

TAPICC: Translation Cases and Classes



Jim Compton
Program Manager,
Partnerships Program,
RWS Moravia

Agenda

- Introductions: Who is Jim? What is GALA?
- Why and What is TAPICC?
- Overview and Progress
 - Track 1
 - Track 2
- Getting Involved
- Let's Chat!

Who is Jim?

- ~25-year veteran of the loc industry.
- Have been deeply focused on CMS/TMS integration.
- Manages **RWS Moravia**'s new Partnership Program.
- My mission:

Build a robust ecosystem of technology and services around our clients' global programs.

Directly ↑ relevant to
TAPICC



Jim Compton,
RWS Moravia

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



GALA is the **foundation** of the TAPICC initiative.

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

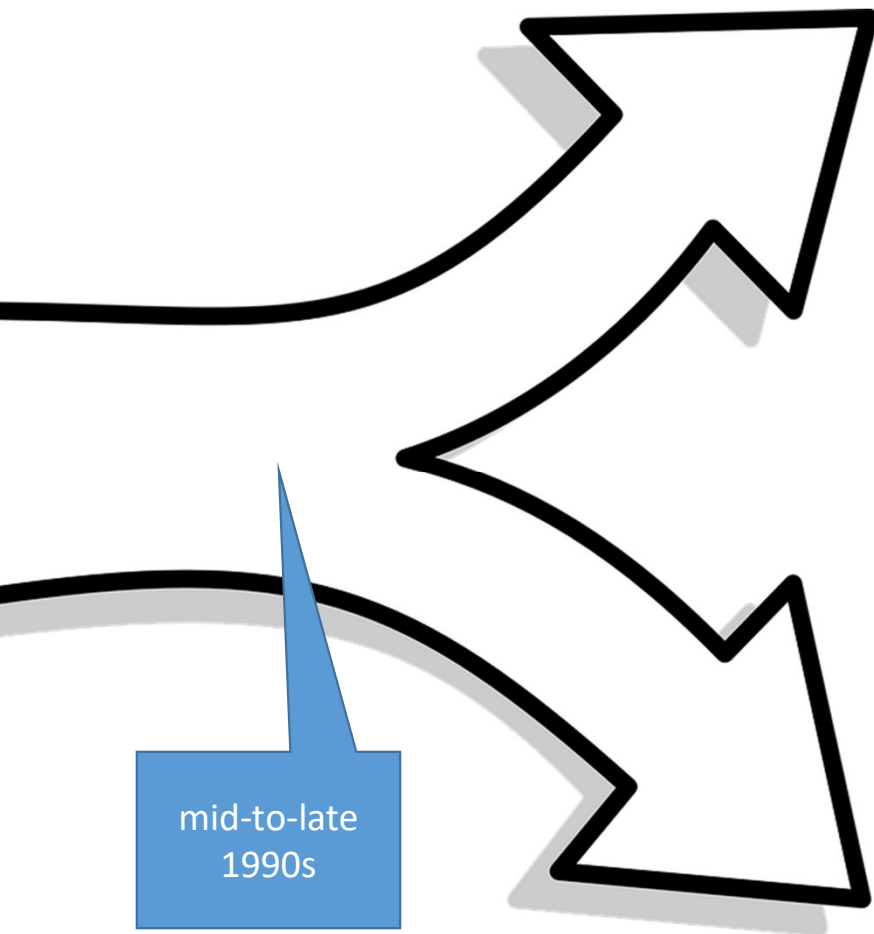
- Represent a CMS, DAM or similar tech?
- Represent a TMS, CAT tool or similar tech?
- Create / manage content that must get localized?
- Responsible for localizing / translating content?
- Any system integrators?



Why (and What Is) TAPICCC?

A low-angle photograph looking up at a utility worker on a metal tower. The worker is wearing a blue shirt and orange safety gear. They are surrounded by a dense, chaotic web of black power lines that crisscross the frame against a clear blue sky. The lines are bundled together in some places and hang loosely in others, creating a complex pattern. The worker is positioned in the lower right, reaching up towards the center of the frame where the lines are most concentrated.

A brief history of the problem.



content management

tech and practices



translation management

tech and practices

Starts with: “How do I translate the content that I’m managing in my CMS?”

