



Università di Padova
Dipartimenti di Studi Linguistici e Letterari

Tecnologie per la Traduzione 2019/2020

Workflow e XML per la traduzione

Giorgio Maria Di Nunzio

Obiettivi

- Capire le varie fasi del flusso di lavoro di un traduzione
- Problemi relativi alla codifica della struttura dei documenti nell'ambito di una traduzione di un testo
 - XML Localisation Interchange Format File (XLIFF)
 - Translation Memory eXchange (TMX)

Localisation Workflow

- Un progetto di localizzazione è composto da una serie di passaggi che coinvolgono vari attori.
- In una visione “tradizionale”, il workflow è il seguente

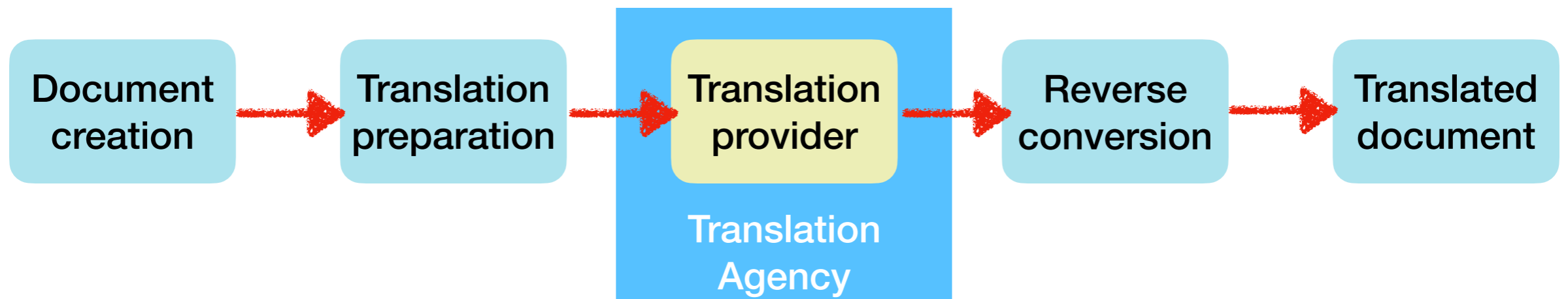


Localisation Workflow

- Problema di formati del documento da tradurre
- Porzioni di testo ripetute in documenti diversi
- Formattazione del testo da riconsegnare
- Passaggio di informazioni riservate

Localisation Workflow (Revised)

- Una soluzione alternativa è la seguente



<https://www.ibm.com/developerworks/library/x-localis/>

Localisation Workflow (Revised)

- Estrarre il testo da tradurre dal documento originale
- Salvare il testo in un documento XML
- Inviare il documento XML per la traduzione
- Recuperare le traduzioni nel file XML tradotto e reinserire il testo nel formato originale

XML Localisation Interchange Format File (XLIFF)

- Formato basato sull'XML per la standardizzazione del processo di localizzazione.
- Organization for the Advancement of Structured Information Standards (OASIS), 2002
 - Versione 1.2 del 2008

