Project acronym: LISE

Grant Agreement Number: 270917

Project title: Legal Languages Interoperability Services

DELIVERABLE

D4.2 Evaluation Plan Update

Dissemination Level: PU

Version No. 7

24/01/2013

Project co-funded by the European Commission within the ICT Policy Support Programme.

Document Information

Deliverable Number	D4.2
Deliverable title	Evaluation Plan Update
Due date of deliverable according to DOW	M24
Actual submission date of deliverable	
Main Author(s)	Joeri Van de Walle
Participants	Lambros Kranias, Michael Wetzel
Reviewer	
Work package	WP4
Work package leader	CL
Dissemination level	PU
Version	6

Revision History

Revision	Date	Author	Organisation	Description
1	**2012-11-28**	**Joeri Van de Walle**	**CL**	**Initial draft**
2	**2012-12-12**	**Joeri Van de Walle**	**CL**	**Update after meeting with ESTeam**
3	**2012-14-12**	**Lambros Kranias**	**ESTeam**	**Green font changes**
4	**2012-12-20**	**Joeri Van de Walle**	**CL**	**Various updates**
5	**2013-01-07**	**Natascia Ralli, Elena Chiocchetti**	**EURAC**	**Feedback and input**
6	**2013-01-08**	**Tanja Wissik**	**UniVie**	**Feedback and input**
7	**2013-01-24**	**Joeri Van de Walle**	**CL**	**Integrate feedback partners and finalise**

Statement of originality:
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Contents

1 Introduction 4
2 Evaluation Plan 5
2.1 Evaluation Set 5
2.2 Evaluator Profile 9
2.3 Preparation 9
2.3.1 Cleanup 9
2.3.2 Omeo 11
2.3.3 Fillup 14
2.4 Evaluation 15
2.4.1 IATE Volunteers Examine Cleanup Results 15
2.4.2 IATE Volunteers Examine Omeo Results 16
2.4.3 IATE Volunteers Examine Fillup Results 16
2.5 Evaluation Results 17
3 Appendix A - Languages Found in IATE Database Export 18
4 Appendix B - Workshop Program 19
5 Appendix C - Evaluation Results Template 20
5.1 Cleanup Results 20
5.2 Omeo Results 20
5.3 Fillup Results 20

1 Introduction

The purpose of this document is to describe how we will evaluate the performance of the tools developed within the LISE project.

This document provides an update on the evaluation strategy that was outlined in deliverable D4.2, “Evaluation Plan”. In that first draft of our evaluation plan, we described an evaluation approach that was based on a matrix of user profiles and use cases. However, since that first draft, it has become clear from the interviews performed in the context of deliverable D3.1, “Report Analysis of existing Terminology Workflows”, and D3.2, “Report Workflow Adaptation for LISE” that clear-cut user profiles are rarely adhered to in the daily practice of terminology management. Different tasks that need to be performed (such as term selection, elaboration of term entries, revision and quality check,…) are not necessarily always entrusted to the same roles/user profiles within different terminology workflows. How organisations fill in terminology management responsibilities seems to depend more on availability of people and available skill sets than on clear-cut role definitions or profile descriptions.

Because of this observation, we have decided to move away from the profile-driven evaluation approach that we proposed in the D4.2 and focus more on the use cases. The actions or tasks involved in enacting the use cases may be performed by any user that has the required skills.

For the use cases, we refer to the scenarios described in deliverable D3.2, “Report Workflow Adaptation for LISE”. The scenarios described in that deliverable illustrate typical situations that owners of terminological data collections might be confronted with and how the LISE tools may assist in those situations. When we describe the tool evaluations below, we will indicate which tools are considered relevant to which use cases.

The evaluation plan that we describe below is targeted towards a very specific data set, namely that of the IATE terminological database. We describe the optimal and most comprehensive workflow for making the most out of this specific data set by making use of all the LISE tools. However, the components and methods described may also be used in different contexts and with different data sets. For instance, an organisation may be happy with the way its current terminological database is structured and the contents that it contains, but may want to expand into additional languages. In that case, the organisation would only need to use Fillup and look into the Fillup use case that applies. For each of the tools, we mention the relevant use cases to direct interested users to the tool(s) they need for the scenario that applies to their specific usage context. It must be noted, though, that the use of different data sets may require specific preparatory steps before the data can be loaded into the tools.

2 Evaluation Plan

In a nutshell, the objective of this usability evaluation is to obtain feedback on the usefulness of the LISE tools by putting users to work with the tools on real data and have them carry out tasks that they would also need to perform in their daily terminology work. By having the users work with their own data, we try to stay as close as possible to the real usage context so as to ensure meaningful evaluation results.

In this section we describe the main components of the updated evaluation approach. We describe (1) how the evaluation set was abstracted from the data received, (2) the profile of the people who performed the evaluation, and (3) the evaluation preparation and script.

2.1 Evaluation Set

Early 2012, the LISE consortium received a complete export of the IATE database. The first step after receipt of the data was to analyse it to understand how exactly the data was structured, what types of information it contained, and what type of problems we might be confronted with while setting up the evaluation.

Table 1 shows some figures on the IATE data; more specifically, it shows the number of terminological entries (records) contained in the export and how many terms are used in various languages to represent the concepts the entries represent. As a general rule, each terminological entry describes a concept, which in turn can be labelled by one or more terms in each language considered.

