

Final report

Grant Agreement number: 250467

Project acronym: ATLAS

Project title: Applied Technology for Language-Aided CMS

Project type: Pilot A Pilot B TN BPN

Period covered: 01.03.2010 - 28.02.2013

Project coordinator name, title and organisation:

Anelia Belogay, CEO, Diman Karagiozov, CTO,

Tetacom Interactive Solutions

Tel: +35924950444

Fax: +35924950443

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

Project website address: www.atlasproject.eu

Introduction

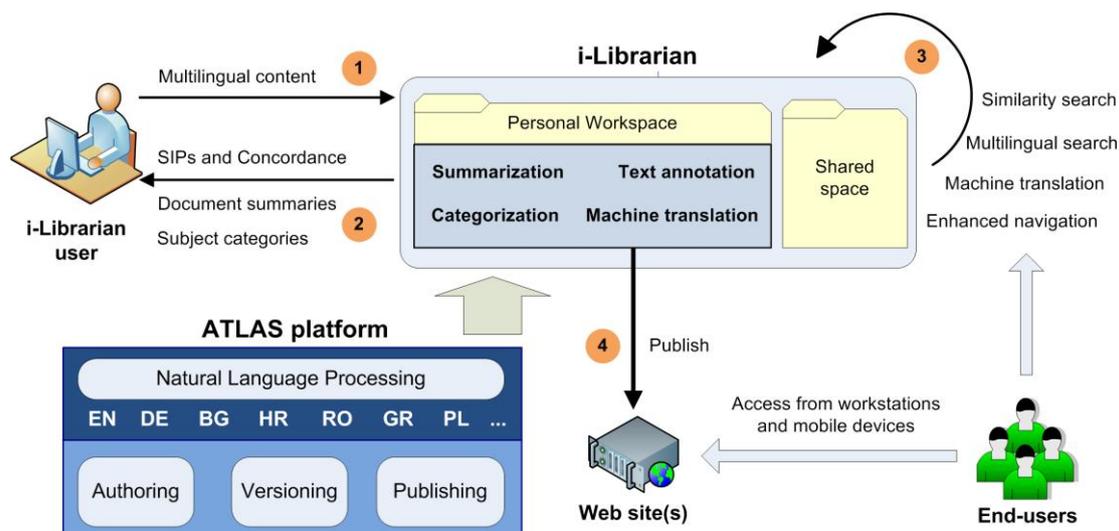
The advent of the Web revolutionized the way in which content is manipulated and delivered. As a result, digital content in various languages has become widely available on the Internet and its sheer volume and language diversity have presented an opportunity for embracing new methods and tools for content 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 summarization and categorization, 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.

The project solutions

i-Librarian – the intelligent content assistant service

The first solution, i-Librarian, is a web-based content assistant service, which allows users not only to store, organize and publish their personal works but also to locate similar documents in different languages and to obtain easily the most essential texts from large collections of unfamiliar documents or search engine results.

i-Librarian is a web-based content assistant 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. Advanced language-based technology is implemented to help users easily navigate between and access both their personal works and unfamiliar documents. After processing a large collection of unfamiliar texts i-Librarian displays short summaries and extracted concepts that enable users to easily decide which documents are worth reading and which could be discarded. Furthermore, i-Librarian interlinks all user documents based on the extracted phrases, words and names, and thus improves significantly content navigation. Finally, the service helps users with no previous experience to publish their own content using the power of a modern content management system but without struggling with the inherent complexity of such systems. The features of i-Librarian will be initially available in seven languages – English, German, Bulgarian, Croatian, Greek, Polish and Romanian. However, as more languages could be easily integrated in the service, the consortium will explore several options to secure the necessary funding after the end of the project for supporting all other major European languages.



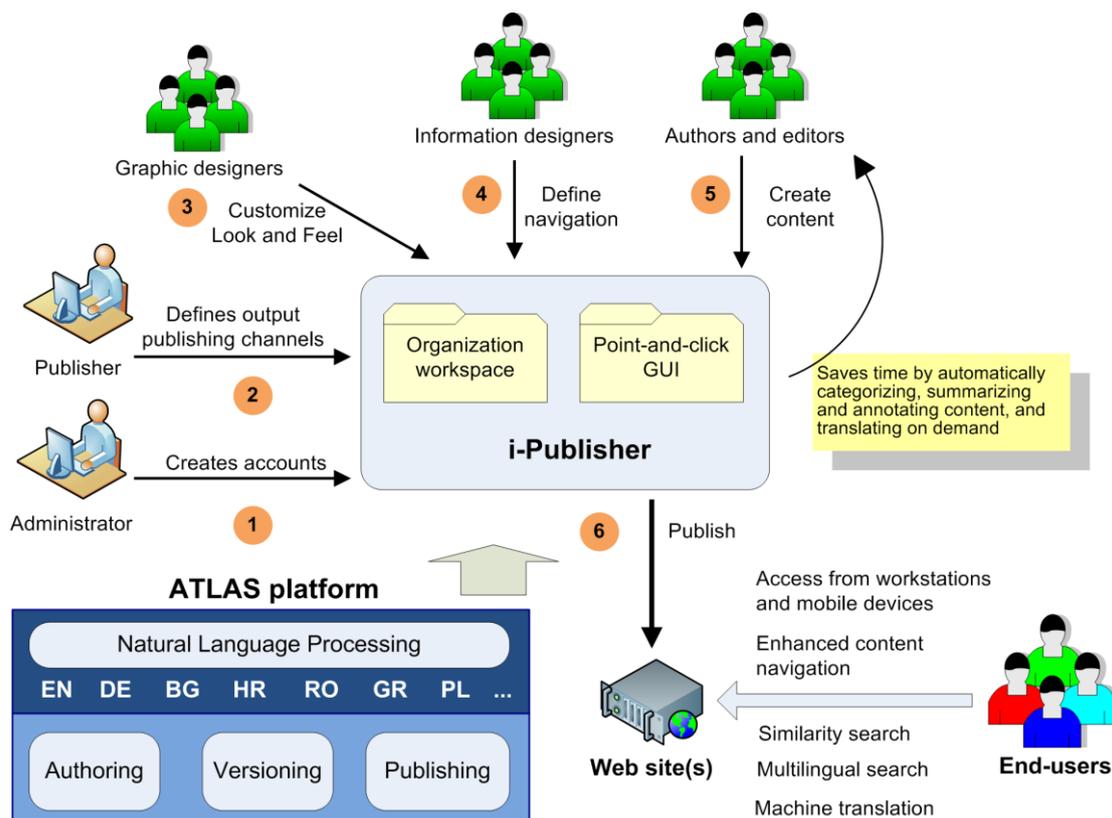
Some of the main characteristics of i-Librarian are summarized below:

- i-Librarian offers multilingual full-text search inside personal or shared documents.
- i-Librarian provides multilingual similarity search, which enables users to easily locate both personal and shared similar documents in different languages.
- i-Librarian implements powerful instruments for computer-aided translation, automatic content categorization, summarization, and annotation of important words, phrases and names.
- Users can rate the quality of automatic translations and improve them, which would help the consortium to build better translation models for future use.
- Users can publish heterogeneous multilingual content on a personal web site hosted by i-Librarian or on existing web sites and portals.
- Users can freely annotate documents in their personal workspace and search through the annotations (possibly in different languages) shared by other users in order to find documents of interest.
- i-Librarian is accessible from a browser or mobile devices such as iPhone.
- i-Librarian includes a mechanism for reporting and removing of materials that violate copyright laws.

i-Publisher – the online web content management solution

i-Publisher is a novel software-as-a-service solution for web content management, which allows both small and large organizations to deploy and manage multilingual web sites without spending time and efforts for installing and maintaining a content management system. This service assists organizations in retrieving, unifying, and packaging heterogeneous pieces of content, and dynamically rendering them on multiple web sites. i-Publisher fosters collaboration in content creation by enabling authors, editors, and other contributors to work together. It also facilitates the process by automatically categorizing, summarizing, and tagging the newly created content. Furthermore, web sites may be built

with i-Publisher with a point-and-click graphical user interface by people with different expertise but no programming experience – publishers, information designers and graphic designers. The service leverages the full benefits of the ATLAS platform and becomes an ideal choice for promoting any type of organization on the Web. i-Publisher will be available free of cost for non-commercial use in order to promote web standards and encourage language diversity in content creation. Different subscription plans will be available to those who desire more storage space and customer support, or who would like to use i-Publisher for commercial purposes.



i-Publisher characteristics :

- i-Publisher is well-suited to both small and large organizations as it is designed with scalability in mind, i.e. if an organization needs to handle more content and users, additional servers will help address these needs and provide the desired results in terms of performance.
- i-Publisher improves content navigation by dynamically interlinking content items based on extracted important words, phrases and names.
- i-Publisher utilizes a flexible user access rights system comparable to that of a modern server operating system – security policies may be set for groups and specific users as well as for specific content items or even content item properties.
- i-Publisher implements an industrial strength versioning system, which supports the versioning of structured content rather than the simple text-based versioning found in most existing solutions.

- i-Publisher allows content to be mass exported to or imported from file systems, databases or file servers.
- Web sites created with i-Publisher offer to end-users multilingual full-text and similarity search as well as clustered, summarized and annotated content.

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.

Major achievements with regard to the management objectives:

A management and coordination framework was established to ensure the smooth progress of the project.

- The consortium agreed on a process through which to monitor the allocation and distribution of project resources, as well as to control the quality and timely delivery of project deliverables.
- Seven project meetings were organized (the kick-off and two WP meetings). A common understanding of the project goals was gradually achieved on these meetings. Furthermore, the consortium was able to smoothly define the next steps needed in order to achieve the objectives for the next period.
- Channels ensuring the good management and technical communication were established.

- The first periodic report covering month one through month six of the project was prepared and submitted to the EC.

Achievements in terms of the technical objectives set for the project:

WP 2

- A set of use cases to be used by user groups to evaluate the ATLAS platform and online services was prepared.
- A specification of the linguistic framework for the language tools to be integrated into the ATLAS platform was drafted, agreed upon and finalized.
- **i-Publisher visualization layer** was built and integrated into ATLAS CMS and two online services i-Librarian and EUDocLib were built with i-Publisher.
- **i-Publisher** was deployed as an online service at <http://i-publisher.atlasproject.eu>
- **i-Librarian and EUDocLib online services** were deployed at: <http://i-Librarian.eu>; <http://eudoclib.atlasproject.eu>
- **i-Publisher Simple Mode** was built and integrated into ATLAS CMS. The simplified layer of i-Publisher provides inexperienced users with ready-to-be-used thematic web sites and themes.
- Improvements in the linguistic platform were done – a new noun phrase extractor was implemented together with new name entity extractor for English. In addition, a new functionality to show the processing status was implemented and as a result the user gets intermediate feedback from language processing framework
- Improvements of usability of i-Publisher and i-Librarian – export/import and reuse of a web site was implemented; content clustering, archive upload and asynchronous widget reload, major performance and stability improvements; improved workflow of the process of invitation and user registration.
- ATLAS was extended to track and aggregate the activity of its users in order to check if and for how long the users evaluating ATLAS fulfill the exercises. Furthermore, technical monitors have been integrated in ATLAS in order to measure the performance and to identify potential bottlenecks in the system.
- The new import tool in ATLAS supports import of data stored in other content repositories though OAI-PMH ORE standard format.
- The user guides of i-Publisher Advanced Mode and i-Librarian are accessible at:
http://www.atlasproject.eu/atlas/documentation/simple/getting_to_know_the_atlas_simple_mode_interface
<http://i-publisher.atlasproject.eu/atlas/documentation/advanced/interface>

The technical documentation of the three services is accessible at:

<http://www.atlasproject.eu/atlas/documentation/tech/intro>

- 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.

WP 3

- The existing categorisation engine was extended to support several algorithms, namely Relative Entropy, Naive Bayesian, and Class-Featured Centroid. Three dimension reduction techniques were implemented – best-N (by tfidf), prune (below and above) and Chi2. Furthermore, the engine now supports combining of different classifiers which results in better predictive performance.
- A new classification tool was implemented in ATLAS and integrated in i-Librarian. The clustering tool organises the documents into groups and can be used as an alternative to the classification tools based on model building
- 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.
- Using the categorization tool we found the optimal parameters of the classification models in i-Librarian in all project languages.
- 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.

WP 4

- Existing linguistic software components, representing best-suited language processing tools available for all project languages, have been adapted to conform to a number of widely recognized standards, making them technically and linguistically interoperable, and further developed to satisfy high requirements of the project quality and performance.
- Language processing chains (LPCs) for text annotation (a separate processing chain for each target language) have been implemented and integrated into the multilingual Web content management platform (ATLAS) and successfully used in several thematic content-driven websites (including i-Librarian, EUDocLib or PLDocLib).
- All expected functionalities of the LPCs (segmentation, tokenization, lemmatization, extraction of important noun phrases, named entities as well as providing linguistic annotation for higher-level functions, such as categorization, summarization or machine translation) have been successfully delivered.
- All LPCs demonstrated the latest achievements in the field of natural language processing for each project language.
- LPCs have been produced and evaluated using standard software engineering procedures (including multi-level validation, regression tests, documentation, performance improvements etc.) ensuring their high quality and suitability for usage in production environments.
- The flexible, UIMA-based linguistic architecture enables easy extension of the framework with new languages.

WP 5

- Update of RARE, an anaphora resolution framework (by UAIC).
- Construction of RARE's localization files (for all languages, by all partners).
- Acquisition of resources used in the training of the summarisation modules (for all languages, by all partners).

- Development of a Multilingual Sentence Segmentation module incorporating 4 components (by UAIC):
 - a Training Module;
 - a Segmenter Module;
 - an Evaluation Module;
 - a Calibration System.
- Training of the Sentence Segmentation models (for all languages, by all partners).
- Development of the Multilingual Discourse Parser (by UAIC).
- Development of the Discourse Parser Calibration System (by UAIC).
- Development of the MarkersEditor, a graphical interface helping to visualise and edit the markers used by the discourse parser (by UAIC).
- Development of the DrawTrees, a graphical interface for handy visualisation of the computed discourse trees (by UAIC).
- Writing of the “Partner’s manual for building corpora” by Dan Cristea, Daniel Anechitei and Eugen Ignat (UAIC).
- Acquisition of the resources used in the evaluation phase of the summarisation chain (by all partners).
- Integration of UIMA compatible LPCs with the summarisation modules (for all languages, by TETRACOM, UAIC and partners).
- Integration of 6 language-specific summarisation chains in UIMA (by TETRACOM, in close correlation with UAIC and partners).
- Complete beginning-to-end tests of the integrated chains and comparison against the manually annotated files (by partners, actively interacting with UAIC and TETRACOM).
- Comparison against state-of-the-art achievements, described in the deliverable D5.1 (by UAIC, TETRACOM, ATLANTIS and partners).
- Integration of the summarisation chains in the ATLAS applications (by TETRACOM).
- Solving the errors reported by partners and ATLANTIS (by UAIC and TETRACOM, assisted by partners).
- Writing of a detailed documentation describing the summarisation chain, in the form of a deliverable (D5.1) and a paper (Anechitei et al., 2013) submitted for publication as a chapter of a Springer Verlag volume (by UAIC and partners).

WP 6

Machine Translation

The development of a machine translation engine for the ATLAS system faced following challenges:

- 15 language pairs (EN-BG, EN-DE, EN-EL, EN-PL, EN-RO, BG-DE, BG-EL, BG-PL, BG-RO, DE-EL, DE-PL, DE-RO, EL-PL, EL-RO, PL-RO). With exception of the language pairs involving English for all other there are very limited training resources, or existent open-source rule-based systems
- Open-domain setting: there is no a-priori limitation on the domain of documents which can be uploaded in the ATLAS CMS. Through the classification module a domain is assigned to each document, and the system searches if a tailored translation-domain is attached to this domain.

Thus the development of the translation engine involved following steps:

- Documentation of existent training resource for the involved language pairs and selection of those resources which ensure a homogeneous behaviour of the system across language pairs
- Documentation of different parameters involved in open-source machine translation systems and selection of best setting
- Development of an example-based machine translation system for handling domains in which pattern-expressions are quite frequent
- Set-up of a methodology for domain adaptation
- Collection of small domain corpora to be used for domain adaptation for all domains and all language pairs. Given the large number of corpora and the unavailability of resources for some language pairs we had to develop a methodology for automatic collection of corpus material
- Training of language and translation models for all language pairs and all domains. These models are all compatible with the EUROMatrix Plus environment.
- An automatic evaluation (BLEU measure) was performed for all trained models. For selected models we performed also a human evaluation of non-translated lexical

Crosslingual Retrieval

The implementation of a crosslingual search engine into the ATLAS-system faced also following challenges:

- Given the large number of languages and the variety of documents which may be uploaded to the ATLAS CMS, a simple search based on word-index would have been inefficient
- pure semantic search approach with an ontology as backbone is also impossible to be implemented given the open-domain setting

Given the above-mentioned restrictions we choose a middle-way solution based on the RDF-index principle of the Semantic web, but avoiding the involvement of domain specific ontologies.

Following steps were followed during the development of the engine

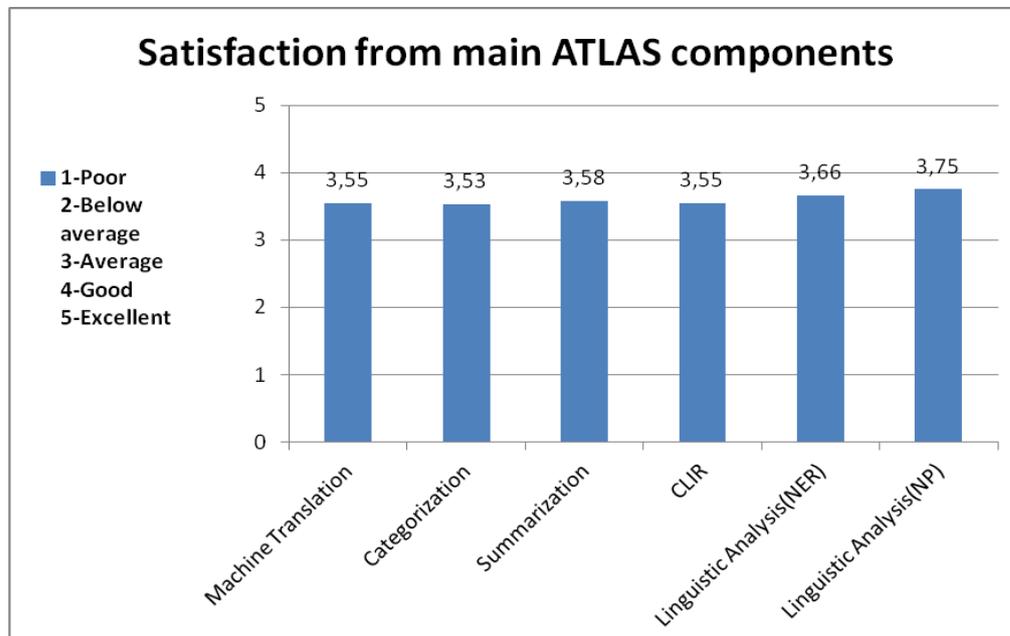
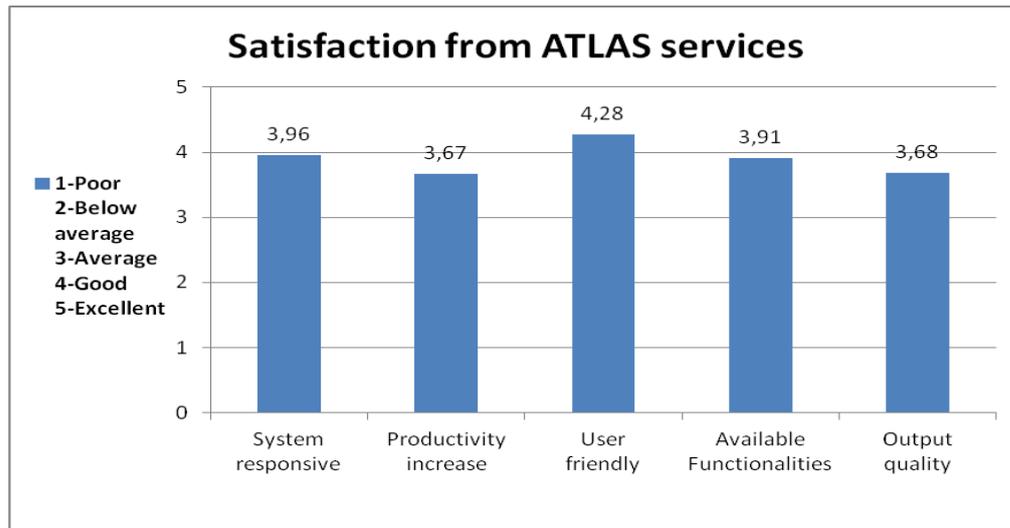
- Definition of an interface between the cross-lingual search engine and the atlas system. This interface collects all information from the ATLAS objects, which have to be indexed. The interface was XML-serialized
- Automatic generation of an RDF index based on the information collected in Step 1
- Adaptation of the NEBULA 5 Framework developed at the University of Hamburg, in order to ensure the generation of the Index Database and the dynamic document search.
- The cross-lingual engine was consequently tested in both monolingual and multilingual environment.

WP 7

The scope of WP7 encompassed two major axes of work, namely user acceptance evaluation and technical evaluation and testing. The major project achievements in each one of them is outlined below:

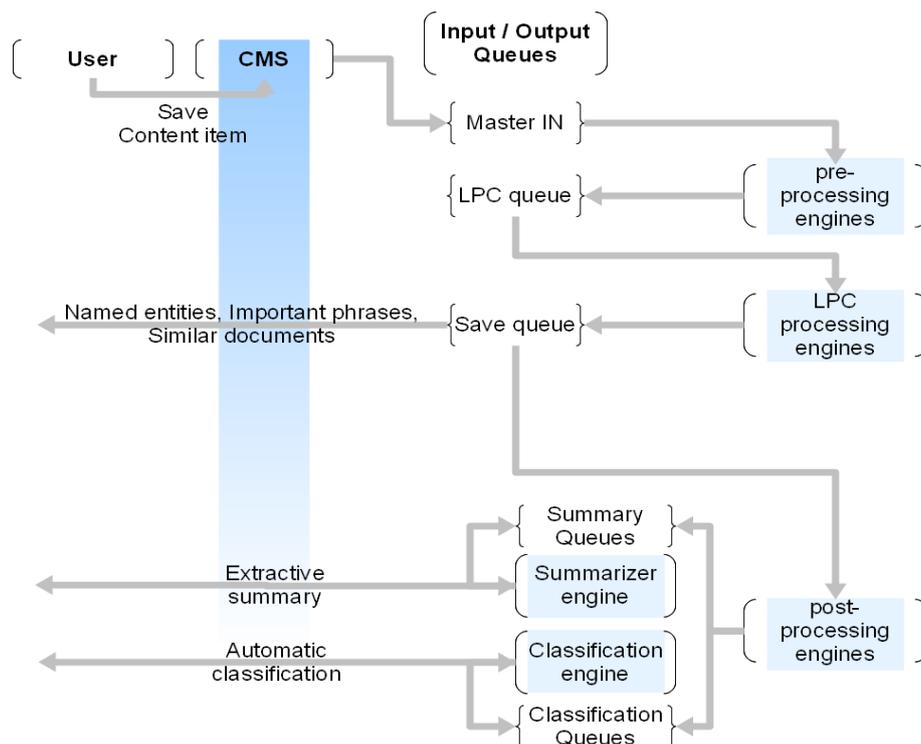
User Acceptance evaluation

- Detailed methodology in order to organise and conduct the user evaluation: methodological approach to the different evaluation challenges, approach for processing of the results, user types and groups, methodology and timing of their involvement, local user groups and Living Labs involvement, etc.
- Established a pilot Living Lab (“Multilingual e-Content and e-Library” - MLeCeL) to leverage the effectiveness of our user evaluation activities and strengthen the user involvement in the project. MLeCeL achievements include:
 - ✓ Training and Testing Seminar of i-Librarian service;
 - ✓ Evaluation workshop on the i-Librarian service;
 - ✓ Two feedback workshops on the full versions of i-Librarian and i-Publisher;
 - ✓ One workshop to test and evaluate the i-Publisher - Advance Mode;
 - ✓ Four testing and evaluation workshops on i-Librarian and i-Publisher.
- Definition of 4 types of user groups and a Living Lab user group:
 - ✓ UG1: students and scholars
 - ✓ UG2: authors, scientists, researchers
 - ✓ UG3: Internet users with moderate WEB experience
 - ✓ UG4: digital content and content publishing professionals (WEB designers, editors, publishers, lawyers, etc.)
 - ✓ LL: professionals, members of the MLeCeL Living Lab.
- Detailed use-case scenarios and exercises for each one of the 4+1 defined user groups, including objectives per scenario and the relevant system features under evaluation, mandatory and optional steps, etc.
- Statistical framework for collecting and analysing the user feedback, including statistical indicators and measures to be used, and relevant thresholds.
- Online questionnaires for each user group type.
- Conducted 3 user evaluation rounds, each round having different objectives and user involvement:
 1. First round: a small number of focus group members (along with ATLAS project members) evaluated a limited set of functionalities of i-Librarian, i-Publisher and EUDocLib in English. Took place during January – March 2011. User participation: 33
 2. Second round: evaluation of i-Librarian, i-Publisher and EUDocLib first complete production versions. Took place during January – March 2012. User participation: 131.
 3. Third round: evaluation of the final version of the public services and the ATLAS software with all planned functionalities available in all partner languages. Took place during November – December 2012. User participation: 139.



