



# It's Time for a Big Idea

**It's time for the Human Language Project**

[www.taus.net](http://www.taus.net)

 **TAUS**

# We live in a time of hyper-globalization

---

- Trade integration:
  - Last 2 decades 75 of developing countries are catching up with economic frontier
- World trade is growing
- From stuff to fluff
- Democratic: openness is widely embraced
- Criss-crossing globalization

*Peterson Institute for International Economics, July 2013*

---



# Global flow will double or triple before 2025

---

*“Global flows are creating new degrees of connectedness among economies—and playing an ever-larger role in determining the fate of nations, companies, and individuals; to be unconnected is to fall behind. Our scenarios show that global flows could reach \$54 trillion to \$85 trillion by 2025, more than double or triple their current scale.”*

McKinsey, April 2014

---



# The History of Translation

**TRANSLATION:  
YESTERDAY,  
TODAY,  
TOMORROW**



# The History of Translation

**TRANSLATION:  
YESTERDAY,  
TODAY,  
TOMORROW**

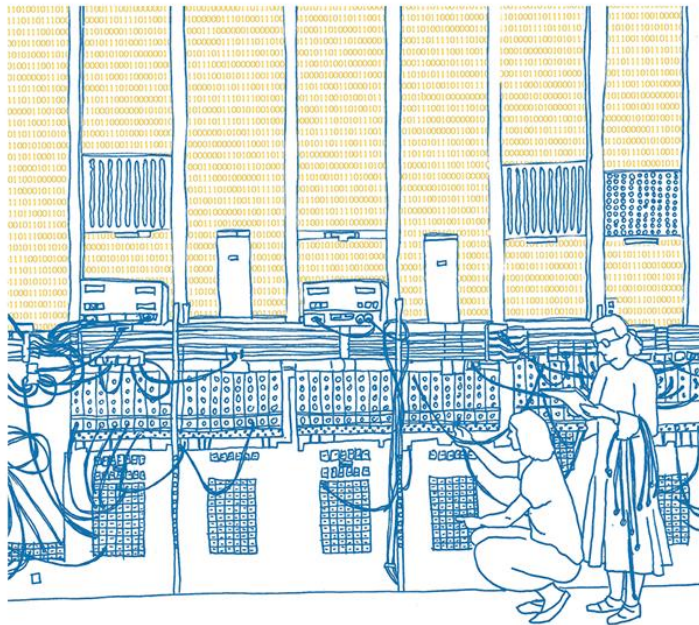


# Print Democratized Knowledge

---

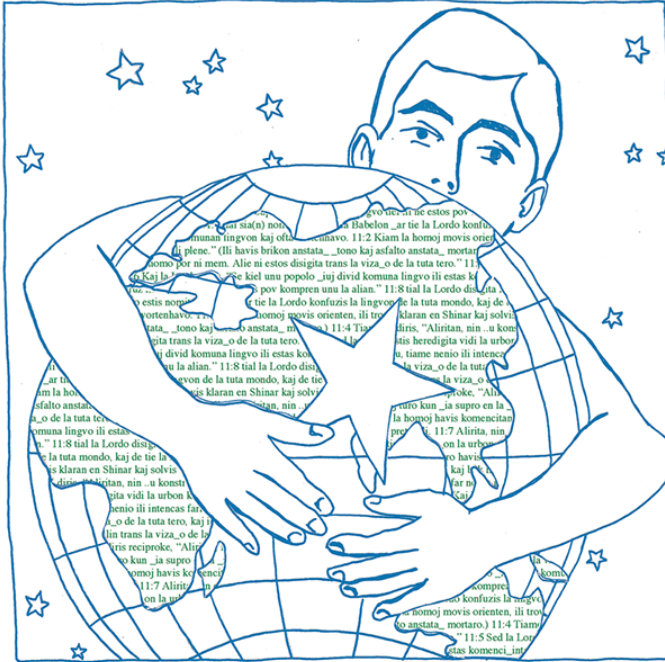