Entries	Terms	Languages
1,470,943	11,163,473	22

Table 1: Amount of entries, terms, and languages found in IATE database export

Although the database export had terms in 22 languages, we found that these terms were not equally distributed across the different languages, as Figure 1 below illustrates.

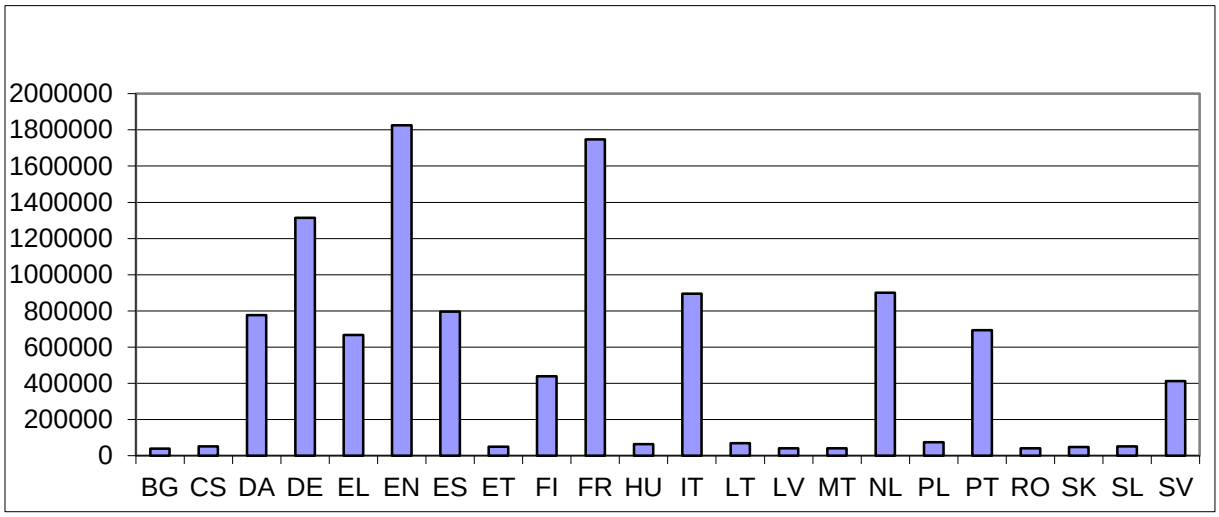

Figure 1: Number of terms per language for the complete data set

Many terminological entries contain terms in French, English or German, but relatively few also have terms in the newer EU languages, such as Estonian, Slovak, or Slovenian.

Looking at the data more closely, we found that a vast majority of terminological entries contain terms in just one or two languages. In addition, quite a few contain terms in up to 11 languages. Entries that contain terms in more than 11 languages are rare. Figure 2 provides an overview of the amount of languages represented in all terminological entries in the data set.

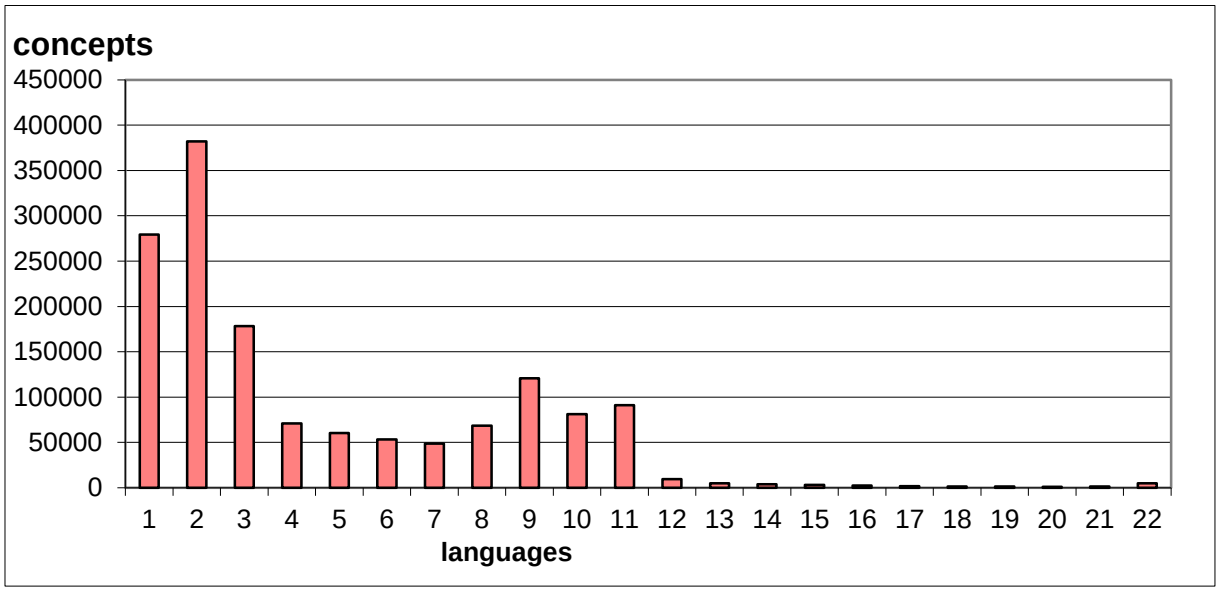

Figure 2: Number of languages per entry for the complete data set

The terminological entries in the database export had a lot of meta data associated with them. We have used this information to select a subset of the provided data to focus on during the evaluation. Since the focus of the LISE project is on legal and administrative terminology in the domain of social security, we limited our evaluation set to those entries that have at least one of the domain attribute values shown in Table 2.

