

Maribor

**Open Translation
a new challenge for
training and research**

Peter Sandrini

14/05/2014



presenter

Position: researcher at the translation studies department of the University of Innsbruck

Main interests: Translation of legal texts, Terminology, Translation technology, Globalization – Localization

Website: <http://www.petersandrini.net>
<http://uibk.academia.edu/PeterSandrini>



outline

- 1) the „Open“ movement
- 2) Open Translation
- 3) Open Tools
- 4) Open Formats
- 5) Open Data



the „Open“ movement

- notion of a free and open society in the humanities, pedagogy, public administration, ICT, research, technology, medicine, law ...
- open sharing of research, ideas, work ...



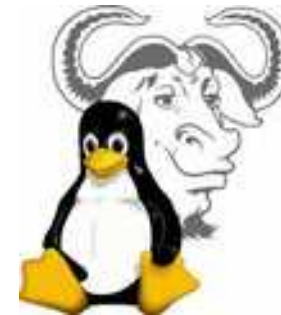
Open Movement: chronological development

- **Free Software Movement**

- cooperation in development and free use of software (ecosystem)
- GNU General Public License
- Linux, WWW

- **Open Knowledge**

- Open Access
- Creative Commons
- Copyleft
- Open Education (OER, MOOC)
- Digital Commons



- **Free Culture**

- freeculture.org, Open Society Foundation, creative commons

- **Open Government**

- Freedom of Information for public documents and processes

- **Open Data**



Open Access

- **unrestricted online access to scholarly research**
 - research can thrive only when based upon other research
 - growing costs for journal subscriptions and publishers
 - traditional publishing models are not sustainable
 - new open medium WWW
- **Austrian Science Fund (FWF)**
 - requires and supports all project leaders and co-workers to make their research results freely available



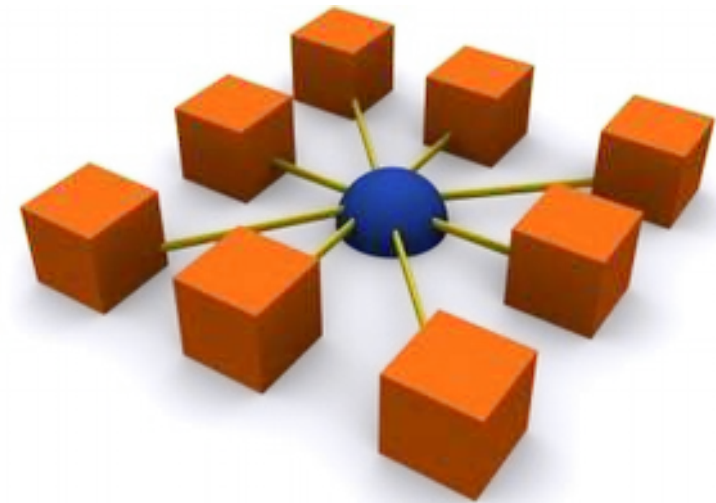
outline

- 1) The „Open“ movement
- 2) Open Translation**
- 3) Open Tools
- 4) Open Formats
- 5) Open Data