Technical evaluation and testing

- Detailed testing methodology: what is to be tested, what is the scope with respect to individual components and indicators, etc.
- Scenarios, testing steps and conditions for failure and success, testing corpora, etc. for the integration and regression testing of the whole platform and its main components (see testing reference diagram below).



- Results from the testing of the ATLAS platform, at the level of isolated components, at the level of integrated platform, and after each new deployment (i.e. regression testing).

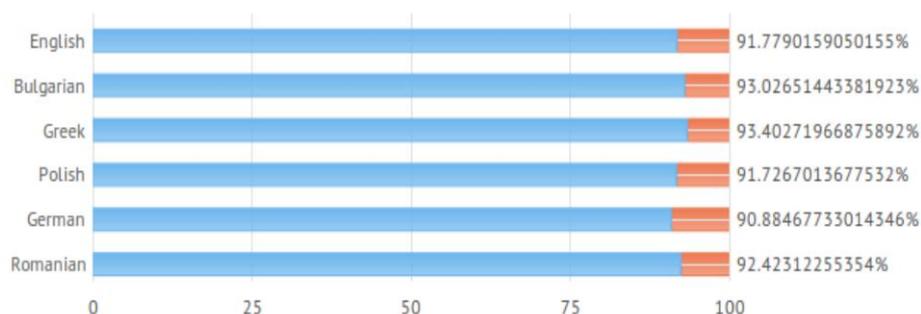


Regression Tests Report

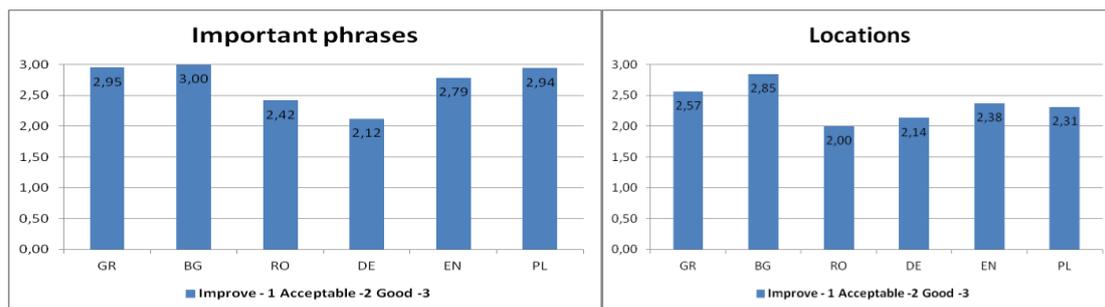
01.02.2013

10:18

Overall results



- Results from assessing the level of ATLAS application specifications fulfilment, including the assessment results for its two main applications (i-Librarian, i-Publisher).
- Collection of documents (corpora) used for both testing and technical evaluation.
- Methodology used for the technical evaluation of the platform components, including test cases and scenarios for each main component and respective technical indicators.
- Results from the technical evaluation of each main component, including comparative assessment results – where this was applicable.



	Language	BG	DE	EN	GR	PL	RO	AVG
ATLAS summarizer	Precision (H)	0,19	0,23	0,27	0,23	0,17	0,22	0,22
	Recall (H)	0,29	0,44	0,41	0,41	0,36	0,32	0,37
	F-measure (H)	0,23	0,30	0,32	0,29	0,23	0,25	0,27
OTS summarizer	Precision (H)	0,16	0,19	0,24	0,27	0,19	0,29	0,22
	Recall (H)	0,25	0,20	0,22	0,33	0,21	0,06	0,21
	F-measure (H)	0,19	0,20	0,23	0,27	0,20	0,10	0,20
LexRank summarizer	Precision (H)	0,15	0,23	0,27	0,24	0,24	0,21	0,21
	Recall (H)	0,18	0,25	0,25	0,22	0,24	0,22	0,18
	F-measure (H)	0,16	0,24	0,26	0,23	0,22	0,21	0,19

We performed automatic evaluation of the machine translation modules through the BLEU measure. A comparison with GOLD standard is not appropriate, as the only existent GOLD standard refers to training and test data from the ACQUIS-Communautaire, while the ATLAS system performed domain adaptation on 13 domains.

Major achievements on dissemination activities

WP 8

- ATLAS project website was created with i-Publisher: <http://i-publisher.atlasproject.eu/>, <http://atlasproject.eu/>
- The Atlas services – i-Publisher and the linguistic platform Asset have their web sites: <http://i-publisher.atlasproject.eu/>, <http://atlasproject.eu/atlas/asset>
- A new vision of the project and its services is applied to the dissemination materials and services websites. New brochures for i-Publisher, i-Librarian and Atlas project were designed and printed
- 31 articles and papers, published in scientific and technical journals and conference proceedings to the Research Community and 10 other publications in online media and etc.
- 10 posters, introduced at conferences and exhibitions for better presentation of the developed services, used technologies, present and further benefits, and widening the scope of the potential users.
- Development of 12 demonstration cases for potential customers
- Developed Sustainability strategy to be applied after the project end.
- Partners` exploitation plans for further use of the ATLAS results
- Interaction with Research Community (linguists, computational linguists, NLP engineers, IT scientists, key players in Language Resources and Technologies and representatives of social sciences and humanities, students and researchers) through:
 - Coorganisation of large dissemination events – Conference "Crosslingual language technology in service of a multilingual integrated

- Europe" , May 2012, Hamburg and CESAR META-NET Road show 2012, Sofia, Bulgaria
- Organisation of ATLAS Workshop "Integration of multilingual resources and tools in Web applications workshop", September 2011, Hamburg;
 - Demonstration of the ATLAS services in posters and demos at scientific conferences
 - Conducting 13 university seminars in partner institutions with students and researchers
 - Presentation and demonstrations of ATLAS achievements at University of Buenos Aires, University of Cordoba, Argentina and University of Santiago, Santiago de Chile, November 2012
 - Interaction with Developers and Users Community (IT experts, media, publishers, librarians, business users, etc.) through:
 - Organisation of 8 evaluation workshops with 112 users carried out in the Living Lab
 - Collected feedback from users and experts, available on the LL website
 - 38 user suggestions, collected through the LL website
 - 7 major ATLAS events organised within Living Lab for users and customers
 - 6 pilot experiments carried within the Living Lab
 - 6 feedback from prototype testing
 - Interaction with potential customers through conducting meetings and building a special demonstration service targeted at the potential customers:
 - Multilingual technologies department at World Health Organisation at IKS Workshop "Semantic Enterprise Technologies in Action", June, Salzburg and presentation of the demo service in Geneva office of the WHO, October 2012, <http://asset.atlasproject.eu/who>
 - Education Unit for use case for using linguistic technologies for multilingual data pools, September 2012, <http://unesco.atlasproject.eu/>
 - Department in Journalism, Sofia University and a meeting with a political observer from Bulgarian National Television, <http://newsdemo.atlasproject.eu/>
 - A Living Lab activity: a response to an end-user request for showing similar video objects for the Bulgarian National Television, <http://videodemo.atlasproject.eu/>
 - M3 Communications Group, Inc. A Hill+Knowlton Strategies Affiliate, www.m3bg.com, <http://mediaimage.tetracom.com/>
 - Foundation "Educational programmes", <http://textmatch.eu/>
 - Exploitation event targeting representatives of Enterprise Community in November 2012. Live presentation and demonstration of the 2 ATLAS services (i-Librarian and i-Publisher) in particular business cases were presented to the Hellenic Mobile Cluster, Athens Technical University, Greek Public Properties Company (ETAD) and big private corporate groups.
 - Presentation of ATLAS and discussion for possible commercial exploitation with European entrepreneurs and international investors during EUREKA Venture Forum, in October, Istanbul, Turkey
 - Discussions on funding opportunities for new online services based on specific ATLAS results (i-Publisher, automatic multilingual summarization) at 2nd Pan European Private Equity and Start-ups Forum in June 2012, Athens with executives from Business Angels, Venture Capitals, Banks, Start-ups, including also entrepreneurs and researchers.
 - Meeting with potential customers in with representatives of the Cluster of Innovation and Culture to demonstrate the Media Talk: a press clipping service in Bulgarian and English: <http://mediaimage.tetracom.com/>, February 2013

- Presentation during a meeting with the Institute for Security Studies at ETH Zurich, August
- Meeting with an industrial customer eMatch during LT Innovative, Brussels, June
- Sharing knowledge and experience within the ATLAS communities:
 - More than 62 presentation and demonstration made by partners at international and national events;
 - Online video tutorials, guides and technical documentation about i-Publisher and i-Librarian
 - about 100 performed dissemination activities (18 exhibitions, 40 conferences and 40 workshops, seminars, meetings and lectures...)
 - Multilingual e-Content and e-Library Living lab (MLeCeL) website for interaction with target communities, <http://livinglab.itd-bg.eu/>
- Atlas was presented and demonstrated at several major dissemination events (large conferences and exhibitions):
 - Presentation of ATLAS at the project village of Human Language Technology Days 2012, September 2012, Warsaw, Poland
 - Presentation of ATLAS at the International Exhibition of Software Bucharest, November, 2012
 - The project was presented with a stand at META-FORUM 2012
 - ATLAS services were demonstrated at the Demo section at the most significant scientific event for the year in Europe - EACL 2012, April 2012, Avignon, France
 - ATLAS was presented with a paper and poster at the biggest linguistic conference LREC 2012, May, Istanbul, Turkey
 - The project presentation was held at CESAR META-NET Roadshow, Sofia, May 2012
 - ATLAS was presented at **EAMT 2012**, May 2012 Trento
 - Atlas services were presented at the world biggest digital event of the year CeBIT 2012 in Hannover, March 2012
<http://www.cebit.de/product/atlas-build-your-website-at-no-cost/291713/C913812>
<http://www.cebit.de/product/asset-adds-value-to-your-content/291715/C913812>
 - 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)
 - ATLAS – Multilingual Language Processing Platform paper presentation at SEPLN 2011
 - The ATLAS project and i-Publisher, i-Librarian were presented and demonstrated in a stand at WEBIT 2011 (webit.bg) to a wider audience (5000 visitors, 40 countries)
 - ATLAS poster presentation at META-FORUM 2011, Budapest
 - Demonstration of i-Publisher, i-Librarian, EUDocLib and PLDocLib at Demo session and a paper at 5th Language & Technology Conference (LTC 2011), November 2011, Poznan, Poland
- Established synergies with relevant 10 projects – META-NET, Cesar, D-SPIN, CLARIN, PROMIS Lingua, OpenScout , ShareTEC, SmartBook , DYLAN, ELLIOT and 3 networks – MUMIA, EUROPE Innova KIS and OpenAIRE, by the organisation of joint events, presentations, demonstrations, exchange of ideas and discussion of future exploitation activities

- A linguistic platform **ASSET** was integrated into the content management platform **Atlas**. The supported languages are Bulgarian, English, German, Greek, Polish and Romanian.
- The linguistic platform contains a **Classification Tool**, a **Summarisation Tool** and a **Machine Translation Engine** that process and analyse texts in all target languages.
- Several demo services show the usability of the implemented services and act as a major dissemination instrument to reach the target groups

<http://asset.atlasproject.eu/who>
<http://newsdemo.atlasproject.eu/>
<http://unesco.atlasproject.eu/>
<http://en.mtalk.eu>



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 not merely meet the needs of modern multilingual content management, but also create value for all users.

Main 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. The simplified layer of i-Publisher provides inexperienced users with ready-to-be-used thematic web sites and themes with the text analysis techniques embedded.
- 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 workspace, shared or published. EUDoLib 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. In addition, the platform allows the integration of machine translation providers and classification algorithms.
- 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.

Impact

The project brings together advanced technologies for multilingual web content management and text mining (such as automated 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 impacts:

Techological

Planned	Achievements	Extent
Integration of text mining tools into content management systems	The ATLAS framework employs technologically and linguistically diverse natural language processing (NLP) tools in a platform, based on UIMA2. The UIMA pluggable component architecture and software framework are designed to analyse content and to structure it. The ATLAS core annotation schema, as a uniform representation model, normalizes and harmonizes the heterogeneous nature of the NLP tools. The language processing tools are integrated in a language processing chain (LPC), so the output of a given NLP tools is used as an input for the next tool in the chain.	100%
Integration of text mining services	The ATLAS platform integrates automatic: categorization; summarisation; text annotation; extraction of important phrases, words and names and translates documents regardless of their language and format. The ATLAS platform is designed with extensibility in mind, which allows for easy addition of tools and services for currently unsupported languages as well as new tools for already supported languages.	100%
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.	<p>Machine Translation:</p> <ul style="list-style-type: none"> • Working version of the Example-based MT engine Training of SMT models for : SL <--> English SL < ---> German based on JRC-Acquis Communautaire, EU-constitution and Europarl. • Setting up a framework for training models for domain adaptation that implied: <ul style="list-style-type: none"> - setting -up the methodology, including a strategy for low resourced language pairs - definition of domains for which models will be trained - collection of training - data (on-going) - proof-of-concept prototype for financial domain. <p>Crosslingual retrieval:</p> <ul style="list-style-type: none"> • Standardization of the interface between the ATLAS-platform and the crosslingual search engine. • Implementation of a first prototype of the crosslingual search engine. 	100%

<p>Contribution to the development of text processing chains for languages, which lack resources at present</p>	<ul style="list-style-type: none"> • New, previously non-existing implementations of important tools have been created (e.g. Spejd-based multiword noun phrase lemmatizer for Polish and the Romanian NP-chunker). • Improved versions of resources have been created (e.g. Nerf – Polish named entity recognizer). • Better, more efficient implementation of existing tools (e.g. Pantera – a Brill tagger for Polish or Spejd – a partial parser, used for NP extraction, an improved multilingual anaphora resolution system – RARE). • Development of completely new tools (e.g. the Clause Splitter and the Discourse Parser, with adequate resources that make them operational for all 6 languages). • Server versions of tools have been prepared to improve overall efficiency by pre-loading of language models (e.g. grammar rules pre-compiled Spejd for Polish) or enabling concurrent execution (e.g. simultaneous processing of input file split into parts for Pantera and Spejd). 	<p>100%</p>
<p>Adherence to and promotion of existing and future web standards</p>	<p>Language Processing Chains have been successfully used to promote the UIMA framework.</p>	<p>100%</p>
<p>Practical and economically viable solutions for nearly-automatic provision of multilingual online content and services for some EU languages</p>	<p>For Publishing houses; Media and PR agencies:</p> <ul style="list-style-type: none"> - News Service Demo, http://newsdemo.atlasproject.eu/ - Video materials demo version, http://videodemo.atlasproject.eu/ - Press clipping service in Bulgarian and English, http://mediaimage.tetracom.com/ - Plagiarism service – TextMatch, http://textmatch.eu/ 	<p>100%</p>

Social Impact

Planned	Achievements	Extent
Facilitate exchange of information and knowledge	<p>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.</p> <p>While metadata identifies formal characteristics of a document like author, date and format, the automatic annotations represent its content. ATLAS processes multilingual data pools, analyses and streamlines them with semantic annotations. As a result the reader gains access to a layer of information which, though important, lies hidden from view:</p> <ul style="list-style-type: none"> • the reader spots information-dense entities like personal names, geographical locations, organisations, common phrases, short summary even before reading the text • the reader can search through these entities and reach all relevant content • the reader gains access to similar documents, pages and information recommended automatically by ATLAS instead of editors. <p>A new categorisation tree, containing 80 categories grouped in 6 top-level groups, enhances the users in cataloguing their documents automatically. Categorisation models for all project languages have been developed.</p> <ul style="list-style-type: none"> • A new categorisation tool provides the users with the functionality to split their documents into groups. • The multiple file upload enhances the users while uploading documents and as a result the user can upload all their documents into i-Librarian at one go. • The “public” library contains a collection of more than 22'000 books in English from Project Gutenberg. The books are processed, analysed and as a result enriched with revealing details like the most commonly used phrases, name entities, suggestion for similar books, short summary and are categorized automatically, according the library catalogue. 	100%

Simplify authoring, management and exploitation of heterogeneous multilingual content	I-Publisher Simple Mode. The complexity of i-Publisher, a powerful instrument for creating complex web sites, made it difficult for the group of inexperienced test users to work with it. As it is a wider and an important target group, The Consortium decided to extend the web based tool i-Publisher with simplified layer for non-experienced users. The users choose from ready-to-be-used websites or they create websites with predefined visualisation (themes) still having the possibility to define the content and its structure. As none of the existing now online services, some of the ready-to use websites provide the functionalities related with the linguistic framework so that the user can benefit from its applications like automatic annotations, automatic translation. The interface of the Simple Mode is localised in all project languages.	100%
Address the needs of a large number of people belonging to different target user groups – individuals and organizations	<p>For all individuals (librarians, researchers, IT experts, students, publishers, etc.) and organisations (NGOs, Libraries; Publishing houses; Media and PR agencies; International organisations; Corporate clients):</p> <ul style="list-style-type: none"> - i-Publisher Live demo: http://i-publisher.atlasproject.eu/atlas/i-publisher/demo - i-Librarian website, http://www.i-Librarian.eu/ - EUDocLib website, http://eudoclib.atlasproject.eu/ <p>For Publishing houses; Media and PR agencies:</p> <ul style="list-style-type: none"> - News Service Demo, http://newsdemo.atlasproject.eu/ - Video materials demo version, http://videodemo.atlasproject.eu/ - Press clipping service in Bulgarian and English, http://mediaimage.tetracom.com/ - Plagiarism service – TextMatch, http://textmatch.eu/ <p>For NGOs and international organizations:</p> <ul style="list-style-type: none"> - UNESCO Website Demo, http://unesco.atlasproject.eu/ - UNESCO Chair Website Demo, http://a2.atlasproject.eu:9090/unesco - World Health Organization Website Demo, http://asset.atlasproject.eu/who - Icon-painters from Bundovtzi family website, http://www.atlasproject.eu/asset_demo/icons/bg/index.html 	100%
Cross the language barrier Facilitate culture exchange	The linguistic platform ASSET employs Natural Language Processing tools, a categorisation tool, a summarisation tool and a machine translation engine. The produced services uses ASSET which facilitate the	100%

	<p>cross of the language barrier:</p> <p>TextMatch recognizes the language of the document using a two-stage language detection system. Specific language tokenizers, lemmatizers and other analysers are utilized for English, Bulgarian, German, French and Russian. For languages that are currently not supported, a language independent comparison algorithm is used.</p> <p>i-Librarian offers the intelligent way to browse, categorize and analyse your electronic books and documents in English, Bulgarian, German, Greek, Polish and Romanian.</p> <p>News Demo service categorizes and analyses the news in English, Bulgarian and Greek. All services using ASSET could be extended applying other languages.</p>	
<p>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.</p>	<p>Liaison with the Europeana and EuroMatrix Plus was established in order to foster language diversity in content creation and distribution.</p> <p>Tools and models developed within EuroMatrix are available by using related input data specification. The input between the MT-Serverland and ATLAS is established and integrated into the ATLAS-MT engine. A bridge to Euromatrix is integrated in the workflow of the ATLAS-engine.</p>	

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:

Impact of dissemination activities

Contribution of the dissemination activities performed by the ATLAS partners to the expected impacts by knowledge dissemination across organizations, target users and communities is presented in the following table:

Dissemination activities	Achievements
Production of	1 general ATLAS brochure

dissemination materials	2 marketing brochures for i-Publisher and i-Librarian Partners` brochures and leaflets 10 other publications to the general public Dissemination kit available
Doing research in the field	31 articles and papers, published in scientific and technical journals and conference proceedings to the Research Community
Development of demonstration cases for potential customers	13 demonstration websites - the main tool in the ATLAS exploitation strategy for targeting customers
Organising events to the Research Community	3 organised large events for Research community 10 posters, introduced at conferences and exhibitions for better presentation of the developed services, used technologies, present and further benefits, and widening the scope of the potential users.
Organising events to the Developers and Users Community /Working with specific target groups through organised user workshops	8 evaluation workshops with 112 users carried out in the Living Lab Collected feedback from users and experts, available on the LL website 38 user suggestions, collected though the LL website Feedback of 3 experts from 3rd round testing 7 organised ATLAS events within Living Lab targeting users and customers 4 pilot experiments carried within the Living Lab
Organising events to the potential customers	8 meetings with potential customers in Bulgaria and abroad
Giving presentations to other events	More than 62 presentation and demonstration made by partners at international events
Sharing experience within the communities	Online video tutorials, guides and technical documentation about i-Publisher and i-Librarian about 100 performed dissemination activities (18 exhibitions, 40 conferences and 40 workshops, seminars, meetings and lectures...)
Communicating within the network	ATLAS project website and Living Lab website for interaction with target users
Passing on the project information in the partners` institutions though seminars and lectures	13 organised seminars within partners` institutions with students and researchers
Exchanging idea and conducting discussions and interaction with other projects and networks in the field	Synergy with 11 relevant projects (META, METANET4U, Cesar, D-SPIN, CLARIN, PROMIS Lingua, OpenScout , ShareTEC, SmartBook , DYLAN, ELLIOT) and 3 networks (MUMIA, EUROPE Innova KIS and OpenAIRE) by organisation of joint events, presentations, demonstrations, exchange of ideas and discussion of future exploitation activities

Generating new projects	1 new project SmartCulture (approved)
Communicated ATLAS results with policy makers	2 presentations at UNESCO Chairs and Bulgarian policy makers: http://livinglab.itd-bg.eu/content/events

The whole list of the dissemination activities of the project is accessible at:

<http://www.atlasproject.eu/atlas/file/6644f85e-4ded-497f-bf7c-7f516f257467/13ccaaa0-a0c0-11e0-8264-0800200c9a66/D8.2%20Dissemination%20kit-1.txt>

More information including project details, news, and contact information can be found at:

www.atlasproject.eu

List of all beneficiaries

Part. No.	Beneficiary name, POC	short name	Country
1 - CO	Tetracom Interactive Solutions, Anelia Belogay CEO, Diman Karagiozov CTO, anelia@tetracom.com , diman@tetracom.com	Tetracom	Bulgaria
2	Atlantis Consulting SA, Polivios Raxis, raxis@atlantisresearch.gr	Atlantis	Greece
3	Institute for Bulgarian Language, prof. Svetla Koeva, svetla@dcl.bas.bg	IBL DCL	Bulgaria
4	Instytut Podstaw Informatyki Polskiej Akademii Nauk, Maciej Ogrodniczuk, maciej.ogrodniczuk@gmail.com	ICS PAS	Poland
5	Universitaet Hamburg, Cristina Vertan, cristina.vertan@uni-hamburg.de	UHH	Germany
6	Universitatea Alexandru Ioan Cuza, prof. Dan Cristea, dcristea@info.uaic.ro	UAIC	Romania
7	Institute of Technologies and Development Foundation	ITD	Bulgaria

ANNEX 1: WP description, deliverables and milestones and final achieved results

WP2 description

The launch of the three services was followed by an evaluation done by users within the first test round. The test groups of 33 users assessed the technical performance, the level of fulfilment of the system specifications and the level of user expectations fulfilment and provided Consortium with their feedback. As a result several major improvements were defined, discussed with the Commission and Tetracom started with their implementation:

WP2 implemented tasks

T2.1 – Specification of the linguistic framework

- A draft of the linguistic framework specification was prepared. It describes the framework for development of language processing chains in Work Package 4. Each partner is responsible for integrating their language tools within the Unstructured Information Management Architecture (UIMA), adjusting the tools if necessary. UIMA has been chosen for this project among other architectures (e.g. GATE), mainly because it allows language processing applications to be decomposed into components, which can be replicated over a cluster of network nodes. This enables UIMA to process very large volumes of data almost in real-time. Furthermore, UIMA manages the data flow between different components. The structure of this document is outlined below:
 - Requirements for all tools that will be integrated into the UIMA framework
 - Core set of annotations that the tools for all project languages must be able to produce
 - Naming convention for UIMA components
 - Details about the SVN repository for language tools
- The linguistic framework specification, was reviewed by the partners, indicated in the detailed project plan, and finalized. Changes in the document are possible and may occur once the integration of language tools into the UIMA framework starts.

