

TeTra III

Dati, formati e strumenti aperti nell'ambito della traduzione

Peter Sandrini

24/01/2014



aperto

aperto (a'perto)

aggettivo part. pass. di aprire

1. |schiuso| spalancato |chiuso non chiuso *La finestra è aperta.*

all'aria aperta - in un luogo non chiuso

vocale aperta -

linguistica pronunciata in modo più aperto rispetto alle vocali chiuse

2. luogo |ampio| vasto| spazioso che si estende in ampiezza *La villa è in aperta campagna.*

3. |accessibile a cui si è ammessi liberamente *servizio aperto al pubblico*

4. *figurato* | sincero | franco | spontaneo che si manifesta con schiettezza *un carattere aperto essere un libro aperto* -

figurato non avere nulla da nascondere

parlare a cuore aperto - con sincerità

5. |flessibile| anticonformista di ampie vedute *essere di mentalità aperta*

6. |sveglio| perspicace |acuto che capisce con prontezza *mente aperta*

7. |insoluto non definitivo, non risolto *una questione aperta*

it.thefreedictionary.com/aperto

aperto

aperto (a'perto)

aggettivo part. pass. di aprire

1. |schiuso| spalancato |chiuso non chiuso *La finestra è aperta.*

all'aria aperta - in un luogo non chiuso

vocale aperta -

linguistica pronunciata in modo più aperto rispetto alle vocali chiuse

2. luogo |ampio| vasto| spazioso che si estende in ampiezza *La villa è in aperta campagna.*

3. |accessibile a cui si è ammessi liberamente *servizio aperto al pubblico*

4. *figurato* | sincero | franco | spontaneo che si manifesta con schiettezza *un carattere aperto*
essere un libro aperto -

figurato non avere nulla da nascondere

parlare a cuore aperto - con sincerità

5. |flessibile| anticonformista di ampie vedute *essere di mentalità aperta*

6. |sveglio| perspicace |acuto che capisce con prontezza *mente aperta*

7. |insoluto non definitivo, non risolto *una questione aperta*

it.thefreedictionary.com/aperto

panoramica

- 1) traduzione 2.0 - cambiamenti
- 2) formati aperti
- 3) strumenti aperti
- 4) risorse aperte



traduzione 2.0

- tendenze:
 - digitalizzazione
 - globalizzazione
 - tecnologizzazione



digitalizzazione

- **testi digitali**
rivoluzione nella possibilità di copiare, trasportare, ricerca e riutilizzo
- **umanistica digitale**
applicazione di strumenti computazionali alla ricerca umanistica e l'uso sistematico di risorse digitali nelle discipline umanistiche



globalizzazione e traduzione

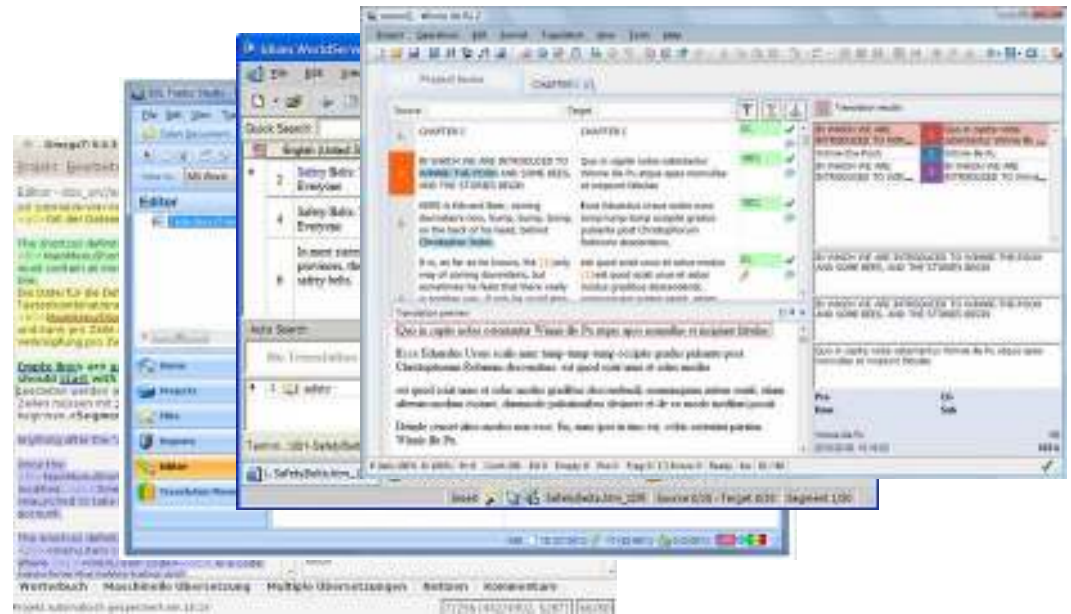
- aumento esponenziale del mercato
- possibilità di cooperazione su scala mondiale
- progetti di traduzione
- esigenze di:
qualità, tempestività,
professionalità e convenienza