DOMAIN_ID	DOMAIN_NAME	NUM_OF_ENTRIES
2431	FINANCE—>Insurance	20393
28	SOCIAL QUESTIONS	68588
2826003	SOCIAL QUESTIONS—>Social affairs—>Social policy	1025
2836	SOCIAL QUESTIONS—>Social protection	2816
2836001	SOCIAL QUESTIONS—>Social protection—>Leave on social grounds	84
2836002	SOCIAL QUESTIONS—>Social protection—>Social security	3104
2836003	SOCIAL QUESTIONS—>Social protection—>Welfare	119
4406004	EMPLOYMENT AND WORKING CONDITIONS—>Employment—>Unemployment	380

Table 2: Amount of entries selected from the IATE database export based on domain

This resulted in subset of 95,544 terms in 21 languages distributed across 21,515 terminological entries. The database export did not appear to contain any terms in the social security domain for Bulgarian. Some terminological entries in the set were associated with more than one of the selected domains.

The distribution of terms across languages follows that of the complete data set, as shown in Figure 3. Terminological entries are likely to contain terms for the older EU languages, more so than for the EU languages of countries that have joined the European Union more recently.

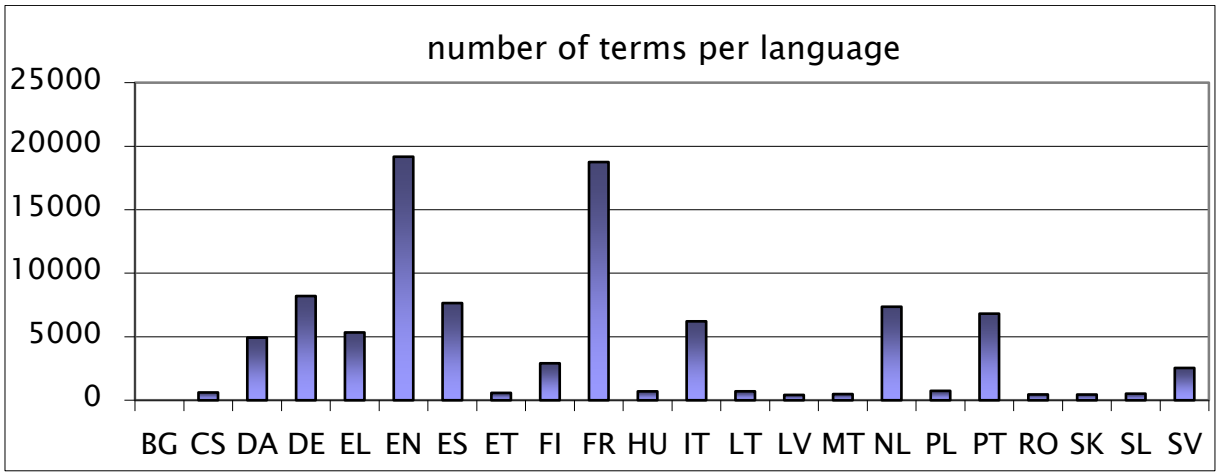

Figure 3: Number of terms per language for the selected evaluation set

Similar to what we found in the complete data set, a large number of terminological entries contain only one term or terms in only one language. Figure 4 shows, that this is the case for almost 9,000 terminological entries out of the 21,515 in the set. Also, in the selected evaluation set there are quite a few entries that contain terms in up to 11 languages. Very few have terms in more than 11 languages.