T2.2 – Use case specifications

A set of detailed use cases for i-Publisher, i-Librarian, and EUDocLib was prepared.

T2.3 – Refinement of software requirements

The set of use cases was reviewed by the partners, indicated in the detailed project plan. Use cases were then used as a basis for the preparation of lists of functional and non-functional requirements.

T 2.4 – Software specification

- The high-level design document was prepared. It includes the following information:
 - General overview of the system
 - Functional requirements
 - Non-functional requirements
 - A set of diagrams providing an overview of the ATLAS architecture
 - A list of entities in ATLAS along with the various relationships between them

- The high-level design document was updated with a set of screens that demonstrate the visual aspects of i-Publisher, i-Librarian and EUDocLib. Screens provide only a visual clue and do not reflect the final outlook of the public services.
- The software specification draft was reviewed by the partners indicated in the detailed project plan and several change requests were sent to Tetracom. Changes were reviewed by the consortium.
- Tetracom modified the software specification in response to the change requests. The final version of the software specification is currently available in the SVN repository of the project.

T2.5 – i-Publisher – implementation and deployment

- The “i-Publisher” user interface to the ATLAS content management functionalities was implemented. It is based on the ZK open source AJAX framework. The new UI is secure, standard based, browser independent and provides rich user experience. The “i-Publisher” architecture is based on the Eclipse Equinox OSGi implementation, thus providing modular and service-oriented design of the application. “i-Publisher” and ATLAS CMS and all their components can now be replicated and deployed on different network locations using a transparent communication layer based on the ECF (Eclipse Communication Framework) OSGi bundles.
- Granular user and system access rights and security layers were implemented in ATLAS. All security-related features are based on the aspect-oriented programming paradigm.
- Tetracom upgraded the text mining engine of the English Language Processing Chain (LPC) to be based on the UIMA framework. ActiveMQ (and respectively the Java Message Service) is used for communication between different components in the engine.
- The English noun phrase extractor was improved in terms of quality and speed. Furthermore, a lemmatizer, syntactic parser and word sense disambiguation tool were added to the English LPC.
- Tetracom created an environment for testing the integration of language tools within the UIMA framework (a.k.a. WebCASDebugger).
- Tetracom built a computer cluster to test various software components including the text mining engine, the distributed databases HBase and GridSQL, and the statistical machine translation system Moses.
- The baseline summarization tool (Open Text Summariser) was integrated in ATLAS core and i-Publisher. ActiveMQ is used for communication between the summarization engine and the ATLAS system.
- The baseline multi-label multi-class categorization tool MULAN was integrated in ATLAS core and i-Publisher. ActiveMQ is used for communication between the automatic categorization engine and the ATLAS system.
- Test deployment of i-Publisher, i-Librarian and EUDocLib web sites was finished on 14 Feb 2011; i-Publisher productive deployment was finished on 26 Feb 2011.
- The launch of the three services was followed by an evaluation done by users within the first test round. The test groups of 33 users assessed the technical performance, the level of fulfilment of the system specifications and the level of user expectations fulfilment and provided Consortium with their feedback. As a result several major improvements were defined, discussed with the Commission and Tetracom started with their implementation:
- I-Publisher Simple Mode. The complexity of i-Publisher, a powerful instrument for creating complex web sites, made it difficult for the group of inexperienced test users to work with it. As it is a wider and an important target group, Tetracom decided to extend the web based tool i-Publisher with simplified layer for non-

experienced users. The users will choose from ready-to-be-used websites or they will create websites with predefined visualisation (themes) still having the possibility to define the content and its structure. As none of the existing now online services, some of the ready-to use websites will provide the functionalities related with the linguistic framework so that the user can benefit from its applications like automatic annotations, automatic translation. The interface of the Simple Mode is localised in all project languages.

- Stability and performance optimizations. I-Publisher and the underlying language processing chains framework were optimized in terms of performance by integration of caching components on different levels of the system architecture. The web sites, powered by ATLAS, were tested with the Apache HTTP server benchmarking tool.
- New widgets – based on the user feedback and change requests Tetracom has implemented new content clustering widget, based on the Bisecting K-Means algorithm; text-mining status widget, showing the process in the language processing chains; archive import widget which improves the productivity of the users when entering a lot of content items.
- Import/export from a website. In order to increase the efficiency while creating a website, Tetracom implemented functionality to import/export the definition of a website. This way, once created, a website and its functionalities can be reused.
- Static website export. Based on a user request, Tetracom implemented a static website export. Thus, an ATLAS web site can be statically exported and deployed on an arbitrary web server.
- OAI-PMH ORE import. The work to support an import of data from other repositories through OAI-PMH ORE standard has been started. The requirements, functional and design specifications are ready.
- 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 evaluation 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.).

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 implementation of i-Librarian was done entirely with i-Publisher as prove of concept. Building the i-Librarian service required implementation of several new functionalities of i-Publisher, such as creating a login widget, improvement of the navigation widget, possibility for the user to create their own categorization tree in order to organize the added content.
- The user adds content in i-Librarian and as a result gets structured, processed, annotated, summarized and categorized content items which are accessible online. In addition, for each entity, the user is provided with a list of similar entities available in the library.
- The description, important phrases and the summary for every library item can be translated in all EU languages.
- Users access the content of the library after successful registration, confirmation of the terms of the service and login. Users can share entries with other readers of the library but not with general public.
- The noun phrase extraction and the name entity extraction in English were significantly improved by using the updated versions of the noun extractors and name entities service.
- A new categorisation tree, containing 80 categories grouped in 6 top-level groups, enhances the users in cataloguing their documents automatically. Categorisation models for all project languages have been developed.
- A new categorisation tool provides the users with the functionality to split their documents into groups.
- The multiple file upload enhances the users while uploading documents and as a result the user can upload all their documents into i-Librarian at one go.
- The “public” library contains a collection of more than 22'000 books in English from Project Gutenberg. The books are processed, analysed and as a result enriched with revealing details like the most commonly used phrases, name entities, suggestion for similar books, short summary and are categorized automatically, according the library catalogue.
- 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 2.7 – EUDoLib - implementation and deployment

EUDoLib is a publicly accessible repository of EU law documents from the EUR-LEX collection. This web site (online service) provides enhanced navigation and easier access to relevant documents in the user's language.

- The implementation of EUDoLib was done entirely with i-Publisher.

- The algorithms used for full-text searching and finding similar items were improved in terms of performance and now they can work within a large pool of documents.
- All documents of EUR-LEX collection, available in all project languages, added up to the 1 January 2011, have been imported into ATLAS (resp. into EUDocLib), processed, analysed and as a result enriched with details like the most commonly used phrases, name entities, suggestion for similar books. The work for finding optimal categorisation models for all documents in EUDocLib has started. TetraCom identified several challenges based on the content of the EUDocLib corpora - there are more than 140'000 documents with various size manually categorized in three different categorization sets - top-level EuroVoc with ~ 665 labels, leaf-level EuroVoc with 5'871 labels and Directory Codes with 475 labels. We approached the first challenge of having various document sizes by clustering documents with similar sizes (in terms of tokens) and building sub-models. The next challenge is related to the big training data and the task to find an optimal set of training data. We decided to pick training documents that featured fewer categories as we assume they are specific for this category. A technical problem, already solved, is the amount of heap space needed to prepare the feature space. We employed high-performance, low-memory-footprint Java collections and caching mechanism in order to calculate the feature space in a timely manner. Still, we had to split the biggest categorization set in three smaller sets. Splitting was done by clustering the documents in groups, based on their relation to the categories.

T2.8 – Technical documentation

TetraCom started writing detailed technical documents describing the ATLAS platform and the online services. Documentation will be made publicly accessible on the project web site. ATLAS technical documentation draft, i-Publisher technical documentation draft and i-Librarian and EUDocLib technical documentation drafts will include a thorough description of the platform architecture and design, and will emphasize the multilingual capabilities of the software.

T2.9 – User guides

TetraCom prepared three level guide system for i-Publisher and i-Publisher Simple Mode. The comprehensive hint system is ingrained in i-Publisher Advanced Mode. Video films complement the documentation and the hint system by enhancing the user to perceive the main complex workflows in i-Publisher and to get the general idea of creating a website in i-Publisher Simple Mode. The major i-Publisher workflows are described in the “Atlas quick user guides”.

The digital library i-Librarian publishes on the main page a video guide providing the users with the detail information on how to use the features of the system.

WP2 deliverables and milestones

The three pilots within the two deliverables D2.1 i-Publisher, D2.2 i-Librarian and EUDocLib were implemented and the M1 was reached.

Final results

A multilingual platform for creating websites ATLAS was created. The powerful SAAS layer i-Publisher Pro builds complex websites while the important target group of inexperienced users use i-Publisher Simple Mode. The users choose from ready-to-be-used websites or

they create websites with predefined visualisation (themes) still having the possibility to define the content and its structure. As none of the existing now online services, some of the ready-to use websites provide the functionalities related with the linguistic framework so that the user can benefit from its applications like automatic annotations, automatic translation. The interface of the Simple Mode is localised in all project languages.

Two website prototypes were created with the i-Publisher Simple Mode: the website for the ACL 2013 that will be held in Sofia: www.acl2013.org and for Gracher group www.gracher.de

Several demo services were implemented with Atlas. They are used as dissemination instruments to attract the Atlas target groups of International organizations with multilingual information flow. The demo services are:

- A demo service for UNESCO - <http://unesco.atlasproject.eu/>
- A demo service for World Health Organisation - <http://asset.atlasproject.eu/who>
- A demo service for ISN at ETH, Zurich

The demo services demonstrate the combination of multilingual content management and semantic technologies that enrich the content.

The enrichment of content with metadata provided by linguistic analysis is of a great importance for the media companies as they can offer better view over their content. Several demo services demonstrate the powerful linguistic tools:

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

A prototype of a press clipping platform MediaTalk was implemented in Bulgarian and English and can be accessed at:

<https://mtalk.eu>

<https://en.mtalk.eu>

WP3 description

The focus of WP3 is implementing a tool that categorizes texts in the target languages and that can be tuned to heterogeneous domains. The ultimate goal of the automatic categorisation is a qualitatively better and more effective organization of multilingual content that is managed in ATLAS and respectively in the online services.

WP3 implemented tasks

T3.1 Prototype and integration

- Three automatic categorisation algorithms are supported by the tool– Naive Bayesian, Relative Entropy and Class-Featured Centroid. Three cascade dimension reduction techniques were implemented – best-N (by tfidf), prune (below and above) and Chi2.
- TetraCom developed a model evaluation tool which compares the results coming from the automatic categorization tool to the manually categorization suggestions for a set of test documents. Several experiments have been conducted to study the categorisation performance in relation to the type of features and reduction algorithm used for vector

space generation. The following features types were used – token based, lemma based, noun phrases based, head token based.

- Tetracom extended the ATLAS categorisation framework with the following changes:
 - The categorisation framework now supports several instances of categorisation algorithms, each instantiated with different feature type.
 - Support for ensemble of categorisation algorithms. Multiple models can be used to obtain better predictive performance.
- Tetracom built an English categorisation tree for i-Librarian service, based on a model for 80 categories. The newly added content is automatically categorized according to the tree. It is accessible on the website ('Topics' categorisation tree) of i-librarian.
- Tetracom implemented a new "Clustering" categorisation tool that splits the uploaded documents in groups depending on their similarity. This tool is based on not-supervised learning Bi-KMeans categorisation algorithm and is used to organize documents when there is no existing categorisation tree to be used for building a categorisation model.

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

Training and Test documents for the 80 categories were delivered by all partners and Tetracom worked on processing the data and building the models for all project languages.

The optimal results have been achieved with head nouns features and using substantial set of features (90%). The highest precision was achieved when using all tokens, however the models size and performance is not suitable for online services, such as i-Librarian. Nevertheless, the work on this task continues with integration of LDA, kNN and SVN algorithms.

- 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_Q7c1mudExmNVJBMHgyTXdaeWhiTGhlcEZIX1E#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)

This task is ongoing. Tetracom implemented part of the UniZD Python tool algorithms in Java to benefit from Python performance and Java maintenance easiness. Furthermore, high-performance and low-memory collections have been used in order to accommodate and process the huge models related to the EUDocLib categorisation sets and hundred thousands of training documents.

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 finding 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 and 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

WP3 deliverables and milestones

Deliverable D 3.1 “Categorisation Tool” was delivered to the Commission.

Final results

The categorisation tool is integrated in ATLAS platform. A categorisation tree with 80 categories categorises the users content in the library i-Librarian. The content of EuroLex in the project languages is processed and analysed and categorised according to the classification system of EuroLex. The press clipping prototype MediaTalk uses the categorisation algorithms to suggest the similarity lists.

WP4 description

The following linguistic tasks have been achieved throughout the project:

- the uniform annotation model has been defined,
- the implementation framework and architecture have been selected,
- the environment for implementation and testing of the adaptation, integration and chaining of existing language tools for each project language has been established,
- test documents in different formats and application domains have been gathered,
- common pre- and post-processing tools have been integrated into ATLAS platform,
- the implementation of the LPCs for all project languages, targeted at providing tools for sentence-splitting, tokenization, lemmatization, POS tagging, NP extraction and NE recognition have been completed,
- the LPCs have been integrated into ATLAS,
- internal verification of the chained tools has been carried out,
- LPC testing environment has been developed in order to measure the performance and stability over time of each individual LPC,
- necessary improvements have been introduced after software reviews,
- regression test environment has been introduced after the review meeting,
- documentation tasks (selection of tools, their properties, interfaces, tagsets, preparation of test plan template etc.) have been completed,
- final D4.1 deliverable document was delivered to the Commission.

WP4 implemented tasks

T4.1 – Implementation of LPCs for the project languages

Bulgarian

At the beginning of the project, a report for Bulgarian Language Processing Tools was elaborated. It contained a brief description of the DCL IBL language processing tools (main functionalities, programming languages, requirements, as well as copy rights).

The DCL IBL made the prototype integration of most of the Bulgarian tools (namely Bulgarian tokeniser, sentence splitter, tagger and lemmatizer) with WebCASDebugger, a visual test interface for running and displaying UIMA annotations on different linguistic levels.

The Linguistic Rule Engine has been developed. It provided a powerful language for the linguistic knowledge representation. The system consists of three major components: the lexicon compiler (it compiles large dictionaries in a single minimal acyclic finite state automata and provides very fast lookup with linear time complexity), the rule compiler (it compiles various types of linguistic rules in finite state transducers), and the rule processor (it applies the lexicon and the transducers over the input data). An NP grammar processed with the Linguistic Rule Engine is itself the NP recognizer for Bulgarian.

The 40 test documents representing the 13 domains specified, with different length, format and encoding have been selected for Bulgarian.

A cascade multi-purpose parser, the ParseEst, that can be of benefit to annotation of different types of syntactic entities: phrase structures, named entities, multi-word expressions, etc., thus to be re-exploited in the implementation of different tasks – i.e. syntactic parsing, NP extraction, named-entity recognition, etc., has been developed. The ParseEst consists of a rule-builder module (lr_builder) and a rule-engine (lr_engine) module. The lr_builder module supports unification and cascade rules. Both the lr_builder and a lr_engine operate with the compiled lexicons from the lexicon builder. Two groups of language rules have been written in order to feed the Bulgarian NP extractor and Bulgarian NER recogniser with the appropriate grammars. The partial parsing approach which was adopted for Bulgarian exploits the finite-state framework to approximate a simplified context-sensitive grammar.

The tools are designed to work together in a language processing chain, for example the sentence boundaries and the token tags are used by the tagger, the tagger's annotations – by the lemmatizer and so on. After the Bulgarian tokeniser, sentence splitter, tagger and lemmatizer the prototype integration into the UIMA framework of the rest of the Bulgarian tools (Bulgarian shallow parser ParseEst (lr_engine and lr_builder), Bulgarian NP extractor, Bulgarian NER extractor, Bulgarian word sense disambiguator and Bulgarian stop list recogniser) has been made. The components are integrated in the UIMA as external components which are invoked by a Java class which represents a UIMA Annotator. The Java class reads the CAS structure and prepares the input data for the components. For each tool there is a UIMA Annotator. Also a UIMA aggregate engine is provided, which aggregates the tools in following order: sentence splitter, tokenizer, tagger, lemmatizer, NP extractor, NER extractor. Each of the language tools can be included as primitive engine, i.e. part of an UIMA aggregate engine, but also as an aggregate engine. In this way any component of the language chain can reuse results produced by a particular tool and exploit its full functionality if required.

The tagset was redefined to integrate agreement between the lexical-grammatical and grammatical categories. Both reduces and full tagsets are supported which required the respective changes in the UIMA part.

The tagger was rewritten to solve some memory management issues, to improve performance and to get rid of the SVMTool libraries. The tagger is able to work in a server mode, it gets data from a TCP port thus the tagger model is not loaded every time when the tagger is invoked. The NP and NER rules and lexicons are tested and further specified.

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. 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 actively contributed to technical management of the WP, particularly by the creation of a testing environment for language processing chains (with demo WebCASDebugger application), involvement in the internal software reviews and implementation of the LPC for English.

Croatian

UniZD initially provided a set of tools for Croatian (tokeniser, sentence splitter, NER tool), but finally it was agreed with the project officer that the Croatian tools will not be implemented/integrated in ATLAS.

English

Initially, an OpenNLP-based language processing chain has been created by Tetracom. In the course of the project the Open NLP noun extractor was replaced by a new one. The new extractor is based on the theory of finite-state language processing and contains set of language-specific rules for recognition of noun phrases described by an English linguist. The NP extractor can be used for other languages as well.

Several updates in the architecture of ATLAS were needed in order for the JRC Names Library to be integrated into Atlas and as a result we provide the user with name recognition covering wider set of person names and organizations. The JRC Names UIMA wrapper is now used in other project Language Processing Chains.

Improvements in the NERs have been made for English. OpenNLP-based NER was 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.

German

UHH adapted the German OpenNLP to be used with the ATLAS framework and Tetracom 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.

UHH tested the GATE components, improved the MUnPEX NE recognizer with lexical resources from the CDG, especially a list of German first names, last names and companies, improved the NP chunker with rules for concatenating NPs and provided wrappers for sentence splitter, tokeniser and PoS Tagger (TnT).

Greek

Atlantis implemented and tested a set of the LPC tools for Greek: sentence splitter, tokeniser, POS tagger, stopword recognizer. A noun phrase extractor has been implemented basing on the Polish Spejd tool. The implementation of the first prototype of the lemmatiser and Named Entity Recognizer for Greek has started and is ongoing.

UIMA compliant primitive engines created for sentence splitter, for tokeniser, for POS Tagger, for NER and the Lemmatizer have been prepared by Atlantis. All Greek primitive engines were revised according to the updated UIMA types and they were re-factored and optimized to support multi-core (over than 30% improvements have been achieved). Encoding was changed from ISO-8859-7 to UTF-8. Debugging and certification testing with input texts from various domains has been carried out.

Greek sentence splitter has been optimized in terms of loading only once the SVM classifier and other critical objects with high memory requirements.

Greek Named Entity Recognizer has been extended with regular expressions and supports names (person, organizations), locations, date, time, money and percentage expressions. It now takes into account the MUC-7 Named Entity Task Definition and integrates the JRCNamer engine. It has been optimized in terms of loading (through absolute or relative paths) the SVM classifier and other critical objects with high memory requirements only once.

Greek lemmatizer has been enhanced by adding more lemmas from printed dictionaries, adding support for MySQL (back-end repository) and for multithreading. The implementation has been enhanced by incorporating a cache approach for the lemmas included in the database (integrating the EhCache open-source tool).

Greek POS tagger has been improved to preload objects and classification files. Tokenization has been improved by treating Greek special characters and adding additional training data from manual annotation.

An improved version of Greek Noun Phrase extractor has been provided with enhanced morphological rules have been enhanced (more than 6 now are available and tested; 2 rules for identifying nested phrases are pending verification and deployment in relation to Anaphora Resolution) and markable code added to denote the HEAD of the noun phrase. Special characters contained in the input, are now treated through an “escaping” mechanism.

Greek Anaphora Resolution based on the RARE tool provided by UAIC has been provided

Pre-processor resources for the Greek Language have been prepared (e.g. Family Names, Pronoun Constraints, markers, etc.).

Polish

ICS PAS actively participated in the creation of internal documents and plans related to implementation of language processing chains, starting with linguistic framework in the form of UIMA properties (document annotations, text annotations) and descriptions of tools to be integrated for all project languages.

ICS PAS adapted the existing language processing tools for Polish to ATLAS and UIMA frameworks by adoption of TaKIPI (tokeniser, lemmatizer, sentence splitter and POS tagger) and Spejd (NP chunker). Prototype stopword filter and list-based NE recognizer have been integrated with the platform to test the annotation model and prevent later implementation issues. The tools have been integrated with WebCASDebugger, a visual test interface for running and displaying UIMA annotations on different linguistic levels

In order to coordinate WP4 work, ICS PAS created the implementation plan with deadlines for integration of certain types of tools and milestones for evaluation of integration status. Following this plan, ICS PAS and Tetracom carried out internal reviews of tools forming language processing chains. A template for registering test documents for the chain has been prepared and distributed among all partners along with a test set for Polish (containing 30 documents in various formats and encodings, corresponding to many domains). ICS PAS prepared the first internal draft version of D4.1 deliverable integrating descriptions of tools and testing data sent by partners.

Spejd-based multiword NP lemmatizer has been implemented from scratch. NP lemmatizer was evaluated on a manually annotated small corpus of multiword expressions. The Polish form generator (the last part of MWU lemmatizer needed to deal with adjectives and participles) and Named Entity Recognizer have been integrated and installed in WebCASDebugger.

Existing statistical named entity recognition tool by Jakub Waszczuk (NERF) was retrained to reflect ATLAS NE model (by supplementing it with ATLAS types not available in the original version – Money, Percentage) and integrated into the platform.

Romanian

Initially, UAIC took part in integration of RACAI Web service into the ATLAS framework. An NP-chunker has also been implemented using the Nooj technology. Following this, UAIC implemented its own POS tagger and Noun Phrase extractor tools. This tools have been

integrated into the ATLAS framework with the help of TETRACOM and are part of the Romanian Language Processing Chain.

The Romanian Language Processing Chain consists of a sentence splitter, a tokenizer, a POS tagger, a lemmatizer, a noun phrase chunker. After the integration task was done, certain problems were corrected (a special case in using a „.” misinterpreted by the tokenizer, „i.Hr” which means before Christ, misinterpreted by the sentence splitter etc.) Another problem was fixed for the badly formatted texts from the EUROLEX corpora. For instance, at some point appeared an enumeration of people’s names (over 1000 names), and one of the names had the „de” in it, which in Romanian is a preposition, while also a defining marker of a noun phrase. The noun phrase chunker, implementing a recursive grammar, either stopped working because of the immense number of matches it found, or crashed. The solution was to impose on the noun phrase chunker a limit over the words it contains. After careful testing a limit was imposed on the possible length of a noun phrases, which, if reached, the on-going development of a noun phrase is cancelled.

Another case of error of the noun phrase chunker was detected when dealing with a text from the EUROLEX corpora. The problem was that a sequence of numbers was chunked as a sentence, and because of the recursive nature of certain NPs (including sequences of numerals), a partial match of a noun phrase was found. However, the sentence ended before meeting a final noun, which made the chunker to enter into an infinite loop. This was fixed by changing the grammar.