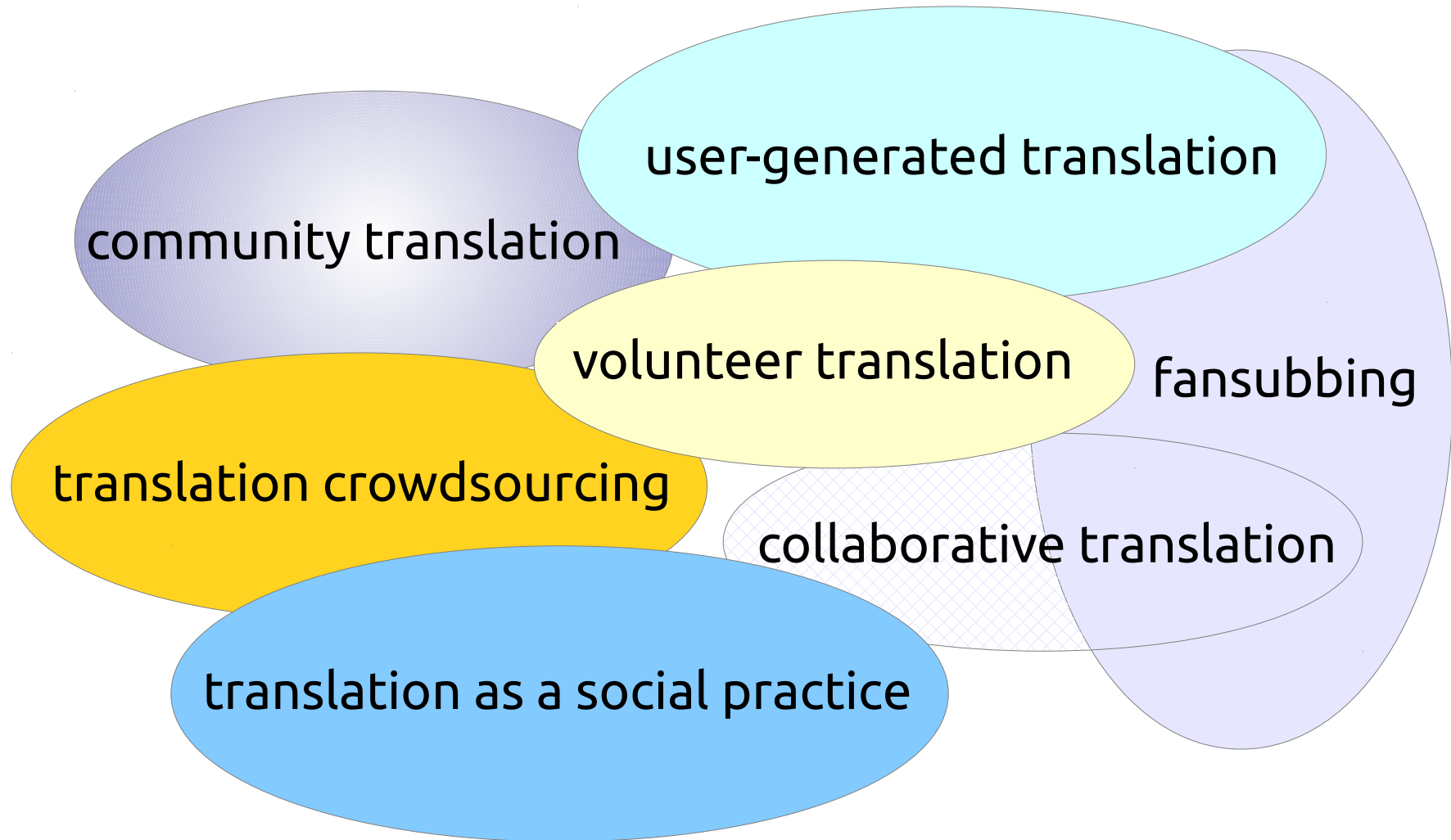


Open Translation

- according to the *FLOSS Manuals: Open Translation Tools* „a nascent field of practice emerging at the crossroads of three dynamic movements“:
 - Open/Peer Production
 - Open Content
 - Free/Libre/Open Source Software (FLOSS)
- difficult definition and delimitation (*O'Hagan 2009, 2011*)



conceptual and semantic overlap



Open Translation and FLOSS

- 1) collaborative translation of FOSS
 - examples Mozilla, OpenOffice
 - gettext framework
 - <http://opentranslators.org/>
 - <http://zanata.org/>
- 2) collaborative development and application of translation technology



Open Translation Technology

- open development of translation technology projects in:
 - <http://sourceforge.net>
 - <https://github.com/>
 - <https://code.google.com/>
 - ...
- open application and use of translation technology
 - machine translation
http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action_en.htm
 - google.translate, GTT
 - ...

Translation Technology Online

- open application and use of translation technology:
 - GTT:
 - „The Google Translator Toolkit ... is not addressed to the professional translator working in localisation, but to ... any web-enabled and motivated bilingual“ (*Garcia 2010: 8*)
 - „The ‘web-enabled and motivated’ users of this tool are, in effect, proofreading MT to the benefit of the tool provider“ (*Malmkjaer 2013:40*)



Open Translation

- 1) collaborative translation of FOSS
- 2) collaborative development and application of translation technology
- 3) collaborative translation for NGOs
„volunteer translation“,
„translation as a social practice“



Open Translation as volunteer translation

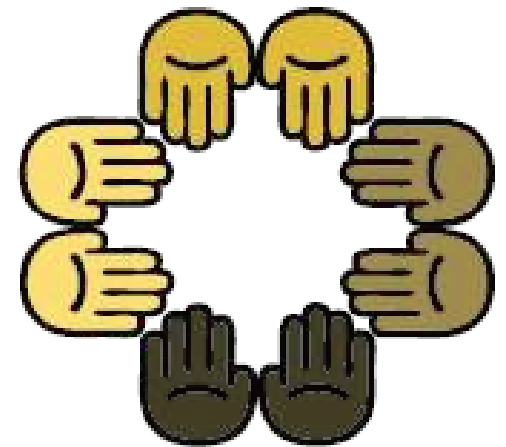
- Bey et al. (2006: 49) identify two main types of online volunteer translator community:
 - 1) *Mission-oriented translator communities*: ... strongly-coordinated groups of volunteers ... involved in translating clearly defined sets of documents.
 - 2) *Subject-oriented translator network communities*: individual translators who translate online documents such as news, analyses, and reports and make translations available on personal or group web page



NGO Translation

1) collaborative translation for NGOs

- Translators without borders
www.translatorswithoutborders.com
- The Rosetta Foundation
www.therosettafoundation.org (O'Brien, Schäler 2010)
- Mondo Lingua Initiative
www.mondo-lingua.org
- Translators and Interpreters for Solidarity, ECOS
<http://ecosteis.wordpress.com>
- Babels
www.babels.org
- ...