# Computers and Internet Democratized Globalization



# One Big Idea: Esperanto

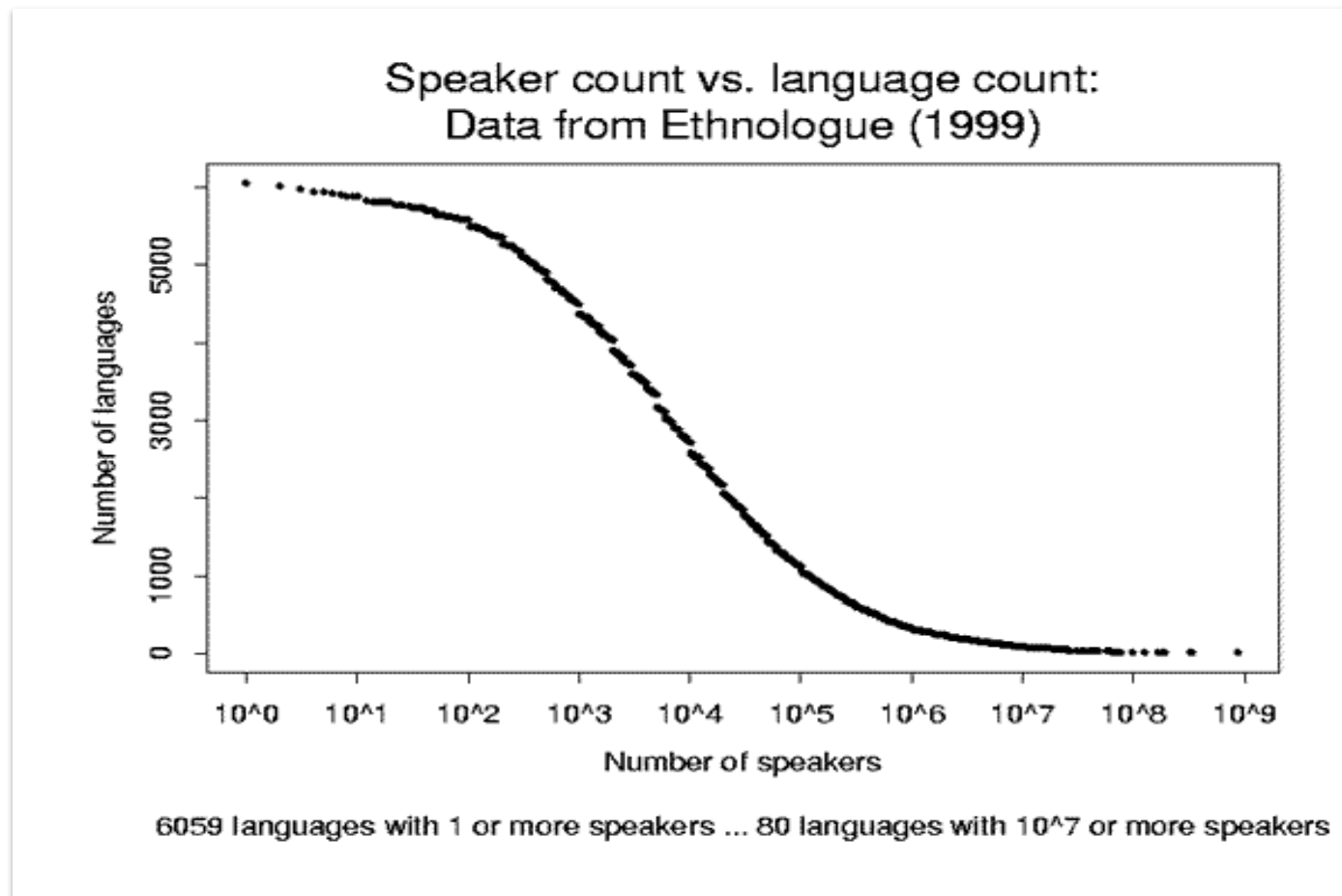
---





# Linguistic Diversity is the new Reality










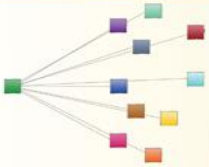
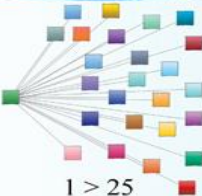
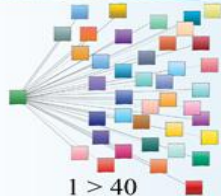
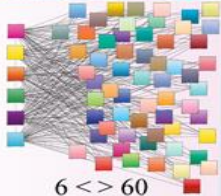

---



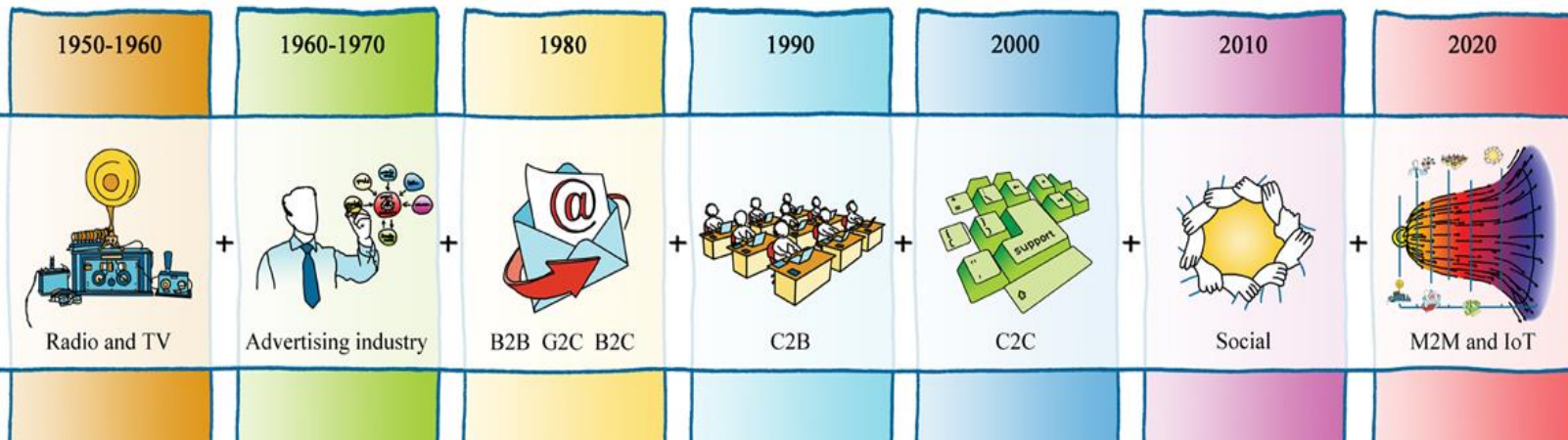
# Another Big Idea....



# Evolution of the Translation Industry











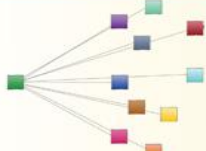
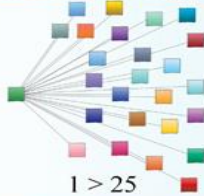
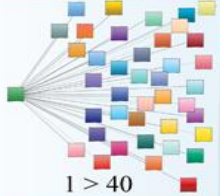
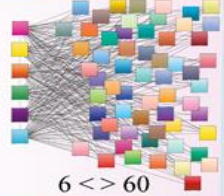





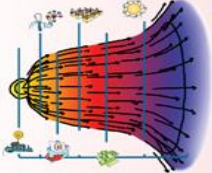
	1980 TRANSLATION	1990 LOCALIZATION	2000 GLOBALIZATION	2010 INTEGRATION	2020 CONVERGENCE
INDUSTRY FOCUS	 Documents	 Software	 Simship	 Integration in enterprise systems	 Embedded in every app, on every screen
TYPE OF CONTENT	 Paper	 Digital	 Static Web	 Dynamic Web	 Personalized
LANGUAGES	 $1 > 10$	 $1 > 25$	 $1 > 40$	 $6 <> 60$	 $150 <> 150$
TECHNOLOGY	None	TM and terminology software	Workflow (GMS)	MT and Advanced Leveraging	Real-time customized MT
DATA	Glossaries	Project TMs	Centralized TMs	Limited shared data	Unlimited big data