tecnologizzazione

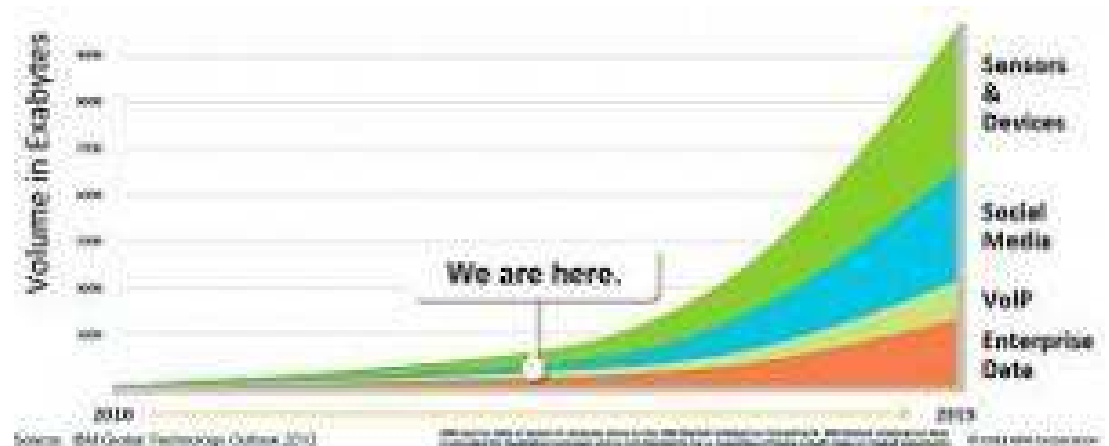
- traduzione assistita
- impiego di 'Translation Environment Tools' (TenT)
- professionalità: la traduzione professionale ed istituzionale si definisce anche attraverso l'impiego di adeguati strumenti tecnologici

"bisogno di guadagnare qualche quattrinello" (Terracini 1957)



datafication

- trasformazione in un'economia data-driven alimentata dalle molteplici fonti di big data (ad esempio social media, settore privato, pubblica amministrazione)
- datafication: "the idea that more and more businesses are dependent on their data for their business"



la traduzione



- definizione semplice:
trasposizione finalizzata di un testo di partenza (tp)
in una lingua in un testo di arrivo (ta) in un'altra
- definizione ampia:
"jede konventionalisierte, interlinguale und transkulturelle
Interaktion [...], die in einer Kultur als zulässig erachtet wird"

"ogni interazione interlinguale e transculturale
convenzionalizzata [...] che si ritiene ammissibile in una
determinata cultura" (*Prunc 1997: 108*)

'datafication' della traduzione

- **traduzione** = produzione di un testo di arrivo + dati
- **dati** = (segmenti del) testo di partenza + (segmenti del) testo di arrivo + relazioni di equivalenza

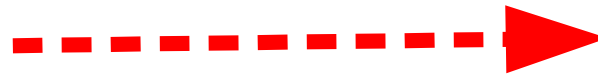


+

<i>n°</i>	<i>segmento tp</i>	<i>equivale a</i>	<i>segmento ta</i>
<i>n°</i>	<i>segmento tp</i>	<i>equivale a</i>	<i>segmento ta</i>
<i>n°</i>	<i>segmento tp</i>	<i>equivale a</i>	<i>segmento ta</i>
<i>n°</i>	<i>segmento tp</i>	<i>equivale a</i>	<i>segmento ta</i>
<i>n°</i>	<i>segmento tp</i>	<i>equivale a</i>	<i>segmento ta</i>

'datafication' della traduzione

input



output



tp

+



dati

**processo
traduttivo**

ta



+



dati

dati (traduzione)

- livelli linguistici
 - a livello lessicale ➤ dati lessicografici e terminologici
 - a livello sintattico ➤ memorie di traduzione
 - a livello testuale ➤ corpora paralleli



- a livello lessicale/terminologico

Schlitzsperre

briglia a fessura

- a livello sintattico

difesa degli argini nei bacini montani

Eliminate gender disparity in primary and secondary education, preferably by 2005, and in all levels of education no later than 2015.

Das Geschlechtergefälle in der Grund- und Sekundarschulbildung beseitigen, vorzugsweise bis 2005 und auf allen Bildungsebenen bis spätestens 2015.

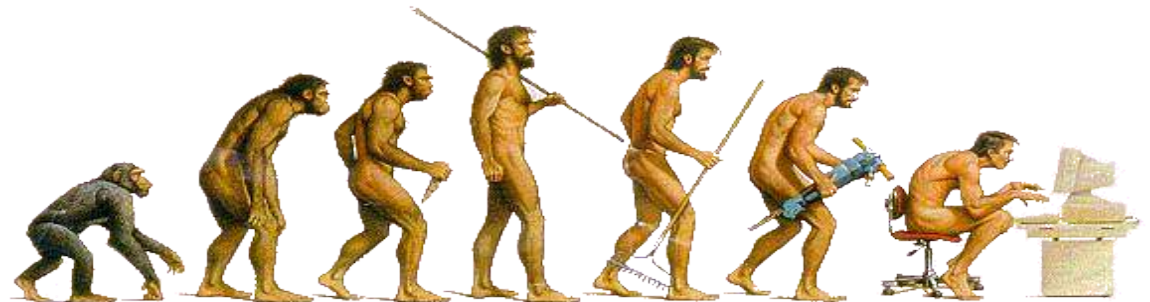
MDG dell' ONU

- a livello testuale

fr	en
<p>DIRECTIVE 95/46/CE DU PARLEMENT EUROPÉEN ET DU CONSEIL du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données</p> <p>LE PARLEMENT EUROPÉEN ET LE CONSEIL DE L'UNION EUROPÉENNE, vu le traité instituant la Communauté européenne, et notamment son article 100 A, vu la proposition de la Commission (1), vu l'avis du Comité économique et social (2), statuant conformément à la procédure visée à l'article 189 B du traité (3), (1) considérant que les objectifs de la Communauté, énoncés dans le traité, tel que modifié par le traité sur l'Union européenne, consistent à réaliser une union sans cesse plus étroite entre les peuples européens, à établir des relations plus étroites entre les États que la Communauté réunit, à assurer par une action commune le progrès économique et social en éliminant les barrières qui divisent l'Europe, à promouvoir l'amélioration constante des conditions de vie de ses peuples, à préserver et conforter la paix et la liberté, et à promouvoir la démocratie en se fondant sur les droits fondamentaux reconnus dans</p>	<p>DIRECTIVE 95/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data</p> <p>THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION, Having regard to the Treaty establishing the European Community, and in particular Article 100a thereof, Having regard to the proposal from the Commission (1), Having regard to the opinion of the Economic and Social Committee (2), Acting in accordance with the procedure referred to in Article 189b of the Treaty (3), (1) Whereas the objectives of the Community, as laid down in the Treaty, as amended by the Treaty on European Union, include creating an ever closer union among the peoples of Europe, fostering closer relations between the States belonging to the Community, ensuring economic and social progress by common action to eliminate the barriers which divide Europe, encouraging the constant improvement of the living conditions of its peoples, preserving and strengthening peace and liberty and promoting democracy on the basis of the fundamental rights recognized in the constitution and laws of the Member States and</p>

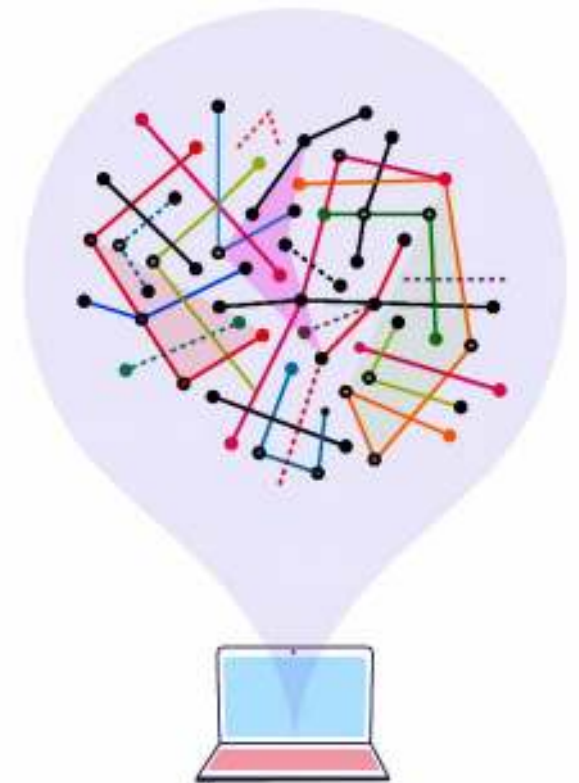
dati traduttivi - considerazioni

- depersonalizzazione della figura del traduttore
- riutilizzo delle traduzioni, e quindi aumento dell'efficienza e della produttività
- traduzione automatica statistica (SMT)
Google translate, Microsoft Translator ...
- ...



'datafication' della traduzione

- **traduzione =**
 - remix (rigenerazione creativa) in base a
 - dati (dati traduttivi) e
 - input (scopo del ta, parametri della commissione)
- **traduzione =**
 - trasformazione del tp +
 - selezione, ricombinazione, adattamento di dati



concezione aggiornata della traduzione (processo)

traduzione = applicazione e produzione di dati traduttivi nel caso concreto in funzione di determinati criteri

la qualità della traduzione = **f** (qualità dei dati traduttivi)
“gigo” (acronimo di garbage in, garbage out)



formati aperti

il formato dei dati digitali si definisce "aperto" quando ne viene resa pubblica, mediante esaustiva documentazione, la sintassi, la semantica, il contesto operativo e le modalità di utilizzo. Tali informazioni, unitamente ad una guida all'uso del formato, orientata alla lettura da parte dell'utilizzatore, devono essere presenti in uno o più documenti rilasciati dall'ente proponente lo standard.



Agenzia per l'Italia Digitale
Presidenza del Consiglio dei Ministri

<http://www.digitpa.gov.it/formati-aperti>

formati aperti: vantaggi

- dati accessibili e perenni
- indipendenza da software specifici
- assenza di dati nascosti (trasparenza)
- interoperabilità
- assenza di virus

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>

formati aperti adatti alla memorizzazione delle traduzioni

- formati standardizzati idonei a registrare segmenti di testo equivalenti in due o più lingue
 - TermBase Exchange Format (TBX)
 - Translation Memory Exchange Format (TMX)
 - XML Localization Interchange File Format (XLIFF)
 - GNU gettext PO



formati aperti: differenze strutturali

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"

...
```

strumenti aperti

- software libero (free software)
 - software open source
- } codice sorgente aperto
+ licenze aperte



in contrapposizione al
software proprietario e commerciale

open source vs software proprietario

differenze strutturali

open source

- molteplicità di progetti
- funzionalità ristretta
- licenza libera
- aggiornamenti irregolari
- supporto da parte della comunità degli utenti

software proprietario

- applicazioni complesse
- funzionalità completa
- licenza commerciale
- aggiornamenti regolari
- supporto commerciale

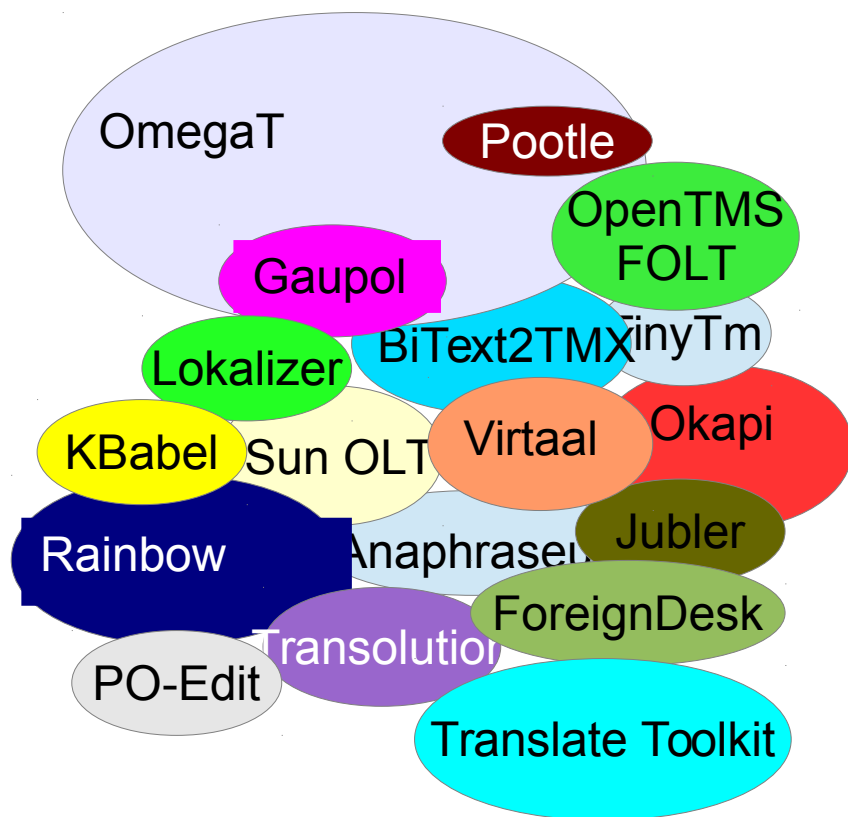
strumenti aperti: le ragioni

- necessità di strumenti aperti per poter
 - accedere
 - leggere e modificare
 - pubblicare i dati aperti
 - ridistribuire e diffondere gli strumenti assieme ai dati
- senza vincoli
 - giuridici
 - economici
 - ...

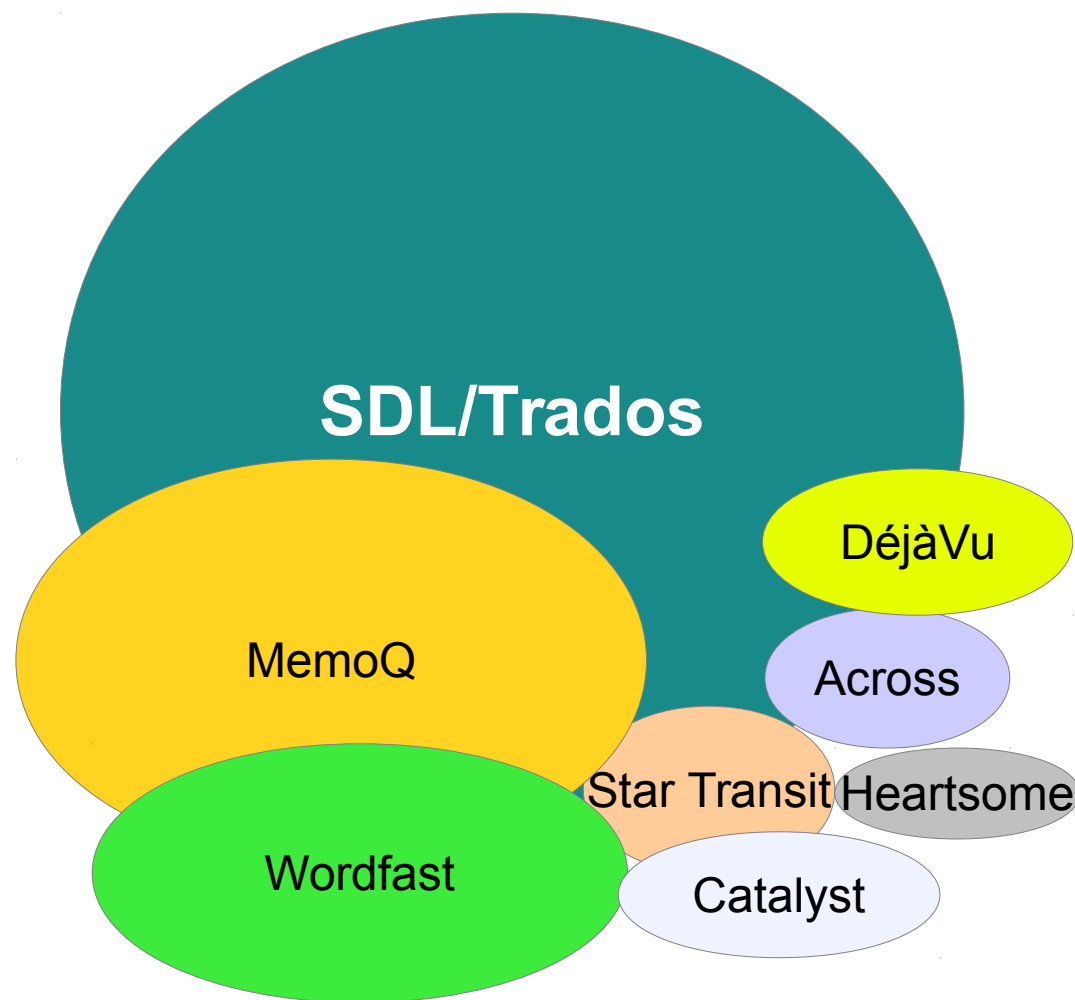


situazione mercato

ambito
'open source'



ambito commerciale



supporto per formati di testo

OmegaT

- Open Document
- Microsoft Office XML
- (X)HTML
- HTML Help Compiler (HCC)
- DocBook
- File di solo testo
- Java bundle.properties
- .po monolingua
- INI (formato "codice=value")
- 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
- XLIFF
- DITA
- W3C ITS.

supporto formati aperti

OmegaT

- completo:
 - TMX
- parziale:
 - XLIFF (Okapi, sdlxliff)
 - PO
- mancante:
 - SRX

SDL-Trados

- completo:
 - TMX
- parziale:
 - XLIFF (sdlxliff)
- mancante:
 - SRX

le nostre proposte

- strumenti aperti per la traduzione assistita
USBTrans (ambiente Windows)

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



- desktop completamente open source
per la traduzione assistita *tuxtrans* (Linux)

<http://www.tuxtrans.org>



dati aperti



WIKIPEDIA
L'enciclopedia libera

[Pagina principale](#)

[Ultime modifiche](#)

[Una voce a caso](#)

[Vetrina](#)

[Aiuto](#)

▼ [Comunità](#)

[Portale Comunità](#)

Voce [Discussione](#) [Leggi](#) [Modifica](#) [Modifica sorgente](#)

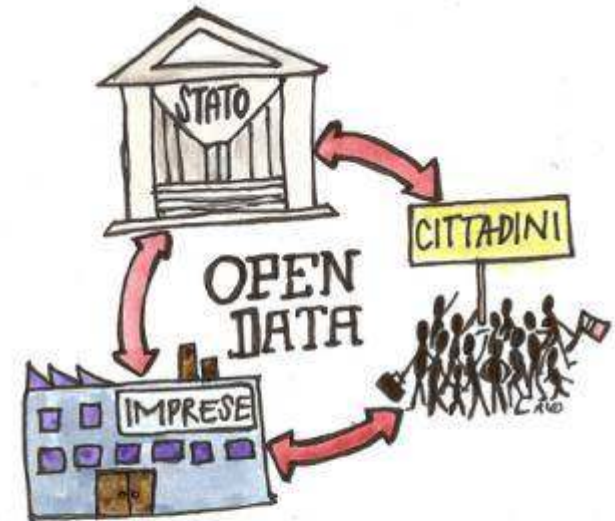
Dati aperti

Da Wikipedia, l'enciclopedia libera.

I **dati aperti**, comunemente chiamati con il termine **inglese open data** anche nel contesto italiano, sono alcune tipologie di dati liberamente accessibili a tutti, privi di brevetti o altre forme di controllo che ne limitino la riproduzione e le cui restrizioni di **copyright** eventualmente si limitano ad obbligare di citare la fonte o al rilascio delle modifiche allo stesso modo. L'open data si richiama alla più ampia

dati aperti: condizioni fondamentali

- riutilizzo e redistribuzione:
*libero utilizzo per tutti gli scopi
anche quelli originariamente non
previsti dal produttore*
- disponibilità e accesso liberi:
(prezzo, licenza, log-in ...)
- possibilità di elaborazione e modificazione:
I dati devono essere disponibili in un formato utile e modificabile



vedasi:

<http://www.opendefinition.org>

<http://opendatahandbook.org/it/>

dati aperti: ragioni

- le opere dell'ingegno finanziate (a fondo perduto) con soldi pubblici e le opere di pubblico dominio siano:
 - pubblicamente accessibili (facilmente reperibili su Internet);
 - universalmente accessibili (accessibili anche per i diversamente abili);
 - liberamente fruibili (non occorre pagare per: leggere un testo, vedere un'immagine, ascoltare una musica);
 - legalmente fruibili (l'utente è certo di poter scaricare un file nella piena legalità);
 - ottimamente fruibili (qualità digitale idonea a garantire una buona visualizzazione e/o un buon ascolto).

citato da: scarichiamoli.org



i dati aperti nel mondo

- World Map of Open Government Data Initiatives

<https://maps.google.com/maps/ms?ie=UTF8&oe=UTF8&msa=0&msid=105833408128032902805.00048bf bba4ecb314e822>

- European Union Open Data Portal

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

- I dati aperti della PA

<http://www.dati.gov.it>

- governo USA

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

- open data in Austria

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

- open data nei comuni della Germania

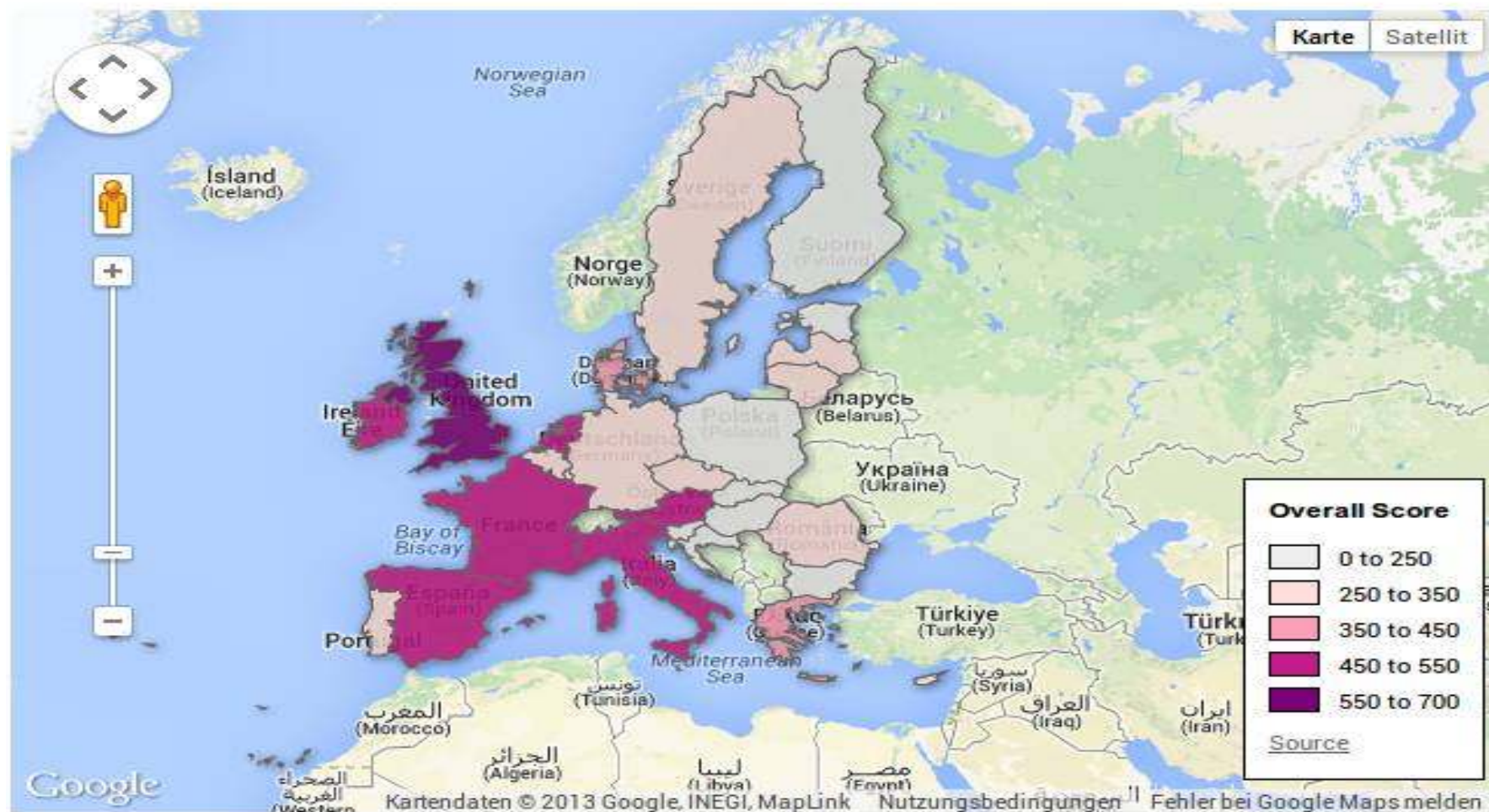
<http://opendata-kommunen.de/>

- ...



Direttiva 2013/37/UE del Parlamento europeo e del Consiglio, del 26 giugno 2013 , che modifica la direttiva 2003/98/CE relativa al riutilizzo dell'informazione del settore pubblico

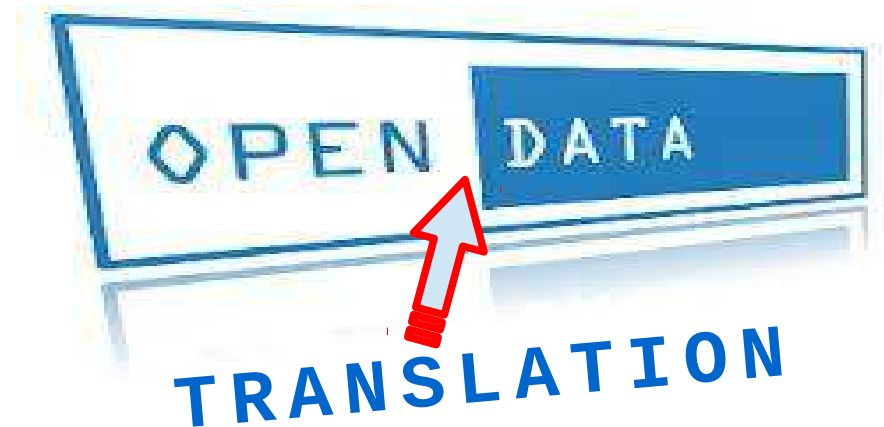
Il quadro di valutazione PSI è uno strumento di 'crowdsourcing' per misurare lo stato di Open Data e riutilizzo in tutta l'UE. Essa non controlla le politiche del governo, ma ha lo scopo di valutare la situazione PSI riutilizzo complessivo, che comprende le attività della comunità open data. Per saperne di più sulla metodologia e come aggiornarlo.



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

le traduzioni come dati aperti?

- le traduzioni eseguite nelle strutture pubbliche a finanziamento pubblico fanno parte del **patrimonio informativo pubblico (PSI)**
- e rientrano dunque nell'ambito dei dati aperti
- in generale con il termine degli Open Translation Data intendiamo tutti i dati liberamente accessibili che memorizzino traduzioni in un formato aperto specifico



Open Translation Data: esempi

- 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



- Direzione Generale della Traduzione
<http://open-data.europa.eu/en/data/dataset/dgt-translation-memory>
traduzione completa del diritto europeo (acquis communautaire) nelle lingue ufficiali dell'UE
- suddiviso per annate 2004-2012 e file trimestrali
- estrazione della combinazione di lingue preferite con TMExtract
`java -jar TMXtract.jar DE IT /home/c61302/temp/DGT-deit.tmx`



Portale Open Data dell'Unione europea

Open Translation Data: ragioni

- efficienza delle traduzioni (riutilizzo)
- aumento della consistenza delle traduzioni
- divulgazione delle traduzioni ufficiali
- enfaticizzazione del ruolo della traduzione
- supporto, accelerazione e controllo della pianificazione linguistica e terminologica
- ...



esempio: Prov. Aut. di Bolzano

Sito ufficiale all'interno della Rete Civica con la legislazione completa nelle 3 lingue ufficiali:

Deutsch - Italiano - Ladin | Rete Civica dell'Alto Adige | Amministrazione Provinciale | Feedback | **Alto contrasto**

AUTONOME PROVINZ BOZEN SÜDTIROL  PROVINCIA AUTONOMA DI BOLZANO ALTO ADIGE

LexBrowser

Home | Direzione Provinciale

In vigore al: 08/01/2014 

RICERCA: Testo  Anno Numero Articolo Natura

Deutsch Italiano Ladin

- Norme costituzionali
- Normativa provinciale
- Delibere della Giunta provinciale
- Sentenze della Corte costituzionale
- Sentenze T.A.R.
- Indice cronologico

Deutsch Italiano Ladin

Ultima edizione Visualizza note Visualizza massime

[Normativa provinciale](#) » [Energia](#) » [Risparmio energetico](#) » [Legge provinciale 7 luglio 2010 , n. 9](#)

a) Legge provinciale 7 luglio 2010 , n. 9¹⁾

Disposizioni in materia di risparmio energetico e energia rinnovabile

- Art. 1 (Finalità e ambito di applicazione) 

esempio: Prov. Aut. di Bolzano

- premesse:
 - testi tradotti e pubblicati
 - traduzioni eseguite con l'ausilio di memorie di traduzioni
 - supporto del formato TMX

The screenshot shows a web interface with a navigation bar at the top containing language options: Deutsch, Italiano, Ladin, and a highlighted 'formato multilingue TMX'. To the right of this bar is a button labeled '| Esporta in TMX |'. Below the navigation bar, there are options for 'Ultima edizione', 'Visualizza note' (unchecked), and 'Visualizza massime' (checked). Further right are links for 'Stampa', 'Esporta in PDF', and 'Invia'. The main content area displays a breadcrumb trail: 'Normativa provinciale > Energia > Risparmio energetico > Legge provinciale 7 luglio 2010 , n. 9'. Below this, the title of the document is shown: 'a) Legge provinciale 7 luglio 2010 , n. 9¹⁾ Disposizioni in materia di risparmio energetico e energia rinnovabile'. At the bottom right of the content area is a link '▼ Documento intero'. At the bottom left, there is a list item: '• Art. 1 (Finalità e ambito di applicazione)'. Red hand-drawn annotations highlight the 'formato multilingue TMX' button and the '| Esporta in TMX |' button.

sintesi

- formati e dati aperti sono imprescindibili
- l'uso di strumenti aperti nella traduzione è possibile ed auspicabile
- tutte le traduzioni ufficiali e pubbliche dovrebbero essere disponibili come dati aperti
- gli Open Translation Data portano dei vantaggi pratici cospicui, nonché aumentano il riconoscimento e la presenza in pubblico dei traduttori



Grazie

sito web:

<http://www.petersandrini.net>

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