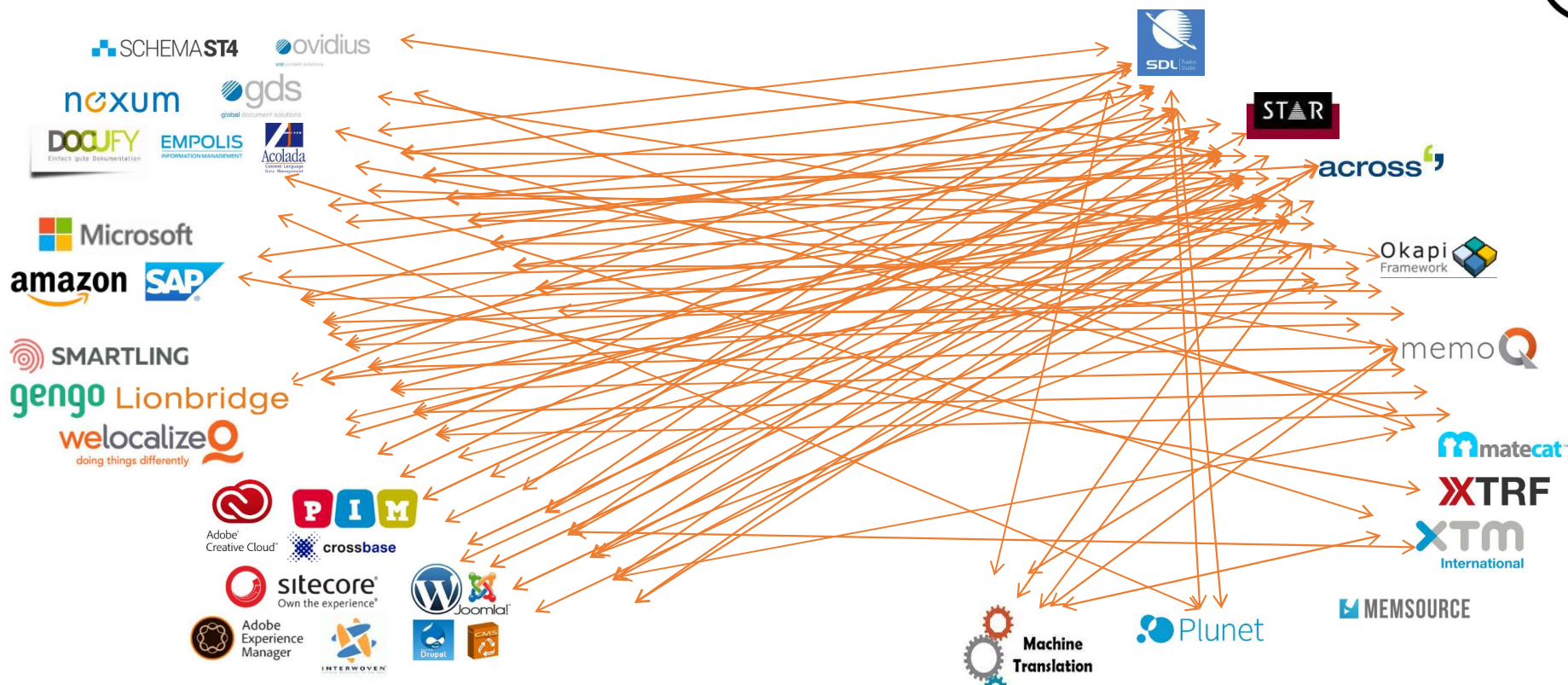


API Integration!