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

Although WP4 activities in the project were planned for months 5-23 only, several additional issues have been completed until the end of the project, especially following the project review in 2012, i.e.:

- 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.
- 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 in order to increase the quality and recall of the extracted locations.
- Several NER engines have been combined in order to improve the quality of the extracted person names and organizations.
- Freebase person name NER has been implemented using lexicons built from the Freebase lists of people.
- BOM UTF-8 text extraction bug has been found and fixed.
- Bug fixes and minor improvements in all LPCs – EN POS tagger and lemmatizer, EL and PL NP extractor, BG NER, DE stability (see more details below).
- Extensions in the LPCs to provide richer linguistics annotations for the summarization tool in WP5 – e.g. gender information for nouns, improvements in NP annotation etc.

Bulgarian

The existing language processing tools for Bulgarian have been adjusted in the following way: the BG tokeniser and sentence splitter was rewritten in C++ using the boost library. The

finite state automata tools (compiling the large dictionaries into acyclic finite state automata and traversing the automata for fast look up) were rewritten in C++. The BG lemmatizer components that do not coincide with components integrated in other tools were rewritten in C++ as well. Each tool is implemented as a system daemon, sockets are used for the interprocess communication. Tests over the performance of the BG tagger have been developed leading to the respective optimization – the improvement of the memory footprint and the performance.

Tests over the performance and precision of the Bulgarian tools have been developed leading to the respective optimisation. The existing language processing tools for Bulgarian have been adjusted to improve the performance of the Bulgarian sentence splitter and tokeniser. Both the precision and performance of the Bulgarian tagger have been improved as well. The precision of the Bulgarian lemmatizer has become higher.

The lexicon builder component of the ParseEst was rewritten from Objective C to C++ fixing at the same time some bugs in lexicons support both in `lr_builder` and `lr_engine`.

The whole LPC as well as the LPC integration was extensively tested with a very large amount of data. The ParseEst tool was extensively tested and respectively improved as well.

German

UHH adapted the German OpenNLP and integrated The German Constraint Dependency Parser (CDG <http://nats-www.informatik.uni-hamburg.de/view/CDG/WebHome>) within the ATLAS framework. It includes a PoS Tagger, Lemmatiser, NP Chunker and NE recogniser that were adapted for ATLAS purposes. The components were isolated into an independent working module to eliminate deep parsing processes that are not needed in ATLAS.

Improvements in NE (geographical locations and NE) and in extraction of nested NPs have been implemented.

Greek

Stability improvements have been introduced to the Greek LPC. Grammatical rules have been designed and defined in the form required by the Spejd. Already implemented primitive engines for Greek have been integrated into the UIMA overall ATLAS framework.

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 as well as 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 in terms of execution time).

Polish

The LPC for Polish was further improved and updated with the most recent version of the language processing tools. To improve performance, a parallel execution of linguistic processing for split data has been implemented.

Data that could be used for NE recognition have been retrieved from other Polish NLP centers in Wrocław and Poznań and the lists of available NE resources has been collated (with 4 resources developed at ICS PAS and 2 external resources) to extend the form base of an existing prototype of Named Entity Recognition tool which was investigated and is going to be included in UIMA framework for Polish. A new mechanism for named entity recognition was integrated with the chain (using the Wikipedia- and GeoNames-based dictionary).

Extending the model suggested to all other partners, ICS PAS implemented the multiword NP lemmatizer (to be included in the NP chunker) for Polish. As compared to the initial version of the MWU lemmatiser, the current version contains further improvements concerning correction of gerund forms and negated nouns. 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 performance of generation of gerundial forms has been improved.

Pantera tagger output was corrected to split the MSD codes into separate POS, gender, number and case attributes.

Spejd shallow parser used for NP recognition was reimplemented in C++ to improve its efficiency. Server version of Spejd (with the rules compiled only once, not once for every input file) was implemented and integrated. Technical problems with compiling and integrating the new version of Spejd tool have been solved. Presentation of Polish NPs has been improved by removing NPs containing pronouns classified as adjectives.

The efficiency of the Polish LPC was improved by reimplementing the integration by making both time-consuming primitive engines (Tokeniser/Sentencer/Tagger Pantera and Noun Phrase Recogniser/Lemmatiser Spejd) work simultaneously on two parts of the split input file (for large files only). Provided that the machine has at least two cores (practically always nowadays) this parallelization makes the whole process faster by 30-40%.

Romanian

The Sentence Splitter has been updated in several steps (for instance, it experienced a segmentation problem on sentences ending on capital letter words).

The Romanian LPC has been improved to avoid sending sequences of characters that were treated erroneously by the NER component (a component which is part of GATE). A filter was placed at the very beginning of the LPC. The whole LPC now features higher stability. Several other issues related to the NER have been fixed (e.g. regarding regular expressions, which in some cases entered an infinite loop and hanged the system). We fixed the `.jape` grammar rule files to have a maximum recursive level; also work was done on the gazetteers to improve the recall of the NER.

The POS tagger was updated at a newer version and, apart from making it faster, it accepts now sentences longer than 1000 words (before this enhancement, the POS tagger working over sentences longer than 1000 was either very slow or even hanging). Also, the POS tagger and the Noun Phrase Extractor have now a better time response.

The NP Chunker was updated at a newer version that implements a bigger grammar, thus recognising more types of NPs.

T4.3 – ATLAS LPC support extension

ATLAS LPC storage engine and the underlying database have been extended to support multilingual content in UTF-8 format. Significant challenges have been solved for the Greek and Romanian language processing chain because most of the internal models were not UTF-8 compliant.

Being a by-product the Language Processing Framework has not notation of the content model and the available languages in ATLAS. Thus, ATLAS was extended to detect and use the language of each uploaded file.

T4.4 – Integration of LPCs in ATLAS

All linguistic tools delivered by partners have been integrated in the ATLAS as external processes invoked by UIMA components, which handle the UIMA types and data transfers between components.

The variety of technologies and platforms used in each language processing chain resulted in a great integration challenge. Atomic NLP tools based on JAVA, C/C++, Perl and Python tools, GATE components and pipes has been integrated and harmonized.

Significant efforts have been made to not only to integrate by to optimize the performance of certain components and whole LPCs. Tetracom has developed parallel processing patterns and constructions to be used in all language processing chains. Furthermore, the post-processing of the annotated data was optimized by using a hybrid data store – PostgreSQL database and Lucene indexes. This decision was taken when the initial approach, based purely on a relational database, reached a point where the overall performance and maintenance of the database was not satisfactory. More information on the fusion between the RDBMS and Lucene can be found at:

- http://www.atlasproject.eu/atlas/documentation/chapter_details_v2/en?k_211a4d10-d142-4e4d-98fe-b0caa7426130=715282a7-1d42-45c8-b955-aa568b0ddb0d&kl=8.database%20structure&ci=642bcc70-401d-4fa4-8fab-5e8891caf81d
- http://www.atlasproject.eu/atlas/documentation/chapter_details_v2/en?k_211a4d10-d142-4e4d-98fe-b0caa7426130=715282a7-1d42-45c8-b955-aa568b0ddb0d&kl=8.database%20structure&ci=93f588c7-740c-4107-8045-15711e0f4fd4

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

D4.1 deliverable document has been collated from partners' input and delivered to the EC on January 4, 2012. Report on integration test results has been prepared basing on performance data made available by Tetracom.

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.

WP4 deliverables and milestones

The deliverable D4.1 – Language Processing Chains (well-documented language processing chains for German, Bulgarian, Croatian, Greek, Polish, and Romanian; the documentation includes detailed testing results and a short guide for integration and usage) was delivered on PM 22 to the Commission.

Milestone 2: Public launch of the online services with support for all project languages (signifying the integration of language processing and text categorization modules for all project languages) has been achieved in M24.

Final results

Language Processing Chains provided the 3 online services (EUDocLib, i-Publisher, i-Librarian) with linguistic capabilities, which was the main goal of the project and the major improvement offered by ATLAS Web content management system.

i-Librarian now features a full range of language-powered features, including extraction of important phrases, named entities (locations, person names, organizations etc.) and linguistically similar documents (content items interlinked based on text annotations). EUDocLib enriches legal EU content with SIPs and similar documents.

Tools created by WP4 provide the most important part of the functionality offered by these services.

Description of the WP5

This work package focused on adjusting and fine-tuning existing software components and producing summarization tools for the project's target languages. At the beginning of the project the consortium had to take some decisions regarding the summarisation method. UAIC was known to have achieved good results with a summarization method based of discourse-parsing, which allows for both general and focused summaries. The consortium decided to implement this method in ATLAS. The summary is obtained at the end of a number of processing steps (Cristea and Postolache, 2005), as follows: the text is first segmented into elementary discourse units (mainly clauses), then for each sentence a discourse tree is composed based on cue-phrases recognizable by a parser. Following, the sequence of sentence trees is arranged into one big tree for the whole discourse, by maximizing a score contributed from Centering transitions (Grosz et al., 1995) and anaphoric links, as proposed by Veins Theory (Cristea et al., 1998). On the global structure tree, any type of summaries can be computed effortless. The method can be applied to any language provided the following processing tools and resources are available for each language: a

sentence splitter, a collection of discourse markers and (optionally) an anaphora resolution tool.

The overall summarization system is truly multilingual in the sense that it first detects the language of the text and subsequently commutes to the specific language processing chain. Apart from small variations, all language versions have a similar design, as displayed in Figures 1, 2 and 3.

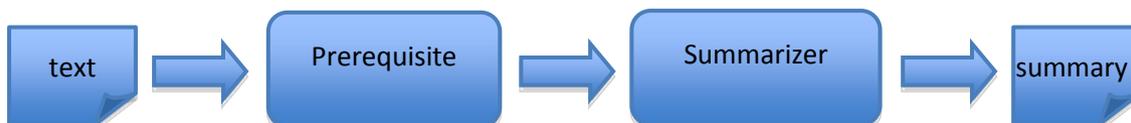


Figure 1: The block architecture of the summarization process

The Prerequisite part is a basic language processing chain (LPC), which includes the steps usually steps usually needed in many applications. In

Figure 2, this is indicated by placing the modules in a pipeline, although slight variations of this chain could be effective in different languages, depending on the proper realization of the envisioned functionalities.

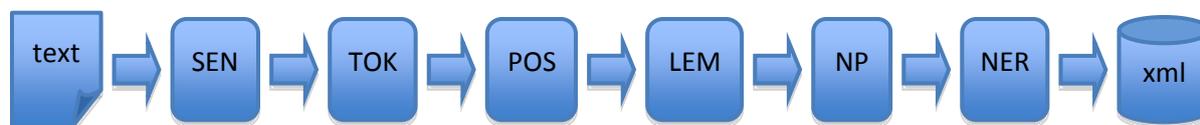


Figure 2: Details of the summarization Prerequisites

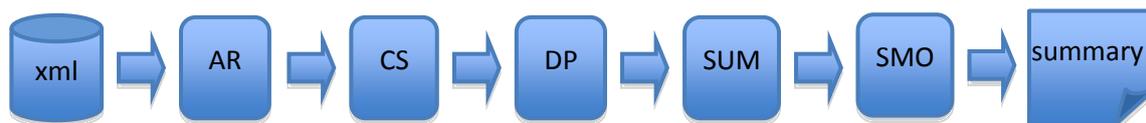


Figure 3: Details of the proper Summarizer chain

The abbreviations in Figures 2 and 3 have the following meaning: SEN = sentence splitter, TOK = tokenizer, POS = part of speech tagger, LEM = lemmatizer, NP = noun phrase chunker, NER = name entity recognizer, AR = anaphora resolver, CS = clause splitter, DP = discourse parser, SUM = summarizer, SMO = smoothing module.

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

- All partners have acquired their corpora to be used in training and evaluation, as follows:
 - o A corpus including clause boundaries and discourse markers (Clause Boundaries and Markers Corpus – CBMC), used by the Clause-splitter module and the Discourse-parser module;

- A summaries corpus (Summaries Corpus – SC), including minimum 3 versions (produced by different human subjects) of summaries of short texts (minimum 10 texts per language), used in the calibration and evaluation of the Summarisation modules.
- The summarisation processing chains, in which the language independent modules were coupled with specific resources, have been tested off-line, first by the WP5 responsible partner (UAIC) and then by each partner.
- A new version of RARE (Robust Anaphora Resolution Engine) was produced, with a much friendlier language localisation interface:
 - all information depending on language are now declarable in a resource file.
- A Multilingual Sentence Segmentation module, able to segment sentences in clauses, trained on manually segmented documents, was built. It incorporates 4 modules:
 - a Training Module extracts a model from a manually built corpus, which will be used by the Segmenter module;
 - a Segmenter Module segments the input, by applying pattern matching and machine learning algorithms;
 - an Evaluation Module evaluates the test file against the gold file;
 - features used by the Training and the Segmenter modules are configurable, so we also developed a Calibration System with the purpose to determine the feature and their values that helped to achieve the best results.
- A Multilingual Discourse Parser was built. The method is based on Veins Theory and combined an incremental type of parsing (by applying two operations taken from LTAG (Joshi and Schabes, 1997) – adjunction and substitution) with a method based on combinatorics (Cristea et al, 2003). Several heuristics were implemented for weighting the resulted discourse trees.
- A graphical interface, called MarkersEditor, was built to visualise and edit the markers_type.xml files, used by the discourse parser.
- A graphical interface (DrawTrees) was built for a handy visualisation of the computed trees at the level of each sentence or at the level of the whole discourse.

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

When errors or malfunctioning were reported, UAIC, assisted by the partners, have solved them. This activity proved to take more time than expected, and necessitated a permanent communication between UAIC and each partner responsible for a particular language. The improvements were addressing the language independent modules of the chain, while others were addressing the language specific resources. Each time a language independent module was changed, the positive tests already done for some languages had to be passed again for consistency checking.

The integration phase was followed by another testing/evaluation phase in which each integrated chain had to pass a beginning-to-end complete test and the results of the summarisation were compared against the manually annotated files. During this phase the partners actively interacted with UAIC and TETRACOM – the partner responsible for integration. At this moment, the evaluation included also a comparison against the state-of-the-art achievements. The deliverable D5.1 describes in details all these results. In general, the ATLAS summarisation system is superior to some well-known systems. And in any case, to our knowledge, there is no other system capable to do summarisation for 6 languages, in an integrated framework, using similar architectures and powered by the same theoretical model and performance mechanisms.

In some cases, bugs were reported and UAIC (if related to modules), TETRACOM (if related to the integration) or even partners (if related to resources) had to solve them, in close interaction.

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

At the moment when each module reached certain stability, as revealed by the off-line (decoupled) tests, a number of 6 language-specific chains integrated in UIMA were produced. The responsible for this task was TETRACOM, which worked in close correlation with UAIC – the responsible for WP5, and each partner.

T5.4 (13 PM) – Documentation

UAIC was in charge of writing a detailed documentation describing the summarisation chain. The language specific information was contributed by all partners. Two documents were produced: the D5.1 deliverable and a paper (Anechitei et al., 2013) submitted for publication as a chapter of a Springer Verlag volume.

The “Partner’s manual for building corpora” was written by Dan Cristea, Daniel Anechitei and Eugen Ignat.

Supplementary to the planned activities, the development and testing phases revealed the necessity to add also the following language-specific resources:

- a RARE localisation resource, including a set of anaphora resolution rules;
- a list of MSD tags for verbs, used by the Clause-splitter module to classify verbs into elementary and compound (asking for a following compulsory verbal role);
- an XML file containing the MSD tags of nominal categories (common noun, proper nouns, pronoun) – used by RARE and the Discourse-parser, and of the punctuation marks (used by the smoothing module).

List of the deliverable and milestones for WP5

Deliverables:

D5.1 Text summarization tools (M30). Leader: UAIC (R, PU, 124 PM). Well-documented text summarization tools for Bulgarian, English, German, Greek, Polish and Romanian. The documentation includes a description of the implementation approach, detailed testing results, and a short guide for integration and usage.

Milestones:

M3 Public launch of the final versions of the online services (M32). Leader: TETRACOM. Milestone 3 designates the integration of text summarization, machine translation and cross-lingual search modules for all project languages. The final public versions of ATLAS and the online services are released..

WP 6 implemented tasks

T 6.1. Fine tuning of existing translation models

For most part of the involved language pairs there was no translation model. In order to ensure a homogeneous implementation (and consequently translation accuracy) over the language pairs, we performed first a series of tests with selected language pairs, namely DE-EN, DE-RO and RO-EN. In these experiments we investigated:

- How good corpus-based approaches are working reasonable with small corpora and Which is the minimum size of training materials needed in order to ensure reasonable translations.
- How robust is the Moses system to domain change as well as text genre?
- The influence of tuning (a very time consuming operation) in the translation quality
- The impact of PoS-factored models on the quality of results?
- The advantages of involving an EBMT system
- How behave the SMT-Moses to rich morphology language pairs.
- How reliable are the automatic measures.

For the experiments we used the following setting:

- Training Corpora:
 - Acquis Communautaire
 - ROGER (Romanian-German-English -Russian) Corpus developed at UHH (technical manuals)
 - Europarl Corpus
- Test sets (10% from the training material) were extracted from these corpora
- Tools:
 - Moses as set up for the Euromatrix Evaluation Campaign 2010
 - EBMT in-house system (UHH)
- Experiments performed
 - SMT: Comparing results with similar research papers
 - SMT vs. EBMT on Acquis Communautaire
 - SMT vs. EBMT on ROGER
 - ROGER as Test-Corpus for SMT trained with Acquis Communautaire
 - SMT trained on Acquis Communautaire and tested with input from Europarl

The results of these experiments allowed us to define the system architecture, which will be described under Task T 6.2.

The main activity under T 6.1 was the domain adaptation.

The methodology used by the ATLAS project implies the usage of a pivot language, in our case English. This means that e.g. assuming we need a Bulgarian-Romanian Corpus for domain X, and we have a Bulgarian-English parallel corpus for domain X, we will use best available Mt-Engine English-Romanian to translate the English part of the corpus into Romanian. Then we will align the original Bulgarian part with the Romanian Translation.

In order to obtain accurate results we used if available more translation engines and interpolated the results. A pilot-testing phase was done before the ATLAS System used

the methodology. The parallel corpora (collected or automatic generated) have at least 2000 aligned units.

Finally we trained the domain adaptation models for 15 language pairs x 13 domains.

This consisted of:

- Translation model adaptation
 - Concatenation of the domain specific parallel corpora with JRC-Acquis (considered here as general domain)
 - Training of concatenated parallel corpora, introducing an additional factor providing the domain to which the word belongs
- Language model adaptation (separate SRLIM models for domain corpus and JRC-Acquis, followed by interpolation)

T.6.2. Input data format specification

The analysis of the results from these experiments allowed us to define the system architecture of the ATLAS-MT engine, as described in Figure 4.

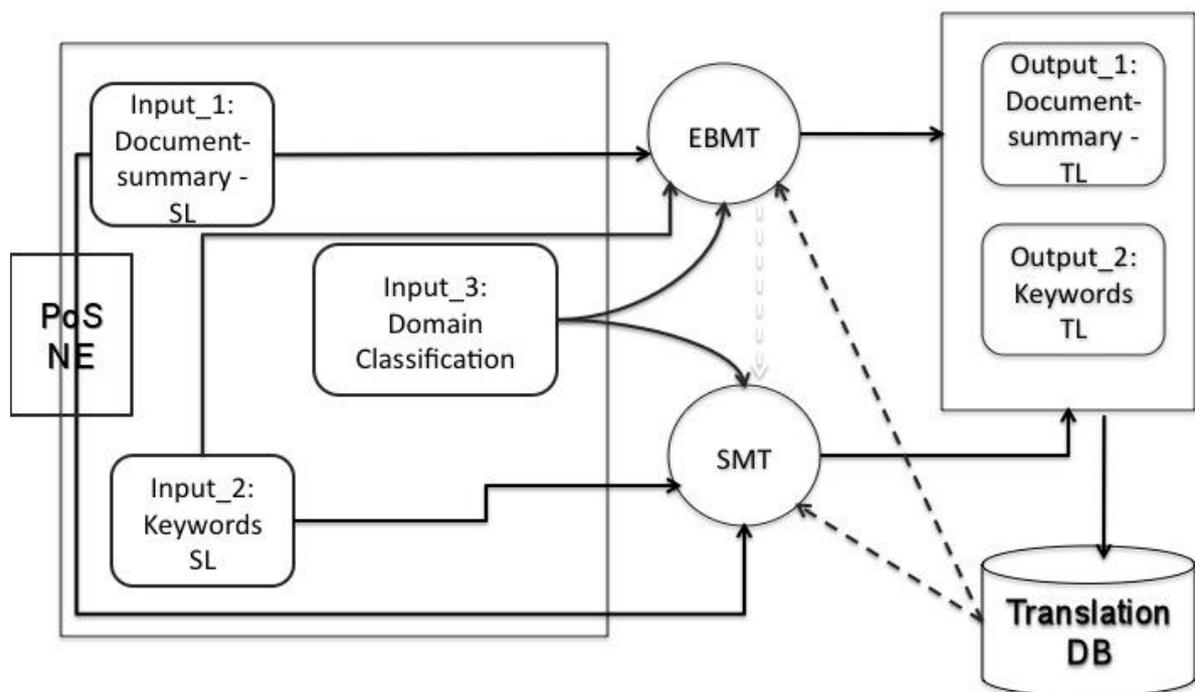


Figure 4. Architecture of the MT-Engine in the ATLAS System

The MT-Engine is a hybrid system using an EBMT component and a SMT component. A document uploaded to the system will be first passed to the linguistic processing tools, the categorisation and the summarization module. The categorization module will classify the

document to one of the 13 MT-Domains (upper levels of the ATLAS hierarchy). Each node in this hierarchy has attached an information regarding the availability of a SMT model trained for the respective domain. Accordingly either the translation of the document abstract and the most important NPs is performed or, a message is passed to the user.

The Example-based machine translation engine is working entirely with surface –forms, which makes it language independent. MT-input is passed first to the EBMT engine: if the entire input or very big parts of it are retrieved in the translation database the EBMT is turned on and delivers the output. If not the input is transmitted further to the SMT.

To summarize the workflow of the MT-engine is the following:

- Prerequisites: each translation request should be accompanied about an information about the domain to which the input belongs (i.e. document was before categorised); and an information about the period when the document was written (metadata extraction)
 - Input from older than 1850 will not be translated, and a message will be shown to the user
 - Input from Documents between 1850- 1950 will be sent further to the translation module, however the user will be supported with information that the translation may be not accurate
 - If for the respective domain and language pair, there is a trained model /translation database the translation model is automatic initiated. If not the user will be informed and the model available for the closest node in the categorisation tree (going up in the hierarchy) will be selected
- The input is first processed by EBMT engine. If the input is entirely found in the translation database (which can often happen in case of restricted domains or keywords, short chunks) then the translation equivalent is provided as result.
- In all other cases the input is sent to the SMT engine, which will provide the result.

T 6.3. Implementation of a bridge to EuroMatrix

All models trained for the ATLAS System can be used with current implementation with Moses System, thus the bridge with EUROMatrix is ensured.

T 6.4. Adaptation of existent cross-lingual search engine

The translation results are then embedded in a document model, which is used further for crosslingual search.

Each document is thus converted to the following format

```
<foaf:Document rdf:about=http://atlas.eu/item#20>
<dc:title>Internet Ethics </dc:title>
<dc:creator rdf:resource=http://atlas.eu/pers#950 />
<atlas:summary xmlns:lang="en">
Default english summary
<atlas:summary>
<atlas:summary xmlns:lang="de">
Deutsche Zusammenfassung
</atlas:summary>
```

```
</foaf:Document>  
<foaf:Personrdf:about=http://atlas.eu/pers#950>  
<foaf:name>Name </foaf:name>  
<foaf:mbox> name@some.address.eu </foaf:mbox>  
</foaf:Person>
```

This is the basis for creation of the RDF-Index. The crosslingual search engine is in this case a classic Lucene search engine, which operates however not with word-indexes but with these RDF-indexes, which automatically include multilingual information.

The implementation of the ATLAS CLIR engine is built in a layered fashion. The top/application layer constitutes the actual Nebula5 part whereas the lower layers are part of the Apache ServiceMix distribution and stem from a multitude of open source projects of whom we will describe the most important ones in this section (mainly by excerpting their description).

Within Nebula5, we aim for Semantic Search (but not for Semantic Web Search) for the following reasons: Search is often limited to searching literal text or URI nodes and is implemented as specific function within a RDF framework. We feel that is an unnecessary limitation because search functionality and RDF framework functionality should be tightly and efficiently integrated. It also should be possible to integrate schema information and use description logics and such when required.