Struttura del file XLIFF

<trans-unit>

<trans-unit approved="no" id="1" resname="res1">

</trans-unit>

Struttura del file XLIFF

<trans-unit>

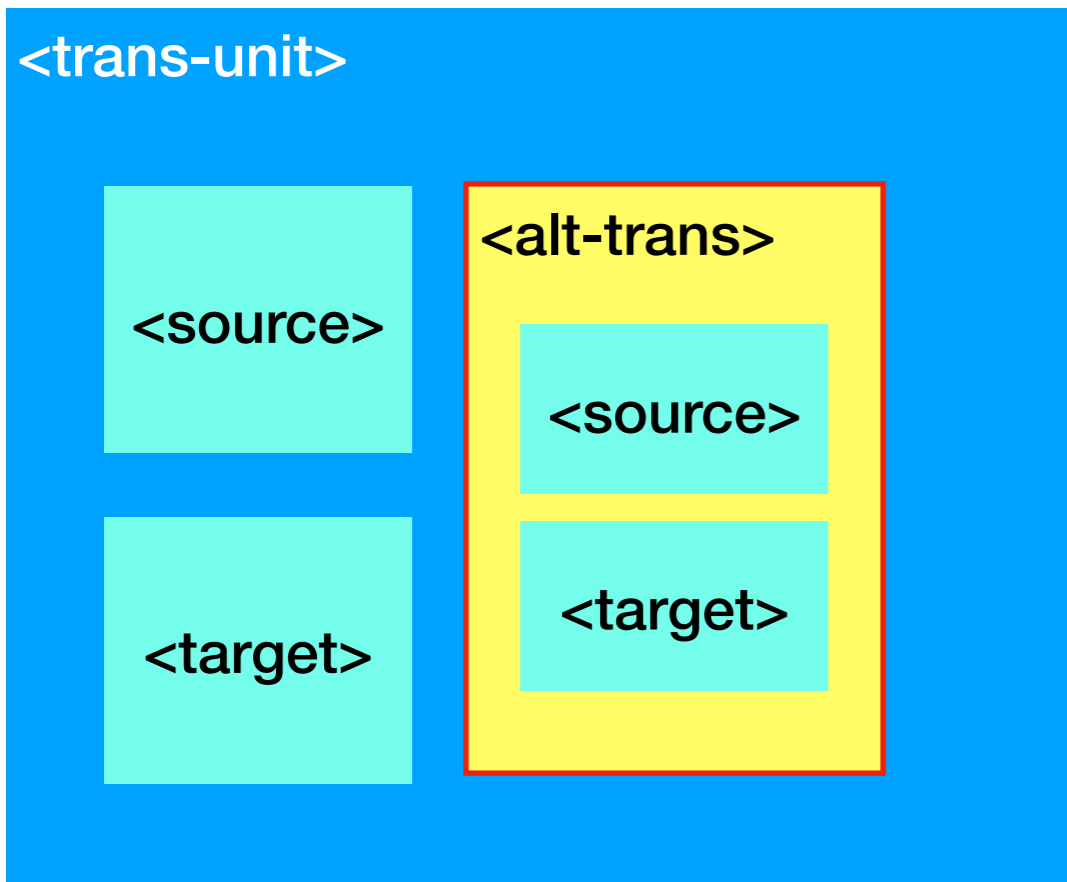
<source>

<target>

```
<trans-unit approved="no" id="1" resname="res1">  
  <source xml:lang="en">Corporate Headquarters</source>  
  <target xml:lang="es">Casa Central de la Compañía</target>
```

