

The technical concept of OSELI

&

Open Source as platform for MQM-based industry-wide benchmarks

Marc Mittag, MittagQI – Quality Informatics, Mössingen

Why Open Source?

The technical reasons

- Open collaboration leads to
 - More creativity
 - More people checking and fixing code
 - More people developing features
 - This leads to
 - Better stability
 - More feature diversity
- of the software

Why Open Source?

The economical reasons

The translation industry is becoming an IT-industry:

- Technical processes of translation will become more complex
 - Automated workflows
 - Automated quality checks & quality estimation
 - Workflows based on the results of automated QA-checks
- Machine translation
- Selling translation processes instead of translation
- Enabling clients to translate content, they never would have translated otherwise
- Web-based systems for every purpose

Why Open Source?

The economical reasons

3 choices for LSPs, to survive and prosper:

1. Developing the processes by themselves
2. Be completely dependent on the tools and processes of others
3. Develop the tools and processes in a shared way based on Open Source

The first 2 choices lead to a market concentration.

The third one will be the option for independence and process control.

Open Source: A shift in paradigms

What's different in Open Source

Some philosophical thoughts

- Our economy is based on “taking” - greed is the maxim
- It's based on competition – working against each other
- This leads to numerous problems: Social ones, environmental ones and economical ones

Open Source: A shift in paradigms

What's different in Open Source

Some philosophical thoughts

- Competition costs a lot
- Competition gives no benefit (at least in software)
- While jointly working on a common sense saves a lot

Open Source: A shift in paradigms

What's different in Open Source

What is needed is a culture based on

giving, sharing and working together –

and receiving back to close the circle.

Open Source: A shift in paradigms

What's different in Open Source

And: It simply makes fun to work together

Across the boundaries of companies

And to cooperate with competitors :-)

Technical concept of OSELI

Open Source Ecosystem for the Language Industry

OSELI is about collaboration

- The diversity of Open Source projects form an ecosystem
- The strength comes from different tools and projects collaborating and interfacing
- The goal is, that all tool and projects are interfacing and cooperating

Enabling plurality of solutions: The idea of an ecosystem

Technical concept of OSELI

Open Source Ecosystem for the Language Industry

Integration is about APIs and Plug-ins

- (A) small core system/s is/are needed
- It/They offer/s only core features
- And most important: It/They offer/s
 - A Plug-in system
 - A complete API

