



Industry Shared Metrics with the TAUS Dynamic Quality Dashboard and API

What About Translation Quality

Old School: “One size fits all”

Since the 1980's

LISA QA Model, SAE J2450 prescribe today's quality processes:

1. Static:

- One quality fits all purposes, all content, all audiences

2. Subjective:

- Evaluations are often subjective and anecdotal

3. Costly:

- QE causes friction, delays
- QE can cost up to 25% of total translation costs

4. Non-transparent:

- Necessity without remedy



Dynamic
Quality
Framework

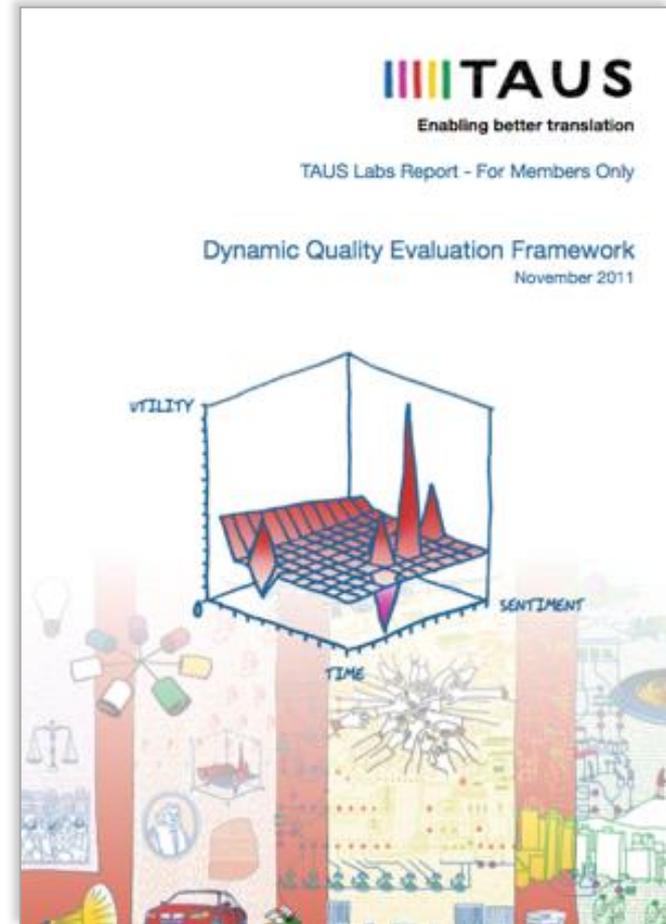


Industry Collaborative Program

DQF started in 2011

Participating members

Adobe	Lionbridge
Appen	Medtronic
Autodesk	Microsoft
AVB	Moravia
CA Technologies	Nikon
Cisco	Oracle
Crestec	Pactera
Crosslang	Pangeanic
Dell	Paypal
DFKI	Philips
eBay	PTC
EMC	Siemens
Google	Spil Games
Hewlett Packard	Systran
Intel	VMware
LDS Church	Welocalize
Lingo24	Yahoo!



From DQF Tools to Quality Dashboard

DQF Tools

Since January 2014

Tools on TAUS web site:

to measure:

- Productivity
- Adequacy
- Fluency

to review and count:

- Translation errors

to get:

- Stats and reports

Used by 100+ members



Quality Dashboard

Launched June 2015

DQF integrated in:

- CAT Tools
- TMS Systems

Use of DQF plug-in provides:

- Enhanced statistics
- Benchmarking

Open to everyone

The Power and Value of the Quality Dashboard

- DQF collects data and generates reports on the Dashboard real-time
- Translators, managers, buyers, developers get their own stats, benchmarks and analytics
- Not only track and benchmark against your own data, but also against industry averages, between translators, customers, projects, technologies

Quality Evaluation



Business Intelligence



What is the average productivity?

Quality Dashboard

Productivity

Efficiency

Adequacy

Fluency

Language

Time

Technology

Process

Content

Industry

Project

Translator/vendor

Customer

Statistics

Language

Time

Technology

Process

Content

Industry

Translator/vendor

Customer

Distribution of segments

Language

Time

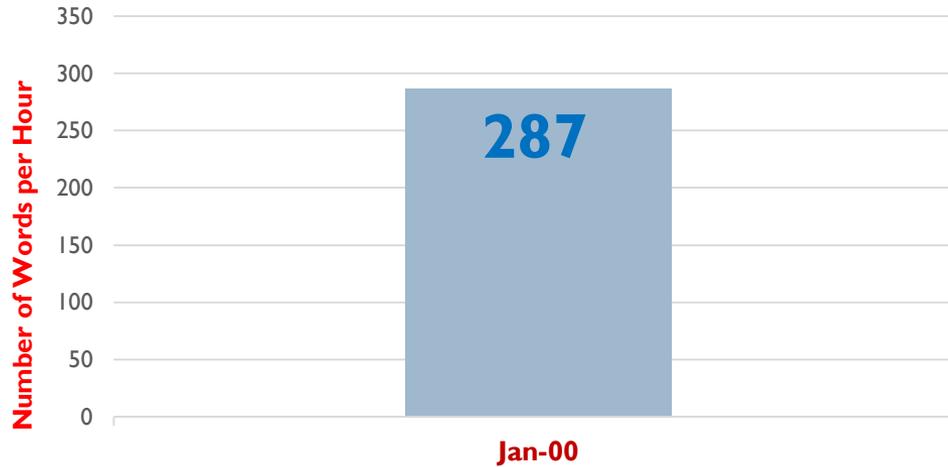
Content

Industry

Project

Productivity

Across all languages, industries, technologies, processes, content, translators



August 31, 2015: 566,987,756 words have been measured

-  More information
-  Benchmark

Quality Dashboard

Content profiling

Quality evaluation

Adequacy/Fluency

Error review

MT Ranking

Productivity measurement

DQF API

What is the average productivity of MT vs. TM?

