

# Standards.

## We Know Where We've Been, But Where Are We Going?

Hans Fenstermacher (Translations.com/TransPerfect/**GALA**)

David Filip (LRC/CNGL, **XLIFF-TC**, **Unicode-ULI TC**, **TAUS**)

Patrick Guillemin (**ETSI**)

Jaap van der Meer (**TAUS**)

Donna Parrish

# Standards

Format:

- Introductions, presentations by each panelist
- Questions to the panelists from each other, from the moderator and from you
- Answers?

## Standards that might come up in discussion today

**EN 15038:** quality standard developed especially for translation services providers, including regular audits and certification.

**SRX (Segmentation Rules eXchange):** provides a common way to describe how to segment text for translation and other language-related processes.

**TBX (Term Base eXchange):** XML-based format for standardized exchange of data from terminology databases.

**TMX (Translation Memory eXchange):** XML-based format for the standardized exchange of translation memory data.

**UNICODE-ULI:** Unicode is a standard for character encoding. ULI is the Unicode Localization Interoperability Technical Committee (ULI), which works to ensure *interoperable data interchange* of localization data.

**XLIFF (XML Localization Interchange File Format):** an XML-based format created to standardize localization across multiple file formats.

**UAX#29 (Unicode Annex 29): Unicode Text Segmentation.** This annex describes guidelines for determining default segmentation boundaries between certain significant text elements: grapheme clusters (“user-perceived characters”), words, and sentences. For line break boundaries, see [[UAX14](#)]

## Organizations involved with standards:

**ETSI (European Telecommunications Standards Institute)** produces globally-applicable standards for Information and Communications Technologies. ETSI is an independent, not-for-profit association whose more than 750 member companies and organizations, drawn from 62 countries across 5 continents worldwide. Of interest today: the ETSI Industry Specific Group on Localisation Industry Standards.

**GALA (Globalization and Localization Association)** is a global organization providing resources, education, ideas and research to enhance the power and performance of the translation and localization sector. Membership includes companies worldwide who specialize in language and translation services or technologies, language service buyers and technology providers.

**OASIS (Organization for the Advancement of Structured Information Standards)** is a global consortium that drives the development, convergence and adoption of e-business and web service standards. With its headquarters in the USA, members of the consortium decide how and what work is undertaken through an open, democratic process. XLIFF is under the aegis of OASIS.

**OASIS XLIFF TC** defines, through extensible XML vocabularies, and promotes the adoption of a specification for the interchange of localizable software and document based objects and related metadata.

**TAUS (Translation Automation Users Society)** is an innovation think tank and interoperability watchdog for the translation industry. Its mission is to increase the size and significance of the translation industry to help the world communicate better. TAUS supports entrepreneurs and principals in the translation industry to share and define new strategies via of events, publications and knowledge tools.

**UNICODE – ULI TC (Unicode Localization Interoperability Technical Committee)** works to ensure interoperable data interchange of localization data including segmentation rules and translation source strings and their translations. Initial focus is improved interoperability through segmentation rules based on UAX#29 using SRX or its successor format for rules exchange, and XLIFF format as the standards data and meta-data exchange format in the localization process.

## Panelists:

**Hans Fenstermacher** co-founded the Globalization and Localization Association (GALA). In 2009, Hans was re-elected to the GALA Board and serves as the organization's current chairman.

**David Filip** is a voting member of OASIS XLIFF TC and its liaison in Unicode Localization Interoperability TC. He also serves on the TAUS Standards Advisory Board.

**Patrick Guillemin** is responsible within the ETSI Secretariat for the development, co-ordination and management of new standardization initiatives.

**Jaap van der Meer** is director of TAUS (Translation Automation User Society) and the TAUS Data Association. He inspired and funded the founding meetings of the LISA.

# Jaap van der Meer, TAUS

# TAUS Interoperability

## Program Management Office

2011 - 2015



# Vision Statement

- We envision translation as a standard feature, a ubiquitous service. Like the **internet**, **electricity**, and **water**, **translation** is one of the basic needs of human civilization.