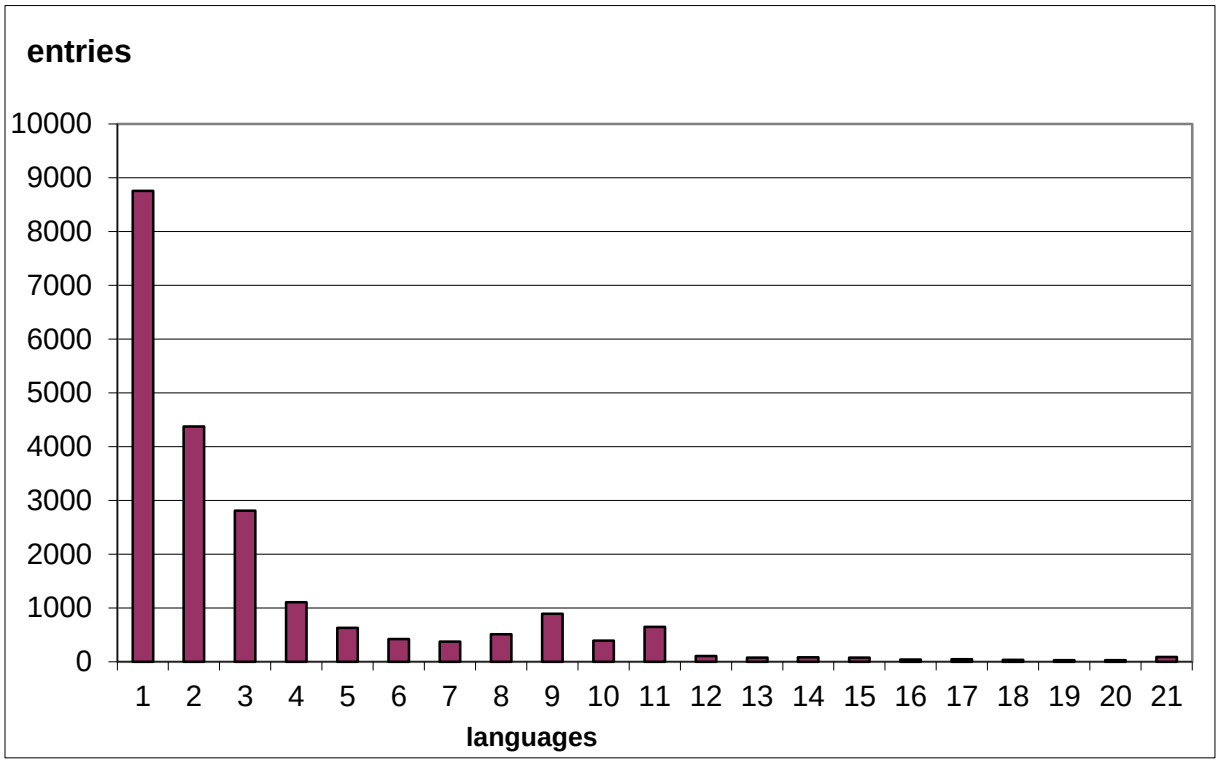

Figure 4: Number of languages per entry for the selected evaluation set

As already mentioned, the domain information that was present in the IATE database was used to limit the data set to those entries with a domain attribute related to social security. However, the terms and entries in the database had a lot more meta data associated with them than just domain information. In total, more than 100 meta data fields existed in the IATE database export. For the purpose of this evaluation, we have limited the set of meta data fields that are taken into account for the evaluation to the following:

- Term Group Id
- Term Language
- Term Type (Term/Abbreviation/Phrase/etc.)
- Term Domain (3 hierarchical levels)
- Term Origin
- Term Reliability
- Term Frequency

Some of these fields have been used for the selection of terminological entries and for decision making (term domain). Others will be used in the tool interfaces to provide contextual information to evaluators (term group id, term language, term domain, term origin).

This set of 21,515 terminological entries (containing 95,544 terms in 21 languages) in the social security domain, henceforth called *the evaluation set*, formed the basis of our

evaluation. Relevant subsets of this evaluation set featured in the evaluations of the LISE tools Cleanup, Omeo, and Fillup.

2.2 Evaluator Profile

The evaluation set was processed by the LISE tools as described in section 2.3, "Preparation". The outcome of each of the processing steps will be examined by human evaluators during a one-day workshop that will be held on February 5th, 2013, at the Translation Centre for the Bodies of the European Union (CdT) in Luxembourg.

Workshop participants will all be people that are actively involved in the maintenance of the IATE terminological database (available at http://iate.europa.eu). Their profile is that of terminologist. That is, these users are accustomed to work on terminology-related tasks as part of their daily work. Typical tasks they perform include: cleaning up existing terminological entries, validating specific term data, checking/adding meta data information to entries/terms, etc.

2.3 Preparation

To assess the performance of the LISE tools, the evaluation set described in section 2.1, "Evaluation Set", was processed by the tools. The results of that processing will be presented to and evaluated by the IATE members that volunteered for the workshop on the 5th of February. The actual processing of the data happened prior to the workshop.

2.3.1 Cleanup

Common use cases that Cleanup addresses are:

- Terminology database needs to be updated (scenario 1 in D3.2)
- Duplicate entries ("doublettes") need to be removed (scenario 6 in D3.2)

The selected evaluation set was processed by the Cleanup tool that tried to identify potentially problematic terms. The type of problems the tool looks for are listed in Table 3 below.

Error Type	Description
Misspelling	Terms that are misspelled (check is based on spelling dictionaries).
Canonisation	Terms with missing or erroneous accents or diacritical signs (check is based on similar term occurrences in the data set).
Language	Terms that are attributed to the wrong language.
Equivalent	Doublettes or terms that look very similar and occur more than once within the same language and within the same terminological entry.
Domain	Terms for which additional domains are suggested.
Translation	Terms that appear to be mistranslations of other terms in the terminological entry.
Subset	Overlapping entries, i.e. a set of terms in a terminological entry occurs as a subset of another terminological entry that has the same terms (or close variants) plus, maybe, one or more additional terms.

Table 3: Cleanup error type descriptions

In preparation of the workshop a number of Cleanup runs were performed to optimise the error identification algorithms. During these preparatory runs we have found that some languages yielded results that were more suitable for use in the workshop than others (see Table 6 below). This is not surprising, considering that the amount of terms for each language varies greatly (see Figure 4 above).