We propose that fusing search engine and semantic web technology at the right level, i.e. enabling semantic annotations and intra-institutionwise distributed extensibility – while maintaining freetext search functionality – will create a certain amount of synergy which can raise the effectiveness of a semantic search approach in an institutional (enterprise) environment.

From our preliminary evaluation of some query logs of our institution we found that queries are strongly biased towards personal information (~28% of all queries) and organizational or structural queries, related to the institution (~36%), such as querying for departments, scripts, e-learning courses, etc.. This enterprise-search related aspect of course will have a great impact on the kind of semantics we need to employ – especially named entity processing should be treated with high priority.

Our approach to semantic enterprise search is based on a distributed modular architecture, named Nebula5. One of the main differences between our architecture and others is this: while the common approach to freetext semantic search (and also to semantic query expansion) aims to translate natural language queries/ keyword queries into formal expressions – which are subsequently used to search a model repository for matching RDF statements, we instead use conventional freetext queries on our RDF documents.

The most important key heuristic is hidden in the postprocessing step of our architecture – by querying the index we encounter three cases:

1. Using conventional keywords only: documents containing these keywords will be discovered and ranked according to the Lucene tf*idf scheme. Additionally, RDF URI nodes can be discovered, too – exploiting the fact that most RDF URIs contain semantically relevant, human readable parts. For example, a keyword search for “bob homepage” will also reward indexed items containing “<foaf:homepage>” – especially when in conjunction to the literal fragment “bob”. This can be quite useful, because many homepages in an institutional environment do not explicitly state that they are homepages!
2. Submitting a mixture of keywords and RDF URIs: queries like “foaf:homepage bob” will find “Bobs homepage” – but not “Jills homepage” with a reference to Bob! Because Lucene query analyzers eliminate non-alphanumeric characters, a domain-less URI is treated like a keyword; i.e. the query “:homepage bob” will not be restricted to the FOAF12 domain but rather work like an ordinary keyword query in the above explained way. In the special case of web documents containing microformats such as RDFa these will be implicitly honoured the same way.
3. Submitting RDF URIs only will exhibit documents with certain semantic properties: the query “foaf:homepage” will return all indexed items that contain homepages in the sense of the homepage element of the FOAF schema, plus the FOAF schema itself (as it also contains the fragment “foaf:homepage”).

The query results (possibly filtered by a predefined document relevance threshold or by a first-N-documents-only heuristic) are merged into a single resulting RDF model that can be searched by means of templates, implemented structured RDF querying languages, e.g. SPARQL in order to provide end-user application functionality.

In this way, a query-centric RDF model is constructed dynamically on each search occasion that reflects the “ignorance-artifact” created by the user. Because schemata are discovered as well (and can be further tracked by using the PREFIX RDF document fields), we are not restricted to structured RDF queries only but can also apply description logics in order to further examine query results. For example, we can deduce subclassing etc. On the other hand, when it’s just a portion of the textual content, it is being searched for, we can simply output the value of the “nie:content” predicate triple. In this way, we are able to defer complex processing until it is really needed.

Nebula5 is realized as set of RDF-centric components that are integrated into the ServiceMix service container infrastructure. The ATLAS CLIR engine itself is basically an independent configuration applied to these components and also deployed to the ServiceMix container.

T 6.5. Integration of the cross-lingual search engine in ATLAS

The Nebula5 system is based on a set of external and publicly available open source frameworks as follows.

ServiceMix

Apache ServiceMix is a multi-purpose service container that builds on an OSGi infrastructure and offers many technologies (such as java messaging, web services, web applications, java business integration etc.) bundled into a single, consistent provider application.

ServiceMix is an open-source development, but private companies offer consulting and supporting services.

Maven

Maven is a net-based development project management tool that is tightly integrated into ServiceMix. Maven enables the service container instance to retrieve dependencies (jar files that are needed by some installed feature) from local disk locations or from public web repositories.

OSGi Component Bundling

ServiceMix contains many individual software components from third-party projects in form of so-called “bundles”. These bundles are like ordinary java jar files in many respects except for a handful of specific manifest entries that define external dependencies and how software may access the classes and services provided by the respective bundle.

The main benefits of OSGi are:

- for developers: OSGi reduces dependency and versioning complexity by providing a modular architecture for (distributed) component-based systems.
- in business terms: The OSGi modularity concept reduces implementation and operational costs and integrates components in a highly dynamic environment, successfully aiming at application development, maintenance and remote service management.

(cf. <<http://www.osgi.org/Technology/WhyOSGi>>, accessed 21.06.2012)

In practical terms, OSGi solves many issues related to jar library versioning, local service provisioning and library access management. All bundle and feature management functions (installing new modules, uninstalling modules, starting and stopping services, dependency management, etc.) within ServiceMix are handled by OSGi.

Apache Camel

Apache Camel is needed for providing connectivity to and from external clients and services as well as routing document-centric data from the external connectors to the RDF index database.

The routing process may also include transformation processes such as templating the RDF search results and returning them as HTML or XML document fragments to the querying party.

ActiveMQ and JMS

ActiveMQ as an implementation of the Java Messaging Services (JMS) specification is the transport technology that connects ATLAS to the Nebula5/ServiceMix environment.

ATLAS Assembly

The ATLAS assembly defines some metadata, by which it can be identified within the ServiceMix container:

1. Bundle-Version = 1.0
2. Bundle-Name = Nebula5 :: ATLAS CrossLingual Information Retrieval
3. Bundle-SymbolicName = nebula5-atlas
4. Bundle-Description = CrossLingual Information Retrieval for the ATLAS EU project
5. Bundle-Vendor = ronald.winnemoeller@uni-hamburg.de

The assembly is implemented as Spring-XML file and located in the `SERVICEMIX.ROOT/deploy` directory. Any changes to the assembly reflect immediately to the running processes in case the ServiceMix container is running.

The assembly should be regarded as “default” starting point – without intercepting the default behaviour, many endpoints and transformations can be added when desired.

T 6.6. Documentation

Both engines are well documented. The crosslingual search engine comes with an installation and deployment manual, inserted as well in the Deliverable D 6.2. For the machine translation a web page containing all experiments and results obtained through the project is available at <http://www.c-phi.uni-hamburg.de/MTATLAS>

WP6 Deliverables and Milestones

D 6.1. Machine Translation Engine was delivered to the commission. It contains a state-of-the-art of the approaches used in the ATLAS engine, the experiments performed, the system architecture as well as the results of the BLEU evaluation

D 6.2. Crosslingual search engine was delivered to the Commission. It describes the architectural framework Nebula 5, the adaptation to ATLAS, the interface with the ATLAS CMS, and includes an installation guide.

No milestones were associated with this workpackage.

Final results

The ATLAS system integrated for the first time a machine translation engine capable of domain adaptation over a large number of domains and language pairs. The domain selection is performed by the document classification component embedded into the ATLAS engine. The MT-engine uses a hybrid paradigm combining two corpus-based approaches. Together with the engine we deliver 13 domains x 15 language pairs translation models.

The cross-lingual search engine is an independent module, which can be easily adapted and implemented with another CMS. It indexes the ATLAS Document via RDF triples automatically constructed from the information provided by all other processing components of the atlas system. The crosslinguality is ensured via the translation component.

WP7 description

The WP7 “Testing and User Evaluation” is a horizontal WP taking input from the results of all technical WPs (WP2, 3, 4, 5, 6). The focus of WP7 is on ensuring the acceptable quality of the final ATLAS system, by assessing the technical performance and the level of fulfilment of the system specifications, as well as assessing the level of user expectations fulfilment. The main WP activities include:

- Design and development of a User Evaluation Plan
- Definition and design of a Technical Test Plan, addressing various testing aspects (isolation, integration, regression).
- Setting up and maintenance of User Groups and a dedicated Living Lab to strengthen the user involvement in the project.
- Definition of technical evaluation methodology (testing corpora, user cases, indicators, etc.)
- Performing of Technical & User Acceptance Evaluation, by applying the Technical Test Plan, the technical evaluation methodology and the User Evaluation Plan.
- Analysis of evaluation findings; respond with system improvements.

These activities are implemented in the following 2 Tasks:

T7.1 – User evaluation and feedback adjustments

The task started with the development and the detailed definition of a User evaluation plan, providing information on all important aspects of the user evaluation exercise:

- Definition of the evaluation scope: what are we going to evaluate, what is the scope with respect to geographical distribution, what are the scenario to be trialled, etc.
- Organisation of the user involvement: what are the expected user types, methodology and timing of their involvement, local user groups and Living Labs involvement, etc.
- Description of the overall methodology to be followed: how the activities will be organised and conducted, what will be the methodological approach to the different evaluation challenges, processing of results, etc.
- Design of the statistical framework for collecting and analysing the user feedback: what will be the statistical indicators, what statistical measures will be used, the relevant thresholds, etc.
- Description of the use-case scenarios: which scenarios are relevant for our target user types, the objectives per scenario and the relevant system features under evaluation, etc.
- Setting up the detailed user evaluation timeplan: how many rounds should we deploy, what will be the timing of each round, round-specific evaluation scope, etc.
- Presenting the user evaluation instruments: what are the questions to be asked, the contents of the questionnaire, how can we maximise user access potential, etc.
- Description of the Multilingual e-Content and e-Library Living Lab – its mission, objectives and sequence of maturity phases, host institutions, infrastructure, partners, methodology, expected results, contribution in the project evaluation activities, etc.

We defined 4 types of user groups and a Living Lab user group:

1. UG1: students and scholars
2. UG2: authors, scientists, researchers
3. UG3: Internet users with moderate WEB experience

4. UG4: digital content and content publishing professionals (WEB designers, editors, publishers, lawyers, etc.)
5. LL: professionals, members of the MLeCeL Living Lab.

We also defined 4 User Test Scenarios and a LivingLab specific Test:

- Scenario 1 (Reader): i-Librarian service will be used in order to evaluate the interface usability and the users' reaction to the language features provided by the service; as well as extended features such as storing and processing collection of documents, retrieval and review of extracted textual details, etc.
- Scenario 2 (User): i-Publisher service will be tested by non-professionals while customising a ready-to-use website.
- Scenario 3 (Creator): general users with moderate web experience will use the i-Publisher service in order to evaluate the easiness, simplicity and convenience of website creation (i.e. through simple wizards) and publishing.
- Scenario 4 (Content provider): where professionals, working with on-line content, editors in news and media agencies, publishers, lawyers, web designers will work on i-Publisher in order to evaluate the features and flexibility of the i-Publisher interface.
- L.L. Scenario (Professional users): professionals such as information designers, web studio members, and web designers from the Living Lab will test the i-Librarian, but they will mainly focus on the advanced functionalities of the i-Publisher service so as to evaluate the available features and flexibility of the i-Publisher interface for creating and customizing a rich content-driven web site.

Each scenario was complemented by an exercise, requesting the user to accomplish various tasks, each consisting of one or more steps (mandatory or optional). The users had the possibility to access online questionnaires in all rounds of the evaluation process. Each question in the questionnaire refers to an indicator. For the measurement of each indicator a 5-point scale was used ("1" representing "very low" and 5 "very high"). Indicators were grouped into 4 main categories, namely: Usability indicators (9), Qualitative (4), Satisfaction (7), Task fulfilment (20).

To leverage the effectiveness of our user evaluation activities and strengthen the user involvement in the project we established a pilot Living Lab ("Multilingual e-Content and e-Library" - MLeCeL), which conducted:

- Training and Testing Seminar of i-Librarian service;
- Evaluation workshop on the i-Librarian service;
- Two feedback workshops on the full versions of i-Librarian and i-Publisher;
- One workshop to test and evaluate the i-Publisher - Advance Mode;
- Four testing and evaluation workshops on i-Librarian and i-Publisher.

The members of the Living Lab, along with local user groups from the partner countries, were our main assessors, as we recorded and analysed their feedback; mainly from 3 evaluation rounds.

1st evaluation round

In the 1st round both services had limited functionalities and only English support. The MEAN value of each indicator was average or above average, which was encouraging for a start. Our strong point was the user interface which was considered good/excellent by the majority of the users.

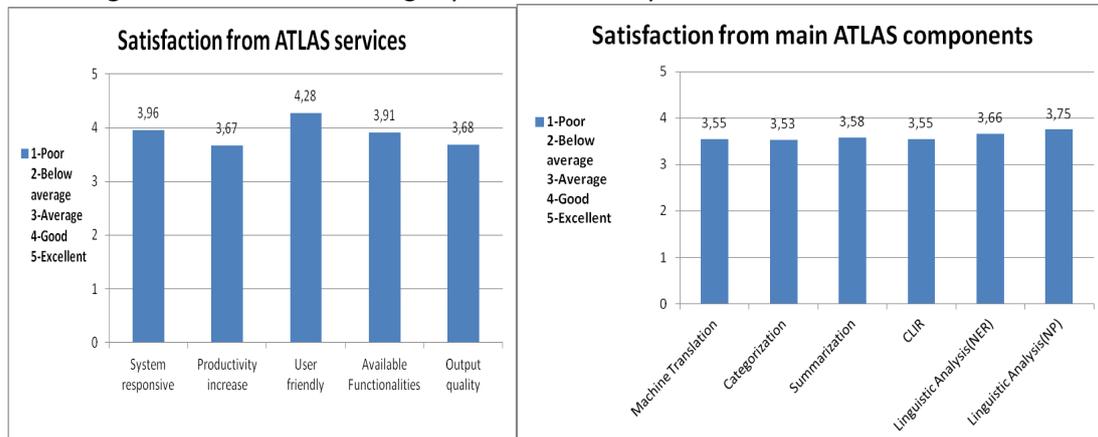
2nd evaluation round

In the 2nd round, for both services, we had complete production versions in English only. For the majority of the indicators the MEAN value was above average, but users indicated that there is still space for improvements (e.g. online help and tooltips could be improved, automatic categorisation, machine translation and summarisation should be improved, etc). 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 with i-Publisher in less than 20 mins.

3rd evaluation round

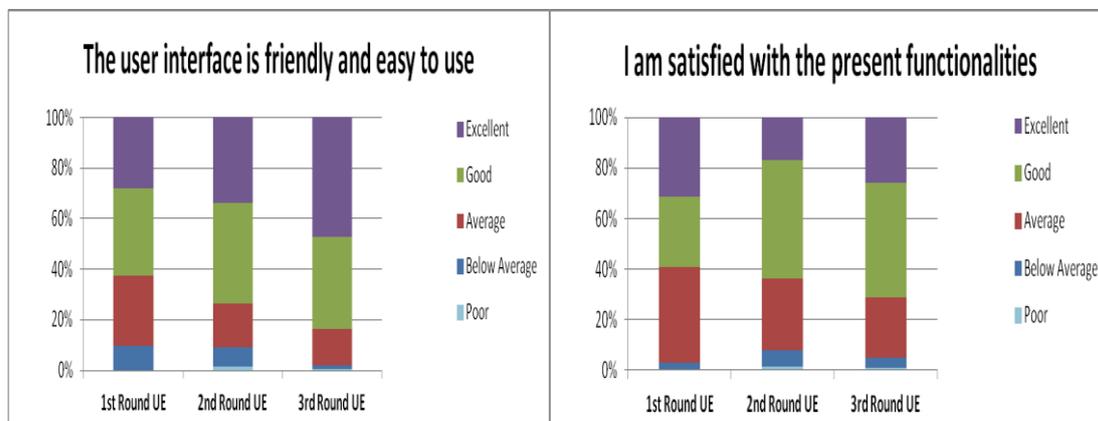
In the 3rd round we had true multilingual support for both services and we evaluated all functions in their latest production versions:

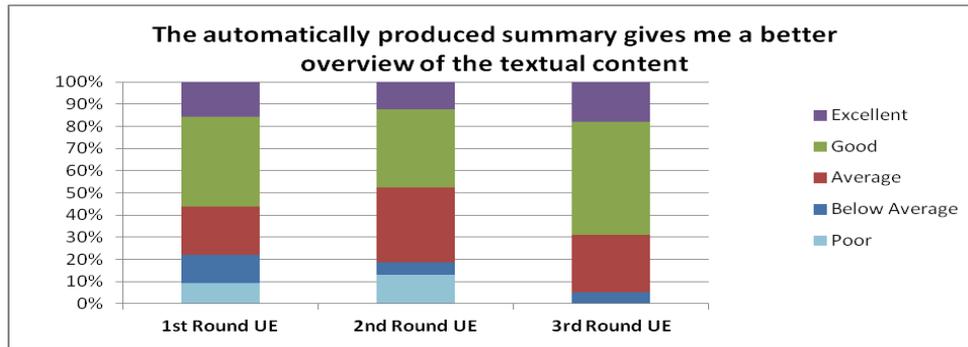
- The “user friendliness” seems to be our strongest point for the online services (it is considered “more than good”).
- The user impression from the overall output quality (groupings, text excerpts, summaries, translations) is above average, but below good.
- The linguistic analysis seems to be almost good; while the machine translation and the categorisation were found slightly less satisfactory.



Evolution of user appreciation across evaluation rounds

By analysing the results from all 3 rounds it becomes obvious that ATLAS was progressively improving its services to the end-users. As one can easily notice from the following diagrams, major improvements were achieved in terms of the user interface friendliness, offered functionalities and automatic summarisation.

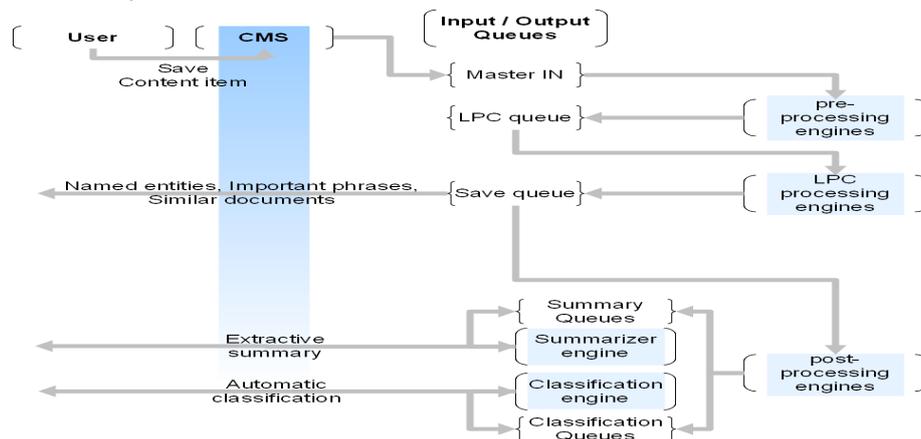




T7.2 – Component, integration and system testing

The task dealt with the testing of the ATLAS platform and its main components, in terms of isolation, integration and regression testing. In more detail:

- We defined in details the testing methodology: what is to be tested, what is the scope with respect to individual components and indicators, etc.
- We developed scenarios, testing steps and conditions for failure and success, testing corpora, etc. for the integration and regression testing of the whole platform and its main components.



- We executed the testing scenarios and scripts at the level of isolated components (unit tests), at the level of integrated platform (integration tests), and after each new deployment (regression tests).

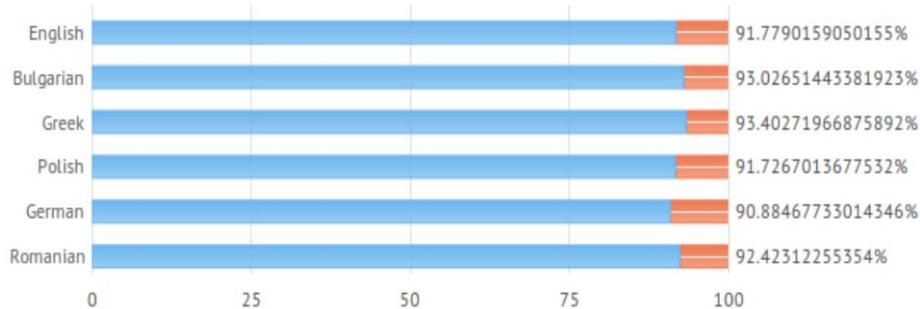


Regression Tests Report

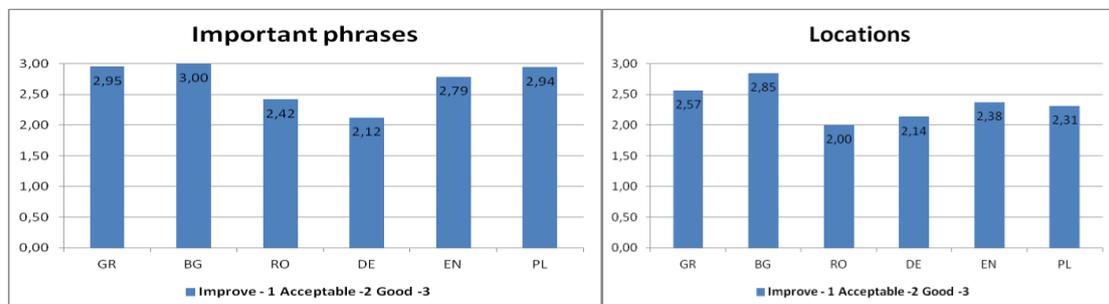
01.02.2013

10:18

Overall results



- We defined “Simple Confirm Indicators” to assess the level of fulfillment of ATLAS specs (mainly the 2 online services, i-Librarian and i-Publisher).
- We collected a significant amount of documents (testing corpora) used for both testing and technical evaluation.
- We defined in details the methodology to use for the technical evaluation of the platform components, including test cases and scenarios for each main component and respective technical indicators – 1 set of indicators for each main ATLAS component (LPCs and CMS, categorisation, MT, CLIR, summarisation).
- We collected and documented results from the technical evaluation of each main component, including comparative assessment results – where this was applicable.



	Language	BG	DE	EN	GR	PL	RO	AVG
ATLAS summarizer	Precision (H)	0,19	0,23	0,27	0,23	0,17	0,22	0,22
	Recall (H)	0,29	0,44	0,41	0,41	0,36	0,32	0,37
	F-measure (H)	0,23	0,30	0,32	0,29	0,23	0,25	0,27
OTS summarizer	Precision (H)	0,16	0,19	0,24	0,27	0,19	0,29	0,22
	Recall (H)	0,25	0,20	0,22	0,33	0,21	0,06	0,21
	F-measure (H)	0,19	0,20	0,23	0,27	0,20	0,10	0,20
LexRank summarizer	Precision (H)	0,15	0,23	0,27	0,24	0,24	0,21	0,21
	Recall (H)	0,18	0,25	0,25	0,22	0,24	0,22	0,18
	F-measure (H)	0,16	0,24	0,26	0,23	0,22	0,21	0,19

Comparative assessment for the ATLAS summarisation component

Lang-Pair	BLEU score for the “Sociology” domain	BLEU score
-----------	---------------------------------------	------------

	for the ATLAS MT component	JRC-Acquis
DE-EN	35.21	27.30
EN-RO	30.46	21.92
PL-EN	39.69	19.43
RO-BG	17.68	7.31

Comparative assessment for the ATLAS Machine Translation component

WP7 deliverables and milestones

Three deliverables were submitted:

1. D7.1 “User evaluation plan”: provides details of the user evaluation methodology to be followed (user groups, evaluation rounds, timing of activities, scenarios and exercises, questionnaires, etc.)
2. D7.2 “Analysis of User Evaluation”: provides details about the results of the user evaluation activities across the 3 evaluation rounds.
3. D7.3 “Final Report on Test Results”: provides details about the testing approach (isolation, integration, regression, etc.) used and the results obtained, the methodology used for the technical evaluation and the results achieved (including comparative assessment – where applicable).

No milestones were associated with this WP.

Final results

The results of this WP are related to the 2 main groups of activities, namely user acceptance evaluation and technical evaluation and testing. An outline of the major results is as follows:

The scope of WP7 encompassed two major axis of work,. The major project achievements in each one of them is outlined below:

User Acceptance evaluation

- Detailed methodology in order to organise and conduct the user evaluation: methodological approach to the different evaluation challenges, approach for processing of the results, user types and groups, methodology and timing of their involvement, local user groups and Living Labs involvement, etc.
- Established a pilot Living Lab (“Multilingual e-Content and e-Library” – MLeCeL) to strengthen the user involvement in the project.
- Detailed use-case scenarios and exercises for each one of the 4+1 defined user groups, including objectives per scenario and the relevant system features under evaluation, mandatory and optional steps, etc.
- Statistical framework for collecting and analysing the user feedback, including statistical indicators and measures to be used, and relevant thresholds.
- Online questionnaires for each user group type.