- system-to-system communication
- automation

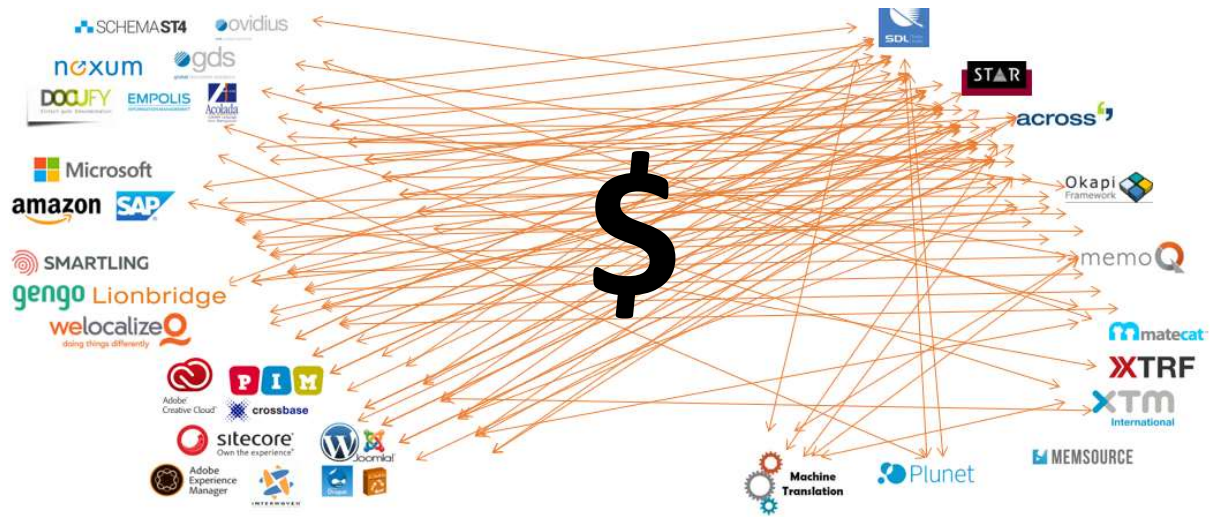
The rise of point-to-point integrations...



the “Wild West”