# Big Bang of Content and Communications





# Evolution of the Translation Industry

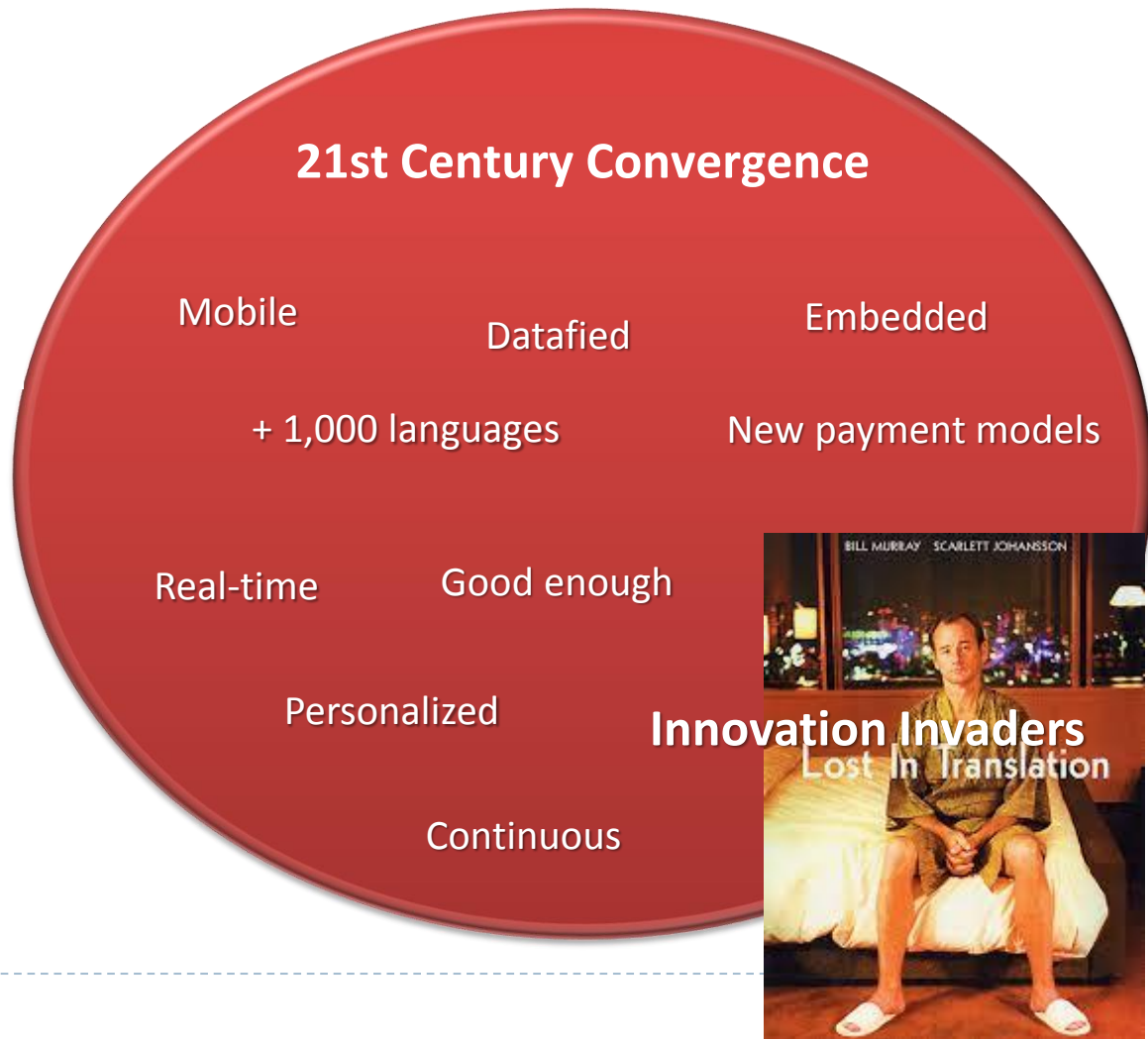
	1980 TRANSLATION	1990 LOCALIZATION	2000 GLOBALIZATION	2010 INTEGRATION	2020 CONVERGENCE
INDUSTRY FOCUS	 Documents	 Software	 Simship	 Integration in enterprise systems	 Embedded in every app, on every screen
TYPE OF CONTENT	 Paper	 Digital	 Static Web	 Dynamic Web	 Personalized
LANGUAGES	 $1 > 10$	 $1 > 25$	 $1 > 40$	 $6 <> 60$	 $150 <> 150$
TECHNOLOGY	None	TM and terminology software	Workflow (GMS)	MT and Advanced Leveraging	Real-time customized MT
DATA	Glossaries	Project TMs	Centralized TMs	Limited shared data	Unlimited big data
COMMUNICATIONS	 B2B G2C B2C	 C2B	 C2C	 Social	 M2M and IoT

# Translation Shifts Gears

From 10,000 customers who buy translation as a 'luxury' product to 6 billion users who consider translation 'free'.



Publisher-driven translation industry



# Is the Translation Industry Ready?

---



We Know  
Better



Together  
We Know  
More

**Who do we (translation operators) work for?  
For companies and governments?**

---



# .... or do we work for Global Citizens?

---

## TAUS MEETS “GLOBAL CITIZEN” TARS

*Moscow, May 24, 2014*

*Tars speaks Bashkir & Russian*

*Drives a taxi in Moscow*

*Plays the qurai in his free time*

*Relies on Yandex for his translations*



Will Tars find the time and ‘go the extra mile’ to learn English?

.... or will he continue to rely on Yandex to communicate with his fellow global citizen?

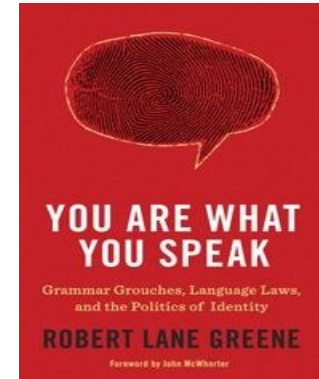
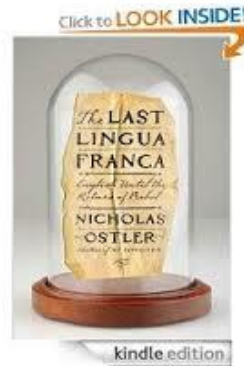
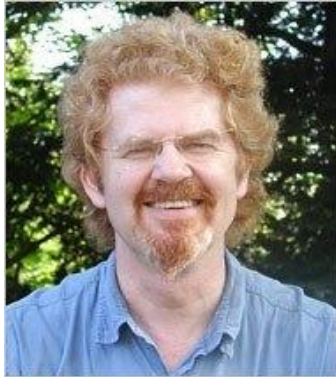
---





