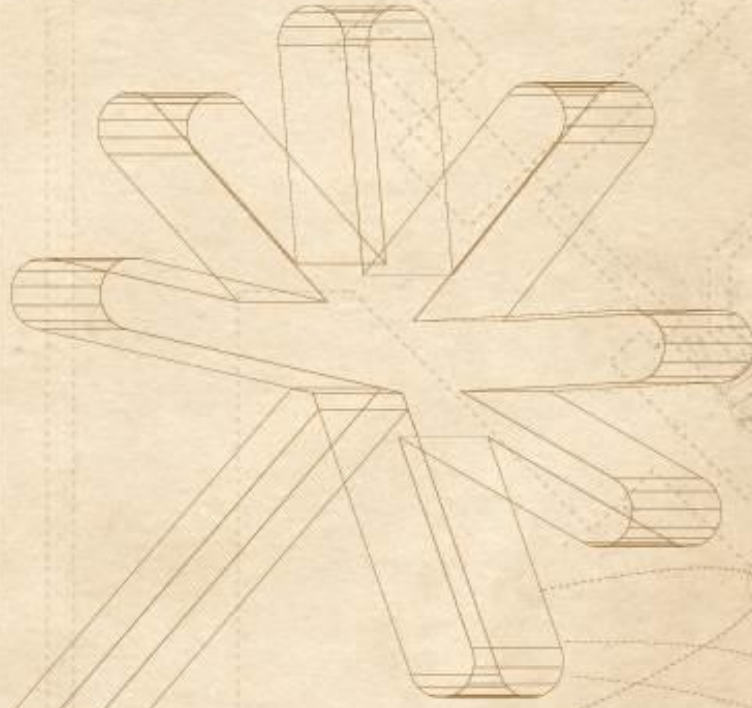


cocomore



Multilingualism & Drupal

The MultilingualWeb-LT project

Denver, CO. March 20th, 2012





Agency for integrated communication and IT services

120 employees

Offices in Germany and Bulgaria

Our clients include Fresenius, Nestlé, Procter & Gamble, PwC, RTL, Sanofi-Aventis

Drupal is our CMS of choice

What we will talk about today



A language technology project connecting CMS, language service providers and other technologies

A new (W3C) standard for the multilingual web to enhance quality of machine and human translations and to save time

How Drupal will become the first CMS to implement this new standard

What we will talk about today

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)



Source: <http://xkcd.com/927/>

commonore

01 | What is the MultilingualWeb-LT project?

02 | What's in it for Drupal?

03 | Discussion / Q&A

The MultilingualWeb-LT project



MultilingualWeb-LT

LT stands for 'Language Technologies'

Bringing CMS and localization chain closer together

Creation of translation metadata standard building on existing standards

Why metadata are important

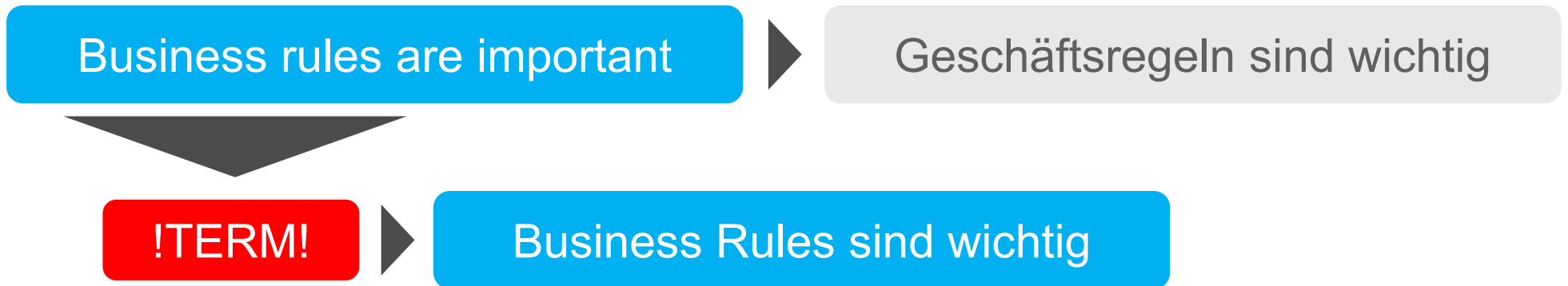
Metadata can assist the choice of LT resources

Domain & genre

Do-not-translate items, e.g.terminology

Example: terms and terminology

Metadata marking of terms can avoid the problem:



Example: domains

Translations may differ depending on the audience:

Could you pass the butter?