This is what we intend with translate5 OSELI project

Technical concept of OSELI

Open Source Ecosystem for the Language Industry

Short discussion break

Technical concept of OSELI

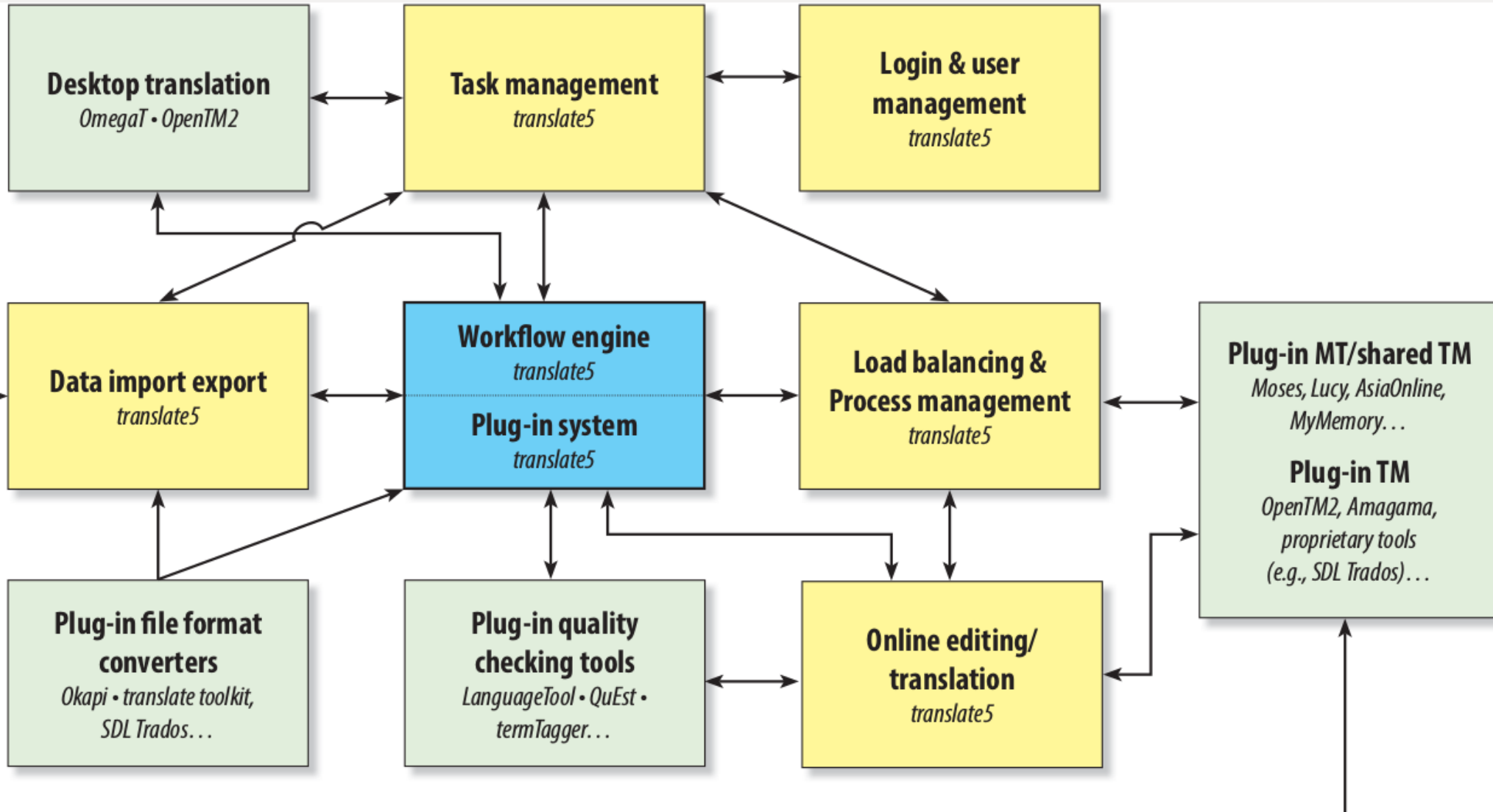
Open Source Ecosystem for the Language Industry

translate5 infrastructure for integration

- REST-API for nearly everything
- Event-based plug-in system
- Events where needed
- Plug-ins can hook in
 - Import/export
 - Workflow
 - QA
 - Back-end & GUI
 - ... nearly everywhere

OSELI translate5 instance

Present & future



OSELI translate5 instance

Present & future

Short live presentation

OSELI translate5 instance

Next steps

- Flexible TM- & MT-integration
- Flexible workflow engine
- MQM-usage to evaluate business benchmarks

OSELI translate5 instance

Next steps

Flexible TM- & MT-integration

- First project set up for fund-raising from multiple companies
- Integration for matches and concordance search of
 - openTM2 as TM-back-end
 - MyMemory as Shared Memory
 - Moses-systems for MT-Matches
- Will be developed as translate5 Plug-in

OSELI translate5 instance



translate5

Eingeloggter Benutzer: Thomas Lauria
Loginname: manager

Abmelden Meine Einstellungen Deutsch
Task Overview User Administration **TM Overview**

TM Overview

Add TM Aktualisieren

TM Name			Color		Resource
MyMemory	Deutsch (de)	English (en)	Green		MyMemory
TM of Client XY	Deutsch (de)	English (en)	Cyan		OpenTM2
Collected Phrases	English (en)	Deutsch (de)	Orange		My Slow TM
TM of Client XY	Deutsch (de)	Italienisch (it)	Purple		OpenTM2
Moses MT	English (en)	Deutsch (de)	Red		Moses MT

OSELI translate5 instance

The screenshot shows the translate5 web interface. At the top left is the 'open source initiative' logo. The main header displays 'translate5' and 'Login name: manager'. A dropdown menu on the right is set to 'English'. The 'Task Overview' sidebar on the left contains 'Add Task' and 'Refresh' buttons, and a grid of task icons. The main content area is dominated by a modal dialog titled 'Properties to task "ENG_Minipaket"'. This dialog has three tabs: 'Associate users to task', 'Task-properties', and 'Associate TMs to task'. The 'Associate TMs to task' tab is active, showing a table of translation memory (TM) entries. The table has columns for a checkbox, a colored name, and source/target language pairs. A 'Save TMs' button is located at the bottom right of the dialog. In the bottom right corner of the overall interface, there is a red 'X' icon with the text 'close window'.

translate5

Login name: manager

English

Task Overview

Add Task Refresh

Actions

Properties to task "ENG_Minipaket"

