

ELRC Data Validation Report



ELRC Data Validation Report

Dissemination Level	Internal
Validation Guidelines version No.	3.1
Date	30 June 2017
Name of LR	Letter of rights for persons arrested on the basis of a European Arrest Warrant
Contact person	Soulos Ioannis
Validator	Kanella Pouli – ILSP
Validation Manager	Kanella Pouli – ILSP
Validation status	<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):

Validation steps	Validated (check box if yes)	Comments
1) ELRC scope (see section Error! Reference source not found. for details)	<input checked="" type="checkbox"/>	
2) Quick content check (see section Error! Reference source not found. for details)	<input checked="" type="checkbox"/>	

ELRC Data Validation Report

3) LR Metadata (see section Error! Reference source not found. for details)	<input checked="" type="checkbox"/>	
4) Legal issues (see section Error! Reference source not found. for details)	<input checked="" type="checkbox"/>	

1. Compliance with ELRC scope

	Validated (check box if yes)	Comments
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 ¹ (not the documentation)	<input checked="" type="checkbox"/>	

2. Quick content check

	Validated (check box if yes)	Comments
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

	Validated (check box if yes)	Comments
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"/>	

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	Letter of rights for persons arrested on the basis of a European Arrest Warrant	<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	Yes	<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	Soulos	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Contact person email	isoulos@pio.moi.gov.cy	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Linguality type	Multilingual	<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	Bulgarian, Dutch ; Flemish, English, French, German, Italian, Latvian, Modern Greek (1453-), Polish, Romanian ; Moldavian ; Moldovan	<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)	UTF-8	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
Size	854	<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
		<input type="checkbox"/>	<input type="checkbox"/>	
		<input type="checkbox"/>	<input type="checkbox"/>	
		<input type="checkbox"/>	<input type="checkbox"/>	
		<input type="checkbox"/>	<input type="checkbox"/>	
		<input type="checkbox"/>	<input type="checkbox"/>	
		<input type="checkbox"/>	<input type="checkbox"/>	

ELRC Data Validation Report

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 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"/>	

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 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"/>	

5. Further comments

During conversion, parallel PDF files were aligned into a parallel corpus in TMX format. 45 language pairs.