Quality Dashboard

Productivity

Efficiency

Adequacy

Fluency

Language

Time

Technology

Process

Content

Industry

Project

Translator/vendor

Customer

Statistics

Language

Time

Technology

Process

Content

Industry

Translator/vendor

Customer

Distribution of segments

Language

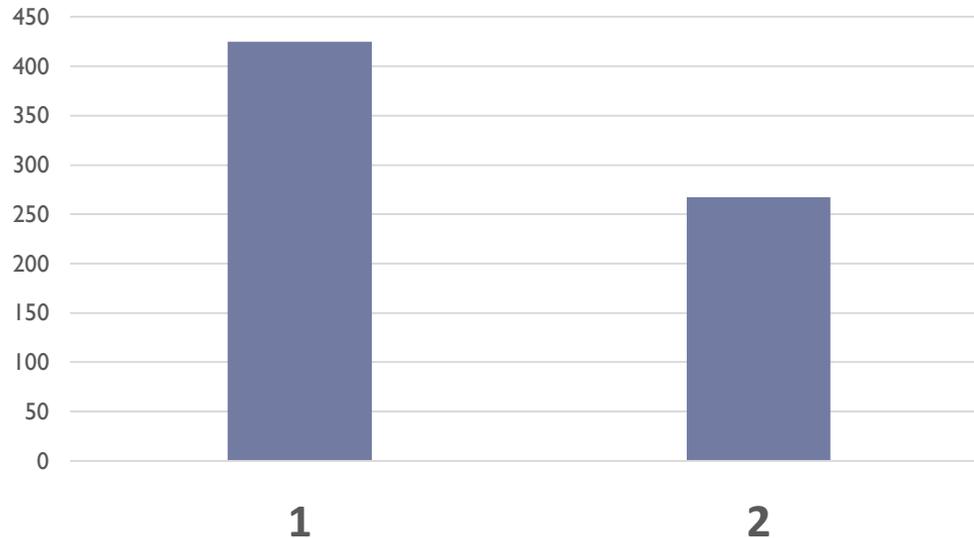
Time

Content

Industry

Project

Chart Title



August 31, 2015: 566,987,756 words have been measured

 **More information**

 **Benchmark**

Quality Dashboard

Content profiling

Quality evaluation

Adequacy/Fluency

Error review

MT Ranking

Productivity measurement

DQF API

API Integrators

What is my productivity compared to industry?

Quality Dashboard

Productivity

Efficiency

Adequacy

Fluency

Language

Time

Technology

Process

Content

Industry

Project

Translator/vendor

Customer

Statistics

Language

Time

Technology

Process

Content

Industry

Translator/vendor

Customer

Distribution of segments

Language

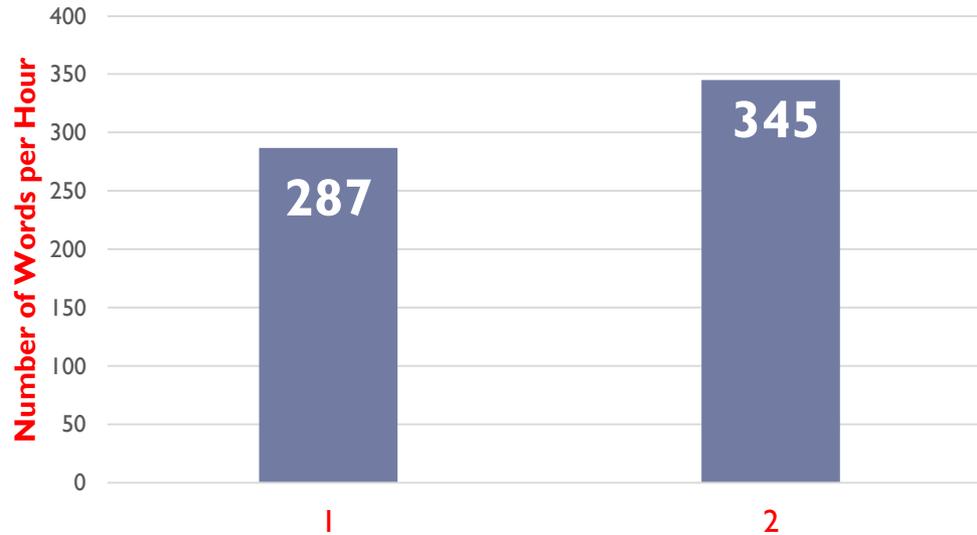
Time

Content

Industry

Project

My Productivity Compared to Industry Average



August 31, 2015: 566,987,756 words have been measured

[i](#) More information

[>](#) **Benchmark**

Quality Dashboard

Content profiling

Quality evaluation

Adequacy/Fluency

Error review

MT Ranking

Productivity measurement

DQF API

Where do my translations come from?

