

**tcworld 2011 Conference**

**Workshop – LOC 24**

***“Interoperability Standards in the Localization Industry – Status Today and Opportunities for the Future”***

***Across Systems***

**Sukumar Munshi**

**Wiesbaden, 19th October 2011**



## Agenda



**11:15** Introduction

..

**11:30** Interoperability Standards in the Localization Industry

..

**12:00** Interoperability and Perspectives

..

**12:30** Options for the Future

..

**12:45** Summary

..

## Before we start ...

- > This is not a presentation
- > It's a workshop to share and exchange
- > Supportive slides for discussion
- > It will be as technical or as lively you make it
- > New content will be compiled and sent to all participants

## Your interests, perspectives & backgrounds

> Business

> Standards

> Processes

> Politics

> Other

> Management

> Projects

> Engineering

> LSP

> End user

> Other

**Share a challenge or area of interest from your perspective**

## Statements to verify

*“While the Localization Industry evolves, many views, perspectives, constraints and maturity issues are competing against each other when it comes to interoperability standards.”*

**What is interoperability?**

## Definitions of interoperability

### >IEEE Glossary:

the ability of two or more systems or components to exchange information and to use the information that has been exchanged.

### >WIKIPEDIA:

Interoperability is a property of a product or system, whose interfaces are completely understood, to work with other products or systems, present or future, without any restricted access or implementation.

**Why does Interoperability matter – what does it mean to you?**

## Agenda



<b>11:15</b>	<b>Introduction</b>
--------------	---------------------

..

<b>11:30</b>	<b>Interoperability Standards in the Localization Industry</b>
--------------	--

..

<b>12:00</b>	<b>Interoperability and Perspectives</b>
--------------	--

..

<b>12:30</b>	<b>Options for the Future</b>
--------------	-------------------------------

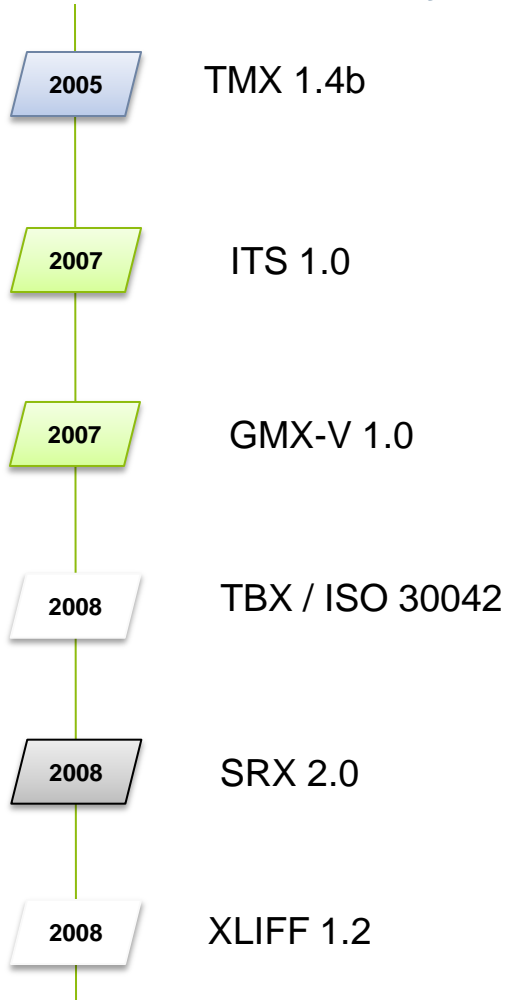
..

<b>12:45</b>	<b>Summary</b>
--------------	----------------

..