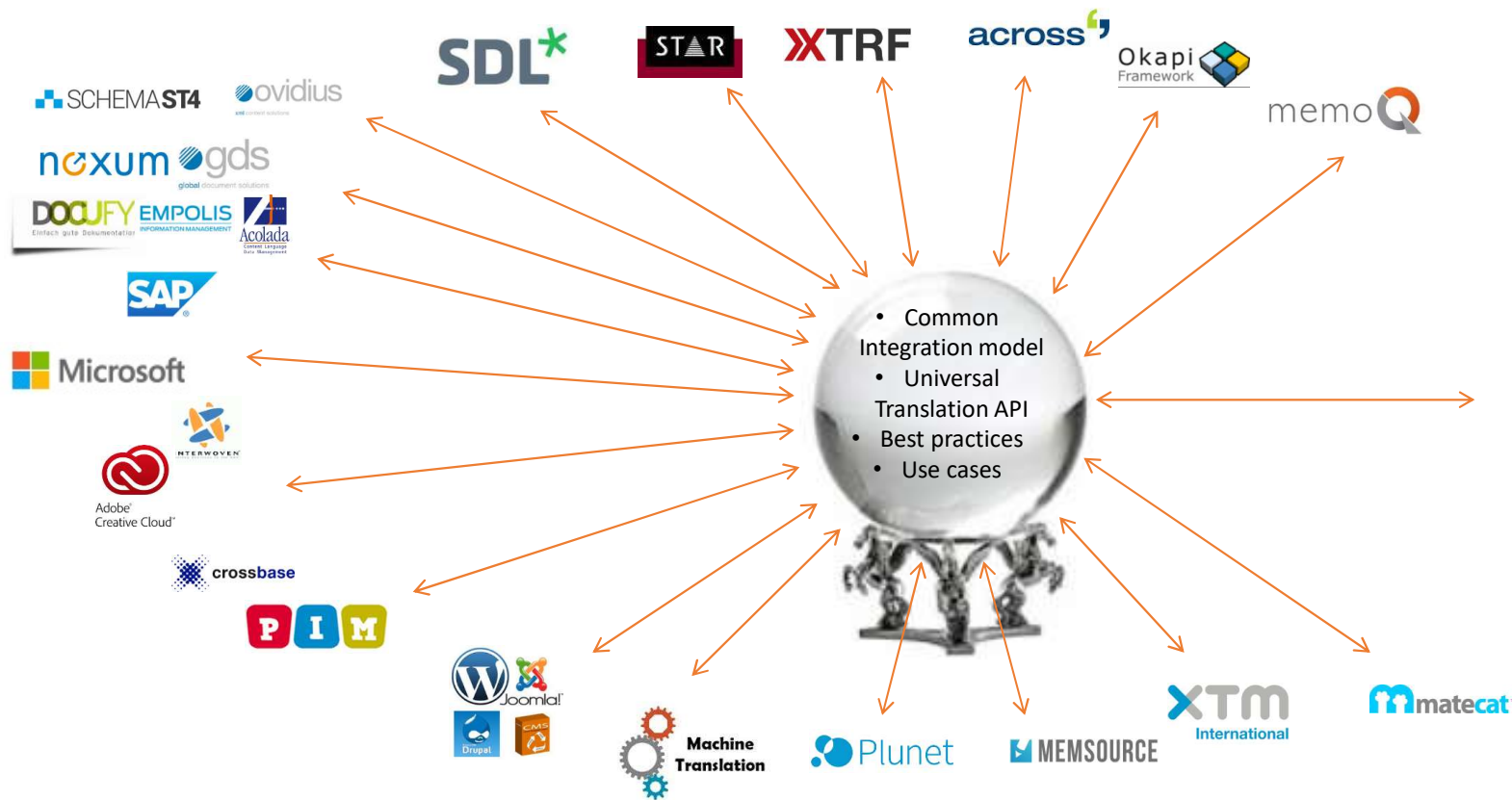


- Unnecessary variation
- Continuous reinvention of the wheel
- Wasted \$\$!
 - For clients
 - For LSPs
 - For tools vendors
- Loss of operational freedom



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

The TAPICC Vision



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

Initiative is Organized into Four Tracks

We've been working on this...

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”

...and are ramping-up with this.

4

Layout representation level

- Support process with visual context

Don't confuse the **tracks** with the **working groups**. The working groups are there to divide-up the work *within* a track.

Track One:

Supply Chain Automation

Track One Overview

1

Business metadata for supply chain automation

- Harmonize existing business metadata models

WG 1: Business Metadata

- Define business-level metadata
- Canonical names and values
- Workflow relevance
- Harmonization
- Compatibility

WG 2: Payload Specification

- Define types of payload
- Define payload-level metadata
- Canonical names and values
- Harmonization
- Compatibility

WG 3: XLIFF Extraction

- Extraction guidance
- ITS decoration
- XLIFF Extractors/Mergers
- XLIFF OM representations, JLIFF

WG 4: API Specification

- Protocols
- Transfer, pull and/or push
- Define classes

One of the principles of TAPICC is to leverage from relevant existing standards and standardization efforts, including (especially) **XLIFF**.

Progress So Far...

Business Metadata (WG 1)

- Have defined a set of standard **tasks types**, and method for creating nuanced task has been defined (reflected in the REST specification).
- Have created a business **glossary of localization terms**.
- Basically done.

Payload Specification (WG 2)

- Has created guidance around payload that will be summarized in a short prose deliverable.
- Evaluating similarities and aligning with the **Language Interoperability Portfolio (Linport)** initiative.

Progress So Far...

XLIFF Extraction (WG 3)

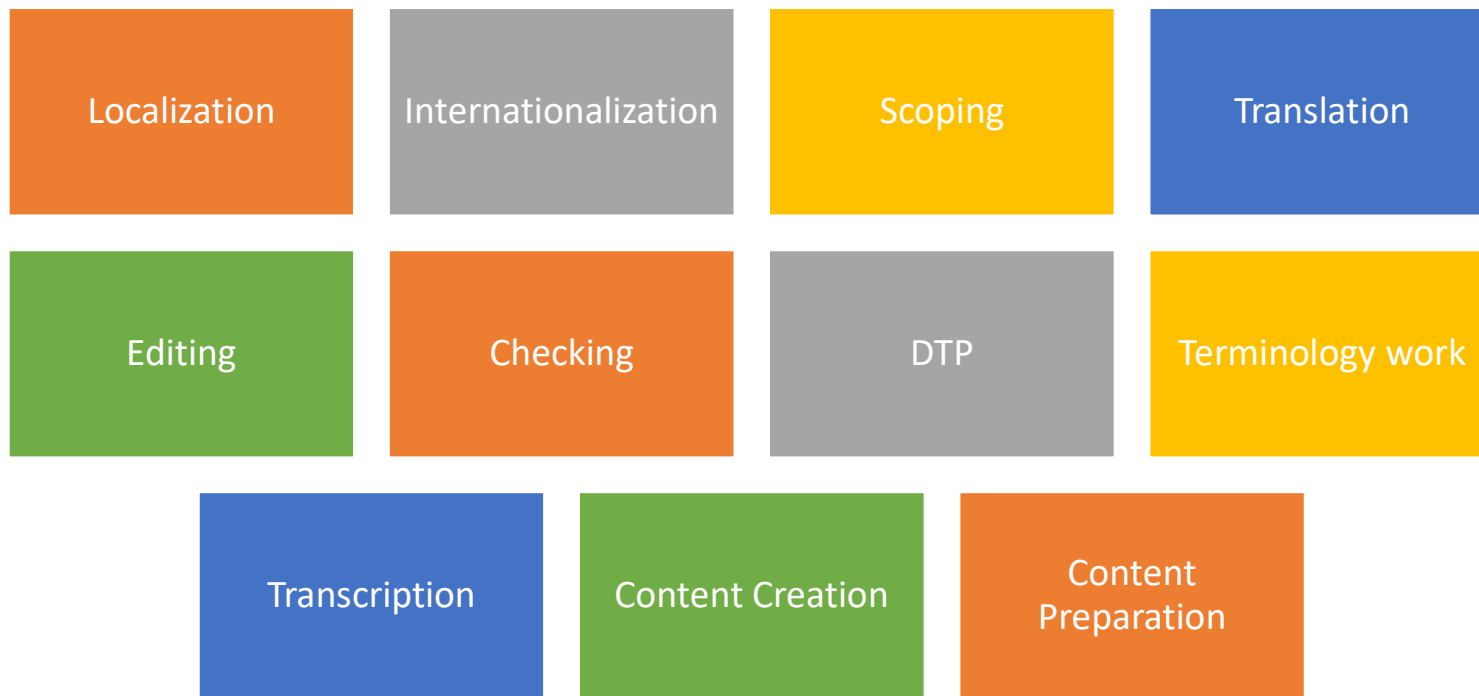
- “[XLIFF 2 Extraction and Merging Best Practice](#)” document being contributed to OASIS (good example of how pre-standardization efforts can become standard).
- Still need to support the community with real-world extraction problems.