Quality Dashboard

Productivity

Efficiency

Adequacy

Fluency

Language

Time

Technology

Process

Content

Industry

Project

Translator/vendor

Customer

Statistics

Language

Time

Technology

Process

Content

Industry

Translator/vendor

Customer

Distribution of segments

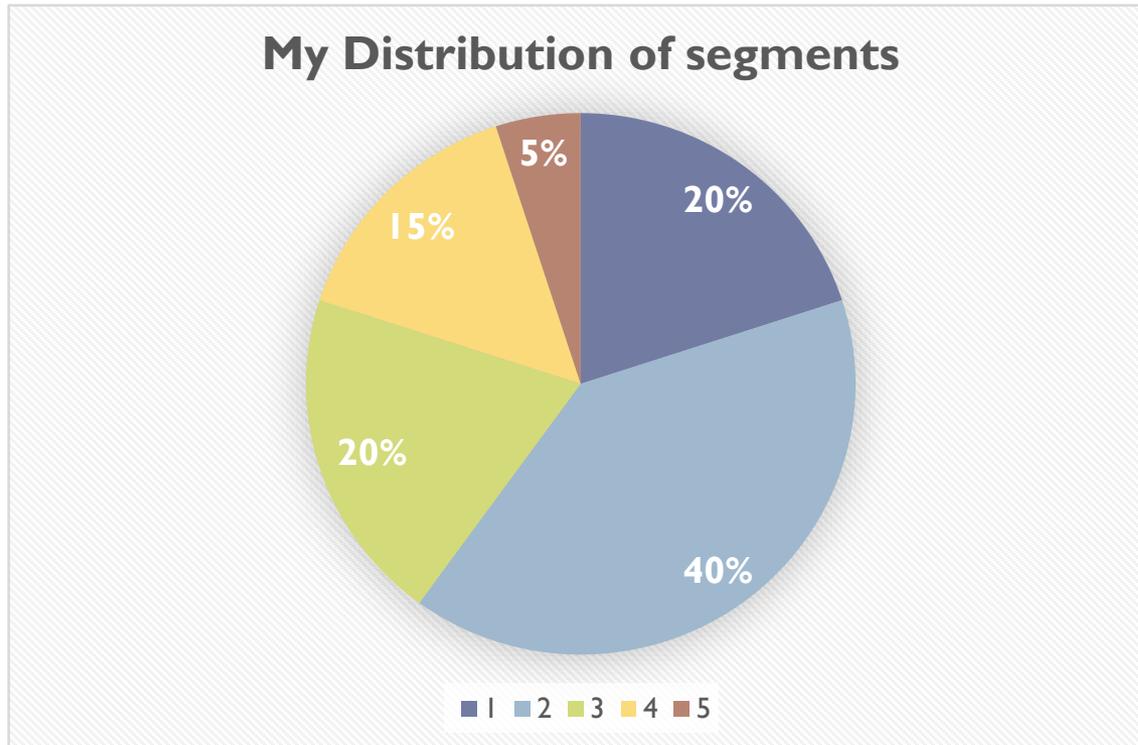
Language

Time

Content

Industry

Project



August 31, 2015: 566,987,756 words have been measured

 **More information**

 **Benchmark**

Quality Dashboard

Content profiling

Quality evaluation

Adequacy/Fluency

Error review

MT Ranking

Productivity measurement

DQF API

Where do my translations come from vs. industry?

Quality Dashboard

Productivity

Efficiency

Adequacy

Fluency

Language

Time

Technology

Process

Content

Industry

Project

Translator/vendor

Customer

Statistics

Language

Time

Technology

Process

Content

Industry

Translator/vendor

Customer

Distribution of segments

Language

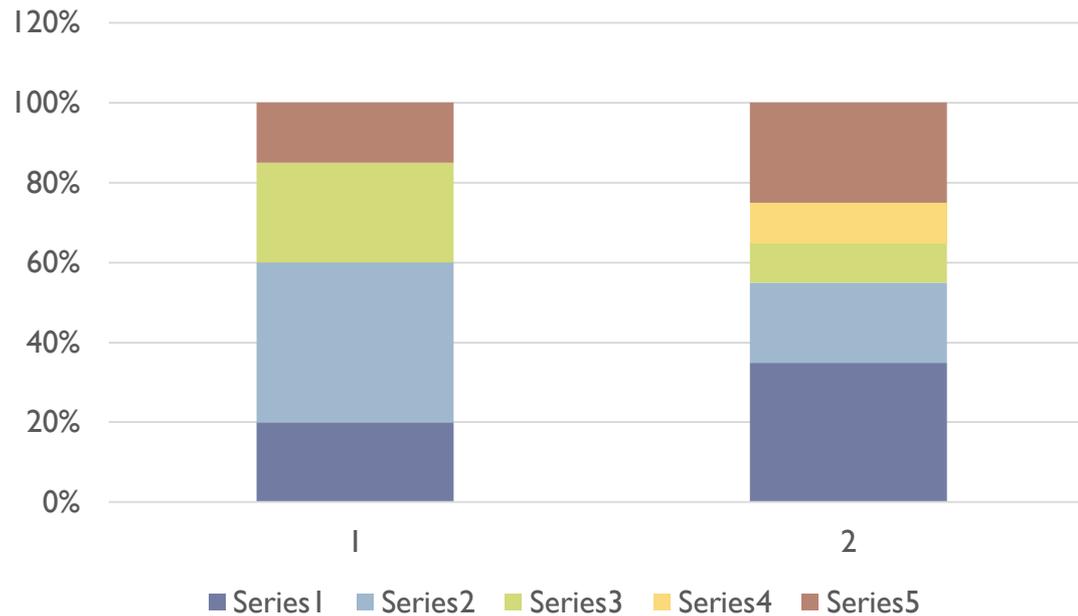
Time

Content

Industry

Project

My Distribution of Segments Compared to Industry



[i](#) More information

[>](#) **Benchmark**

Quality Dashboard

Content profiling

Quality evaluation

Adequacy/Fluency

Error review

MT Ranking

Productivity measurement

DQF API

August 31, 2015: 566,987,756 words have been measured

What is my productivity by language and by project?