# 20<sup>th</sup> Century Translation

Top-down globalization

Export mentality – pushing out



**One big world**

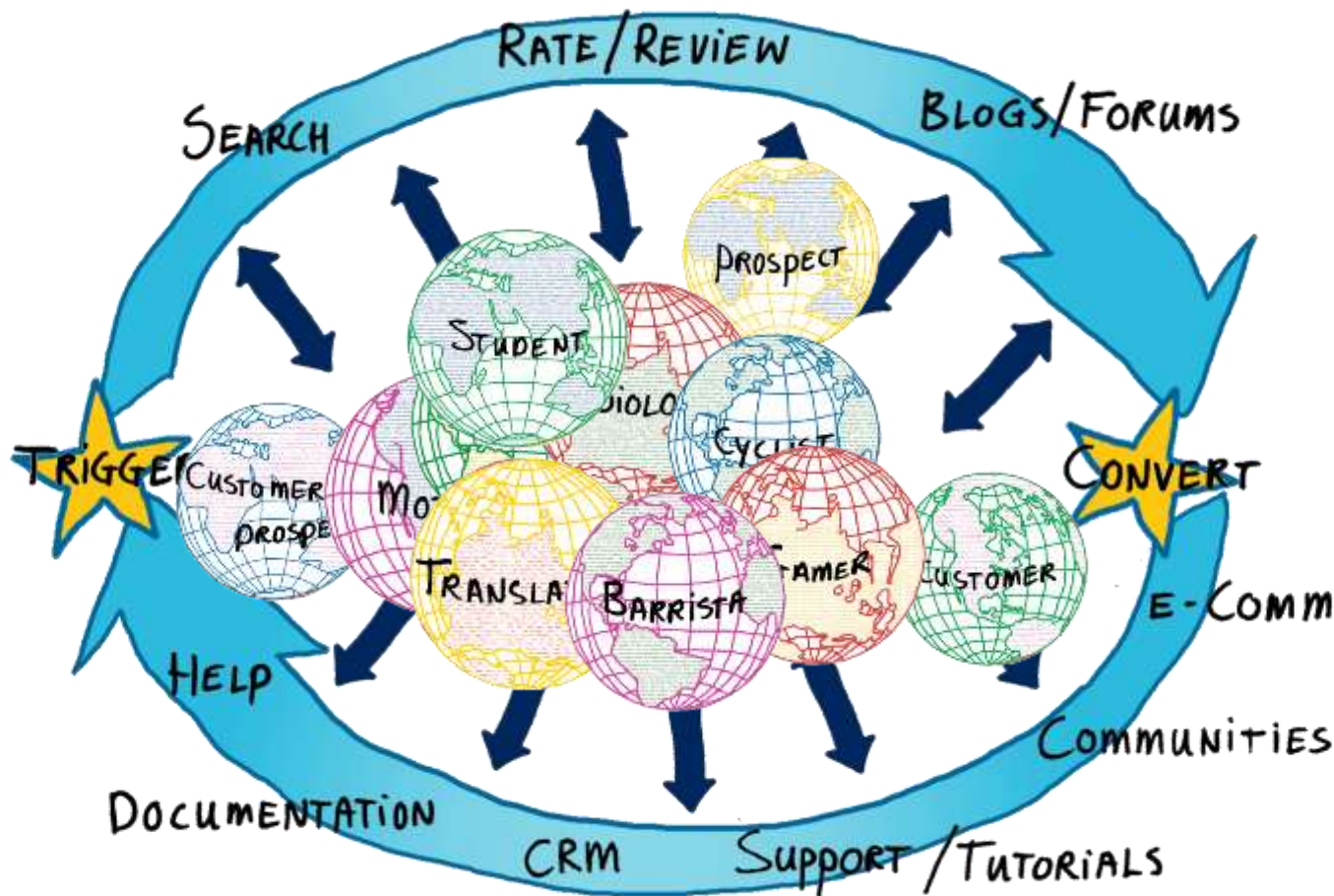
1. One translation quality fits all
2. Selecting locales – limited languages
3. Counting words – owned content
4. TM is core
5. Project-based translation
6. Cascaded supply chain
7. Publisher-driven
8. One directional