API Specification (WG 4)

- RESTful **specification candidate** (in Swagger format) evolving though healthy debate on GitHub.
- Underlying data model continues to be refined.
- **Prose deliverable** being authored and edited by a small team.
- Various parties discussing reference implementations.

WG1 Complete Set of Task Types

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



Commenting on WG1 Results

- Elementary Linguistic Task Types

https://docs.google.com/document/d/1POWcolrwKvUQpgq_ImgcpK_p6PZ35k7g7mieKARGgR4/edit?usp=drive_web&oid=113136557825967529407

- Glossary of terms for the L10N industry

<https://docs.google.com/document/d/1OvYksscrYGtz7gUK-25AR28jobRN7gtkNtu1W3rvgaA/edit>

WG2 Consensus

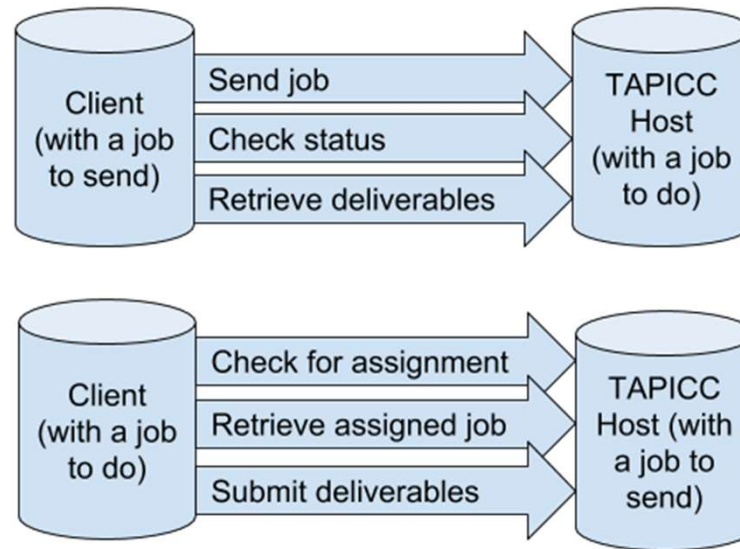
- Prose deliverable:
 - Primary goal of TAPICC is translation round trip. Other task are considered only as far as they support achieving the primary goal.
 - Payload is driven by task type.
 - If task type involves bitext, XLIFF2 is mandated. Legacy bitext is considered as source format and will be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