	TOTAL	SPELLING	CANONISATION	LANGUAGE	EQUIVALENT	DOMAIN	TRANSLATION	SUBSET
EN	**310**	**22**	**0**	**4**	**6**	**225**	**54**	**102**
FR	**387**	**24**	**7**	**25**	**10**	**282**	**42**	**154**
DE	**121**	**14**	**0**	**7**	**7**	**68**	**25**	**67**
ES	**170**	**22**	**4**	**11**	**8**	**118**	**8**	**28**
IT	**158**	**14**	**0**	**12**	**1**	**122**	**9**	**33**
DA	81	9	0	28	2	38	4	19
NL	**113**	**14**	**0**	**20**	**3**	**63**	**13**	**33**
SV	40	6	0	13	0	16	5	8
EL	**124**	**7**	**1**	**0**	**2**	**101**	**15**	**18**
PT	**119**	**8**	**2**	**9**	**3**	**95**	**2**	**27**
FI	28	9	0	8	0	8	3	7
CS	17	5	0	6	0	5	1	1
HU	14	7	0	4	0	2	1	1
PL	14	4	0	3	0	4	3	0
SK	13	6	0	2	0	5	0	1
SL	6	3	0	1	0	2	0	2
ET	10	8	0	1	1	0	0	1
LT	12	3	0	4	0	4	1	2
LV	10	6	0	2	0	2	0	1
MT	7	0	0	4	0	2	1	0
BG	0	0	0	0	0	0	0	0
RO	8	1	0	3	0	2	2	1

Table 4: Preliminary Cleanup results

Based on these findings, the results of the following languages were found to be most appropriate for use during the workshop:

- English
- French
- German
- Spanish
- Italian
- Dutch
- Greek
- Portuguese

IATE was informed about these results prior to the workshop. They were asked to select the languages to be reviewed during the workshop.

The results of the Cleanup run were loaded in the Cleanup client tool for the selected languages.

2.3.2 Omeo

Common use cases that Omeo addresses are:

- New terminology data collections need to be imported (scenario 2 in D3.2)
- Terminology databases need to be merged (scenario 3 in D3.2)
- Duplicate entries ("doublettes") need to be removed (scenario 6 in D3.2)
- Language coverage needs to be monitored (scenario 8 in D3.2)
- Synonyms and variants need to be controlled (scenario 7 in D3.2)

The selected evaluation set was processed by the Omeo tool to try and create entry groups. Entry groups are created on the basis of terms that are found to be similar across different terminological entries.

Omeo was actually run twice. During the first round of processing, Omeo tried to identify related entries based on similarities that exist between terms for the same language in different terminological entries. Each of the languages in the evaluation set is used as the language for finding similar terms, so this first run creates as many potential entry groups as there are languages in the database.

For example:

```
Entry 1
   EN: old age insurance
   EN: old-age insurance
   EN: insurance in respect of old-age


Entry 2
   EN: poverty eradication
   EN: eradication of poverty
   EN: eradicating poverty


Entry 3
   DE: Europäisches System der integrierten Sozialschutzstatistik
   DE: Europäisches System der integrierten Sozialschutz-
Statistik
   DE: Europäisches System integrierter Sozialschutz-Statistiken
```

At this stage monolingual entry groups had been generated that are broader than the original IATE terminological entries. However, the original IATE terminological entries are (often) multilingual. Therefore, we can create even broader entry groups by combining the results of the first round of Omeo processing with the original multilingual IATE terminological entries. For example:

```
IATE Original Terminological Entry 1

   EN: provision of benefits
   EN: granting of benefits
```

```
   DE: Gewährung einer Leistung
   DE: Leistungsgewährung


IATE Original Terminological Entry 2

   EN: entitlement to benefits

   DE: Gewährung von Leistungen


OMEO Step-1 Entry Group 1

   DE: Gewährung einer Leistung
   DE: Leistungsgewährung
   DE: Gewährung von Leistungen



OMEO Step-2 Entry Group 1

   EN: provision of benefits
   EN: granting of benefits
   EN: entitlement to benefits
   DE: Gewährung von Leistungen
   DE: Gewährung einer Leistung
   DE: Leistungsgewährung
```

In this example, the original IATE terminological entries can be combined using the Omeo round-1 German group as a pivot point. Combining entry groups that were created during the first round of processing using the original terminological entries from IATE is what the second round of Omeo processing does. We call this 'multilingual grouping'.

Table 5 shows the number of monolingual groups that were produced during the first round of Omeo processing. Note that these groups do not take into account links that may have existed between terminological entries in the IATE database.

EN = 766	FR = 904	HU = 7	PL = 12
DE = 476	ES = 310	SK = 6	SL = 8
IT = 326	DA = 227	ET = 4	LT = 9
NL = 347	SV = 128	LV = 6	MT = 6
EL = 206	PT = 269	BG = 0	RO = 4
FI = 108	CS = 17		

Table 5: Preliminary results of first round of Omeo processing

Then, having merged the original IATE terminological entries during Omeo round 1 processing and Omeo round 2 processing (the multilingual grouping), we end up with 19,684 multilingual groups. The size of these groups, measured in number of distinct languages, is presented in Figure 5 below.