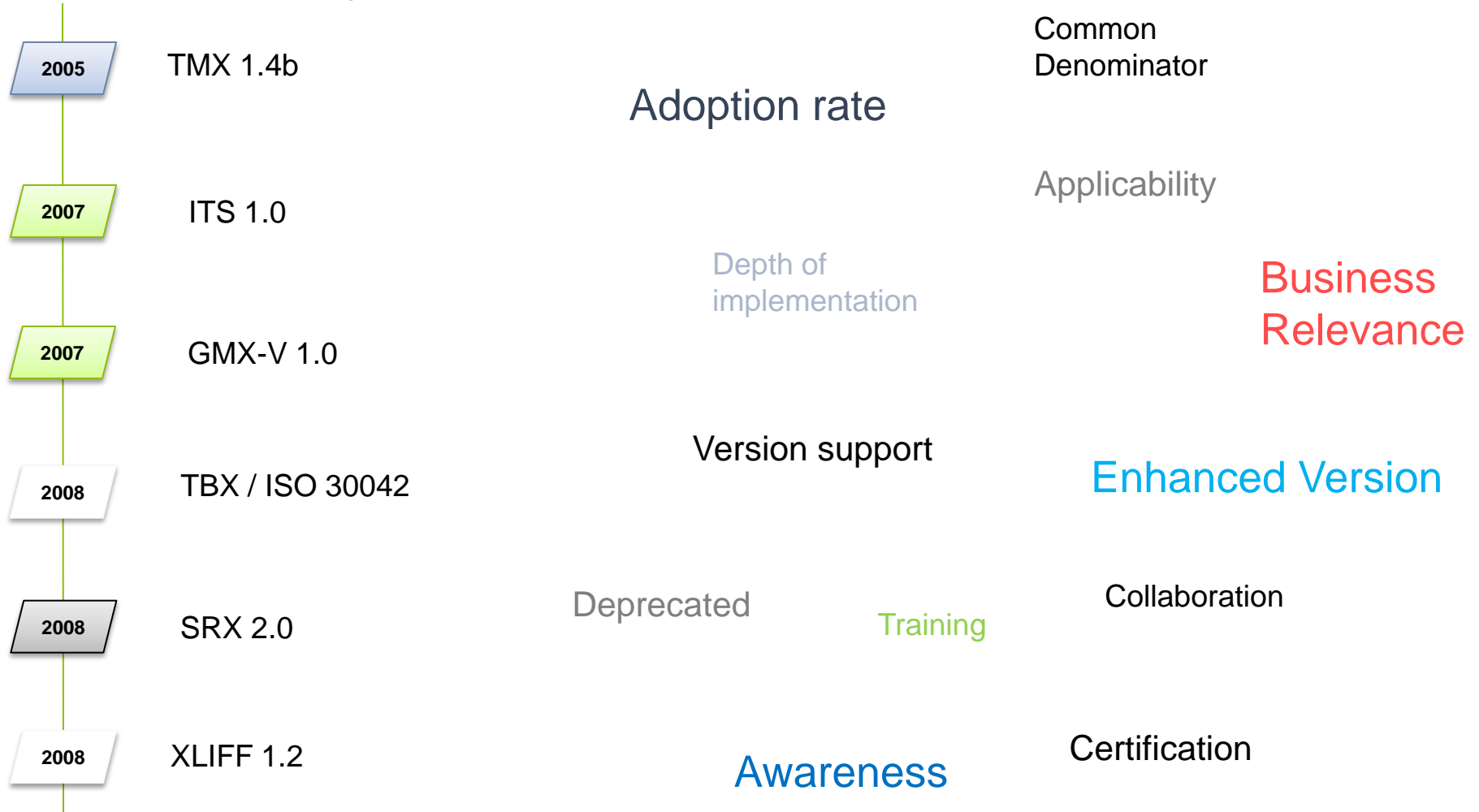
**What are interoperability standards in our industry?**

## Interoperability standards in localization



How would you rate these IO standards?

## Interoperability standards in localization



## Translation Memory eXchange (TMX)

<b>Origin</b>	LISA
<b>Maintenance</b>	Open
<b>Version</b>	TMX 1.4b / OSCAR Rec., 26 April 2005
<b>Applicability</b>	Most translation tools
<b>Technology</b>	XML
<b>Tools</b>	Most tools support TMX
<b>Benefits</b>	Exchange of translation memory data
<b>Caveats</b>	Segmentation issues. Quality of TM Data. Availability of attributes.
<b>Limitations</b>	Not all tools support same TMX Versions, show the same behavior.
<b>Rating</b>	High relevance to the industry.
<b>Conclusion</b>	Established standard that is widely supported.
<b>Ref</b>	<a href="http://www.lisa.org/tmx/tmx.htm">http://www.lisa.org/tmx/tmx.htm</a>

## TMX example

```
<xml version="1.0">  <!-- Example of TMX document -->
<tmx version="1.4">
  <header [attributes]> ... </header>
  <body>
    <tu tuid="0002" srclang="*all*">
      <prop type="Domain">Cooking</prop>
      <tuv xml:lang="FN">
        <seg>The menu</seg>
      </tuv>
      <tuv xml:lang="FR-CA">
        <seg>Le menu</seg>
      </tuv>
      <tuv xml:lang="DE-DE">
        <seg>Das Menü</seg>
      </tuv>
    </tu>x
  </body>
</tmx>
```

## Revisiting TMX

- > Not all the tools deal with TMX the same way.
- > Taking a closer look at TMX, most processes lead to losses in leveraging

Example: TMX 1.4b

- > Tool 1 export: information about the sequence of placeholders are lost
- > Tool 3 for backups and cleanups
- > Exchange with TMX
  - Tool 1 import: <ph><bpt><ept><ph> are consolidated
  - Tool 2 server converts <ph> to <bpt><ept> in some cases
  - Tool 2 has issues with hierarchical tag constructs
  - Tool 2 offers 2 variants for TMX
    - > One conforming to the standard TMX 1.4
    - > One for internal backups and processing, following TMX 1.3 including domains and additional proprietary metadata

Key statements for validation:

- > TMX is successful, because so many tools support it
- > It works fine with simple flat TMs
- > Standard has been constant over the last years

## Main standards: SRX

<b>Origin</b>	<b>LISA</b>
Maintenance	Open
Version	SRX 2.0 / April 2008.
Applicability	Defines segmentation rules, conjunction with the TMX Format
Technology	XML, ICU
Tools	SRX Editor, MemoQ, etc.
Benefits	Can improve leverage when associated with TMX import
Caveats	Not all tools support SRX
Limitations	ICU is not supported by all technologies
Rating	Segmentation rules can be set using localization tool functionality.
Conclusion	Unless TMX and SRX are not converging, lower enforcement
Ref	<a href="http://web.archive.org/web/20090523015600/http://www.lisa.org/Segmentation-Rules-e.40.0.html">http://web.archive.org/web/20090523015600/http://www.lisa.org/Segmentation-Rules-e.40.0.html</a>