# English or Machine Translation as the New Lingua Franca

---



**More and more people will learn to speak English and English will become the lingua franca of the world.**

**43,3%**

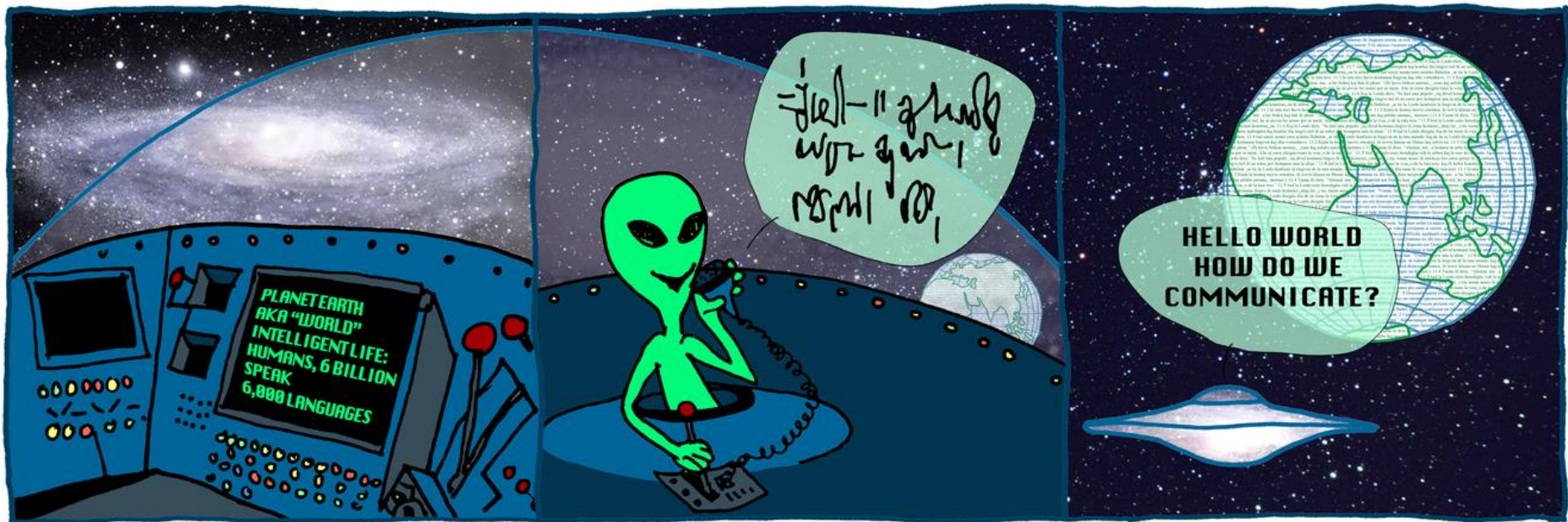
**Linguistic diversity is the new reality and MT technology will become the lingua franca of the world**

**56,7%**



# Hello World! How do we communicate?

---



**It's Time for another... Big Idea!**

---



## IT'S TIME FOR A BIG IDEA: THE HUMAN LANGUAGE PROJECT

05 December 2013



It's time for a big idea. It's time for the Human Language Project. A hundred million Euro investment over the next five years in the Human Language Project will boost the language services sector to a hundred billion Euro plus industry, create the conditions for a truly open market in the European Union and spur growth in international trade and ecommerce.

Machine Translation (MT) has come a long way helping people with their day-to-day simple communications across language barriers. The Human Language Project will establish MT as the new lingua franca realizing effective and adequate communication across the eighty languages spoken in Europe in first instance and across a thousand or more languages spoken in the world in the next phase.

The Human Language Project is an open platform of language resources and tools, consisting of at least (and maybe more):

1. Fearless sharing of language and translation data (speech and text) in all languages and language pairs, not hindered by outdated copyright law. European legislators must modernize copyright regulations on translation data. (See [TAUS article published](#) in January 2013),
2. A library of translation, language and reordering models covering all languages and a wide scope of domains to help fast-track and fine-tune the development and customization of machine translation engines.
3. A translation quality evaluation platform to help assess, benchmark and predict the right translation quality for different content types and different purposes of communication.
4. A library of language tools – such as parsers, chunkers, lemmatizers, taggers – to assist service and technology providers to improve and customize their solutions.
5. Common translation web services API's to ensure that all services and technologies work seamlessly together.

---

## Human Language Project

Collaboration between business,  
government and academia  
worldwide

### ***Think of:***

- ✓ *the Human Genome Project: a \$3.8 Billion investment in sharing data about the human genome drove \$796 Billion in economic impact, and spurred growth in the life sciences industry.*



---

**Translation is  
Becoming a  
Utility**



---

**MT as the New  
Lingua Franca**





# Ingredients of Human Language Project

---



## The Human Language Project consists of (at least):

1. **Fearless sharing of language and translation data** (speech and text) in all languages and language pairs, not hindered by outdated copyright law. European legislators must modernize copyright regulations on translation data. (See [TAUS article published](#) in January 2013)
2. **A library of translation, language and reordering models** covering all languages and a wide scope of domains to help fast-track and fine-tune the development and customization of machine translation engines.
3. **A translation quality evaluation platform** to help assess, benchmark and predict the right translation quality for different content types and different purposes of communication.
4. **A library of language tools** – such as parsers, chunkers, lemmatizers, taggers – to assist service and technology providers to improve and customize their solutions.
5. **Common translation web services API's** to ensure that all services and technologies work seamlessly together.

# Building Blocks

---





# TAUS Objectives

---

Our **aspiration** is *to enable better translation*, to help the world communicate better.

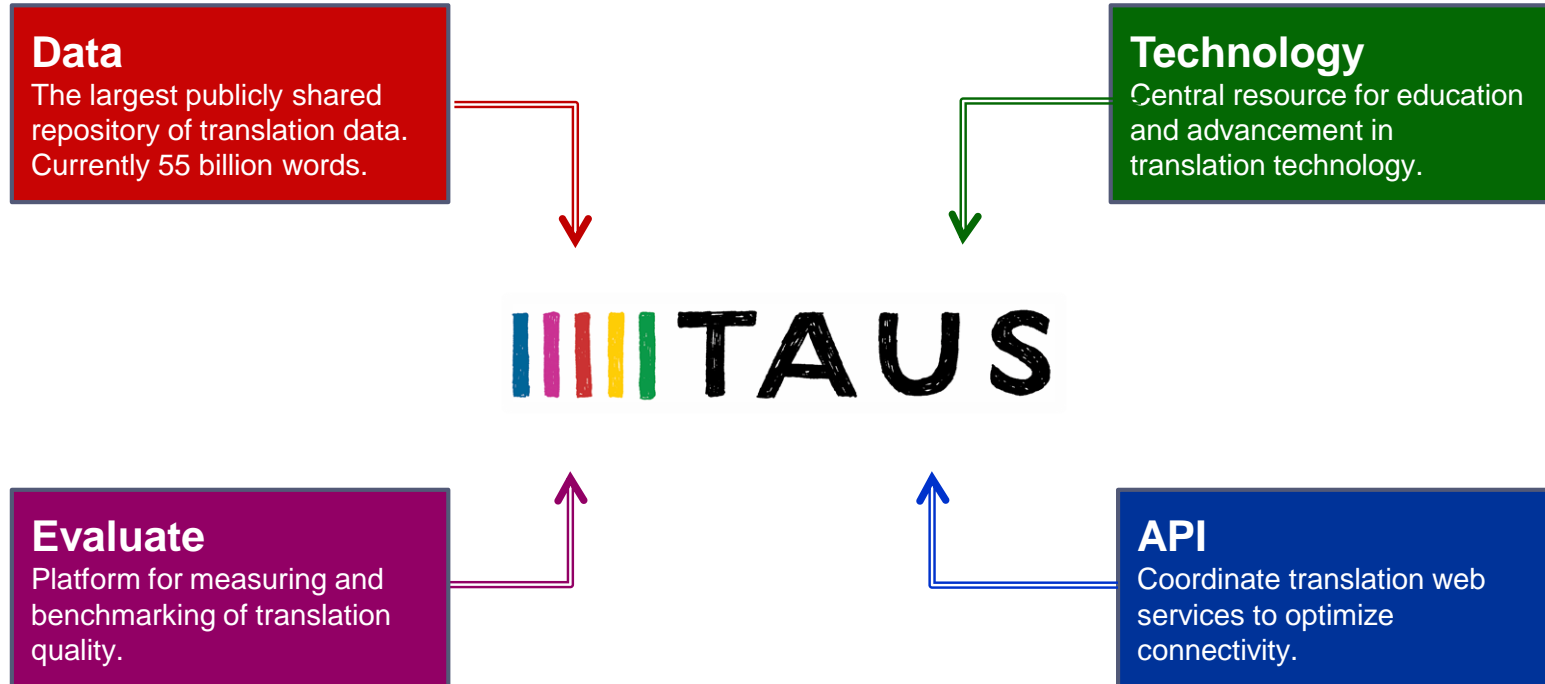
Our **business** is to support buyers and providers of translation *with industry-shared resources and technology*.