- Conducted 3 user evaluation rounds.

Technical evaluation and testing

- Detailed testing methodology.
- Scenarios, testing steps and conditions for failure and success, testing corpora, etc. for the integration and regression testing of the whole platform and its main components (see testing reference diagram below).
- Documentation and appraisal of results from the testing of the ATLAS platform, at the level of isolated components, at the level of integrated platform, and after each new deployment (i.e. regression testing).
- Assessment of the level of ATLAS application specifications fulfilment.
- Collection of documents (corpora) used for both testing and technical evaluation.
- Methodology used for the technical evaluation of the platform components, including test cases and scenarios for each main component and respective technical indicators.
- Technical evaluation of each main component, including comparative assessment results – where this was applicable.

Description of WP8

The activities performed by all project partners within WP8 resulted in:

- Wide national and international promotion of the ATLAS platform, encouraging the use of i-Publisher, i-Librarian and EUDoCLib.
- A lot of promotional materials (including brochures, posters, leaflets and etc.) were produced to demonstrate the real value of the ATLAS platform and the online content services.
- ATLAS website was developed and updated with information and direct access to the ATLAS services. It is the main tool for interaction with target communities.
- The scientific results achieved within ATLAS were communicated with scientific community through presentation of scientific articles at national and international conferences and workshops.
- Interaction with relevant projects was performed by organisation of joint events, presentations, demonstrations, exchange of ideas and discussion of future exploitation activities
- MLeCeL Living Lab website was developed and many testing and evaluation activities with users and IT experts were performed in order to improve the performance and usability of the ATLAS services.
- Users were actively involved in the evaluation workshops, organised in the MLeCeL Living Lab, experts` and users` feedback collected.
- ATLAS consortium partners elaborated strategy for ensuring project sustainability and planned exploitation activities to provide a basis for future project funding and extension.
- ATLAS services were demonstrated at large exhibition and conferences in order to establish contacts with potential customers.

- Potential customers were communicated by presentation of ATLAS at meetings and exploration events and production of demonstration websites in response to the identified needs of the customer.

Updated description of the tasks implemented in WP8

T 8.1: Project website

According to the project work plan, in Month 2 a portal was created to promote the ATLAS project. It was built with i-Publisher as an experiment within MLeCeL LL. In the second year, a new vision of the project website was applied and the website was updated with information about the Atlas services – i-Publisher and the linguistic platform Asset (<http://i-publisher.atlasproject.eu/>, <http://atlasproject.eu/atlas/asset>) and developed additional demonstration services - i-Librarian, EUDoLib, News Service Demo, UNESCO Website Demo, and World Health Organization Website Demo. The ATLAS website is dynamically updated and has several important functions: from one side, it plays a role of the main project information repository and from another – it is a tool for dissemination and interaction with users.

T 8.2: Detailed Dissemination Plan

Initial dissemination plan was developed during the first year and was updated regularly. It included preparation of brochures, website development and distribution of dissemination materials and presentations the ATLAS project at the identified national and international events (conferences and exhibitions). During the 2nd year, the planning consisted of participation in larger conferences and presentation of ATLAS services through scientific papers, posters and demonstrations to the target users. Last year, planned dissemination activities of the consortium partners were focused on organization of events targeted at ATLAS potential users and demonstration of ATLAS project achievements at exhibitions and large international conferences.

T 8.3: Sustainability plan preparation

Developed Sustainability strategy to be applied after the project end and specific partners' exploitation plans for further use of the ATLAS results are presented in Deliverable 8.4.

- Description of ATLAS products and services (their main characteristics and functionalities as well as some of their competitive advantages)
- Description of ATLAS target users and respective use cases
- Performed search in the market of web content management systems, market trends and identification of main players on the market.
- Elaboration of comparative matrix for products, available on the market and ATLAS.
- Generic SWOT analysis of ATLAS products.
- Identification of IPR issues, concerning ATLAS exploitable results
- Description of commercial non-commercial exploitation cases of ATLAS products and services
- Paths for Community building, including description of communities, tools for interaction and knowledge transfer
- Promotion activities, tools and plans associated with sustainability strategy and exploitation.

T 8.4: Preparation of primary outreach materials

In order to increase the visibility of the ATLAS project and its identity among potential end-users and public and to facilitate their access to project services, variety of dissemination materials were produced within the project. Their detailed description has been done in *D8.2 Dissemination kit* according to the project work plan.

- The brochures of Atlas, i-Librarian and the general project brochure are available on the project website: <http://www.atlasproject.eu/atlas/project/dissemination/en>
- 10 posters were developed for visualising and disseminating the project results. They were introduced at conferences and exhibitions for better presentation of the developed services, used technologies, present and further benefits, and widening the scope of the potential users.
- The ATLAS project results are presented in 11 articles, press releases, newsletters and other publications addressed to the users and the general public.
- The results of the Atlas project are made available through 31 articles and papers, published in scientific and technical journals and conference proceedings and presented at annual conferences of the world's major professional communities for content management, and other relevant areas.
- Online video tutorials, guides and technical documentation about i-Publisher and i-Librarian were developed.
- Development of 12 demonstration cases for potential customers:
 - i-Publisher Live demo: <http://i-publisher.atlasproject.eu/>
 - i-Librarian website, <http://www.i-Librarian.eu/>
 - EUDocLib website, <http://eudoclib.atlasproject.eu/>
 - PLDocLib, <http://www.atlasproject.eu/pl/>
 - News Service Demo, <http://newsdemo.atlasproject.eu/>
 - UNESCO Website Demo, <http://unesco.atlasproject.eu/>
 - UNESCO Chair Website Demo, <http://a2.atlasproject.eu:9090/unesco>
 - World Health Organization Website Demo, <http://asset.atlasproject.eu/who>
 - Video materials demo version, <http://videodemo.atlasproject.eu/>
 - Press clipping service in Bulgarian and English, <http://mediaimage.tetracom.com/>
 - Plagiarism service: <http://textmatch.eu/>
 - Icon-painters from Bundovtzi family website, http://www.atlasproject.eu/asset_demo/icons/bg/index.html

T 8.5: Dissemination of primary materials

In order to reach to the target customers and users and to maximize the visibility of the project, ATLAS partners used different channels and tools and cooperated with other relevant national and international organisations and projects, sharing similar objectives. A list of appropriate events and conferences that bring together the relevant stakeholders had been preliminary developed in order to facilitate the selection of the most important channels for disseminating ATLAS information and achievements.

- During the 1st year, the Dissemination activities included preparation of brochures and website development. Partners mainly distributed dissemination materials and presented the ATLAS project at **39** national and international events. The objective was to raise awareness of the ATLAS by presenting the project to the content providers, linguists, relevant representatives of the scientific community and potential users.
- During the 2nd year, the results of the Atlas project were made available through **13** scientific papers to the target users. Ten newsletters and articles addressed the users and the wide public as well. Through technology and tool demonstrations, case study presentations, the ATLAS services were presented at **32** relevant events

(exhibitions and big conferences) in order to explore possible ways of their exploitation by demonstrating their real benefits to the different target groups.

- During the 3rd year services' presentations, demos and **18** scientific papers were disseminated at **38** national and international events, with the focus on large conferences and exhibitions and organisation of target workshops to increase visibility of the ATLAS services among potential customers – individuals and organisations and to strengthen the partnership between ATLAS members and the potential participants in the established ATLAS communities.

T 8.6: Interaction with target communities

Dissemination activities performed in ATLAS project were also oriented towards establishment of ATLAS 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);
- ▶ Business Angels from all over Europe (11th Congress of European Business Angels Network);
- ▶ IT experts, PhD students and researchers in Software engineering, web developers (Usability seminar). Most of them participated in the 3 testing rounds of the prototypes;
- ▶ Web developers, e-commerce companies, hosting companies, content management and CRM providers, ad networks from CEE and the rest of the world (WEBIT'11)
- ▶ Living Lab users and experts during ATLAS evaluation workshops

The communication and interaction with the different communities was achieved by participation in international events and a number of **coordinated workshops** (international workshops for research community; meetings with potential users; user evaluation workshops with developers and users and joint events within MLeCeL Living Lab and etc.) with both international and national scope, **hosted and organised by ATLAS partners**.

The Multilingual e-Content and e-Library Living lab (MLeCeL) has been launched in the frames of the ATLAS Project and 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.

Living Lab website (<http://livinglab.itd-bg.eu/>) developed in September 2011, is a communication tool for the members of the Living Lab communities and it is very appropriate environment for attracting these participants in the ATLAS communities as well. Testing and Evaluation questionnaire was developed and published on the LL website (November 2011) for collecting feedback from users, developers and experts - <http://livinglab.itd-bg.eu/content/testing-evaluation-questionnaire>. Living lab platform is update regularly with current LL activities, including dissemination events, results, feedback, and reports from testing, user suggestion box. It is going to be used further as one of the main dissemination tools after the project expiration in order to keep the communities informed about latest developments of the ATLAS services.

Interaction with 10 relevant projects (META-NET, Cesar, D-SPIN, CLARIN, PROMIS Lingua, OpenScout, ShareTEC, SmartBook, DYLAN, ELLIOT) and 3 networks (MUMIA, EUROPE Innova KIS and OpenAIRE), by the organisation of joint events, presentations, demonstrations, exchange of ideas and discussion of future exploitation activities

T8.7: Project impact measurement

In this task we started with the determination of the requirements of the main target user's groups and their respective networks. We used various techniques to assess the relationships between different organizations and how to leverage the knowledge dissemination across organizations and increase the effectiveness and productivity of the content workers. As a result we establish close relations with each target group and network, plan and implement many activities to disseminate the potential impact of ATLAS tools and collected feedback from many professionals and organisations. The main vehicle for supporting this type of activities was the Living Lab created and connected with relevant other networks of Living Labs. By continuing monitoring, operation and analysis of the users and networks through the Living Lab, we contributed to the experience of building communities of knowledge and practice.

Expected Impact

The project brings together advanced technologies for multilingual web content management and text mining (such as automated annotation, mark-up and translation) in a united platform. The developed software-as-a-service architecture demonstrates the capabilities of the ATLAS platform, and the open-source license facilitates the spread of the project output.

The ATLAS system offered the following 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
 - Facilitate exchange of information and knowledge
 - Simplify authoring, management and exploitation of heterogeneous multilingual content
 - Address the needs of a large number of people belonging to different target user groups – individuals and organizations
 - Cross the language barrier
 - Facilitate culture exchange
 - Liaise with Europeana and EuroMatrix Plus – The liaison with EuroMatrix

Most of the technological expected impacts were measured using the defined in WP7 evaluation criteria and scenarios. There were additionally proved by providing selected in-depths interviews with leading experts in the field as well as with a series of targeted implementations of the ATLAS platform in various domains like University content management portals, research management solutions, media information systems, etc.

The LL contributions to impact measurement and analysis were channeled through:

- User interviews (112) and user suggestions (38);

- Feedback from idea generation workshops with lead-users (3);
- Feedback from idea generation workshops with experts (4);
- Online suggestion box (ongoing moderation and summarization);
- Feedback from workshops with users and prospective users (8);
- Feedback from prototype testing (6);
- Feedback from real settings service deployment (at SULTSIT and Sofia University);
- Usability Lab controlled assessment of user acceptance (3);
- Feedback from education and training sessions at university level (2);
- Feedback from training workshops with developers (3);
- Feedback from workshops with researchers (2);
- Catalysing research activities based on ATLAS (3 PhD study and 12 MSc studies);
- Ongoing support of ATLAS communities and summing up of their feedback;
- Regular workshops for testing, evaluation and feedback from users of all ATLAS software tools and system as a whole (6)
- Final report of the LL activities.

In particular, data was collected with surveys that ask users (from the user groups described in WP7) to evaluate different aspects of their perception of access to content. Access to content reach the standard level during the first year of the project, and increased during the second and third years respectively with 20% and 30%.

The achievement of the planned milestone of 20/25/30 people attending dissemination events is a guarantee for the significant level of success of the ATLAS project.

The number of web site hits during the first year was increased with 75% and 50% over the next two years.

The number of media citations during the first year increased with 200% by the end of the third year.

T8.8 and T8.9: Nationwide and Trans-European project promotion

ATLAS team used the opportunities of relevant **exhibitions and scientific events** to disseminate the project results by giving **presentations and performing demonstrations**, thus reaching potential customers and users. The **most important** events of ATLAS consortium are presented as follows:

- CeBIT 2012, 6-10 March 2012, Hanover, Germany
- EAACL 2012, 23 - 27 April 2012, Avignon, France
- CESAR META-NET Road show 2012 Sofia, Bulgaria May 2nd 2012
- LREC 2012, 21-27 May 2012, Istanbul, Turkey
- EAMT (Summit of the European Association for Machine Translation), 28-30 May 2012, Trento
- 17th International conference on Applications of Natural Language Processing to Information Systems (NLDB 2012), 26-28 June 2012, Groningen, The Netherlands
- International Exhibition of Software Bucharest, 7-9 November, 2012

The ATLAS project results were **presented using presentations, posters and scientific papers at national and international conferences, workshops and seminars** among different target users. The most important events of ATLAS consortium are presented as follows:

- University of Winterthur, Winterthur, Switzerland, February 2013
- University of Buenos Aires, University of Cordoba, Argentina and University of Santiago, Santiago de Chile, Chile, 17-29 November 2012
- UNESCO IITE and UNITWIN/UNESCO Chairs International Conference "UNESCO Chairs Partnership on ICTs use in Education", 5 – 10 September 2012 St.-Petersburg, Russian Federation

- Human Language Technology Days 2012, 27-28 September 2012, Warsaw, Poland
- META-FORUM 2012, 19-20 June 2012, Brussels, Belgium
- 2nd Pan European Private Equity and Start-ups Forum, 21-22 June 2012, Athens, Greece
- Cross lingual Language Technology 2012, 7-8 May 2012, Univ. Hamburg, Germany
- NLP Seminars on Computational Linguistics, Faculty of Computer Science, "Al. I. Cuza" University of Iasi, 3 May 2012 and 10 May 2012, Iasi, Romania
- Bring-IT on! 2012, 17-18 May 2012, Iasi, Romania
- International Exhibition of Software, ExpoBinary, 7-7 November 2012, Bucharest, Romania.
- Language Technologies in Romanian Diaspora Research & Development, 26-27 September, Bucharest, Romania.
- 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
- META-FORUM 2011, 27-28 June 2011, Budapest, Hungary
- 5th Language & Technology Conference (LTC 2011): Human Language Technologies as a Challenge for Computer Science and Linguistics, 25-27 November 2011, Poznan, Poland
- WEBIT'11, 26-27 October 2011 in Sofia, Bulgaria
- Conference on Resources and Tools for Processing Romanian Language - ConSLR-2012, 27 April, 2012, and 2011, Bucharest, Romania
- Conference on "measures to support enterprises in implementing investments in digital solutions". 17 January 2011, Drama, Greece
- Corporate business day (Atlantis group of companies) with the participation of managers from all over Greece and Cyprus, 10 February 2011 Thessaloniki, Greece
- LREC 2010, 17-23 May 2010, Valetta, Malta
- MOBIP 2010 and Investment in Mobile and IT Services, 10-16 June 2010, Valencia, Spain
- FASSBL07: The Seventh International Conference Formal Approaches to South Slavic and Balkan Languages, 3-7 October 2010, Dubrovnik, Croatia
- Bring-IT on! 2010, 10 December 2010, Iasi, Romania
- Corallia Conference for the mi-clusters.job, 20 December 2010, Athens, Greece

List of the deliverable in WP8

- D8.1 Project website (M2)
- D8.2 Dissemination kit (M31)
- D8.3 Project Dissemination report (M35)
- D8.4 Sustainability plan (M35)

References

Daniel Anechitei, Dan Cristea, Ioannidis Dimosthenis, Eugen Ignat, Diman Karagiozov, Svetla Koeva, Mateusz Kopeć, Cristina Vertan (2013). Summarizing Short Texts Through a Discourse-Centered Approach in a Multilingual Context. In Neustein, A., Markowitz, J.A. (eds.), *Where Humans Meet Machines: Innovative Solutions to Knotty Natural Language Problems*. Springer Verlag, Heidelberg/New York.

- Dan Cristea, Oana Postolache, Georgiana Puşcasu, Laurenţiu Gheţu (2003). Summarizing documents based on cue-phrases and references, in Proceedings of the International Symposium on Reference Resolution and Its Applications to Question Answering and Summarization, Venice, Italy, June.
- Cristea, D., Postolache, O., Pistol, I. (2005). Summarisation through Discourse Structure. In Alexander Gelbukh (Ed.): Computational Linguistics and Intelligent Text Processing, 6th International Conference CICLing 2005, Mexico City, Mexico, February 2005, Proceedings, Springer LNSC, vol. 3406, ISBN 3-540-24523-5, pp. 632-644.
- Cristea D., Ide N., Romary L. (1998). Veins Theory: A Model of Global Discourse Cohesion and Coherence. In Proceedings of the 17th international conference on Computational Linguistics.
- Grosz, B.J. Joshi, A.K. and Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. Computational Linguistics, 12(2), 203-225.
- Joshi A., Schabes Y. (1997). Tree Adjoining Grammars. In Handbook of formal languages.

ANNEX 2: Dissemination report

Introduction

The present Deliverable: 8.3 Project dissemination report is a part of the Work package 8 (WP8): Dissemination and Exploitation of the ATLAS project. The objectives of this work package are to:

- Promote the ATLAS platform and encourage the use of the i-Publisher, i-Librarian and EUDocLib;
- Plan the project sustainability and demonstrate the real value of the ATLAS platform and the online content services, and so provide a basis for future project funding and extension;
- Disseminate project results through various channels – advertising materials, publications, workshops, etc;
- Present the project on both national and international forums.

The main objective of the Deliverable 8.3 Project Dissemination Report is to present a detailed description of all events organised by the consortium as well as the participation of the partners in both national and international expos and conferences. The report also covers the preparation and dissemination of primary outreach materials, interaction with target communities, and the project impact.

Deliverable 8.3: Project Dissemination Report contains **nine** chapters:

The **first** chapter presents **dissemination strategy** approved by the team and implemented in the ATLAS project. It includes description of dissemination objectives, target groups of users and dissemination tools. The **second** chapter describes the ATLAS project website developed as a main tool for dissemination of the project achievements.

The **third** chapter includes **ATLAS services` website**, which demonstrates the ATLAS technologies and functionalities offered to the potential users. It also presents other demonstration websites developed to respond to the identified needs of the potential customers and users. The **forth** chapter is dedicated to the **MLeCel Living Lab website**. It is not only a dissemination tool but also a tool for interaction with target communities within ATLAS.

The **sixth** chapter describes in general the **primary outreach materials** (brochures, posters and flyers) produced in the ATLAS workshops and used for dissemination among target users. According to the project Work Plan, the detailed description of the dissemination materials had been done in Deliverable 8.2: Dissemination kit, submitted in month 31 of the project implementation.

The **seventh** chapter presents the **performed dissemination activities** including publication in online media, scientific publication at international events and workshops, participation in exhibitions and large forums to demonstrate ATLAS prototypes and participation in other events. The most important events are listed and described in this chapter. All events, performed during the three project years, are included in the Annex 1 enclosed to this document as its integral part.

The **eighth** chapter is about the **interaction activities with Atlas communities** performed through organisation of events dedicated to the **industry and academia, workshops with users** and **synergies** with relevant projects and networks. The **ninth** chapter describes ATLAS achievements towards expected **technological and social impact**.

The document contains the following Annexes:

Annex 1: List of dissemination events for the project duration

Annex 2: List of dissemination materials produced

1. Dissemination Strategy

1.1 Objectives

The overall broader objective, to which the dissemination will contribute, is to spread knowledge about the project efforts and gained experience in using the developed tools and infrastructure among the members of the target communities, both in Europe and worldwide and to provide a basis for sustainable exploitation of the project results long time after its expiration.

Based on this, the following specific objectives were defined:

- To promote the ATLAS platform;
- To encourage the use of i-Publisher, i-Librarian and EUDocLib;
- To demonstrate the real value of the ATLAS platform and the online content service;
- To disseminate project results through various channels – advertising materials, publications, workshops, etc.;
- To present the project achievements on both national and international forums.

1.2 Target Users

The ATLAS platform as a whole and also some of its standalone components are beneficial to different groups of users. Thus, the consortium classified the potential users of each major software component into several target groups. This allowed the team to pay special attention to the specific needs and requirements of each of these groups. The table below summarises this distribution made:

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

EUDocLib(thematic content-driven web site built with i-Publisher) The general public

Table 1: Target users

1.3 Dissemination tools

Efficient dissemination requires making use of a variety of **dissemination tools**. The ATLAS Dissemination Plan developed under Task 8.2 defined specific dissemination tools for each of the target groups.

Variety of major national, European and international forums, organised both by other institutions and projects and by the ATLAS consortium itself, played an important role for the active participation of ATLAS members in them. These forums can be structured as:

- Exhibitions and fairs;
- Conferences, seminars and workshops.

The tables below present the types and objectives of the events, the target groups attended them and the actions undertaken by the ATLAS members:

Events, organised by other institutions/projects in which ATLAS partners have participated:

Event	Target groups	Objective	Action
Exhibitions and fairs	Industry, IT community, Potential Users	Exploring possible ways of exploitation of the ATLAS service by demonstrating their real benefits to the different target groups.	Technology and Tool Demonstration Use case presentations www Leaflet Distribution
International and National conferences, Workshops and Seminars	Policy makers, Scientific community, general public, Users	Raise awareness of the ATLAS services by presenting the achievements of the project to the content providers, linguists, relevant representatives of the scientific community.	Scientific papers, Poster Leaflet Distribution

Dissemination events organised by the ATLAS partners:

Event	Target groups	Objective	Action
International workshops	Industry, IT community, policy makers	Presenting the achievements of the project and sensitizing the industry about prospects on the European market, dynamic networked organizations and the value of strategic international	Technology and Tool Demonstration Use case presentations www

		partnerships	
	Scientific community	Presenting the achievements of the project to the content providers, linguists, relevant representatives of the scientific community.	
User-targeted events at national level	Potential Users (Table 1)	Presenting the ATLAS services, their real benefits to the different target groups and exploring possible ways of their exploitation.	Technology and Tool Demonstration Use case presentations
User evaluation workshops	Focus groups for evaluation	Used for disseminating ATLAS results and attracting users to participate in the Living Lab activities.	Demonstration and real testing
Smaller adhoc workshops	All target users	An arena for information dissemination and experience sharing. Providing a forum for discussion, information sharing, and social exchange, experience sharing, and collecting (external) advice.	Use case and Best practice presentations
Internal seminars		Informing partner staff about the project.	Presentations, Leaflets, Initiatives for students of the Universities participating to the ATLAS project, Internal email lists

3. ATLAS website

ATLAS website: <http://www.atlasproject.eu/>

According to the project work plan, in Month 2 a portal was created to promote the ATLAS project. It is dynamically updated and has several important functions: From one side, it plays role of the main project information repository and from another - it is a tool for dissemination of news. Another important function of the portal is the provision of an access point to the ATLAS platform and its repositories.

ATLAS website provides direct access to **three web-based services**, produced by the Atlas multilingual platform:

- i-Librarian
- i-Publisher
- EUDocLib

ATLAS website also provides direct access to the websites for demonstration of ATLAS and ASSET functionalities.

4. ATLAS services' websites

4.1 i-Publisher website

i-Publisher website: <http://i-publisher.atlasproject.eu/>

i_Publisher is a novel software-as-a-service solution for web content management, which allows both small and large organisations to deploy and manage multilingual web sites without spending time and efforts for installing and maintaining a content management system.

Additional in order to contribute to the marketing objectives, a website of I-Publisher was created. It presents the functionalities offered by i-Publisher service, the needs of the main target groups and the benefits that the service offers.

4.2 i-Librarian website

i-Librarian website: <http://www.i-librarian.eu/>

i-Librarian is a web-based content assistant service, which allows users not only to store, organize and publish their personal works but also to locate similar documents in English, Bulgarian, Greek, Romanian, German and to easily obtain the most essential texts from large collections of unfamiliar documents or search engine results.

4.3 ATLAS Demonstration websites

Different services produced with i-Publisher and demonstrating the functionalities provided by ASSET were produced in order to show the real benefits to the target group of users and customers. Demonstration websites are the main tool in the ATLAS exploitation strategy.

