Project acronym: LISE

Grant Agreement Number: 270917

Project title: Legal Language Interoperability Services

DELIVERABLE

D6.2 Updated Exploitation and Business Plan

"identifying potential exploitation elements and tools within the LISE project; maintenance and sustainability of the service"

Dissemination Level: PU

Final Version - No. 04

31/07/2013

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

Document Information

Deliverable Number	D6.2
Deliverable title	Update Exploitation Plan
Due date of deliverable according to DOW	31 July 2013
Actual submission date of deliverable	31 July 2013
Main Author(s)	Michael Wetzel, ESTeam; Jochen Hummel, ESTeam, Gudrun Magnusdottir, ESTeam;
Participants	
Reviewer	Tanja Wissik; Elena Chiocchetti
Work package	WP 6
Work package leader	ESTeam AB
Dissemination level	PU
Version	Final

Revision History

Revision	Date	Author	Organisation	Description
01	**31 June 2013**	**Michael Wetzel**	**ESTeam**	**Draft**
02	**30 July 2013**	**Michael Wetzel**	**ESTeam**	**Final for Review**
03	**31 July 2013**	**Elena Chiocchetti**	**EURAC**	**Review**
04	**31 July 2013**	**Tanja Wissik**	**UniVie**	**Review**

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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

1. Table of Contents

2. List of Figures

3. Executive Summary

The ESTeam tool suite used in LISE has been successfully applied and used throughout the European IPR community. End users of the client tools available in LISE are all National offices of Patents and Trademarks, in all member states of the EU and Switzerland as well as WIPO, including the Office for the Harmonisation of the Internal Market (OHIM). The tools have paved the way for advanced structuring and harmonisation of terminology in the IPR community making interoperability in the field of IPR possible. The purpose of LISE is to provide the tools to a wider community and focussing on perhaps the most difficult area of all EU law and terminology, namely the terminology in the domain of social security within the IATE terminological database.

The technology developed through LISE is now applied and available to IATE, and can be made available as a terminology service to other customers (outside of IATE). The approach is now strengthened through terminology best practice research as well as market analysis. This clear insight supports reaching the right audience, i.e. exactly the users that have the need and the customers that fit into the LISE offering.

User and market research also revealed that the pure value proposition and functionality benefits of LISE are so appealing that it asks for a longer term future of the technology, namely its evolution into a comprehensive product offering. Thus, LISE's technology will not be a privilege of very large institutions and corporations, but available to the very wide audience of terminologists, taxonomists and knowledge workers. The LISE project triggered the foundation of a new SME, the Coreon GmbH (www.coreon.com).

4. Introduction

This is the final version of the LISE Exploitation Plan. The early version D6.1 had been developed for month 6 (M6) of LISE. While the D6.2 Exploitation Plan deliverable was originally – according to DoW – expected at M30, the LISE Consortium, together with LISE Reviewers, decided to anticipate D6.2 and had a draft version ready around M24. Thus, LISE could react on further input from the Advisory Board (in M25) as well as from the Reviewers (in M26). This Exploitation Plan for M30 now includes also a way more elaborated version of a Business Plan. A first Business Plan had been delivered at M12, but a more precise view on LISE and its market asked for an update of this – see section BUSINESS MODEL IN A LEAN STARTUP APPROACH.

All tasks foreseen in D6.1 (market research, competitive analysis, methodological exploitation, business potential) have significantly progressed and are now reflected in this document. While D6.1 was early, most of its anticipated estimations turned out to be true. They are now confirmed through research and input from several sides. Particularly the user research from project partners EURAC and University of Vienna (see D3.2 Report Workflow Adaptation for LISE) elaborates deeply about the use cases and feature benefits that the LISE technology provides. It clearly describes the needs and situations where LISE plays an indispensable role. WP4 leader CrossLang confirmed and proved the concrete benefits of the technology.

All LISE partners verified the LISE value proposition during several conferences, trade shows and face-to-face prospect visits. All this lead to a comprehensive and realistic estimation of LISE and shaped its value proposition. See the MARKET ANALYSIS RESULTS for details.

LISE Exploitation itself emerges into two directions. The theoretical outcome (WP3) is available as an e-book with ISBN id. The technological outcome is part of a 3-phase business and development plan: First, as a terminology service to larger institutions and corporations; (LISE technology and applied to IATE and other customers). In the second phase, the technology behind LISE evolves into the product offering of Coreon, a company that has been created to address exactly the needs of the LISE audience. In the third phase, the Coreon software will become more and more an off-the-shelf, standard software product, so that also smaller, lower-budget audiences can benefit from the value that LISE technology delivers.

5. Market Analysis Results

LISE tackles a challenge that no other software vendor nor consulting business addresses today: improving the quality of terminology resources, particularly large volume resources. There is no language technology that can provide the services proposed.

"Data cleansing" however is an undertaking every organisation is confronted with (e.g. cleaning address book data in CRM systems, removing duplicates in ERP managed inventory lists, etc.) Yet, no one has addressed this for terminology resources.

In terminology work 'quality' meant mainly two things until today:

1) Guaranteeing the formal consistency within one single "entry", normally by using input templates (for instance, by making a Definition field 'mandatory').
2) Treating terminology resources as a key factor for improving the consistency of texts when writing and translating. This justifies the ROI into terminology; this is also the reason why we see many terminology tools as part of authoring and translation tools (see COMPETITIVE LANDSCAPE below).

Yet, no service nor product looks in a holistic view onto a resource. No one can measure nor assure the overall quality of a terminology database. In simple words, the answer to the question: "Is this resource good to use, or should I refrain from using it?" We illustrate now why data cleansing is required for terminology resources, who could compete in this market, and where today's strength and weaknesses of LISE are.

5.1. User Needs

When starting with terminology work, data maintainers are initially faced with the challenge to quickly populate the database. This is to achieve the early goal to relatively fast cover the targeted domain. Population happens manually by entering one entry after the next or in "batch", namely by importing already existing lists and other terminology databases. This happens often together with terminology extraction, i.e. text harvesting technologies to spot new term candidates. Population can have two different focuses: 1) adding new entries (*concepts*) to better cover the domain and 2) multilingual expanding to cover more languages and cultures.

After years of collecting and storing terminology data organisations are today faced with the challenge that the quality of their resources is degrading. For instance, users suffer from "noise" (doublettes, spelling variants) and uncertainty (inconsistencies). This negatively impacts on users' trust into the resource. This often means that the investment made over years is at risk!

Many interviews and feedback on conferences in the LISE context reveal two large capabilities that are not covered by existing tools. Users either simply do not address the issue or develop workarounds for both: a) semi-automated maintenance, cleaning and

improving large terminological resources; and b) terminology-focused collaboration platforms that cater for an inter-institutional context.

Workflow research (see D3.2) outlined several major use cases:

1) Update of a terminology database
2) Import/addition of a terminology collection into an existing database
3) Merging of terminology collections
4) Enlargement of a terminology database

All these use cases call for technologies that LISE offers: cleaning, improving, and enhancing a terminology resource, particularly when bringing two or more existing resources together. Further, it asks for a collaboration platform, because the owners of the previously distinct resources will have to work together to put their expertise / knowledge into one unified collection.

Interviews conducted by ESTeam in 2011/2012 also recorded needs expressed like[1]:

- "Need a meta-engine that helps comparing and harmonizing terminologies from different sources"
- "Need improved workflow modeling"
- "Need for cross-departmental cooperation" and "terminology must reach users outside my department"

Harmonizing and maintaining termbases manually is not efficient and ultimately fails its purpose. ESTeam's experience shows that if the amount of entries in a database reaches a ~2,000, it is already no longer possible to maintain and to supervise the content manually(human activity of leafing through the terms or by searching and filtering). Only software with sophisticated linguistic algorithms can deliver this.

It has also been continuously confirmed that terminology activity involves many contributors next to the "linguists". The terminology group is often organizationally part of the translation department, but delivers value to many other departments. In other words, usually we do not see a dedicated "owner" of the terminology resource, acting with a dedicated cross-departmental focus.

5.1.1. Example: IATE, EU

The termbase of the European Institutions is the concrete focus and pilot case for LISE. The data contained in IATE has grown over decades and covers more than 20 languages. It is the result of bringing several European termbases together into one common system. This is a continuous process. Initially the existing, larger resources, such as Eurodicautom or Euterpe,

[1] See D6.1 Exploitation Plan (Early version)

have been brought together. But as it happened before, due to the entrance of Croatia into the European Union, today new domains as well as a new language must be covered in IATE.

Throughout WP2, ESTeam has analyzed and illustrated to what degree the quality of IATE needs improvement. This hypothesis has also been confirmed during several user workshops and sessions with IATE maintainers. Concrete figures are available in the WP4 (evaluation results) run by LISE partner CrossLang together with ESTeam. Main issues, presented by a IATE representative at the LSP conference in Vienna on 8 July 2013 are:

> *"Duplicates, Incomplete entries, Typos, Broken hyperlinks, Obsolete data, Lexical items (LGP) instead of terms (LSP), Non canonical forms, Phrases instead of terms (particular collections), Multiple terms in term field."*
>
> IATE maintainers have
>
> *"DIFFICULTY TO ENSURE CONSISTENCY, IN TERMS OF QUALITY, COVERAGE, COMPLETENESS"*[2].

Today, this problem is not tackled in an appropriate manner. Ad-hoc intervention when consulting data or when receiving a comment from users is happening, but maintainers "… lack automation in problem detection and solving".

5.1.2. Example: Video and Imaging Company, UK

A Japan based manufacturer of cameras and video capturing hardware with its European headquarter in the UK experiences similar challenges. Over the years two separate termbases have been maintained, one for the terminology of still cameras, one for video cameras. Nowadays, the technologies are coming together: every recent camera is equipped with features to also capture motion pictures. The company therefore merged the business units. By consequence, also the maintainers of the terminology resources are now asked to bring their resources together into one.

While on a smaller scale than IATE, the challenges are similar: Duplicates, inconsistencies, gaps and depth of domain coverage. A maintainer of the resource, whom ESTeam interviewed at the TKE conference in Madrid, June 2012, expressed a serious need for automatically spotting such weaknesses. Without, users suffer from low efficiency in the translation process, with a concrete impact on higher operational costs and delayed time-to-market.

5.1.3. Example: Austrian Parliamentary Administration, AT

The Austrian Parliamentary Administration is maintaining a list of keywords (terms) to index and to analyze the content of parliamentary materials. The list contains ~13500 records. While not developed as a classic terminology database, the list shares many characteristics of a terminology resource. In analogy, also this list suffers from problems known from

[2] Slides available from LSP Conference: http://lsp2013.univie.ac.at/

terminology resources. The paper "Legal and Legislative Taxonomy Management" developed by the Austrian Parliamentary Administration as part of the LISE project illustrates these cases in detail[3]:

- Typing errors: *Afghanisatan-Konferenz → Afghanistan-Konferenz*
- Term formation, plural vs. singular: *Autobahnen → Autobahn*
- Inconsistent notation: *Bioethikkommission → Bioethik-Kommission*, *Akademiker → Akademiker/innen*, *Bundesgesetz über das Verbot von Anti-Personenminen → Verbot von Anti-Personenminen, Bundesgesetz über das*
- Incorrect categorization
- Inconsistent acronyms and abbreviations: *BKA*, *Bundeskanzleramt → Bundeskanzleramt (BKA)*
- Duplicates: *Bundes-Gleichbehandlungskommission / Bundesgleichbehandlungskommission*
- Missing permutations: Publishing acronym-term connections in two ways: acronym first, followed by term full form and vice versa

"More than 13 thousand keywords (main, index and sub-subject keywords) and more than 18 thousand references (autonomous and dependent references) are reason enough why the manual revision of the keyword lists are not possible any more." Concretely, the Austrian Parliamentary Administration expresses two different needs to improve the resource: "The first is the area of data management (settling of doublets, spelling mistakes, etc.) and the second is the area of developing the German terminology to a multilingual terminology."[4]

Currently the search interface on www.parlament.gv.at is monolingual. Yet, a

> *"… multilingual search could be very useful in regard of multilingual parliamentary materials like treaties or documents of the European Union."*[5]

5.1.4. Common Patterns of User Needs

We recognize several common patterns amongst these prototypical cases:

- **Grown resources**: The data resources do exist and have been developed over several years. Their size is significant so that manual spotting of quality issues is not possible.
- **Combination of data**: The resources are often a result of merging or integrating several domains into one; this activity was triggered by some higher level, organizational mandate.
- **Beyond organizational boundaries**: There were and/or are the owners of the previously separate resources. Cross-organizational or cross-departmental collaboration is required. Different expertise and interest is coming together.

[3] Paper is available at www.lise-termservices.eu/downloads . We reference a draft version from 16 July 2013.
[4] Page 22.
[5] Page 24.

- **Multilingualism**: Maintainers and users of the resources are not working in one single language. Resources cover many languages (20+). Some very global operations today have to cover up to almost 100 languages, for instance Microsoft.
- **Need for Data Maintenance**: All express the clear need for data maintenance. Yet, human spotting of errors is not possible.
- **Different individual cases**: While there are common problems, each organization has different setups, workflows and individual needs.

5.2. Competitive Landscape

LISE's goal is to provide tools and best practices to clean and improve terminology resources. Who else is in this market? Who is developing software technologies that can also address what LISE tries to resolve?

We can treat the terminology technology market as a "sub-market" of language technologies. The LT-Innovate report "Status and Potential of the European Language Technology Markets"[6] identifies three segments to structure the market: *Speech*, *Translation*, and *Intelligent Content*. While the report informs about the size of each of these segments, there are no figures about a "terminology technology market" on its own. Terminology (and very related approaches such as taxonomy, thesaurus, and ontology management) are mentioned under both Translation and Intelligent Content; also, dictionary / lexical work is part of Speech. ESTeam did not find any other research or study on the terminology market size of its own. This makes it also very difficult to project revenue figures in the context of the business plan.

Looking closer to the terminology technology market landscape, we identify four types of suppliers: Translation tools companies, authoring tools companies, ontology / taxonomy / thesaurus tools companies, and dedicated terminology technology companies.

	Value Proposition	Strengths and / or Opportunities	Weaknesses and / or Threats	Known Players
Translation tools (CAT)	"Approved multilingual term lists help you to translate efficiently and consistently"	+ Big user base + Proven off-the-shelf software	- Focus on terminology consumption not maintenance	• SDL Language Technologies • Across • Kilgray • STAR

[6] See http://www.lt-innovate.eu/resources/document/press-release-lt-innovate-publishes-landmark-report-state-european-language-techn

	Value Proposition	Strengths and / or Opportunities	Weaknesses and / or Threats	Known Players
Authoring tools	"Approved, monolingual good-versus-bad list help you to write consistent source texts, thus increasing readability and reducing translation costs"	+ Confirmed market trend to move terminology work from the translator to the author	- Focus on terminology consumption but not maintenance	• SDL Structured Content • Acrolinx • Congree
Ontology / Taxonomy / Thesaurus tools	"Find smarter by expanding search keywords to broader or narrower concept terms"	+ Enterprise search is a big market	- Usually monolingual - Besides (ex-) Autonomy, all of them rather still "garage" companies	• Semantic Web Company • Hewlett Packard (Autonomy) • SmartLogic • Top Quadrant
Terminology tools	"Capture your organization's language and knowledge"	+ Clear focus on terminology needs	- Little integration into CAT or authoring tools - Focus on term storage, not maintenance - Termbases managed as silos, not for inter-departmental collaboration	• InterverbumTech • Acolada • Flashterm

From these four software types only the last one, namely the dedicated terminology tools, may one day see a functionality that covers what LISE delivers: not only storing isolated entries, but a collaborative development and maintenance of one or more termbases plus software assisted identification and processing of language inconsistencies and redundancies.

To summarise, the closest competition today is quality control software working at document level and providing terminological consistency, as well as monolingual terminological expansion and data mining software. Yet, none of these are targeted enough to harmonising and improving grown, multilingual, interinstitutional terminology resources.

5.3. LISE SWOT Analysis

Looking closer onto LISE itself: What are the helpful, what are the harmful aspects of today's LISE offering?

	Helpful	Harmful
Internal	**Strengths** • Unique linguistic algorithms to semi-automatically improve the quality of terminology resources in many different aspects • Proven to work for large data • Knowledge and experience of staff • Secure; can run in-house • Covers all EU languages	**Weaknesses** • Limited scalability: Every customer requires customisation • Tools not so user friendly • Separateness of Collaboration Portal, Tools, and master data itself
External	**Opportunities** • After years of collecting, more and more terminology resources are getting "messy". Needs will grow. • Increasing cross-border, cross- and inter-organisational "one language" initiatives that trigger merging of previously separate resources. Market will grow. • Recognition that low-quality resources cause damage • Recognition that language resources build leverageable knowledge (and go beyond simple CAT-Tools' needs)	**Threats** • Competitors to develop similar algorithms • Less cross-institutional activities • Less mergers and acquisitions • Little demand for high quality resources

The strong advantage of LISE is its unique technology, which addresses concrete user needs; i.e. LISE is not only a marketing story, but there is concrete, valuable technology behind. The algorithms have been fine-tuned over the years. Competitive offerings will have a hard time to catch up.

The biggest adoption obstacle of LISE is probably that it cannot be used out-of-the-box. It requires customisation and terminology service; therefore, today it is an offering that cannot so easily be mapped to the different needs of thousands of customers.

6. Opportunities: Technical Exploitation Elements of LISE

6.1. Recap – the LISE Workflow

To see how the different technologies and data come together, the top level LISE workflow should be kept in mind:

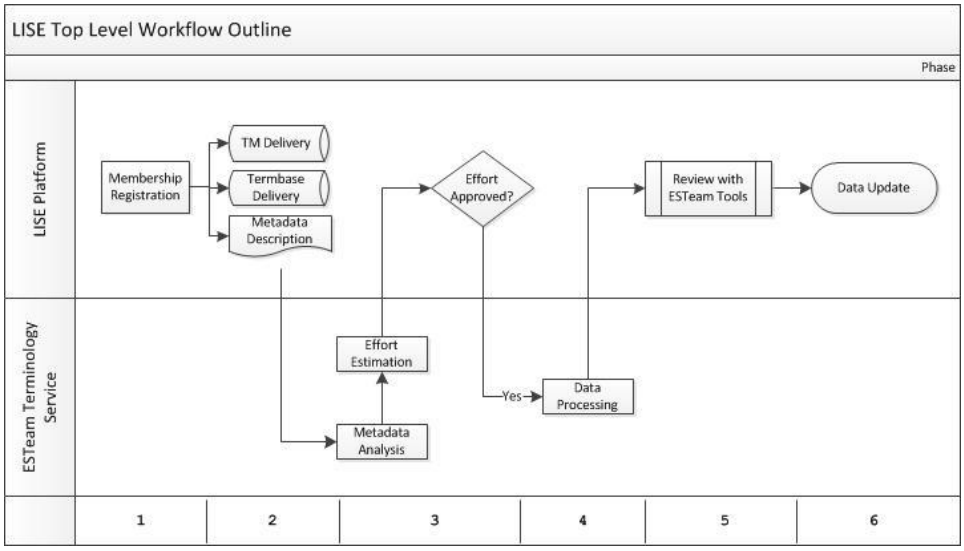


Figure 1 LISE Workflow Outline

1) A user or an organisation applies for membership in the LISE Collaboration Portal. The account is set up.
2) The member delivers terminology and translation memory data for processing, together with a metadata description based on which the processing takes place.
3) Estimation Phase
    a. ESTeam Terminology Service team reviews the data and estimates the effort.
    b. The LISE member approves the effort and estimated timeline for delivery.
4) The data is being processed by ESTeam AB through the ESTeam Language Server
5) The processing results are made available to the user, together with the customised ESTeam Tools client software for user processing and review of the results. Terminology Expert Services may contribute here.
6) Target databases are updated.

6.2. LISE Technology: Exploiting the ESTeam Tools

The ESTeam Tools are data editing tools used on a Microsoft Windows workstation. They allow post-processing, reviewing, and finalization of the initial results created earlier by the ESTeam Language Server. ESTeam Language Server is run as a service by ESTeam AB.

- **ESTeam Cleanup** helps to remove errors in resources. It verifies data against its metadata description, such as domain assignments. It identifies misspellings or

wrong translations. ESTeam Cleanup has been admired as “a terminologist’s dream” during a workshop with IATE users in February 2013.

- **ESTeam Omeo**: ESTeam Omeo helps to harmonize and streamline terminologies; it identifies doublettes and it groups similar terms, for instance variations in spelling like *eye glass* and *eyeglass*. It compares units in one language to find alternatives that share the same meaning but are written differently. It is a pre-step to make subsequent tasks, such as the Fillup process, even more efficient.
- **ESTeam Fillup**: ESTeam Fillup finds translations of existing terminology in translation memories and proposes the translations. This is a bilingual terminology extraction that benefits immensely from previous ESTeam Omeo runs.

It is important to understand that the ESTeam Tools present suggestions. It is always the human terminologist that has the final saying. This has been confirmed as a top priority characteristic of terminology workflows (see the guidelines as developed in WP3).

The ESTeam Tools are exploitable individually or together. Applying them in the above sequence is generally recommended: first clean the data, then group concepts, and only then add further languages. Yet, deviations from this sequence may be very advantageous, as the processing of the IATE showed[7] where a fourth step was introduced, namely applying ESTeam Omeo a second time.

6.3. LISE Technology: Exploiting the Collaboration Portal

The LISE Collaboration Portal addresses the needs for inter-institutional terminology work. It takes into account that all contributors and members come from different institutions with different backgrounds. A common ground of knowledge or user training cannot be assumed. The portal therefore has been developed with a strategy to make it as user friendly as modern social media applications and with the goal to reduce administrative overhead as far as possible – which concretely responds to the fact that there is often no dedicated, cross-departmental owner of the resource.

The portal caters for both data and information exchange needs: topics and reply-to-posts, bookmarking and tagging of topics; attachments of files and linking to master data entries (in LISE this is the online link to the http://iate.europa.eu site); a voting feature to drive decisions, as well as user defined visibility / privacy of a topic.

The LISE Collaboration Portal is exploitable separately from the ESTeam Tools. Yet, both play together as they complement each other in making terminology workflows more efficient.

6.4. Supplementary Services

Terminology Expert Services: ESTeam AB is providing data analysis, data conversion, and terminology services in an efficient and highly automated way. The main goal of these

[7] See the report on WP2 data processing activities in D2.2.3 LISE Service Version Three

services is to analyse unclean terminological data in order to identify and report on inconsistencies or missing term translations, undesired duplicates or clusters of term variants. The results of the service can then be processed and reviewed within the ESTeam Tools.

LISE Partners provide consulting and terminology expert services on how to best and most efficiently clean and review terminology sources.

Support Services: Companies and organisations contributing to LISE offer both technical and workflow related support and consulting services.

7. Exploitation Concept and Strategy

7.1. Methodological Exploitation

The current dissemination activities already proved that LISE has a great potential regarding the methodological exploitation. Terminology management workflows are scrutinized in WP 3. The tools can be integrated for a certain task into existing methods and workflows, since they are task-oriented; they do not have to support the whole complex terminological workflow of a certain institution.

Consulting activities arise from this methodological exploitation and from the development of best practices. Consulting activities regarding workflow and methods in terminology work are part of the future exploitation of the project results.

7.1.1. Raising Awareness for High Quality Terminology Resources

Terminology resources enable high quality and efficient documentation and translation. LISE efficiently brings high quality into these terminology resources, thus making sure that users are not using faulty or mediocre data. As a consequence, users trust their resources better, which leads to higher acceptance and use of the resources.

Clean data and less noise also means that data maintainers can focus on the right records and are not bothered with tedious double and redundant work. This gain in efficiency leads to a better ROI of the terminology system and its data. The evaluation report dives into such aspects.

Generally, neither academic research nor market research of terminology tools companies have delved deeper into this topic. All LISE partners see quite a potential to raise awareness among terminology maintainers and consumers for the effect of assuring quality in terminology resources. This has a concrete economic (better ROI through efficiency gains) as well as a psychological effect (higher trust, less frustration).

One task in this context is to identify the “low hanging fruits”. LISE is focusing on IATE and also on data from the Austrian Parliament. But who else is concretely facing the same challenges? Who has language resources that were growing and growing and are nowadays seeing a decline in quality, which leads to inefficiencies and frustration? Workflow research gives concrete answers on where to search for. And as part of the exploitation, the whole idea of “only high quality resources can deliver value” must become a topic of conferences, industry events, and scientific papers. Potential customers will thus become aware of an unprecedented solution for an often existing - yet still hidden - problem.

7.1.2. Outlook: Develop a Teaser

A clearly convincing method is to develop some kind of “teasing” capability. Nothing is more convincing than seeing the theoretical arguments proven by real life facts. It was astonishing

to see how the energy in the workshops with the IATE users changed as soon as the users touched the results with their own data.

Every terminology resource looks different and also every workflow looks different. Yet, it is well feasible to develop, in a future version of the ESTeam Tools, a kind of a demo mode that, without requiring large customization work, analyses a terminology resource and reports on identified problems. This demo mode could for instance report on the overall number of found problems and display the first five ones.

This activity significantly helps to foster the hypothesis and to evangelise a new, quality-driven view of terminology work.

7.2. Software and Services Exploitation

Software and software services exist for terminology as simple terminology databases, mostly used for translation support. Maintenance and enhancement of the terminology resources is normally carried out manually using normal search to view and edit the contents of the term database.

The LISE software is unique on the market: it supports cleaning errors in terminology resources and provides the means for internal consolidation and interoperability across terminology in the same language and domain.
There is other software doing translation data mining, but the level of accuracy provided by the LISE technology by using metadata is much higher. The combination of software and services provided by LISE, namely the terminology expertise, user support and training, are also new.

7.3. Answer to Competitive Landscape

LISE technology is unique. There is no direct competitor tackling the problems that LISE addresses. Furthermore, the algorithms underlying the ESTeam technology have been developed over more than a decade; their origin is in machine translation research and have been tuned and adapted over years. Within LISE they now have been successfully adapted from the IP domain to more general purpose, “human” terminology, with a concrete focus onto the legal / administrative domain. In contrast to all terminology and related offerings that only store and retrieve data, the ESTeam algorithms process the data. This maturity of the algorithms means also a high entry barrier for others.

Therefore, all other players in the market are rather to be seen as partners than as competitors.

- **CAT and authoring tools**: LISE complements the low profile terminology tools that ship with translation memory or technical writing software. As with IATE, LISE receives the resource, ideally in TBX format to minimize conversion and analysis effort, then analyses and processes it.

- **Terminology tools**: Unless vendors like Interverbum Tech or Flashterm dive deep into computational linguistics, start developing parsers, morphological analyzers etc., LISE is a unique value added complement for their customers.
- **Ontology/taxonomy tools**: Larger ontologies and taxonomies suffer similar data consistency problems that are known for terminology resources. However, their advantage is that they are used from the very beginning with a sound method aimed at keeping structure in the data. This is in contrast to all terminology tools, where thousands of records reside next to each other and where maintainers run the risk of losing control.
  Again, LISE complements such tools; yet, their more "intelligent", meaningful organization of data is a big algorithmic helper to create even better suggestions. Since ontologies and taxonomies are usually largely monolingual today, ESTeam Fillup is a big efficiency improvement when populating such resources with multilingual data – as the analysis of the Austrian Parliamentary Administration stated (see above under USER NEEDS).
- **Homegrown solutions**: IATE, Microsoft, IHTSDO Snomed – such proprietary solutions are going beyond what an off-the-shelf software delivers (or was able to deliver upon the time when the decision was made to develop an individual solution). Their strength lies usually in their extreme customization to an organization's business processes. Nevertheless, the resources suffer the same flaws as other terminology databases. Even more, usually designed and developed by IT project companies or internal IT departments but not by experts in language technology, functionalities to linguistically improve the data are absent.
  Again, LISE complements such solutions. Efforts may be a bit higher, since custom solutions not always need or rely on exchange standards such as TBX[8].

7.4. Exploitation Strategy: Sustainability through Three Phases

While the LISE technology is unique, we have also identified its weaknesses due to the lack of cohesion in terminology tools: first, the necessary disconnectedness of the ESTeam Tools/ESTeam Language Server from the normally on-site running master terminology management systems. Second, LISE is currently a relatively exclusive offering. This has its economic advantages regarding profitability, since it requires customization for every new customer/terminology database scenario. It cannot be used out-of-the-box.

In LISE's market research a question that came up repeatedly concerned what would be necessary to overcome these weaknesses. The need for it was expressed very explicitly on the one hand due to the appeal of the offering, and on the other hand since almost all visible competitors from the CAT world seem to no longer invest into terminology technologies – their business focus is on consuming terminology but not on maintaining it efficiently.

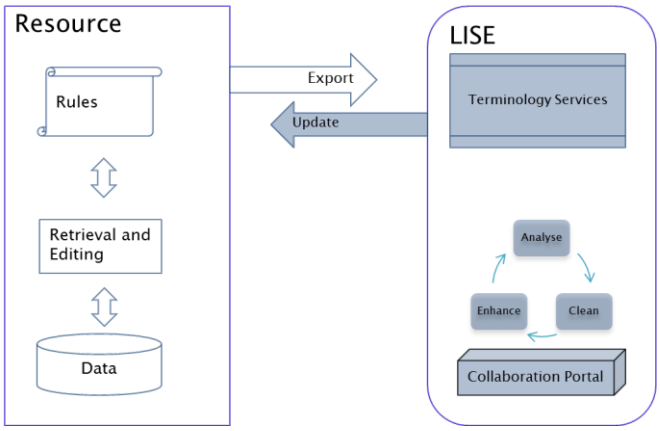
[8] And indeed the IATE data was delivered in CSV-like Excel files that required quite some analysis and then conversion effort. See D2.2.3 – Final LISE Service Version Report

Authoring tools also only consume terminology data, taxonomy/ontology tools do not manage language. And dedicated terminology software, like TermWeb from Interverbum, is also not known to tackle this problem.

So, LISE has confirmed the need for such a technology, but the market continues to lack an offering that brings together terminology storage, terminology maintenance/workflow, and terminology consumption in a user friendly, collaborative approach. As an answer to this, ESTeam follows a clear market opportunity for such a comprehensive offering, focusing strongly on the terminology maintenance / workflow needs by incorporating and productizing algorithms that make LISE unique.

7.4.1. Phase One: Applying LISE Today

1) A technologically independent system acts as a master. It stores the data ("Resource"), it has data manipulation features, and often, workflow and business rules.
2) The data is exported. This "snapshot" of data is analysed through LISE, improved. Results are discussed and, after approval, reflected back in the master system.



1 – Applying LISE Today

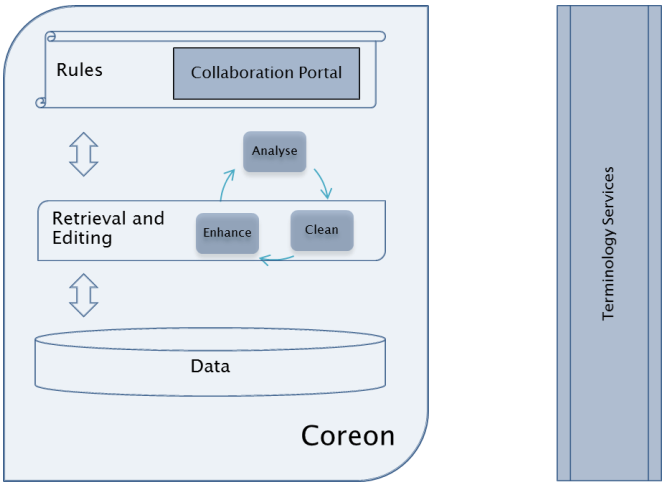
As mentioned under the LISE SWOT Analysis, this approach has some flaws. The LISE process resides outside of the master data system. This requires a) export / import runs, and b) the LISE process always works on a snapshot; the processed data may be out-of-sync from the master data. Therefore, a better integrated approach was expressed by users.

Nevertheless, LISE partners can execute this approach already today and can generate concrete business value.

7.4.2. Phase Two: Applying LISE through Coreon

Coreon – Knowledge Meets Language, a spin-off from ESTeam AB, is the answer to a gap in the language technology market landscape. Initiated through LISE's substantial market research, in 2012 ESTeam had decided to create a dedicated company that addresses the need to manage and explore large volumes of multilingual terminological, taxonomical and ontological resources. While terminology tools as well as taxonomy and ontology tools exist, LISE identified the need and delivered modules for getting large, multilingual resources under control. Through Coreon, a comprehensive product offering for managing multilingual taxonomies and terminologies, the LISE technology will be maintained and sees a sustainable future.

1) The LISE technology is evolving into modules of the master system. For instance, functionalities delivered through cleanup are continuously available while editing or changing data.
2) Terminology workflows are equipped with an integrated collaboration platform, to discuss and decide on data changes.

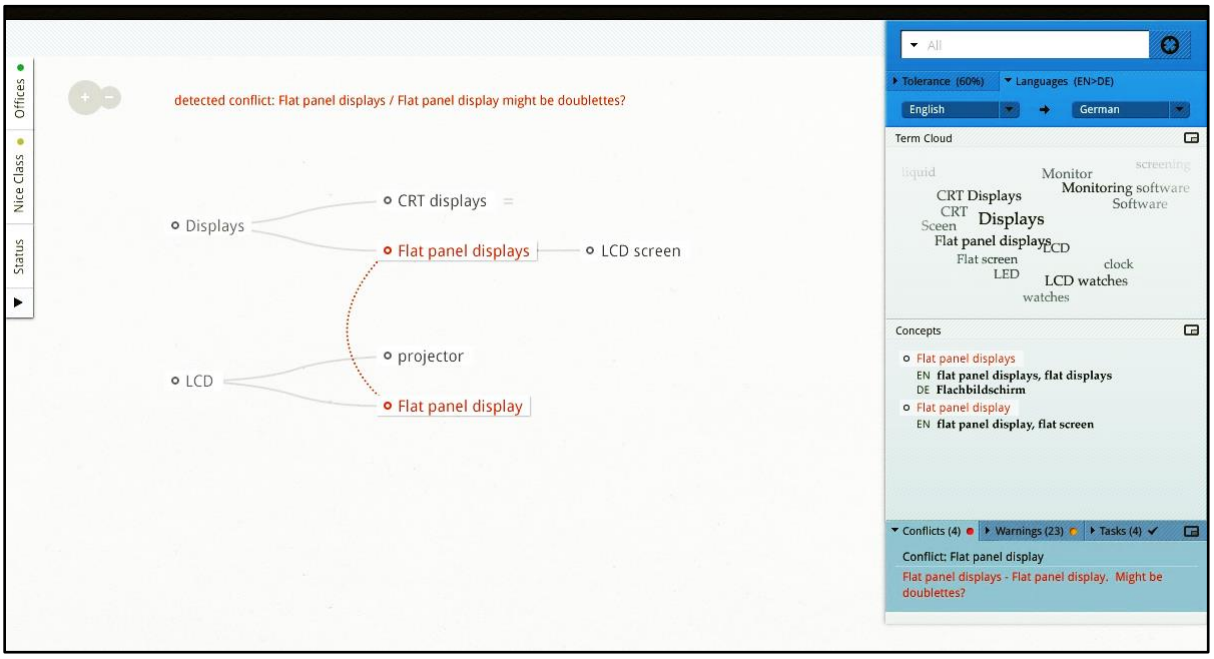


2 – Applying LISE through Coreon

The LISE Collaboration Portal as well as the ESTeam Tools become part of a unified system and user experience. This addresses one of the identified major weaknesses of LISE, namely that the technology resides outside of the master data technology (see LISE SWOT ANALYSIS).

Coreon, a system to manage multilingual taxonomies and terminologies, is delivered as a browser solution, developed with most modern web technologies and following UI standards. As a side-effect, the solution is platform neutral and no longer limited to Microsoft Windows workstations. In this context the whole user interface of the ESTeam Tools is rethought and integrated into terminology management workflows.

It is the normal case in taxonomies but not yet known from terminology management systems: Coreon encourages the organization of concepts in a hierarchical, structured way. Now, when spotting flaws in such a so-called *concept map,* the user benefits from a much more visual and semantically enriched information. For instance, when spotting a doublette, the UI would always present broader concepts; through this transparency and understanding of the meanings, the data maintainer can much easier and faster come to a decision.



3 – Cleanup-like doublette spotting rendered in Coreon

Every single data change – i.e. just adding one term – may cause side-effects that impact the overall quality of the resource. An always visible notification widget keeps the data maintainer informed about conflicts and serious flaws, less serious warnings, and also points him to collaboration activities, tasks, recent votings etc. The algorithms from the ESTeam Tools work in the Coreon user interface like "linguistic alerts" in a user's notification and todo list.

4 – Alerts and Tasks in Coreon

This Phase Two of exploiting LISE has a dependency on the Coreon development roadmap. The LISE technology evolves step-by-step into Coreon. First, ESTeam Cleanup routines are ported, together with integrating the LISE Collaboration Portal. In Phase One, the Collaboration Portal required a customization for every different master terminology system (esp. when the feature of direct links from a topic to a term entry are needed). In Coreon this is no longer necessary. The Collaboration Portal and the master data are parts of one system. Cleanup spottings feed the Collaboration Portal's task list. Omeo and Fillup features then follow.

LISE technology adds several absolutely unique selling points to Coreon. These functionalities will significantly contribute to the economic success of Coreon.

7.4.1. Phase Three: Towards Commoditization

The strategy during Phase One is to generate some early success stories; yet, it is a relatively exclusive audience. In Phase Two several functionalities are more standardized, so that they

can be positioned as product features and no longer as a customizable service. The audience is still an exclusive circle, but definitely larger.

Then, in Phase Three, Coreon and its built-in LISE technology shall evolve towards a commodity. In other words, routines like Cleanup or a voting support should no longer produce a wow effect, but rather be expected from state-of-the-art terminology and taxonomy management systems. Coreon with its LISE technology is the reference application for this.

7.5. Strategy Summary

As of today, the LISE exploitation strategy is a service-oriented strategy. LISE is a terminology service that improves the quality and coverage of large terminological resources through a collaborative, semi-automated approach with the goal to increase reach of and trust into the resource, thus strengthening its value.

Yet, with a service-oriented approach only some markets can be reached. Therefore, it has been decided to evolve LISE into a product offering so that wider markets can benefit from LISE technology through an off-the-shelf software offering.

8. Business Model in a Lean Startup Approach

8.1. The "Lean Startup" Approach

Steve Blank, Consulting Associate Professor at the Stanford University, developed the so-called *Lean Startup* method. It is the answer to today's fast evolving businesses; it analyses particularly the criteria that make startups successful. This method is to replace more and more "established" procedures when creating new enterprises. In his book "The Four Steps to the Epiphany"[9] Steve Blank outlines: instead to finalize a complete business plan and develop a product over years "in stealth mode" nowadays ideas, features, modules are tested and developed together with potential users. The value propositions are discovered and developed together with potential customers.

The so-called *Business Model Canvas* that we intensively use below is described in the book "Business Model Generation", by Alexander Osterwalder and Yves Pigneur[10]. Osterwalder and Pigneur pick up Steve Blank's method and develop a tool to concretize key aspects of a new business. Steve Blank praises the business model canvas as the new "standard framework" for business model development[11].

8.2. Phase One – LISE as a Service

A service oriented model is the approach to start exploiting today's LISE offering. This takes the three main technical and logical conditions into account: 1) Master data resides in a different system (the customer's terminology system), 2) ESTeam Language Server and ESTeam Tools require customisation to the metadata, and 3) Workflow analysis is needed. Therefore, a simple product / software-license sale driven approach is not sufficient; a service driven approach seems more appropriate.

8.2.1. The Value Proposition

Combining all the research, user feedback and advisory board recommendations as well as the technological offering, the value proposition in Phase One is the following:

LISE protects the investment that has been made into terminology resources by identifying areas of problematic quality in termbases, which propagate to produce problems in legal texts and erroneous information.
LISE is available as a service; it semi-automatically detects errors and flaws, spots redundancies and suggests translations for not yet fully covered languages. LISE is indispensable to collaboratively assure a continuously high quality in termbases. It increases the trust, the usefulness and the impact of such precious resources.

[9] Steven Gary Blank: The Four Steps to the Epiphany. 3rd edition, 2007.
[10] Alexander Osterwalder, Yves Pigneur: Business Model Generation. 2010, Wiley, Hoboken, New Jersey.
[11] See also Steve Blank: Schneller Gründen. Harvard Business Manager, German edition, 07/2013, page 22-31.

8.2.2. Target Markets

Focus segments are these environments where the use cases that have been identified in the user needs' analysis often take place: merging / updating termbases, enhancing termbases with other languages. These are very typical for environments where two or more organisations are required to cooperate or are even merging business units or departments. The European Union is obviously a very apparent example for this. Therefore LISE focuses on identifying opportunities that share the following characteristics:

- **Inter-institutional / cross-border initiatives**: Situations where two or more organisations have to collaborate
- **Need for high quality language resources**: Verticals where language precision is mandatory, such as the governmental, legal, finance/insurance, or health care markets
- **Have large amounts of data**: Organisations with terminology resources that are big enough to make a human approach to get them under control will fail.

The above characteristics indeed match the IATE situation. Also other potential prospects identified during the LISE market research share more or less these characteristics; to name some: Swiss Federal Chancellery, Austrian Parliament, World Intellectual Property Organisation.

8.2.3. Phase One Business Model Canvas

A business model can well be described through nine major building blocks; they cover the four main areas of a business: customers, offer, infrastructure, financial viability. This is summarised in the already mentioned business model canvas:

<table>
<tr><td rowspan="2">Key Partners

Language experts</td><td>Key Activities

Server maintenance
Effort estimations
Data processing
Tools customisation
Marketing & sales</td><td rowspan="2">Value Proposition

Protect your investment by getting notified about flaws in your terminologies and resolve them efficiently in a semi-automatic, collaborative way.</td><td>Customer Relationship Management

Dedicated direct
Training
Feedback to development</td><td rowspan="2">Customer Segment

Institutions
Corporations, pre-/post-merger
Corporations with significant intra-departmental communication needs</td></tr>
<tr><td>Key Resources

LISE / ESTeam server infrastructure
Technical staff
Linguists</td><td>Channels

EU initiatives
Cross-border initiatives</td></tr>
<tr><td colspan="3">Cost Structure

Data conversions and processing
Server infrastructure
Marketing & Sales
Terminology work</td><td colspan="2">Revenue Stream

Service subscription
Data processing (value-oriented)
Usage of expert staff (usage-oriented)</td></tr>
</table>

8.2.4. Revenue Stream

As the business model canvas outlines, there are three items that make up the concrete revenue stream: a) subscription to the service, b) the amount of data processing and b) actual usage of expert staff.

- **Service subscription**: Continued software updates and data improvements required by the customer will be provided through an annual subscription.
- **Data processing**: The model for pricing the data processing services is based on a combination of person days and success rate. The software processing as well as the user interaction is logged and the customer only charged for what is actually useful. This guarantees an extremely transparent, yet value-driven price model. The revenue volume here depends on the size of the resources, the type of the software service selected and the actual software performance. Customers are only charged for a useful result.
- **Usage of expert staff**: The customization, training and terminology expert support is charged on the basis of person days. The ESTeam and LISE Partners Expert Services comprise:
    - Terminology and technology expert support: to analyse and process terminology databases.
    - Customization of ESTeam Language Servers and ESTeam Tools to adapt to customer metadata.

This is charged based on person days spent for each customer, depending on the status of their metadata. To propagate the standard, it can be said that the closer to TBX, the less work is required.

- Processing of data: The concrete technical execution of the processing.
- Customer support and software training: Support Services, to give users hands-on help as well as Training Services, to give users initial training on workflow and technologies.

Besides above, the ESTeam Tools client software and the Collaboration Portal are provided free of charge to all customers and all partners. They come without licensing nor license enforcement.

Maintaining terminology resources is not a once-in-a-lifetime activity but is an on-going continuous process[12]. A subscription based model is appropriate. Membership is obligatory and will be free during the first working phase, but for continuity and upgrades the annual membership fee will provide that service.

8.3. Phase Two – Productisation of LISE Technology in Coreon

In Phase Two, LISE technology becomes an integral part of Coreon. Coreon means the entry of a new player into the market of terminology and taxonomy management tools, but with yet unknown capabilities in data maintenance and collaboration.

The LISE features move towards an off-the-shelf offering. In the ideal case, namely with Coreon as the master system, tedious data mirroring steps from another system and back are no longer needed. Yet, data migration, analysis, preparation and consultancy activities remain a requirement. The revenue stream consists not only of a service and data processing component, but in Coreon contains also a SaaS licensing component.

8.3.1. The LISE Value Proposition in a Coreon Context

The value proposition shifts towards the overall Coreon proposition. LISE technology becomes a unique selling point.

With Coreon enterprises and institutions can build and maintain taxonomies and terminology within one single system. This facilitates the management of large and long-living resources.
Linguistic maintenance and collaboration tools keep the data clean and consistent all the time, thus the data investment is always protected**.** Coreon is delivered as a SaaS-solution. It works in browsers and tablets.

[12] IATE was recently enhanced (winter 2012/2013) with more resources.

8.3.2. Target Markets

The target markets for Phase Two do not differ significantly from Phase One. The more product-like approach however helps to address already some lower volume / smaller budget prospects.

8.3.3. Phase Two Business Model Canvas

<table>
<tr>
<td rowspan="2">Key Partners

Language experts
Business & pre-sales consultants
SaaS hosting company
ESTeam</td>
<td>Key Activities

Server operations
Coreon development
Data conversions
Marketing & Sales</td>
<td rowspan="2">Value Proposition

Manage large and long-living resources. Via linguistic maintenance tools you keep the data clean and consistent all the time.</td>
<td>Customer Relationship Management

Dedicated, direct
Self-service
Consultancy
Feedback to development</td>
<td rowspan="2">Customer Segment

Institutions
Corporations, pre-/post-merger
Corporations with significant intra-departmental communication needs</td>
</tr>
<tr>
<td>Key Resources

Coreon infrastructure
Expert staff</td>
<td>Channels

EU initiatives
Cross-border initiatives</td>
</tr>
<tr>
<td colspan="2">Cost Structure

Coreon infrastructure
Marketing & sales
Expert staff (development, conversion, data analysis)
Terminology work</td>
<td colspan="3">Revenue Stream

Coreon subscription (SaaS license model)
Data maintenance module (value-driven)
Usage of tech staff / expert service (usage driven)</td>
</tr>
</table>

Coreon is a legal entity on its own, therefore ESTeam AB is then a key partner to deliver services, technology and expertise.

8.3.4. Revenue Stream

The effort moves away from customizing the tools, but towards improving the user interface. Still, tuning the analysis algorithms remains a significant item; the then integrated modules are being optimized as well as more and more parameterized with every new customer.

- **Coreon subscription / SaaS License revenue:** Coreon is positioned as a Software-as-a-Service offering as well as available on premise. As a fully featured terminology and taxonomy management application it comes with its own pricing and revenue model. Classic SaaS-models favor a user-based pricing. Yet, to accommodate the LISE audiences, Coreon could for instance differentiate between “power users” (who edit and maintain data) from “consumption users” (large user groups with just read-only

rights, incl. third party API-driven access). The LISE collaboration portal is fully integrated and part of the Coreon subscription.
All these parameters depend from the Coreon business development, though.

- **Data maintenance capabilities:** An extension module to Coreon will host the capabilities that are currently known from the ESTeam Tools as developed in LISE. Since absolutely unique and not yet a commodity, this module sees an additional, value-based price tag.
- **Usage of expert staff:** This revenue stream remains similar to the one in Phase One. Customization, training and terminology expert support is charged on the basis of person days. The ESTeam and LISE Partners Expert Services comprise:
    - Terminology and technology expert support: to analyse and process terminology databases.
    - Improvement and tuning of data maintenance algorithms.
      This is charged based on person days spent for each customer depending on the status of their metadata. Coreon has a direct support for TBX import. This reduces the conversion effort.
    - Processing of data: concrete technical execution assistance while processing.
    - Coreon customer support and software training: Support Services, to give users hands-on help as well as Training Services, to give users initial training on workflow and technologies.

8.4. Phase Three – Towards Commoditisation

ESTeam and Coreon follow the "Lean Startup" method. Upon entering Phase Three, many experiences with early adopters will have led to a revised definition of the business model and how to execute on it. Projections today are still speculative.

8.4.1. Value Proposition

While the core of the value proposition remains the same, a much wider audience of users is targeted at:

With Coreon SMEs and large organizations can build and maintain taxonomies and terminology within one single system. This facilitates an efficient management of large and long-living resources.
Linguistic maintenance tools keep the data clean and consistent all the time, thus the data investment is always protected. Coreon is delivered as a SaaS-solution. It works in browsers and tablets.

8.4.2. Target Markets

In phases One and Two we concentrate on customers with very large data, acting in a cross-border, inter-institutional context. The business case is triggered through a political or organisational mandate.

In Phase Three we widen the view. As ESTeam experiences show, every maintainer of terminological and taxonomical resources owning more than 2,000 entries suffers more or less significantly from a quality problem. Coreon establishes itself as the one solution that not only stores and retrieves data, but that has quality modules in its core. This becomes an appealing benefit for every customer of terminology and taxonomy tools.

8.4.3. Phase Three Business Model Canvas

<table>
<tr>
<td rowspan="2">Key Partners

Language experts
Third party VARs
Business & pre-sales consultants
SaaS hosting company
ESTeam AB</td>
<td>Key Activities

Server operations
Coreon development
Marketing & Sales</td>
<td rowspan="2">Value Proposition

Manage large and long-living resources easily on-online. Via linguistic maintenance tools you keep the data clean and consistent all the time.</td>
<td>Customer Relationship Management

Dedicated direct and self-served
Consultancy
Coreon user conferences</td>
<td rowspan="2">Customer Segment

SMEs
Small and large organizations</td>
</tr>
<tr>
<td>Key Resources

Coreon infrastructure
Coreon development</td>
<td>Channels

Globalisation activities
Cross-border initiatives worldwide</td>
</tr>
<tr>
<td colspan="3">Cost Structure

Coreon infrastructure
Marketing & Sales
Expert staff</td>
<td colspan="2">Revenue Stream

Coreon subscription (SaaS-model)
Coreon extension modules (value driven)
Usage of expert staff (usage driven)</td>
</tr>
</table>

A detailed revenue model for Phase Three is developed only when we reach this phase. The Coreon product offering has then learnt from first reference customers; the LISE technology is integrated into Coreon. It needs to be seen to what degree data maintenance workflows can be standardized or still require significant manual customization.

8.5. Marketing

Along all three phases the starting pain point when it comes to messaging is the following: after years of collecting terminology data and storing huge amounts of terms, we nowadays see more and more terminology databases that suffer from degrading quality. Even when merging existing resources the quality problem is inherent, due to unavoidable data conflicts. This poses an often invisible, yet serious problem: the investment that had been made into collecting and polishing terminological entries is in danger of being lost. Resources that have been built over time with a lot of effort are losing their value.

8.5.1. Audience Characteristics

The customer acquisition focuses on EU organizations, international organizations and global corporates. The LISE partners ESTeam and CrossLang have a large network of corporate and institutional clients. The revenue channels will be highlighting the following benefits:

- **Terminology Quality Control** – to improve the quality of existing terminology
- **Terminology Translation Expansion** – to add new translations to existing terms
- **Terminology Harmonisation** – to streamline and clean up terminologies to avoid redundancy in indexing and information retrieval

LISE functionalities concretely target the terminologists; the business value is for the users of an organization that rely on the quality of the resource. LISE Partners have a good network in the terminologist / language management world. The visibility outside this network is probably the main challenge. But partners from the Austrian Parliament have a good relationship with the legal world and ESTeam is a known technology player in the trademark and patent world. EURAC and University of Vienna are best connected in the terminology world. CrossLang is best connected in the language and content management world. Connections to the finance or healthcare businesses are being established[13].

The LISE and then the Coreon marketing therefore follows two strategies: 1) horizontal dissemination and visibility through language, terminology, and knowledge management conferences and publications; 2) establishing new connections to the identified key vertical markets through partners and channels, for instance by approaching language service providers that are focusing on the finance market.

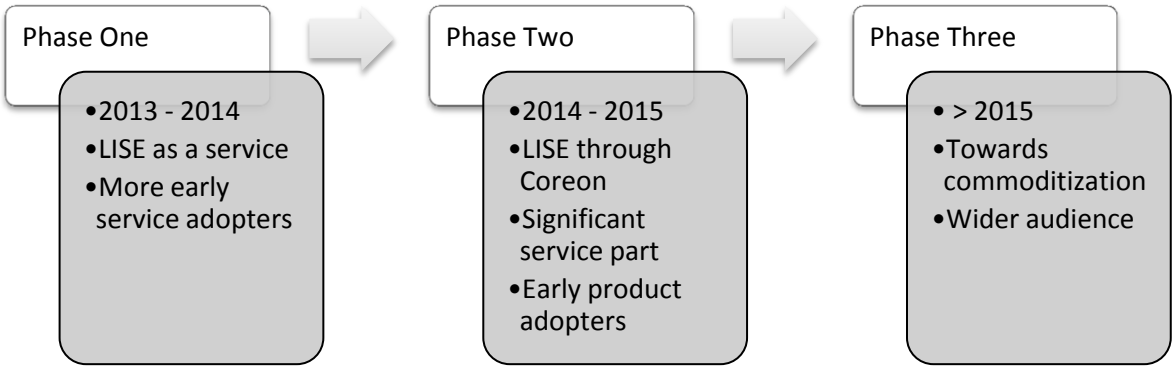
8.5.2. Marketing Plan

Phase One focuses heavily on evangelising the market, on shaping the value proposition even more, and on fine tuning the messaging. The goal is to discover 2-5 early adopters. This phase started in early 2013 and is planned to run into 2014.

In 2014 the marketing will shift to a more product-oriented messaging. The benefit of a "unified system" (LISE as part of Coreon) makes the solution appealing to more customers, the early product adopters.

By 2015, Coreon (including its LISE part) will be a mature and feature rich product, so that it can be rolled out to a very wide audience, namely as an off-the-shelf offering.

[13] At the SEMIC2013 conference late June in Dublin, ESTeam managed to connect with representatives from SSI, Denmark. Corrently EsTeam is establishing follow up contacts with IHTSDO.
Contacts to the European Central Bank have been revitalized during ETUG 2013 conference in Berlin.



8.5.3. Supporting Marketing and Sales Activities

Sales-supporting and marketing activities have to be seen under the context of the Coreon roadmap and its marketing plan. As Coreon evolves, we see more and more supporting materials and activities: growing and maturing web-site content, continuous blog posts, analyst quotes, case studies, demo and presentation material etc. Concretizing this plan in detail is beyond the scope of this document. Yet, there are several noteworthy items that guarantee LISE's sustainability in the context of Coreon marketing.

The following activities help to evangelise the market during Phase One:

- Developing a Conceptual White Paper "Valuing Quality in Terminology Resources to Protect Investment". This outlines the concrete reasoning why to value and how to measure quality in terminology resources. It outlines how to address the problems. It takes a lot of WP3 research as a foundation and maps it then to concrete technology approaches.
- Developing a Business White Paper "How language precision is key to effective and efficient cross-border legal / healthcare / finance processes". This sets the value proposition into a political and economic context.
- Events to spread the message:
    - i-know 2013 (September 2013, Graz): Knowledge management conference together with i-semantics. This event is outside of the pure terminology world; it rather focuses on taxonomy and ontology resources, semantics and semantic web topics. Yet, this is the opportunity to make right this audience aware of methods and reasons how to improve large resources.
    - META Forum 2013 (September 2013, Berlin): This is one of the most inspiring language technology networking events in autumn 2013.

- CHAT2013 (November 2013 at tekom/tcworld conference, Wiesbaden): *Creation, Harmonization and Application of Terminology Resources* – LISE partners will give a pure LISE presentation in this workshop. Above mentioned conceptual white paper targets particularly this audience. ESTeam exhibits in parallel at this conference, as part of a joined LT-Innovate booth.
- SemTechBiz 2014 (February 2014, Berlin): *Semantic Technology and Business* – The Business White Paper helps particularly in this audience.

Entering Phase Two, LISE/Coreon marketing then emphasizes the product aspect:

- First version of a "Teaser" in Coreon, to quickly prove the methodological value – as described under METHODOLOGICAL EXPLOITATION. While this is rather a development task, it significantly supports marketing and pre-sales.
- Demo material: The completely re-worked version of the original ESTeam Tools demos as available today on youtube.com, but now in Coreon (like earlier in the document the screenshots under PHASE TWO: APPLYING LISE THROUGH COREON).
- Coreon benefits: This will, amongst other aspects, particularly highlight the value that LISE delivers. LISE technology is an absolutely unique selling point for Coreon compared to all other terminology and taxonomy management tools.
- Case studies: The early adopters' success from Phase One are referenced in case studies and are presented on conferences and trade shows. This raises the awareness for the 'why' and the 'how'. The presentation of IATE users at the LSP conference in Vienna in July demonstrated how convincingly this works.
- Events in the planning for this phase:
  - DTT 2014 (April 2014, Mannheim): *Deutscher Terminologie-Tag* – The meeting of the worldwide largest terminologists association. This will particularly target the market of terminology management systems. We message the pragmatic, semi-automatic approach to regain control over large terminology databases.
  - TKE 2014 (June/July 2014, Berlin): *Terminology and Knowledge Engineering* – While this is a slightly more academic conference, the goal is to repeat the methodological and best practices results from LISE. The event falls together with the ISO TC37 annual meeting, hence all leading experts will be there.

In Phase Three the focus is on increasing the market and audience.

- Social media activities: Very active messaging on twitter, business platforms etc. takes place in Phase Three. During the previous phases the customers were found through other channels, such as direct contact or conferences. It is worth to note that from the WP3 interviewees almost no one is active on any social media platform; this means, during the early phases social media do not represent a very efficient channel. In Phase Three we intend to make this thrive.

- Case studies: Product and Service case studies now substantially support messaging the value proposition. A rather vertical oriented messaging can evolve from this.

8.6. Market Vision

The LT-Innovate report "Status and Potential of the European Language Technology Markets"[14] estimates a growth of the Language Technology Software and Services Market from €19.3B in 2011 to nearly €30B in 2015. Around one quarter of this is for software sales, the rest for services. Terminology and taxonomy applications are indispensable core technologies for two of the three segments that the LT-Innovate report identified: *Translation* as well as *Intelligent Content*. In other words, terminology and taxonomy resources are everywhere and continue to grow. Hence, the need for maintaining and improving such resources will grow at least with the same intensity. In these markets alone the demand will grow.

With the rise of big data and text analytics the need for high quality multilingual resources will grow even more. Multilingual taxonomies or thesauri are the backbone, i.e. the algorithm behind smart, deep linguistic analysis of large textual data. In other words, without a properly maintained language resource, the quality of enterprise search and data analytics drops. Companies such as Smartlogic, Semantic Web Company or Top Quadrant see this business potential already and have, by consequence, developed taxonomy editors. But all of these tools lack sophisticated quality assurance algorithms; their focus is on managing the taxonomy, but not on mastering a potentially large amount of terms in there. Yet, equipping such taxonomies with large amount of terms and making them multilingual would work like an amplifier. Maintainers of these resources will run into similar problems as now understood for "classic" terminology databases.

International classification and cross-border interoperability activities require merging and harmonisation of resources. In such multilingual and multicultural contexts pure human data maintenance approaches will fail. As proven with OHIM and now through LISE for IATE, semi-automatically improving the quality is the only method towards an efficient and effective result. Initiatives such as ESCO[15] (European Skills, Competencies and Occupations taxonomy) or the IES (Institute for Environment and Sustainability) harmonisation research prove that the demand is concretely there: "Consistency of language is vitally important to semantic interoperability … For data access and to facilitate understanding it is useful to develop cross-language information retrieval strategies" [16].

[14] See http://www.lt-innovate.eu/resources/document/press-release-lt-innovate-publishes-landmark-report-state-european-language-techn - pages 15-16.
[15] ESCO is part of EURES, the European Job Portal. See http://ec.europa.eu/social/BlobServlet?docId=9171&langId=en
[16] See JRC Reference Report: A Conceptual Model for Developing Interoperability Specifications in Spatial Data Infrastructures (JRC Publication #69484), pages 29-30.

In a much wider context even, as highlighted by the LISE Advisory Board, the whole idea behind LISE can be mirrored to initiatives around the topic of “data and information quality” and/or of “cleansing” of textual data. The quality demands that apply to terminology resources are transferrable to any kind of collections of textual data. For that purpose it is for instance necessary to establish contacts to organisations such as the Deutsche Gesellschaft für Informations- und Datenqualität e. V.[17] So far, database cleaning focused mainly on the numeric aspects. Due to a lack of algorithms it was almost impossible to improve human, textual data. The technology in LISE delivers this now.

All this goes beyond LISE, but should be seen as a future path where the core technology has a very positive, even wider impact.

[17] See http://88.198.68.171:8080/confluence/display/homepage/Home

9. LISE Exploitation Conclusion and Outlook

The ESTeam Tool Suite and surrounding services have to date provided revenue in the seven-figure region in Euro. The tools and services provided in combination with the tools have already proved extremely profitable. The value of the result for the customers has also proven to be very high. The quality improvements and usefulness of the terminology after being treated by the ESTeam tools has provided the basis for making it possible for the IPR domain to harmonise across EU borders.

The technology is now proven to be applied and has been mapped to use cases outside of the IPR domain. It addresses concrete user needs, “a terminologist’s dream” according to IATE users. ESTeam Tools and LISE services are a sustained service offering that is already attracting new customers outside the domain it was originally designed for.

Through the LISE market analysis, ESTeam identified a gap in product offerings in the market landscape of terminology/taxonomy management systems and created a new SME, the Coreon GmbH. The technology deployed in LISE will be part of Coreon and become one of its unique selling points.