DITA and Translation
Reduce translation costs and speed the time to market of your global content

JoAnn T. Hackos, PhD
President
Comtech Services Inc.
Executive Director
The Center for Information-Development Management (CIDM)
October 24, 2006
Translation costs

• DITA Business Case

• Many organizations can win support for their business case to move to DITA by demonstrating that the costs of localization and translation can be dramatically reduced
Translation costs

- Current: 100%
- +TM: 75%
- +XML/DITA: 65%
- +Reuse: 55%
- +WF at transl: 50%
- +Integr.WF: 40%
- +Pre-transl: 30%

20% of cost savings dependent on the CMS
20% of cost savings dependent on XML & DITA
Translation costs

- Translation memory (TM) reduces translation and quality assurance time [question 2]
- XML/DITA reduces desktop publishing time
- DITA reuse reduces translation time but may add administrative overhead
- Workflow at the localizers may reduce administrative overhead
- Integrated workflow with a content management system reduces administrative costs
- Pre-translation in the CMS with a DITA source further reduces translation and quality assurance (xml:tm)
Translation costs

• Translation costs may be reduced even further by adjustments to content quality
  • Consistent terminology
  • Consistent phrasing
  • Investment in editorial review of the source language before translation
  • Investment in controlled language tools
  • Managing DITA for translation [question 3]
What are the opportunities?

- Enhanced standards at the source
- More opportunities for standardized, consistent content
- More consistent branding of content
- Customized, branded look-and-feel for OEMs and VARs
- More reliable use of translation memory
- Better schedules, less chaos at the end
- Reduced time and cost of production in multiple languages
- Automated topic interrelationships to eliminate hand-coding of references and links
- More languages and greater customization for global audiences
What does DITA mean for localization?

• Topics approved and ready for localization sooner – faster
• Only new or changed topics to be translated
• Only new or changes content within topics to be translated
• Topics reused in multiple contexts, reducing duplication or near-duplication
• More opportunities to control source language through disciplined, structured authoring
• Coming (?), ability to edit/translate at the map level
What does DITA mean for production?

• No formatting in the XML source
• Styles added after assembly
• Topics assembled into ditamaps
• Automated production of mapped content into multiple media
• Filtered content for language- and locale-specific information through the ditaval file
Discussion topics

- How do you handle “not to be translated” elements in a topic?
- What are the advantages of multi-language generated text?
- How do you handle text in a ditamap?
- What should you avoid in your source topics?
- How do you make your service partners DITA-aware?
Managing DITA for translation

• Using DITA attributes in your source content
  • xml: lang
  • translate="yes/no"
xml:lang

• What is it?
  • The xml:lang attribute can be associated with any DITA topic and with most DITA elements
  • The xml:lang attribute provides a label for each language used in a topic
  • The default attribute is xml:lang="en-us" for US English
xml:lang

• Why use it?
  • Identify the language of the topic, i.e., xml:lang="fr-ca"
  • Identify the language of an element if it is different from the language of the topic, i.e.,
    • first warning in en-us
    • second warning in fr-fr
    • third warning in fr-ca
    • fourth warning in it-it
translate=

• What is it?
  • The translate attribute can be associate with any DITA element
  • The translate attribute provides instructions to the translator
  • The translate attribute preserves certain words, phrases, or sentences in the source language
Why use it?

Preserve a word, phrase, or sentence that should remain in the original source language, i.e.,

- `<title>Winter Wren</title>`
- `<shortdesc translate="no">Troglodytes troglodytes</shortdesc>`
- `<author>JoAnn</author>` in English
- `<author>Giovanna</author>` in Italian
- `<author translate="no">JoAnn</author>`
Generated text

- DITA automatically generates some text upon processing and allows you to generate additional text of your own
- Generated text simplifies author text entry and ensures consistency in terminology
- Generated text should be placed in a resource file for each language
Map text