# 21<sup>st</sup> Century Translation

Bottom-up and top-down globalization

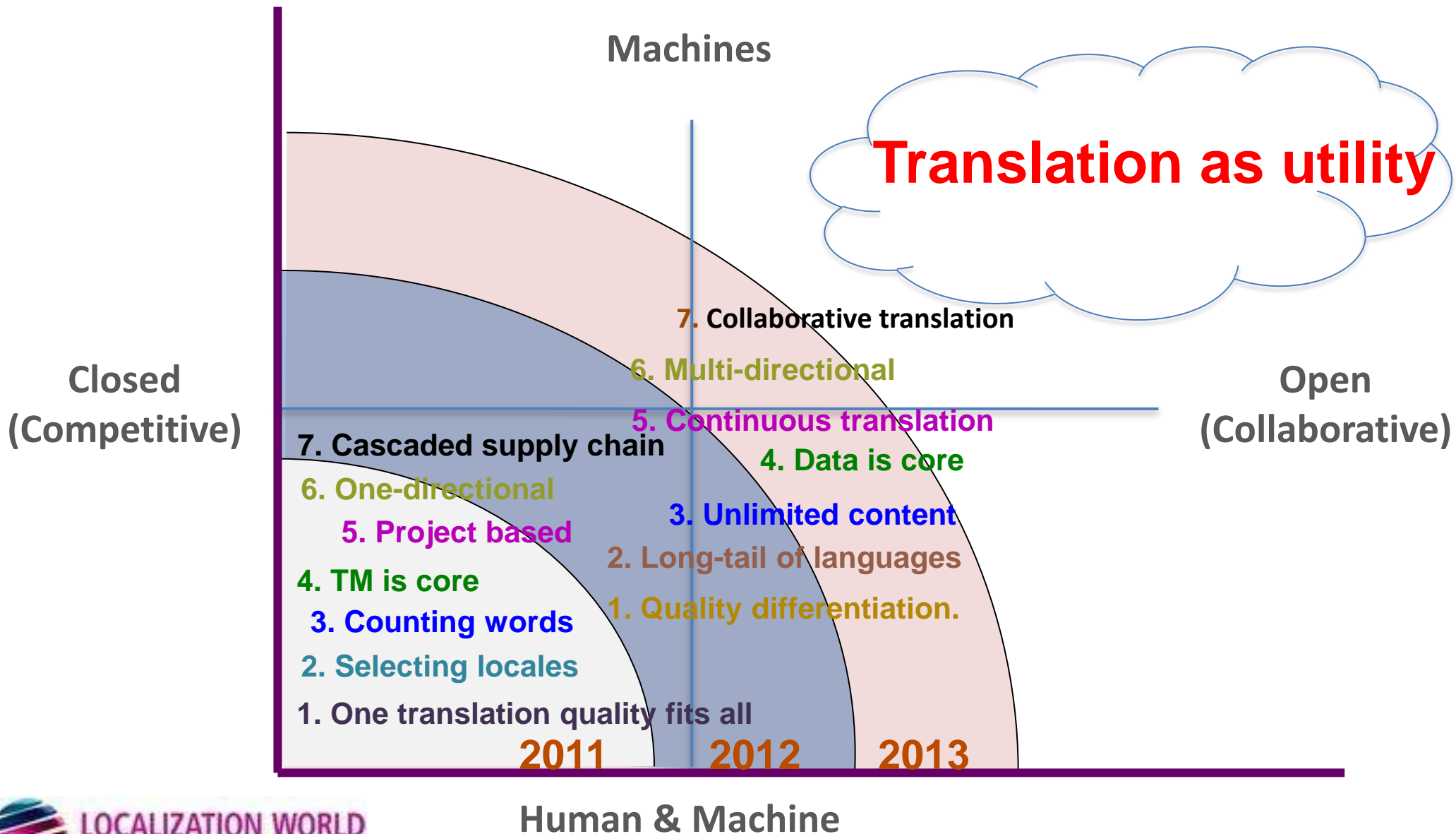
Information is omnipresent – people are connecting

1. Quality differentiation
2. Long-tail– unlimited languages
3. Unlimited content– owned, shared, earned
4. Data is core
5. Continuous translation
6. Collaborative translation
7. User-driven
8. Multi directional



**Many big worlds in one small planet**

# Strategic Direction



# TAUS PgMO Objectives

1. Consolidate **fragmented functions** in global translation industry
2. Ensure an open environment to **stimulate innovation** and automation
3. Reduce friction and **improve translation efficiency** for everyone

# Action Lines

1. Drive translation interoperability
2. Deliver dynamic quality evaluation framework
3. Develop market for open source MT

# Interoperability Action Line

## Work Packages:

- ◆ Market outreach & promotion
- ◆ Interoperability standards coordination
- ◆ Interoperability dashboard

# Focus of Interoperability Action Line

- Ensure that jobs, resources and content can travel without any loss of value, ensuring a competitive industry, facilitating automation and innovation and freedom of choice of tools and vendors.

# Dynamic QE Action Line

## Work Packages:

- ◆ Benchmarking QE models
- ◆ Content profiling & draft dynamic QE model
- ◆ Development of QE dashboard

# Focus of Dynamic QE Action Line

- Development of a quality evaluation model based on content and communication profiles allowing for a more dynamic approach to translation, fewer bottlenecks, price differentiation and greater inclusion of end users.

# Develop Market for Open Source MT Action Line

## Work Packages:

- ◆ Research adoption, user needs and cases
- ◆ Market outreach and education
- ◆ Academic and industry collaboration
- ◆ Communication platforms of resources and tools

# Focus of Develop Open Source MT Market

- Help grow effective use of open-source machine translation through industry outreach events, greater academic/industry collaboration, and high profile, authoritative research and communications.