Könntest du mir die Butter reichen?

!DOMAIN!

Könnten Sie mir (...)?

Könntet ihr mir (...)?

The MultilingualWeb-LT project: metadata are needed

- Metadata can assist the choice of LT resources
 - Domain & genre
 - Do-not-translate items
 - Terminology
- Information might be present in original content or known by the content manager/creator
 - CMS should be able to store this information
 - CMS need to be able to provide this information to LSP
 - Metadata need to be standardized to allow wide-spread adoption & interoperability
- Current stage in the project: standard requirement gathering

The MultilingualWeb-LT project: the goals

MultilingualWeb-LT will create:

The MultilingualWeb-LT standard for content metadata

Reference implementations in Drupal and in the systems of LSPs and research institutions

Enable spread and interoperability across systems through a well-defined standard

The MultilingualWeb-LT project: the consortium

The consortium

Thirteen partners working on the standard and its implementation until the end of 2013

We will make Drupal the first open-source CMS to handle MultilingualWeb-LT metadata



Consortium even wider at W3C Working Group level. Goal: Make MultilingualWeb-LT a W3C recommendation

commonore

01 | What is the MultilingualWeb-LT project?

02 | What's in it for Drupal?

03 | Discussion / Q&A

MultilingualWeb-LT and Drupal: the inclusion of the metadata

Drupal will be enhanced to allow the inclusion of the MultilingualWeb-LT metadata. We envision three ways in which metadata could be applied:

- by the content manager
- through automatic process either in the CMS or through a service
- by the processes and technologies of the LSP

