

Open European Location Services (Open ELS)

Arrangements for European Data Portal

Deliverable Open ELS task 2.3

Authors: Roy H. Mellum

Date: May 20, 2019

Version: 1.0



Co-financed by the European Union
Connecting Europe Facility

Contents

1	Introduction.....	3
2	Linking to the European Data Portal	3
3	The Automated Translation Building Block (eTranslation).....	3
3.1	Initial considerations	3
3.2	Testing	4
3.3	Validation.....	4

1 Introduction

The work described in this document has been carried out by the Open ELS Task 2.3 “Arrangements for the European Data Portal”. The task included two subtasks (based on requirements from INEA during the negotiation of the Grant Agreement):

1. Linking the Open ELS platform to the European Data Portal (based on The Gold Book and harvesting guidance),
2. Testing, validation and consideration of the Automatic Translation Building Block (eTranslation).

The main results of the Open ELS Task 2.3 are:

- The Open ELS metadataservices (CS-w) serving metadata for the Open ELS services are harvested by the European Data Portal.
- The report on the testing, verification and consideration of The Automated Translation Building Block (eTranslation).

2 Linking to the European Data Portal

The task of the ELF Cascading WFS service is to facilitate access to national download services.

The Open ELS metadata catalogue and metadataservice has been established in the ELS technical infrastructure. It is based on the open source software Geonetwork.

This task has a strong link to Activity 3, task 3.1 where the metadata for the Open ELS products has been created and made available for harvesting by EDP.

The access point (capabilities request) for the Open ELS catalogue service web (CS-w) is:

<https://demo.locationframework.eu/geonetwork/srv/eng/csw-opensels?request=GetCapabilities&service=CSW&version=2.0.2>

3 The Automated Translation Building Block (eTranslation)

3.1 Initial considerations

The initial considerations included the decision on the test approach, selection of appropriate documents for testing, selection of the domain provided by eTranslation for testing and selection of languages (all test originals in English) to translate to.

The following approach for testing was decided:

- 1) Translation of selected documents into selected languages with the One-off translation method from the eTranslation website.
- 2) Validate the results.
- 3) If results from 1) are positive and valuable, continue testing in machine-to-machine mode by interfacing to the eTranslation web-service API.

Three types of documents were selected for testing:

- 1) Open ELS metadata records INSPIRE profile in plain XML-format.
- 2) The 21 pages report for Open ELS task 2.5 'On the fly edge matching' in .docx format; including table of contents and several figures and illustrations.
- 3) Two web articles from an Open ELS Project Partner (Kartverket) in .docx format, text only; provided by the Partners information officer.

All the chosen documents are regarded as typical for the geospatial domain and are domain specific in the sense that they include technical expressions and wordings.

The availability of already available domains for translation was considered as a main issue for achieving the best possible test results. Not surprisingly there is no specific translation domain provided for 'The geospatial domain'. Based on description of the domains provided by eTranslation, the Cutting Edge domain was chosen.

The test should include translation of the selected documents (not web articles) from the project working language English to: German, Dutch, Lithuanian, Finnish and Norwegian. The two web articles were only translated from Norwegian to English.

3.2 Testing

The documentation provided on the eTranslation website is good and guided the project team in an intuitive way through the man-machine test process. It was straightforward for us to obtain credentials for use of eTranslation. The documentation for the API (machine-to-machine) has been evaluated by experts concluding that it can be integrated in a wide range of applications with reasonable efforts.

An overview of typical experienced response times is indicated in the table below.

Document type	Format	Typical response time (Man-Machine)*
Metadata record	xml	15 secs
Open ELS task 2.5 report	docx	150 - 350 secs
Web article	docx	15 secs

*From document send to received translation by e-mail.

The typical response times are considered to be fairly good for a service of this type.

All translations and tests were done during week 51, 2018. The eTranslation man-to-machine service was stable and reliable all through the test period.

3.3 Validation

The translated documents were validated/ reviewed by representatives of Project Partners in the various languages. The results has been classified into 3 categories:

- 1 – Fairly good useable with little editorial work,
- 2 – Medium useable require more editorial work,
- 3 – Unusable use of human translator required.

Validated results for translation of two Open ELS documents to 5 languages are shown in the table below:

Document type	GE	NL	FI	LT	NO
Metadata record	1	2	2	2	2
Open ELS task 2.5 report	1	2	3	2	3

For the two web articles translated from Norwegian to English the validation unfortunately is of category 3.

The results for the testing and validation of the metadata records does not vary much between languages. This is not surprising because the document contains very little text. Only three of the mandatory fields are free text and often of limited length. The rest of the contents come from code lists and/or vocabularies. The validators comment for improvements for translation of metadata documents is that significant improvements for eTranslation can be achieved by translating the code lists and vocabularies.

The translation of the technical report had more severe weaknesses, in particular limited capabilities of translation technical ‘language and expressions’ but also constructing meaningful sentences from the original English version.

From our expectations before the work started, the overall results are not surprising because:

- The geospatial domain is a small domain compared to e.g. health and justice and resources for supporting eTranslation in this domain from countries is probably not high priority,
- The geospatial domain is very technical with special words and also use of words and sentences,
- The use of eTranslation in some of the partner countries Finland and Norway has probably been limited, thus the machine learning parts of eTranslation has not proven the potential value,
- For Norway, being an EEA country.

Based on the test results and validations it the project team decided not to implement the machine to machine interface (API):

- The translation to most of the languages for performed tests on typical documents in the geospatial domain has several weaknesses – translation of words and phrases- and also building meaningful understanding of the text in a broader.
- The Open ELS User Interface was under development and built at the end of the project, therefore Open ELS had not have any proper application or website available to integrate with. A small test application with some more or less fake test data could have been developed, but it was not found meaningful because of the reason mentioned above (currently low level of maturity for geospatial domain).

An eTranslation for the geospatial domain could be very beneficial, a proposed approach is to agree with a geospatial organisation like EuroGeographics to moderate the translation work.