Open Translation

- 1) collaborative translation of FOSS
- 2) collaborative development and application of translation technology
- 3) collaborative translation for ONG
- 4) collaborative translation and subtitling



Open Multilingual Media

- collaborative translation and subtitling
fansubs, digisubs, fandubs, scanlators
(O'Hagan 2009)
- Wikipedia translation
(McDonough-Dolmaya 2011, 2012)
- TED translation
„Technology, Entertainment and Design“
a nonprofit devoted to spreading ideas
(Olohan 2014)



Open Translation

- 1) collaborative translation of FOSS
- 2) collaborative development and application of translation technology
- 3) collaborative translation for ONGsgemeinschaftliches Übersetzen von FOSS
- 4) collaborative translation and subtitling
- 5) Availability and exchange of translation data



Open Translation

- characteristics:
 - volunteer unpaid translation
 - lay translators
 - unclear context of translations
 - skopos?
 - quality management?
 - team translation?
 - workflow management
 - data exchange



Training?

- „The vision for Open Translation is predicated on the notion that anyone can be a translator“
(FLOSS-Manual Open Translation Tools)
- „demonetization and deprofessionalization of translation“ *(Olohan 2014: 18)*
- Training goals?



professionalism

according to Pfadenhauer 2005: 14

- qualification (formal training)
- disposition (service offer)
- authorization (accreditation, certification)

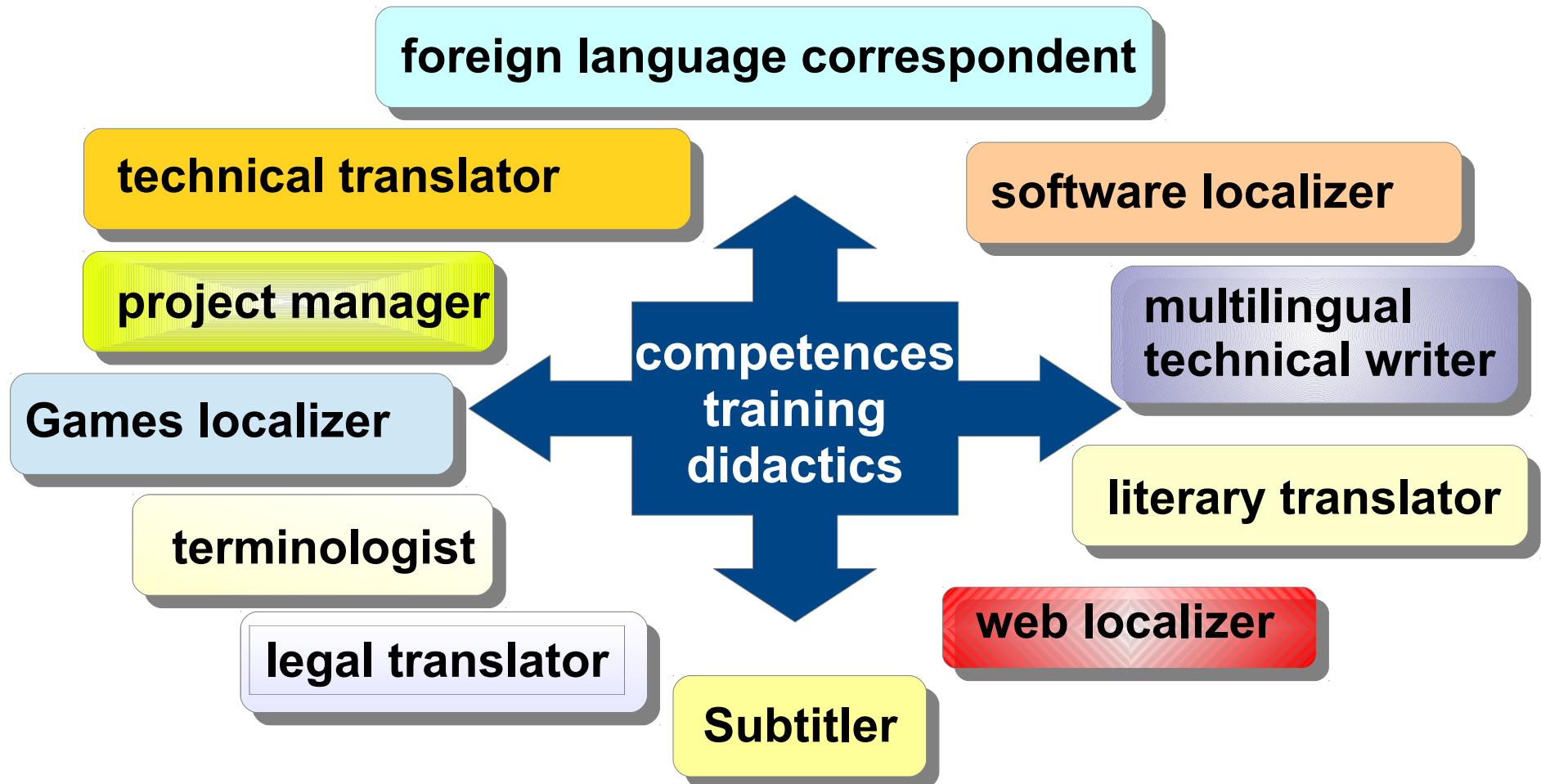


world of translation

- Diversification and Differentiation of the profession
(Adams 2013)
 - Linguistic diversification
 - terminology services
 - post-editing of mt-output
 - voice-over and subtitling
 - copywriting
 - linguistic validation
 - cross-cultural consulting
 - transcription
 - editing of non-native texts
 - online language teaching
 - ...
 - Extra-linguistic diversification
 - project management
 - cat tool consulting
 - specialization
 - web-related services
 - teaching, coaching
 - multilingual desktop publishing and OCR services
 - ...



