PROGRESS REPORT

Grant Agreement number: 250467

Project acronym: ATLAS

Project title: Applied Technology for Language-Aided CMS

Project type: ☐ **Pilot A**   **V Pilot B**   ☐ **TN**   ☐ **BPN**

Periodic report: **1ˢᵗ** ☐   **2ⁿᵈ** ☐   **3ʳᵈ V   4ᵗʰ** ☐

Period covered: **from   01.03.2012 to 31.08.2012**

Project coordinator name, title and organisation:

Anelia Belogay, CEO, Diman Karagiozov, CTO,

Tetracom Interactive Solutions

Tel: 35924950444

Fax: 35924950443

E-mail:anelia@tetracom.com, diman@tetracom.com

Project website address: www.atlasproject.eu

DECLARATION BY THE PROJECT COORDINATOR

I, as coordinator of this project and in line with my obligations as stated in Article II.2 of the Grant Agreement declare that:

- The attached periodic report represents an accurate description of the work carried out in this project for this reporting period;

- The project (tick as appropriate):

  V   has fully achieved its objectives for the period;
  has achieved most of its objectives for the period with relatively minor deviations;
  has failed to achieve critical objectives and/or is deviating significantly from the schedule.

- The public Website is up to date;

- [this point only applies to projects with actual cost reimbursement] To my best knowledge, the information contained in the financial statement(s) submitted as part of this report is in line with the actual work carried out and consistent with the reported resources and if applicable with the certificates on financial statements.

Name and position of Coordinator: Anelia Belogay, CEO

Date: 31.10.2012

Signature:

PUBLISHABLE SUMMARY

Introduction

The advent of the Web revolutionized the way in which content is manipulated and delivered. As a result, digital content in various languages has become widely available on the Internet and its sheer volume and language diversity have presented an opportunity for embracing new methods and tools for content creation and distribution. Although significant improvements have been made lately in the field of web content management, there is still a growing demand for online content services that incorporate language-based technology. Mechanisms such as automatic annotation of important words, phrases and names, text summarisation and categorisation, and computer-aided translation could facilitate the process of manipulating heterogeneous multilingual content as well as enhance end-user experience by allowing for better content navigation. This project unifies such mechanisms in a common software platform called ATLAS and builds three separate solutions around this platform.

Summary description of project objectives

The consortium will adjust and integrate several existing software components, assembling a platform for multilingual web content management called ATLAS, and a visualization layer called i-Publisher, which adds to the platform a powerful web-based point-and-click tool for building, reusing and managing multilingual content-driven web sites. An instance of i-Publisher will be made publicly available as an online service. i-Publisher will also be used to build two thematic content-driven web sites – i-Librarian and EUDocLib.

The ATLAS project aims to meet the following objectives:

- Software platform and services, demonstrating the latest achievements in the field of multilingual web content management and addressing the needs of individuals and organizations for easier web site building and content publishing.
- Liaison with the Europeana and EuroMatrix Plus initiatives in order to foster language diversity in content creation and distribution
- Interoperability by conforming to a number of widely recognized web, natural language processing, and content management standards
- Sustainable management format to ensure the progress of the project
- Mechanisms and procedures that enable and simplify the addition of new languages to the ATLAS platform, thus targeting all major European languages after the successful completion of the project.

Expected final results

The primary goal of the ATLAS project is to facilitate organizations and individuals who manage and publish multilingual content. Thus, the project solutions will not merely meet the needs of modern multilingual content management, but also create value for all users.

Main expected final results:
- The software solutions built during the project reveal the true value, capabilities and power of several existing tools for web content management, multilingual versioning, and natural language processing by combining them in an innovative manner and offering the end results to the general public at no cost.
- With i-Librarian and i-Publisher users can easily create, manage and publish multilingual content without installing and maintaining a standalone system. Nevertheless, they retain full control over their content regardless of whether it is in their private, shared or published workspace. EUDocLib provides easy and intuitive access to a vast collection of EU law documents.
- The ATLAS platform is designed with extensibility in mind, which allows for easy addition of tools for currently unsupported languages as well as new tools for already supported languages.
- Furthermore, ATLAS significantly reduces the time and efforts for content authoring and editing because it automatically categorizes, summarizes, annotates and translates documents regardless of their language and format. The software platform enables i-Librarian users to find the most essential texts from large document collections by displaying text summaries and extracted important phrases, words and names.
- Finally, ATLAS improves content navigation by interlinking content items based on text annotations and by automatically placing the content items in appropriate subject categories.

Potential impact

The project brings together advanced technologies for multilingual web content management and text mining (such as automatic annotation, mark-up and translation) in a united platform. The intended software-as-a-service architecture of the envisaged solutions, which demonstrate the capabilities of the ATLAS platform, and the open-source license, will facilitate the spread of the project output.

Main expected impacts:
- Technological
  - Integration of text mining tools into content management systems
  - Integration of text mining services
  - Stable and more efficient Machine Translation modules for the project languages. The language pairs considered in ATLAS are covered by Google Translation but with very low quality. On the other hand, these language pairs have strong relevance for the Central- and East-European commercial space.
  - Contribution to the development of text processing chains for languages, which lack resources at present
  - Adherence to and promotion of existing and future web standards
  - Practical and economically viable solutions for nearly-automatic provision of multilingual online content and services for some EU languages
- Social

- o Facilitate exchange of information and knowledge
- o Simplify authoring, management and exploitation of heterogeneous multilingual content
- o Address the needs of a large number of people belonging to different target user groups – individuals and organizations
- o Cross the language barrier
- o Facilitate culture exchange
- o Liaise with Europeana and EuroMatrix Plus – The liaison with EuroMatrix Plus will be established at the beginning of the project. Europeana will be approached by the end of the first year, when the consortium will be able to demonstrate the potential value of ATLAS to the European digital library.

Use

The ATLAS platform as a whole and also some of its standalone components are beneficial to different groups of users. Thus the consortium has distributed the potential users of each major software component into several target groups while paying special attention to the needs and requirements of each group. The table below summarizes this distribution:

Target groups

Component	Target group
ATLAS (includes KMS Content Management System, Text Mining engine, Search engine, Summarisation Engine, Machine Translation engine) + i-Publisher (ATLAS web-based graphical user interface for building interactive, content-driven web sites)	Web design companies – faster prototyping, web design and site building
	Hosting companies – as part of hosting packages
	Education, Media, Publishing, Non-profit, Government
Text Mining engine	Online bookstores
	Digital libraries/repositories
	News agencies/websites
i-Publisher (as online public service)	Small enterprises
	Non-profit organizations
i-Librarian (thematic content-driven web site built with i-Publisher)	Students, Researchers
	Readers
EUDocLib (thematic content-driven web site built with i-Publisher)	The general public

Table 1: Target groups

Project web site

More information including project details, news, and contact information can be found at http://www.atlasproject.eu

PROJECT PROGRESS

Project objectives for the period

With regard to the management objectives set for the reported period the following tasks have been completed:

- The management and coordination framework ensured the smooth progress of the project.
- The control of the quality and timely delivery of project deliverables as well as monitoring of the allocation and distribution of the project resources resulted in implementation of all tasks as planned.
- One project meeting was organized during the reported period. It was held in Luxembourg. The Consortium checked all documents related to the upcoming review meeting and discussed the project progress and defined the next steps needed to achieve the objectives for the next periods.
- The second year report covering month twelve to month twenty four of the project was prepared and submitted to the EC.
- The second review meeting was held in Luxembourg.

The work done in terms of the technical objectives set for the period includes the following:

- WP 2
    - ATLAS improvements - major performance and stability improvements; improved workflow of the process of invitation and user registration.
    - i-librarian extension made possible the internal evaluation of ATLAS on corpus containing 186 manually annotated documents. The evaluation results can be accessed at: http://www.i-librarian.eu/ilib/i-librarian/eval_list using the following user account: username: eval@i-librarian.eu, password: eval1234
    - Regression test module - a strategy for regression tests was built as a result of the reviewer recommendations. It was followed by the implementation of a regression test module.
    - An integration module for one of the most popular open-source content management system Alfresco was done. The module provides the linguistic annotations in XML and/or JSON format; the annotations are parsed by the implemented Alfresco extensions;
    - Several demo services were implemented to prove the qualities of ATLAS services and to be used as main dissemination instruments. The demo services are:
        - A demo service using publication from UNESCO website
        - A demo service for World Health Organisation
        - A demo service for ISN at ETH, Zurich
        - Media demo services:
            a. A multilingual media service – the corpora of CETimes is processed, analysed and presented in a service powered by ATLAS in Bulgarian, English, Greek and Romanian.

  b. A Bulgarian news service – all big digital media provider's news is collected, processed and presented in a web site.

  c. A Bulgarian video news service – video materials from the Bulgarian Television are collected, the related texts are processed in order to semantically enriched the videos and to provide text search functionalities for them.

- o The technical documentation was updated with the newly integrated modules (categorization and summarization).
- o Four demo services were implemented with Atlas. They will be used as dissemination instruments to attract the Atlas target groups.

- WP 3
  - o The final fine tuning of the categorization tool was done.
  - o Standalone evaluation tools for finding the optimal model settings was implemented. Using the tool we achieved results above the state-of-the-art reported figures on the Reuters-21578 corpus.
  - o Using the categorization tool we found the optimal parameters of the classification models in i-Librarian in all project languages.
  - o A new approach to categorise content by using metrics to measure the similarity between content items and category vectors was implemented and applied in the demo services.
  - o The Deliverable D 3.1 was sent to the Commission.
- WP 4
  - o multilingual resource GEO-Names (ATLAS-GEO) which provides content similar to JRC-Named content but focusing on geographical locations were implemented in order to increase the quality and recall of the extracted locations;
  - o Name Entities improvements – several NER engines has been combined in order to improve the quality of the extracted person names and organizations;
  - o Freebase person name NER has been implemented using lexicons built form the Freebase lists of people;
  - o BOM UTF-8 text extraction bug has been found and fixed;
  - o Bug fixes and minor improvements in all LPCs – EN POS tagger and lemmatizer, EL and PL NP extractor, BG NER, DE stability;
  - o Extensions in the LPCs to provide richer linguistics annotations for the summarization tool in WP5 – "gendered" nouns, NP improvements;

- WP 5
  - o The summarisation model has got significant improvements in different modules (anaphora resolution: interpreter of rules that allow description of resolution constraints; clause segmentation and discourse parser: evaluation and calibration loops, that allow optimisation of the parameters for achieving the best results; consideration of the summariser as a separate module from the discourse parser);

- o the whole summarisation chain has been tested and evaluated for each of the languages: BG, DE, EN, GR, PL, RO;
- o all chains have been integrated in U-Compare, a variety of UIMA;
- o the D5.1 deliverable was written, improved, and continuously updated

- WP 6
  Machine Translation
    - o definition of a methodology for semi-automatic acquisition of domain-corpora
    - o acquisition of domain corpora for all 15 language pairs for 13 domains (195 parallel corpora with minimum 2000 sentence pairs)
    - o model training for domain data
    - o final implementation of EBMT engine and adaptation for ATLAS integration
    - o final implementation of the CLIR engine
    - o monolingual tests with the CLIR
    - o first integration in the ATLAS-System.
  Crosslingual retrieval
    - o final implementation of the CLIR engine and monolingual tests with the CLIR;
    - o Standardization of the interface between the ATLAS-platform and the cross-lingual search engine which led to the first integration in the ATLAS-System.

- WP 7
    - o Finalisation of the 2nd user evaluation round; statistical analysis and conclusions
    - o Comparative assessment between user evaluation results from the 2 rounds
    - o L.L. evaluation workshops; drawing of conclusions

- WP 8 – work carried out on dissemination for this period includes:
    - o "ATLAS" stand at the biggest world digital fair Cebit 2012
    - o ATLAS services were demonstrated at the Demo section at the most significant scientific event for the year in Europe - EACL 2012
    - o ATLAS was presented with a paper at the biggest linguistic conference LREC 2012
    - o The project was presented with a stand at META-FORUM 2012
    - o The project presentation was held at CESAR META-NET Roadshow, Sofia
    - o Dissemination activities with target groups:
        - An initial meeting with World Health Organisation representative in Salzburg, June
        - A meeting with the Bulgarian TV representative discussing possible ways of integration of the Atlas services in the internal environment of the Bulgarian Television, July
        - A presentation for the association of the Bulgarian Libraries in a conference , May
        - A presentation during a meeting with the Institute for Security Studies at ETH Zurich, August

- A meeting with an industrial customer eMatch during LT Innovative, Brussels, June
  o Performed dissemination activities during the reported period (01.03.2012-31.08.2012), are summarised in Annex 1.

Work progress and achievements during the period

Work package 1 – Project Management
The package description is placed in chapter "Project management".

Work package 2 – Software Specification, Implementation and Deployment
The work package focused on improvements of the functionalities and usability of the Atlas services. In addition, the package includes tasks to keep the technical documentation updated. Finally, the package covers the implementation of the first prototypes of services to be used as dissemination instruments of the project.

T2.5 (23 PM) – i-Publisher – implementation and deployment
- As requested after the second-year project review, Tetracom invested time and resources in optimizing the performance of i-Publisher and web sites built with i-Publisher. The performance improvements have been achieved by optimization of the database queries and minimization of the communication between different ATLAS components. Smarter caching strategies have been added to i-Publisher as well.
- i-librarian extension made possible the internal evaluation of ATLAS on corpus containing 186 manually annotated documents. The results of the evacuation were used to improve the quality of the semantic extracts provided by the LPCs.
- Regression test module. As requested in the review report, the Consortium built a strategy for regression tests and Tetracom implemented the regression test module. The module tests the integrity of the Atlas components. The test infrastructure helps the ATLAS consortium to identify problems and deviations in the quality caused by the deployment of new versions of components, changes and bug fixes, communication failures. The regression test module was integrated in Atlas.
- New widgets – based on the user feedback and change requests, Tetracom implemented new widget maintaining the web forms. The information filled out in a web form can be automatically sent to predefine email(s) upon configurable set of events (e.g. user confirms their registration, user logs in for the first time, etc.).
- The technical documentation was updated with the newly integrated modules.
- Four demo services were implemented with Atlas. They will be used as dissemination instruments to attract the Atlas target groups. The demo services are:
  o A demo service for UNESCO
  o A demo service for World Health Organisation
  o A demo service for ISN at ETH, Zurich
  o Media demo services:

.1. A multilingual media service – the corpora of CETimes is processed, analysed and presented in a service powered by Atlas in Bulgarian, English, Greek and Romanian.

