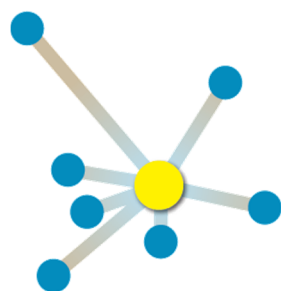


European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2014/1074 programme.



**European Language
Resource Coordination**
Connecting Europe Facility

ELRC Data Validation Report

Dissemination Level	Internal
Validation Guidelines version No.	2.0
Date	30 June 2017
Name of LR	Corpus of State-related content from the Latvian Web
Contact person	Jānis Ziediņš, KISC
Validator	Roberts Rozis, TILDE
Validation Manager	Aivars Bērziņš, TILDE
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 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**

	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"/>	Checked. No such info.
Is there any information mentioning Pre-processing done through ELRC services	<input type="checkbox"/>	Checked. No such info.
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"/>	

¹ 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	Corpus of State-related content from the Latvian Web	<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-SA	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Contact person surname	Ziediņš	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Contact person email	janis.ziedins@kis.gov.lv	<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)	-	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Language(s) name	en lv	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Encoding level (n/a for corpora)		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Character encoding (applicable for corpora only)	UTF-8	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Size	6861	<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	Politics	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Conformance to classification scheme	EUROVOC	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Multilinguality type	Parallel	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Attribution text	The Corpus of State-related content from the Latvian Web has been donated to the European Language Resources	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

ELRC Data Validation Report

	Coordination Action (ELRC) (http://lr-coordination.eu/) by Culture Information Systems Centre and is licensed under "CC-BY-SA 4.0" (https://creativecommons.org/licenses/by-sa/4.0/).			
Allows Uses Besides DGT	Ticked "yes"	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
IPR Holder	Culture Information Systems Centre	<input checked="" 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”)	<input checked="" type="checkbox"/>	
If attribution is required, IPR Holder(s) is identified in the “IPR holder” field	<input checked="" 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"/>	No Privacy / Confidentiality

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"/>	n/a
If attribution is required, IPR Holder(s) is identified in the “IPR holder” field	<input type="checkbox"/>	n/a
Privacy/Confidentiality (if the resource is identified as private or confidential, is “Personal Data Included” or “Sensitive Data Included” box ticked?)	<input type="checkbox"/>	n/a

5. Further comments

During conversion, parallel Plain text files were converted into a parallel corpus in TMX format.