Quality Dashboard

Productivity

Efficiency

Adequacy

Fluency

Language

Time

Technology

Process

Content

Industry

Project

Translator/vendor

Customer

Statistics

Language

Time

Technology

Process

Content

Industry

Translator/vendor

Customer

Distribution of segments

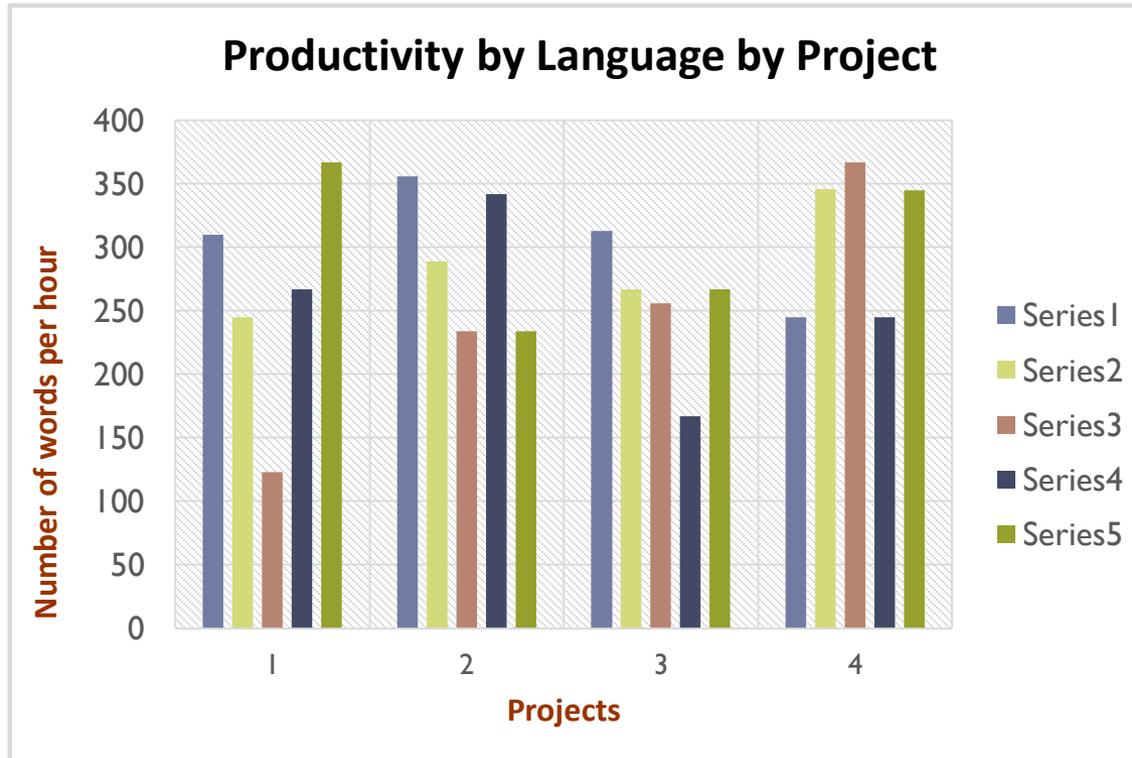
Language

Time

Content

Industry

Project



August 31, 2015: 566,987,756 words have been measured

 **More information**

 **Benchmark**

Quality Dashboard

Content profiling

Quality evaluation

Adequacy/Fluency

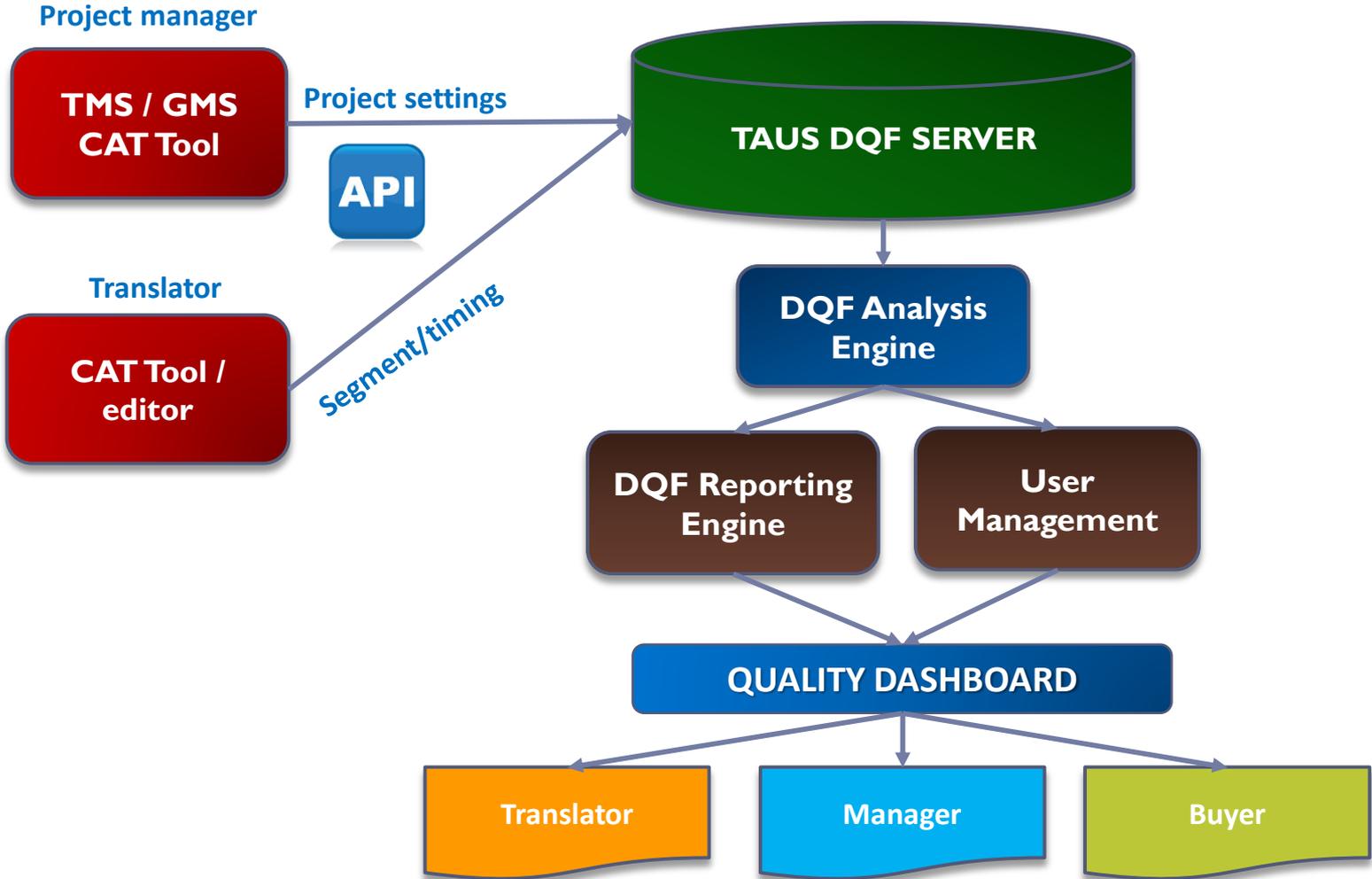
Error review

MT Ranking

Productivity measurement

DQF API

TAUS DQF Infrastructure



DQF Data Instrumentation

- ▶ Milliseconds per segment
- ▶ Source segment
- ▶ Target segment
- ▶ Edited target segments
- ▶ Time
- ▶ Language pair
- ▶ Project key
- ▶ Translator key



Open API

- ▶ **Test Environment**

- ▶ <https://dqf.taus.net/assets/api/v1/index.html>

- ▶ **Open API on GitHub**

- ▶ <http://github.com/TAUSBV/dqf-api>

- ▶ Specification

- ▶ Test Code

- ▶ Documentation

- ▶ Issue Tracker

- ▶ **Available under the MIT Open Source License**



Quality Dashboard Integrators



“Microsoft Office International team is committed to the DQF model and approach and are actively partnering with TAUS to investigate how best to integrate TAUS Quality Dashboard API into our translation tool set.”