## SRX Example

```
<xml version="1.0" encoding="UTF-8">
<srx xmlns="http://www.lisa.org/srx20" version="2.0">
  <header segmentsubflows="yes" cascade="no">
    <formathandle type="start" include="no"></formathandle>
    <formathandle type="end" include="yes"></formathandle>
    <formathandle type="isolated" include="no"></formathandle>
  </header>
  <body>
    <languagerules>
      <languagerule languagerulename="default">
        <rule break="no">
          <beforebreak>([A-Z]\.){2,}</beforebreak>
          <afterbreak>\s</afterbreak>
        </rule>
        <rule break="yes">
          <beforebreak>\.</beforebreak>
          <afterbreak>\s</afterbreak>
        </rule>
      </languagerule>
    </languagerules>
    <maprules>
      <languagemap languagepattern=".*" languagerulename="default"></languagemap>
    </maprules>
  </body>
</srx>
```

## TBX – TermBase eXchange

<b>Origin</b>	LISA (OSCAR)
<b>Maintenance</b>	ISO 30042; Framework: ISO 12620 (MARTIF), ISO 12200 and ISO 16642
<b>Version</b>	ISO 30042:2008
<b>Applicability</b>	designed to processes involving terminological data: analysis, representation, dissemination, and interchange. The primary purpose of TBX is for interchange of terminological data.
<b>Technology</b>	XML
<b>Tools</b>	Many tools support TBX
<b>Benefits</b>	Eased setup for Terminology Databases, lower migration cost
<b>Caveats</b>	Data quality, proper categories.
<b>Limitations / Rating</b>	Key IO standard
<b>Conclusion</b>	ISO backing. Widely supported standard.
<b>Ref</b>	<a href="http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?number=45797">http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?number=45797</a>

## TBX Example

```
<termEntry id="c180">
  <descrip type="subjectField">Restaurant Menu Item</descrip>
  <descrip type="conceptPosition" target="foodskos">s79</descrip>
  <descrip type="definition">the thigh of a frog leg (the foot has been removed)</descrip>
  <langSet xml:lang="en">
    <ntig id="t365">
      <termGrp>
        <term>frog legs</term>
      </termGrp>
    </ntig>
  </langSet>
  <langSet xml:lang="fr">
    <ntig id="t364">
      <termGrp>
        <term>cuisses de grenouilles</term>
      </termGrp>
    </ntig>
  </langSet>
</termEntry>
```

## TBX, tools and alternatives (LSP situation from practice)

### *Terminologie: TBX, MultiTerm XML, CSV*

Key statements for validation:

- > TBX not yet fully supported by major tools on the market. The main exchange format is often CSV.
  - Tool 1 (version 1) cannot import or export TBX, Tool 1 (version 2) can only import.
  - Tool 2 imports Tool 1's XML (incl. XDT), exports TBX, but does not export Tool 1's XML
  - Hence Interoperability between these tools requires CSV
  - Since many tools use CSV and offer comfortable features, motivation is not very big to move
  - End users prefer to use CSV since it's easier to manage than TBX
- > What is your experience with these formats?

## GMX-V (Global Information Management Metrics eXchange – Volume)

<b>Origin</b>	LISA (OSCAR)
<b>Maintenance</b>	Open
<b>Version</b>	GMX-V1.0 / 2007
<b>Applicability</b>	quantifying the workload for a given localization or translation task
<b>Technology</b>	XML (XLIFF)
<b>Tools</b>	Globalsight, Heartsome, etc.
<b>Benefits</b>	Information beyond Word Counts
<b>Caveats</b>	--
<b>Limitations / Rating</b>	Low adoption in the industry
<b>Conclusion</b>	Relevance yet to be proven
<b>Ref</b>	<a href="http://www.lisa.org/fileadmin/standards/GMX-V.html#SectionReferences">http://www.lisa.org/fileadmin/standards/GMX-V.html#SectionReferences</a>

## GMX-V Example

The following is an example of a GMX-V instance for a single file/resource:

```
<metrics:metrics version="1.0" source-language="en-GB" tool-name="XYZ Tool" tool-
version="1.23">
  <metrics:stage phase="initial" date="20041218T13:06:52Z">
    <metrics:notes from="auser@company.com">
      Initial count based on source document.
    </metrics:notes>
    <metrics:count-group name="non-verifiable">
      <metrics:count type="OC-TestingFiles" value="99"/>
      <metrics:count type="OC-DTPFiles" value="99"/>
      <metrics:count type="ScreenCount" value="99"/>
    </metrics:count-group>
    <metrics:count-group name="verifiable">
      <metrics:count type="TotalWordCount" value="99"/>
      <metrics:count type="TotalCharacterCount" value="99"/>
      <metrics:count type="TranslatableLinkingInlineCount" value="99"/>
    </metrics:count-group>
  </metrics:stage>
</metrics:metrics>
```