Formal training goals?



outline

- 1) The „Open“ movement
- 2) Open Translation
- 3) Open Tools**
- 4) Open Formats
- 5) Open Data



Open Tools

- Openness of use (licensing model)
- Openness of development (source code)
- Free software
- Open Source Software

in contrast to proprietary software



FLOSS vs proprietary software structural differences

FLOSS

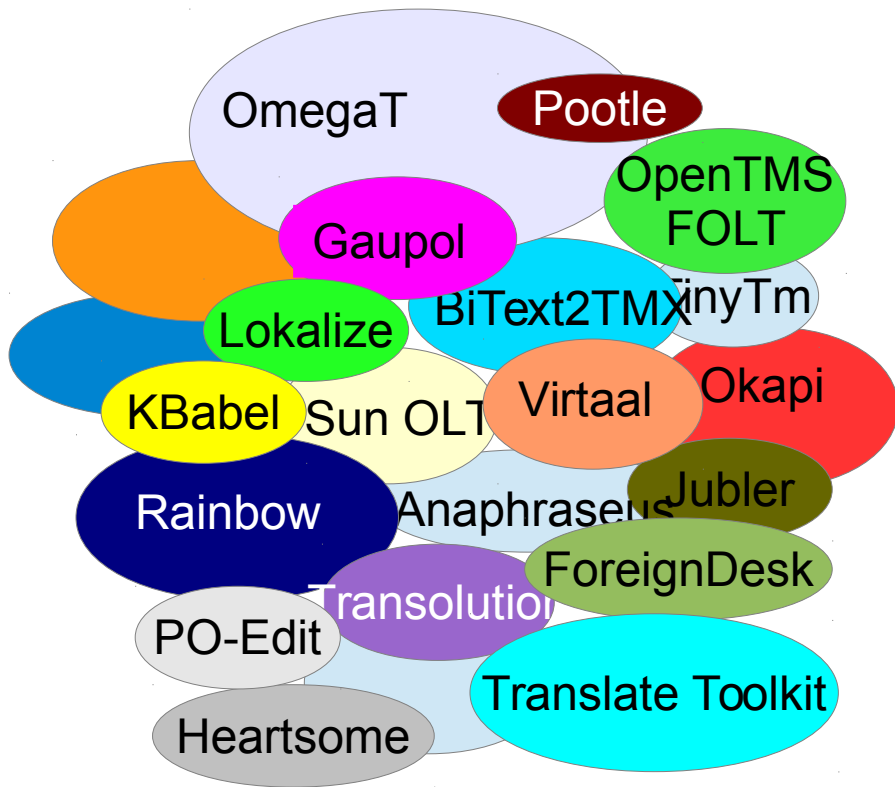
- multitude of projects
- concentrates on one function
- free licenses
- irregular updates
- support from the community of users

proprietary software

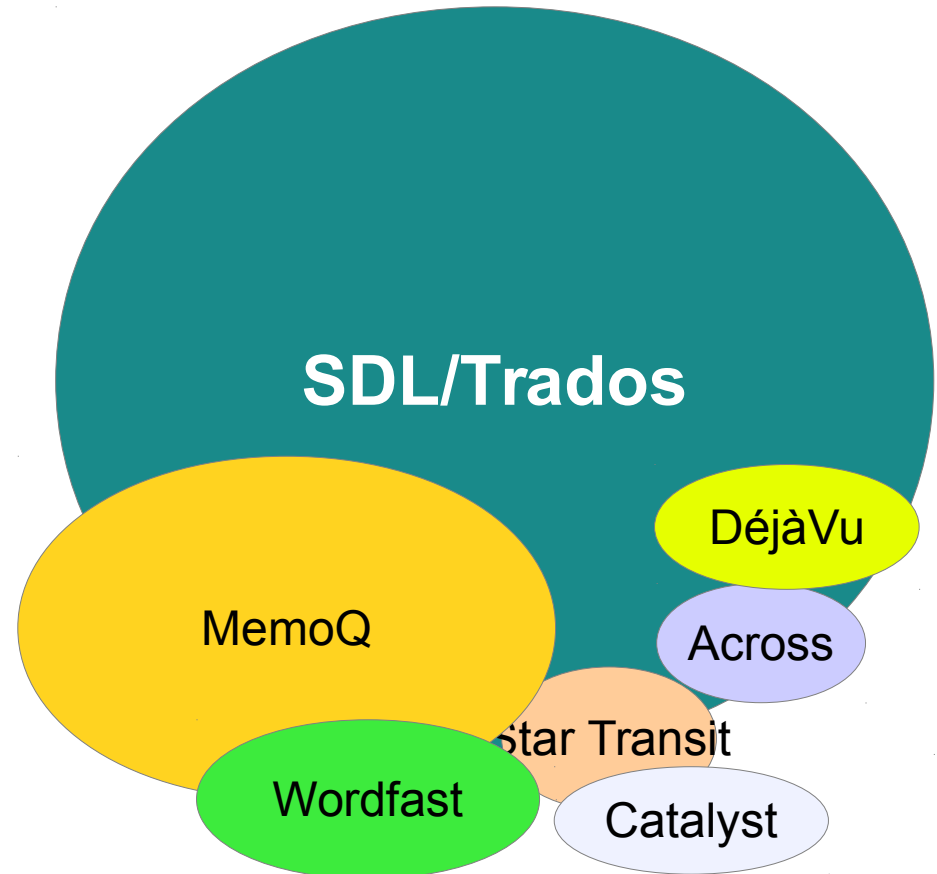
- complex applications
- comprehensive functionality
- commercial license
- constant updates
- commercial support