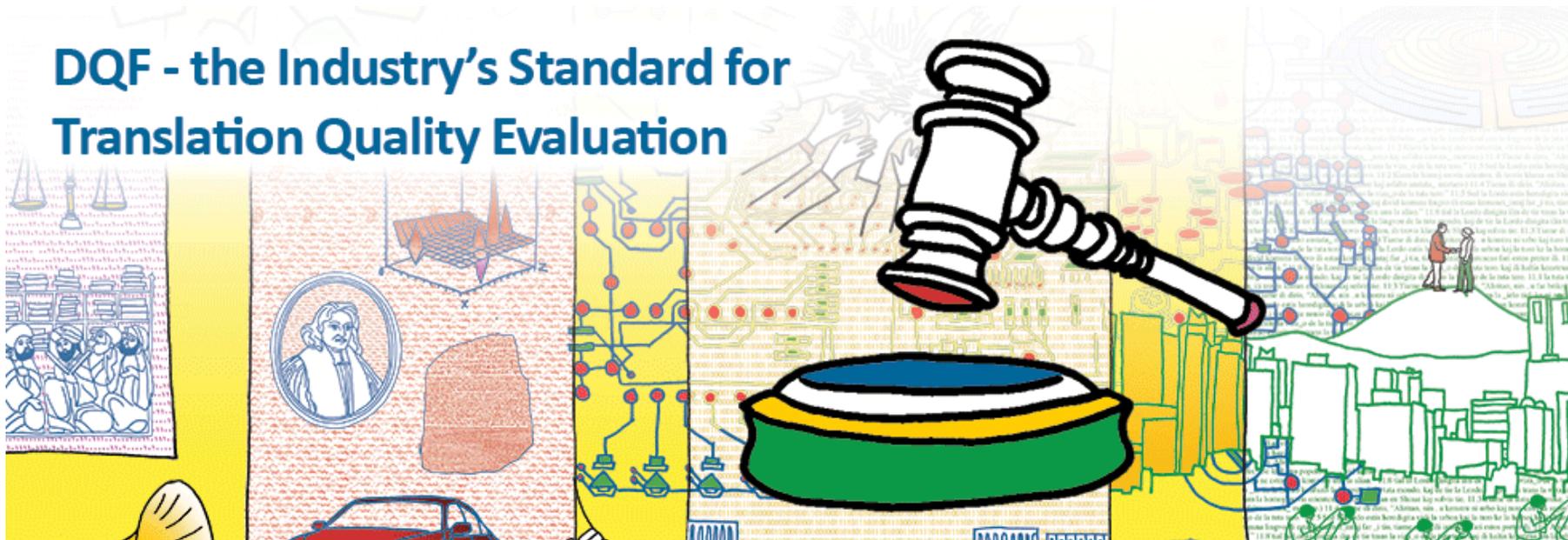
The TAUS Efficiency Score

Introducing a new score for measuring productivity

- ▶ **2 Core variables:**
 - ▶ Words per Hour - **WPH**
 - ▶ Edits per Hour - **EPH**
- ▶ **Efficiency = WPH + EPH**
 - ▶ **Normalized** using **Min-Max**
- ▶ **Credit:** Nikos Argyropoulos

Productivity

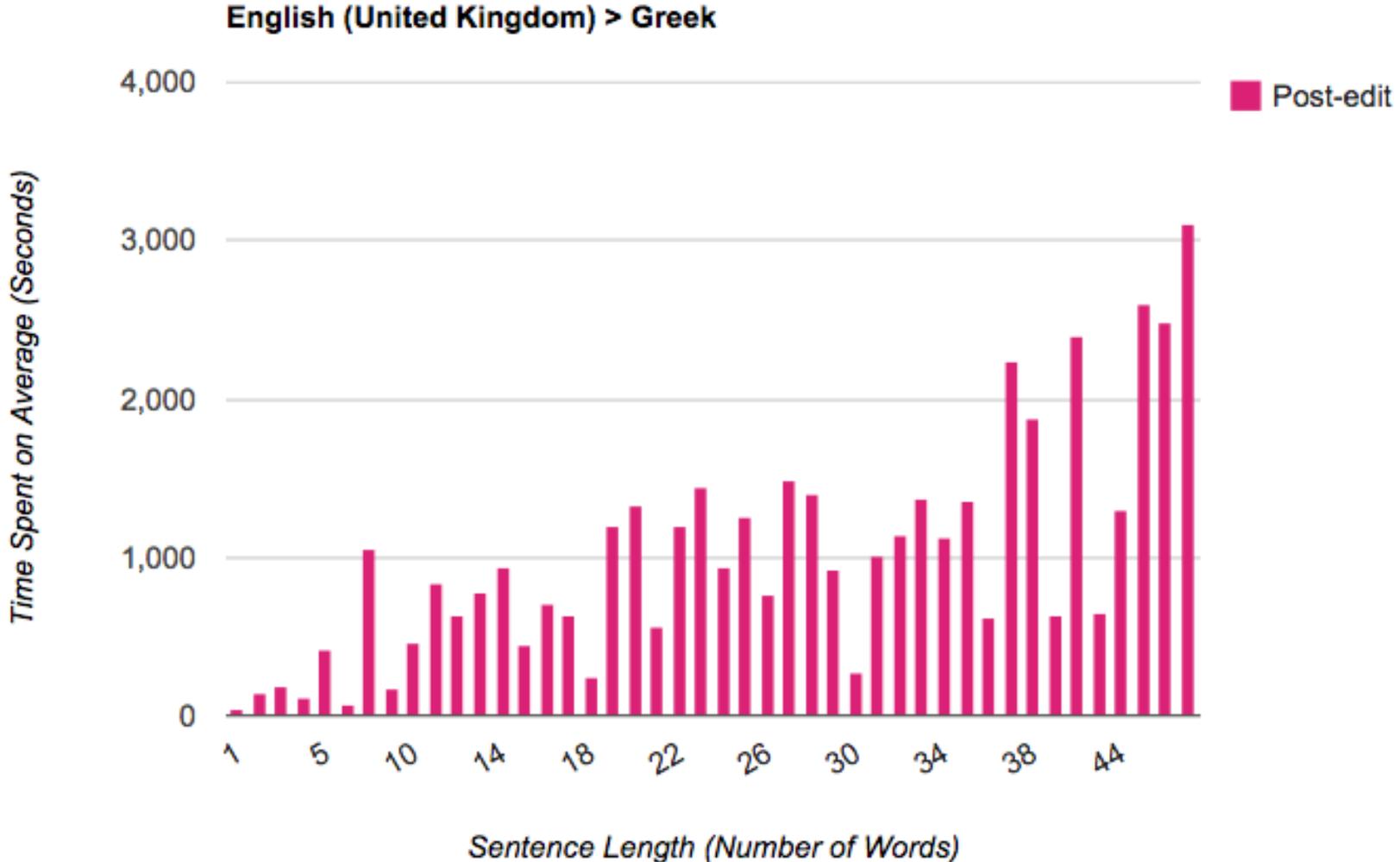
DQF - the Industry's Standard for Translation Quality Evaluation