## ITS Internationalization Tag Set

<b>Origin</b>	W3C
<b>Maintenance</b>	<a href="http://www.w3.org/International/its/ig/">http://www.w3.org/International/its/ig/</a>
<b>Version</b>	ITS 1.0 / W3C Recommendation 03 April 2007
<b>Main features</b>	XML Vocabulary addressing topics like directionality, selective translation, language association, inline text behavior, terminology tagging, notes.
<b>Applicability</b>	XML, XLIFF, Web
<b>Technology</b>	XML
<b>Tool support</b>	ITS2XLIFF, Spritser, Rainbow, XTM, others
<b>Benefits</b>	Support for localization issues with XML based content, covers main localization issues
<b>Caveats / Limitations</b>	Little enforcement, not widespread use. Depends on how XML is used
<b>Rating</b>	Good standard with little application. Commonalities to XLIFF.
<b>Conclusion</b>	More „evangelization is needed“ to promote the standard.
<b>Ref</b>	<a href="http://www.w3.org/International/its/ig/">http://www.w3.org/International/its/ig/</a>

## ITS example

```
<text xmlns:its="http://www.w3.org/2005/11/its">
  <head>
    <revision>Sep-10-2006 v5</revision>
    <author>Ealasaidh Mclan</author>
    <contact>ealasaidh@hogw.ac.uk</contact>
    <title
      its:translate="yes">The Origins of Modern Novel</title>
    <its:rules version="1.0">
      <its:translateRule translate="no" selector="/text/head"/>
    </its:rules>
  </head>
  <body>
    <div xml:id="intro">
      <head>Introduction</head>
      <p>It would certainly be quite a <span its:translate="no">faux
pas</span> to start a dissertation on the origin of modern novel without
mentioning the <tl>Epic of Gilgamesh</tl>...</p>
    </div>
  </body>
</text>
```

## Standards: XLIFF

<b>Origin</b>	<b>IBM, Novell, Berlitz, ...,</b>
Maintenance	OASIS XLIFF TC
Version	XLIFF Version 1.2 / 02.2008
Main features	interoperability between tools, one format for all. Interchange of data and metadata. Inline tag, segmentation support, alternate translation, references to glossaries
Applicability	Proprietary formats localization, storage of localization data. Usage as data format for localization, as translation editor format
Technology	XML
Tool support	Many tools support XLIFF, but at different levels
Benefits	Standardization. XML-Automation.
Caveats / Limitations	No enforcement on application. Many ways of implementation. Loss of semantic data. XLIFF is still unknown. Some limitations, tool support for XLIFF (picking preferred parts), broad usage, low focus
Rating	Increased application, enhancing collaboration
Conclusion	High relevance for the industry – but still some consolidation needed
Ref	<a href="http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xliff">www.oasis-open.org/committees/tc_home.php?wg_abbrev=xliff</a>

## XLIFF example

Example of a "minimalist" representation (un-mapped parts are stored in the Skeleton file):

```
<xliff xmlns='urn:oasis:names:tc:xliff:document:1.2'
  xmlns:xsi='http://www.w3.org/2001/XMLSchema-instance'
  xsi:schemaLocation='urn:oasis:names:tc:xliff:document:1.2 xliff-core-1.2-transitional.xsd'
  version='1.2'>
<file original='sample.htm' source-language='en' datatype='html'>
  <header><skl><external-file href='the_skeleton_file.skl' /></skl></header>
  <body>
    <trans-unit id='1' restype='x-html-h1'>
      <source xml:lang='en'>Report</source>
    </trans-unit>
    <trans-unit id='2' restype='cell'>
      <source xml:lang='en'>Text in cell r1-c1</source>
    </trans-unit>
    <trans-unit id='3' restype='cell'>
      <source xml:lang='en'>Text in cell r1-c2</source>
    </trans-unit>
    <trans-unit id='4' restype='cell'>
      <source xml:lang='en'>Text in cell r2-c1</source>
    </trans-unit>
    <trans-unit id='5' restype='cell'>
      <source xml:lang='en'>Text in cell r2-c2</source>
    </trans-unit>
    <trans-unit id='6' restype='x-html-p'>
      <source xml:lang='en'>All rights reserved (c) Gandalf Inc.</source>
    </trans-unit>
  </body>
</file>
</xliff>
```

## XLIFF revisited

Key statements for validation:

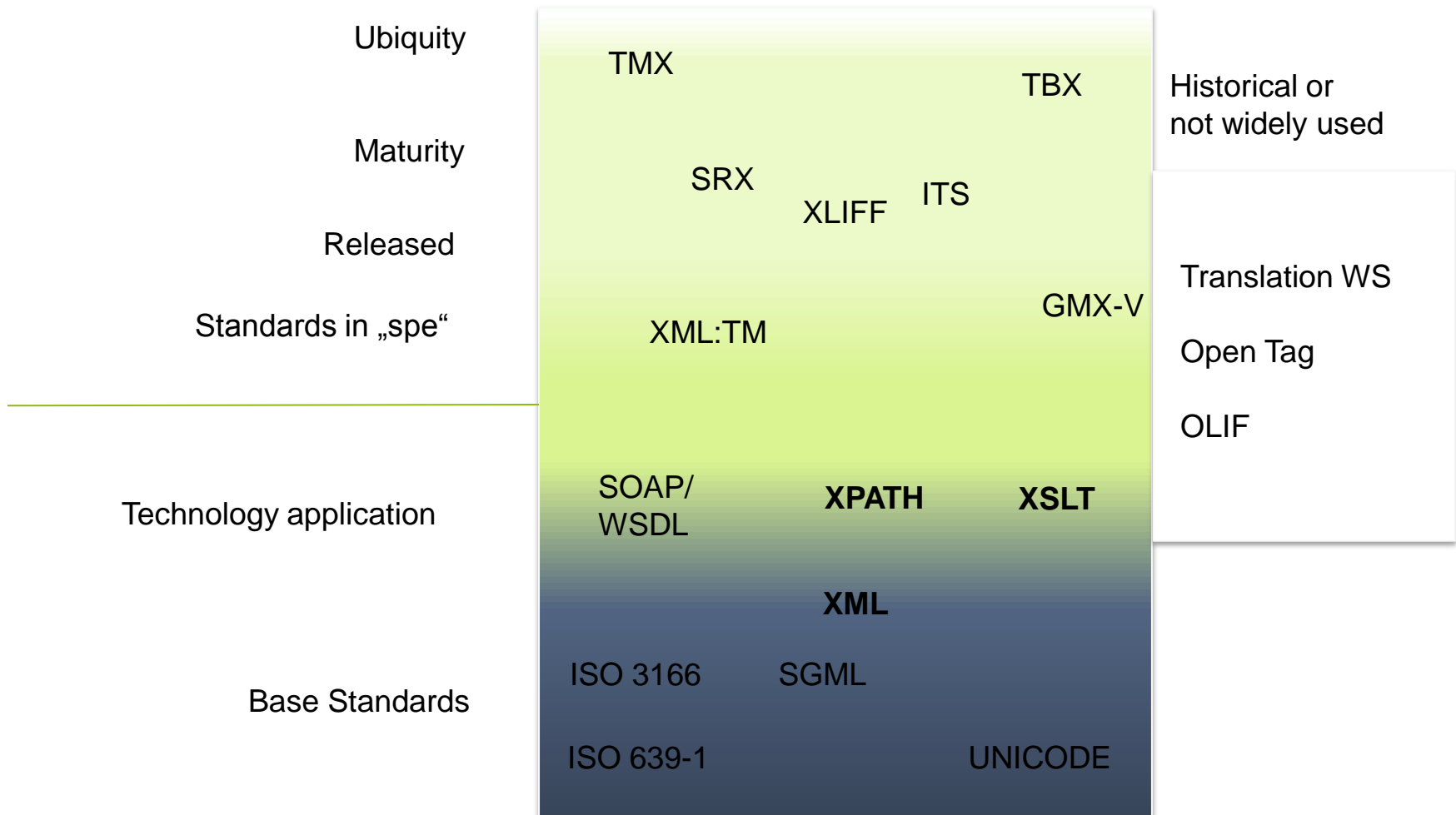
- > Leading software industry publishers still prefer using **proprietary file formats** rather than the XLIFF Standard. Reasons for this are
  - Lack of awareness about XLIFF standard
  - Limitations and issues of the standard
  - Improper implementation
  - Lack of tool support
  
- > Many XLIFF features are not supported
  - Business case requirements lead to partial implementation
  - Complexity, limitations of the Standard
  - Some features are difficult to implement
  - “I am doing my own XLIFF Syndrome”
  - No defined criteria
  
- > Formal XLIFF compliance is easy
  - Generate XLIFF files that conforms to XLIFF specification.
  - Custom Extensions, Custom Attribute Values are allowed
  - XLIFF compliance does not mean all features are implemented
  - XLIFF TC is working on conformance criteria for the standards

LOC 17: **Applied XLIFF**  
Bryan Schnabel, Thursday  
20th October – Lecture room  
12B, 9:45

## What could be the wider picture of interoperability standards?

- > Past
- > Precursors
- > Base technologies

## From standard to maturity – an attempt categorization



## **Interdependency of standards – Is this relevant at all?**

> What do you think of TMX and SRX?

## Statements to verify

*“While the Localization Industry evolves, many views, perspectives, constraints and maturity issues are competing against each other when it comes to interoperability standards.”*

*„Interoperability is real when tools can seamlessly process files and interpret their content with no significant data loss.“*

## Points for discussion (1)

- ▶ How to distinguish trends and a standard?
- ▶ How important is interoperability to your business?
- ▶ Are current standards covering the needs?
- ▶ What are requirements for implementation for standards?

## Agenda

**11:15 Introduction**

..

**11:30 Interoperability Standards in the Localization Industry**

..

**12:00 Interoperability and Perspectives**

..

**12:30 Options for the Future**

..

**12:45 Summary**

..

## **IO standards in the perspective of the different users...**

> Who cares ...

...Who should?

And why?