---

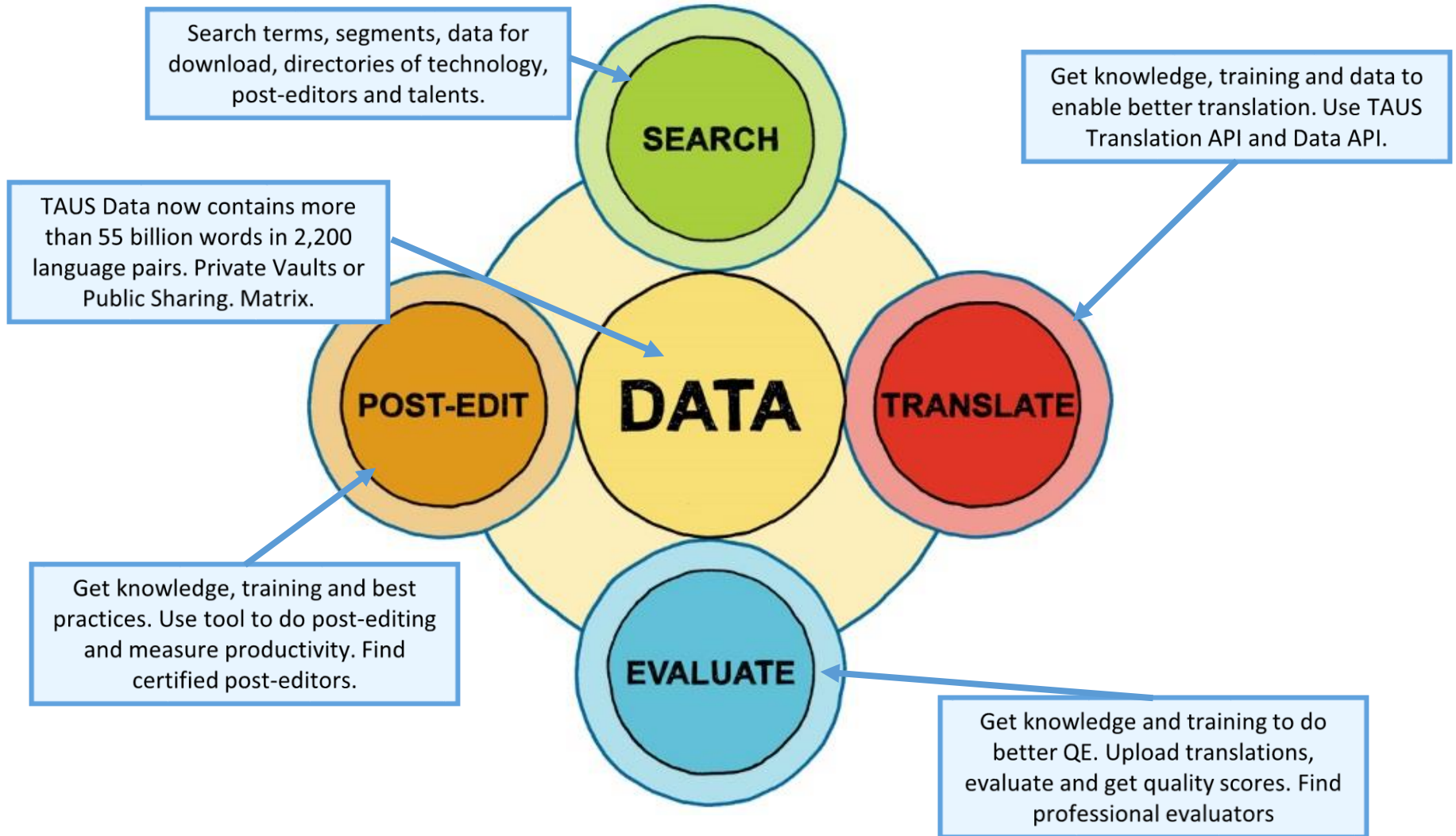


# Four Action Lines

---



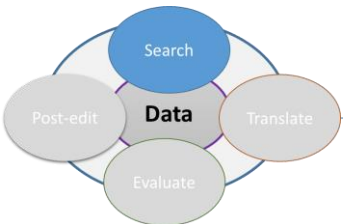
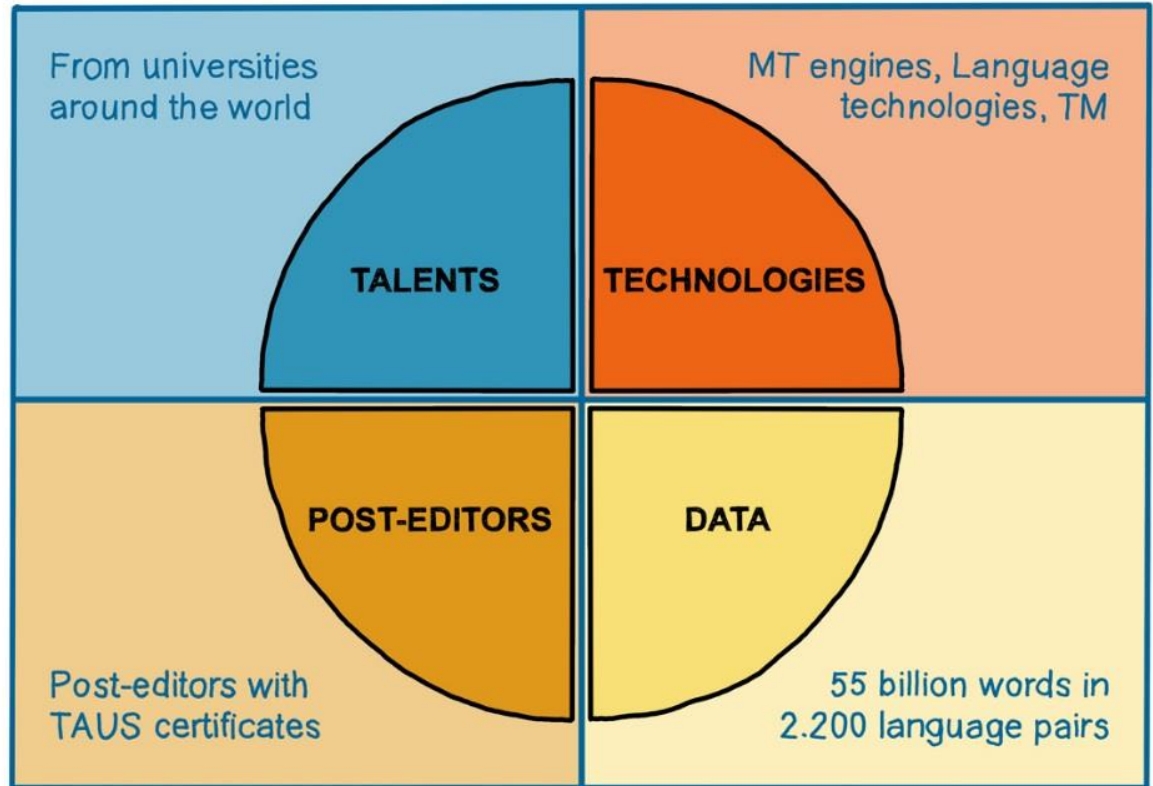
# TAUS Service Overview



# Search

## Use scenarios

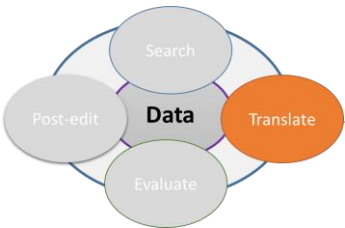