market reality

open



proprietary



Open Translation Tools

- translation memory
- text alignment
- software localization
- web localization
- subtitling
- machine translation
- translation workflow
- dictionaries
- terminology management
- ...



Translation Memory: text formats

OmegaT

- Open Document
- Microsoft XML
- (X)HTML
- HTML Help Compiler (HCC)
- DocBook
- text files
- Java bundle.properties
- .po monolingual
- INI
- XLIFF (Okapi, sdlxliff)
- MediaWiki (Wikipedia)

SDL-Trados

- OpenDocument
- Microsoft Office (2000-2003, 2007)
- (X)HTML
- DocBook
- RTF
- XML
- Adobe FrameMaker 8.0/9.0
- Adobe InDesign CS4
- PDF
- XLIFF
- DITA
- W3C ITS.

Why do we need FLOSS?

- to be able to do Open Translation without
 - financial barriers
 - legal barriers

In the digital era the Information and Knowledge Society needs to be able to

- access knowledge and data
- edit and modify translation data
- publish translation data
- redistribute the tools



FLOSS and Universities

„More than a acquisition of competencies the training of professional translators must be a internalization of **efficient strategies**, in the form of **flexible mental models**, which can be applied to changing situations. In other words, the capability of adaptation should be encouraged allowing students – based on previous knowledge and skills - to **compare, to abstract, to generalize**, and ultimately to decide“

„la formación de traductores profesionales tenga menos que ver con la adquisición de competencias que con la interiorización, en la forma de modelos mentales flexibles, de estrategias eficaces para actuar en situaciones cambiantes. EN otras palabras, debe estimularse una capacidad de adaptación que permita, a partir del contraste con los conocimientos y las habilidades previos, comparar, abstraer, generalizar, y en ultimo extremo, decidir.“ (Diaz-Fouces 2011: 13)

„In general, the way to advance within the profession usually involves more **conceptual control over technology**, not less. [...] Only when translators are critically aware of the available tools can they hope to be in control of their work“ (Biau/Pym 2007: 19)

FLOSS and Universities

- Teaching principles, methods and evaluation criteria
- Translation technology as a means to freedom and empowerment
- Universities should not compete with commercial suppliers of product training
- Lower access barriers to translation technology
- No dependencies from software companies
- No begging for educational or campus licenses, trial versions
- Unlimited licenses and installations



Open Tools ready to use

- USBTrans – Open Translation Tools for Windows

<http://homepage.uibk.ac.at/~c61302/fsftrans.html>



- *tuxtrans* – Open desktop system (Linux) for translators

<http://www.tuxtrans.org>



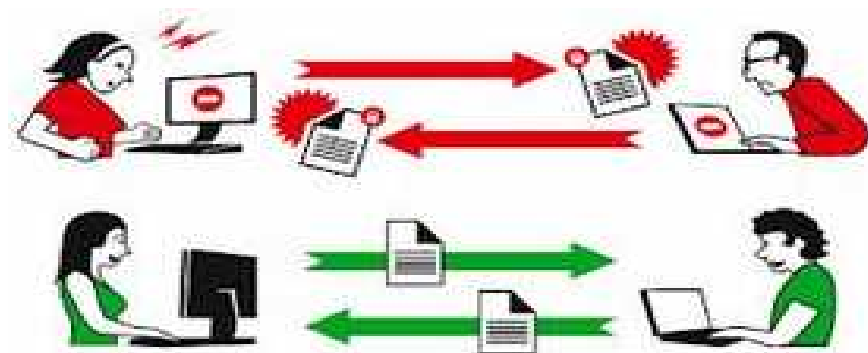
outline

- 1) the „Open“ movement
- 2) Open Translation
- 3) Open Tools
- 4) Open Formats**
- 5) Open Data



Open Formats

- The format is based on an underlying open standard
- The format is developed through a publicly visible, community driven process
- The format is affirmed and maintained by a vendor-independent standards organization
- The format is fully documented and publicly available
- The format does not contain proprietary extensions



Open Formats: benefits

- accessible data
- reusing information
- independence from producers
- transparency of data
- no viruses and malware
- interoperability

www.openformats.org



If you value the freedom to use your files

with **any** app,
on **any** device,
on **any** platform,
without lock-in,
without patent fees:

Insist on open standards!