**Sponsored by:**



**EUROPEAN  
COMMISSION**

# Patrick Guillemin, ETSI

# ETSI ISG LIS

to enhance and maintain LISA Standards TMX, TBX, SRX and GMX with former LISA stakeholders and re-established liaisons with TAUS, GALA, ISO TC 37, OASIS, W3C and to create new localisation standards as appropriate

Founded by common ETSI-LISA Members: WhP (convenor), HP, Huawei, Alcatel-Lucent, Lionbridge, Cisco, IBM..for all

# At what cost?

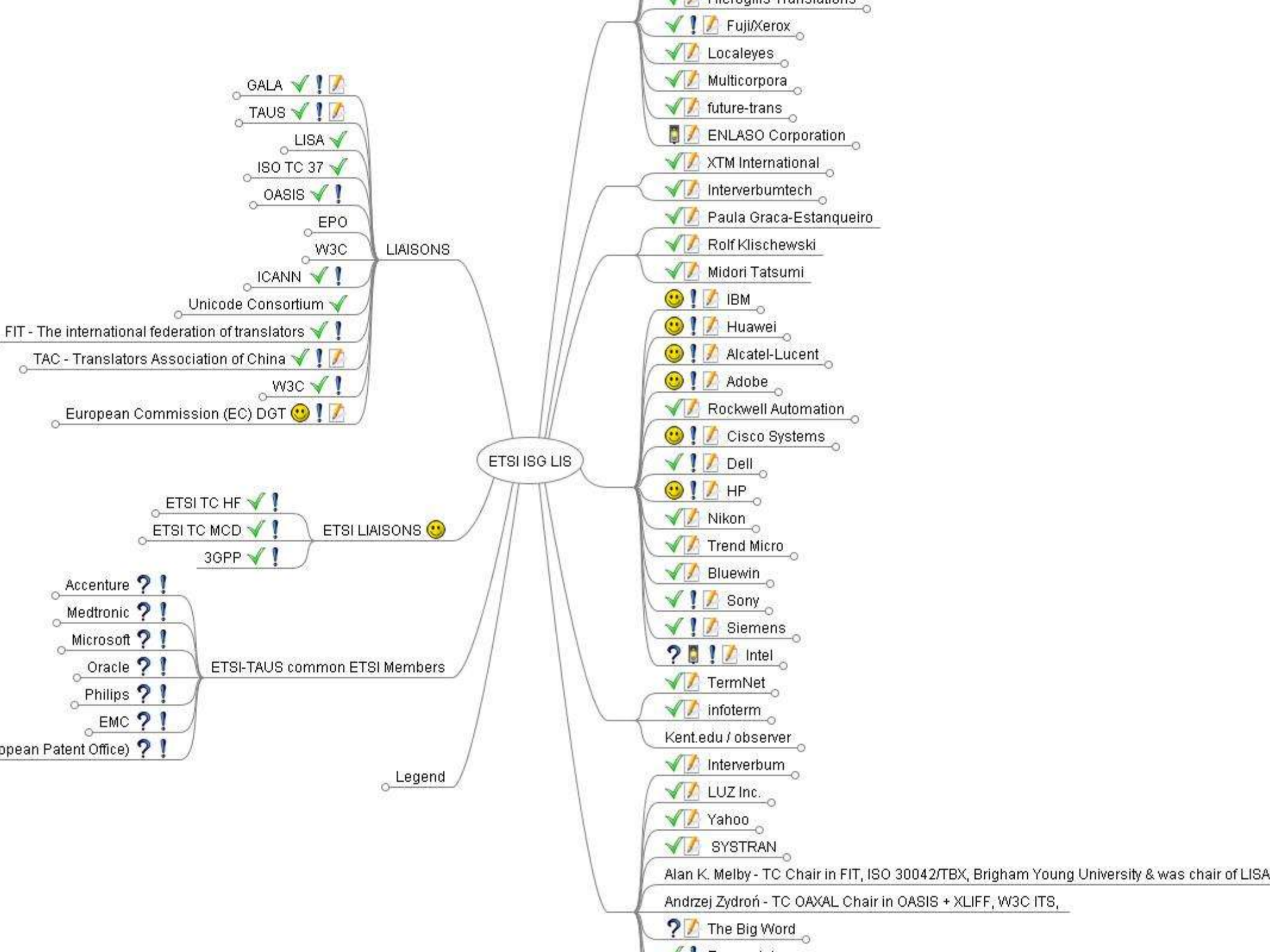
- No cost for LISA, TAUS, GALA members who are already ETSI Members like
  - the 8 ISG founders,
  - Nokia, Intel, Sony, Siemens, Microsoft, Accenture, Medtronic, Philips, EMC, Oracle, EPO, European Commission DGT...
- 100 Euros per delegate, per face-to-face meeting day only for ISG Participants (non-Members of ETSI)

# How can I join and participate?

- 1) Download and sign the ISG Agreement ETSI ISG LIS KIT FOR SIGNATURE.zip  
<http://docbox.etsi.org/HF/Open/ISGLIS/>

\*and\*

- 2) Subscribe to ISG LIS email exploder list and participate in conference calls, virtual meetings and face-to-face meetings collocated with Localisation events in all regions.