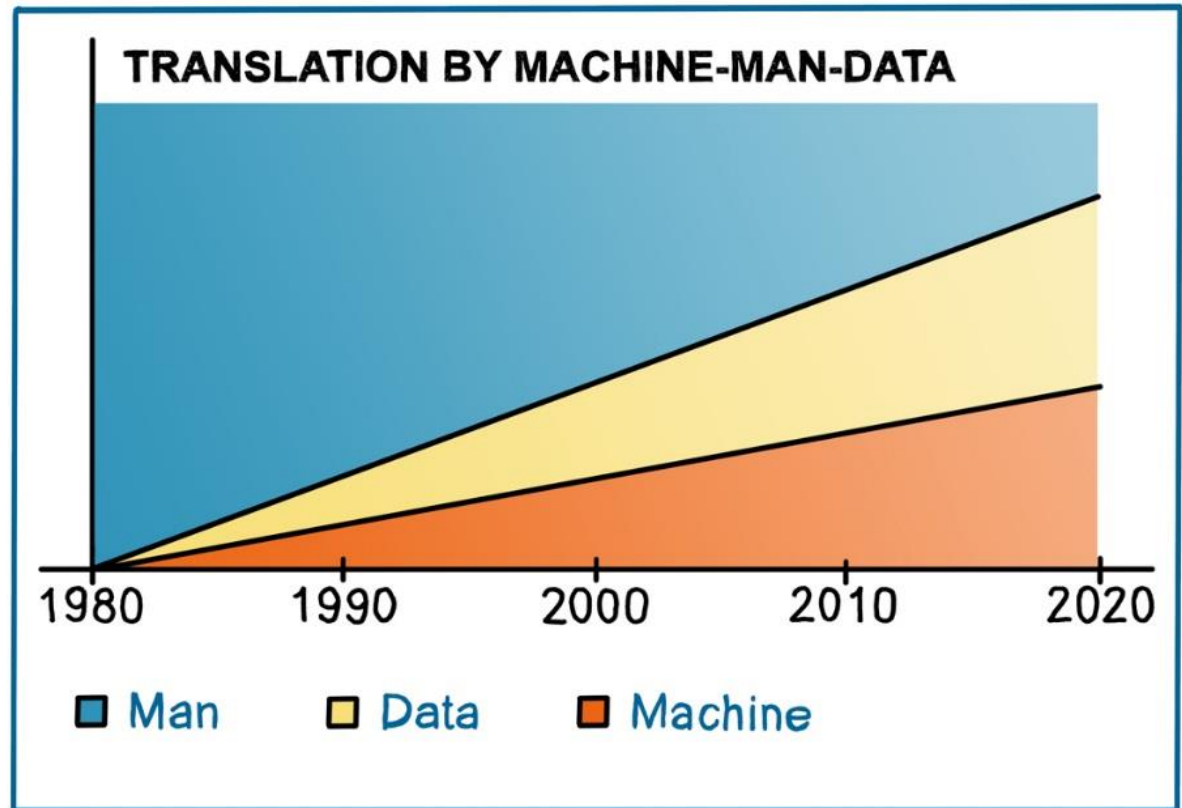
- ▶ TAUS Data API
- ▶ Search TMs
- ▶ List software and MT engines
- ▶ Find tools and MT engines
- ▶ Find post-editors
- ▶ Find specialists
- ▶ Get data