## Interoperability – who cares / who should

- > Customers
  - More flexibility onboarding new solutions
  - Freedom of choice
- > Language Service Providers
  - Securing investments in current software
  - Less pre- /post processing
  - Flexibility with suppliers
- ▶ Software Vendors
  - ▶ Create less parsers
  - ▶ Concentrate on the core product
- ▶ Organizations
  - ▶ Dispersed activities
  - ▶ Interoperability matters to the industry
- ▶ Governmental institutions
  - ▶ Economical concerns
  - ▶ EU Case

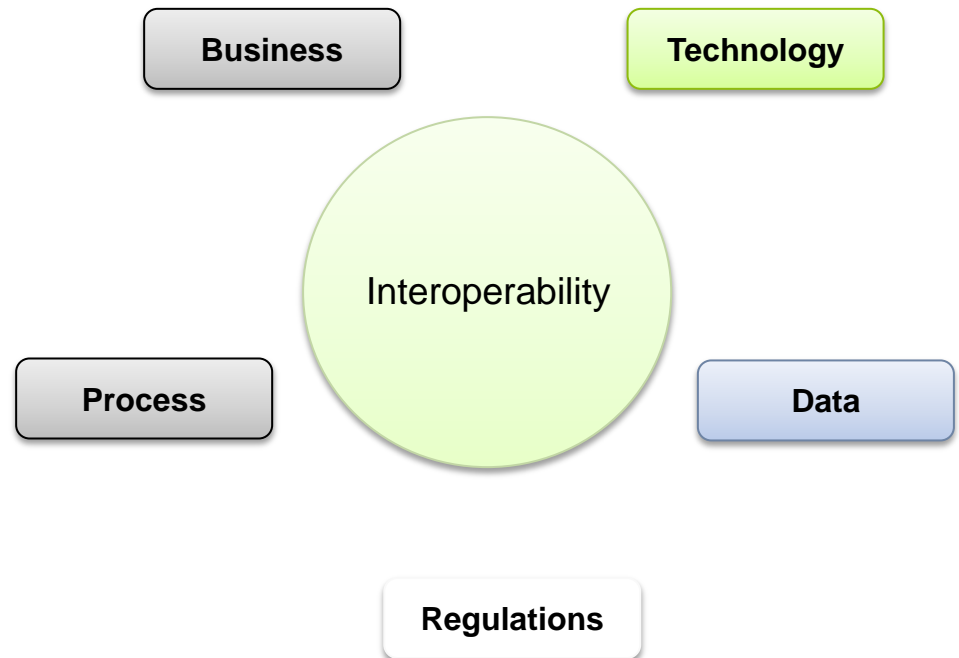
## Points for discussion (2)

- ▶ What is the current consensus regarding interoperability?
- ▶ What are the different positions to be reconciled?
- ▶ How should the different parties deal with the topic?

**What does it take to make an IOP Standard successful?**

## Aspects of Interoperability

- > Data management
  - Access (barrier free)
  - Readability
- > Technology usage
  - No impact on tools
- > Process Benefits
  - Requirements
  - Transaction cost
- > Business purpose
  - Clear value add
  - Substitute process optimization with productive tasks
- > Regulatory aspects
  - Promotion of standards
  - Macro economy



## **Interoperability in the real world ... does it work out for you?**

> What are the issues you encountered?

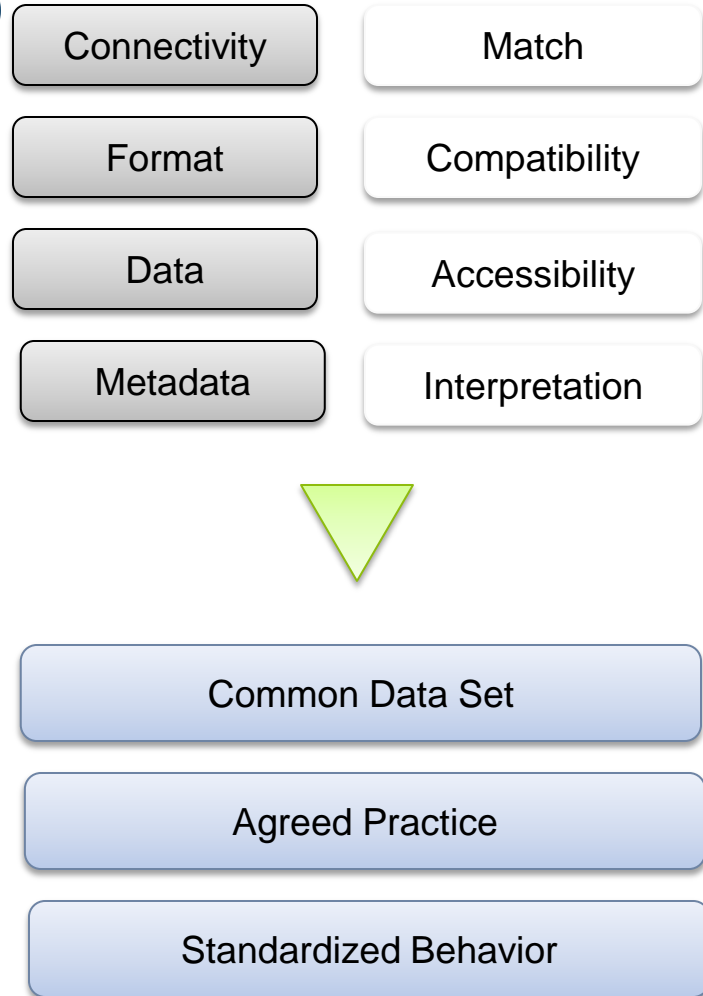
## Interoperability issues (XLIFF or other)