MultilingualWeb-LT and Drupal: possible formats

HTML5 + RDFa



On-the-Web content and their 'on-the-fly' translations by MT systems

'Content harvesting' for MT training purposes

Recent 'translate' attribute inclusion in HTML5
great step forward

MultilingualWeb-LT and Drupal: possible formats

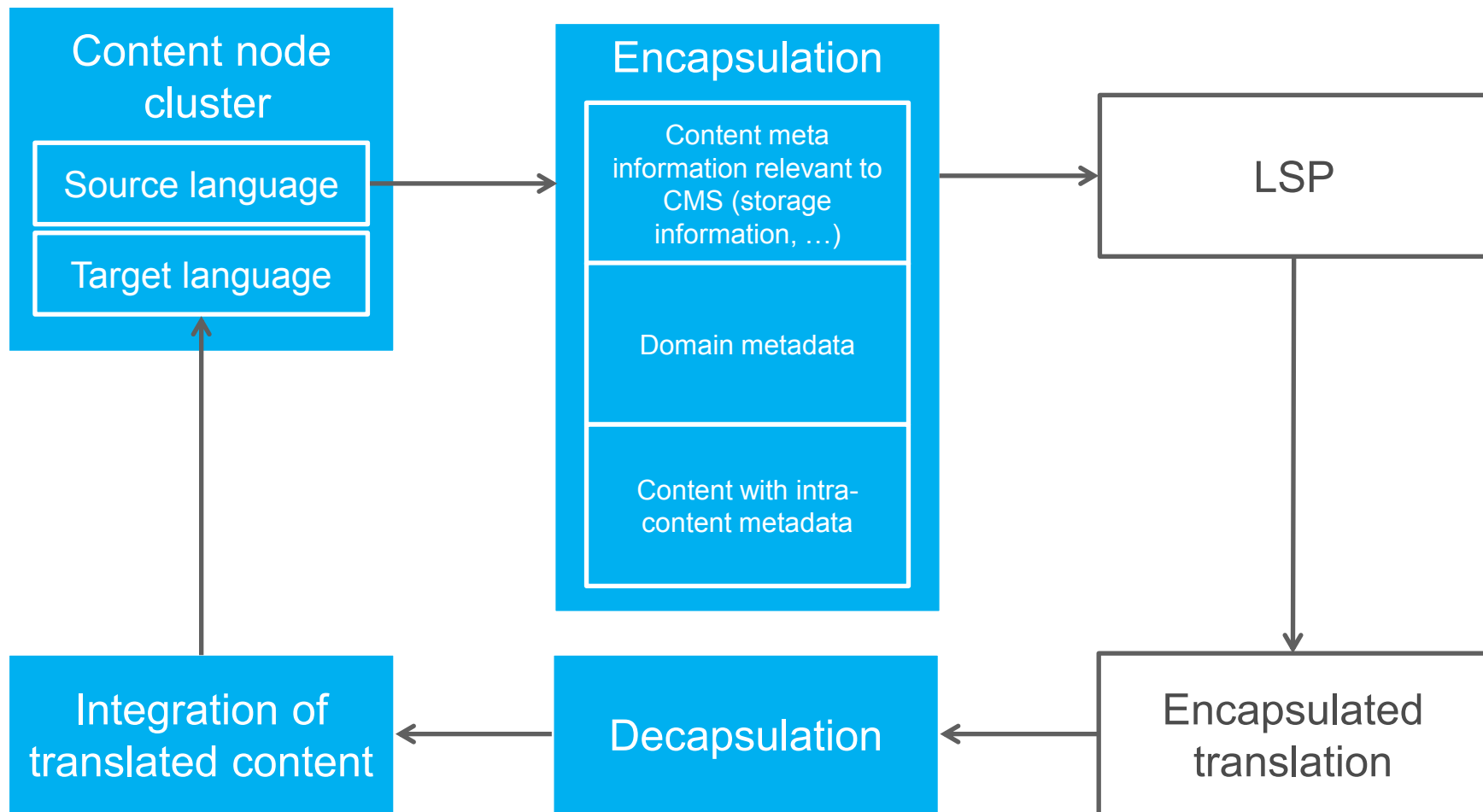
XLIFF

For systems that require parallel corpora, because both source and target languages can be relayed

Often used broker format in LSP systems

Possible meta-data integration solutions for XLIFF are currently being evaluated.

Example: possible workflow integration



Links and contact information

W3C

Survey on International Web Content Requirements:

<http://bit.ly/ypBVHJ>

(Please fill the survey out until March 31st, 2012.)

W3C MultilingualWeb-LT Working Group:

<http://bit.ly/zuGPOW>

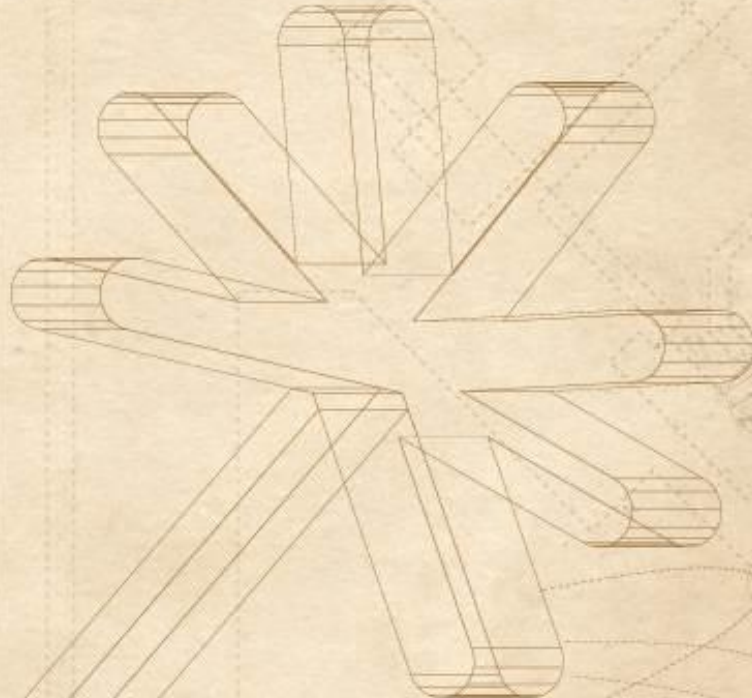
cocomore 
multilingual-web@cocomore.com

mlw@cocomore.com



The MultilingualWeb-LT Working Group receives funding by the European Commission (project name LT-Web) through the Seventh Framework Programme (FP7) Grant Agreement No. 287815 (<http://bit.ly/z33T67>)

cocomore



Discussion and Q&A