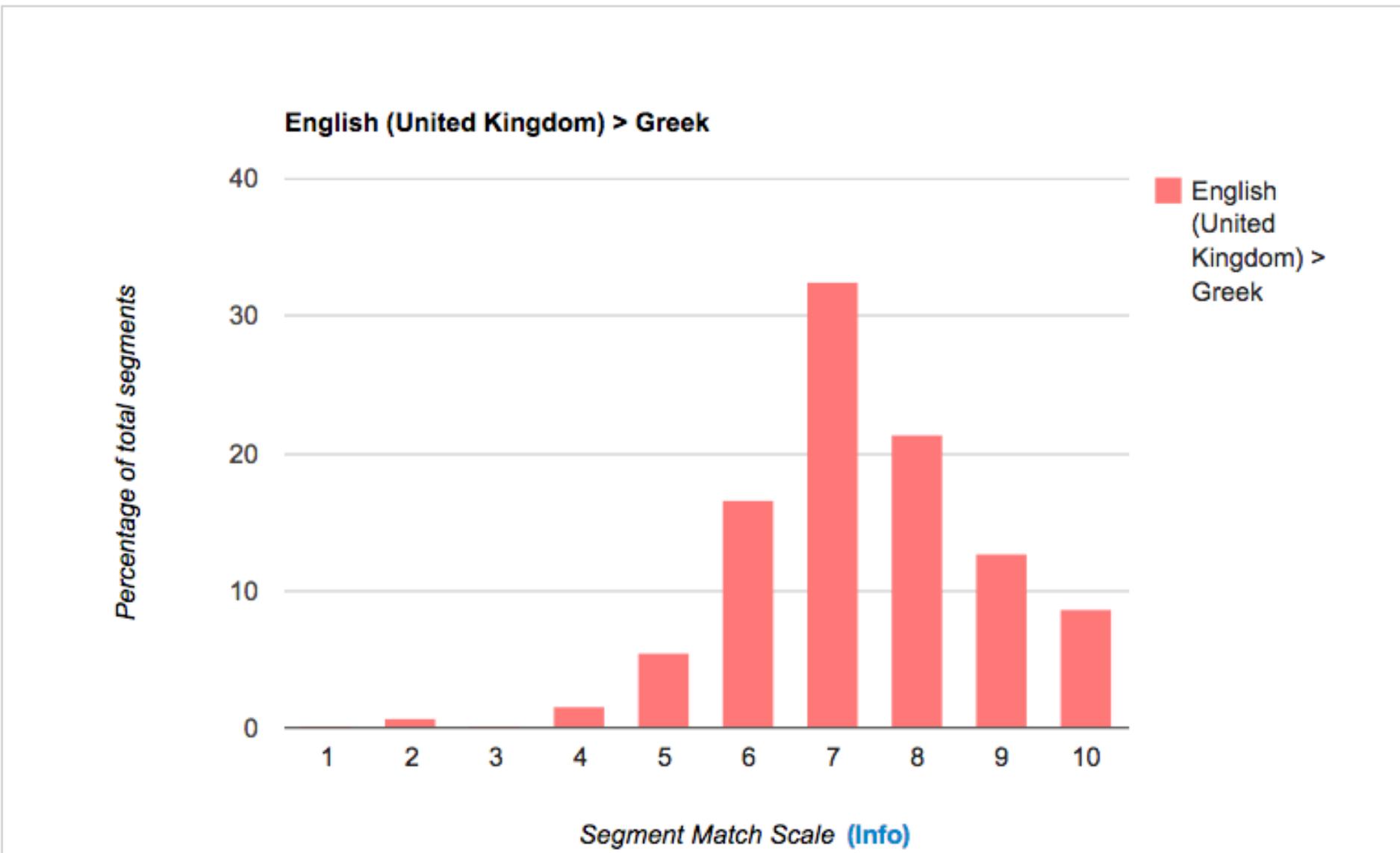
Average Productivity [\(Info\)](#)

Language Pair	Number of Segments	Number of Words	Post-edit (WPH)	Time Spent for Post-edit (seconds)	MT Engine
English (United Kingdom) - Greek	126	2,941	838	12,621	Not Specified

Average Time Spent by Sentence Length [\(Info\)](#)



Edit Distance Graph [\(Info\)](#)



Edit Distance

Levenshtein distance

The Levenshtein distance calculates how many operations are necessary to modify one sentence into another one. The number of single ***character edits*** (insertion, deletion, replacement) needed, is called the Levenshtein distance.



Efficiency = WPH + EPH

Name	Number of Words	Time (seconds)	Words per hour (WPH)	Edit distance	Edits per hour (EPH)
Translator 1	100	120	3000	50	1500
Translator 2	150	140	3857	80	2057
Translator 3	80	120	2400	70	2100
Translator 4	120	130	3323	30	831



Min-Max Normalization

$$X' = \frac{X - \min_{WPH}}{\max_{WPH} - \min_{WPH}} (\text{new}_{max} - \text{new}_{min}) + \text{new}_{min}$$

To normalize the value 3000 to a new range [0.0, 1.0] the following should be calculated,

$$X' = \frac{3000 - 2400}{3857 - 2400} (1.0 - 0) + 0 = \frac{600}{1457} 1.0 = 0.411$$

So, by min-max normalization, the value 3000 in the WHP metric will be transformed to 0.411.