- A DITA map is an xml file and may contain text for elements such as
  - Map title
  - Navigation title
  - Map short description
  - Metadata and index terms
- Ensure that your localization vendor receives the map file in addition to the topic files
Inline elements

- Inline elements are used within a sentence to label a word or phrase semantically
  - `<b>ENTER</b>`
  - `<uicontrol>Format menu</uicontrol>`
  - `<userinput>www.comtech-serv.com</userinput>`
- For translation, use the semantic inline elements rather than the typographic elements like bold and italic
- Remember that typographic elements do not necessarily translate in other languages
Segmentation

• When you send topics for translation, the localization vendor transforms the text into a form ready for the translator
• Your text is split into blocks of text
• Blocks of text can be further segmented into sentences
• Translation tools working with XML split segments by working with the xml tags
• Most inline elements should be ignored by the segmentation rules so that they don’t break text into awkward sentence parts
• Best practice suggests that you think carefully about your xml tagging so that you don’t create awkward breaks
Document design example

• Good structure
  <p>Here is a list of the ingredients to mix:</p>
  <ul>
   <li>2 tablespoons peanut butter</li>
   <li>2 tablespoons soy sauce</li>
   <li>juice of 2 limes</li>
  </ul>
  <p>Mix thoroughly until all lumps are removed</p>

• A list separated from the paragraph will be simple to translate
• Poor structure
  <p>Here is a list of the ingredients to mix:
  <ul>
  <li>2 tablespoons peanut butter</li>
  <li>2 tablespoons soy sauce</li>
  <li>juice of 2 limes</li>
  </ul>
  Mix thoroughly until all lumps are removed</p>
• A list inside a paragraph should still translate easily
Document design example
(continued)

• Bad structure
  
  
  <p>Mix these ingredients together</p>
  <ul>
    <li>2 tablespoons peanut butter</li>
    <li>2 tablespoons soy sauce</li>
    <li>Juice of 2 limes</li>
  </ul>

  so that they are thoroughly combined.</p>

• A list inside a sentence will cause serious problems for translation
Indexing

• Place index items carefully to avoid creating segments that are difficult to translate
  • Place in the topic prolog metadata
  • Place at the beginning of a block, immediately following the start tag
  • Don’t relocate an inline index tag from one version of a topic to a new version
Associated Standards

Reduce costs further by using complementary standards:

- LISA OSCAR xml:tm – XML based Text Memory (author and translation memory)
- LISA OSCAR TMX – Translation Memory eXchange
- LISA OSCAR SRX – Segmentation Rules eXchange
- OASIS XLIFF – XML Localization Interchange File Format
- W3C ITS – Internationalization Tag Set
Working with your localization service partners

- Involve them in your early planning process
- Ensure that they are familiar with handling XML content
- Ask about translation-industry standards (see previous slide)
- Help them understand your business goals
- Recognize that DITA will decrease their revenue
- Explain that by reducing the costs for one language, you will be able to translate into more languages
Interested in learning more?

- Join the OASIS DITA Technical Committee
  - Translation subcommittee
- Reference [http://dita.xml.org](http://dita.xml.org) for translation best practices in the future
- Refer to the standards work being done by the W3C working group on translation (ITS) and the LISA OSCAR standards committee
Resources

- Comtech at [www.comtech-serv.com](http://www.comtech-serv.com)
  - DITA, XML, Structured Writing, Minimalism, Content Management Strategies workshops
  - Consulting – information architecture, process redesign, change management, requirements and vendor selection
- CIDM (Center for Information-Development Management) at [www.infomanagementcenter.com](http://www.infomanagementcenter.com)
  - Conferences, membership, senior management community
Conferences

• DITA Europe
  • November 2-3, 2006 Frankfurt, Germany

• Content Management Strategies 2007
  • March 26-28, 2007 Boston, MA
  • [http://www.cm-strategies.com/index.htm](http://www.cm-strategies.com/index.htm)

• Best Practices 2007
  • September 24-26, 2007 Atlanta, GA