DocumentFreedom.org



Document Freedom Day 2013
March 27th
see why it matters at documentfreedom.org



DocumentFreedomDay

LIBERATE YOUR DOCUMENTS



<http://www.documentfreedom.org>

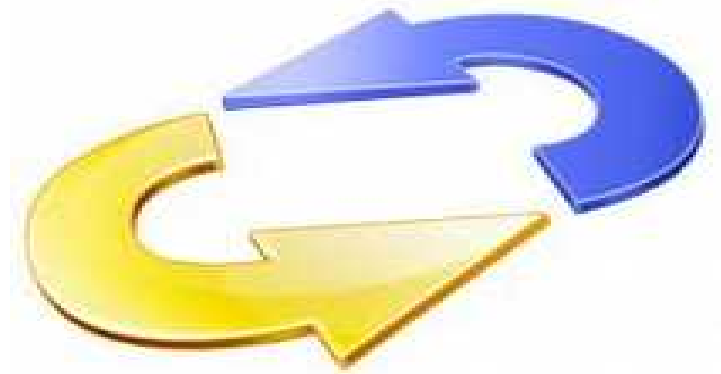
Standardized, vendor-neutral formats in translation

- clear division between format and application
- usable in different applications
- exchange of data
 - text files
 - translation memory data TMX
 - segmentation rules SRX
 - terminology TBX
 - localization files XLIFF
 - translation volume metrics GMX-V



Open Formats as translation repository

- formats which can store text segments in two or more languages
 - Translation Memory Exchange Format (TMX)
 - XML Localization Interchange File Format (XLIFF)
 - GNU gettext PO



Open Formats: structural differences

TMX

```
<tmx>
  <tu>
    <tuv>
      <seg> TP</seg>
    </tuv>
    <tuv>
      <seg> TA</seg>
    </tuv>
  </tu>
  ...
</tmx>
```

XLIFF

```
<xliff>
  <file>
    <trans-unit>
      <source> TP
      <target> TA
    </trans-unit>
    ...
  </file>
  ...
</xliff>
```

PO

```
# comments
msgid "TP"
msgstr "TA"

# comments
msgid "TP"
msgstr "TA"

...
```


Translation Memory Tools: support of open formats

OmegaT

- complete support:
 - TMX
- partial support:
 - XLIFF (Okapi, sdlxliff)
 - PO
- no support:
 - SRX

SDL-Trados

- complete support:
 - TMX
- partial support:
 - XLIFF (sdlxliff)
- no support:
 - SRX

outline

- 1) the „Open“ movement
- 2) Open Translation
- 3) Open Tools
- 4) Open Formats
- 5) Open Data**



Open Data



WIKIPEDIA
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia
- Wikimedia Shop

▼ Interaction

- Help
- About Wikipedia
- Community portal

Article [Talk](#) [Read](#) [Edit](#) [View history](#)

Open data

From Wikipedia, the free encyclopedia

Open data is the idea that certain [data](#) should be freely available to everyone to use and republish as they wish, without restrictions from [copyright](#), [patents](#) or other mechanisms of control.^[1] The goals of the open data movement are similar to those of other "Open" movements such as [open source](#), [open hardware](#), [open content](#), and [open access](#). The philosophy behind open data has been long established (for example in the [Mertonian tradition of science](#)), but the term "open data" itself is recent, gaining popularity with the rise of the [Internet](#) and [World Wide Web](#) and, especially, with the launch of open-data government initiatives such as [Data.gov](#) and [Data.gov.uk](#).

Open Data

- Open Government Data: Public money was used to fund the work and so it should be universally available
- in scientific research, the rate of discovery is accelerated by better access to data
- restrictions on data re-use create an anticommons.
- can improve productivity and efficiencies
- enhances data analytics
- considerable benefits for communities



Open Data worldwide

- World Map of Open Government Data Initiatives

<https://maps.google.com/maps/ms?ie=UTF8&oe=UTF8&msa=0&msid=105833408128032902805.00048fbba4ecb314e822>

