Status Update on the TAPICC Initiative

David Filip
Klaus Fleischmann
(Allison Ferch,
Jim Compton,
Serge Gladkoff)
Agenda

Why TAPICC?

Status Track 1

Launch Track 2
What is GALA?

• Global non-profit trade association with membership of 400 companies (LSPs, tech developers, buyers of translation)

• Non-biased platform for information-sharing and collaboration, training and professional development

• Many organizational partnerships, including TAUS and LTI
## Multinational Translation Companies (50% of Common Sense Advisory “Top 100”)

<table>
<thead>
<tr>
<th>Lionbridge</th>
<th>TRANSPERFECT</th>
<th>LLs™</th>
<th>SDLQ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>welocalize</td>
<td>RWS Moravia</td>
<td>CYRACOM</td>
<td>AMPLEXOR</td>
<td>Semantix</td>
</tr>
<tr>
<td>RR DONELLEY</td>
<td>pactera</td>
<td>CAPITA</td>
<td>CSOFT</td>
<td>VERZTEC</td>
</tr>
<tr>
<td>CERTIFIED LANGUAGES INTERNATIONAL</td>
<td>Akorbi</td>
<td>SeproTec</td>
<td>CONCORDE</td>
<td>xplanation</td>
</tr>
</tbody>
</table>
### Translation Technology Developers

| Across | KantanMT.com | Kilgray | Lucy Software
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Matecat</td>
<td>Memsource</td>
<td>Plunet</td>
<td>SDL</td>
</tr>
<tr>
<td>Smartling</td>
<td>Wordbee</td>
<td>XTTRF</td>
<td>XTM Cloud</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Translation Technology for a Globalized World.*
• Represent a CMS, DAM or similar tech?
• Represent a TMS, CAT tool or similar tech?
• Create / manage content that must get localized?
• Responsible for localizing / translating content?
• Any system integrators?
Why TAPICC?
The Current API Wild West

[Diagram showing various API providers and their interconnections]

www.gala-global.org/tapicc
The Current API Wild West

- Unnecessary variation
- Continuous reinvention of the wheel
- Wasted $$!
  - For clients
  - For LSPs
  - For tools vendors
- Loss of operational freedom

Can be a “deal breaker” for making content available worldwide!
The TAPICC Vision

- Common Integration model
- Universal Translation API
- Best practices
- Use cases
TAPICC stands for: “Translation API Cases & Classes.”

It is:
• an open, volunteer-driven pre-standardization initiative.
• focused on use cases based on translation and language work.

It isn’t:
• an attempt to replicate the functionality of CMSes or TMSes.
• an attempt to dictate the actual flow of work through a supply chain or how work gets executed.
Features & Benefits

- Agreed upon metadata, use cases, best practices, classes
- Go-to place for information and education for ALL stakeholders
- Quickly implementable classes and use cases
- Reduce cost of integration
- Quickly onboard new clients, systems, LSPs
- Easily embed L10N in content processes and enterprises
What we have done so far
### Foundation

<table>
<thead>
<tr>
<th>Legal Framework (Open Source)</th>
<th>Community Engagement</th>
<th>Organizational Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The 3-Clause BSD License (BSD-3 Clause)</td>
<td>• GALA forum</td>
<td>• Project charter</td>
</tr>
<tr>
<td>• Creative Commons Legal Code (CC-BY 2.0)</td>
<td>• Working groups</td>
<td>• Working Group Playbook</td>
</tr>
<tr>
<td></td>
<td>• GitHub wiki</td>
<td>• Numerous presentations and marketing collateral</td>
</tr>
<tr>
<td></td>
<td>• Wide representation from the language industry</td>
<td></td>
</tr>
</tbody>
</table>
Initiative is Organized into Four Tracks

1. Business metadata for supply chain automation
   - Harmonize existing business metadata models

2. Exchange on unit level
   - Pass a segment/unit from an editor to a TM/MT or other tool

3. Semantic enrichment of units
   - Terminology, TM, MT, layout for “good enough”

4. Layout representation level
   - Support process with visual context

We’ve been working on this...

...and are ramping-up with this.