**David Filip, OASIS XLIFF TC, Unicode-ULI TC**

# OASIS XLIFF TC 2011-2012 Focus Areas

## ❑ XLIFF 2.0 core specification:

- Committee Draft within 2011
- Public review Q12012
- OASIS standard Q22012

## ❑ Characteristics of 2.0

- Non-negotiable core
- Optional modules for specialized functionalities
- Explicit conformance criteria and processing requirements for both core and modules
- Although the XLIFF 1.x standard was intended primarily as an exchange format the industry practice shows that the defined format is also suitable for storage and legacy content leverage purposes.

# Proposed ULI 2011-2013 Focus Areas

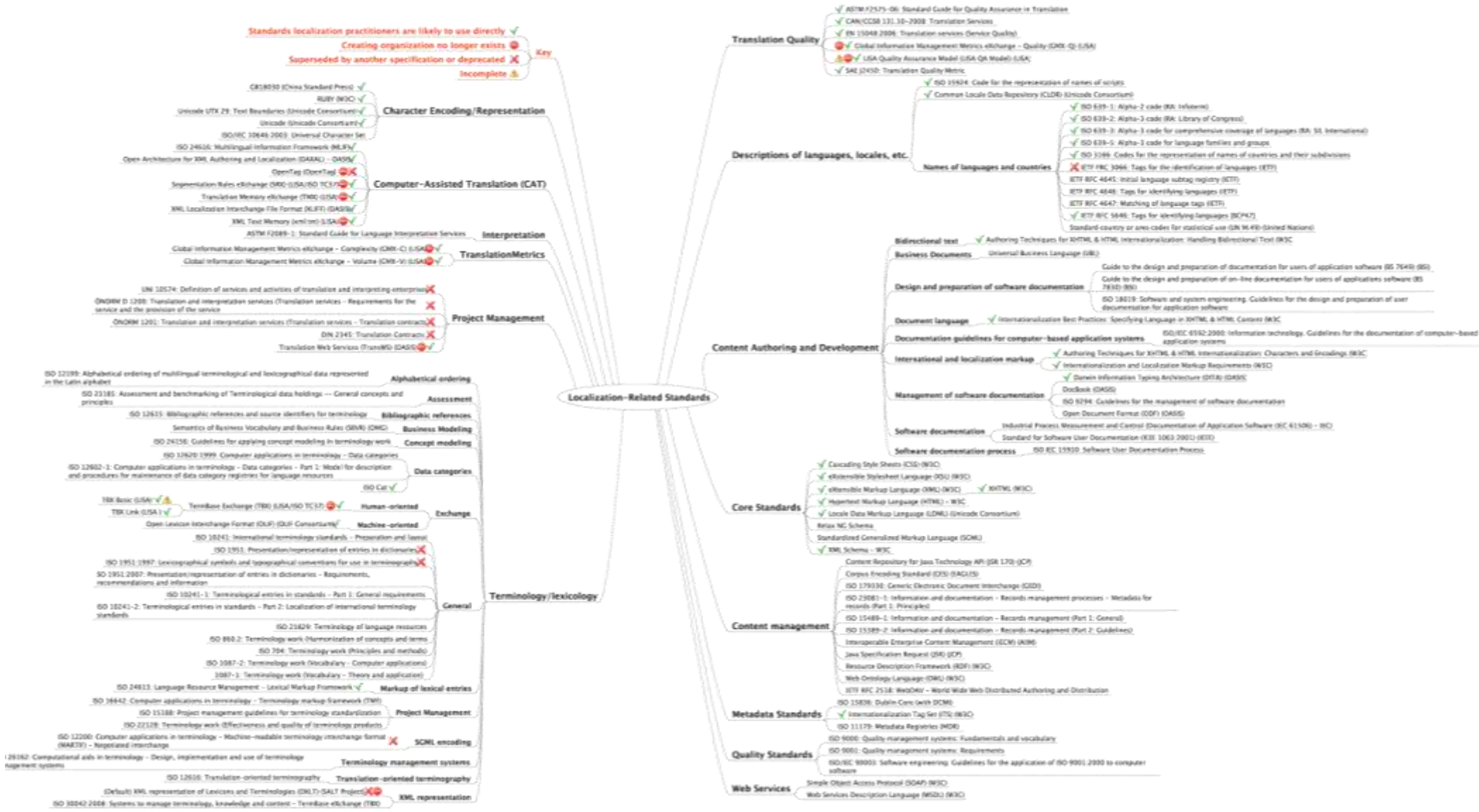
- Consolidate text segmentation standards:
- Leverage [UAX#29](#) as the rule specification standard
- Update SRX (Segmentation Rule eXchange) based on UAX29 for interchange between systems
- Establish core definition at per language level thru [CLDR](#) (Unicode Common Locale Data Repository)
- Allow for extension possibilities per language per domain thru ULI segmentation repertoire (TBD)
- Provide reference implementation where applicable:
- Deliver basic implementation of segmentation thru [ICU](#) (International Component for Unicode)
- Integrate ICU in public domain translation workbench and management software such as [OpenTMS/OpenTM2](#) or [GlobalSight](#)
- Shared reference implementation of XLIFF for localization data transfer