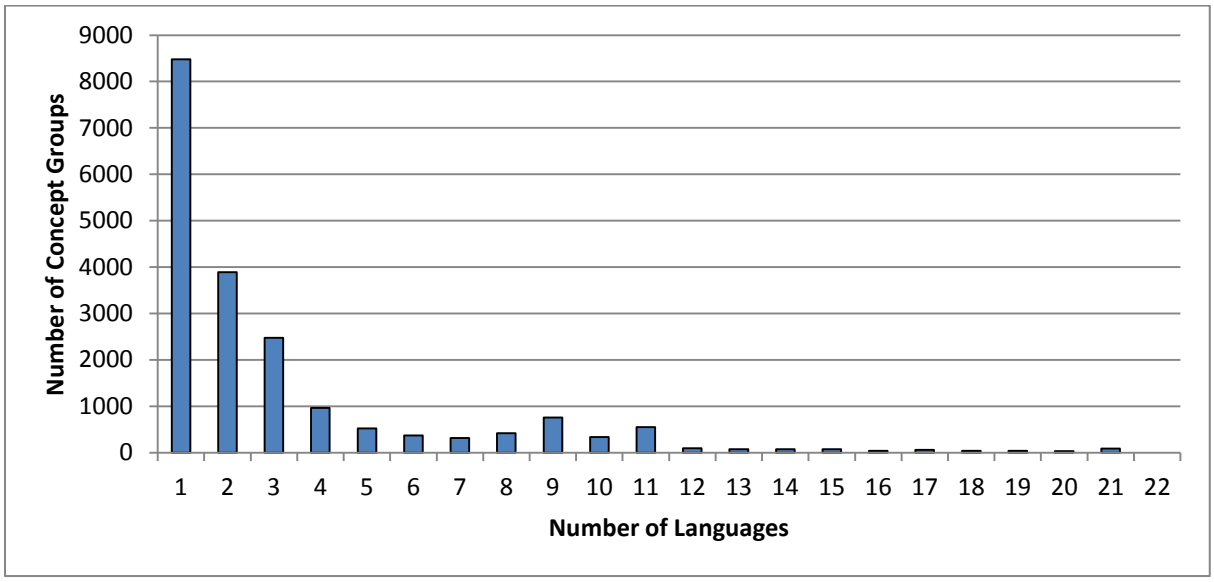

Figure 5: Number of languages per entry group after multilingual grouping with Omeo

This figure shows, for example, that 8,481 entry groups have terms in only one language; 3,892 entry groups have terms in two languages, etc..

Figure 6 shows how languages are represented in the entry groups.

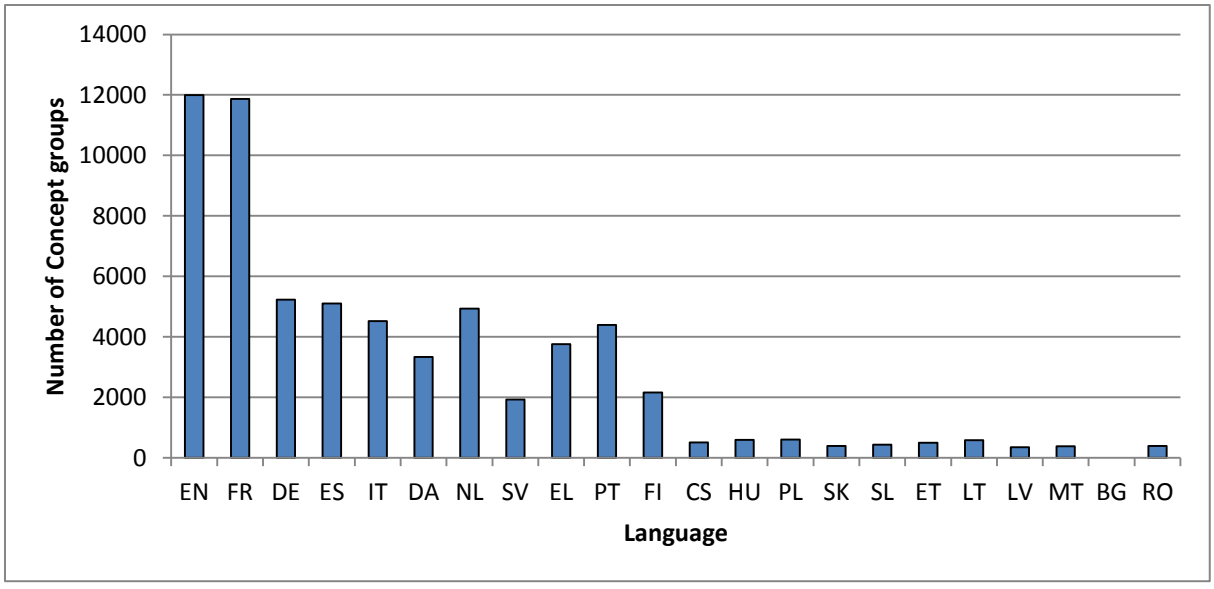

Figure 6: Languages representation per entry group after multilingual grouping with Omeo

This figure shows, for example, that English is present in 11,989 entry groups; French in 11,865 entry groups, German in 5,222, etc..

The purpose of making these groups as broad as possible is to facilitate the Fillup process. The broader the groups, the wider translation suggestions that are found during the Fillup process can be applied. If a translation suggestion is found, it can be applied to all IATE terminological entries that are part of the entry group. During the first round of processing

Omeo will try to create entry groups by linking existing IATE terminological entries based on terms that happen to be the same (or similar) across different terminological entries for specific languages. During the workshop, all suggestions from this first round of processing will be evaluated. During the second round of processing, Omeo will automatically make decisions as to which suggested entry groups can actually be grouped further. During the workshop, IATE volunteers will inspect the accuracy of these automatic decisions as they are reviewing the Fillup results that are based on the multilingual groupings.

So, the results of the first round of Omeo processing will be loaded in the Omeo client tool for the selected languages. The results of the second round of Omeo processing will be used as input for Fillup processing and will then be validated indirectly in the Fillup client tool, as users are reviewing the Fillup results.