Target customers of these services are:

- Libraries;
- Publishing houses;
- Media and PR agencies;
- International organisations;
- Corporate clients.

MediaTalk service (<http://mediainage.tetracom.com/>). Customers of this service can sample the media flow for their public image in the press. The news is screened in real time using the company's profile so that the user can get company related news only. This approach differs significantly from the current one which matches keywords to filter the news. ASSET monitors the company's website to detect changes that automatically affect the company's profile. Any news reporting the appointment of a new CEO or the start of a new project will be acknowledged by ASSET.

Video materials demo version, <http://videodemo.atlasproject.eu/>

Plagiarism service: <http://textmatch.eu/>

News Service is another service developed and a demo of it is available at: <http://newsdemo.atlasproject.eu/>. News service demo is multilingual. It presents news in English, Bulgarian, and Greek languages. News service is suitable for media.

Both demonstrations of News service and **World Health Organization website** (<http://asset.atlasproject.eu/who>) show semantic indexing and automatic categorization of ASSET Language Framework. The automatically generated blocks "Topics in the focus", "Hot Regions, "People in the focus" are based on the text analysis. Every publication is enriched with a list of similar items, together with the list of people, organizations and important concepts used in the text which increases the reader's receptive capacity.

Another demo web site presents a set of news, publications, and events from **UNESCO website**, which have been semantically annotated and structured in dynamic blocks by the linguistic platform ASSET: <http://unesco.atlasproject.eu/>

Demo of the **UNESCO Interfaculty Chair website** is available at: http://i-publisher.atlasproject.eu/itd/unesco_chair/en. A multilingual website of UNESCO Interfaculty Chair "ICT in Library Studies, Education and Cultural Heritage" was created with i-Publisher. It presents the Department activities, maps the education workflow and organizes multilingual digital content in a library. The granular user access rights ensure the fluent and secure workflows. The content editors work only on the content they are responsible for.

Icon-painters from Bundovtzi family website –

(http://www.atlasproject.eu/asset_demo/icons/bg/index.html) is made using i-Publisher and the technologies of the ATLAS platform.

EUDocLib (<http://eudoclib.atlasproject.eu/>) addresses the needs of people who require easier access to EU law documents from the EUR-LEX collection in their own language. Users of this web site can easily find similar documents, read the summaries of desired documents, or read extracted important phrases and words or their translations in the user's language.

PLDocLib (<http://www.atlasproject.eu/pl/>) is a language processing chain-powered website offering full-text search and category-based browsing of around 1000 acts of Polish Sejm (the lower chamber of the Polish Parliament). For each document a set of recognized named entities, automatically clustered important noun phrases (with their weights) and a list of similar documents is produced. For presentation, base forms of multiword units are generated and manually assigned categories (retrieved from the document source) are used.

5. MLeCeL Living Lab

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>).

Living Lab website, <http://livinglab.itd-bg.eu/>, is a communication tool for the members of the Living Lab communities and it is very appropriate environment for attracting these participants in the ATLAS communities as well. It is going to be used further as one of the main dissemination tools after the project expiration in order to keep the communities informed about latest developments of the ATLAS services.

MLeCeL website was developed in September 2011. The following activities were performed:

- Communication tools developed within LL website – groups (November 2011)
- Testing and Evaluation questionnaire developed and published on the LL website (November 2011) for collecting feedback from users, developers and experts - <http://livinglab.itd-bg.eu/content/testing-evaluation-questionnaire>
- Update of the Living lab platform with current LL activities, including dissemination events, results, feedback, reports from testing, user suggestion box.

MLeCeL Activities

MLeCeL Living Lab activities performed are:

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. Collection of **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 **three experiments in the Living Lab** and collecting feedback from LL experts and developers.
4. Identifying new companies, users and networks relevant as ATLAS partners and to involve them as registered users in the ATLAS Living Lab (LL) platform.
5. Designing 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. Organisation of **exploitation events and marketing activities**.
8. Strengthening the cooperation between MLeCeL LL and other Living Labs

6. Primary outreach materials

In order to increase the visibility of the ATLAS project and its identity among potential end-users and public and to facilitate their access to project services, variety of dissemination materials were produced within the project. Below a summary of the developed materials is

presented. Their detailed description has been done in D8.2 Dissemination kit according to the project work plan.

6.1 ATLAS brochures

In order to raise awareness about ATLAS project among various target groups, the following brochures were produced and disseminated within the project:

- Two versions of the ATLAS project brochure – one general brochure designed and used at the beginning of the project and one marketing brochures for project focused on the benefits of the ATLAS services and innovative aspects of the ATLAS solutions;
- i-Publisher and i-Librarian brochure for marketing purposes.

The brochures of Atlas, i-Librarian and the general project brochure were developed by TETRACOM and they are available on the project website:

<http://www.atlasproject.eu/atlas/project/dissemination/en>

6.2 ATLAS posters

Different posters were developed for visualising and disseminating the project results. They were introduced at conferences and exhibitions for better presentation of the developed services, used technologies, present and further benefits, and widening the scope of the potential users.

1. Poster addressed to the MT Service in ATLAS at EAMT 2012 titled: Embedding Machine Translation in ATLAS Content Management System. It had been presented at EAMT 2012, the 16th Annual Conference of the European Association for Machine Translation, 28-30 May 2012 Trento;
2. Poster presenting ATLAS project and its services at LREC 2012, 21-27 May, Istanbul, Turkey, titled: Merging heterogeneous resources and tools in a digital library, Anelia Belogay, Diman Karagiozov, Cristina Vertan, Svetla Koeva, Adam Przepiórkowski, Maciej Ogrodniczuk, Dan Cristea, Eugen Ignat and Polivios Raxis;
3. Another ATLAS poster was produced and presented at the annual conference of the European chapter of the Association for Computational Linguistics on 25.04.2012 in Avignon, France. The idea of the poster and the accompanying demonstration was 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. The theme of the ATLAS demonstration was "Harnessing NLP Techniques in the Processes of Multilingual Content Management". Reference URL: <http://eacl2012.org/system-demonstration/index.html>

The poster is publicly available at the ATLAS project website: <http://www.atlasproject.eu/atlas/project/news/details/en?ci=005926e9-4346-4cc3-ae4b-c742e07dbe9f>;

4. ATLAS poster presented at RANLP 2011 titled: Training Data in Statistical Machine Translation – The More, the Better, developed by Monica Gavrila, Cristina Vertan,

- from University of Hamburg, Germany. RANLP 2011, Recent Advances in Natural Language Processing was held on 12-14 September 2011, Hissar, Bulgaria;
5. ATLAS poster presented at META-FORUM on June 27/28 2011 in Budapest, Hungary. The poster presents the ATLAS solution and its innovative aspects, i-Publisher and i-Librarian services, functionality description, target groups: <http://www.atlasproject.eu/atlas/project/news/details/en?ci=005926e9-4346-4cc3-ae4b-c742e07dbe9f>;
 6. ATLAS poster presented at GSCL 2011 Conference titled: ATLAS a Robust Multilingual Platform for the Web. This poster was developed by Maciej Ogrodniczuk, Cristina Vertan, Svetla Koeva and Adam Przepiorkowski. GSCL 2011 Multilingual Resources and Multilingual Applications, Conference of the German Society for Computational Linguistics and Language Technology (GSCL) 2011 was held on 28 September - 30 September 2011, Hamburg University, <http://www.corpora.uni-hamburg.de/g scl2011/en/?download=GSCL-Programm.pdf>;
 7. Poster related to the MT Service in ATLAS presented at EAMT 2011 titled: Machine Translation in ATLAS: Applied Technology for Language Aided Content Management Systems. EAMT 2011, the 15th Annual Conference of the European Association for Machine Translation was held on 30-31 May 2011, Leuven, Belgium;
 8. ATLAS poster in German language titled: ATLAS Intelligent Web Content Management System;
 9. CMS – Project poster presented at the European Summit on Machine Translation, Leuven, June 2011
 10. ATLAS posters in Bulgarian and English were shown at the Permanent exhibition of the Bulgarian Academy of Sciences by DCL – IBL in Sofia.

6.3 ATLAS flyers

ATLAS flyers, produced by the partners, have been used to present the project and to disseminate the ATLAS innovative advantages among potential target users in their countries.

- ATLANTIS partner has made a flyer for presenting ATLAS project objectives and expected results. The flyers were distributed during different the dissemination activities carried out;
- ITDF partner developed two flyers devoted to i-Publisher and i-Librarian services. The two flyers were distributed during project dissemination events.

7. Performed Dissemination Activities

In order to reach to the target customers and users and to maximize the visibility of the project, ATLAS partners used different channels and tools and cooperated with other relevant national and international organisations and projects, sharing similar objectives. A list of appropriate events and conferences that bring together the relevant stakeholders had been preliminary developed in order to facilitate the selection of the most important channels for disseminating ATLAS information and achievements.

During the **1st year**, the Dissemination activities included **preparation of brochures** and **website development**. Partners mainly distributed dissemination materials and presented

the ATLAS project at national and international events. The objective was to raise awareness of the ATLAS by presenting the achievements of the project to the content providers, linguists, relevant representatives of the scientific community and potential users.

During the **2nd year**, the results of the Atlas project were made available through scientific papers to the target users. Newsletters and articles addressed the users and the wide public as well. Through technology and tool demonstrations, case study presentations, the ATLAS services were presented at relevant exhibitions and fairs in order to explore possible ways of their exploitation by demonstrating their real benefits to the different target groups.

Based on the gained achievements, the dissemination materials developed during the **3rd year** became more reach in content and more varied. These services' presentations, demos and scientific papers were disseminated at national and international events, with the focus on large conferences and exhibitions.

In addition, each partner organised events at national level, targeted at potential users and customers. The aim of these events was to present the ATLAS services, their real benefits to the different target groups and exploring possible ways of their exploitation and collaboration.

The following dissemination activities were performed during ATLAS project implementation:

- Distribution of introducing materials on national and international conferences, workshops, and meetings;
- Online publications (articles and newsletters) and scientific articles to reach target audiences over a long period of time;
- Participation in national and international events and exhibitions to demonstrate ATLAS achievements among potential users;
- Organisation of target workshops to increase visibility of the ATLAS services among potential customers – individuals and organisations and to strengthen the partnership between ATLAS members and the potential participants in the established ATLAS communities;
- Synergies with other relevant EU projects through presentation of the Atlas project at project meetings and organisation of joint workshops in order to establish closer collaboration between ATLAS project and other EU projects in the relevant area.

The most important dissemination activities are described in the next chapters of the present report (D8.3). Complete list of all dissemination activities, performed by the partners during the project duration, are also presented in **Annex 1**.

7.1 Publications in online media and journals

The ATLAS project results are presented in the following articles, press releases, newsletters and other publications addressed to the users and the general public, are:

- Cristea, D., Ignat, E., Anechitei, D.: "ATLAS Project: The romanian Component", in BringITon! 2012 Catalog. Edited by Alboaie, L., Cristea, D., Forăscu, C., Gîfu, D. ISSN 2285-0929, ISSN-L: 2285-0929, Pp. 8-9, Iasi, 17-18 May, 2012.
- ATLAS project description, objectives, list of participants in Polish Potential and Contribution to EU Success in ICT, pp. 200-201.
See <http://www.kpk.gov.pl/pliki/12357/pc-net.pdf>

- A paragraph on ATLAS in a broader text "UE supports processing of Polish" (in Polish in Biznes raport – a publication about innovative projects added to Polish daily newspaper Dziennik – Gazeta Prawna
- ATLAS project description, objectives, list of participants in Official Polish record of ICT projects: Polish Potential and Contribution to EU Success in ICT; pp. 200-201. See <http://www.kpk.gov.pl/pliki/12357/pc-net.pdf>
- EU fördert cross-linguale Suchmaschine der Arbeitsstelle Computerphilologie, Dr. Cristina Vertan, Article in German in Online-Newsletter der Universität Hamburg from April 2010, Nr. 13, pp 8
- A paragraph on ATLAS in a broader text "UE supports processing of Polish" (in Polish) in Biznes raport – a publication about innovative projects added to Polish daily newspaper Dziennik – Gazeta Prawna.
- ATLAS article in UHH Hochschulmagazin (in German), page 40
<http://www.uni-hamburg.de/newsletter/EU-foerdert-cross-linguale-Suchmaschine-der-Arbeitsstelle-Computerphilologie-.html>
- Presentation of ATLAS Project in electronic brochure in Program CIP-ICT PSP Przykłady polskiego uczestnictwa, prepared by the Institute of Computer Science, Polish Academy of Sciences
http://www.kpk.gov.pl/pliki/11429/Broszura_wersja_elektroniczna.pdf
- Press release of ATLAS in the electronic newsletter of ATLANTIS (more than 85.000 recipients).
- An article about ATLAS project was published by ICS PAS on the Polish portal for Success Stories:
http://en.kpk.gov.pl/index.php?option=com_sobi2&catid=5&Itemid=142&lang=pl
The portal presents Polish successful projects financed under FP5, FP6 , FP7 and the CIP programme.

7.2 Scientific publications

Major ATLAS issues and results achieved are analysed in variety of reports, articles, monographs published in different scientific and technical journals worldwide, electronic books, collections and other editions of scientific relevance and presented at annual conferences of the world's major professional communities for content management, and other relevant areas.

The results of the Atlas project are made available through 31 articles and papers, published in scientific and technical journals and conference proceedings.

1. Daniel Anechitei, Dan Cristea, Ioannidis Dimosthenis, Eugen Ignat, Diman Karagiozov, Svetla Koeva, Mateusz Kopeć, Cristina Vertan (2013, to appear). **Summarizing Short Texts through a Discourse-Centered Approach in a Multilingual Context** in Neustein, A., Markowitz, J.A. (eds.), *Where Humans Meet Machines: Innovative Solutions to Knotty Natural Language Problems*. Springer Verlag, Heidelberg/New York.
2. Cristina Vertan (2013), **Melania Duma, Integration of Machine Translation in On-line Multilingual Applications – Domain Adaptation**, article in journal *Open Access Journal Translation: Computation, Corpora, Cognition* (<http://www.t-c3.org>)

3. Anelia Belogay, Damir Ćavar, Dan Cristea, Diman Karagiozov, Svetla Koeva, Roumen Nikolov, Maciej Ogrodniczuk, Adam Przepiórkowski, Polivios Raxis and Cristina Vertan (2012) **i-Publisher, i-Librarian and EUDocLib - linguistic services for the Web**. P. Pęzik (ed.): Corpus Data across Languages and Disciplines, Lodz Studies in Language Series, vol. 28, pp. 203-212, Peter Lang. ISBN 978-3-631-62547-7.
4. Anelia Belogay, Diman Karagyozov, Svetla Koeva, Cristina Vertan, Adam Przepiórkowski, Polivios Raxis, & Dan Cristea: **Harnessing NLP technologies in the processes of multilingual content management**. [EACL 2012] Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France, April 23-27, 2012; pp. 6-10. <http://www.mt-archive.info/srch/authors-C.htm>
5. Anelia Belogay, Diman Karagiozov, Cristina Vertan, Svetla Koeva, Adam Przepiórkowski, Maciej Ogrodniczuk, Dan Cristea, Eugen Ignat and Polivios Raxis (2012): **Merging Heterogeneous Resources and Tools in a Digital Library**, Proceedings of the "Merging LR" workshop at 8th International Conference on Language Resources and Evaluation (LREC 2012), pp. 41-44. ELRA, Istanbul, Turkey, www.lrec-conf.org/proceedings/lrec2012/workshops/06.LREC%202012%20Merging%20Proceedings.pdf
6. Ignat, E. and Anechitei, D.: **"Multilingual summarisation system based on analysing the discourse structure"** in Language Technologies in Romanian and Diaspora Research & Development. Edited by Forăscu, C., Ioniță, A., Tufiș, D., Cristea, D. ISBN 978-973-703-813-5, Pp 34-35, Bucharest, 26-27 September, 2012.
7. Koeva, Svetla (2012): **ATLAS - Human Language Technologies integrated within a Multilingual Web Content Management System**. In Proceedings of the Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra), Avignon: Association for Computational Linguistics, 2012, pp. 72-76. ISBN: 978-1-937284-19-0. <http://dl.acm.org/citation.cfm?id=2387966&CFID=250398417&CFTOKEN=38583902>
8. Koeva, Svetla, Ivelina Stoyanova (2012): **Toward automatic generation of parallel corpora**. In Proceedings from the 8th International Conference Formal Approaches to South-Slavic and Balkan Languages. September 19-21, 2012, Dubrovnik, Croatia.
9. Cristina Vertan (2012): **Machine Translation support for digital libraries - Problems, limitations and practical solutions**, submitted at Digital Humanities 2012, July 2012, Hamburg
10. Cristina Vertan (2012): **Two approaches for integrating translation and retrieval in real applications**. EACL Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches

- to Machine Translation (HyTra): Proceedings of the workshop, 23-24 April 2012, Avignon, France; pp. 59-64.
11. Cristina Vertan (2012): **Embedding Machine Translation in ATLAS Content Management System**, EAMT 2012 , Trento, 28-30 May 2012
 12. Cristina Vertan, Walther v. Hahn, Monica Gavrila (2012): **Same domain different discourse style-A case study on Language Resources for data-driven Machine Translation**, LREC 2012, 21-27 May, Istanbul
 13. Cristina Vertan (2012), **Integrarea unui sistem de traducere automată în platforme multilinguale exemplificată în cazul perechii de limbi Română-Germană**, abstract in Workshop Booklet of Diaspora in Cercetarea Stiintifica si Invatamantul Superior din Romania, 26-27 September 2012, Bucharest, Romania
 14. Dan Cristea, Eugen Ignat și Daniel Anechitei (2012): **"ATLAS Project - the summarisation system"**, at ConsILR - The International Conference on "Resources and Tools for Romanian Language", Bucharest, April 2012.
 15. Radu Simionescu (2012): **"Graphical grammar studio as a constraint grammar solution for part of speech tagger"**, at ConsILR – The International Conference on "Resources and Tools for Romanian Language", Bucharest, April 2012.
 16. Ogrodniczuk, Maciej and Przepiórkowski, Adam (2012) **Polish Language Processing Chains for Multilingual Information Systems**. G. Bouma et al. (ed.): NLDB 2012, LNCS 7337, pp. 152-157. Springer, Heidelberg.
 17. Diman Karagiozov, Anelia Belogay, Dan Cristea, Svetla Koeva, Maciej Ogrodniczuk, Polivos Raxis, Emil Stoyanov, and Cristina Vertan (2012). **i-Librarian – Free Online Library For European Citizens**, in INFOtheca – Journal of Librarianship and Informatics, no. 1, vol. XIII, May, BS Print, Belgrade, pages 27-43, ISSN: 1450-9687.
 18. Maciej Ogrodniczuk and Diman Karagiozov (2011): **"ATLAS – Multilingual Language Processing Platform"**, SEPLN 2011 conference. The conference proceedings were published in Procesamiento del Lenguaje Natural journal, vol. 47, 2011. ISSN 1135- 5948, pp. 241–248. More information: <http://www.uhu.es/sepln2011/index.php/programme.html#dos>
 19. Diman Karagiozov (2011): **"Integration of Natural Language Processing Chains in Content Management Systems"**, S3T 2011 Conference, in Advances in Intelligent and Soft Computing Vol. 101, Springer. <http://www.atlasproject.eu/atlas/project/dissemination/en>
 20. Maciej Ogrodniczuk (2011): "i-Publisher, i-Librarian and EUDocLib - Linguistic services for the Web", PALC 2011
 21. Degórski, Łukasz (2011) **Towards the Lemmatisation of Polish Nominal Syntactic Groups Using a Shallow Grammar**. P. Bouvry et al. (ed.): SIIS 2011, LNCS 7053, pp. 370-378. Springer, Heidelberg.
 22. Anelia Belogay, Dan Cristea, Eugen Ignat, Diman Karagiozov, Svetla Koeva, Maciej Ogrodniczuk, Adam Przepiórkowski, Polivos Raxis and Cristina Vertan (2011): **"Language Processing Chains in ATLAS"**, 5th Language & Technology Conference (LTC 2011): Human Language Technologies as a Challenge for Computer Science and Linguistics, 25-27 November 2011, Poznan, Poland.

23. M. Ogrodniczuk and D. Karagiozov (2011): **ATLAS – Multilingual Language Processing Platform**, SEPLN 2011: 27th Conference of the Spanish Society for Natural Language Processing, 5-7 September 2011, Huelva, Spain
24. C. Vertan (2011): **Machine Translation in ATLAS - Applied Technology for Language Aided**
25. C. Vertan, M. Gavrilă (2011): **Using manual and parallel aligned corpora for Machine Translation Services within an Online Content Management System**, EPC 2011: proceedings of the Second Workshop on Annotation and Exploitation of Parallel Corpora, associated with the 8th International Conference on Recent Advances in Natural Language Processing (RANLP 2011), 15th September 2011, Hissar, Bulgaria; pp.53-58.
gavrilă, vhahn@informatik.uni-hamburg.de, cristina.vertan@uni-hamburg.de
26. P. Raxis (2011): **Towards the integration of monolingual NLP frameworks for multilingual applications: A case study for the Greek Language within the ATLAS project framework**, GSCL 2011 Conference, 28-30 September, Atlantis, Greece.
27. Radu Simionescu (2011): **"Hybrid POS tagger"**, at the international Summer School "NLP Technology Goes Industrial" EUROLAN 2011, Workshop on "Language Resources and Tools with Industrial Applications", August- September 2011, Cluj-Napoca.
28. Dan Cristea (2011): **"Romanian Linguistic Resources On Very Large Scale"**, in IIS 2011 - International Workshop on Intelligent Information Systems, Chişinău, 13-14 September.
29. Dan Cristea, Eugen Ignat and Daniel Anechitei (2011): **"ATLAS Project - the language processing chain"**, ConsiLR - The International Conference on "Resources and Tools for Romanian Language", Bucharest, December 2011.
30. Anechitei, D., Cristea, D., Ignat, E.,: **"Clause Level Multilingual Segmentation"**, at ConsiLR - The International Conference on "Resources and Tools for Romanian Language", Bucharest, December 2011.

Publication in Books:

31. Cristina Vertan, Walther v. Hahn, **Multilingual Processing in Eastern and Southern EU Languages: Low - resourced Technologies and Translation – to appear in Cambridge Scholar Press, June 2012**

7.3 Exhibitions and forums

ATLAS team used the opportunities of relevant **exhibitions and scientific events** to disseminate the project results by giving **presentations and performing demonstrations**, thus reaching potential customers and users. The most important events of ATLAS consortium are presented as follows:

1. CeBIT 2012, 6-10 March 2012, Hanover, Germany

Event Type: 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

Event type: 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, Plovios Raxis (2012) Harnessing NLP Techniques, in Walter Daelemans, Mirella Lapata Lluís 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"

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

Event type: International conference and exhibition

Target users: Computational linguists, industry people, mass media

Participants: 300 +

Performed dissemination activities:

- Paper presentation at LREC 2012

Cristina Vertan, Walther v. Hahn, Monica Gavrilă, 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. CESAR META-NET Road show 2012 Sofia, Bulgaria May 2nd 2012

Event type: National event

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:

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.

- 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 event and presentation:

http://dcl.bas.bg/en/cesar_roadshow_program_en.html

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

Event type: 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.

6. Human Language Technology Days 2012, 27-28 September 2012, Warsaw, Poland

Event type: Exhibition, project village

Target users: linguists, computational linguists, IT experts

Participants: 200

Performed dissemination activities:

ATLAS exhibition booth of ICS PAS. This is an Event presenting projects related to CESAR ICT PSP project with the participation of linguists, computational linguists, IT experts.

7. 2nd Pan European Private Equity and Start-ups Forum, 21-22 June 2012, Athens, Greece

Event type: International event

Target users: Executives from Business Angels, Venture Capitals, Banks, Start-ups, including also entrepreneurs and researchers

Performed dissemination activities:

Discussions on funding opportunities for new online services based on specific ATLAS results (i-Publisher, automatic multilingual summarization)

Distribution of ATLAS brochure and custom flyer

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

Event type: 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, Chair holder, UNESCO Interfaculty Chair on ICT in Library Studies, Education and Cultural Heritage

9. International Exhibition of Software Bucharest, 7-9 November, 2012

Event type: International exhibition

Target users: Decision makers and IT managers

Performed dissemination activities:

The Binary Conferences are a series of events that aim to present specific IT&C solutions for the SMB sector. Binary targets decision makers and IT managers from different areas.

