DITA Best Practices for Localization

Andrzej Zyroń: azydron@xml-intl.com
DITA TC 20-December-2005
General Principles

• Generic rules for XML localization:

• Keep it simple – KIS
  – Keep granularity level as low as practically possible
  – Try not to nest elements

• Use XML aware CMS system for workflow and document management
What to avoid

• Breaking linguistic integrity for nested elements:
  – `<p> PCDATA <ol><li> PCDATA <p> PCDATA</li></ol>PCDATA</p>`
• Unnecessary nesting of elements
• Translatable attributes
• Be careful with:
  – `Translate="no"`
• Use `xml:lang` only if you absolutely have to
What to do

• Clearly define *true* inline elements
• Clear guidelines re linguistic integrity and nested elements
• Use Automated Localization Workflow
  – Use Integrated CMS TM
  – Use automated TM
  – Control your own matching and word counts
• Produce XLIFF files for translation
  – Protect DITA document structure
  – Language Service Providers have to handle only one standard format
Traditional Electronic Document Translation Scenario

Publishing

- source text

Translation

- extract
- extracted text
- tm process
- prepared text

- target text
- merge
- translated text
- translate

- target text
- QA
- source text
- extract

- target text

- translated text

- translated text
- tm process
- extracted text
- extract
- source text
- source text
- source text
- source text
- source text
- source text
Breakdown of Translation Costs

Source Professor Reinhard Schäler LRC - ASLIB 2002
Automated Translation Scenario

Publishing

xml:tm source text

extract

Extracted text

tm process

leveraged matching

perfect matching

XLIFF file

Automatic Process

Web

QA

Web service/ interface

Translate

Translator

xml:tm target text

Automatic Process

merge