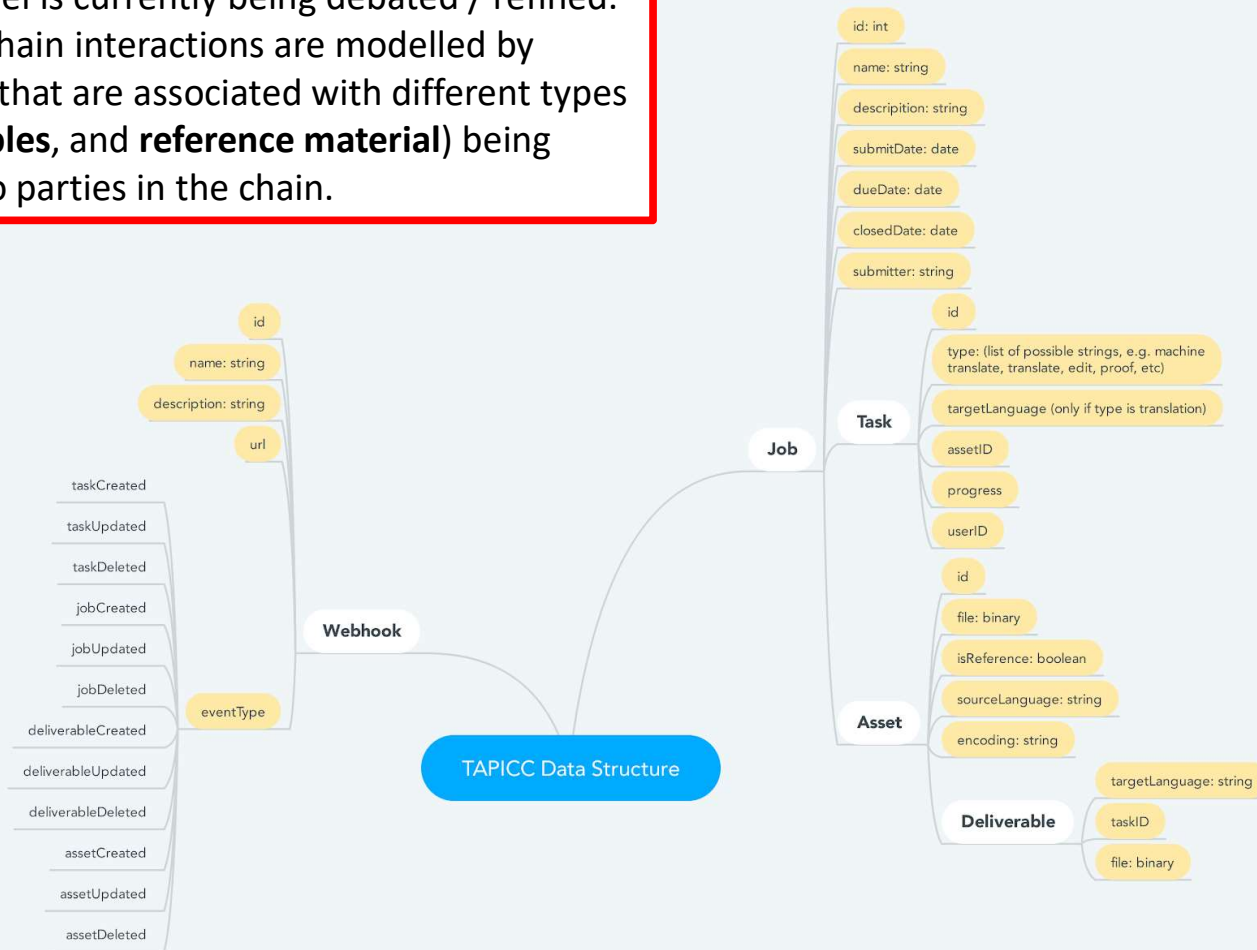
WG4 Model in a nutshell

- 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

moravia-te... | tapicc | 0.0.0

Editor Split UI

Last Saved: 07:15:29 pm Mar 13, 2018 VALID Save

```
1 swagger: '2.0'
2 info:
3   title: TAPICC
4   description: a Sails application
5   version: 0.0.0
6   contact:
7     name: alino
8   license:
9     name: ''
10  tags:
11    - name: Swagger
12    - name: Asset
13    - name: Job
14    - name: Task
15    - name: User
16    - name: Webhook
17  definitions:
18    contact:
19      properties:
20        group:
21          type: integer
22          format: int32
23        id:
24          type: integer
25          format: int32
26        createdAt:
27          type: string
28          format: date-time
```

Asset

- GET** /asset/{id} Read Object(s)
- POST** /asset/{id} Create Object(s)
- PUT** /asset/{id} Update Object(s)
- GET** /job/{parentid}/assets/{id} Read Object(s)
- POST** /job/{parentid}/assets/{id} Create Object(s)
- DELETE** /job/{parentid}/assets/{id} Destroy Object(s)
- GET** /asset Read Object(s)