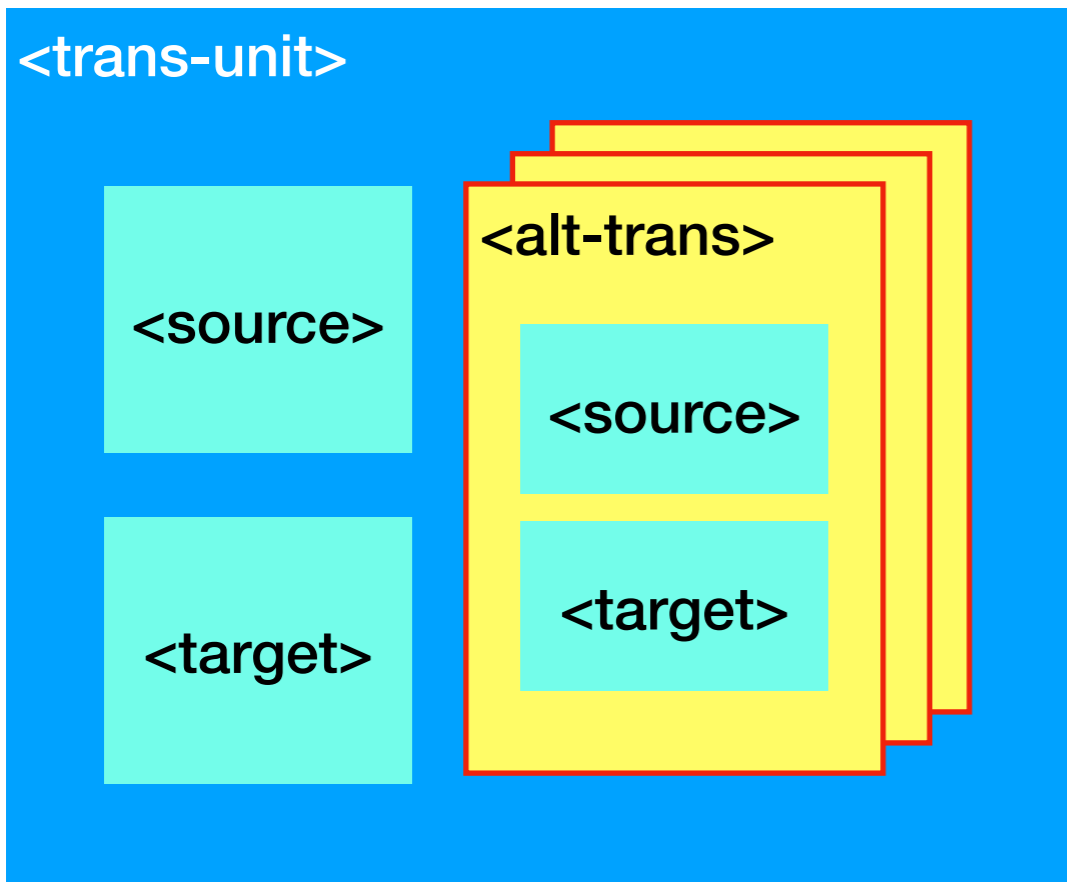
</trans-unit>

Struttura del file XLIFF



```
<trans-unit approved="no" id="1" resname="res1">  
  <source xml:lang="en">Corporate Headquarters</source>  
  <target xml:lang="es">Casa Central de la Compañía</target>  
  <alt-trans match-quality="100" origin="web" tool="TM Search">  
    <source xml:lang="en">Corporate Headquarters</source>  
    <target xml:lang="es">Casa Central de la Compañía</target>  
  </alt-trans>  
</trans-unit>
```

Struttura del file XLIFF



```
<trans-unit approved="no" id="1" resname="res1">  
  <source xml:lang="en">Corporate Headquarters</source>  
  <target xml:lang="es">Casa Central de la Compañía</target>  
  <alt-trans match-quality="100" origin="web" tool="TM Search">  
    <source xml:lang="en">Corporate Headquarters</source>  
    <target xml:lang="es">Casa Central de la Compañía</target>  
  </alt-trans>  
  <alt-trans match-quality="92" origin="web" tool="TM Search">  
    <source xml:lang="en">"Corporate Headquarters"</source>  
    <target xml:lang="es">"Casa Central de la Compañía"</target>  
  </alt-trans>  
</trans-unit>
```

Translation Memory eXchange (TMX)

- *“TMX (Translation Memory eXchange) is the vendor-neutral open XML standard for the exchange of Translation Memory (TM) data created by Computer Aided Translation (CAT) and localization tools. The purpose of TMX is to allow easier exchange of translation memory data between tools and/or translation vendors with little or no loss of critical data during the process.”*
- OSCAR (Open Standards for Container/Content Allowing Re-use), a special interest group of LISA (Localization Industry Standards Association), and first released in 1997.
 - Version 1.4b 2005

Translation Memory eXchange (TMX)

- Both specifications, XLIFF and TMX, contain all the elements necessary to store source document formatting information in XML format.
- A good CAT tool will allow a translator to reuse a sentence from an HTML page when translating a Rich Text Format (RTF) document, keeping text layout intact. This is the key reason why TMX format should be used as a complement to XLIFF.