2.3.3 Fillup

Common use cases that Fillup addresses are:

- Terminology database needs to be enlarged (scenario 4 in D3.2)
- Language coverage needs to be monitored (scenario 8 in D3.2)

The next step after entry consolidation is to try and find translations for terms that have no equivalent in certain languages. As already mentioned and as illustrated in Figure 4 above, some languages are much better covered in the IATE terminology database (and consequently in our evaluation set) than others. With the Fillup tool, we use external resources to try and find translations for terms for which no translation exists in the terminology database. Fillup will try to find these translations in external parallel data sources, typically translation memories.

In view of the evaluation of the Fillup tool, IATE and the Austrian Parlement provided the consortium with a number of translation memories. Unfortunately, Fillup only found very few translations in these translation memories. Therefore, consortium proprietary translation memories in the trademark domain were also used to find matches. To assess which languages were most suitable for inclusion in the workshop, Fillup was run on all available translated materials for eight language directions. Table 6 shows how many entry groups are to be filled up per language direction and for how many of those Fillup managed to suggest at least one translation.

Translation Direction	# of Groups to Fill up	# of Groups with a Suggestion
EN - DE	4280	356
EN - FR	3121	228
EN -ES	4187	375
EN - EL	4622	366
DE - EN	606	45
FR - EN	2492	258

ES - EN	755	50
EL - EN	299	14

Table 6: Number of groups with translation suggestions found by Fillup during preparatory runs

Based on the amounts of translation suggestions that Fillup was able to find in the provided translation memories during these first preparatory runs, the following language directions appeared to be most suitable for use during the workshop:

- English-German
- English-French
- English-Spanish
- English-Greek
- German-English
- French-English
- Spanish-English

As with the Cleanup results, IATE was informed about these results prior to the workshop in order to be able to find appropriate human resources to participate in the workshop to review the results.

The results of the Fillup run were loaded in the Fillup client tool for the selected languages.

2.4 Evaluation

The results obtained from processing the evaluation set were loaded in the respective client tools and will be presented to and evaluated by the IATE members that volunteered for the workshop of February 5th in Luxembourg.

2.4.1 IATE Volunteers Examine Cleanup Results

As a first step in the workshop, IATE volunteers will review the Cleanup results that resulted from processing the selected evaluation set. Cleanup results will be presented in the Cleanup client tool as a list of potentially problematic terms. Evaluators will have dedicated accounts to work with. This will allow them to select the language they want to review results for and the language(s) they want to see as reference(s).

IATE volunteers will have three options when reviewing a Cleanup suggestion from the tool:

- Accept the Cleanup suggestion as is
- Accept the Cleanup suggestion with editing
- Reject the Cleanup suggestion

Evaluators' choices will be tracked and results will be presented in D4.3, "Evaluation Result Report", as per the evaluation results template presented in Appendix C.

Evaluators will also use Cleanup a second time, later in the workshop, to find potentially wrong translation suggestions that have been made during the second round of Omeo

processing. During this second round of Omeo processing translation suggestions are added to entries by combining the results of the monolingual grouping with the IATE terminological entries and propagating translations across all entries of the multilingual groups that are created. During this second review, Cleanup is used in Advanced mode and only shows potential translation errors.

Also for this second pass, evaluators' choices will be tracked and results will be presented in D4.3, "Evaluation Result Report", as per the evaluation results template presented in Appendix C.

2.4.2 IATE Volunteers Examine Omeo Results

As a second step in the workshop, IATE volunteers will examine all entry groups that Omeo had identified and will assess in how far the groupings were valid or not. Some of the grouping decisions are regarded by the system as safe (high degree of linguistic similarity) and are automatically pre-accepted by the system, while the rest needs to be confirmed/rejected by the user. The user has the option, in any case, to also confirm/reject the automatically pre-accepted cases.

So, the IATE volunteer will have two options:

- Accept the Omeo suggestion
- Reject the Omeo suggestion

Evaluators will also examine the results from the second round of Omeo processing. Whereas the inspection of the first round results involves one language only (evaluators are assessing the results of the monolingual groupings), inspecting the results of the second round involves multiple languages (evaluators are assessing entry groups that have been created by combining all single language groups that resulted from the first round of processing with the original IATE terminological entries). Examining the results of the second round of Omeo processing will therefore not happen in the Omeo client. It will take place in the Cleanup client (second pass, looking for wrong translations) and the Fillup client (assess validity of new translation suggestions) instead.

Evaluators' choices will be tracked and results will be presented in D4.3, "Evaluation Result Report", as per the evaluation results template presented in Appendix C.

2.4.3 IATE Volunteers Examine Fillup Results

As a final step in the workshop, IATE volunteers will review the translation suggestions generated by Fillup. This assessment will take place in the Fillup client tool. Evaluators will have dedicated accounts to work with that will ensure they only see the Fillup results for the language they are supposed to validate.

IATE volunteers will have the following options when reviewing the translation suggestions from the tool:

- Accept the Fillup suggestion as it is and assign it to the relevant term(s)/entry(s)
- Edit the Fillup suggestion and assign it to the relevant term(s)/entry(s)
- Reject the Fillup suggestion and provide a new translation to be assigned to the relevant term(s)/entry(s)

IATE volunteers will be able to choose to validate translation suggestions either on group level, entry level, or term level.

Evaluators' choices will be tracked and results will be presented in D4.3, "Evaluation Result Report", as per the evaluation results template presented in Appendix C.