Track Two: Unit-level Exchange

Track Two Overview

2

Exchange on unit level

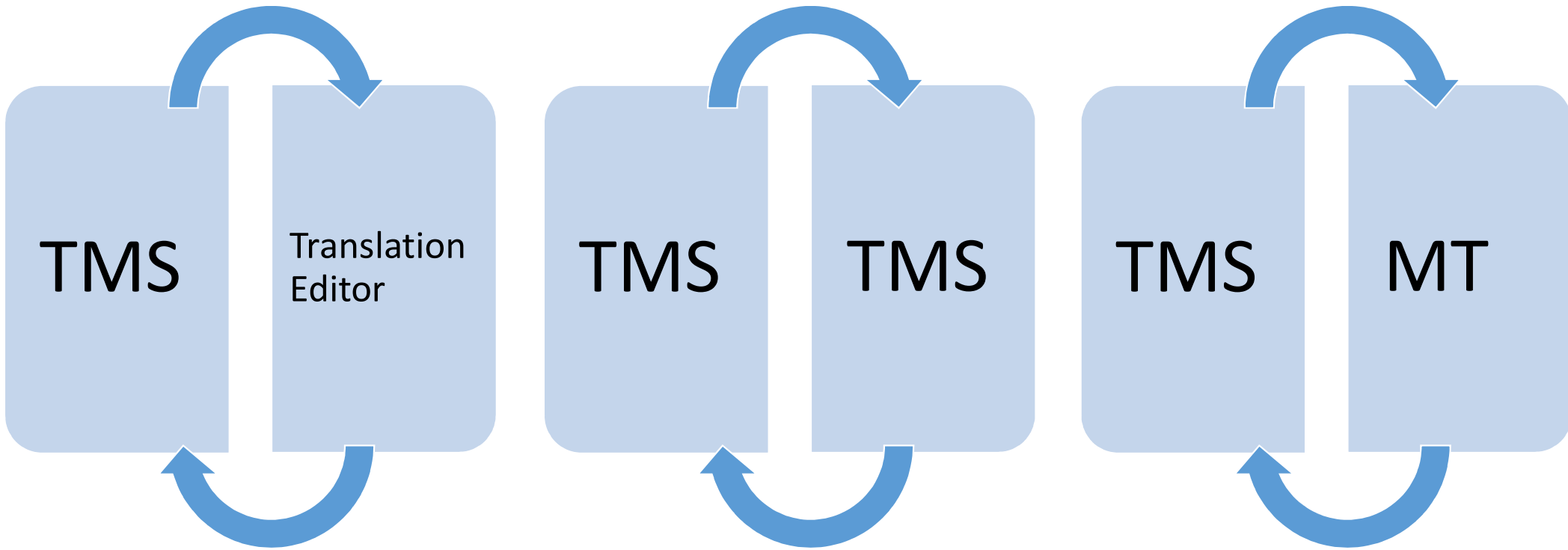
- Pass a segment/unit from an editor to a TM/MT or other tool

- T2 addresses **instantaneous fragment level exchange** between agents operating on bitext units such as TMS, MT Broker MT Engine, Translation editor etc.
- JRARTEBU!*
- *JLIFF REST API for Real Time Exchange of Bitext Units
- Track 2 charter: https://www.gala-global.org/sites/default/files/uploads/pdfs/TAPICC_Track2_Charter.pdf
- Track 2 WG1 Connect Group: <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



Other Resources for Track 2

- 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>

Getting Involved

How You Can Help

- Track 1 is ready for
 - Feedback on the deliverables
 - 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



How to Join

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

- Set up an account:
 1. <https://www.gala-global.org/user/register>
 2. Register as a new member.
- Join the Connect Group for TAPICC:
 1. <https://www.gala-global.org/community-groups>
 2. Search for “TAPICC”.
 3. Select the group.
 4. Click on “Join this group”.
 5. Click through and accept the legal.

(If you are representing an organization, you probably need to talk to your legal department.
<https://www.gala-global.org/tapicc-legal-agreement>.)

Thanks!

TAPICC Resources

<https://www.gala-global.org/TAPICC>

- Project Charter
- Open Source Legal agreement
- Project Statement
- TAPICC Groups
- Various links