Serge Gladkoff graduated from an elite Russian nuclear science college, after which he was engaged in the localization industry for more than 25 years, first as a deputy director for a software distribution company (Dialogue-MEPHI) and then as a localization manager for the Moscow office of Borland International. For the past 20 years he has been President and co-owner of Logrus International Corporation, and he is now CEO of Logrus Global LLC.

Serge has been a speaker and presenter at many events, and served as an elected GALA Board member from 2009-2012 and again from 2014-2015. Serge is the founder of the largest professional online community in the localization industry, the Localization Professional group on LinkedIn.
Stop Reinventing the Wheel! The TAPICC Pre-Standardization Initiative for Translation APIs

If you have ever lamented the time and money you spent on building yet another connector for some part of the translation technology stack, this session is for you.

The Translation API Cases and Classes (TAPICC) initiative officially launched only seven months ago, yet significant progress has been made by the four Track 1 working groups, which have been focused on supply chain automation. The second track, or phase, is about to begin.

This session describes the achievements, challenges, and the work ahead for this collaborative, community-driven, open-source initiative to advance API standards for multilingual content delivery.
Status Update on the TAPICC Initiative

Laura Brandon
David Filip
Klaus Fleischmann
Serge Gladkoff
Agenda

Why TAPICC?
So far
Status
Specific Asks
Use Case Collector
Announcement Track 2

Session 1 – 45 min
Session 2 – 45 min
Why TAPICC?
The Current API Wild West

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

My Content

My Tool

[Logos and Emblems]
Partnerships and Complementary Initiatives

(COTI specification)
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

Legal Framework (Open Source)
- The 3-Clause BSD License (BSD-3 Clause)
- Creative Commons Legal Code (CC-BY 2.0)

Community Engagement
- GALA forum
- Working groups
- GitHub wiki
- Wide representation from the language industry

Organizational Documents
- Project charter
- Working Group Playbook
- Numerous presentations and marketing collateral
Four Tracks

1. Supply chain automation
   Transporting a translation “job” through the people & systems in the supply chain

2. Exchange on unit level
   Real-time interaction between different systems, for example between 2 TMS tools

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

4. Layout representation level
   Support process with visual context

We are here
## Track 1: Four Working Groups

<table>
<thead>
<tr>
<th>Track</th>
<th>Description</th>
</tr>
</thead>
</table>
| 1. Business Metadata | • Define business-level metadata  
• Canonical names and values  
• Workflow relevance  
• Harmonization  
• Compatibility |
| 2. Payload Specification | • Define types of payload  
• Define payload-level metadata  
• Canonical names and values  
• Harmonization  
• Compatibility |
| 3. XLIFF Extraction | • Extraction guidance  
• ITS decoration  
• XLIFF Extractors/Mergers  
• XLIFF OM representations, e.g. JLIFF |
| 4. API Specification | • Protocols  
• Transfer, pull and/or push  
• Define classes  
• Code samples/  
• Snippets  
• Specify API |

Each WG has co-chairs and 20+ participants. Join WG groups on GALA’s website: www.gala-global.org/tapicc
Overview of TAPICC Deliverables so far

**WG 1**
- Business Glossary of Translation Industry
- Task types plus parameters

**WG 2**
- Potential candidates for payloads

**WG 3**
- List of XLIFF extraction and merging Best Practices

**WG 4**
- Data model and RESTful API data model draft by WG4

Use Case Collection
Current Status / Workshop Results
Workshop Results

WG2 Consensus
- Payload is driven by task type
- If task type involves BITEXT, XLIFF2 is required
- Primary goal of TAPICC is translation roundtrip
- Other task types only support primary one
- Terminology options:
  1. Use in metadata
  2. Tag in payload
  3. Glossary model in XLIFF2
  4. Precedence in metadata

T1: 1. Metadata
    2. Payload
    3. XLIFF
    4. API
WG1 Results of Workshop

- Decision to include service tasks
- Cleared a lot of questions on the lists of tasks and combined some, more logical groupings
- Established that some tasks are high level and have no actual data (compound tasks)
- Others are very details (atomic), some are even both
- With this model, ANY workflow can be built
- The API will be built to support this complete list of task types with parameters
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

Localization (project and defined project plan exist)

Content preparation
Translation with Parameters Level of detail Resources etc.
Translation
Languages Resources etc.
Process

Editing
Linguistic / Technical Monolingual / Bilingual Etc.
Etc

www.gala-global.org/tapicc
WG2 Consensus

- Primary goal of TAPICC is translation round trip. Other tasks 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.