- European Union Open Data Portal

<http://open-data.europa.eu>

- US government

<http://www.whitehouse.gov/open>

- Open Data in Austria

<http://www.data.gv.at/>

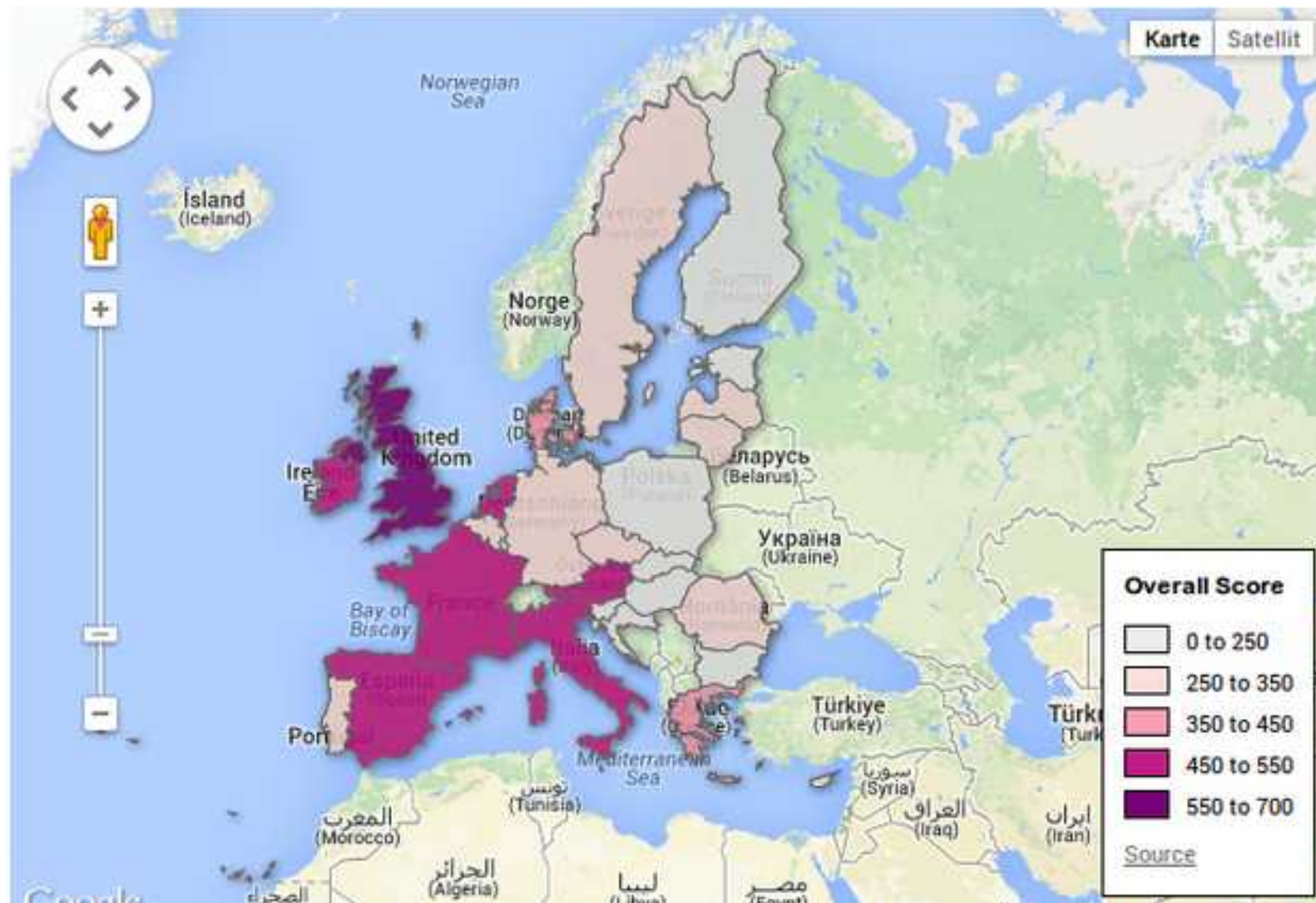
- Open Data in Germany

<http://opendata-kommunen.de/>

- ...



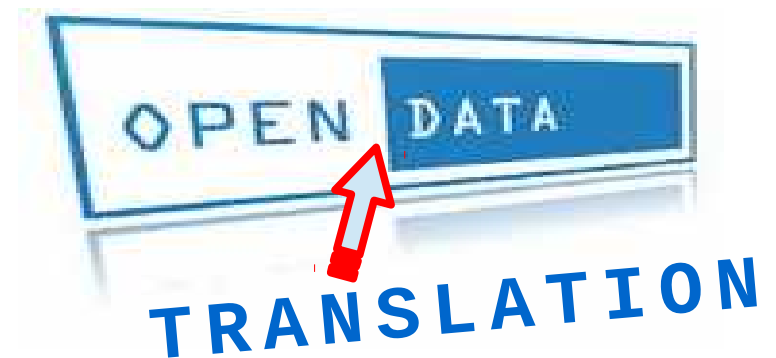
Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information



<http://www.epsiplatform.eu/content/european-psi-scoreboard>

Open Translation Data

- Availability and reuse of translation data:
 - 1) Research data about translation processes: *TAP, keystroke logging, eye tracking ...*
 - 2) Translated text or text segments (TM)**



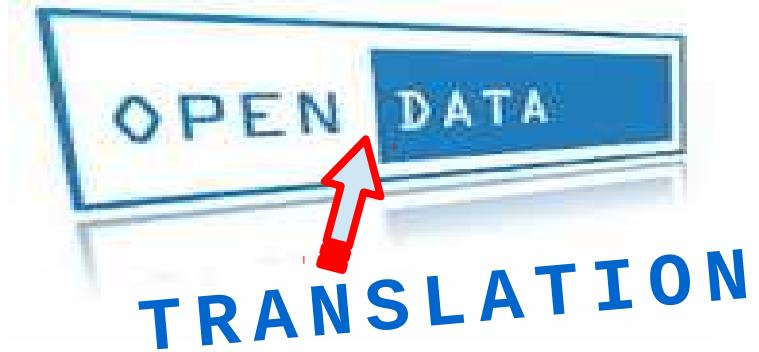
Open Translation Data

- Availability and reuse of translation data:

1) Research data about translation processes: *TAP, keystroke logging, eye tracking ...*

2) Translated text or text segments (TM)

freely accessible data storing source and target texts or text segments in a structured and machine readable open format



Open Translation Data: examples

- Open Data Euskadi
http://opendata.euskadi.net/w79-contdata/es/contenidos/ds_recursos_linguisticos/memorias_traduccion/es_izo/memorias_traduccion_izo.html
Memorias de traducción del Servicio Oficial de Traductores
- United Nations General Assembly Resolutions
<http://www.uncorpora.org/>
- ECDC-TM European Centre for Disease Prevention and Control' (ECDC)
<http://ipsc.jrc.ec.europa.eu/?id=782>
- JRC DGT



DGT-TMX



- Directorate-General for Translation

<http://open-data.europa.eu/en/data/dataset/dgt-translation-memory>

translation of the entire body of European legislation, comprising all the treaties, regulations and directives adopted by the European Union in 23 languages

- TMX-Format
- subdivided by years 2004-2012 and quarterly files
- you may extract language pairs with TMExtract

```
java -jar TMXtract.jar DE IT /home/c61302/temp/DGT-deit.tmx
```



European Union Open Data Portal

Open Translation Data: benefits

- efficiency of translations (reuse)
- increase of consistency
- dissemination of official translations
- emphasizing the role of translation
- support, acceleration and control of language planning and terminology harmonization
- ...



Summary

- ✓ Open Translation includes diverse activities: collaborative, volunteer translation
- ✓ Openness in translation technology: tools, formats, data
- ✓ Open Content in translation – translation data (PSI)
- ✓ Increased public awareness of translators through openness



Some conclusions for training and education

- education and training of translators must not be oriented purely towards current market requirements
- open translation varieties should be included
- diversification
- encourage critical thinking
- translation technology is a key competence
„convertir el manejo de tecnologías en una habilidad transversal en la formación de traductores, y no reservada a una única materia [...] esa transversalidad debería complementarse necesariamente con una apertura a la pluralidad de opciones, que lleva aparejada una imprescindible formación en criterios para elegir“
(Díaz-Fouces 2011: 14)
- not product training, but empowerment



References

- Adams, Nicole (2013): Diversification in the Language Industry. Success beyond translation. NYA Communications, Australia.
- Bey, Youcef; Boitet, Christian; Kageura, Kyo (2006): The TRANSBey Prototype: An Online Collaborative Wiki-based CAT Environment for Volunteer Translators In: E. Yuste (ed.) Proceedings of the 3rd International Workshop on Language Resources for Translation Work, Research and Training (LR4Trans-III), Paris: ELRA, 49–54.
- Diaz Fouces, Oscar (2011): ¿Merece la pena introducir el software libre en la formación de traductores profesionales? In: Universitat de Vic (ed.): CDROM Language and Translation Teaching in FacetoFace and Distance Learning (2011). Facultat de Ciències Humanes, Universitat de Vic.
- Garcia, Ignacio (2010): Is Machine Translation Ready Yet? In: Target 22(1): 7–21.
- Maeve Olohan (2014) Why do you translate? Motivation to volunteer and TED translation, Translation Studies, 7:1, 17-33
- Malmkjaer Kirsten (2013): Where are we (FromHolmes Map until now) In: Routledge Handbook of Translation Studies. Routledge, 31-44.
- McDonough Dolmaya, Julie (2011): Wikipedia survey IV (motivations). Some thoughts on translation research and teaching. <http://mcdonough-dolmaya.ca/2011/08/24/wikipedia-survey-iv-motivations/>.
- McDonough Dolmaya, Julie (2012): Analyzing the crowdsourcing model and its impact on public perceptions of translation. The Translator 18, no. 2: 167-191.
- O'Hagan, Minako (2009): Evolution of user-generated translation: Fansubs, translation hacking and crowdsourcing. Journal of Internationalisation and Localisation 1, no. 1: 94-121.
- O'Hagan, Minako (2011): Community translation: Translation as a social activity and its possible consequences in the advent of Web 2.0 and beyond. Linguistica Antverpiensia 10: 1-10.
- O'Brien, Sharon, and Reinhard Schaler (2010): Next generation translation and localization: Users are taking charge. Paper presented at Translating and the computer 32, London, November 18-19. http://doras.dcu.ie/16695/1/Paper_6.pdf.
- Pfadenhauer, Michael (2005) (ed.): Professionelles Handeln. Wiesbaden: VS Verlag für Sozialwissenschaften.



Web:

<http://www.petersandrini.net>

<http://uibk.academia.edu/PeterSandrini>