Associate users to task Task-properties Associate TMs to task

	Name		
OpenTM2 TMs			
<input type="checkbox"/>	TM of Client XY	Deutsch (de)	Englisch (en)
<input checked="" type="checkbox"/>	TM of Client YX	Deutsch (de)	Englisch (en)
My Slow TM			
<input type="checkbox"/>	Collected Phrases	Deutsch (de)	Englisch (en)
MyMemory			
<input checked="" type="checkbox"/>	MyMemory	Deutsch (de)	Englisch (en)
Moses MT			
<input checked="" type="checkbox"/>	Moses MT	Deutsch (de)	Englisch (en)

Save TMs

close window

OSELI translate5 instance



translate5

Files << Segment list and editor

Work files - Editor modes ▾ Hide tags Short tag view Full tag view Reset sorting / filtering QM Subsegment Statistics

Nr.	Source text	Pivot language	Target text	Target text	Comments
1	This file is a based on a part of the php-online-Dokumentation. It's translation is done by a pretranslation based on a very fast winalign-Project and is not at all state of the translation art. It's only purpose is the generation of demo-data for translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	
2	Apache 2.x on Unix systems - Manual	xxxxxxxxxxxxxxxxxxxxxApache 2.0 auf Unixsystemen - Manual	Apache 2.0 auf Unixsystemen - Manual	Apache 2.0 auf Unixsystemen - Manual	
3	PHP Manual	xxxxxxxxxxxxxxxxxxxxxPHP Manual	PHP Manual	PHP Manual	
4	Installation and Configuration	xxxxxxxxxxxxxxxxxxxxxInstallation und Konfiguration	Installation und Konfiguration	Installation und Konfiguration	
5	Installation on Unix systems	xxxxxxxxxxxxxxxxxxxxxInstallation auf Unix-Systemen	Installation auf Unix-Systemen	Installation <u>auf Unix-Systemen</u>	
6	Apache 1.3.x on Unix systems	xxxxxxxxxxxxxxxxxxxxxApache 1.3.x auf Unix-Systemen	Apache 1.3.x auf Unix-Systemen	Apache 1.3.x auf Unix-Systemen	
7	Apache 2.x on Unix systems				

MT / TM Matches Concordance Search

	Source Lang	Target lang	MatchRate / TM
1	Installation on Unix systems	Installation auf Unix-Systemen	100%
2	Installation on Unix systems	Installation auf Unix-Systemen	100%
3	Installation on Linux systems	Installation auf Linux-Systemen	95%
4	Installation on a Unix system	Installation auf einem Unix-System	85%
5	Installation on PC systems	Installation auf PC-Systemen	85%
6	Attributes: Last Edit: Marc Mittag 07/05/2015 Created: Thomas Lauria 07/05/2014	Another Result....	80%
7		Another Result....	80%
8		Another Result....	80%
9	(This is shown as ToolTip on MouseOver as long the mouse is over the icon)	Another Result....	80%
10		Another Result....	80%

Reference files +

OSELI translate5 instance



translate5

Dateien << **Segmentliste und Editor**

Arbeits-Dateien - DummyDirectory - install-unix.apach...

Editormodi ▾ | Tags verbergen | **Tag-Kurzansicht** | Tag-Vollansicht | Sortierung / Filterung zurücksetzen | QM Subsegment Statistik

Nr.	Ausgangstext	Relaissprache	Zieltext	Zieltext ✎	Kommentare
1	This file is based on a part of the php-online-Documentation. It's translation is done by a pretranslation based on a very fast winalign-Project and is not at all state of the translation art. It's only purpose is the generation of demo-data for translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	
2	Apache 2.x on Unix systems - Manual	xxxxxxxxxxxxxxxxxxxxxApache 2.0 auf Unixsystemen - Manual	Apache 2.0 auf Unixsystemen - Manual	Apache 2.0 auf Unixsystemen - Manual	
3	PHP Manual	xxxxxxxxxxxxxxxxxxxxxPHP Manual	PHP Manual	PHP Manual	
4	Installation and Configuration	xxxxxxxxxxxxxxxxxxxxxInstallation und Konfiguration	Installation und Konfiguration	Installation und Konfiguration	
5	Installation on Unix systems	xxxxxxxxxxxxxxxxxxxxxInstallation auf Unix-Systemen	Installation auf Unix-Systemen	Installation auf Unix-Systemen	
6	Apache 1.3.x on Unix systems	xxxxxxxxxxxxxxxxxxxxxApache 1.3.x auf Unix-Systemen	Apache 1.3.x auf Unix-Systemen	Apache 1.3.x auf Unix-Systemen	
7	Apache 2.x on Unix systems				
8	Lighttpd 1.4 on Unix systems	xxxxxxxxxxxxxxxxxxxxxLighttpd 1.4 on Unix systems	Lighttpd 1.4 on Unix systems	Lighttpd 1.4 on Unix systems	

MT / TM Matches | Concordance Search | **TestTM One Results** x

Searchfield for Source: Searchfield for Target: 2 of 3 TM Resources answered

	Source Lang	Target Lang	TM - Resource
1	Installation on Unix systems	Installation auf Unix-Systemen	TM of Client XY - OpenTM2
2	Installation on Unix systems	Installation auf Unix-Systemen	TestTM One - MyTM
3	Installation on Linux systems	Installation auf Linux-Systemen	TM of Client XY - OpenTM2
4	Installation on a Unix system	Installation auf einem Unix-System	TM of Client XY - OpenTM2
5	Installation on PC systems	Installation auf PC-Systemen	TestTM One - MyTM
	loading...		Collected Phrases - My Slow TM

Referenz-Dateien +

OSELI translate5 instance

Authenticat
Login name:
Task: testfiles
TaskGuid: {35

Logout



Segment list and editor

Editor modes | Hide tags | Short tag view | Full tag view | Reset sorting / filtering | QM Subsegment Statistics | Settings

Nr.	Source text	Target text	Autostatus
1	This file is a based on a part of the php-online-Dokumentation. It's translation is done by a pretranslation based on a very fast winalign-Project and is not at all state of the translation art. It's only purpose is the generation of demo-data for translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	
2	Apache 2.x on Unix systems - Manual	Apache 2.0 auf Unixsystemen - Manual	
3	PHP Manual	PHP Manual	

MT / TM Matches | Concordance Search

Nr.	Source text	Target text	Score
1	Installation on Unix systems	Installation auf Unix-Systemen	100%
2	Installation on Unix systems	Installation auf Unix-Systemen	100%
3	Installation on Linux systems	Installation auf Linux-Systemen	95%
4	Installation on a Unix system	Installation auf einem Unix-System	85%
5	Installation on PC systems	Installation auf PC-Systemen	85%

OSELI translate5 instance

Next steps

Short discussion break

OSELI translate5 instance

Next steps

Flexible TM- & MT-integration

- translate5 project is looking for 10 companies to support the development
- Each supporting company
 - Can influence the development
 - Will get 6 months of exclusive usage before the features will be public
- Beo did fund the specification (incl. architecture concept)
<https://mittagqi.atlassian.net/browse/TRANSLATE-421>

Building OSELI bottom-up

Building OSELI the grass-root way:

- Organizing in open networks
- Open, transparent communication and coordination
- Users and developers joining efforts to reach the goal

This is the way OSELI translate5 project is going to be organized

Who supports OSELI so far

Open Source Ecosystem for the Language Industry

Organizations

- GALA
- LT-Innovate
- TAUS
- META-NET

Open Source Projects

- openTM2
- translate5
- OmegaT

Companies

- DFKI
- Beo
- Text & Form
- Iolar
- Logrus
- Memsource
- ABBYY
- MittagQI

Business benchmarks

based on translate5 & MQM

- Based on MQM calculation of a translation quality indicator is possible
- translate5 project
 - Supports MQM markup
 - Aims to support calculation of the translation quality indicator
- For comparability common MQM-sets are needed per
 - Language pair
 - Subject

Business benchmarks

based on translate5 & MQM

A number of companies will cooperate to

- Agree on a number of MQM-sets for specific subjects (e. g. automotive and medical instruments)
- Agree on algorithms to get samples from translation jobs
- Aim to broaden that approach towards an industry standard

Business benchmarks

based on translate5 & MQM

Based on these MQM-sets, samples and translate5 workflow integration

- QA-calculation can become standard in translation industry
- Can be uploaded in anonymised form to a server, which calculates benchmarks
- The Open Source ensures everyone, that the data is indeed anonymised

OSELI translate5 project

Thank you for your attention!

And last, but not least ...

OSELI translate5 project

Get involved

- Contribute your ideas
- Support the development
- Contribute your plugins & get involved in development

Download translate5: <http://www.translate5.net/downloads/>