# Comparison of Possible LT-Standardization homes

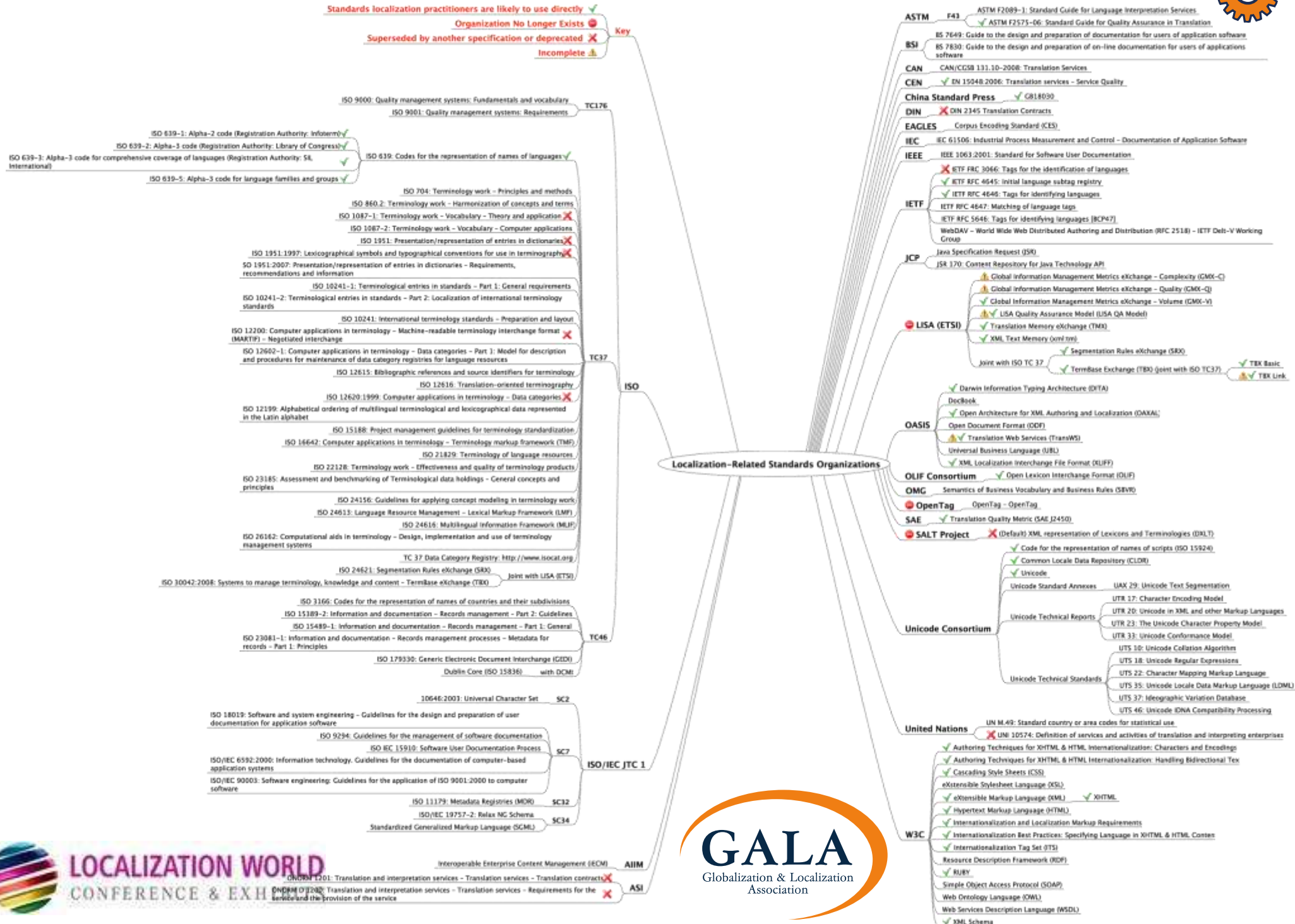
	IPR modes	Mailing lists	Fees	LSP below 50m Euro	proportional vote	Can non-members vote?	Individual or other low cost option?	with TC voting rights?
<b>ETSI</b>	<b>FRAND</b>	restricted	<a href="http://www.etsi.org/WebSite/membership/fees.aspx">http://www.etsi.org/WebSite/membership/fees.aspx</a>	€6,000	yes	NO	free	no
<b>OASIS</b>	RF on RAND	public	<a href="http://www.oasis-open.org/join/categories-dues">http://www.oasis-open.org/join/categories-dues</a>	\$7500 (~€5300)	NO	NO	\$300/\$1200	yes
<b>W3C</b>	RF	public	<a href="http://www.w3.org/Consortium/fees">http://www.w3.org/Consortium/fees</a>	€7,800	NO	NO	free	yes
<b>Unicode</b>	<b>RAND</b>	public	<a href="http://www.unicode.org/consortium/levels.html">http://www.unicode.org/consortium/levels.html</a>	\$7500 (~€5300)	yes	NO	\$75	no