# Translate

## Use scenarios

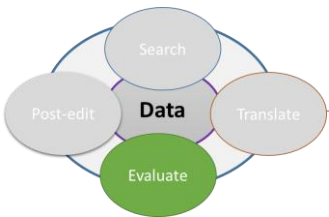
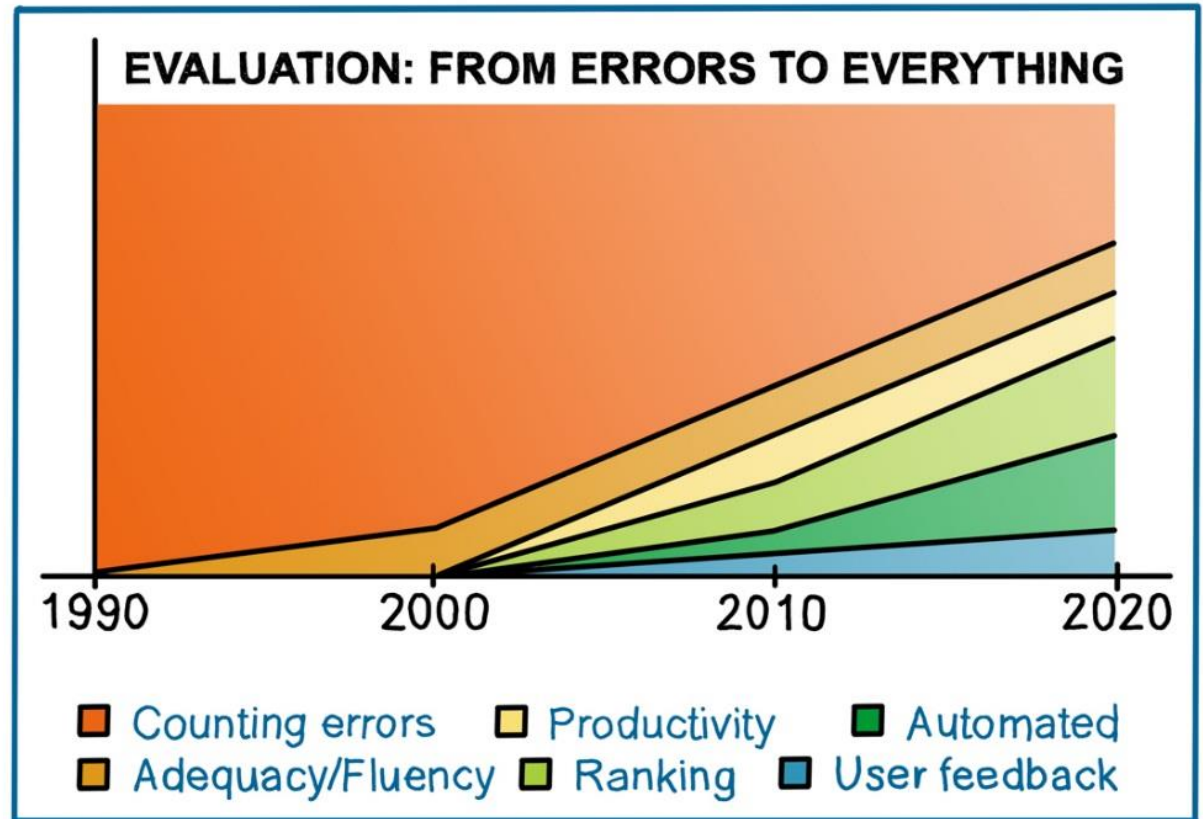
- ▶ Knowledge Base
- ▶ MT Users Call
- ▶ Translate Community
- ▶ Question-answering
- ▶ Training
- ▶ Translation API



# Evaluate

## Use scenarios

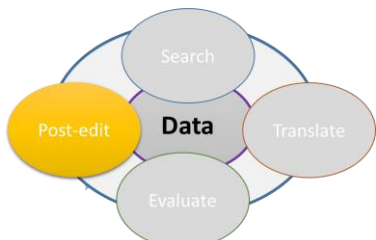
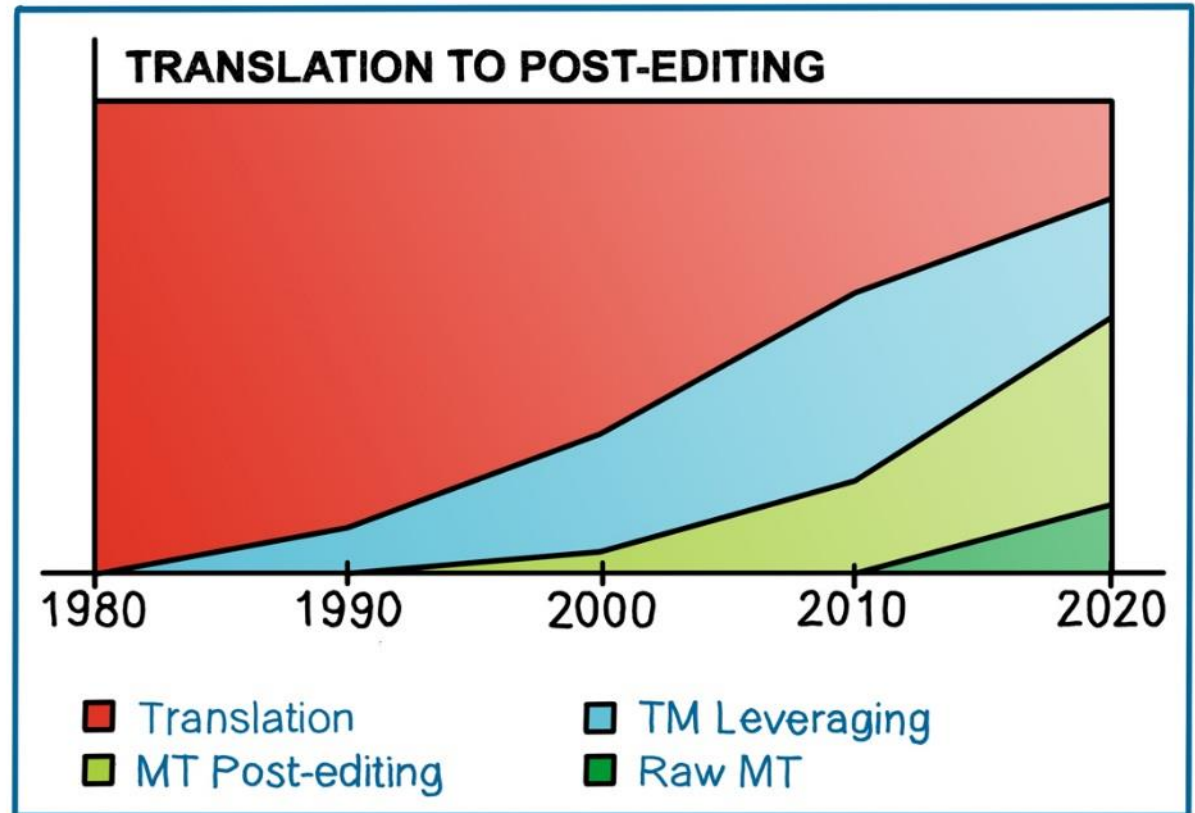
- ▶ Knowledge Base
- ▶ Content profiling
- ▶ Evaluate Community
- ▶ Question-answering
- ▶ Professional evaluators
- ▶ Training