.2. A Bulgarian news service – all big digital media providers's news is collected, processed and presented in a service.

.3. A Bulgarian video news service – video materials from the Bulgarian Television are collected, the related texts are processed in order to semantically enrich the videos and to provide text search functionalities for them.

T 2.6 - i-Librarian – implementation and deployment

i-Librarian is a thematic web site (online service), which encourages visitors to register and get a personal workspace where they can store, share and publish various types of documents and have them automatically categorized into appropriate subject categories, summarized and annotated with important words, phrases and names.

The latest extension of i-Librarian made possible the internal evaluation of Atlas services. The purpose of the evaluation was to assess the quality of the linguistic extracts provided by the NLP tools used in i-Librarian after the second year. The corpus contains 186 test documents, evenly distributed among the six project languages was processed. The results of the processing are shown in a similar to i-Librarian way but includes several improvements:

- the list of named entities (person names and organisations, location, dates), as well as the list of important phrases (incl. their head nouns) show up to 100 entities (standard i-Librarian shows max 20; head nouns are not shown to the user);
- new facility to evaluate different aspects of the i-Librarian functionalities – text extraction, different types of named entities, important phrase, categorization and summarization.

The evaluation results can be accessed at: http://www.i-librarian.eu/ilib/i-librarian/eval_list using the following user account: username: eval@i-librarian.eu, password: eval1234

T 2.8 - Technical documentation

ATLAS technical documentation draft, i-Publisher technical documentation draft and i-Librarian and EUDocLib technical documentation drafts include a thorough description of the platform architecture and design, and emphasize the multilingual capabilities of the software. The documentation is kept updated. The chapters for the summarisation and the machine translation will be added when the tools are implemented and integrated in Atlas.

Work package 3 – Automatic Categorisation
The focus of WP3 during the reported period was to finalise the implementation of the categorisation tool and the corresponding deliverable.

T3.2 (25 PM) – Fine-tuning to specific languages

- Tetracom implemented an external to ATLAS evaluation tool to formally measure the quality of the classification algorithms included in the Classification Tool. Using this tool we evaluated the Reuters 21578 corpus with 3 algorithms and one meta-algorithm (ensemble). Results, above the state-of-the-art, have been achieved. The evaluation results can be checked at https://docs.google.com/spreadsheet/ccc?key=0AuWVg_Q7c1mudExmNVJBMHgyTXdaeWhiTGhlcEZlX1E#gid=0
- The evaluation tool helped us to find optimal categorization parameters for each of the language models used in i-Librarian. These results can be significantly improved in the future if more training data becomes available.

T3.3 (18 PM) – Software adjustments (in resp. to technical reviews)
Faced with the use case when the training data is not enough Tetracom introduced a new way of unsupervised categorization – the available training data, for each category, is represented as a single document which is then processed by the LPCs. Then we index the full text of the document as well as the annotations (tokens, NPs, heads, lemmas) using Lucene, thus we are capable of fining the most significant features and their relative weight for each category. In the same way, new uncategorized documents are processed and the final step is to perform a query in the Lucene indexes and find the closest categories profiles (top-N terms and their weights). This method was firstly used in the WHO demo web site where available training data for each category was not more than 5 sentences. The WHO staff checked the results are their feedback was very positive. There is, however, still a space for improvements, mainly related to the proper setting of a similarity threshold.

T3.4 (14 PM) – Documentation
The categorisation modules were documented and the documentation was integrated in the technical documentation of Atlas

The deliverable D 3.1 has been sent to the Commission.

Work package 4 – Language Processing Chains
The main focus on WP 4 for the reported period was to:

Strategy for regression testing has been prepared and adopted. Respective document has been sent to the EC on 24 May 2012 together with ATLAS assessment after performing the internal evaluation.

The consortium identified a potential problem related to the foreign names in local texts. For example, a Greek name may appear in German text but the German gazetteers contain mainly German names. Thus, the Greek name may not be recognized. This problem is partly solved by JRC-Names as it focus on person names (96%) and organization (4%). The locations however and un-popular local names may not be recognized correctly.

A multilingual resource GEO-Names (ATLAS-GEO) which provides content similar to JRC-Named content focusing on geographical locations has been created and all LPCs extended with it.

T4.2 – Software adjustments (in resp. to technical reviews)

ICS PAS
A series of follow-up actions have been carried out by ICS PAS to improve quality of the Polish LPC, mainly regarding the recognition of named entities:

- The multi-word unit lemmatizer has been improved. Now non-positive adjectives are not lemmatised to their positive forms (as in "better" → "good"), what is not natural in this context, but is forced by the way Pantera tagger works. This has been achieved by using Morfeusz form generator, the same that generates correct non-masculine gender forms for adjectives. The forms are replaced in the ann_morphosyntax file using TEI API.
- The gerundial forms are generated faster.
- Solved technical problems with compiling and integrating the new version of Spejd tool.

Separate work is being carried out by Jakub Waszczuk (outside of ATLAS project) on development of MWU lemmatization for proper names. If the work is completed before end of February 2013, it will be integrated in ATLAS, possibly replacing the current solution.

ICS PAS collected and structured the multilingual geographical locations for the multilingual GEO-Names resource by harvesting the Wikipedia/DBPedia, CIA factbook and/or other resources.

Presentation of Polish NPs has been improved by removing NPs containing pronouns classified as adjectives.

Tetracom
Improvements in the NERs have been made for English. OpenNLP-based NER is extended with rules that resolve ambiguities. Different NERs have been unified to resolve the problem with the foreign names in local texts. Person names, organization and locations NER have been extended with additional lexicon based on Freebase.

Tetracom created an UIMA primitive engine for the multilingual GEO-Names resource and included it in all language processing chains.

Tetracom found a problem related to "UTF-8 with BOM" plain text files, which blocked some of the LPC or did not provide the expected annotations. The process of text extraction from DOCX, RTF, ODT and PDF documents was significantly improved by resolving this issue. The ICU4J library is used to detect the character encoding of the input text. Occasionally, the library failed to correctly detect the encoding of some Greek and Cyrillic texts. Tetracom integrated another library that supplements the ICU4J for these two types of texts.

Tetracom enabled new functionality providing access to the textual content of a document in i-Librarian.

Tetracom also improved the general stability of the German LPC by decomposing the monolithic GATE component into smaller UIMA sub-components. The communication between the UIMA components and the external dependency parser ParZu has been improved.

IBL

A series of follow-up actions have been carried out by DCL-IBL to improve quality of the Bulgarian LPC, mainly regarding the recognition of named entities:

- The ParseEst has been improved by means of improving the underlying finite state technology, the xml based language for rules definition and lexicon compiler.
- Quality of the named entity extraction component has been improved by streamlining the rules and enlarging the lexicons. The ambiguity between different types of named entities included in the lexicons were automatically cleaned up and manually validated. Large lexicons for the NER were compiled automatically and validated semi-automatically.
- The NER and NPE regression tests were implemented.

UIAC

Romanian LPC has been improved to avoid sending problematic texts to the NER component. The whole LPC now features higher stability.

Performance optimizations have been introduced to the POS tagger and NounPhrase extractor.

Atlantis

- Stability improvements have been introduced to the Greek LPC.
- Noun Phrase Extractor problem (hanging up the whole LPC processing in certain exceptional cases) has been fixed.
- Named Entity Recognizer now recognizes genitive of location names coming from GEO-Names multilingual lexicon. To this end, 2 lists have been prepared:
  - 2.676 entries with geonyms, hydronyms, countries, areas, etc. in nominative and genitive
  - 60 entries with continents, regions, etc. in nominative and genitive - Several test and evaluation files were prepared for assessing the performance of the updated NER.
- Improvements were applied to the Greek Sentence Splitter Primitive engine (performance I terms of execution time).

UHH

- Improvements in NE (geographical locations and NE)

- Improvements in extraction of nested NPs

T4.4 – Integration of LPCs in ATLAS

Tetracom created a standalone UIMA LPC wrapper so that the LPCs can be tested outside ATLAS. The standalone wrapper tool is also used for processing of the corpora for categorization, summarization and machine translation tasks.

T4.5 – Documentation

As requested after the second-year project review, **Tetracom** added a chapter in the technical ATLAS documentation about adding a new Language processing chain in ATLAS.

Work package 5 - Text Summarisation

Months March-September 2012 have been used by the ATLAS partners to build the first operational summarisation chain and to continuously improve the results. UAIC, as the main developer, has adopted the final architecture, has revised and improved the functionality of the summarisation chain and has coordinated the work of the contributing partners to improve their specific resources. The contributors in this WP, namely ATLANTIS, IBL-DCL, ICS-PAS and UHH have built the specific resources for their languages (respectively: GR, BG, PL and DE) and have transmitted to UAIC the output of the XML test files in their languages. Tetracom has coordinated the integration process and had intensive communication with UAIC in order to perform the integration in ATLAS.

T5.1 (74 PM) – Implementation of text summarization tools

All modules of the summarisation model (as reported in D5.1) have been tested separately and in the complete chain. The complete chain include prerequisite modules (actually developed as part of WP4: SEN = sentence splitter, TOK = tokeniser, POS = part of speech tagger, LEM = lemmatiser, NP = noun phrase chunker, NER = name entity recogniser) and proper summarisation modules (AR = anaphora resolver, CS = clause splitter, DP = discourse parser, and SUM = summariser).

T5.2 (6 PM) – Software adjustments (in resp. to technical reviews)

To fulfill the needs of the integration, at the junction point between the prerequisite modules and the proper summarisation modules, a wrapper has initially been built. The wrapper was responsible to adapt the output of the POS-tagger in different languages to the input required by the anaphora resolution module (RARE).

In the proper summarisation chain, a number of modules require specific resources. These modules are: RARE, CS and DP. There resources have been developed by the partners and influenced a training process, intended to adapt the general architecture of the summariser to the specific needs of each language. The optimisation of parameters of each module was automatised in a parameter fixing loop (calibration). This process is not ended.

T5.3 (6 PM) – Integration of the summarization tools in ATLAS

Intensive discussions went on between the implementer of the modules (UAIC) and Tetracom, responsible for the integration of the ATLAS summarisation module in the general ATLAS machinery. A preliminary U-Compare UIMA integration has been performed by UAIC, tested and reported in D5.1. To acquire the final line of completely integrated ATLAS tools some efforts are still necessary.

T5.4 (13 PM) – Documentation
A great part of the documentation for the summarisation tool has been written and reported as part of D5.1. This preliminary documentation will make easier the final documentation writing phase in the project.

Work package 6 – Machine Translation and Crosslingual retrieval:
  o   Standardization of the interface between the ATLAS-platform and the crosslingual search engine.
  o   Implementation of a first prototype of the crosslingual search engine.

T6.1. Fine tuning of existing translation models
This task involves fine-tuning of existing translation models to the needs of the project and creating new models (e.g. German-Romanian). The Acquis Communitaire as well as the SETimes and Europarl, as well as small manual aligned corpora (2000-5000 sentences) corpus will be used as training corpora.

This task is extended from fine-tuning to Development of translation models. This is on-going task. In the months 24-30 following activities took place under this task
  1. Definition for methodology for automatic collection of parallel corpora
  2. Collection of training data: overall 195 corpora of at least 2000 sentences per language for each domain included in the main hierarchy of iLibrarian. The corpora are sentence aligned
  3.   Training of the models necessary for EBMT (word-aligned files by means of GIZA++)
  4. Preparation of the data for running with the EBMT-Engine
  5. Preparation of PoS factored data, by means of the LPCs developed under WP4
  6. Training of Moses Models on JRC-Acquis, Europarl and SE-TIMES
  7. Integration of the EBMT engine

DCL-IBL
Compilation of parallel corpora for 13 domains and 15 pairs of languages
  • Parallel corpora are compiled for all pairs of the following 6 languages: Bulgarian - BG, English - EN, German - DE, Greek - EL, Polish - PL, Romanian – RO
  • The domains are as follows: Business, Computing, Fiction, History, Politics, Mathematics, Biology, Physics, Chemistry, Arts, Music, Geography, Sociology. The domains have been selected in compliance with the text categorisation module and for the purposes of example based machine translation.
  • The parallel texts contain minimum 2000 parallel sentences per domain per pair of languages

At least one counterpart in any pair of parallel texts is original and not generated via Machine Translation (MT).

The following requirements are followed:

Fill as many categories as possible with original parallel data from several main sources (European Central Bank corpus, SETimes, BulNC, etc.) and other texts from the Internet is followed.

For each language pair, the number of filled and unfilled domain corpora (cols. 1-4)

For each domain, the number of filled and unfilled language pairs (cols. 5-8)

Lang. Pair	Full	Partially filled	Empty	Domain	Full	Partially filled	Empty
EN-BG	6	4	3	Business	15	0	0
EN-DE	6	4	3	Computing	9	3	2
EN-EL	4	3	6	Fiction	6	2	7
EN-PL	6	3	4	History	0	15	0
EN-RO	6	3	4	Politics	15	0	0
BG-DE	4	4	5	Maths	0	0	15
BG-EL	4	4	5	Physics	15	0	0
BG-PL	6	1	6	Biology	0	0	15
BG-RO	5	3	5	Chemistry	0	0	15
DE-EL	5	1	7	Arts	0	8	7
BG-PL	5	1	7	Music	0	7	8
BG-RO	5	2	6	Geography	1	5	9
EL-PL	5	1	7	Sociology	15	0	0
EL-RO	4	3	6				
PL-RO	6	1	6				

Collect monolingual texts and generate parallel corpora via MT for unfilled categories – all 195 corpora have been filled

Where necessary parallel corpora are generated using MT. This includes the following stages:

(1) Automatic sentence segmentation of the source language (SL)

(2) Selection of appropriate sentences for translation

(3) Obtaining MT from the source

(4) Evaluating and selecting parallel pairs

Different approaches have been examined for selection of suitable sentences for MT considering length (80-800 symbols), punctuation (end of sentence) and contents (at least 70% letters)

MT systems used are:

(1) Google Translate. Free access with extended daily limit provided via University Research Program.

(2) WebTrans by SkyCode. Free access was provided by SkyCode.

(3) Bing (Microsoft Translator). Limited access – 2 million symbols per month.

8. MyMemory (translation memory). Free access.

Several approaches are exploited for the selection of proper parallel sentences after MT: full or partial (>50% or >70%) overlap of translations from two MT systems – based either on string, tokens or lemmas; BLEU between translations from two MT systems (>0.4 or >0.7); BLEU score measured between the SL text and back translation of MT.

T 6.4. Adaptation of existent cross-lingual search engine

Ongoing work.

In the Months 24-30 work concentrated on:

- Final development of the search engine and adaptation to the ATLAS platform respectively:
  1. development of the installer package
  2. development of an individual Demo-client within the development framework
  3. File-based processing within the Server
  4. Error handling
  5. Error correction in the main(basic) System
  6. Translating indexing techniques from Lucene into Solar

T 6.5. Integration of cross-lingual search engine into ATLAS

Ongoing work; first preliminary discussions took place

T 6.6 Documentation

On-going work.

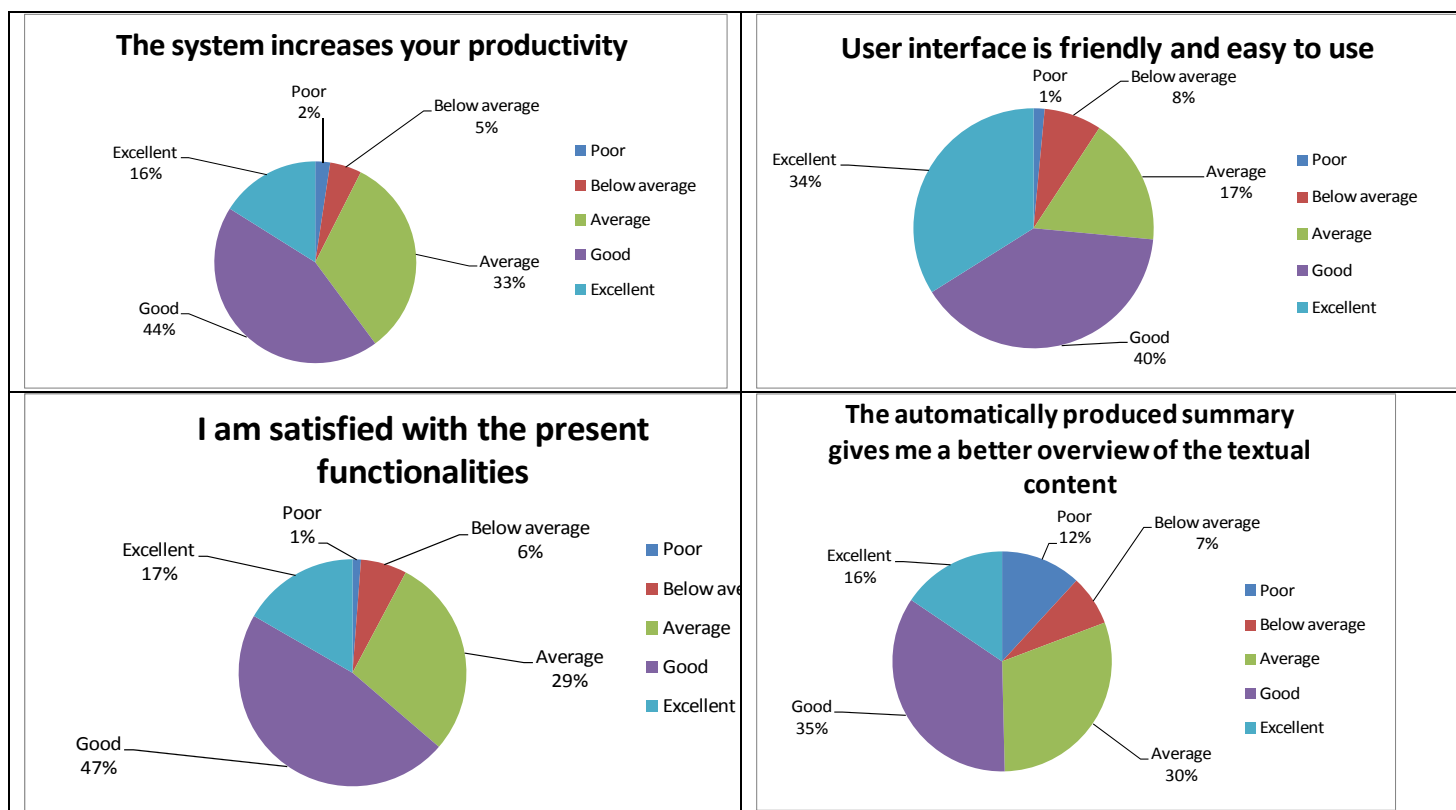**Work package 7 - Testing and User Evaluation**

The focus of WP7 is on ensuring the acceptable quality of the final ATLAS system, by assessing the technical performance and the level of fulfillment of the system specifications, as well as assessing the level of user expectations fulfillment. The WP includes 2 tasks; the progress in each one of them during the reporting period is the following:
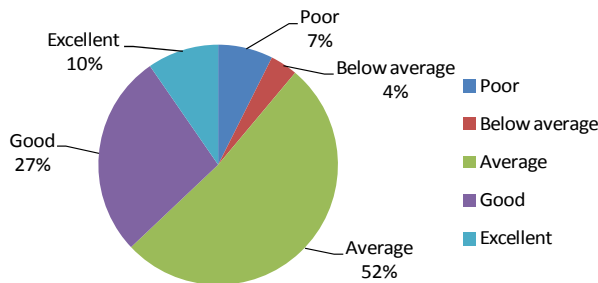

T7.1: User Evaluation and feedback adjustments

The Task is ongoing. Activities and results within the period:

- The 2nd round of the ATLAS system User Acceptance Evaluation has been conducted (following the methodology described in the updated User Evaluation Plan):
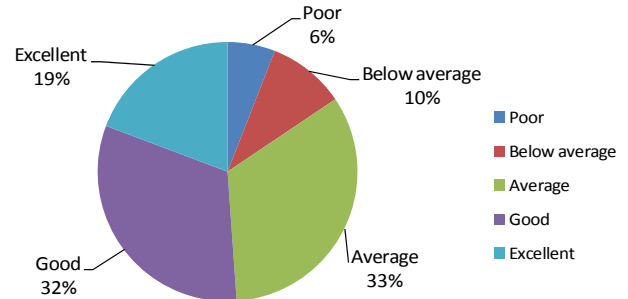
- o 134 unique users (based on IP) participated from 5 countries.
- o The overall user impression is positive. For the majority of indicators the MEAN VALUE is above average, however there are indicators with rather low MEAN VALUE ("I am satisfied I found relevant documents in a language I do not use", "The automatically produced summary gives me a better overview of the textual content", "I am satisfied I found similar to mine documents in the library", "How helpful was the automatic production of summaries?").
- o Some early conclusions show that:
  - ▪ ATLAS is user-friendly and increases the productivity of users.
  - ▪ Online help and tooltips could be improved (MEAN VALUE above "average", but below "good").
  - ▪ The value offered by automatic categorization, machine translation and summarization to the users should be improved.
  - ▪ In a rather complex scenario, as the 3rd one (i.e. the 2nd i-Publisher scenario which has 4 steps and 1-step exercise), only 35% of the users managed to complete 4 or 5 steps.
  - ▪ The vast majority of users managed to add a new book to i-Librarian in less than 3 mins, and create their own simple WEB Site in less than 20 mins.

The system increases your productivity

Poor 2%
Below average 5%
Excellent 16%
Average 33%
Good 44%

Poor
Below average
Average
Good
Excellent

User interface is friendly and easy to use

Poor 1%
Below average 8%
Excellent 34%
Average 17%
Good 40%

Poor
Below average
Average
Good
Excellent

I am satisfied with the present functionalities

Poor 1%
Below average 6%
Excellent 17%
Average 29%
Good 47%

Poor
Below ave
Average
Good
Excellent

The automatically produced summary gives me a better overview of the textual content

Poor 12%
Below average 7%
Excellent 16%
Good 35%
Average 30%

Poor
Below average
Average
Good
Excellent

I am satisfied I found relevant documents in a language I do not use

- Poor 7%
- Below average 4%
- Average 52%
- Good 27%
- Excellent 10%

Legend:
- Poor
- Below average
- Average
- Good
- Excellent

I am satisfied I found similar to mine documents in the library

- Poor 6%
- Below average 10%
- Average 33%
- Good 32%
- Excellent 19%

Legend:
- Poor
- Below average
- Average
- Good
- Excellent

Time needed by user to add new book in iLib

- 0-3 min 82%
- 4-7 min 10%
- 8-11 min 5%
- >11 min 3%

Legend:
- 0-3 min
- 4-7 min
- 8-11 min
- >11 min
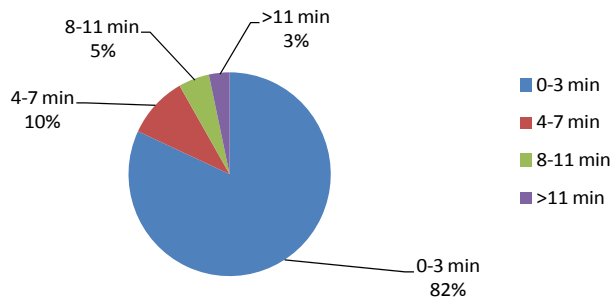
Time needed by user to setup simple web site

- 0-20 min 76%
- 20-40 min 17%
- 40-60 min 3%
- >60 min 4%

Legend:
- 0-20 min
- 20-40 min
- 40-60 min
- >60 min
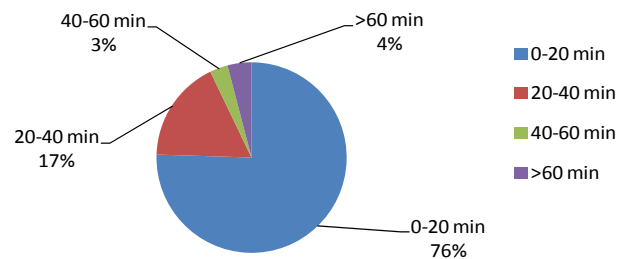
Steps completed successfully by users in 4stps-scenario & 1step exercise (#3 iPubl)

- 1step 8%
- 2steps 24%
- 3steps 33%
- 4steps 20%
- 5steps 15%

Legend:
- 1step
- 2steps
- 3steps
- 4steps
- 5steps
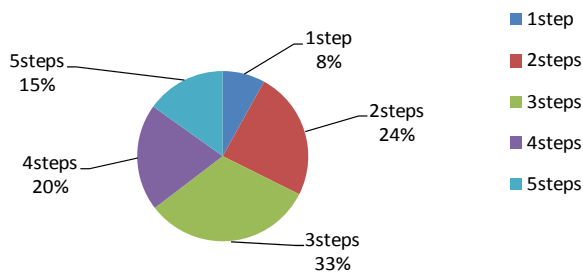
2011-2012 comparison

2011 (1st rnd)   2012 (2nd rnd)
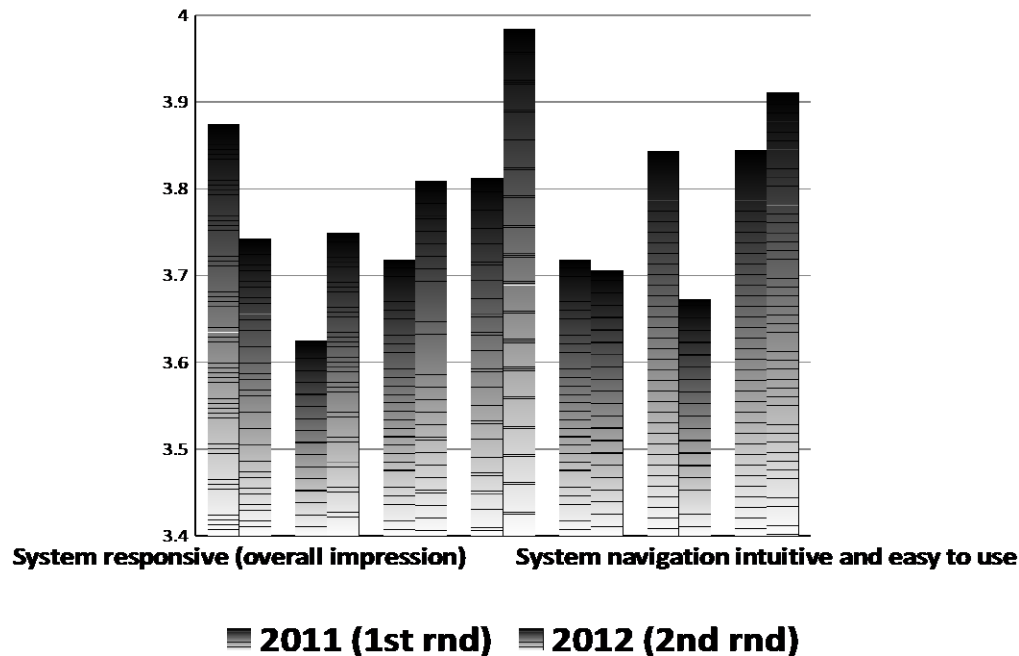
- Living Lab related activities:
  - Testing, evaluation and feedback workshops were organized, with the following conclusions:
    - ATLAS is user-friendly and increases the productivity of users.
    - The System produces adequate web sites with multilingual content.
    - The websites created are in line with user expectations.
    - The system offers most of the needed functions.
    - The system should be improved in order to offer comprehensive well grouped documents.
    - The Automatic document classification could be improved.
    - The system could be improved in order to increase even more the users productivity.
    - The User Interface of the system should be improved.
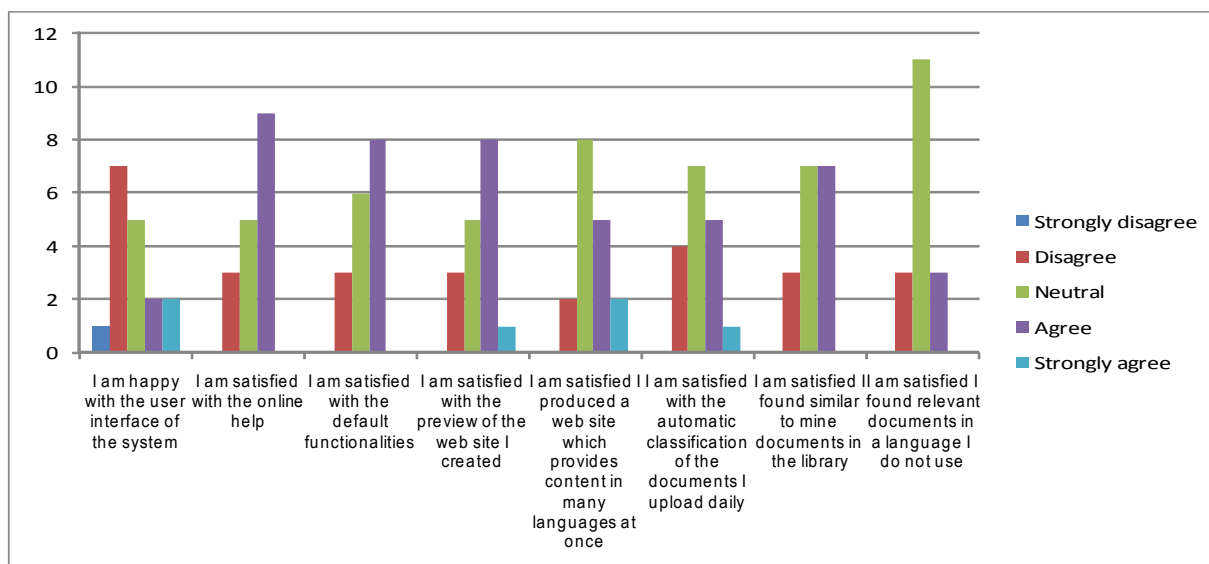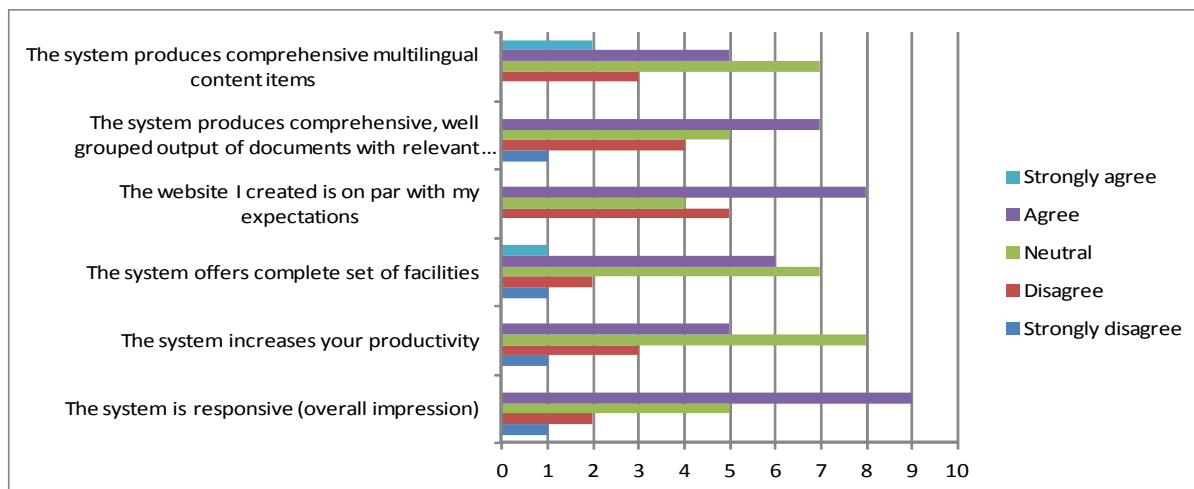
T7.2: System Testing (component, integrated system)

The Task is ongoing. Activities and results within the period:

- Technical Indicators have been finalised for the following system modules.
- CMS indicators (single request / single request in environment of 100 concurrent):
  - CMS_SWDi: time needed to render a static widget i (i = 1..3)
  - CMS_DWDi: time needed to render a dynamic widget i (i = 1..3)
    Need to predefine length / magnitude of associated dynamic content (e.g. a list with 20 entries)
  - CMS_SPi: time needed to render a sample page i (i = 1..3)
  - CMS_SQ: average time and Java heap usage and limits for rendering a search query within a poll of 1.000, 100.000 and 1.000.000 content items.
  - CMS_DB: time needed to render a database response of 100 records.
  - CMS_LG: time needed to respond to a user login request.
- Automatic classification indicators:

Let L be a set of labels and D a multi-label evaluation data set, consisting of |D| multi-label examples (xi, Yi), i=1..|D|, Yi⊆L . If H is a multi-level classifier and Zi=H(xi) is the set of labels predicted by H for the xi example, then:

- o CLS_PREC: the sum of correct results divided by the number of returned results for all evaluation documents:

$$\frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|Y_i \cap Z_i|}{|Z_i|}$$

- o CLS_RECL: the sum of correct results divided by the number of results that should have been returned for all evaluation documents:

$$\frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|Y_i \cap Z_i|}{|Y_i|}$$

- Machine Translation indicators:
  - o MT_WDi: % of non-translated words for each project language pair (i=1..lang_pairs).
  - o MT_BLEU: use widely accepted metrics such as BLEU score, TER, etc and compare against reference values for each project language pair.
- Cross-lingual Content Retrieval indicators (result set of 100 items):
  - o CLR_INS: time needed to process and index a single 10K text document insert.
  - o CLR_BINS: time needed to process and index a batch document (i.e. 10 text documents of 10K each) insert.
  - o CLR_EXEC2: time needed to render the results of a search, using a query with 2 input terms.
  - o CLR_EXEC4: time needed to render the results of a search, using a query with 4 input terms.
- Summarisation indicators:
  - o SUM_RESP: response time to provide a summary of a short text (length to be decided in terms of number of sentences or clauses) in each of the project languages.
  - o SUMP_PR: precision / recall rate, based on comparisons between auto generated summaries and gold summaries (i.e. summaries produced by humans). Three (3) human summarisers will be engaged, with 3 consecutive length reduction rates: 20% → 10% → 5%

Work package 8 - Dissemination and Exploitation

During the reporting period, the main objectives in Work package 8 "Dissemination and Exploitation" are:

- Organisation of National events in each partner country focused on presenting the ATLAS services, their real benefits to the different target users and exploring possible ways of their exploitation and collaboration.

- Participation of ATLAS project in big exhibitions and conferences
    - Cebit 2012
    - EACL 2012
    - LREC 2012
    - META-FORUM 2012
    - CESAR META-NET Roadshow
- Demonstration and presentation of ATLAS platform and services at big exhibitions and conferences in order to promote the ATLAS platform and its services among different target users groups and public.
- Disseminate the research results gained within the project and ATLAS benefits through:
    - Papers at International conferences;
    - Publications in journals and books;
    - Organisations of special workshops in national and international level;
    - Poster presentations;
    - Demo presentations.

Task 8.2: Detailed Dissemination Plan

During next project period, ATLAS partners plan to continue ATLAS services` presentations, demos and scientific papers at national and international events (conferences and exhibitions).

Dissemination Plan was updated for the next project period. Dissemination activities of the consortium partners will be focused on:

- Organization of events targeted at ATLAS potential users.
- Presentations of ATLAS project results and services at national and international events

Organization of events targeted at ATLAS potential users:
Each partner will organize special event at national level, targeted at potential users and customers. The aim of these events is to present the ATLAS services, their real benefits to the different target groups and exploring possible ways of their exploitation and collaboration. During the events, representatives of the ATLAS target users will be invited.

- ICS-PAS plans to organize ATLAS Road Show in Warsaw, Poland in February 2013. Project demonstration targeted to potential customers and users of the ATLAS platform and language processing chains will be made to the publishers and publishing houses.
- DCL – IBL plans to organize 2 events in Sofia, Bulgaria. Demonstration of the ATLAS services and technologies will be made at both events in October 2012 for publishers, publishing houses and in January 2013 for electronic media.
- ATLANTIS plans to organize workshop, titled "Multilingual e-libraries, content management and e-publishing" in Athens, Greece in February 2013. Representatives of universities, research centers, libraries, museums, and publishing houses and

- research projects with similar topics, in order to achieve a multiplier effect in terms of participation will be invited.
- ITD plans to organize special workshop at European day of entrepreneur – 2012 in Sofia during October 2012. ATLAS workshop at EDE is organized every year for demonstration of ATLAS platform benefits to the target audience. Representatives of university libraries, municipal libraries, and media experts will be invited to participate.
- UAIC plans to organize ATLAS Road Show in Iasi/Bucharest, Romania during February 2013 for publishers, publishing Houses, libraries. UAIC will make demonstration of the ATLAS platform to potential users, will show language-processing chains applications dedicated to publishing houses, will explore ways for business or scientific partnerships.
- In November 2012, UHH plans to organize an event in Hamburg. ATLAS platform will be demonstrated and language processing chains technologies will be presented to the librarians and scientific community (linguists, computational linguists).

Presentations of ATLAS project results and services at national and international events (exhibitions, synergy with other projects, distributions, etc.)

Event: Human Language Technology Days 2012 in Warsaw, Poland on 27-28 September 2012, Event related to CESAR ICT PSP project
Target users: Linguists, computational linguists, IT experts
Expected number: 150-200 participants
Dissemination activity: ATLAS exhibition booth.
Project partner: ICS-PAS

Event: CESAR Road Show in Serbia, Hungary, Croatia in October-November 2012, Event related to CESAR ICT PSP project
Target users: Scientific and business participants
Expected number: Hundreds of participants
Dissemination activity: Project exhibition and distribution of project information at the exhibition.
Project partner: ICS-PAS

Event: International Conference: Formal Approaches to South Slavic an Balkan Languages in Dubrovnik Conference, September 2012
Target users: Linguists
Dissemination activity: Paper at a conference
Project partner: DCL – IBL

Event: Workshop: UAIC days, Iasi, Romania in October, 2012
Target users: IT professors and PhD students
Expected number: 20+
Dissemination activity: Presentation of Eugen Ignat, Daniel Anechitei, Dan Cristea: The multilingual summariser in ATLAS
Project partner: UAIC

Task 8.3: Sustainability plan preparation
Sustainability Plan was updated. NGO will not be established.

Task 8.4: Preparation of primary outreach materials
Preparation of Deliverable 8.2 Dissemination kit, including produced dissemination materials (brochures, posters, flyers, articles, presentations at exhibitions and fairs)

Brochures, and other dissemination materials is available at: http://www.atlasproject.eu/atlas/project/dissemination/en

Task 8.5: Dissemination of primary materials
The dissemination activities, performed by the partners during the reporting period, include:

1. Demonstration and presentation of Atlas services at conferences and exhibitions among potential users.

   - Cebit 2012, 06- 10 March 2012, Hannover, Germany

Atlas presentation: http://www.cebit.de/product/atlas-build-your-website-at-no-cost/291713/C913812
Asset presentation: http://www.cebit.de/product/asset-adds-value-to-your-content/291715/C913812

   - EACL 2012, 23 - 27 April 2012, Avignon, France

Demonstration and presentation of NLP processing in ATLAS: http://eacl2012.org/system-demonstration/index.html

Poster presentation "Harnessing NLP Techniques in the Processes of Multilingual Content Management": http://www.atlasproject.eu/atlas/file/748ce0e5-cc64-4fdb-b30d-cd595cf2eb02/13ccaaa0-a0c0-11e0-8264-0800200c9a66/atlas_eacl2012_a0.pdf

   - LREC 2012, 21-27 May, Istanbul, Turkey

Paper and poster presentations: http://www.lrec-conf.org/lrec2012/?List-of-accepted-papers

   - CESAR META-NET Roadshow 2012 Sofia, Bulgaria May 2nd 2012

Programme of the events:http://dcl.bas.bg/en/cesar_roadshow_program_en.html

Presentation of the ATLAS project made by TETRACOM: Widening the Scope. Harnessing NLP Techniques in Content Management Systems

   - EAMT 2012, 28-30 May 2012 Trento

Poster presenting MT Service in ATLAS at EAMT 2012

   - Conference "Crosslingual language technology in service of a multilingual integrated Europe" 4-5 May, University of Hamburg, http://www.c-phil.uni-hamburg.de/view/Main/MaiProgramme

2. Papers at International conferences:

- Cristina Vertan, Walther v. Hahn, Monica Gavrila, Same domain different discourse style-A case study on Language Resources for data-driven Machine Translation, presented at LREC 2012, 21-27 May, Istanbul

- Cristina Vertan, "Two approaches for integrating translation and retrieval in real applications" at EACL Workshop 2012, Avignon France, 23-27 April 2012

- Cristina Vertan, Embedding Machine Translation in ATLAS Content Management System at the EAMT 2012 , Trento, 28-30 May 2012

- Anelia Belogay Diman Karagyozov Svetla Koeva Cristina Vertan Adam Przepiórkowski Maciej Ogrodniczuk Dan Cristea Eugen Ignat , Polivios Raxis, Merging heterogeneous resources and tools in a digital library, in Proceedings of the Workshop "Merging Language Resources", at LREC 2012

- Anelia Belogay, Diman Karagyozov, Svetla Koeva, Cristina Vertan, Adam Przepiórkowski, Dan Cristea, Polivios Raxis (2012) Harnessing NLP Techniques, in Walter Daelemans, Mirella Lapata Lluis Marquez (Eds.), Processes of Multilingual Content Management, Proceedings of EACL 2012 – the 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France, April 23-27, pages 6-10, ISBN: 978-1-937284-19-0.

- Polish Language Processing Chains for Multilingual Information Systems by Maciej Ogrodniczuk and Adam Przepiórkowski, LNCS. See http://www.nldb.org/

3. Publications in journals:

- i-Librarian – Free online library for European citizens by Diman Karagiozov, Anelia Belogay, Dan Cristea, Maciej Ogrodniczuk, Polivios Raxis, Emil Stoyanov and Cristina Vertan. INFOtheca: Journal of Informatics and Librarianship 13(1), 2012, pp. 27-42, ISSN 1450-9687. See http://infoteka.bg.ac.rs/.

- Dan Cristea, Eugen Ignat, Daniel Anechitei: ATLAS Project - the Romanian Component, printed in the Bring IT on! 2012 Catalog, "Alexandru Ion Cuza" University Editing House, ISSN 2285-0929, pages 26-27.

4. Publications in Books:

- Book: Cristina Vertan, Walther v. Hahn, Multilingual Processing in Eastern and Southern EU Languages.- Less - resourced Technologies and Translation – Cambridge Scholar Press, 2012

5. Talks:
- Cristina Vertan, Machine Translation in a content management system – a multilingual case study involving Polish, Bulgarian, Romanian, and Greek, Hamburg ,4 May 2012, http://www.c-phil.uni-hamburg.de/view/Main/MaiProgramme
- Cristina Vertan, „Cross lingual retrieval in a content Management System", University of Granada, May 2012

- Presentation of ATLAS System, University of Bucharest, March 2012

Task 8.6: Interaction with target communities

Demonstrations and promotion of ATLAS services and results were made among different target users:

- Linguists, computational linguists, NLP engineers, IT scientists, key players in Language Resources and Technologies and representatives of social sciences and humanities (in many conferences, workshops and special demo sessions);

- Industry, IT community and policy makers

Task 8.8: Nationwide project promotion

ATLAS project was promoted and presented among different users on national level. The nationwide ATLAS promotion activities include organization of national workshops and presentations at national events:

- Cristina Vertan, Organisation of Conference "Cross lingual Language Technology in the service of a multilingual Europe -20 years on-" Hamburg 4-5 May 2012 (UHH) http://www.c-phil.uni-hamburg.de/view/Main/MaiProgramme

- Joint ATLAS/ MLeCel LL/DSLL workshop "Multilingual Digital Repositories and Services" , Sofia, 3-5 May 2012. (ITD, TETRACOM)

Participants: Professionals in developing Multilingual Digital Repositories and Services. The aim of the Workshop is to bring together the developers and end users of Multilingual digital repositories and services as well as to demonstrated the ATLAS services in order to show real benefits and explore potential exploitations. The participants presented achievements in the area and discussed opportunities for further collaborations.

Presentations from the workshops are available at: http://livinglab.itd-bg.eu/content/events

- Presentation of the modules: RARE (anaphora resolution), clause segmentation and discourse parser, by Eugen Ignat and Daniel Anechitei at NLP Seminars on Computational Linguistics, Faculty of Computer Science, "Al. I. Cuza" University of Iasi in front of Master students in Computational Linguistics

- Presentation of Dan Cristea, Eugen Ignat şi Daniel Anechitei: Proiectul ATLAS (in Romanian) at The 8th Conference on Resources and Tools for Processing Romanian Language - ConsILR-2012, 27 April, 2012 Bucharest, Romania

Task 8.9: Trans-European project promotion

The dissemination activities performed by partners and promoting ATLAS project in European and international level include:

Demonstrations and presentations at International exhibitions and conferences:

1. Cebit 2012, 06- 10 March 2012, Hannover, Germany

Type of the event: Trade show
Target users: international ICT industry

Participants: 1000 +

Performed dissemination activities:

ATLAS project and software platform were present and demonstrated @CeBIT 2012 in Hanover, which was held this year from 6 to 10 March, Hall 2, stand 58. All visitors were able to sample the content management platform i-Publisher and Linguistic Platform ASSET, which processes and analyses text. Reference URL-s:

Atlas: http://www.cebit.de/product/atlas-build-your-website-at-no-cost/291713/C913812

Asset: http://www.cebit.de/product/asset-adds-value-to-your-content/291715/C913812

News and pictures of the event is published on: http://www.atlasproject.eu/atlas/project/dissemination/en

2. EACL 2012, 23 - 27 April 2012, Avignon, France

Type of the event: International conference and exhibition

Target users: Linguists, computational linguists, IT experts

Participants: 300 +

Performed dissemination activities:

- Demonstration and presentation of NLP processing in ATLAS (http://eacl2012.org/system-demonstration/index.html)

ATLAS was demonstrated in the most prestigious linguistic event for 2012 in Europe – the annual conference of the European chapter of the Association for Computational Linguistics on 25.04.2012 in Avignon , France. The demonstration aims to prove that people reading websites powered by our multilingual web management platform can easily find documents, kept in order via the automatic classification, find context-sensitive content, find similar documents in a massive multilingual data collection, and get short summaries in different languages that help the users to discern essential information with unparalleled clarity.

- Poster presentation

The theme of ATLAS demonstration is "Harnessing NLP Techniques in the Processes of Multilingual Content Management", available at http://www.atlasproject.eu/atlas/file/748ce0e5-cc64-4fdb-b30d-cd595cf2eb02/13ccaaa0-a0c0-11e0-8264-0800200c9a66/atlas_eacl2012_a0.pdf

- Paper in Processes of Multilingual Content Management, Proceedings of EACL 2012

Anelia Belogay, Diman Karagyozov, Svetla Koeva, Cristina Vertan, Adam Przepiórkowski, Dan Cristea, Polivios Raxis (2012) Harnessing NLP Techniques, in Walter Daelemans, Mirella Lapata Lluis Marquez (Eds.), Processes of Multilingual Content Management, Proceedings of EACL 2012 – the 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France, April 23-27, pages 6-10, ISBN: 978-1-937284-19-0.

http://www.atlasproject.eu/atlas/file/3c952233-5129-44b0-9ed8-257d98cf0bec/13ccaaa0-a0c0-11e0-8264-0800200c9a66/eacl2012_2.pdf

- Paper at EACL Workshop 2012

Cristina Vertan, "Two approaches for integrating translation and retrieval in real applications" at EACL Workshop 2012, Avignon France, 23-27 April 2012, http://www-lium.univ-lemans.fr/esirmt-hytra/programme.php

- Invited presentation

prof. Svetla Koeva (IBL DCL) had an invited presentation on the topic: "Exploiting Synergies between Information Retrieval and Machine Translation and Hybrid Approaches to Machine Translation" at EACL 2012.

3. LREC 2012, 21-27 May, Istanbul, Turkey

Type of the event: International conference and exhibition
Target users: Computational linguists, industry people, mass media
Participants: 800 +
Performed dissemination activities:
- Paper presentation at LREC 2012

Cristina Vertan, Walther v. Hahn, Monica Gavrila, Same domain different discourse style-A case study on Language Resources for data-driven Machine Translation, presented at LREC 2012, 21-27 May, Istanbul, http://www.lrec-conf.org/lrec2012/?List-of-accepted-papers

- Paper at LREC 2012 Workshop and poster presentation

Anelia Belogay Diman Karagyozov Svetla Koeva Cristina Vertan Adam Przepiórkowski Maciej Ogrodniczuk Dan Cristea Eugen Ignat , Polivios Raxis, Merging heterogeneous resources and tools in a digital library, appeared in Proceedings of the "Merging LR" workshop at 8th International Conference on Language Resources and Evaluation - LREC 2012, ELRA, Istanbul, Turkey, pp. 41-44.

- Workshop Organization "Adaptation of Language Resources and Tools for Cultural Heritage" at LREC 2012 (Cristina Vertan)

4. META-FORUM 2012, 19-20 June 2012, Brussels, Belgium

Type of the event: International conference and exhibition
Target users: Scientific and business participants
Participants: 200 +
Performed dissemination activities:
Participation in the poster session, distribution of ATLAS leaflets.
Synergies with CESAR ICT PSP project.

5. CESAR META-NET Roadshow 2012 Sofia, Bulgaria May 2nd 2012

Type of the event: National event
The meeting is organized by CESAR, part of META-HET (Multilingual Europe Technology Alliance), a Network of Excellence consisting of 57 research centres in 33 countries and funded by the European Commission. The four projects within the META-NET – T4ME, CESAR, META-NET4U and META-NORD – will provide access to many Language Resources and Language Tools for many European languages. The event in Sofia is the first in a series of official presentations of CESAR in a number of European countries.
Target users: representatives of research centres, small and large technology corporations, translation services and other users or producers of Language Technology, language

communities and societies, and policy makers responsible for supporting research and innovation, economy and ICT.

Participants: 80 +

Performed dissemination activities:

Synergies with CESAR ICT PSP project.

Presentation of the ATLAS project made by TETRACOM: Widening the Scope. Harnessing NLP Techniques in Content Management Systems

Programme of the events:http://dcl.bas.bg/en/cesar_roadshow_program_en.html

6. UNESCO IITE and UNITWIN/UNESCO Chairs International Conference "UNESCO Chairs Partnership on ICTs use in Education", 5 – 10 September 2012 St.-Petersburg, Russian Federation

Type of the event: International event

Target users: Librarians, researchers, Policy makers

Performed dissemination activities:

Presentation of ATLAS project in the power point presentation titled: OPEN EDUCATION AND OPEN INNOVATION IN A GLOBAL LEARNING ENVIRONMENT, Prof. Roumen Nikolov, Chairholder, UNESCO Interfaculty Chair on ICT in Library Studies, Education and Cultural Heritage

Complete list of the dissemination activities performed by partners during the reporting period is presented in Annex 1.

Annex 1: Report of Dissemination Activities for the period: 01.03.2012-31.08.2012

Performed dissemination activities for the period: 01.03.2012-31.08.2012 by all project partners within the ATLAS project:

Type of Dissemination Activity	Title	Dates and Place	Participant Name and Institution	Performed Dissemination Actions	Available Dissemination Materials	Target Groups and Number of People Reached
Trade Show	Cebit 2012	06- 10 March 2012 Hannover, Germany	Tetracom, Bulgaria	Demonstrations and presentations of ATLAS and Asset, Hall 2, stand 58	Atlas: http://www.cebit.de/product/atlas-build-your-website-at-no-cost/291713/C913812 Asset: http://www.cebit.de/product/asset-adds-value-to-your-content/291715/C913812	international ICT industry No of people reached: 1000 +
International conference and exhibition	EACL 2012	23 - 27 April 2012 Avignon, France	TETRACOM ICS PAS UHH DCL – IBL ATLANTIS UAIC	Demonstration and presentation of NLP processing in ATLAS Poster: "Harnessing NLP Techniques in the Processes of Multilingual Content Management", available at http://www.atlasproject.eu/atlas/file/748ce0e5-cc64-4fdb-b30d-cd595cf2eb02/13ccaaa0-a0c0-11e0-8264-0800200c9a66/atlas_eacl2012_a0.pdf Paper	Paper: Anelia Belogay, Diman Karagyozov, Svetla Koeva, Cristina Vertan, Adam Przepiórkowski, Dan Cristea, Polivios Raxis (2012) Harnessing NLP Techniques, in Walter Daelemans, Mirella Lapata Lluis Marquez (Eds.), Processes of Multilingual Content Management, Proceedings of EACL 2012 – the 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France, April 23-27, pages 6-10, ISBN: 978-1-937284-19-0. http://www.atlasproject.eu/atlas/file/3c952233-5129-44b0-9ed8-257d98cf0bec/13ccaaa0-a0c0-11e0-8264-0800200c9a66/eacl2012_2.pdf	Linguists, computational linguists, IT experts No of people reached: 300 +

International conference and exhibition	Merging LR Workshop at the 8th international conference on Language Resources and Evaluation (LREC 2012)	22 May Istanbul, Turkey	TETRACOM ICS PAS UHH DCL – IBL ATLANTIS UAIC	Paper at LREC 2012 Workshop and poster presentation	Anelia Belogay Diman Karagyozov Svetla Koeva Cristina Vertan Adam Przepiórkowski Maciej Ogrodniczuk Dan Cristea Eugen Ignat , Polivios Raxis, Merging heterogeneous resources and tools in a digital library, appeared in Proceedings of the "Merging LR" workshop at 8th International Conference on Language Resources and Evaluation - LREC 2012, ELRA, Istanbul, Turkey, pp. 41-44.	Computational linguists, industry people, mass media No of people reached: 300 +
Article in Journal	Infotheca Open Access journal	June, 2012 Belgrade, Serbia	TETRACOM ICS PAS UHH ATLANTIS UAIC	Journal Article	i-Librarian – Free online library for European citizens by Diman Karagiozov, Anelia Belogay, Dan Cristea, Maciej Ogrodniczuk, Polivios Raxis, Emil Stoyanov and Cristina Vertan. INFOtheca: Journal of Informatics and Librarianship 13(1), 2012, pp. 27-42, ISSN 1450-9687. See http://infoteka.bg.ac.rs/.	Linguists, computational linguists, Librarians No of people reached:
Organized special event	CESAR META-NET Roadshow 2012	Sofia, Bulgaria May 2nd 2012	TETRACOM DCL – IBL	Presentation of ATLAS project	Presentation of the ATLAS project made by TETRACOM: Widening the Scope. Harnessing NLP Techniques in Content Management Systems Presentation available on: http://dcl.bas.bg/en/cesar_roadshow_program_en.html	representatives of research centres, small and large technology corporations, translation services and other users or producers of Language Technology, language

						communities and societies, and policy makers responsible for supporting research and innovation, economy and ICT No of people reached: 80+
Pan-European Forum	2nd Pan European Private Equity and Startups Forum	21-22 June 2012, Athens, Greece	ATLANTIS	Discussions on funding opportunities for new online services based on specific ATLAS results (i-Publisher, automatic multilingual summarization)	ATLAS brochure and custom flyer	More than 50 executives from Business Angels, Venture Capitals, Banks, Startups, including also entrepreneurs and researchers
Project exhibition	META-FORUM 2012	19-20 June 2012 Brussels, Belgium	ICS PAS, Poland UAIC, Romania	Event related to CESAR ICT PSP project	Participation in the poster session, distribution of ATLAS leaflets.	Scientific and business participants *No of people reached: 200*
Conference	17th International conference on Applications of Natural Language Processing to Information Systems (NLDB 2012)	26-28 June 2012 Groningen, The Netherlands	ICS PAS, Poland	Conference article	Polish Language Processing Chains for Multilingual Information Systems by Maciej Ogrodniczuk and Adam Przepiórkowski. To be published in LNCS. See http://www.nldb.org/.	Scientific and business participants *No of people reached: 100*
Conference	The 8th Conference on Resources and Tools	27 April, 2012 Bucharest,	UAIC, Romania	Oral presentation	Dan Cristea, Eugen Ignat şi Daniel Anechitei: Proiectul ATLAS (in Romanian)	Linguists, computational

	for Processing Romanian Language - ConsILR-2012	Romania				linguists, IT experts No of people reached: 50
Series of seminars	NLP Seminars on Computational Linguistics, Faculty of Computer Science, "Al. I. Cuza" University of Iasi	3 May 2012 and 10 May 2012 Iasi, Romania	UAIC, Romania	Presentation of WP5-related tasks	Presentation of the modules: RARE (anaphora resolution), clause segmentation and discourse parser, by Eugen Ignat and Daniel Anechitei	Master students in Computational Linguistics No of people reached: 15-20
Conference	Crosslingual Language Technology 2012	7-8 May 2012 Univ. Hamburg, Germany	UAIC, Romania	Power Point presentation	Dan Cristea and Ionut Pistol: Multilingual linguistic workflows. A case study for two languages	Linguists, computational linguists, IT experts No of people reached: 40
Workshop for prospecting the links between university research and industry	Bring-IT on! 2012	17-18 May 2012 Iasi, Romania	UAIC, Romania	Poster and article in Bring IT on! 2012 Catalog	Dan Cristea, Eugen Ignat, Daniel Anechitei: ATLAS Project - the Romanian Component; Also printed in the Bring IT on! 2012 Catalog, "Alexandru Ioan Cuza " University Editing House, ISSN 2285-0929, pages 26-27.	Computational linguists, industry people, mass media No of people reached: 80-100
Conference	EACL 2012	23-27 April, 2012 Avignon, France	UHH, Germany	Paper	Cristina Vertan, "Two approaches for integrating translation and retrieval in real applications", submitted at EACL Workshop 2012, Avignon France, 23-27 April 2012	Linguists, computational linguists, IT experts
Conference	LREC 2012	21-27 May 2012	UHH,	Paper and workshop	Cristina Vertan, Walther v. Hahn, Monica	Computational

		Istanbul, Turkey	Germany	organization	Gavrila, Same domain different discourse style-A case study on Language Resources for data-driven Machine Translation, presented at LREC 2012, 21-27 May, Istanbul	linguists, industry people, mass media
Conference	EAMT (Summit of the European Association for MAchine Translation)	28-30 May 2012 Trento	UHH, Germany	Paper	Cristina Vertan, Embedding Machine Translation in ATLAS Content Management System at the EAMT 2012 , Trento, 28-30 May 2012	Linguists, computational linguists, IT experts
Conference	Crosslingual Language Technology for an integrated multilingual Europe	4-5 May 2012 Hamburg	UHH, Germany	An event partially organized by ATLAS partners Talk of Dr. Cristina Vertan	Cristina Vertan, Machine Translation in a content management system – a multilingual case study involving Polish, Bulgarian, Romanian, and Greek, Hamburg ,4 May 2012	Linguists, computational linguists, IT experts
Conference	UNESCO IITE and UNITWIN/UNESCO Chairs International Conference "UNESCO Chairs Partnership on ICTs use in Education	5 – 10 September 2012 St.-Petersburg, Russian Federation	ITDF	Presentation	Presentation of ATLAS project in the power point presentation titled: OPEN EDUCATION AND OPEN INNOVATION IN A GLOBAL LEARNING ENVIRONMENT, Prof. Roumen Nikolov, Chairholder, UNESCO Interfaculty Chair on ICT in Library Studies, Education and Cultural Heritage	Librarians, researchers, Policy makers

Deliverables and milestones tables

N	Name	WP	Lead participant	Nature	Dissemination level	Due delivery date from Annex I	Delivered Yes/No	Actual / Forecast delivery date	Comments
D 1.3	Second year progress monitoring report	1	1	R	R	M 26	yes	M 26	-
D3.1	Categorisation tool	1	1	R	PU	M27	yes	M28	
D 5.1	Summarisation Tool	5	7	R	PU	M30	yes	M 31	
D 6.1	Machine Translation Tools	6	6	R	PU	M30	no		Delay due to additional implementation
D 6.2	Cross-lingual Search Engine	6	6	R	PU	M 28	yes	M 29	

Table "Deliverables"

Project management

The Consortium established a structure for communication with the EC and the beneficiaries regarding contractual and financial issues, consortium communication, coordination and support of work package leaders, as well as preparation of reports for the EC. In addition, the activities within this package cover regular assessment of project progress and the organization of project meetings.

Deviation from work

- New Regression test module
- Tetracom involvement in implementation and integration of the Summarisation tools
- Classification voting system, new classification algorithm
- Redistribution of resources of Atlantis (see table 3.3)
- Redistributions of resources of DCL- IBL in WP 3, 5, WP 6 and WP 7.
- Reported possible delay of D 7. 2 and D 7.3 at the review meeting .
- Reported delay of D 6.1 at the review meeting
- Redistribution of resources for ICS PAS reported and agreed in the report October 2011

Work package 1– Project Management

The tasks included in the first working package, together with their status for the reported period, are listed below.

T1.3 – Technical communication

Technical communication is an ongoing task.

T1.4 – Project coordination and management

The project management activities took place from M24 to M30. A flow of information is organized as follows:

- management mailing list. The Project CEO and the partners staff executives discuss management topics in the Management mailing list;
- internal reports. An internal reporting system was established and a template for an internal report was approved. It is planned for the internal reports to be issued every three months. However 3rd PM internal report was inapplicable due to the lack of significant activities. The first year report was submitted to the European Commission.
- collaboration area. A project web site internal area is set to enable the partners to work on tasks and topics in protected environment. More details are given in the section Project web site.
- discussion forum. The discussion forum is part of the collaboration area and it is a place for the discussion of technical issues.

T1.5 – Project meetings

One project meeting was held during the reported period. The meeting was in Luxembourg and the first part focused on the preparation for the yearly review. The second part of the meeting was dedicated to the upcoming tasks within WP 5 and WP 6 and the organization of the post review implementation.

T1.6 – Work on contractual deliverables

Resource Monitoring

- The actual state of financial expenditures is constantly monitored through reports, prepared by beneficiaries every three months.
- Quality control and work plan monitoring

The coordinator monitors the project activities, ensuring that they lead to the required deliverables and keep up with the project program. No delays were reported or experienced for the reported period.

- Reporting to the European Commission

The second year progress report was delivered by the Coordinator to the European Commission.

The project planning is kept updated.

Use of resources

Legend: A= actually spent person months for the period 01.03.2012- 31.08.2012, P = Planned for the period 01.03.2012-31.08.2012

	WP1		WP2		WP3		WP4		WP5		WP6		WP7		WP8		Total	
	A	P	A	P	A	P	A	P	A	P	A	P	A	P	A	P	A	P
Tetracom	6.6	6.6	0	0	7.8	3.9	0.3	0.3	17.3	5.6	5.4	5.4	2.8	2.8	4.28	4.28	44.48	28.88
DFKI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Atlantis	0.76	0.7	0	0	0.73	1.20	2.50	0	5.15	4.24	0.27	0.72	5.83	5.25	0.81	0.83	16.05	12.94
IBL DCL	0.5	0.5	0	0	2.2	0	3	0	7	17.5	6.1	17	1	1	1.5	1.5	21.3	37.5
ICS PAS	2.14	1.08	0	0	0.00	0.00	0.00	0.00	11.99	6.05	0.47	0.24	3.89	1.96	2.57	1.30	21.06	10.63
UHH	0,5	0,5	1	2	0	1	0	0	3	3	6	5	0	0	1	1	11	12
UAIC	0	0	0	0	0	0	0	0	11.7	11.7	0	0	0.79	0.79	0.41	0.41	12.9	12.9
UniZD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ITD	0.05	0	0	0	0	0	0	0	0	0	0	0	0	1	3.89	3	3.94	4
TOTAL	10.05	8.88	1	2	10.73	6.1	5.8	0.3	56.14	48.09	18.24	28.36	14.31	12.8	14.46	12.32	130.73	118.85

Table 3.1 Personnel, subcontracting and other major cost items for Beneficiary 1 TETRACOM INTERACTIVE SOLUTIONS for the period 01.03.2012-30.09.2012			
Work Package	Item description	Amount	Explanations
	Personnel costs	86 615.00	
	Subcontracting		
	workshops, dissemination meetings	4389.53	
	Bank fee	13.34	
	Hardware		
	Software		
	Server rent	2 065.15	
	Deprecations	1 021.79	
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT		94 104.81	

Table 3.2 Personnel, subcontracting and other major cost items for Beneficiary 2 DFKI for the period 01.03.2011-28.02.2012			
Work Package	Item description	Amount	Explanations
	Personnel costs	0	
	Subcontracting	0	

	Travel	0	
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT	0		

Table 3.3 Personnel, subcontracting and other major cost items for Beneficiary 3 - ATLANTIS for the period (March-August 2012)			
Work Package	Item description	Amount	Explanations
1, 2, 3, 4, 5, 6, 7, 8	Personnel costs	74.564,08	
	Subcontracting	-	
	Workshops, meetings	1.378,05	• Project meeting + Project review, Luxembourg 7-09/05/2012
WP8	Dissemination events		
	Deprecations	-	
	Hardware	631,67	Purchase of a new PC unit
	Software	-	
	Server rent	-	
	Indirect costs	22.369,22	
TOTAL DIRECT COSTS AS CLAIMED IN FINANCIAL STATEMENT		98.943,02	

ATLANTIS is requesting the following modifications in their budget distribution:

1) Within "Other specific direct costs" category: Increase Travelling subcategory budget from 10.500 to 13.000 Euros, by reducing accordingly the budget of other subcategories (consumables, h/w, meetings).

2) From "Other specific direct costs" to Personnel: Move 5.000 Euros (to cope with increased effort required for improvements in NER and for building summarization corpora)

3) From "Other specific direct costs" to "Indirect Costs": Move 1.500 Euros

Table 3.4 Personnel, subcontracting and other major cost items for Beneficiary 4 IBL DCL for the period 01.03.2012-31.08.2012		
Item description	Amount	Explanations
Personnel costs	46904.52	
Subcontracting	no	
Workshops, dissemination meetings	5503.08	Project Meetings and dissemination: -Sofia-Luxembourg-Sofia -Sofia-Paris-Avignon-Paris-Sofia Dissemination materials
Hardware	no	
Materials	no	Office materials

TOTAL DIRECT COSTS AS CLAIMED IN FINANCIAL STATEMENT	52407.6	

Table 3.5 Personnel, subcontracting and other major cost items for Beneficiary 5 – ICS PAS for the period 01.03.2012-31.08.2012		
Item description	Amount	Explanations
Personnel costs	75 788,53	
Subcontracting	-	
workshops, dissemination meetings	2 196,00	Project Meetings:Luxembourg; Conference: Amsterdam, NLDB 2012; Hamburg, Crosslingual Language Technology 2012
Deprecations	-	
Hardware	1 542,97	
Software	-	
Server rent	-	
TOTAL DIRECT COSTS AS CLAIMED IN FINANCIAL STATEMENT	79 527,50	

Table 3.6 Personnel, subcontracting and other major cost items for Hamburg University for the period 01.03.2012-31.08.2012		
Item description	Amount	Explanations
Personnel costs	15790.74	
Subcontracting		
Other direct costs	5182.78	Travel
	1864.14	Dissemination
	122.74	Materials
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT	22 960.40	

Table 3.7 Personnel, subcontracting and other major cost items for Beneficiary 7 - UAIC for the period for the period 01.03.2012-31.082012			
Work Package	Item description	Amount	Explanations
5, 7, 8	Personnel costs	24861.78	
	Subcontracting	-	
	workshops, dissemination meetings	1414.30	Project Meetings: Hamburg, Luxembourg
	Deprecations	1622.42	Desktop and server
	Hardware	-	
	Software	-	
	Server rent	-	

TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT	27898.50		

Table 3.8 Personnel, subcontracting and other major cost items for University of Zadar		
Item description	Amount	Explanations
Personnel costs	0	
Subcontracting		
workshops, dissemination meetings	0	
Hardware	0	
Software	-	
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT	0	

Table 3.9 Personnel, subcontracting and other major cost items for Beneficiary 9 ITD for the period 01.03.2012-31.08.2012		
Item description	Amount	Explanations
Personnel costs	9675.98	Payments to Krassen Stefanov, Viktoriya Damyanova and Pavel Boichev
Travel	962.55	Travel for Krassen Stefanov to Luxemburg

Other	0	
Equipment	0	
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT	10638.53	

Annex 1: Report of Dissemination Activities for the period: 01.03.2011-29.02.2012

Performed dissemination activities for the period: 01.03.2011-29.02.2012 by all project partners within the ATLAS project:

Type of Dissemination Activity	Title	Dates and Place	Participant Name and Institution	Performed Dissemination Actions	Available Dissemination Materials	Target Groups and Number of People Reached
Conference	S3T 2011 Conference	September 2011, Bourgas, Bulgaria	Tetracom, Bulgaria	Paper Presentation: "Integration of Natural Language Processing Chains in Content Management Systems" Diman Karagiozov	http://www.atlasproject.eu/atlas/file/1bfe6f91-081c-4ab3-85fe-f6043e112b47/13ccaaa0-a0c0-11e0-8264-0800200c9a66/s3t2011_submission_25.pdf	Researchers
Information portal	Success Stories portal	1 March 2011 Poland	ICS PAS, Poland	Information about the project	Initial information on the project as an article in the portal (http://en.kpk.gov.pl/index.php?option=com_sobi2&catid=5&Itemid=142&lang=pl).	General public *No of people reached: available in the Internet*
Conference	8th International Conference Practical Applications in Language and Computers (PALC 2011)	13-15 April 2011, Łódź, Poland	ICS PAS, Poland	Conference article "i-Publisher, i-Librarian and EUDocLib - Linguistic services for the Web.	Distribution of project leaflets More information: http://palc.ia.uni.lodz.pl ."	Linguists, representatives of social sciences and humanities *No of people reached: 50*
Seminar	Linguistic tasks in ATLAS project (with particular focus on lemmatization of multiword units) @ Natural Language Processing	18 April 2011 Warsaw, Poland	ICS PAS, Poland	ATLAS presentation, showcase of the multiword prototype	A presentation by Adam Przepiórkowski, Maciej Ogrodniczuk and Łukasz Degórski	Computational linguists, NLP engineers No of people reached: 30

	Seminar					
Conference	FLaReNet Forum 2011	26-27 May 2011 Venice, Italy	ICS PAS, Poland	Information about the project	Distribution of project leaflets	Key players in Language Resources and Technologies No of people reached: 120
Conference	19th International Conference Intelligent Information Systems and 26th International Conference on Artificial Intelligence	13-14 June 2011 Warsaw, Poland	ICS PAS, Poland	Conference article: Łukasz Degórski: Towards the Lemmatisation of Polish Nominal Syntactic Groups Using a Shallow Grammar on lemmatization of multiword expressions. More information: http://iis.ipipan.waw.pl	Distribution of project leaflets	IT scientists, linguists No of people reached: 100
Project meeting	CESAR project meeting	26 June 2011, Budapest, Hungary	ICS PAS, Poland DCL – IBL, Bulgaria	Information about the project	Distribution of project leaflets	CESAR project members No of people reached:20
Conference	META-NET Forum 2011	27-28 June 2011 Budapest, Hungary	ICS PAS, Poland DCL – IBL, Bulgaria	Information about the project More information: http://www.meta-net.eu/events/meta-forum-2011/exhibition	Presentation of the project leaflets, poster and online demo in the dedicated ATLAS booth.	Key players in Language Resources and Technologies No of people reached: 300
Exhibition	Permanent exhibition of the Bulgarian Academy of Sciences	April 2011 - April 2012 Sofia, Bulgaria	DCL – IBL, Bulgaria	ATLAS posters in Bulgarian and English	ATLAS posters in Bulgarian and English	General public

Conference	SEPLN 2011: 27th Conference of the Spanish Society for Natural Language Processing	5-7 September 2011, Huelva, Spain	ICS PAS, Tetracom	ATLAS – Multilingual Language Processing Platform article by Maciej Ogrodniczuk and Diman Karagiozov. More information: http://www.uhu.es/sepln2011/.	Conference article, distribution of project leaflets	IT scientists, linguists No of people reached: 60-70
Conference	SlaviCorp 2	12-14 September 2011, Dubrovnik, Croatia	ICS PAS, UniZD	Distribution of project leaflets	Distribution of project leaflets	IT scientists, linguists No of people reached: 40
Meeting	ATLAS @ INFSO.E1 consultation meeting	6-7 December 2011, Paris, France	ICS PAS	Distribution of new i-Publisher leaflets.	The INFSO.E1 (Language Technologies & Machine Translation) consultation meeting in ELDA.	EC representatives, LT/MT experts No of people reached: 30
Conference	RANLP 2011	11-17 September 2011, Hissar, Bulgaria	UHH, Germany	Conference article	http://aclweb.org/anthology-new/R/R11/	Researchers and students in Natural Language Processing
International Seminar	International seminar "Educational Publishing Futures seminar"	17-18 February 2011, Brussels, Belgium	ITDF, Bulgaria	Distribution of brochures	Distribution of brochures	Researchers, publishers, policy makers No of people reached: 80

Seminar	Usability Seminar	18-19 May 2011, Sofia	ITDF, Bulgaria	i-Librarian and i-Publisher prototypes were included in the programme for demonstration among participants	ATLAS project presentation	IT experts, PhD students and researchers in Software engineering, web developers No of people reached: 20
Workshop	International Workshop: Re-designing Institutional Policies and Practices to Enhance the Quality of Education through Innovative Use of Digital Technologies	14-16 June, 2011, Sofia	ITDF, Bulgaria	More information: http://qed.unibit.bg	Distribution of project leaflets	No of people reached: 100 from different EU countries, USA, Russia and Bulgaria
International Network Conference	Dylan Final Conference	22 September 2011	ITDF, Bulgaria	Link of the Network:http://www.dylan-project.org/	Flyer distribution	EC policy makers, Universitites, Companies and NGOs in the field of Language Dynamics and Management of Diversity Number of people reached: 150
International Network Conference	MUMIA - Multilingual and Multifaceted Interactive Information Access (funded by EU COST Action IC1002)	22 September 2011, Amsterdam, Netherlands	ITDF, Bulgaria	Link of the Network: http://www.mumia-network.eu/	Flyer distribution	Network members Number of people reached: 100

Exhibition	WEBIT'11	26-27 October 2011 in Sofia, Bulgaria	ITDF, Bulgaria	ATLAS video presentation on stand	ATLAS video presentation, brochure distribution	Web developers, e-commerce, hosting, content management and CRM provider, ad network, affiliate system for networking with colleagues from the region and learn from the best practices and professionals form CEE and the rest of the world. No of people reached: 150
International Conference	Library and Cultural management and ICT	3-4 September, 2011, Bourgas, Bulgaria	ITDF, Bulgaria	Distribution of project brochures	Distribution of project brochures	Librarians and researchers, students No of people reached: 60
Organised Workshop	Digital Business Perspectives for the Publishers and Librarians within EDE 2011	26 September, 2011, Sofia, Bulgaria	ITDF, Bulgaria	The panel session is organized within EDE 2011 and discussed the possibilities provided by Share.TEC portal and ATLAS project to the publishers and librarians.	Presentation of ATLAS services and Living lab, Distribution of project brochures	Librarians and researchers, students No of people reached: 30
Conference	7th Conference on e-Education, "E-Learning – Real Experience through Innovations"	17 November, 2011, Sofia	ITDF, Bulgaria	Project presentation.	Distribution of project brochures	Students, researchers, companies

						No of people reached: 50
Round table	E-Infrastructures: New pathways for European Development of Bulgaria, Round table, initiated by Mr. Georgi Parvanov, President of Bulgaria and under the auspices of Ms. Irina Bokova, General Director of UNESCO,	18 November 2011, Sofia	ITDF, Bulgaria	Project presentation.	Distribution of project brochures	Policy makers No of people reached: 50
Participation in Workshop	Open Innovation and Service Product development For SMEs, an ENoLL event	5-6 December 2011, Rome Italy	ITDF, Bulgaria	Exploring opportunities for Living Lab networking and collaborations	Distribution of project brochures	Developers, users Living Labs representatives and policy makers No of people reached: 70
Workshops	Three Testing seminars of i-Librarian and i-Publisher within Living Lab	21-22 December 2011, Sofia, Bulgaria	ITDF, Bulgaria	Presentation of i-Librarian and i-Publisher services	Conducting Testing experiments Distribution of project brochures http://livinglab.itd-bg.eu/content/events	Researchers, Authors, Publishers, Librarians - users No of people reached: 20

Participation in Workshop	IPR and Living Labs - contradiction in terminus or way for new opportunities, http://ipr4livinglabs.eventbrite.com/	27 January, 2012	ITDF, Bulgaria	The European Network of Living Labs is organising together with the project APOLLON and IBBT, a workshop dedicated to issues of IPR for Living Labs	Distribution of project brochures	Developers, users Living Labs representatives and policy makers No of people reached: 50
Project meeting	Kick-off meeting of the PROMIS Lingua international project	13-15 April 2011, Luxembourg	Atlantis, Greece	Distribution of EN flyer.	Flyer	Business Consultants, Linguistic and IT experts, as well as a European SME umbrella organisation. *No of people reached: 30*
European Congress	11th Congress of the European Business Angels Network (EBAN). *Theme: "Syndication and co-investment: partnerships and vision for the future"*	12-13 May 2011, Warsaw, Poland	Atlantis, Greece	Discussions on funding possibilities for ATLAS take-up. Distribution of EN flyer.	Flyer and brochure	Business Angels from all over Europe. *No of people reached: 100*
Business development meeting	Corporate business day (Atlantis group of companies) with the participation of managers from all over Greece and Cyprus	10 February, Thessaloniki, Greece	Atlantis, Greece	The 2 ATLAS online services were presented and discussions took place about business development and commercial exploitation	ATLAS presentation, flyer, brochure	New business development managers No of people reached: 12
Conference	GSCL 2011 Conference	28-30 September	Atlantis, Greece	Paper presentation on 26 September 2011	Paper titled "Towards the integration of monolingual	Computational Linguistics and

		2011, Hamburg			NLP frameworks for multilingual applications: A case study for the Greek Language within the ATLAS project framework".	Language Technology providers and researchers
Organized Workshop	ATLAS Workshop "Integration of multilingual resources and tools in Web applications workshop" @ Conference of the German Society for Computational Linguistics and Language Technology (GSCL 2011)	26 September 2011, Hamburg, Germany	TETRACOM, ICS PAS, UHH, ATLANTIS, DCL – IBL	http://www.corpora.uni-hamburg.de/gscl2011/?Workshops:Integration_of_multilingual_resources_and_tools%0Ain_Web_applications	Co-organization of the workshop, several papers, presence in the organising and programme committee	Computational Linguistics and Language Technology providers and researchers No of people reached: 20
Poster Session	ATLAS Poster session @ Conference of the German Society for Computational Linguistics and Language Technology (GSCL 2011)	29 September 2011, Hamburg, Germany	TETRACOM, ICS PAS, UHH, DCL – IBL	ATLAS — A Robust Multilingual Platform for the Web poster by Maciej Ogrodniczuk, Cristina Vertan, Svetla Koeva, and Adam Przepiorkowski. More information: http://www.corpora.uni-hamburg.de/gscl2011/en/?download=GSCL-Programm.pdf	Poster at the main conference	Computational linguists No of people reached: 50
Demo session	ATLAS @ Demo session at the 5th Language & Technology Conference	25-27 November 2011, Poznan, Poland	TETRACOM, ICS PAS, UHH, DCL – IBL,	Paper in the proceedings, software and resource demonstration, information about the	Language Processing Chains in ATLAS by Anelia Belogay, Dan Cristea, Eugen Ignat, Diman	IT scientists, computational linguists

		ATLANTIS, UAIC	project, roll-up exhibition.	Karagiozov, Svetla Koeva, Maciej Ogrodniczuk, Adam Przepiórkowski, Polivios Raxis and Cristina Vertan: demonstrating i-Publisher, i-Librarian, EUDocLib and PLDocLib	No of people reached: 100
(LTC 2011): Human Language Technologies as a Challenge for Computer Science and Linguistics					

Annex II: Multilingual e-Content and e-Library Living lab (MLeCeL) Report

General introduction

The Multilingual e-Content and e-Library Living lab (MLeCeL) is a pilot sectoral Living lab dedicated to contribute and promote digital multilingual on-line content and libraries as an important instrument for development of knowledge-based and multicultural society. The living lab is considered as a strong instrument for establishing an innovation ecosystem which supports quality assurance of innovative services and products, sustainability and business exploitation of such services and products.

The living lab has been launched in the frames of the ATLAS Project, but it has a wider scope of activities. It attracted developers, researchers and prospective end-users of the ATLAS prototype services (I-Publisher, I-Librarian and EUDocLib), as well as ones related to other projects, having been involved in development and deployment of multilingual e-content and e-library services, such as: OpenScout (http://livinglab.itd-bg.eu/content/open-scout), Share.TEC (http://www.share-tec.eu/) and SmartBook (http://dse.fmi.uni-sofia.bg/SmartBook/index.htm). The living lab is open for cooperation with other developers of similar services, end-users, researchers, etc. MLeCeL LL will benefit from the cooperation and the capacity of the regional Virtual Services and Open Innovation Living Lab (VirtSOI LL), from the partnership with the Digital Spaces Living Lab (http://www.digitalspaces.info/) and the European Network of Living Labs (ENoLL). MLeCeL LL will make efforts of joining different networks and research intensive clusters in the area of Cultural and Creative Industries (CCIs) - one of Europe's most dynamic sectors, contributing around 2.6 % to the EU GDP, with a high growth potential, and providing quality jobs to around 5 million people across EU-27 (Green Paper "Unlocking the potential of cultural and creative industries", COM(2010) 183). Such efforts will be directed to strengthening participation in ENoLL, establishment of a network of CCI Living Labs in South-East Europe (together with DSSL and ITI-CERTH), and to joining the ALADIN network of Living Labs along the Danube (http://eLivingLab.org/CrossBordereRegion/InterMunicipality).

MLeCeL Planned activities

1. Ongoing evaluation exercises of the ATLAS services (i-Publisher, i-Librarian, EUdocLIB) by using the official ATLAS evaluation instruments and test scenarios, as well as by applying a number of specialized living labs instruments developed in the framework of the European Network of Living Labs (ENoLL).

2. Collect feedback from users, LL experts, and developers through organized evaluation workshops and dissemination events and analyze the preferred ways for communication of all intended ATLAS LL users, and implement new communication channels if needed.

3. Conducting Two experiments in the Living Lab and collecting feedback from LL experts and developers.

4. Identify new companies, users and networks relevant as ATLAS partners and to involve them as registered users in the ATLAS Living Lab (LL) platform.

5. Design a questionnaire for identifying end user needs in respect to the expected ATLAS impacts by main ATLAS instruments and services and collect filled questionnaires from ATLAS LL users regarding intended ATLAS impact and improve the LL platform to be able to better measure the impact of ATLAS tools and services

6. Organizing and participation in dissemination events.

7. Participation in preparation of an FP7 Regions of Knowledge project (SmartCulture) targeting establishment of a European Network of Research Intensive Clusters and developing a Joint Action Plan in the area of CCIs (ITD, Tetracom and DSLL are members of the Sofia Cluster.

8. Strengthening the cooperation between MLeCel LL and Digital Spaces Living Lab (http://www.digitalspaces.info/) through use of platforms and tools and organization of joint workshops:

   a. Digital Spaces Living Lab will explore the opportunity to use ATLAS prototypes (I-Publisher, I-Librarian and EUDocLib) and will test the prototypes.

   b. Organizing a joint ATLAS/ MLeCel LL/DSLL workshop "Multilingual Digital Repositories and Services" , Sofia, 3 May 2012. Draft programme is available in Bulgarian on the http://livinglab.itd-bg.eu/content/events

   c. Joint MLeCel LL/DSLL "Innovative Social Media and Web Content Processing Applications and Services" workshop in Thessaloniki June/July 2012 hosted by ITI-CERTH (member of DSLL).

MLeCel LivingLab achieved results until April 2012

1. Improving Living Lab platform

- MLeCel website – developed and currently updated (September 2011)
- Communication tools developed within LL website – groups (November 2011)
- Testing and Evaluation questionnaire developed and published on the LL website (November 2011) in order to collect feedback from users, developers and experts - http://livinglab.itd-bg.eu/content/testing-evaluation-questionaire

2. Evaluation Workshops

Testing, evaluation and feedback workshops with users of i_Librarian and i_Publisher, organized in the Living Lab by ITDF (http://livinglab.itd-bg.eu/content/atlas-test-workshops-mlecel-ll):

February-March 2012, Sofia

- One evaluation was performed from the BSc students (Year 2 of their study) from the University of Sofia, during the course Human-Computer Interaction. On February

27th there was a workshop where the ATLAS system was presented to all students. On February 29th students started testing the ATLAS system. All the data was collected from TETRACOM.

- Two workshops for testing and evaluation of ATLAS services were organized by ITDF in cooperation with Sofia University, Department "Software engineering", partner organization of MLeCel Living Lab and Tetracom. They were conducted on 15 March and 17 March 2012. The testers were students in Informatics, Software engineering and Information systems. Most of the testers work in IT companies. Four scenarios were tested by 50 students. The data from the testing was collected by TETRACOM.

- During February and March 2012, i-Publisher - Advance Mode was evaluated by web designers and developers. The data from the evaluation was collected by MLeCel Living Lab. 17 users have tested 4 scenario – i_Publisher Advance Mode. Results at http://livinglab.itd-bg.eu/sites/default/files/Results_March%202012.pdf

December 2011, Sofia

- One Testing Seminar of i-Librarian service was conducted on 21/12/2011 and organised by the MLeCel Living Lab and ITDF. Researchers and young scientists, representatives of Scientific Research Department, partner organisation of the MLeCel Living Lab participated in the Testing Seminar and evaluated the i-Librarian - the interface usability and provided languages features.

- Tree Testing workshops were also organized during December 2011 in State University of Library Sciences and IT, a partner organisation of the MLeCel Living Lab. General Internet users and web designers evaluated I-Publisher service. i-Librarian services were tested by students, and librarians.

Users' feedback was collected with a questionnaire: http://livinglab.itd-bg.eu/content/testing-evaluation-questionaire

Report including users' feedback from Dec. 2011 workshops is available here: http://livinglab.itd-bg.eu/sites/default/files/Results_Dec2011.pdf

3. Pilot experiments in the Living Lab

Feedback from LL experts and users and suggestions were collected during two pilot experiments in the Living Lab:

- **Integration of i-Librarian with DSpace -** Scenarios for possible integration of i-Librarian with DSpace were suggested by the LL experts.

  LL expert feedback and suggestions are available here: http://livinglab.itd-bg.eu/sites/default/files/Scenarios%20for%20integration%20with%20Dspace.pdf

- **A multilingual website of UNESCO Interfaculty Chair** "ICT in Library Studies, Education and Cultural Heritage" was created with i-Publisher within MLeCel Living Lab - http://livinglab.itd-bg.eu/content/i-publisher-show-case

  Report from workshop with LL developers and experts from SVUBIT is available at: http://livinglab.itd-bg.eu/sites/default/files/Creating%20a%20multilingual%20website-en.pdf

4. Performed Dissemination events

Performed dissemination events, organized by ITD, are listed on the Living Lab website: http://livinglab.itd-bg.eu/content/events. Performed dissemination activities include organization of workshops and participation in international events

- **December 2011** - The UNIBIT students in MSc degree in eBusiness tested MLeCeL Living lab services and developed quality course works. In their research documents, they recommended new services, proposed new content, and LL improvements. The Living lab developers and organizers will analyze their high-quality input and will provide soon some new services. Some of the new ideas include adoption of e-learning services and platforms, ideas for closer collaboration with students and companies and others. An aggregated report will be published on the LL web site.

- **Open Innovation and Service Product development For SMEs**, December,5-6, Rome Italy. The Industrial Forum 2011 was a two days event. The 5th of December program was organized into two sessions: a morning session, built around two main panels of representatives from major international enterprises, SME's Clusters, Professional Associations and Policy makers, providing their own view on the emerging practices for Domain Networks to support Industry and SMEs Services; and an afternoon session, organised in 2 parallel workshops, introducing Innovation Platform typologies, one on "Domain Living Lab Networks - Instruments and Service in support to SMEs" and the second focusing on "Transversal Critical Themes for Living Labs". The 6th of December was dedicated to the Match Making of Living Lab and SMEs. SMEs holding a Business idea will chair dedicated Multi-stakeholder Tables to identify practical steps forward for the launch of joint initiatives and to discuss and negotiate terms and conditions for conducting reality checks trials. ATLAS Brochures were distributed. Living Lab was introduced.

- **ATLAS and MLeCel Living Lab@Webit Congress 2011**, October 26-27, 2011. Webit 2011 is one of the Europe's premier digital marketing and IT industry events with special focus on Central and Eastern Europe, covering markets with population of more than 450 million people with more than 5000 visitors from over 40 countries; 7 parallel conference tracks; top world-class speakers; trade show exhibition; unparalleled networking.

  ATLAS project was presented at a stand; brochures of i-Publisher and i_Librarian were distributed. Video, presenting ATLAS services was presented as well.

- **Digital Business Perspectives for the Publishers and Librarians, September 26, 2011** The panel session was organized within EDE 2011 and discussed the possibilities provided by Share.TEC portal and ATLAS project to the publishers and librarians.

- **The EUROPEAN DAY OF THE ENTREPRENEUR, September 26-27**, 2011. Sofia University "St.Kliment Ohridski" and Sofia Municipality is organizing the ninth issue of the European Day of Entrepreneur (EDE). It was held in the frame of the European Week of the SMEs under the patronage of the Ministry of Economy, Energy and Tourism and Minister Traicho Traikov. ATLAS brochures were distributed.

- **Library and Cultural management and ICT, September 3-4, 2011**, organized by State University of Library Studies and Information Technologies in collaboration with Association of the University Libraries in Bulgaria and Free University of Bourgas.

- **SOFTWARE, SERVICES & SEMANTIC TECHNOLOGIES, September 1-2, 2011**. S3T 2011 provided a forum for connecting researchers and international research communities for worldwide dissemination and sharing of ideas and results in the areas of Software and Services and Intelligent Content and Semantics.

- **The Future of e-Book and e-Libraries Workshop, 28 October 2010**, Sofia, Bulgaria. The workshop was organised in the framework of the eight edition of the EUROPEAN DAY OF THE ENTREPRENEUR (EDE) as a traditional collaborative initiative of Sofia Municipality and Sofia University "St. Kliment Ohridski" (28-29 October 2010). The participants were: researchers, e-book companies, Sofia University Library. City Library, National Centre for Information and Documentation, ASTEA Solution and StockPodium (technological companies).

  Presentation of ATLAS services to the potential customers and users, getting feedback from the users about offered services during The Future of e-Book and e-Libraries Workshop, 28 October 2010, Sofia, Bulgaria

5. Planned events for dissemination and getting feedback from the experts

Future expert workshops for getting feedback and dissemination, organized by ITDF and other Living Labs:

- Joint Workshop on Multilingual Digital Repositories and Services, 3 May 2012, Sofia at State University of Library Sciences and IT. Invitation and draft programme in Bulgarian are available here: http://livinglab.itd-bg.eu/content/events

  ATLAS and its services will be presented at the workshop. Other presented projects and products will be: CLARIN: Common Language Resources and Technology Infrastructure (BAS), Europeana and Bulgariana (Ontotext), Multimedia newsroom in Bulgaria (TV), Vocabler - всички ваши думи на всички ваши езици (Vocabler), Querying RDF Semantic Repositories in Natural Language (Ontotext) and others. The objective of this workshop is to explore different opportunities for collaboration between projects and services and to match make between identified needs for services and offered solutions presented by the participants.

- Joint MLeCel LL/DSLL "Innovative Social Media and Web Content Processing Applications and Services" workshop in Thessaloniki June/July 2012 hosted by ITI-CERTH (member of DSLL).

6. Indicators progress

Indicators	Quantity	Progress (April 2012)	Comments
User interviews	> 20	15 (Dec 2011) 17 (Febr. – March 2012)	Report including users' feedback from Dec. 2011 workshops is available here: http://livinglab.itd-bg.eu/sites/default/files/Results_Dec2011.pdf Results are at: http://livinglab.itd-bg.eu/sites/default/files/Results_March%202012.pdf
User suggestions	> 30	38	Collected from all the events described
Feedback from idea generation workshops with lead-users	at least 3	2	Feedback from joint idea generation workshops with lead-users, organized by ITD and Tetracom were presented in relation with Two experiments. During joint workshop on 03/05 in the Living Lab, ideas will be collected also from potential users and experts.
Feedback from idea generation workshops with experts	at least 2	1 for Unesco Interfaculty Chair website 1 for Integration with DSpace	Report from workshop with LL developers and experts from SVUBIT is available at: http://livinglab.itd-bg.eu/sites/default/files/Creating%20a%20multilingual%20website-en.pdf Feedback from developer from Sofia University: http://livinglab.itd-bg.eu/sites/default/files/Scenarios%20for%20integration%20with%20DSpace.pdf
Online suggestion box	ongoing moderation and summarization	Available	To collect suggestions from the casual users in the database
Feedback from workshops with users and prospective users	at least 4	8	Description of Workshops is on: (http://livinglab.itd-bg.eu/content/atlas-test-workshops-mlecel-ll)

			Feedback is collected through questionnaires, available on the website. Several more workshops to be planned
Feedback from prototype testing	at least 2	3	Feedback was collected by questionnaires during prototype testing on evaluation workshops.
Feedback from real settings service deployment	at SULSIT and Sofia University		Constantly collected with a special questionnaire at Living Lab during the service deployment.
Usability Lab controlled assessment of user acceptance	at least 1	1	Usability Lab controlled assessment of user acceptance was organized during December 2011 for i-Publisher and i-Librarian. The participants were introduced to the services by ITDF and then they experimented with the services. The results are consolidated in two reports: http://livinglab.itd-bg.eu/sites/default/files/Results_Dec2011.pdf http://livinglab.itd-bg.eu/sites/default/files/Results_March%202012.pdf
Feedback from education and training sessions at university level	at least 2	1	Received feedback from education and training sessions at university level, published at: http://livinglab.itd-bg.eu/content/atlas-test-workshops-mlecel-ll
Feedback from training workshops with developers	at least 2	2	Training Workshops were conducted by TETRACOM with developers in order to be introduced with i-Publisher-Advance Mode.
Feedback from workshops with researchers	at least 2	1	Received feedback from education and training sessions at university level, published at: http://livinglab.itd-bg.eu/content/atlas-test-workshops-mlecel-ll
Catalyzing research activities based on ATLAS	at least 1 PhD study and 10 MSc studies	3 PhD students and 12 MSC students	All studies related to different research problems related to ATLAS project.