# Hans Fenstermacher, GALA

# Current Localization Standards



# Current Standards Organizations





# Key Pillars

- **Coordination**
  - Too big a problem to handle without a central body
  - Standards bodies know they need help
  - Problems are common: solutions should be, too
- **Education and Training**
  - Making standards useful for people
  - Helping people find the information they need
- **Promotion and Development**
  - Best practices
  - New resources to solve collective needs



# Collaboration

- Need to address needs of entire community of users:
  - Clients
  - Language Services Providers
  - Tools Developers
  - Translators
  - Organizations
- Example: Working with TAUS on
  - Standards taxonomy (current landscape)
  - Development of technical profiles for XLIFF
  - Data and quality models (will make sure our efforts are unified)

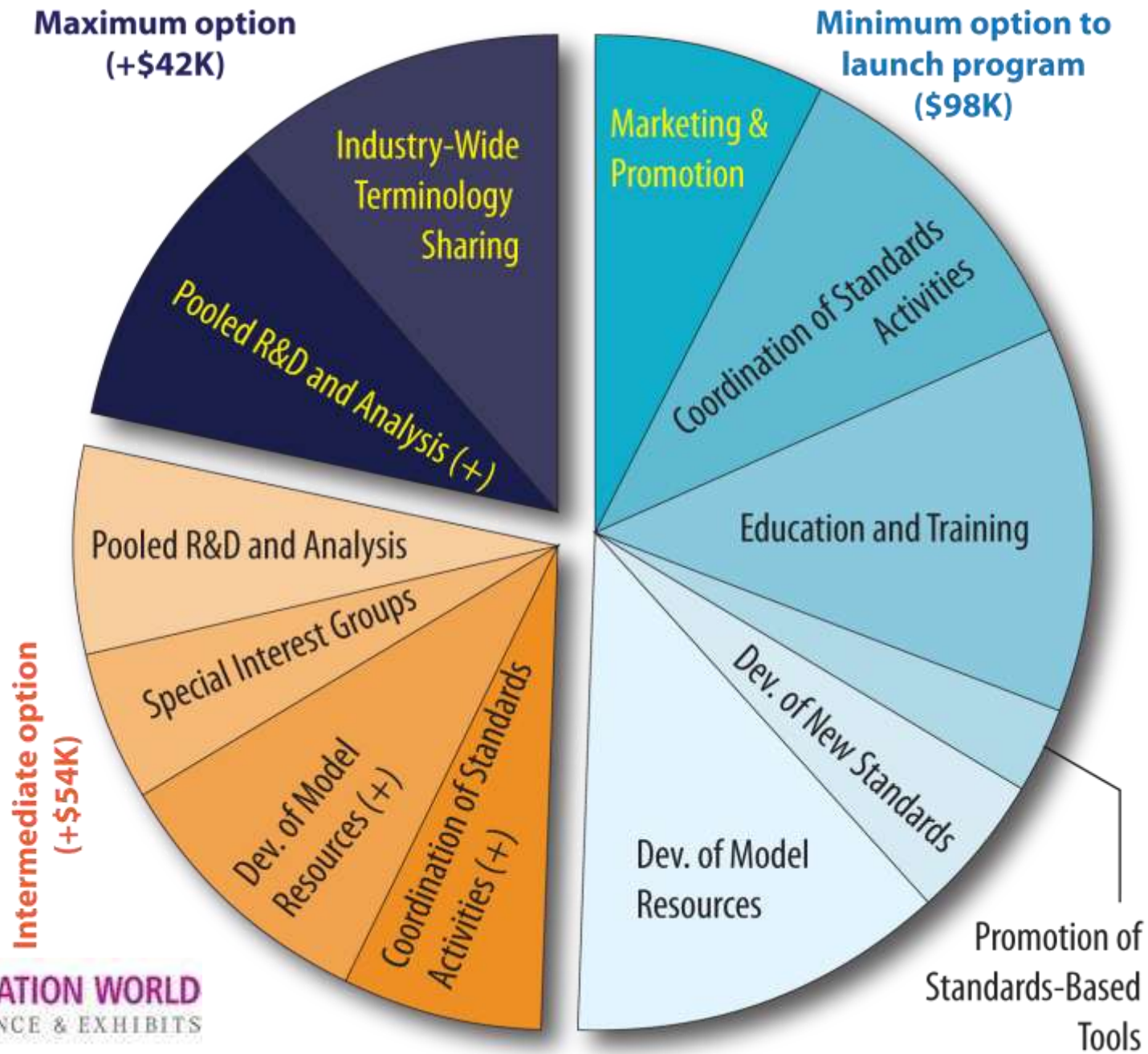


# Deliverables

- Platform for coordination of standards activities
- Education and training materials
- Promotion of standards-based tools
- Development of new standards (Container project, MFAQ Quality Assessment)
- Development of data models
- Special interest groups on regional standards/regulations
- Pooled research, development and analysis
- Industry-wide terminology sharing
- Facilitating professional certification



# Preliminary GALA Standards Initiative Budget



# Example

## Model Service Definitions



- Deliverables
  - Public definitions of key industry terms
  - Establishment of a group to define elementary and complex service profiles
  - Publication of service profiles on the GALA website and promotion of these resources to the public