Normalized scores & Efficiency Score

Name	WHP normal	EPH normal.	Sum	Efficiency Score
Translator 1	0.411	0.527	0.938	0.469
Translator 2	1.0	0.966	1.966	0.983
Translator 3	0.0	1.0	1.0	0.5
Translator 4	0.633	0.0	0.633	0.317



Post-editor profiles

Evaluate Post-editors

The evaluation of the translators is based on the following selection:

Content Type: Website Content

Industry: Computer Software

Source Language: English (United States)

Target Language: Dutch (Netherlands)

Evaluator Name	Email	Number of Segments	Number of words	Time Spent for Post-edit (seconds)	Words per Hour	Total Edit distance	Score
User 1	user1@email.com	4	11	59.01	671.07	20	1
User 2	user2@email.com	4	11	61.48	644.11	8	0.33
User 3	user3@email.com	4	11	66.96	591.4	15	0.29



Post-editor profiles

	WHP	Post-editing	PE quality	Post-editing type	Result
1	High	High	Full	Fast & Aggressive	Good
2	High	Low	Light	Fast & Passive	Good
3	High	High	Light	Fast & Aggressive	Average
4	High	Low	Full	Fast & Passive	Average
5	Low	High	Full	Slow & Aggressive	Average
6	Low	Low	Light	Slow & Passive	Average
7	Low	High	Light	Slow & Aggressive	Not suitable
8	Low	Low	Full	Slow & Passive	Not suitable



Limitations and further work

- ▶ More data for benchmarking
- ▶ From relative to absolute scores
- ▶ 0 score theoretically possible = discouraging
- ▶ Eliminating outliers
- ▶ **Additional variables to include**



Additional variables to include

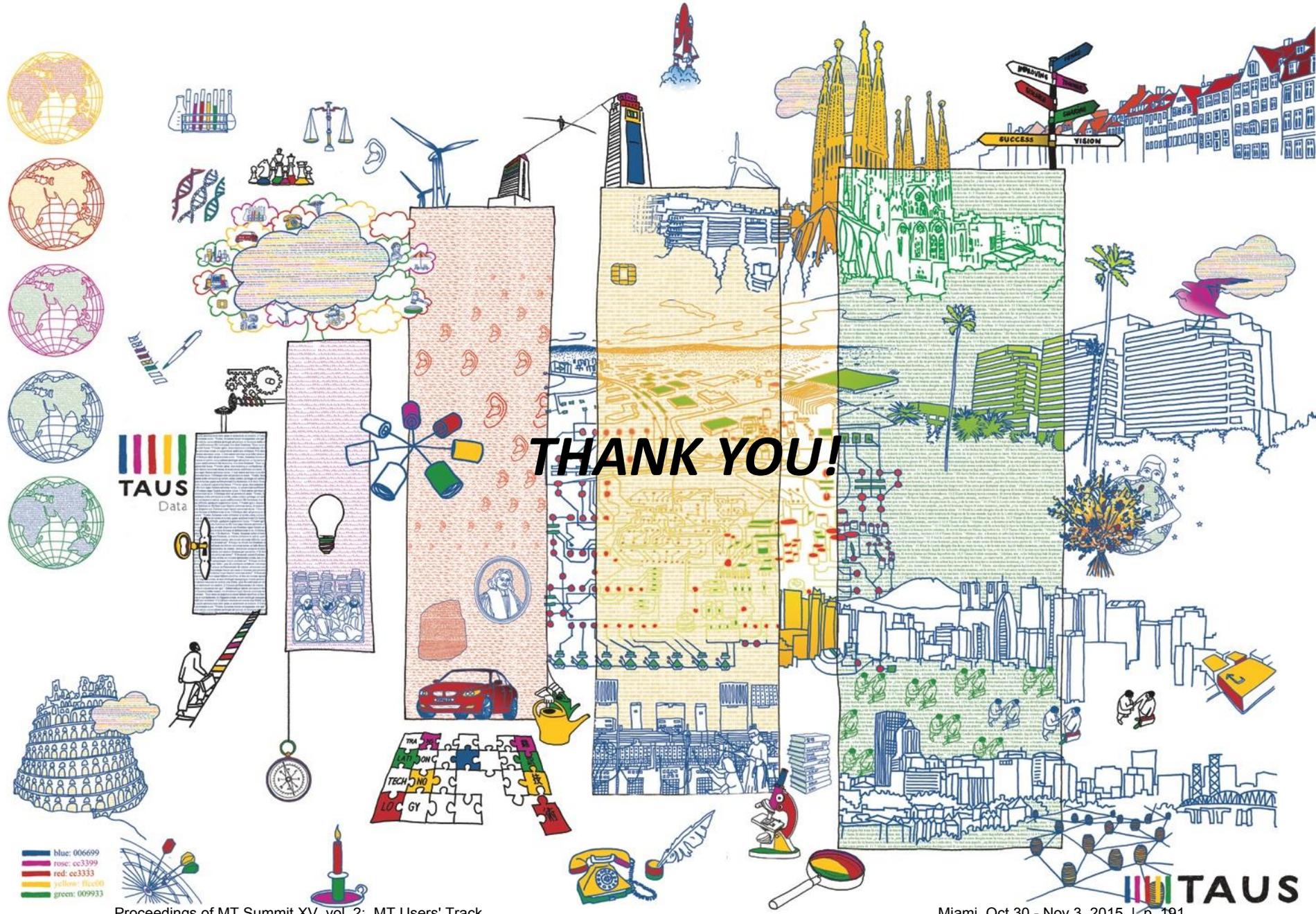
- ▶ **Keystrokes** – number of keystrokes
- ▶ **Mouse clicks** – number of clicks
- ▶ **TM fuzzy** – 0-100%
- ▶ **MT confidence** – 0-100%
- ▶ **Quality** – Review, automatic QA or manual QE
- ▶ **Difficulty of Source**
- ▶ **Experience** – number of words produced

Harmonized error-typology

DQF & MQM Harmonization

Cooperation with **DFKI** to harmonize **DQF** with **MQM** and standardize **Error categories and metrics**. A deliverable in the **EU project Q21**.





THANK YOU!

TAUS
Data

- blue: 006699
- rose: cc3399
- red: cc3333
- yellow: ffcc00
- green: 009933

