

# *Common Sense* **Advisory**

## **Standards for Global Content Applications**

Donald A. DePalma, Ph.D.  
Chief Research Officer



# Who is Don DePalma?

- An academic: historical phonology, generative grammar, computational linguistics
- A technical writer: database, 4GL, repository, technology standards
- An entrepreneur: database management
- An analyst: database, application development, content management, data warehousing, globalization
- A strategist: global content management
- An analyst: global business, services, technology, and best practices – and standards instigator



# Standards are documented agreements

- William Sellers – “On a Uniform System of Screw Threads,” 4/21/1864 (60° angle)
- Cadillac: “Standard of the World” branding due to early 20c ability to disassemble cars and re-assemble with interchanged parts
- 1978: EEC highway specifications for building E20 in ČSSR (enabled use of standard materials)
- Dimensional building materials; paper; electrical plugs; etc.
- Technology: programming languages, database, remote procedure calls, web services

Fig I  
Graphic Representation of  
Formulae for the pitch of Threads of  
Screw Bolts.

*horizontal scale 1/8" vertical scale 1/32" diam full size*

*The dotted lines show the usual best form of Threads of non-ferrous bolts.*

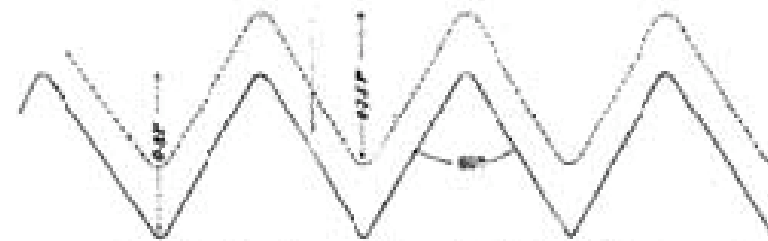


Fig. II. Usual Form of Threads best Workshop practice.



Fig. III. Form of Thread introduced Mr. Whitworth.



Fig. IV. Form of Thread proposed by Mr. Sellers.



# Where do content-related globalization standards and specifications come from?

- De jure:
  - Governments, multilateral organizations, industry associations, purpose-created bodies such as ASTM, ISO, JCP, LISA, UC
  - Preferably open, public, non-profit, supported, transparent
- De facto:
  - Standards can be set by business entities like IBM (e.g., SNA), Microsoft (Windows and its APIs)
  - Less open, but frequently documented and supported
- Reality check:
  - You have no choice on some
  - Some are about about good hygiene, but little used
  - Others are not ready for prime time in their current pfor



# The code and content ecosystems need an injection of globalization DNA

- Code:
  - For platform, middleware integration, services
  - De facto: Windows, HTTP, .Net objects, Java, libraries,
  - De jure: Unicode, XML buffet
- Content:
  - For dealing with format, representation, tagging, segmentation, and other attributes rather than about the words themselves.
  - De facto: FrameMaker, Epic, Word, DITA, S1000D
  - De jure: XML-based TMX and TBX dealing with format, representation, tagging, segmentation, and other attributes rather than about the words themselves.
- Process:
  - Six Sigma, ISO 9000
  - ASTM, CEN
  - Medical à la MDD 93/42/EEC, HIPAA



# The character, language, and locale nits of globalized content applications

- Character codes
  - Unicode
  - ANSI
  - GB18030
- Language codes
  - ISO 639: 2- and 3-letter codes
- Locale identifiers
  - IETF 3066
  - ISO 15924
  - IEEE POSIX
  - UC CLDR

## The WØRD:

- Character codes are basic building blocks.
- Language codes provide practical identification for large-scale translation efforts.
- Locale identifiers take language ids to the next level by adding details about where the language is used
- Bottom line: necessary, useful and actually used



# The plumbing of globalization

- Translation memory
  - TMX (Translation Memory eXchange)
- Segmentation
  - SRX (Segmentation Rules eXchange)
- Interchange
  - XLIFF (XML Localization Interchange File Format)
  - OLIF (Open Lexicon Interchange Format)
- Terminology
  - TBX (TermBase eXchange)

## The WØRD:

- Rather than functioning as middleware, these specifications often enable export and replace rather than interchange.
- Second-generation specs slow in coming to fruition.
- Open interfaces not a given.
- The bottom line: Industry has failed at creating a vital ecosystem for data and content interchange.



# The Holy Grail of globalization: Better process

- Quality control and assurance:
  - ISO 9000:2000
  - FMEA
  - SAE J2450
  - LISA QA Model
- Metrics:
  - GMX-V / C / Q
- Process specification:
  - ASTM
  - CEN

## The WØRD:

- Quality can be extraordinarily subjective, app-specific.
- Metrics depend on extra-cost, volunteer, or ad hoc resources.
- Certification in ISO can be a way to pave cowpaths of inefficient, ineffective, targeted processes.
- Process specs are more guidelines on how to develop a process spec than a useful exercise as currently instantiated.
- The bottom line: Nice to have specs but far from gotta-have for more companies.



# Will we see the end of Taylorism in globalization processes?

- The problem with all candidate process specifications is that they formalize an age-old process of translate-edit-proofread (TEP).
- Quality improvement methodologies like Six Sigma teach us that adding steps to a process only increases the probability of adding errors. The TEP model forces each individual and task to work in lockstep rotation according to time-motion studies. ASTM, CEN, and ISO perpetuate this approach.
- An alternate future: Innovators will incorporate collaboration across borders, communities of translators, updates of centralized translation memories and terminology bases, in-situ use of style and grammar guides, translation more intimately tied to authoring, and Agile development methods.
- The bottom line: Modern technologies will make TEP's sequential process redundant.



# Summary

- Standards are agreed-upon conventions that allow “things” to work with other things.
- Content owners must actively address some, but not all, specifications related to globalization
- Some globalization specifications and standards are no-brainers, others require a major stretch, still others are theoretical exercises



# What to read

- Global Watchtower (Bob Dylan reference) blog at [www.globalwatchtower.com](http://www.globalwatchtower.com)
- Research at [www.common sense advisory.com](http://www.common sense advisory.com)
- “Business Without Borders: A Strategic Guide to Global Marketing”
- “Perspectives on Localization,” (ATA) edited by Keiran Dunne
- *MultiLingual* magazine, AIIM *E-DOC* magazine
- Proceedings of ASTM, CEN, ISO, etc.



**Thank you.**

Don DePalma

[don@commonsenseadvisory.com](mailto:don@commonsenseadvisory.com)

+1.978.275.0500 x1001