- > Data set divergence
- > Process and practice diversity
- > Implementation depth and quality
- > Tool to tool interoperability
- > Process relevance

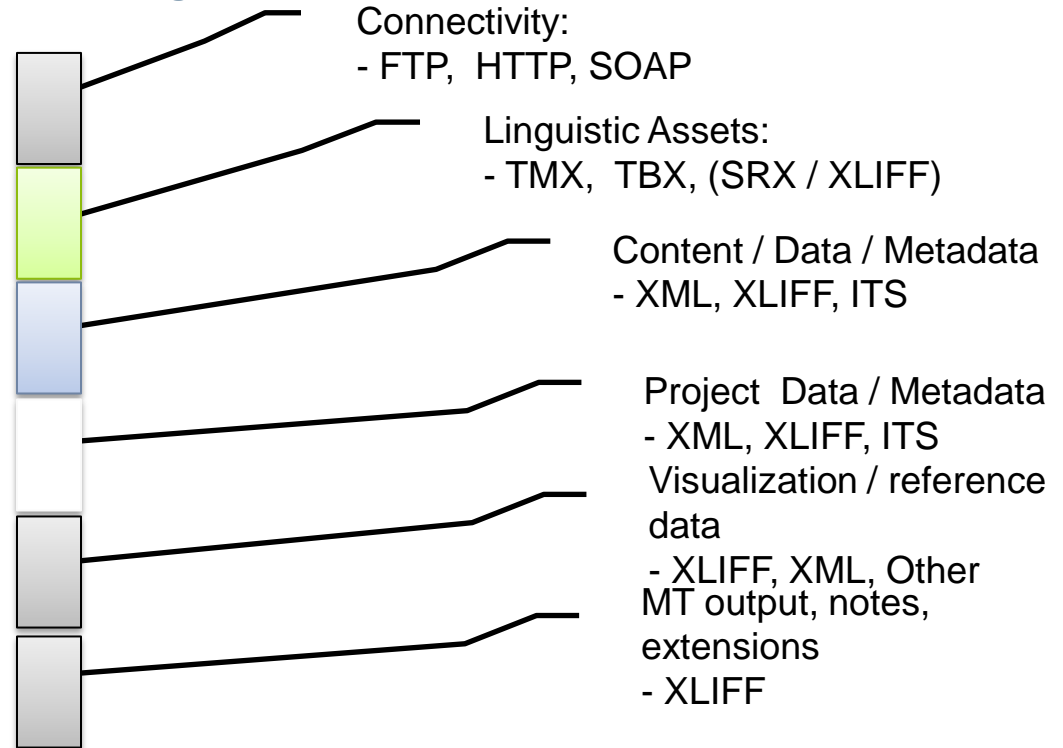
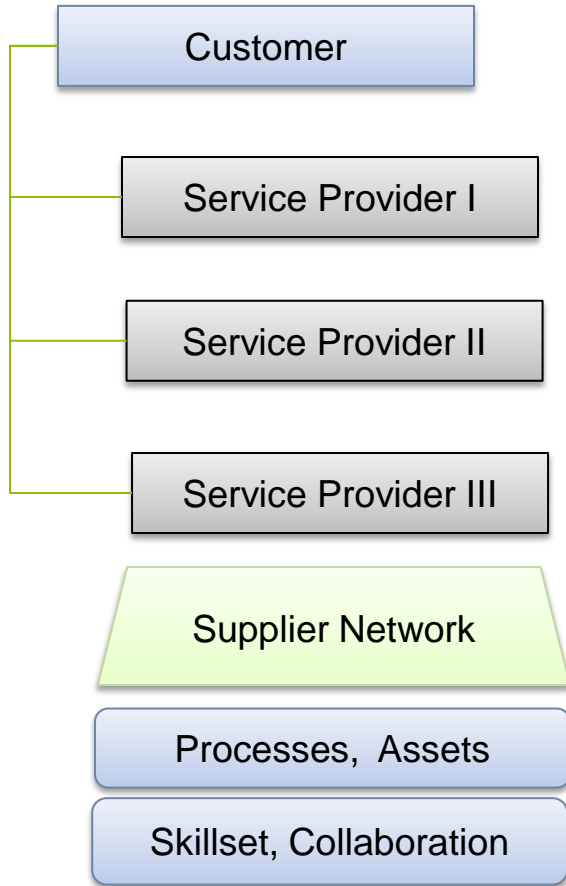
*With these issues in mind, what are the key things that you would expect interoperability does?*

## Interoperability issues (XLIFF or Other)

- > Data set divergence
- > Process and practice diversity
- > Implementation depth and quality
- > Tool to tool interoperability
- > Process relevance



## Interoperability: relevant technologies



## Statements to verify

*“While the Localization Industry evolves, many views, perspectives, constraints and maturity issues are competing against each other when it comes to interoperability standards.”*

*„Interoperability is real when tools can seamlessly process files and interpret their content with no significant data loss.“*

*“Interoperability in a wider sense means working together in a more efficient manner with support of standard technologies.”*

## The benefits and cost of interoperability

### Benefits

- > Prevent being locked-in
- > Repurpose data
- > Avoid costly pre-post processing
- > Concentrate on core task
- > Improve process efficiency