  - Development and publication of model contracts, SLAs, RFPs, etc.
  - Creation of “Certified Localization Procurement Specialist” certificate program and tests
- Sample Benefits
  - Buyers of localization services
    - Greater clarity in bid/procurement processes
  - Language services companies
    - Standardized service descriptions prevent confusion
  - Localization tools developers
    - Standard process models simplify interfaces between workflow components
  - Individual translators
    - Improved business definitions will help translators clarify their offerings
  - Standards organizations
    - Consistent industry framework for standards projects



# Funding and benefits

## Funding

- Not pay-for-play – no membership requirement
- Three voluntary funding levels
  - Contributor (\$200/year)
  - Supporter (\$1,000/year)
  - Charter (\$5,000/year)
- Shared costs/shared benefits model

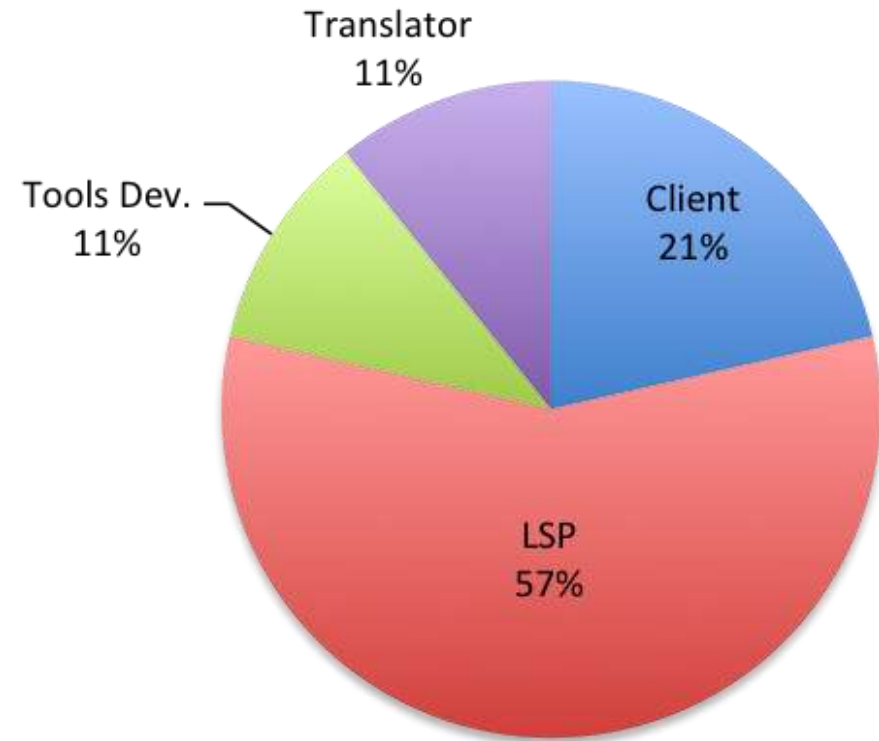
## Benefits

- Improved standards efficiency
- Common solutions to common problems
- Open platform for contribution
- Improved access to information and best practices



# Contributors

- Early contributions coming in
  - ADAPT
  - Amway
  - A-CLID
  - Clear Word Translations
  - The Level
  - TO-IN
  - Welocalize
- Approximately \$25K received or in the pipeline, with another \$45K requested
- Breakdown of pledges:





# Standards Initiative Collaboration



**Brigham Young University  
Translation Research Group**



More in the pipeline...



# Learn more

[\*www.gala-global.org/standards\*](http://www.gala-global.org/standards)

[\*standards@gala-global.org\*](mailto:standards@gala-global.org)