- Reference data:
  - Options for passing terminology references are:
    1: URI in Metadata
    2: Glossary module in XLIFF 2
    TBX in payload
    (precedence is established in the metadata)
  - This will be generalized to support other reference types
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://github.com/GALAglobal/TAPICC/tree/master/extraction_examples
TAPICC Working Group 4: API Specification

What have we done?
• created a draft REST-based architecture
• reviewed against use cases
• started recruiting for pilot

What is left to do?
• incorporate outputs from WG1
• run pilots
• collect feedback and make adjustments
TAPICC Working Group 4:
API Specification

Guiding principles:

• Keep it simple.
• Be harmonious with existing approaches.

Here’s what we’ve come up with....
Model in a nutshell

- **TAPICC Exchange types:**

  ![Diagram of TAPICC exchange types]

  Client or host can be *any type of system*: CMS, TMS, etc.
Jobs contain Tasks and Assets, and eventually Deliverables.
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
  • Register as a new member

• Join the Connect Group for TAPICCC
  • https://www.gala-global.org/community-groups
  • Search for „TAPICCC“
  • Select the group
  • Click on „Join this group“
  • Click through and accept the legal.
Your Feedback And Input
Use Case Collector
TAPICC T[rack]2 Announcement

David Filip on behalf of the TAPICC SC
Agenda of the Announcement

• What is “bitext” and what is “unit”?
• Difference between T1 and T2
• T2 use cases
• T2 next steps
Tyger Tyger, burning bright,
Tygře, tygře, ohnivou
In the forests of the night;
září svítíš lesní tmou!
What immortal hand or eye,
Kdo ten nesmrtelný byl,
Could frame thy fearful symmetry?
že z ní tvůj souměr sestrojil?

William Blake / Jaroslav Skalický
Bitext

Tyger Tyger, burning bright,
In the forests of the night;
What immortal hand or eye,
Could frame thy fearful symmetry?

William Blake

Tygře, tygře, ohnivou
září svítíš lesní tmou!
Kdo ten nesmrtelný byl,
že z ní tvůj soumér sestrojil?

Jaroslav Skalický
Bitext

<source>Tyger Tyger, burning bright, </source>
<target>Tygře, tygře, ohnivou </target>

<source>In the forests of the night; </source>
<target>září svítíš lesní tmou! </target>

<source>What immortal hand or eye, </source>
<target>Kdo ten nesmrtelný byl, </target>

<source>Could frame thy fearful symmetry? </source>
<target>že z ní tvůj souměr sestrojil? </target>
Bitext

msgid "Tyger Tyger, burning bright, "
msgid "In the forests of the night; "
msgid "What immortal hand or eye, "
msgid "Could frame thy fearful symmetry? "

msgstr "Tygře, tygře, ohnivou "
msgstr "září svítíš lesní tmou! "
msgstr "Kdo ten nesmrtelný byl, "
msgstr "že z ní tvůj souměr sestrojil? "
Tyger Tyger, burning bright,
Tygře, tygře, ohnivou

In the forests of the night;
září svítíš lesní tmou!

What immortal hand or eye,
Kdo ten nesmrtelný byl,

Could frame thy fearful symmetry?
že z ní tvůj souměr sestrojil?
<unit id=1>

<segment>
  <source xml:lang="EN">Tyger Tyger, burning bright, </source>
  <target xml:lang="CS">Tygře, tygře, ohnivou </target>
</segment>

<segment>
  <source xml:lang="EN">In the forests of the night; </source>
  <target xml:lang="CS">září svítíš lesní tmou! </target>
</segment>

<segment>
  <source xml:lang="EN">What immortal hand or eye, </source>
  <target xml:lang="CS">Kdo ten nesmrtelný byl, </target>
</segment>

<segment>
  <source xml:lang="EN">Could frame thy fearful symmetry? </source>
  <target xml:lang="CS">že z ní tvůj souměr sestrojil? </target>
</segment>
</unit>
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  Translation Editor  TMS  TMS  TMS  MT

www.gala-global.org/tapicc
Next steps?

• Chartering new Track including Working Groups as necessary
• Call For Participation is coming!
• Look at the resources
  • OASIS XLIFF OMOS TC
  • https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xliff-omos
  • 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