### Caveats

- ▶ Invest in precursors of standards
- ▶ Mix innovation with standard application
- ▶ Not invest at all
- ▶ Standardize and inflate, while disconnecting from the beneficiaries
- ▶ Make a poor implementation

## Examples from other industries (1/3)

- > DOCSIS – Data over Cable Service Interface Specification
  - > EDI - Electronic Data Interchange
  - > HL7 – Health Level 7
- ▶ Cable providers felt locked-in to proprietary technology
  - ▶ Motivation to promote interoperability standards
  - ▶ Price of \$500 per modem in 1994
  - ▶ Introduction of DOCSIS by CableLabs – Certification Program
  - ▶ In 2005 Price fell \$40

## Examples from other industries (2/3)

- ▶ One of the NIST studies estimates that lack of interoperability led to a cost of US\$ 15.8 billion to the US governmental infrastructure capital investments, i.e. on all governmental building and construction infrastructures.
- ▶ Excess of that amount in public money and therefore taxpayers' contributions (Gallagher et al, 2004).

## Examples from other industries (3/3)

- > (Brunnermeier et al, 1999 and White et al, 2004) - estimates of cost for the US economy, arising from lack of interoperability:
  - US\$1 billion/year: engineering data in automotive
  - US\$5 billion/year: all supply chain data in automotive industry
  - US\$3.9 billion/year: all supply chain data in electronics industry.

## Statements to verify

*“While the Localization Industry evolves, many views, perspectives, constraints and maturity issues are competing against each other when it comes to interoperability standards.”*

*„Interoperability is real when tools can seamlessly process files and interpret their content with no significant data loss.“*

*“Interoperability in a wider sense means working together in a more efficient manner with support of standard technologies.”*

***„The wealth interoperability can bring to localization remains an untapped potential, as many factors of a successful standard introduction are not taken into account.“***

### Points for discussion (3)

- ▶ What is the current consensus regarding interoperability?
- ▶ What are requirements for implementation for standards?
- ▶ What should the industry do to become more interoperable?
- ▶ What could be a good roadmap for interoperability standards moving further

## Agenda

**11:15 Introduction**

..

**11:30 Interoperability Standards in the Localization Industry**

..

**12:00 Interoperability and Perspectives**

..

**12:30 Options for the Future**

..

**12:45 Summary**

..

## Organizations and interest in interoperability

- > ISO TC 37
- > ETSI
- > W3C
- > OASIS
- > GALA
- > TAUS

- > EU
- > Unicode Consortium
- > OMG
- > SAE
- > SIG

- > [LISA]

### The Gala Standards Initiative

- > The industry does not suffer from lack of standards, more from lack of awareness
- > Standards may not be in sync with evolving requirements
- > Standards relevant to the industry are developed by many bodies
- > Lack of training, promotion and best practices about localization standards
- > Initiative to collaborate and facilitate work on standards, bring together disparate constituencies that are impacted by standards

## Recent initiatives - moving towards „business interoperability“

### > Linport

Vendor-independent format to package translation materials (specification)

<http://www.linport.org>

### > Interoperability now

Initiative fostering interoperability and the use of open standards (1. Memorandum IO)

<http://interoperability-now.org/>

### > Model Service Elements

- Terminology
- Defining process elements
- „Not the how, but the what“
  
- Allows to build models based on activities
- Decrease complexity of translation and localization task
- Help with dysfunctional procurement processes, improve collaboration / transparency
- Educational aspects

<http://www.gala-global.org/>

## The future of Interoperability

By 2015 ...

- >The pressure on interoperability has increased due to cloud based processing.
- >The community of organizations, customers, suppliers and tool vendors will fully endorse supporting standards
- >All interested parties will have agreed on how IO standards should be applied
- >Tools will support the basic concepts of all current IO standards
- >Added value will be on performance, throughput and solutions for specific applications of the standards



## Agenda

**11:15 Introduction**

..

**11:30 Interoperability Standards in the Localization Industry**

..

**12:00 Interoperability and Perspectives**

..

**12:30 Options for the Future**

..

**12:45 Summary**

..



## Conclusions

- > A released specification does not make a standard successful.
- > Standard generation and adoption is an immature process in the localization industry (or it just needs time).
- > A successful standard is widely used in the industry and delivers benefits to all the interest groups involved.
- > As the localization industry is progressing, pressure increases, the visibility of challenges in the industry and the need for standards is growing.

## Q & A



## References

- > GALA Standards Initiative Plan – Public Draft, 1.2 – A. Lommel, H. Fenstermacher, S. Gladkoff
- > LCX and XLIFF: a Comparison- Asanka Wasala Dag Schmidtke  
Reinhard Schäler - 1st XLIFF International Symposium –  
September 2010
- > Thomas Imhof – XLIFF a Bilingual Interchange Format - May,  
2010
- > Pearson White Paper - Next-Generation Assessment  
Interoperability Standards - November 2010

# Agenda

**11:15 Introduction**

..

**11:30 Interoperability Standards in the Localization Industry**

..

**12:00 Interoperability and Perspectives**

..

**12:30 Options for the Future**

..

**12:45 Summary**

..

