



**European Language  
Resource Coordination**  
*Connecting Europe Facility*

# ELRC Data Validation Report

<b>DisseminationLevel</b>	Internal
<b>Validation Guidelines version No.</b>	3.2
<b>Date</b>	24.05.2018
<b>Name of LR</b>	Convention on the transfer of sentenced persons (English - Greek) (Processed)
<b>Resource ID</b>	496
<b>Resource Version No.</b>	2.0
<b>Contact person</b>	Archontidis Garyfallos
<b>Validator</b>	Vassilis Papavassiliou – ILSP
<b>Validation Manager</b>	Kanella Pouli – ILSP
<b>Validation status</b>	<input type="checkbox"/> Changes required <input checked="" type="checkbox"/> Validated <input type="checkbox"/> Rejected

## Summary sheet

The validation results for this resource are as follows (please refer to the Validation Guidelines for the meaning of the various items):

<b>Validation steps</b>	<b>Validated (check box if yes)</b>	<b>Comments</b>
1) ELRC scope (see section 1 for details)	<input checked="" type="checkbox"/>	
2) Quick content check (see section 2 for details)	<input checked="" type="checkbox"/>	
3) LR Metadata (see section 3 for details)	<input checked="" type="checkbox"/>	
4) Legal issues (see section 4 for details)	<input checked="" type="checkbox"/>	

## ELRC Data Validation Report

### 1. Compliance with ELRC scope

	<b>Validated (check box if yes)</b>	<b>Comments</b>
Data origin (comes from public institutions or relevant to the general administrative/regulatory domain and does not come from the European Commission)	<input checked="" type="checkbox"/>	
Language(s) of the data content <sup>1</sup> (not the documentation)	<input checked="" type="checkbox"/>	

### 2. Quick content check

	<b>Validated (check box if yes)</b>	<b>Comments</b>
Readability of files	<input checked="" type="checkbox"/>	
Data content acceptability (no empty files, correct alignment for parallel corpora, ...)	<input checked="" type="checkbox"/>	

### 3. Validation of LR Metadata

#### a. General information

	<b>Validated (check box if yes)</b>	<b>Comments</b>
Language used in free text fields are CEF languages	<input checked="" type="checkbox"/>	
Does the “resource name” field contain an English version	<input checked="" type="checkbox"/>	
Does Language(s) in “description” field contain an English version	<input checked="" type="checkbox"/>	
Is there any information mentioning Pre-processing done by the provider	<input type="checkbox"/>	
Is there any information mentioning Pre-processing done through ELRC services	<input type="checkbox"/>	
Has any converting been performed on this resource to make it directly useful for training MT engines of the Automated Translation platform	<input checked="" type="checkbox"/>	

<sup>1</sup>Should contain at least one of the following languages: Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Icelandic, Irish, Italian, Latvian, Lithuanian, Maltese, Norwegian, Polish, Portuguese, Romanian, Slovakian, Slovenian, Spanish, Swedish

**ELRC Data Validation Report**

**b. Accuracy of completed metadata with respect to provided LR**

Mandatory metadata field names	Current value	Correct	Wrong	Missing	Comments
Resource name	Convention on the transfer of sentenced persons (English - Greek) (Processed)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Resource type	Corpus	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
PSI - Public Sector Information	NO	<input checked="" type="checkbox"/>	<input type="checkbox"/>	n/a	
Licence	CC-BY	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Contact person surname	Archontidis	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Contact person email	<a href="mailto:garhodidis@justice.gov.gr">garhodidis@justice.gov.gr</a>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Linguality type	Bilingual	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Lexical conceptual resource or Language description type (n/a for corpora)	n/a	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Language(s) name	Modern Greek (1453-)/English/	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Encoding level (n/a for corpora)	n/a	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Character encoding (applicable for corpora only)	UTF8	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
Size	120	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Size unit	Translation Units	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Mime type	TMX	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Other metadata field names (to be listed if completed by submitter)	Current value	Correct	Wrong	Comments
Domain	LAW (rights and freedoms)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Multilinguality type	Parallel	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Attribution text	Convention on the transfer of sentenced persons (English - Greek) was created for the European Language Resources Coordination Action (ELRC) ( <a href="http://lr-coordination.eu/">http://lr-coordination.eu/</a> ) by the Ministry of Justice, Transparency and Human Rights of the Hellenic Republic	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

### ELRC Data Validation Report

	( <a href="http://www.ministryofjustice.gr">http://www.ministryofjustice.gr</a> ) and is licensed under "CC-BY 4.0" ( <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a> ).			
Allows Uses Besides DGT	Yes	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Relation type and ID of related resource	Is Aligned Version of #225	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

## 4. Legal validation

### a. If “PSI - Public Sector Information” metadata checkbox is ticked

	Validated (check box if yes)	Comments
“Licence field” value is identified (any value except “under review”) or indicated as “not available” if information about the licence is not available	<input type="checkbox"/>	
If attribution is required, IPR Holder(s) is identified in the “IPR holder” field	<input type="checkbox"/>	
Privacy/Confidentiality (if the resource is identified as private or confidential, is “Personal Data Included” or “Sensitive Data Included” box ticked?)	<input type="checkbox"/>	

### b. If “PSI - Public Sector Information” metadata checkbox is not ticked

	Validated (check box if yes)	Comments
“Licence field” value is identified (any value except “under review”)	<input checked="" type="checkbox"/>	
If attribution is required, IPR Holder(s) is identified in the “IPR holder” field	<input checked="" type="checkbox"/>	Indicated in the attribution text
Privacy/Confidentiality (if the resource is identified as private or confidential, is “Personal Data Included” or “Sensitive Data Included” box ticked?)	<input type="checkbox"/>	

**ELRC Data Validation Report****5. Further comments**

The dataset was provided as a collection of 3 MO and 2 PDF files. Text was extracted semi-automatically, document alignment and sentence alignment were applied. As a post-processing task several filters were applied to discard/annotate alignments that might be incorrect or of limited use for training MT systems.

**6. Declaration on the list of pre-existing rights**

No.	Options	Selected option
1	The results of this LR are free of rights or claims from creators or from any third parties for any use the contracting authority may envisage and declare that the results do not contain any pre-existing rights to the results or parts of the results or to pre-existing materials as defined in the above-mentioned contract.	<input checked="" type="checkbox"/>
2	The results of this LR and the pre-existing material incorporated in the results are free of rights or claims from creators or from any third parties for any use the contracting authority may envisage and declare that the results contain the following pre-existing rights:	<input type="checkbox"/>

**For Option 2 complete the table below – one line per pre-existing right**

Result concerned	Pre-existing material concerned	Rights to pre-existing material	Identification of rights' holder