# Post-edit

## Use scenarios

- ▶ Knowledge Base
- ▶ Post-editing Community
- ▶ Question-answering
- ▶ Professional post-editors
- ▶ Training



# New TAUS Investments

---

## 1. Free membership for universities

- ▶ Use of 40 billion words in 2,000 language pairs for research only
- ▶ Use of DQF & metrics for education

## 2. Translation QE standardization

- ▶ Dynamic Quality Framework: shared industry platform
- ▶ Work with ISO to convert best practices into standards for QE

## 3. Translation API standardization

- ▶ Work with industry to enhance and standardize Translation API

## 4. New features for Data platform

- ▶ Data Collector
- ▶ Term Extractor
- ▶ Matching Scores





# Returns on Investment

---

## 1. Reduce the shortage of skills and talents

- ▶ Free university program prepares new generation of language workers

## 2. Bring credibility, reliability and flexibility to industry

- ▶ QE standard and industry benchmarking promotes scalability

## 3. Stimulate growth and innovation

- ▶ Common translation API reduces friction and stimulates innovation

## 4. Proliferate use and development of MT

- ▶ More shared data improves the performance of MT

*Benefits both for individual companies and for the industry as a whole.*



# TAUS Events in Vancouver

---



## **TAUS Annual Conference**

Vancouver, BC (Canada)

October 27-28

### **TAUS Moses Industry Roundtable**

Vancouver, BC (Canada)

October 26 Afternoon

### **TAUS Quality Evaluation Summit**

Vancouver, BC (Canada)

October 29

### **TAUS Machine Translation Showcase**

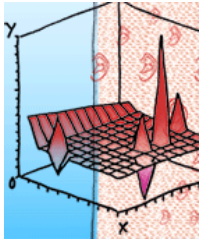
Vancouver, BC (Canada)

October 29

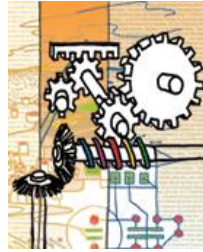


# TAUS Webinars

---



**Translation Quality  
Webinar**  
Bimonthly



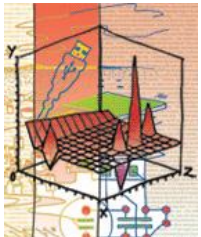
**Translation Automation  
Users Call**  
Monthly



**Post-editing Webinar**  
Weekly



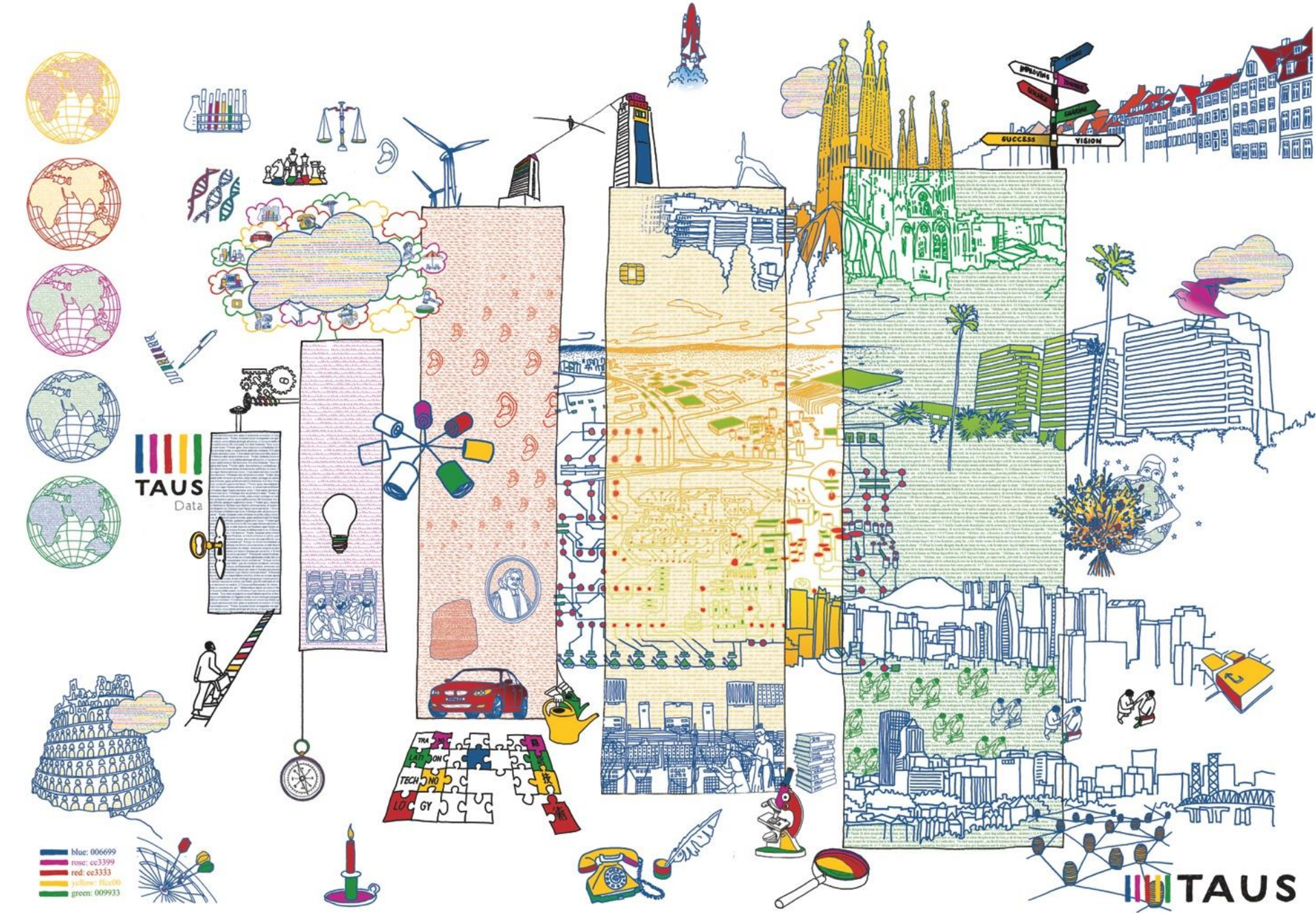
**Translation Technology  
Showcase Webinar**  
Monthly



**DQF Users call**  
Bimonthly







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

**TAUS**  
 Data

**TAUS**



This slide may not be used or copied without permission from TAUS