Don’t confuse the tracks with the working groups. The working groups are there to divide-up the work within a track.
Track 1
Supply Chain Automation
One of the principles of TAPICC is to leverage from relevant existing standards and standardization efforts, including (especially) XLIFF.
WG1 Complete Set of Task Types

- The final deliverable will include the parameters for each of these task types

- Localization
- Internationalization
- Scoping
- Translation
- Editing
- Checking
- DTP
- Terminology work
- Transcription
- Content Creation
- Content Preparation
WG1 Assemble and define tasks
Commenting on WG1 Results

• Elementary Linguistic Task Types
  https://docs.google.com/document/d/1POWcoIrKvUQpgq_lmgcpK_p6PZ35k7g7mieKARGgR4/edit?usp=drive_web&ouid=113136557825967529407

• Glossary of terms for the L10N industry
  https://docs.google.com/document/d/1OvYksscrYGtz7gUK-25AR28jobRN7gtkNtu1W3rvgaA/edit
WG2 Consensus

• Prose deliverable:
  • Primary goal of TAPICC is translation round trip. Other task are considered only as far as they support achieving the primary goal.
  • Payload is driven by task type.
  • If task type involves bitext, XLIFF2 is mandated. Legacy bitext is considered as source format and will be extracted into XLIFF2.

• Details around reference data

• Evaluating similarities and aligning with the Language Interoperability Portfolio (Linport) initiative / ASTM
WG3 Output

• WG3 focuses on XLIFF extraction from native format and providing buyers with guidance on how to create XLIFFs that simplify satisfying their needs.

• Examples of problematic approaches and suggested better alternatives can be found in https://galaglobal.github.io/TAPICC/

• The release of this specific output „XLIFF 2 Extraction and Merging Best Practice“ version 1.0 is here: https://galaglobal.github.io/TAPICC/T1/WG3/rs01/XLIFF-EM-BP-V1.0-rs01.xhtml
WG4 Model in a nutshell

- **TAPICC Exchange types:**

  Client or host can be *any type of system*: CMS, TMS, etc.
The nuance of the data model is currently being debated / refined. But here’s the gist: supply chain interactions are modelled by bodies of work (jobs, tasks) that are associated with different types of content (assets, deliverables, and reference material) being exchanged between any two parties in the chain.
Available as a Swagger document
What we need from you

• Feedback on the deliverables
  • WG1 Documents through Connect / SmartSheet
  • WG2-4 through GitHub

• Joining does NOT require GALA membership, but it DOES require setting up an account and accepting the legal agreement

• Notes:
  • Conference participants already have an account
  • If you are in an organization, you probably need to talk to your legal department
  • https://www.gala-global.org/tapicc-legal-agreement
How to join

• Set up an account
  • [https://www.gala-global.org/user/register](https://www.gala-global.org/user/register)
  • Register as a new member

• Join the Connect Group for TAPICC
  • [https://www.gala-global.org/community-groups](https://www.gala-global.org/community-groups)
  • Search for „TAPICC“
  • Select the group
  • Click on „Join this group“
  • Click through and accept the legal.
Track Two:
Unit-level Exchange
Track Two Overview

Exchange on unit level
• Pass a segment/unit from an editor to a TM/MT or other tool

• T2 addresses **instantaneous fragment level exchange** between agents operating on bitext units such as TMS, MT Broker MT Engine, Translation editor etc.

• JRARTEBU!

• Track 2 WG1 Connect Group: [https://www.gala-global.org/tapicc-t2wg1-jliff-rest-api](https://www.gala-global.org/tapicc-t2wg1-jliff-rest-api)
T2 as opposed to T1

• T1 scope is **Supply Chain Automation**, job level exchange between
  - **Organizations** such as customer, MLV, SLV, MT Provider, Tool Provider, Freelancer, etc. and
  - **Agents** such as CMS, TMS, MT Broker, MT Engine, Translation Editor etc.

• T1 transactions are always **asynchronous** and exchanging the whole bulk of the job.

• T2 addresses instantaneous fragment level exchange between
  - **Agents** operating on **bitext units** such as TMS, MT Broker MT Engine, Translation editor etc.
  - In real time, as **synchronous** transactions
T2 Use Cases

TMS - CAT - TMS - TMS - TMS - MT
Other Resources for Track 2

- OASIS XLIFF OMOS TC

- OASIS XLIFF OMOS TC - JLIFF Repository on GitHub
  https://github.com/oasis-tcs/xliff-omos-jliff

- Vistatec’s Phil Ritchie – JLIFF C# library (MIT License)
  https://github.com/vistatec/JliffGraphTools
Conclusion / Takeaways

• Track 1 is ready for
  • Feedback
  • Input
  • Help with finalizing the API
  • Reference implementations

• Track 1 goal is to be finished by GALA Munich

• Track 2 is looking for volunteers
  • Face2face meeting of Track 2 team at GALA Munich preconference day