2.5 Evaluation Results

The outcome of the workshop, i.e. the evaluation results, will be described in Deliverable D4.3, "Evaluation Result Report". The report we will have a quantitative and a qualitative section.

The quantitative section will contain information on the accuracy of the tool suggestions. For each tool, the number of suggestions made by the tool will be compared to the number of suggestions evaluators agreed or disagreed with. Quantitative information will be reported using the template presented in Appendix C.

The qualitative section will contain information on the workshop participants' experiences with the tools. Information about the user experience will be collected using the survey presented in Appendix D.

3 Appendix A - Languages Found in IATE Database Export

The following languages were found in the IATE database export:

Language Code	Language Name
EN	English
FR	French
DE	German
ES	Spanish
IT	Italian
DA	Danish
NL	Dutch
SV	Swedish
EL	Greek
PT	Portuguese
FI	Finnish
CS	Czech
HU	Hungarian
PL	Polish
SK	Slovak
SL	Slovenian
ET	Estonian
LT	Lithuanian
LV	Latvian
MT	Maltese
BG	Bulgarian
RO	Romanian

4 Appendix B - Workshop Program

Time	Event
9:00 – 9:10	Welcome and Introductions
9:10 – 9:30	Introduction to LISE project + presentation of current state of affairs
9:30 – 9:45	Introduction to workshop program
9:45 – 11:00	Introduction to Cleanup and review of Cleanup suggestions by IATE volunteers
11:00 – 11:15	-break-
11:15 – 12:30	Introduction to Omeo and review of Omeo suggestions by IATE volunteers
12:30 – 13:30	-lunch-
13:30 – 14:45	Introduction to Fillup and review of Fillup suggestions by IATE volunteers
14:45 – 15:00	-break-
15:00 – 15:45	Review of Cleanup suggestions by IATE volunteers (Omeo round 2 output)
15:45 – 16:00	Wrap-up

5 Appendix C - Evaluation Results Template

For each review round performed by the IATE volunteers, the number of correction or translation suggestions made by the tool was compared to the number of suggestions that were (1) accepted by the evaluators without changes, (2) accepted by the evaluators after some editing, (3) rejected by the evaluators.

In our assessment of the accuracy of the tools, we assume that both entries accepted without modifications (1) and entries accepted with modifications (2) mark issues that were correctly flagged by the tools. The fact that changes were required before an entry could be accepted should not be counted against the tool. The main goal of the tools is to correctly flag problematic entries. Therefore, only terms or entries that were incorrectly marked as problematic by the tools are counted as errors ('false positives').

The sections that follow provide an example of how results will be reported.

5.1 Cleanup Results

Results per language:

Error Type	**Tool Suggestions**	**Evaluator Response**		
		Accepted	**Accepted w/**	**Rejected**
Misspelling				
Canonisation				
Language				
Equivalent				
Domain				
Translation				
Subset				

5.2 Omeo Results

Results per language:

Tool Suggestions	**Evaluator Response**		
	Accepted	**Accepted w/**	**Rejected**

5.3 Fillup Results

Results per language:

Tool Suggestions	**Evaluator Response**		
	Accepted	**Accepted w/**	**Rejected**

6 Appendix D - Usability Evaluation Survey Template[1]

For each of the tools used in the workshop (Cleanup client, Omeo client, and Fillup client), workshop participants will be invited to provide feedback on the following statements.

USEFULNESS			**1**	**2**	**3**	**4**	**5**	
1.	It helps me be more effective.	strongly disagree						strongly agree
2.	It is useful.	strongly disagree						strongly agree
3.	It gives me more control over the tasks I have to perform.	strongly disagree						strongly agree
4.	It makes the things I want to accomplish easier to get done.	strongly disagree						strongly agree
5.	It saves me time when I use it.	strongly disagree						strongly agree
6.	It meets my needs.	strongly disagree						strongly agree
EASE OF USE			**1**	**2**	**3**	**4**	**5**	
7.	It is simple to use.	strongly disagree						strongly agree
8.	It is user-friendly.	strongly disagree						strongly agree
9.	It requires the fewest steps possible to accomplish what I want to do with it.	strongly disagree						strongly agree
10.	I can use it without written instructions.	strongly disagree						strongly agree
11.	Both occasional and regular users would like it.	strongly disagree						strongly agree
12.	I can recover from mistakes quickly and easily.	strongly disagree						strongly agree
EASE OF LEARNING			**1**	**2**	**3**	**4**	**5**	
13.	I learned to use it quickly.	strongly disagree						strongly agree
14.	I easily remember how to use it.	strongly disagree						strongly agree
15.	It is easy to learn to use it.	strongly disagree						strongly agree
SATISFACTION			**1**	**2**	**3**	**4**	**5**	
16.	I am satisfied with it.	strongly disagree						strongly agree
17.	It works the way I want it to work.	strongly disagree						strongly agree

[1] This survey template is based on: Lund, A.M. (2001) Measuring Usability with the USE Questionnaire. STC Usability SIG Newsletter, 8:2.

18.	I feel I need to have it.	strongly disagree						strongly agree
19.	It is pleasant to use.	strongly disagree						strongly agree
			1	**2**	**3**	**4**	**5**	

The most **negative** aspect(s) of the tool are:

1.	
2.	
3.	

The most **positive** aspect(s) of the tool are:

1.	
2.	
3.	

Other comments:
