



Project no. 027391

Project acronym: LT4eL

Project title: Language Technology for eLearning

Instrument Specific Targeted Research Project

Thematic Priority Information Society Technology

### **D6.2c Exploitation plan and report**

Due date of deliverable: 31-05-2008

Actual submission date: 08-07-2008

Start date of project: 1-12-2005

Duration: 30 Months

Organisation name of lead contractor for this deliverable: Utrecht University (UU)

Revision [1]

<b>Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)</b>		
<b>Dissemination Level</b>		
<b>PU</b>	Public	x
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	

# D6.2c - Exploitation plan and report

## Contents

- 1 Executive Summary
- 2 Introduction
  - 2.1 Exploitable results
  - 2.2 Potential users
  - 2.3 Main exploitation activities
  - 2.4 IPR issues
  - 2.5 Division of tasks
- 3 Exploitable results
  - 3.1 Know-how
  - 3.2 Corpus of learning objects
    - 3.2.1 Description
    - 3.2.2 Use and exploitation
  - 3.3 Tools for metadata generation: keyword extractor
    - 3.3.1 Description
    - 3.3.2 Use and exploitation
  - 3.4 Tools for glossary creation: glossary candidate detector
    - 3.4.1 Description
    - 3.4.2 Use and exploitation
  - 3.5 Domain specific ontology and language specific vocabularies
    - 3.5.1 Description
    - 3.5.2 Use and exploitation
  - 3.6 ILIAS integrated functionalities
    - 3.6.1 Description
    - 3.6.2 Exploitation and Use
  - 3.7 Validation methodology
    - 3.7.1 Description and exploitation
- 4 Exploitation report and plan
  - 4.1 Learning Management System development community
  - 4.2 Technology Enhanced Learning (TEL) community
  - 4.3 eLearning community in the Mediterranean countries
  - 4.4 Semantic Web (SW) community
  - 4.5 Language Technology community
  - 4.6 Commercial organizations
  - 4.7 Higher education institutions
  - 4.8 The user panel
  - 4.9 PhD projects and Master theses
- 5 Overview table

## 1 Executive Summary

According to the annex, the aim of work package 6 is to present and disseminate results of the project, to foster the transfer of knowledge and a broad take-up of programme results. The following groups have been the special targets:

- end users at universities: learners and teachers
- scientific community
- national educational authorities

The exploitation plan had to be delivered in M12 (D6.2), M24 (D6.2b) and M30 (D6.2c). This report is the M30 exploitation plan (D6.2c). According to the revised list of deliverables, the M30 exploitation plan and

report has to contain:

- an exploitation plan and updates;
- a usability study of the project developments in TEL field;
- future planning activities.

The goal of the report is to provide a plan and a report of the way the resources developed are and can be exploited.

The purpose of deliverable D6.2c is to present the exploitation plan of the LT4eL project at the end of the project (M30). More specifically, D6.2c describes the consortium's strategy for exploiting project outputs during and after the last six months of the project. The purpose of this document is to describe how the results of the projects will be and have been used or exploited by the LT4eL partners and by other potential users after the project. The document is organized as follows:

- The first section identifies the exploitable results, the potential users, and the main exploitation activities.
- The second section contains a description **per exploitable result** on how it can be exploited or used in further research.
- The third section contains a more detailed description of the exploitation channels we use for the exploitation of the results described in the second part.

## 2 Introduction

This introductory section identifies and briefly describes the exploitable results, the potential users, the main exploitation activities and IPR protection measures.

### 2.1 Exploitable results

- Know-how in Language Technology (LT), Natural Language Processing (NLP), Semantic Web and eLearning technologies;
- Corpus of learning material of at least 1000 pages for eight of the languages represented in the consortium, except for Maltese (WP1)
  - corpus of learning material for eight languages annotated with part of speech tagging;
  - corpus of learning material for eight languages annotated with definitions;
  - corpus of learning material for eight languages annotated with keywords;
  - corpus of learning material for eight languages annotated with concepts;
- Tools and methodology for semi-automatic metadata generation (WP2)
  - key word extractor developed for eight languages (Bulgarian, Czech, Dutch, English, German, Polish, Portuguese and Romanian);
- Tools and methodology for the creation of glossaries related to learning objects (WP2)
  - glossary candidate detector able to extract definitions in eight languages (Bulgarian, Czech, Dutch, English, German, Polish, Portuguese and Romanian);
- Tools and methodology for semantic search and cross-lingual search (WP3):
  - domain specific ontology of more than 1000 concepts in the domain of computing;
  - language specific domain vocabularies for nine languages - Bulgarian, Czech, Dutch, English, German, Maltese, Polish, Portuguese and Romanian;
  - annotation grammars;
  - ontology-based search engine);
- ILIAS with integrated functionalities (keyword extractor, glossary candidate detector and cross-lingual retrieval via ontology) (WP4);
- Validation methodology (WP5).

### 2.2 Potential users

- LT4eL partners who will use the collected knowledge and resources for their research activities and the developed functionality integrated in the ILIAS system for research and their didactic activities;
- Scientific community (LT, NLP, Semantic Web and eLearning), including other ICT/TEL projects, that will use the collected resources, the developed methodology for their research activities.

- LMS developers that will use the developed concepts and methodology to integrate new functionalities within their systems.
- LMS end-users who will use the new functionalities integrated in the ILIAS or other systems.
- Local authorities (e.g. ministries of education), and higher education institutions (e.g. universities) who will become aware of the potential of eLearning.

## 2.3 Main exploitation activities

The project results have been marketed in different communities and in different ways.

Communities:

- eLearning community
- LT community
- NLP community
- SW community
- Commercial institutions
- Higher education institutions
- EC-TEL community
- User panel
- PhD and MA projects

A more detailed description of these communities and the ways in which we use them, can be found in section 4 of this deliverable.

Dissemination of the results has been and will be ensured in different ways also beyond the end of the project:

- through the multilingual Web portal ([www.lt4el.eu](http://www.lt4el.eu));
- by participation at scientific events, notably eLearning, LT, NLP and Semantic Web conferences and workshops;
- by means of scientific publications;
- by means of reports;
- by means of mailing lists;
- via informative material (leaflets and videos);
- via the user panel.

In order to promote the use of the LT4eL functionalities, and in general of open source Learning Management Systems (LMS), we have established contacts with local educational authorities (ministries of educations, universities and schools) in particular in the new member states, for example via the organization of special meetings or awareness events on a national level.

The technical results of the project have been made available in well-documented releases. The resources and tools developed are freely available via the web portal as well as through the SourceForge.net portal for open source projects at <https://sourceforge.net/projects/lt4el/>. The new functionalities as well as the enhanced LMS will be freely available under the GNU General Public Licence (GPL) and distributed also through the ILIAS channels, as well as through the “Language GRID” initiative in which the University of Tübingen participates and the Common Language Resources and Technology Infrastructure (CLARIN) project (<http://www.clarin.eu/>) of which Utrecht University is coordinator.

For a more detailed description of our (planned) dissemination activities see deliverable D6.1c.

## 2.4 IPR issues

All the IPRs (Intellectual Property Rights) concerning publications and other know-how resulting from the project have been handled in agreement with all partners. The IPR status of the learning objects collected, that is our corpus of learning material has been investigated in detail. We have assessed that material used for validation is in general IPR free, but this is not the case for the whole corpus. However, we have found more sources of archives for open content and for validation purposes we have added new material to our existing corpus given that the annotation process is rather straightforward at this point. It is the task

of the WP-managers to ensure that all issues related to the intellectual property are respected. Basically our project has dealt only with freely available resources and tools. The project deliverables are publicly available and can be downloaded from the website of the project (<http://www.lt4el.eu>).

## 2.5 Division of tasks

Exploitation activities are part of WP6 with Utrecht University as WP leader. All partners however are committed to the widespread dissemination and active exploitation of the results of the project.

# 3 Exploitable results

This section contains a list of all exploitable results delivered by the LT4eL project. For each of the results, a description of the result and a plan for its use or exploitation are included.

## 3.1 Know-how

Through the LT4eL project, each partner acquires knowledge and skills in state-of-the-art technologies in LT, NLP, Semantic Web and eLearning. It is only natural for all partners to exploit this know-how for their future research and for the acquisition and execution of future projects. This has been the case for the UU, UTU and IPP-BAS partners which will bring the know-how acquired in this project in a new EU project **Language Technology for LifeLong Learning**.

## 3.2 Corpus of learning objects

### 3.2.1 Description

This corpus consists of more than the 200000 words (1000 pages) of learning objects we are supposed to deliver for each language represented in the consortium. The broad domain of this collection is computing. More specific subdomains include teaching academic skills, creating webpages, basic computer skills. All material has been linguistically annotated at least up to the level of part-of-speech tagging and morphological analysis. Besides, it has been marked up with (i) key words, (ii) definitory contexts. The complete annotated corpus has been delivered at month 12 (December 2006). During the second year of the project, an extra annotation layer was added to the objects with ontological annotation of concepts. In the last phase of the project, several learning objects have been added and have been annotated with the various layers of linguistic information.

### 3.2.2 Use and exploitation

The relevance of this corpus lies essentially in its use for the development of the new functionalities within the project. However, a subcorpus has been created with IPR free documents in order to make it publicly available in the ILIAS library, but also on the project website. As potential users we identify ILIAS end-users (teachers and students) interested in the content of the documents, but also the whole scientific community, notably other eLearning projects like iCamp (<http://www.icamp-project.org/>), as well as NLP and SW projects. The latter can use the multilingual corpus - which, if necessary, could be easily annotated with other levels of annotation - as a data base for further experimental research, for example, in language technology and natural language processing. The corpus is particularly useful because it encodes eight languages and it has been annotated not only with morphological information but also with definitions, keywords and concepts. The **Babylon and Ontology** European project (<http://www.bonynetwork.eu/>) might also be interested in the corpus with annotated concepts. In addition, some students of Qusay H. Mahmoud (University of Guelph - Canada) are using our English objects annotated with concepts to investigate the possibility of learning object reusability. Knowledge about the availability of this corpus will be disseminated via the usual channels (project webportal, mailing lists such as the corpora list).

## Corpus of Learning Objects

Language	No of words	Can be used					Cannot be used		Ca u
		Author	Permission	Gov	GPL	free for research	copyright	unknown	
<b>Bulgarian</b>	<b>207.239</b>			76.793		118.223		12.223	<b>19</b>
<b>Czech</b>	<b>1.316.988</b>	332.803	707.990	104.571	162.859			8.765	<b>1.31</b>
<b>Dutch</b>	<b>442.559</b>	127.963		42.867	25.860	115.408		130.461	<b>31</b>
<b>English</b>	<b>1.162.564</b>	219.133		2.336	33.587	162.758	497.343	247.407	<b>41</b>
<b>German</b>	<b>355.043</b>	1.600	60.200			72.194	50.165	170.884	<b>13</b>
<b>Maltese</b>	<b>52.668</b>	16.226				25.491		10.951	<b>41</b>
<b>Polish</b>	<b>656.800</b>	30.660			83.425	189.855	232.545	120.315	<b>30</b>
<b>Portuguese</b>	<b>548.998</b>	315.359			11.400	57.474		164.765	<b>38</b>
<b>Romanian</b>	<b>644.292</b>	439.631		43.889		147.500		13.272	<b>63</b>

### 3.3 Tools for metadata generation: keyword extractor

#### 3.3.1 Description

One of the aims of the project is to improve the retrieval and accessibility of content through the identification of the learning material by means of descriptive metadata. To this end, we have employed available language technology tools and resources to develop a keyword extractor that facilitates the semi-automatic generation of metadata. The keyword extractor, is a tool that supports authors and content managers in selecting, in the chosen learning objects, the keywords that best represent the topic(s) of these learning objects. The tool analyses a set of annotated documents and returns the best keyword candidates for each learning object. The user of this functionality decides on the inclusion of these candidates into the metadata. The tool has been tested on eight languages (i.e. Bulgarian, Czech, Dutch, English, German, Polish, Portuguese, Romanian) and it can be used stand-alone or within the ILIAS Learning Management System.

We have assumed that the addition of this functionality will improve the retrieval of the learning objects in terms of their content. The validation results available with this deliverable confirm this assumption. The tool has been fully documented. The documentation is essential not only for a proper integration within the ILIAS system, but also for the integration within other LMSs that will be interested in adopting it. The final version of the tool is delivered at month 30 (May 2008). Documentation can be found on the project website (<http://www.lt4el.eu>), the lasi portal ([http://consilr.info.uaic.ro/uploads\\_lt4el/](http://consilr.info.uaic.ro/uploads_lt4el/)) and sourceforge.net open source community (<http://sourceforge.net/projects/lt4el/>).

#### 3.3.2 Use and exploitation

The keyword extractor has been integrated into the Learning Management System ILIAS. To favor further exploitation, it has been offered independently as web service so that it can be used by other applications as well. Full integration into other LMSs is thus also possible. To foster implementation into other systems, a document describing the integration process into ILIAS (a step-by-step integration manual) has been provided targeted at developers of other LMSs. Beyond direct exploitation in the eLearning industry, the development of the keyword extractor has had an impact on the scientific community. The functionality uses state-of-the-art LT and NLP technologies and therefore can provide useful feedbacks to these communities and be the starting point for utilization of these technologies into other domains. Several papers describing the keyword extractor have been accepted at international conferences (cf. deliverables 6.1b and 6.1c)

## **3.4 Tools for glossary creation: glossary candidate detector**

### **3.4.1 Description**

Dictionaries constitute an important support to the learning process. One of the assumptions of the LT4eL project is that they can be especially useful if they are constructed on the basis of the definitions extracted from a given learning object. To this end, we have employed available language technology tools and resources to develop a glossary candidate detector. The glossary candidate detector is a tool that supports authors and content managers in creating glossaries by identifying definitory contexts in a text that contains the term to be defined and its definition. The input of this tool are annotated texts (and background resources). The output of this tool is a set of candidates for definitions that are extracted by means of pattern based grammars. For certain languages (i.e. Dutch, English, Polish and Portuguese) we have experimented with filtering wrong definitions by using Machine Learning techniques. The user decides on the inclusion of the candidate definitions into the dictionary which is then attached to the learning object. The tool has been tested on eight languages (i.e. Bulgarian, Czech, Dutch, English, German, Polish, Portuguese, Romanian) and it can be used stand-alone or within the ILIAS Learning Management System.

The validation results available with this deliverable show that the possibility to search for definitions based on the available glossaries has been highly appreciated by learners. The tool has been fully documented. The documentation is essential not only for a proper integration within the ILIAS system, but also for the integration within other LMSs that will be interested in adopting it. The final version of the tool is delivered at month 30 (May 2008). Documentation can be found on the project website (<http://www.lt4el.eu>), the Iasi portal ([http://consilr.info.uaic.ro/uploads\\_lt4el/](http://consilr.info.uaic.ro/uploads_lt4el/)) and sourceforge.net open source community (<http://sourceforge.net/projects/lt4el/>).

### **3.4.2 Use and exploitation**

The glossary candidate detector has been integrated into the Learning Management System ILIAS. To favor further exploitation, it has been offered independently as web service so that it can be used on the internet by other applications as well. Full integration into other LMSs is thus also possible. To foster implementation into other systems, a document describing the integration process into ILIAS (a step-by-step integration manual) has been provided targeted at developers of other LMSs. Beyond direct exploitation in the eLearning industry, the development of the glossary candidate detector has had an impact on the scientific community. The functionality uses state-of-the-art LT and NLP technologies and therefore can provide useful feedbacks to these communities and be the starting point for utilization of these technologies into other domains. Several papers describing the glossary candidate detector for the various languages have been presented at international conferences (cf. deliverables 61.b and 6.1c)

## **3.5 Domain specific ontology and language specific vocabularies**

### **3.5.1 Description**

The domain specific ontology (i.e. computing) is developed in a language independent way and comprises more than 1000 concepts. An English vocabulary is mapped to the concepts and the relations within the ontology. In addition, language specific domain vocabularies have been developed and linked to the ontology for all the nine languages of the consortium. The domain specific ontology was delivered at month 12 (December 2006), and the language specific vocabularies at month 18 (May 2007). In month 24 we delivered a new version of the ontology which contains additional 200 concepts and related lexica. In month 30 (May 2008), we deliver the final version of the ontology. At the moment, the ontology contains 1002 domain concepts, 169 concepts from OntoWordNet and 105 concepts from DOLCE Ultralite. It also contains more than 100 object properties.

The ontology is being used to structure, query and navigate through the learning objects that are part of a Learning Management System. The ontology can play two major roles:

- Classification of learning objects. Each learning object is connected to a set of concepts in the

ontology. This classification allows ontological search, i.e. search based on concepts and their interrelations within the ontology.

- Multilingual search for learning objects. In this case the ontology plays the role of Interlingua between the different languages. Thus the user might specify the query in one language and get learning objects in other language(s).

The innovative aspects of this part of the project consist firstly in the application of semantic web technologies (ontologies) to facilitate learning processes and, secondly, in their use (linked to language specific vocabularies) to address problems of multilingual nature, in particular multilingual search.

The ontology in combination with the lexica and the learning objects have been used for ontology browsing and semantic search. The software developed to this end has been fully documented. The documentation is essential not only for a proper integration within the ILIAS system, but also for the integration within other LMSs that will be interested in adopting it. The final version of the ontology browsing and semantic search is delivered at month 30 (May 2008). Documentation can be found on the project website (<http://www.lt4el.eu>), the lasi portal ([http://consilr.info.uaic.ro/uploads\\_lt4el/](http://consilr.info.uaic.ro/uploads_lt4el/)) and sourceforge.net open source community (<http://sourceforge.net/projects/lt4el/>).

### **3.5.2 Use and exploitation**

From the perspective of our project, the relevance of these semantic web technologies lies essentially in their use within an LMS to improve the organization and the retrieval of the learning objects across languages. Potential users in this respect will be LMS (notably ILIAS) end-users, but mostly LMS developers, open source as well as commercial. Documentation of the ontology development and a report for its integration within an LMS has been provided to facilitate its utilization in any eLearning platform. The results obtained have also had a clear scientific impact. The use of ontologies for multilingual retrieval is in its infancy. We have contributed to the development of these techniques and to provide important feedbacks to the semantic web community on the potential of their technology in this field. The ontology and the lexica developed in the LT4eL project have been distributed to:

- OU-UK and the KMI group (Enrico Motta and Mathieu d'Aquin) for reuse through the Watson ontology search engine which is being maintained by OU-UK.
- DFKI (Paul Buitelaar) for reuse through Ontoselect, which is an ontology library being maintained by DFKI;
- CNR Rome - Laboratory of Applied Ontology (Aldo Gangemi, Alfio Gliozzo) to be reused in their project eu project Babylon and Ontology (<http://www.bonynetwork.eu/>);
- APOSDLE (<http://www.aposdle.tugraz.at>) (Stefanie Lindstaedt and Luciano Serafini) for reuse within their project.

In this way, we ensure exploitation of the ontology beyond the project and reuse by the Semantic Web community. To further promote the use of semantic knowledge in eLearning we organized a workshop on this topic at the EUROLAN 2007 which has been held in Romania in the summer 2007, and we have been and will be present at various Semantic web and eLearning events. The developed ontologies, and in particular the methodology to link these to language specific vocabularies, will be further disseminated via the usual channels to be used by the scientific and semantic web community for further research. We have already written several papers on this topic, which have been presented at conferences (cf. Deliverable 6.1b and 6.1c). We believe that the use of ontologies in eLearning is one of the most innovative aspect of our project with much potential as it appears from our validation results as well as the peer review comments that we receive for submitted papers on the subject. It was particularly this aspect that was found innovative and a motivation for the best scientific paper award we have received at EC-TEL 2007.

## **3.6 ILIAS integrated functionalities**

### **3.6.1 Description**

The tools described above (keyword extractor, glossary candidate detector and ontologies for multilingual retrieval) have been integrated within the ILIAS system. ILIAS is a web-based learning management system and allows users to create, edit and publish learning and teaching material in an integrated system

with a normal web browser. Tools for cooperative working and communication are included as well. ILIAS is available as open source software under the GNU General Public License. The software development worldwide is coordinated by the team at the ETHZ.

The last release (3.5.0) of ILIAS before the project already offered content authors the possibility of annotating learning objects with metadata based on the LOM standard. Also a metadata based search is available. However, the metadata had to be provided manually by the author. This made the annotation process very time consuming and thus only a few authors provide useful metadata. Tools for semi-automatic metadata generation could help to solve this problem. Furthermore, ILIAS does not provide semantic web based functionalities, but it already offers the possibility of reusing learning objects like media objects or glossary items in the creation process of learning material. Ontology based retrieval of learning objects will considerably improve the task of reusing learning objects since ontologies will allow for intelligent searching and navigation in huge amounts of data. Metadata annotation and ontology driven search and navigation allow for individual content assemblance for learners. Learners will be able to build individual learning paths by entering key terms of concepts they need to learn. The validation results showed that ILIAS enhanced with the functionalities developed in the project provides a suitable environment for personalization of the learning path.

### **3.6.2 Exploitation and Use**

The main users of these integrated functionalities are ILIAS end-users (teachers and students). The scenarios produced in the context of our validation have also been developed having these users in mind, cf. deliverable 5.1c for further details.

To promote the use of the newly developed functionalities, the ILIAS English user manual has been extended with detailed descriptions of these functionalities. The user documentation has been distributed online via the ILIAS website as well as via the project website. To distribute ILIAS with the newly integrated functionalities, the project is also maintained at two open source communities, namely SourceForge.net and Ohloh.net. Clearly the functionalities could be integrated in other systems and applications since they have been implemented as web services. The package for installation of ILIAS enhanced with LT4eL functionalities can be also downloaded through the project web page.

## **3.7 Validation methodology**

### **3.7.1 Description and exploitation**

A suitable validation methodology has been developed and applied to the validation of the new functionalities as well as to their integrated set into ILIAS. eLearning applications are very much an emerging field, and there are no standard, general methodologies that can be used to validate effectiveness of the learning process in our specific context. We expect the methodology developed within the LT4eL project to be at least a first step towards this missing standard. To promote dissemination, and thus exploitation, papers on the subject will be written and presented at conference even beyond the end of the project.

## **4 Exploitation report and plan**

As can be seen from the previous section, the LT4eL project has delivered and will deliver different types of results, ranging from tools to scenario descriptions. For each of the individual results, some ways in which they can be exploited in future are already shown in the previous section. The first section mentioned the exploitation channels that are used for the exploitation of the project results (e.g. research institutions and commercial institutions). In this section we provide a more detailed description of the channels via which the project results are exploited and the communities we address. For the communities, we provide a description of the way in which we foresee the exploitation of our results within this community and the activities we have undertaken already.

### **4.1 Learning Management System development community**

The functionalities developed within the project have been integrated in the ILIAS Learning Management System (LMS). We will however market our approach with other open source or commercial systems as well, such as Moodle, OLAT, ML3, ATutor, Sakai, Claroline, Blackboard, WebCT. To create interest of potential users and prospects, as well as direct contact with other LMS developers, the consortium participated at various eLearning events or other LMS-related exhibitions and demonstrated the developed tools and methodology (cf. deliverable 6.1b and 6.1c).

To further favor exploitation, the developed functionalities are also offered independently as web services, so that they can be used by other applications as well. Besides, the project decided to make the source code available under an open source licence and to host it on the SourceForge.net portal (<https://sourceforge.net/projects/lt4el/>) and the Ohloh.net portal (<http://www.ohloh.net/projects/9785?p=LT4eL>), as well as the project website. We have also installed the final version of ILIAS with the integrated functionalities on a server of the University of Utrecht. This includes also a copy of the portal of the developed resources maintained at the University of lasi to guarantee access via the web portal to potential users of the functionalities that would like to try them out as well as partners who would like to carry on research related to LT4eL beyond the end of the project. In this way, it will also be possible to access the web services. The aim of this initiative is to guarantee access to the functionalities developed beyond the end of the projec. In this way, the results are immediately available to the general public.

ZHAW and the ILIAS team act as link with the LMS development community.

The ILIAS team is part of a German network of eLearning developers called Campussource (<http://www.campussource.de/org/>), they have been using this channel for their dissemination. However, we will not be able to rely on it in the future because of missing project funding, Campussource is in a kind of re-structuring phase.

The E-learning Center ZHAW maintains close contact with the Competence Service Production center of St Gallen, Chur and Rapperswil. Those contacts provide a basis for cooperation and quality management in the field of e-learning in the German-speaking part of Switzerland. Presentation of relevant parts of the LT4eL project and the evaluation of certain scenarios through cooperating experts of the e-learning center ZHAW themselves served the project with the expertise and quality management of the Center being used.

Several activities have been carried out to encourage exploitation of the tools in this community:

- our first official release (M24), was featured by ETHZ in the LMS communities (forums) and by sending e-mails to leading LMS developers (such as Moodle, Olat, WebCT).
- The new functionalities were explained and advertised in the ILIAS newsletter.
- At the Langtech 2008 conference (<http://www.lang-tech.org/>), dissemination material on the project has been distributed. Besides, a demo has been given of ILIAS with the new functionalities developed within LT4eL.
- We cooperated more closely with the SHARE project ([share.uni-koeln.de](http://share.uni-koeln.de)). The SHARE project aims to promote several freely available tools that teachers can use for producing, sharing and re-using learning content and metadata issues. On the website of this project, a description will be added of our project and the tools we developed.

Within the LT4eL project, language technology based functionalities have been integrated into ILIAS and validated. A crucial issue is whether these functionalities will be offered as part of ILIAS in the immediate future. In order to accomplish this, we need to ensure a certain maturity of the functionalities that are included within the core system. Within the LT4eL project, it has been possible to create a prototype but integrating the tools within a stable ILIAS release would need improvements in various aspects:

- The performance is still a major problem. We experienced this already in our validation trails, and ILIAS must run with up to 100 (or 200 on multi-server machines) users online at the same time.
- The Linguistic Processing Chain (LPC) is not yet in a real stable phase (it is not in-depth tested), and at the moment it is only available for a subset of the languages in the project, that is Czech, Dutch, English, Polish, Portuguese and Romanian. It will cost some months of development to get it into a usable (for productive environments) state. Unfortunately, during the project we did not have enough time to put too much priority on the LPC.

- The whole architecture of the project is quite complex. This is mainly due to the LPC which needs a higher number of sub-tools to work. Thus a simplification would be required.
- An obvious bottleneck is constituted by the ontology. The effort to make the semantic search work for other domains is too high at the moment. However, as more ontologies become available, this problem might be overcome in the near future.
- If the functionalities are made available through web services for other systems (to prevent the need to individually setup the whole system) to access them, the performance requirements are even higher.
- The sustainability of the maintenance of the tools is also a crucial issue.

We conclude that the LT4eL project proved that the developed functionalities can be useful for eLearning. To make them mature, will take another one to two years development and additional funding.

## 4.2 Technology Enhanced Learning (TEL) community

Our project results have been presented at several conferences, including eLearning conferences. The work done within our project is appreciated and considered to be innovative by the TEL community. The fact that our paper on the EC-TEL 2007 conference won the best paper award, is already a strong indication that the work done within the LT4eL project is considered to be valuable by the TEL community. However, we have taken pains to favour exploitation of our results in the TEL community.

The results of the LT4eL project will be employed in the new EU STREP project 'Language Technology for LifeLong learning' which will be carried out in collaboration with the Open University in The Netherlands (Peter van Rosmalen), that is coordinator of the TenCompetence (<http://www.tencompetence.org/>) IP project. The new STREP project has started in March 2008. Three of the LT4eL partners participate in the new project (i.e. Utrecht University, University of Tuebingen and IPP-Bulgarian Academy of Sciences) which will combine the results and expertise gained within the LT4eL project, the TenCompetence project and the iCamp project. In particular, the new project will offer the possibility to explore the role of ontologies and networks in eLearning further, as well as the possibility to create ontologies semi-automatically on the basis of available techniques (a remark made by the reviewers which can thus be explored in a subsequent phase). In addition, it will be possible to explore the use of more lightweight knowledge representations such as MindMaps.

Results have been exchanged and contacts have been reinforced with the FP6 APOSDLE eLearning project (<http://www.aposdle.org/>) in the area of keyword extraction and ontologies for eLearning. Stefanie Lindstaed, Viktoria Pammer and Luciano Serafini have been informed of the latest results of the LT4eL project and they are investigating whether they can use some of our results.

The possibility of collaboration of UU (Paola Monachesi) in an IP proposal submitted by Claudia Roda (AUP) in the area of *attention* has been discussed. This was the result of various discussions carried out in the past with respect to results in the LT4eL project as well as in the FP6 AtGentive eLearning project (<http://www.atgentive.com/>) in which she was a partner. Unfortunately, the timing of the proposal submission (i.e. March 2008) was not optimal for UU due to the final activities related to the LT4eL project and the start of the new project LTfLL. However, we do hope it will be possible to establish new collaborations in the future to further exploit the results of our projects.

In addition, we have had collaborations and established new collaborations with the following projects:

- I\*teach (<http://i-teach.fmi.uni-sofia.bg/>). We have started adding documents produced by the I\*teach project to our corpus. These documents are particularly suitable for our purposes because they constitute a mini-corpus in a domain which is very close to ours (use of computers in education) of material translated in almost all languages of our consortium (mainly: Bulgarian, Dutch, English, German, Polish, Romanian). In the second year, we added more documents produced by the I\*teach project to our corpus.
- ELENA (<http://www.elena-project.org/en/>). We have exchanged research results on metadata extraction with the ELENA project (finished in May 2005).
- CALIMERA (<http://www.calimera.org/>). We have used the documents collected within this European project to assemble our corpus of learning objects.
- PROLEARN (<http://www.prolearn-project.org/>). All LT4eL electronic meetings are organized using their developed application Flashmeeting (<http://flashmeeting.open.ac.uk/>).

- iCamp (<http://www.htk.tlu.ee/icamp/>). The iCamp project is planning to use our corpus of LOs for their research.
- BAZAAR (<http://www.bazaar.org/>). We have used the Bazaar community to disseminate our project results. Bazaar supports a community for teachers and trainers by providing information and services for teachers and trainers in the development, creation, exchange and use of e-learning materials.
- SHARE (<http://www.share.uni-koeln.de/>). A contact has been established with this project through Matthias Kunkel because they are interested in our tools for metadata generation. They are also informed about our latest results.
- MELT (<http://info.melt-project.eu/>). A contact has been established with this project (Frans Van Assche) in order to provide information about the results of our project. They are interested in multilinguality of LOM instances and they will investigate to which extent our multilingual keyword extractor could be used in the context of their work.
- KP-Lab (<http://www.kp-lab.org/>). A contact has been established with this project (Liisa Ilomaki and Seppo Toikka) in order to inform them of the results achieved within the LT4eL project and investigate to which extent it is possible to establish a collaboration in this respect.

### 4.3 eLearning community in the Mediterranean countries

The tools and methodology developed within the LT4eL project can have an impact and can be exploited also in the context of other languages and cultures. We have investigated the impact that Language Technology can have for facilitating the integration among Mediterranean countries on the basis of eLearning. To this end contacts have been established in several Mediterranean countries which might lead to the creation of new collaborations and projects.

**Egyptian eLearning:** initial contacts have been established by Paola Monachesi at the INFOS 08 conference in this respect with Prof. Galal Hassan, chair of the department of Information System at University of Cairo. We discussed the need for eLearning at his university due to the exponential growth of students and the increasing availability of good infrastructure and the plans for the organization of elearning standards for Egyptian universities to be set by the national eLearning center. After the conference, a meeting was organized with Prof. Galal Hassan and the new responsible for the National eLearning center (<http://www.nelc.edu.eg/>), Prof. Tarek Gharib to discuss possibilities of collaboration between the LT4eL project and the eLearning center. Paola Monachesi has sent the relevant material about the project and at the moment we are investigating the possibility to submit an elearning project based on a collaboration between Utrecht University and University of Cairo to a Dutch funding agency targeted at developing countries.

**Palestinian eLearning:** Contacts have been established by Paola Monachesi at the IMCL 08 conference with M. Qusay (Canada - Palestinian Admn. Area) wrt. research to be carried out by his students in adopting the semantic search implemented in LT4eL for reusability of content and distribution via mobile devices. In addition, we are investigating possibility of grants proposal with Jamil Itmazi (Palestinian Admn. Area) in order to use results of the LT4eL project in blended learning which will allow easier access to education in Palestinian Admn. Areas, which is rather problematic at the moment due to the numerous check points that make mobility of students an issue.

**Jordanian eLearning:** Contacts have been established by Paola Monachesi during the IMCL 08 conference with Razan Khatib and Alma Khasawnih, they have created a Web 2.0 elearning site called Questler ([www.questler.com](http://www.questler.com)) with the aim to promote collaborative learning and discussion especially in the Middle East community. They are considering our semantic search driven by the ontology as a way to access information on their site. We have discussed the Jordan eLearning situation with Eman Y. Tobail from the Jordan Education Initiative ([www.jei.org.jo](http://www.jei.org.jo)) and exchanged information about relevant activities which could be carried out in the framework of their funding also using LT4eL tools such as semantic search. A dinner was organized in this respect with Fabrizio Cardinali (CEO of Giunti Labs) and Said Jahama, general manager of Moodle Rooms, Arabia, to discuss the elearning situation in Jordan and the Middle east and their commercial plans in the future.

## 4.4 Semantic Web (SW) community

The LT4eL project makes use of ontologies in order to carry out cross-lingual retrieval. Ontologies are also a key component of the Semantic Web vision. The results achieved by the project can thus have an impact also for the SW community. In this respect several actions have been carried out.

The collaboration established by Paola Monachesi with the Open University UK (KMI group) in particular, Enrico Motta and Mathieu d'Aquin (member of the user panel) has been reinforced. The ontology and the lexica produced in the LT4eL project have been sent to Mathieu d'Aquin for distribution through the Watson ontology search engine which is being maintained by OU-UK. In this way, we ensure exploitation of the ontology beyond the project and reuse by the Semantic Web community. In addition, the use of Watson to develop new domain ontologies in the context of the LT4eL project has been investigated in the Master thesis of Jantine Trapman.

Our collaboration with Paul Buitelaar (member of the user panel) in the area of ontology development has been reinforced with discussions carried out with several members of the project during the LREC08 conference. The ontology and the lexica produced in the LT4eL project have been given to Paul Buitelaar for distribution through Ontoselect, which is an ontology library being maintained by DFKI. In this way, we ensure exploitation of the ontology beyond the project and reuse by the Semantic Web and NLP community.

The collaboration with CNR Rome - Laboratory of Applied Ontology (Aldo Gangemi, Alfio Gliozzo - user panel member) established by Paola Monachesi has been reinforced during the LREC08 conference. New areas of collaborations have been envisaged and LT4eL results have been exchanged with Alfio Gliozzo in the area of semantic search to be reused in their project eu project Babylon and Ontology (<http://www.bonynetwork.eu/>). In particular, they are interested in the semantic search functionality the way it has been implemented in the LT4eL project, as well as on the use of ontologies in retrieval of Learning objects. We expect the collaboration to continue in the new project LTfLL in the area of social web and eLearning. In addition, we have distributed the ontology and the lexica produced in the LT4eL project to CNR since we have carried out a simplification of the DOLCE ontology (developed by Aldo Gangemi) which might be useful for additional real life applications.

## 4.5 Language Technology community

Language technology has played an important role in the LT4eL project which has made extensive use of tools and resources to enhance eLearning. The close relation with the Language Technology community is reflected in the fact that the project has been very successful in conferences such as the one in Language Resources and Evaluation (LREC08). The project was present at this conference with eight papers. We believe that in this way we have made an effort to raise awareness wrt. the potential of Language Technology for eLearning.

More concretely, we plan to carry out activities in collaboration with the CLARIN ([www.clarin.eu](http://www.clarin.eu)) project aiming at building an infrastructure of language resources for the Humanities. Our goal is to use our project to suggest that eLearning could be a valuable application. Most partners of the LT4eL project are part of this project and eLearning could be considered an interesting application. Utrecht University of the coordinator of CLARIN facilitating thus the connection.

## 4.6 Commercial organizations

The Lt4eL project was quite successful in being able to develop the planned Language Technology functionalities and to integrate them into ILIAS. However, there is still some work to do to develop our prototype into a commercial product and some actions have been taken in this respect.

In the context of exploitation and commercialization of the LT4eL results, new contacts have been established by Paola Monachesi with Giunti Labs through Fabrizio Cardinali (CEO Giunti Labs) during the IMCL08 conference in Amman. The project software has been provided and Giunti will investigate the LT4eL functionalities in more detail. They believe that the functionalities developed by the LT4eL project (especially the keyword extractor and glossary candidate detector) could be in principle integrated in their

Content Management System and might be relevant especially for some of the Eastern European languages. However, this will involve resources on their side which are not available at the moment; they have decided to keep the documentation in standby and to get in touch with the project in case any of their clients would be interested in having their Content Management System enhanced with our functionalities. It was agreed that due to the common research interests, it could be desirable to collaborate more closely in future projects in order to have results which could be directly integrated in their products.

More generally, it seems difficult to commercialize the results of this kind of European Projects. This is a difficulty that cannot be overcome even if an interested SME is part of the project. Usually the lack of funding is a big obstacle for the further development of the results obtained into a commercial enterprise. The SME cannot commercialize the product while it is being developed and after the project has finished it often doesn't have the funding to develop the prototype into a product.

We believe that a solution to this problem might be the creation of venture capital companies which could constitute a bridge between industry and research institutions. They would employ the results obtained in such projects in a way that they could be used by the commercial enterprise. To this end, we are exploring possibilities to create a spin-off of the University of Utrecht which could have this task. The spin-off should be in a position to translate the research results and the prototypes into a format that could be commercialized with little effort.

## **4.7 Higher education institutions**

The project is now ripe to organize meetings with higher education institutions and university to show the potential of eLearning. Attempts have been made in this direction at Utrecht University but it is rather impossible to have an impact on the commercial policy with respect to eLearning tools that Universities have adopted against Open Source products. Nevertheless, some more attempts in this direction have been made in the last phase of the project.

- We hope that it would be possible to reach higher institution through activities carried out in collaboration with the CLARIN ([www.clarin.eu](http://www.clarin.eu)) project aiming at building an archive of language resources for the Humanities. Most partners of the LT4eL project are part of this project and eLearning could be considered an interesting application. Utrecht University is the coordinator of CLARIN facilitating thus the connection.
- We are investigating the possibility to organize a final workshop of the LT4eL project at Utrecht University either in the fall 2008 or at the beginning of 2009 targeted at a broad public including Dutch ministries and higher education institutions in order to disseminate the results of our project. We plan to organize the workshop in collaboration with OU-NL (Peter van Rosmalen) in order to show the continuity of the TEL research in the Netherlands and the relation between the two projects. We aim in this way to raise awareness for this kind of research in the Netherlands.

## **4.8 The user panel**

The user panel consists of representatives of the different fields addressed within the LT4eL project. We aimed at an equally representation of eLearning, Semantic Web and Language Technology experts. Besides, the panel contains both representatives from academic and commercial institutions. The list with members of the panel can be found in the dissemination report.

The panel has been informed on the activities and results of the project on a regular basis the last year. Besides, some of them were in June 2007 present at the workshop in Prague organized by the project. They provided us with feedback and ideas on the basis of their expertise in their research areas. The results of the project can be exploited in different ways by the users of the panel. Whereas the Language Technology experts are more interested in the technologies used within the keyword extractor, the commercial representatives are interested in the use of the tools within their own products.

## **4.9 PhD projects and Master theses**

The project has been the starting point for three PhD theses. They focus on different aspects of the project.

- Rosa Del Gaudio, University of Lisbon, Portugal: As a follow up of the project Rosa Del Gaudio asked a grant to do a PhD. The grant was asked to "Fundação para a Ciência e a Tecnologia" (FCT), the Portuguese research council. The main focus of the research project is the automatic extraction of definition for Portuguese language. This work includes the construction of a system for the automatic extraction of definition as well as its integration in some applications where it can play a key role, in particular in the context of a question answering system, and for semi-automatic extraction of ontologies. The starting point will be the glossary candidate detector developed during the project, that will be improved trying to exploit deep linguistics analysis and machine learning algorithms. Furthermore, the corpus collected for the project will be used, at least in a first phase, to validate new versions of the system.
- Ionut Pistol, UAIC, Romania: Ionut Cristian Pistol is working on a Ph.D. at the "Alexandru Ioan Cuza" University in Iasi, under the supervision of prof. Dan Cristea. The Ph.D. thesis will be completed at the end of 2008 and will be named "The Automated Processing of Natural Language Discourse" and will involve the development of ALPE, a LP meta-system improving on several aspects existing systems, such as GATE and UIMA. Since our project requires the employment of several linguistic processing chains as a functionality of the ILIAS management system, it was decided in a workshop held in June 2007 that ALPE will be used as a management system for the processing chains which will be deployed. These processing chains take the input documents (in plain txt format, UTF-8 encoded) and bring them automatically to the format required for the KWE and GCD tools used in the ILIAS system. The chains available (for English and Romanian languages are available online at <http://thor.info.uaic.ro/~ipistol/ALPE/res/chains/ENGLPC.php> and <http://thor.info.uaic.ro/~ipistol/ALPE/res/chains/ROMLPC.php>).
- Eline Westerhout, Utrecht University, The Netherlands: The precision obtained with the Glossary Candidate Detector (GCD) developed within the LT4eL project is not very high. Therefore, the Dutch team started experimenting with Machine Learning techniques to try to improve the results. In her PhD project, Eline will go on with this work and investigate in which way the best results can be obtained. The focus of the project will be on the extraction of different types of definitions, such as definitions with a form of 'to be', definitions in which another verb is used as connector, punctuation patterns and pronoun patterns. In the project she will explore which approach gives the best results: only machine learning, only a grammar or a combination of the two. We expect to get the best results by using a combination of the two approaches. As for the grammar, she has investigated within the LT4eL project how far we can get by using only a grammar. On the basis of manually extracted definitions rules have been designed to match definition patterns. The rules rely heavily on the linguistic structure of definitions. However, because a definition pattern not always means that we really have to do with a definition, the main problem is to get an acceptable precision with this approach. This is where the machine learning comes in. We will investigate which classifier gives the best results and what kind of information plays a role in filtering. The combination of the two approaches will be done in several ways. We will combine at least two types of grammars with machine learning. The first grammar will be a very basic grammar, classifying the sentences only on the basis of the presence of a certain verb or punctuation mark in the sentence. The second grammar will be a more sophisticated one capable at extracting very specific structures. We will also try machine learning as a separate approach. The final GCD will be evaluated in the context of learning and semi-automatic glossary creation, this is the way in which the tool is used within the LT4eL project. In her PhD project Eline will use many of the results of the project: the corpus created in WP1 for the validation of the GCD, the work on the GCD performed within WP2, and the results of the validation (WP5) to get an idea of what kind of results users appreciate.

Two other persons are writing their Master thesis related to the project:

- Claudia Borg, UOM, Malta: Within the LT4eL project, we look at definition extraction using a rule-based approach. This thesis explores the possibility of using Genetic Algorithms and Genetic Programming to rank definitions and learn new grammatical patterns which could be then included within the rule-based approach. These learning algorithms will be fed the manually written patterns within the project as the initial seed to their population, so as to develop new improved rules with the aim of classifying definitions correctly. The aim of this work is to compare this technique to other techniques used in definition extraction. This work will use the manually annotated definitions from within the English corpus in the LT4eL project as part of the training definitions for the learning algorithms.
- Jantine Trapman, UU, The Netherlands: In the developmental phase, the domain of the LT4eL ontology and the lexicons is restricted to Computing. This thesis describes how other domains can

be included in a semi-automatic way using semantic web techniques. A new domain ontology is created using algorithms described in (Basili et al, 2007). It yields concepts from WordNet via which we can map the ontology easily to LT4el. At the same time, also using the WordNet identifiers, the required terms of the domain can be extracted from the Dutch lexical database Cornetto Database in order to add them to the Dutch lexicon in LT4eL. The additional concepts and the lexical information originating from Cornetto can be used to improve the search and for word sense disambiguation. The thesis has been defended in May 2008.

## 5 Overview table

<b>Exploitable Material</b>	<b>Sector of Application</b>	<b>Timetable for use</b>	<b>Patent or other IPR protection</b>	<b>Owners and other partners involved</b>
Corpus of LOs	Research & eLearning industry	2007	IPR free	all partners
key word extraction & glossary candidate detector	Research & eLearning industry	first version: 2006; second version: 2007 & final version: 2008	IPR/patent free	all partners
ontology & language specific vocabularies	Research & eLearning industry	2007	IPR/patent free	all partners
ILIAS integrated functionalities	eLearning	first version: 2007; final version: 2008	IPR/patent free	all partners
Validation methodology	Research + eLearning industry	2008	IPR/patent free	all partners