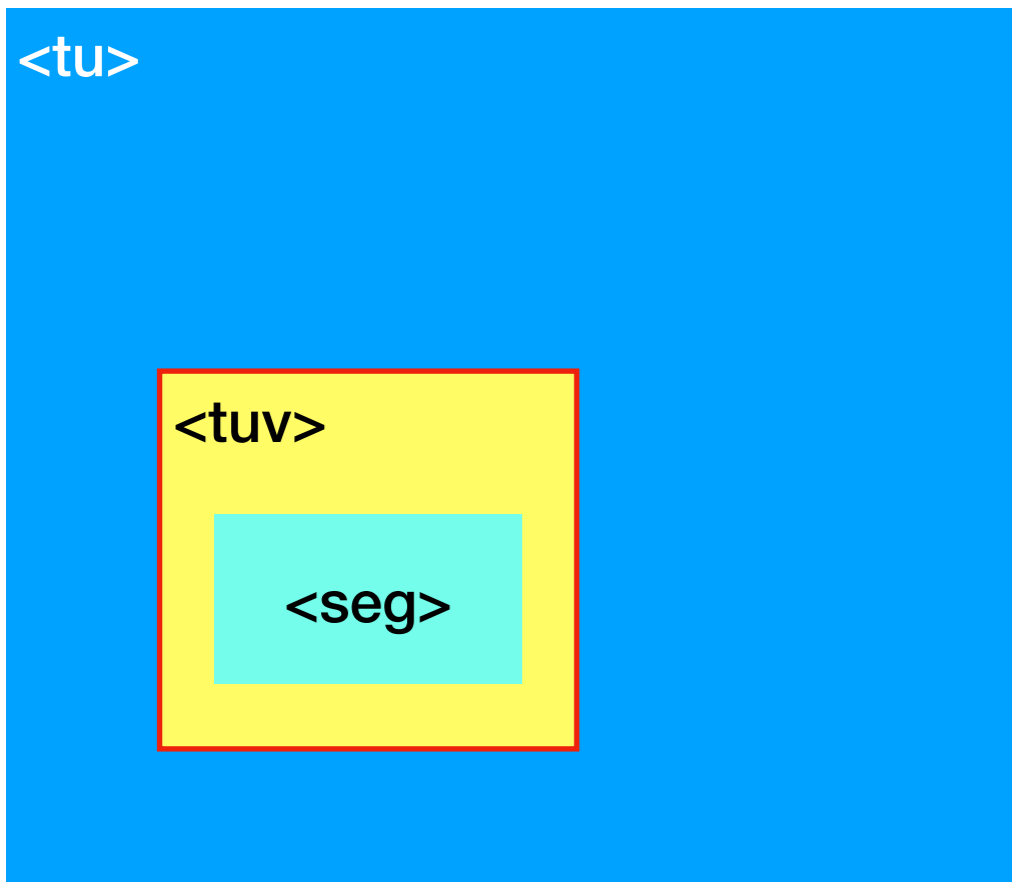
Struttura del file TMX

<tu>

<tu tuid="1090682451312" creationdate="20040313T224601Z">

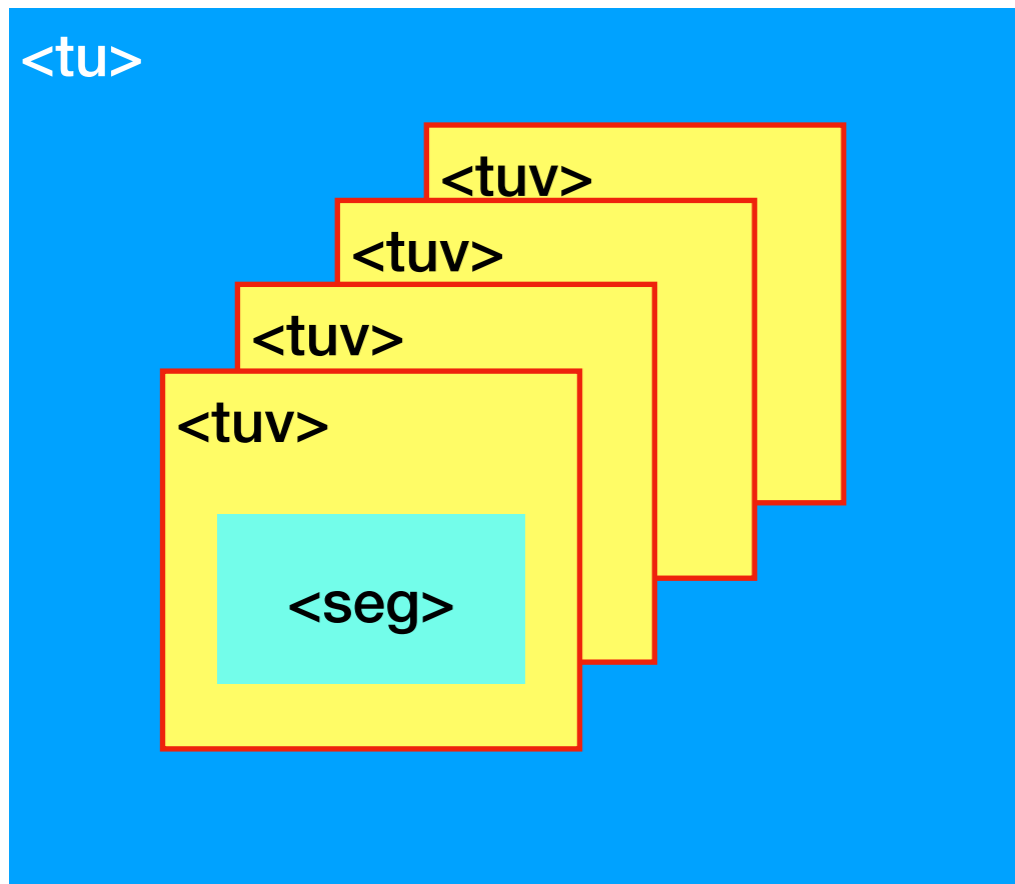
</tu>

Struttura del file TMX



```
<tu tuid="1090682451312" creationdate="20040313T224601Z">  
  <tuv xml:lang="en">  
    <seg>Corporate Headquarters</seg>  
  </tuv>  
</tu>
```

Struttura del file TMX



```
<tu tuid="1090682451312" creationdate="20040313T224601Z">  
  <tuv xml:lang="en">  
    <seg>Corporate Headquarters</seg>  
  </tuv>  
  <tuv xml:lang="es">  
    <seg>Casa Central de la Compañía</seg>  
  </tuv>  
</tu>
```