<http://binary.aries.ro/en/>

10. META-FORUM 2011, 27-28 June 2011, Budapest, Hungary

Event type: International event

Target users: Key players in Language Resources and Technologies

Performed dissemination activities:

Presentation of poster (<http://www.atlasproject.eu/atlas/project/dissemination/en>) and online demonstration was made by ICS PAS and DCL – IBL among key players in Language Resources and Technologies.

META-FORUM 2011 – Solutions for Multilingual Europe is an exhibition space in which the participants display and demo various aspects of the work being done across the entire META community and beyond. The exhibition reflects both the research and industry aspects of the community: <http://www.meta-net.eu/events/meta-forum-2011/exhibition>

11. 5th Language & Technology Conference (LTC 2011): Human Language Technologies as a Challenge for Computer Science and Linguistics, 25-27 November 2011, Poznan, Poland

Event type: International conference

Target users: IT scientists, computational linguists

Performed dissemination activities:

- Demonstration of i-Publisher, i-Librarian, EUDocLib and PLDocLib at Demo session.
- Software and resource demonstration, information about the project and roll-up exhibition.
- Paper: Language Processing Chains in ATLAS by Anelia Belogay, Dan Cristea, Eugen Ignat, Diman Karagiozov, Svetla Koeva, Maciej Ogrodniczuk, Adam Przepiórkowski, Polivos Raxis and Cristina Vertan: demonstrating i-Publisher, i-Librarian, EUDocLib and PLDocLib

12. WEBIT'11, 26-27 October 2011 in Sofia, Bulgaria

Events type: Exhibition

Target users: Web developers, e-commerce, hosting, content management and CRM provider, ad network, affiliate system and professionals from CEE and the rest of the world.

Performed dissemination activities:

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 included as well.

13. 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

Event type: European congress

Target users: Business Angels from all over Europe

Performed dissemination activities:

Discussions on funding possibilities for ATLAS take up made by ATLANTIS.

7.4 Conferences, workshops and seminars

The ATLAS project results were **presented using PowerPoint and oral presentations, posters and scientific papers at national and international conferences, workshops and seminars** among different target users.

Complete list of conferences, workshops and seminars is presented in Annex 1.

The most important events are listed below:

1. University of Winterthur, Winterthur, Switzerland, February 2013

Event type: Seminar

Target users: Professional Translators

Performed dissemination activities:

Presentation and demonstration made by Cristina Vertan, Walther v. Hahn, Embedding Language Technology Tools in Web Applications - Focus on Machine Translation and Domain Adaptation

2. University of Buenos Aires, University of Cordoba, Argentina and University of Santiago, Santiago de Chile, Chile, 17-29 November 2012

Event type: International workshop and seminars

Target users: NLP Researchers and Students

Performed dissemination activities:

- Invited talk and demonstration made by Cristina Vertan, Walther v. Hahn, Embedding Language Technology Tools in Web Applications - Challenges, current limitations, and further perspectives

3. 17th International conference on Applications of Natural Language Processing to Information Systems (NLDB 2012), 26-28 June 2012, Groningen, The Netherlands

Event type: International conference

Target users: Scientific and business participants

Performed dissemination activities:

Ogrodniczuk, Maciej and Przepiórkowski, Adam. Polish Language Processing Chains for Multilingual Information Systems. G. Bouma et al. (ed.): NLDB 2012, LNCS 7337, pp. 152-157. Springer, Heidelberg.

4. EAMT (Summit of the European Association for Machine Translation), 28-30 May 2012, Trento

Event type: International conference

Target users: Linguists, computational linguists, IT experts

Performed dissemination activities:

Paper titled: Cristina Vertan, Embedding Machine Translation in ATLAS Content Management System

5. Conference on Resources and Tools for Processing Romanian Language - ConsILR-2012, 27 April, 2012, and 2011, Bucharest, Romania

Event type: National conference

Target users: Linguists, computational linguists, IT experts

Performed dissemination activities:

Presentation of Dan Cristea, Eugen Ignat și Daniel Anechitei: Proiectul ATLAS (in Romanian) and three papers.

6. NLP Seminars on Computational Linguistics, Faculty of Computer Science, "Al. I. Cuza" University of Iasi, 3 May 2012 and 10 May 2012, Iasi, Romania

Event type: Series of seminars

Target users: Master students in Computational Linguistics

Performed dissemination activities:

Presentation of the modules: RARE (anaphora resolution), clause segmentation and discourse parser, by Eugen Ignat and Daniel Anechitei

7. Cross lingual Language Technology 2012, 7-8 May 2012, Univ. Hamburg, Germany

Event type: Conference

Target users: Linguists, computational linguists, IT experts

Performed dissemination activities:

Presentations made by:

- Dan Cristea and Ionut Pistol titled: Multilingual linguistic workflows. A case study for two languages
- Cristina VERTAN: Machine Translation in a content management system – a multilingual case study involving Polish, Bulgarian, Romanian, and Greek
- Degórski, Łukasz. Towards the Lemmatisation of Polish Nominal Syntactic Groups Using a Shallow Grammar. P. Bouvry et al. (ed.): SIIS 2011, LNCS 7053, pp. 370-378. Springer, Heidelberg.

8. Bring-IT on! 2012, 17-18 May 2012, Iasi, Romania

Event type: Workshop for prospecting the links between university research and industry

Target users: Linguists, computational linguists, IT experts

Performed dissemination activities:

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

9. Corporate business day (Atlantis group of companies) with the participation of managers from all over Greece and Cyprus, 10 February 2011 Thessaloniki, Greece

Event type: Business development meeting

Target users: New business development managers

Performed dissemination activities:

The 2 ATLAS online services were presented by Atlantis and discussions took place about business development and commercial exploitation

10. MOBIP 2010 and Investment in Mobile and IT Services, 10-16 June 2010, Valencia, Spain

Event type: International business and investment event

Target users: Stakeholders and experts from the financing, business consulting, IT, etc. sectors

Performed dissemination activities:

Exchange ideas about exploitation prospects and distributed project material by Atlantis.

11. Corallia Conference for the mi-clusters.job, 20 December 2010, Athens, Greece

Event type: Conference

Target users: Business experts, IT experts

Performed dissemination activities:

ATLANTIS distributed project 20 flyers in English. Discussions took place about the possibility to use ATLAS for online publishing of documents from microelectronics projects of the Greek mi-cluster.

12. Conference on “measures to support enterprises in implementing investments in digital solutions”. 17 January 2011, Drama, Greece

Event type: Conference

Target users: Stakeholders and experts from the financing, business consulting, IT experts

Performed dissemination activities:

- Short presentation of the project, along with the distribution of 60 EN flyers.
- Investigation of the financial environment that the support program Digital Convergence offers in relation to ATLAS.

13. FASSBL07: The Seventh International Conference Formal Approaches to South Slavic and Balkan Languages, 3-7 October 2010, Dubrovnik, Croatia

Event type: Conference

Target users: International research community, representatives from other ICT PSP projects

Performed dissemination activities:

Presentation of the ATLAS project; ATLAS poster; distribution of the ATLAS brochures made by Institute for Bulgarian Language, Bulgaria, <http://hnk.ffzg.hr/fassbl2010/>

14. LREC 2010, 17-23 May 2010, Valetta, Malta

Event type: Conference and workshops

Target users: Linguists

Performed dissemination activities:

Initial presentation of the project made by Institute of Computer Science, Polish Academy of Sciences at the conference

Presentation of ATLAS project and distribution of the brochure made by University of Hamburg, Germany at LREC 2010 Workshop on “Exploitation of multilingual resources and tools for central and (south-) Eastern European languages.

15. Bring-IT on! 2010, 10 December 2010, Iasi, Romania

Event type: Workshop for prospecting the links between university research and industry

Target users: Linguists, computational linguists, IT experts

Performed dissemination activities:

Presentation of the project, poster session and distribution of the brochures made by Alexandru Ioan Cuza University, Romania

8. Target Communities

Dissemination activities performed in ATLAS project were also oriented towards establishment of target communities: ATLAS open source community, content users, learners, companies, universities. The enterprise community is initiated by the consortium partners Tetracom, Atlantis and ITD. The university network is initiated by the universities and research institutes – members of the consortium. The establishment of ATLAS communities and the effective interaction with- and between their members are important precondition for achieving sustainability and increasing the potential impact of the project.

The members of the Open Source Communities are activated by the ATLAS project partners within the project implementation and are enlarged by establishment a network of exploitation partnerships from where the early adopters of ATLAS will emerge after the project end.

The Open Source Communities includes Communities of Developers (CoD) and Community of End-users (CoEU)¹. CoD consists of contributors who share broad commit access to the open source ATLAS platform. CoEU includes users – people who use the services and whose employers may pay for and put it to productive use.

8.1 Interaction with target communities

The communication and interaction with the different communities was achieved by participation in international events and a number of **coordinated workshops** with both international and national scope, **hosted and organised by ATLAS partners**:

- International workshops for:
 - Industry, IT community, policy makers for presenting the achievements of the project and sensitizing the industry about prospects on the European market, dynamic networked organizations and the value of strategic international partnerships;
 - Scientific community to present the achievements of the project to the content providers, linguists, relevant representatives of the scientific community.
- Bilateral meetings with potential users;
- User evaluation workshops with developers and users and joint events within MleCel Living Lab;
- Smaller adhoc workshops designated to users as forums for discussion, information sharing, and social exchange, experience sharing, and collecting (external) advice;
- Internal seminars for informing the partner staff about the project progress.

8.2 Organised events for industry and scientific community

Demonstration 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 and companies through organisation of workshops and seminars.

CESAR META-NET Road show 2012, Sofia, Bulgaria May 2nd 2012

¹ <http://webmink.com/essays/community-types/>

The workshop is organised by DCL-IBL. The participants were 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.

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

Presentation and programme: http://dcl.bas.bg/en/cesar_roadshow_program_en.html

International Conference: Cross lingual Language Technology 2012, 7-8 May 2012, University of Hamburg, Germany

The conference was partially organized by ATLAS partners and the target audience included Linguists, computational linguists, IT experts.

Presentations were made by:

- Dan Cristea and Ionut Pistol titled: Multilingual linguistic workflows. A case study for two languages
- Cristina VERTAN: Machine Translation in a content management system – a multilingual case study involving Polish, Bulgarian, Romanian, and Greek
- Degórski, Łukasz. Towards the Lemmatisation of Polish Nominal Syntactic Groups Using a Shallow Grammar. P. Bouvry et al. (ed.): SIIS 2011, LNCS 7053, pp. 370-378. Springer, Heidelberg.

Conference programme: <http://www.c-phil.uni-hamburg.de/view/Main/MaiProgramme>

ATLAS Workshop “Integration of multilingual resources and tools in Web applications workshop”, 26 September 2011, Hamburg, Germany

The workshop was organised within the Conference of the German Society for Computational Linguistics and Language Technology (GSCL 2011) in Hamburg, Germany, 26 September 2011.

Main organizers of the workshop were: Tetracom (Diman Karagiozov), Polish Academy of Sciences (Maciej Ogrodniczuk) and University of Hamburg (Cristina Vertan).

The Programme Committee consists of Damir Cavar (University of Konstanz, Germany), Dan Cristea (University of Iasi, Romania), Svetla Koeva (Bulgarian Academy of Sciences), Vladislav Kubon (Charles University Prague), Lothar Lemnitzer (Academy of Sciences Berlin), Adam Przepiorkowski (Polish Academy of Sciences), Polivos Raxis (ATLANTIS, Ltd, Greece), Roland Winnemöller (University of Hamburg).

Target audience consisted of Computational linguists, experts and researchers.

The aim of this workshop is to bring together researchers and users working on and with such types of systems and discuss approaches to:

- pro's and con's of using Web services within content management systems
- monolingual language processing chains integrating heterogeneous components
- minimal linguistic framework for assembling multilingual linguistic processing chains
- reusability of language processing chains
- frameworks for building language processing chains (UIMA, GATE etc.)
- performance versus quality in WCMS systems
- domain adaptability of language processing chains

More information: 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

- ATLAS Poster session @ Conference of the German Society for Computational Linguistics and Language Technology (GSCL 2011): ATLAS — A Robust Multilingual Platform for the Web poster by Maciej Ogrodniczuk, Cristina Vertan, Svetla Koeva, and Adam Przepiorkowski: <http://www.corpora.uni-hamburg.de/gscl2011/en/?download=GSCL-Programm.pdf>
- P. Raxis: “Towards the integration of monolingual NLP frameworks for multilingual applications: A case study for the Greek Language within the ATLAS project framework”.

8.3 Meetings with potential customers

Different meetings with representatives of potential users and customers of the ATLAS platform were organised by Tetracom. ITD contributed with establishing contacts with potential interested organisation to exploit ATLAS and its services. These activities are presented below:

- Meeting with the head of Multilingual technologies department at World Health Organisation Yousef Elbes at IKS Workshop "Semantic Enterprise Technologies in Action", June, Salzburg.
 - Building a demo service for World Health Organisation: <http://asset.atlasproject.eu/who>
 - Presentation of the demo service in Geneva office of the WHO, October, 30 participants.
 - Follow-up paper
- Meeting in Education Unit for use case for using linguistic technologies for multilingual data pools.
 - Building a demo service for Unesco, <http://unesco.atlasproject.eu/>
 - Presentation of the demo service done by prof. Rumen Nikolov in September
- Meeting with the Head of the Department in Journalism, Sofia University and a meeting with a political observer from Bulgarian National Television.
 - Dissemination of the linguistic platform, discussion of the possible usage of the linguistic technologies for the enterprise , Living Lab activity - gathering user requirements
 - Building a demo web site for media, <http://newsdemo.atlasproject.eu/>
- A Living Lab activity: a response to an end-user request for showing similar video objects for the Bulgarian National Television.
 - Building a demo version for video materials: <http://videodemo.atlasproject.eu/>
- Meeting with Account Manager at M3 Communications Group, Inc. A Hill+Knowlton Strategies Affiliate, www.m3bg.com

- Dissemination of the linguistic platform, discussion with a potential commercial customer. of the possible usage of the linguistic technologies for the enterprise , Living Lab activity - gathering user requirements
- Building a press clipping service in Bulgarian and English:
<http://mediaimage.tetracom.com/>
- Meeting with the CEO of Foundation "Educational programmes"
 - Purpose: Gathering requirements, building a strategy for an integration of the service in Moodles system. Contribution to Cesar project
 - Building a plagiarism service: <http://textmatch.eu/>

ATLANTIS has performed or participated in several dissemination activities for paving the way for the exploitation activities:

- Atlantis organised exploitation event targeting representatives of Enterprise Community on 19-21 November 2012. Live presentation and demonstration of the 2 ATLAS services (i-Librarian and i-Publisher) in particular business cases were presented to the Hellenic Mobile Cluster, Athens Technical University, Greek Public Properties Company (ETAD) and big private corporate groups.
- Atlantis presented ATLAS to the European entrepreneurs and international investors and discussed with them opportunities for commercial exploitation of ATLAS achievement during EUREKA Venture Forum, 18-19 October, Istanbul, Turkey

8.4 User evaluation workshops and joint events

Seminars and workshops for interaction with developers and users and for collaboration with other Living Labs were organised within the MLeCeL Living Lab. The objectives of these seminars were: to popularize the ATLAS services among target communities; to share experience and knowledge with other local researchers and developers of multilingual e-content and e-library services; to explore possible exploitation of the ATLAS services; to collaborate with other Living Labs.

The interaction with target communities was carried out through the organisation of:

- User evaluation workshops - IT experts, PhD students and researchers in Software engineering, web developers;
- Joint workshops with members of MLeCeL Living Lab and Digital Spaces Living Lab - web developers, e-commerce companies, hosting companies, content management and CRM providers
- Pilot experiments within Living Lab – web developers and users

All events organised within Living Lab are published on: <http://livinglab.itd-bg.eu/content/events>

Evaluation by Users

During the 3rd Evaluation round (December 2013) ATLAS i-Librarian service was tested by the **Living Lab experts** and **80 users** from the SULSIT and Faculty of Mathematics and Informatics, Sofia University. Both organizations are active members of the MLeCeL Living Lab.

- Interviewing Living Lab experts to provide their feedback about i-Librarian service and used technologies: <http://livinglab.itd-bg.eu/content/3rd-atlas-testing-round-living-lab-experts-feedback>

- Conducting testing workshops with users. All users in the testing filled in the questionnaire on the website of the project: <http://ue.atlasproject.eu/>
- Feedback in a form of interviews from users: <http://livinglab.itd-bg.eu/content/3rd-atlas-testing-round-living-lab-experts-feedback>

Evaluation of ATLAS services was organized in **February and March, 2012** in Sofia as a part of Living Lab activities including: testing workshops with users and interviews with Living Lab experts to provide their feedback about i-Librarian service and used technologies.

One evaluation session 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” 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.

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

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 organised during December 2011 in the 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.

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

Users’ feedback was collected with a questionnaire: <http://livinglab.itd-bg.eu/content/testing-evaluation-questionnaire>

Joint workshops within Living Lab

ATLAS workshop 2012: Applied Technology for Language-Aided CMS, 16 October 2012, Sofia

ITDF organised annual ATLAS workshop, titled ATLAS: Applied Technology for Language-Aided CMS. This annual event was held within the European Day of Entrepreneur 2012 (EDE). The aim of the workshop was to present:

- Possibilities provided by ATLAS project to the publishers and media.
- How ATLAS platform enables publishers, information and graphic designers to easily collaborate as well as saves authors, editors and other contributors

valuable time by automatically processing textual data and allows them to work together to produce high-quality content.

- How ATLAS enables users to find the most essential texts from large document collections by displaying text summaries and extracted important phrases, words and names; improves content navigation by interlinking content items based on text annotations and by automatically placing the content items in appropriate subject categories.

The workshop programme is published on the Living Lab and EDE website: http://ede.uni-sofia.bg/docs/Program_Atlas.pdf

Presentation from the workshop are published on the Living Lab website: <http://livinglab.itd-bg.eu/content/events>

Workshop on the Process of creation and development in Living Labs, 19 September 2012, Veliko Tarnovo.

The workshop presented the Living Labs idea, perspectives, and few Living Labs. The workshop was organised within “Digital presentation and preservation of cultural and scientific heritage” conference - <http://dipp2012.math.bas.bg/>. ATLAS was specially presented.

Joint ATLAS/MLeCel LL/DSL Workshop on Multilingual Digital Repositories and Services, 3 May 2012, Sofia at State University of Library Sciences and IT.

Participants: Professionals in developing Multilingual Digital Repositories and Services.

ATLAS and its services were presented at the workshop by TETRACOM. Other presented projects and products are: CLARIN: Common Language Resources and Technology Infrastructure (BAS), Europeana and Bulgariana (Ontotext), Multimedia newsroom in Bulgaria (TV), Vocabler, Querying RDF Semantic Repositories in Natural Language (Ontotext) and others.

The objective of this workshop was 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.

Programme and all presentations in Bulgarian: <http://livinglab.itd-bg.eu/content/events>

International Workshop: Re-designing Institutional Policies and Practices to Enhance the Quality of Education through Innovative Use of Digital Technologies, June 2011, Sofia

ITD established closed collaboration with SULSIT for planning future partnership activities with ATLAS project.

SULSIT through its UNESCO Interfaculty Chair provided access to its partner organisations: UNESCO Chair, The All Russian State Tax Academy of the Ministry of Finance of the Russian Federation, Moscow, UNESCO Chair Holder, University of Tampere, Finland, UNESCO Chair, St Petersburg State University of Aerospace Instrumentation, Russian Federation, The Association of the University Libraries in Bulgaria (AUL), Municipality Cultural Centres, National Institute of Archaeology and Museum, Bulgarian Academy of Science, National Library “St. St. Cyril and Methodius”, National Museum of History in Bulgaria.

Digital Business Perspectives for the Publishers and Librarians, 26 September, 2011.

The panel session was organized within the European Day of the Entrepreneur, Sofia 2011. The possibilities provided by Share.TEC portal and ATLAS project to the publishers and librarians were discussed: http://ede.uni-sofia.bg/EDE2011/panel3_en.htm

The Future of e-Book and e-Libraries Workshop, 28 October 2010, Sofia, Bulgaria.

The participants were: researchers, e-book companies, Sofia University Library. City Library, National Centre for Information and Documentation, ASTEA Solution and StockPodium (technological companies).

ATLAS services to the potential customers and users were presented, getting feedback from the users about offered services was collected during the workshop: http://www-it.fmi.uni-sofia.bg/ede2010/programme_en.htm

Usability Seminar, 18-19 May 2011, Sofia

ITD demonstrated ATLAS products at Usability Seminar, organized in 18-19 May, 2011 in Sofia to 20 participants. The lecturers were Johann Schrammel and Regine Müller from USECON - The Usability Consultants GmbH. The training programme included the following topics: Usability and User Experience, Benefits of Usability, Usability Heuristics and User Interface Principles, Methods of Usability Engineering, Analysis, Innovation, Design, Prototyping, Evaluation, Trends and Special Topics in HCI, Exercises in Usability Reviews & Testing.

Pilot experiments within Living Lab

During project duration, pilot experiments were performed in the Living Lab. As a result, the following prototypes were developed and ideas were collected from the LL developers:

- **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>
http://i-publisher.atlasproject.eu/itd/unesco_chair/en
- **Icon-painters from Bundovtsi family website**. This website is made with i-Publisher and uses the technologies of the ATLAS platform.
http://www.atlasproject.eu/asset_demo/icons/bg/index.html
- A Living Lab activity: a response to an end-user request for showing similar **video objects for the Bulgarian National Television**: <http://videodemo.atlasproject.eu/>

8.5 Synergies with relevant projects and networks

Creating synergies between ATLAS project and other relevant EU projects was one of the major activities of the team for reaching project sustainability. The exchange of ideas and exploring possible ways of partnerships were performed through organisation of joint events and presentation of ATLAS services. Below the partnerships established with different projects are described.

FP7 Regions of Knowledge project: Smart Culture

Participation in the preparation of SmartCulture project proposal (already approved and financed) under FP7 Regions of Knowledge programme targeted at establishment of a European Network of Research Intensive Clusters and developing a Joint Action Plan in the

area of CCIs (ITD, Tetracom and DSLI are members of the Sofia Cluster). The ATLAS services can be of use for SmartCulture project implementation.

META-NET FP7 Network of Excellence: Multilingual Europe Technology Alliance,
<http://www.meta-net.eu/>

META-NET, a Network of Excellence consisting of 60 research centres from 34 countries is dedicated to building the technological foundations of a multilingual European information society. META-NET is forging META, the Multilingual Europe Technology Alliance.

- Participation in the poster session, distribution of ATLAS leaflets in META-FORUM 2012, 19-20 June 2012, Brussels, Belgium
- Presentation of the project leaflets, poster and online demo by ICS PAS and DCL – IBL in the dedicated ATLAS booth at META-NET Forum 2011, 27-28 June 2011, Budapest, Hungary

CESAR - Central and South-East European Resources, FP7 ICT

The CESAR project, in close harmony with META-NET, will make available a comprehensive set of language resources and tools covering the Bulgarian, Croatian, Hungarian, Polish, Serbian, and Slovak languages. Resources will include interoperable mono- and multilingual spoken and written databases, corpora, dictionaries and wordnets, as well as tools: tokenisers, lemmatisers, taggers, and parsers.

-
- Development of Multilingual dictionaries and the web service – TextMatch by Tetracom and
- Presentation of the ATLAS project made by TETRACOM: Widening the Scope. Harnessing NLP Techniques in Content Management Systems at CESAR META-NET Road show 2012, 2 May 2012, Sofia, Bulgaria. Presentation available on: http://dcl.bas.bg/en/cesar_roadshow_program_en.html
- ATLAS exhibition booth at Human Language Technology Days 2012, 27-28 September 2012, Warsaw, Poland. This is an Event related to CESAR ICT PSP project with the participation of linguists, computational linguists, IT experts.
- Resources prepared for ATLAS (parallel corpora, corpus of summaries) made available for CESAR to be included into META-SHARE open access facility.
- DCL – IBL - Synergies with CESAR project through:
 - Meta-Forum 2012: A Strategy for Multilingual Europe – June 20/21, 2012, Hotel Le Plaza, Brussels, Belgium
 - CESAR road-show – May 2, 2012, Hotel Sheraton, Sofia, Bulgaria
 - CESAR road-show – September 27, 2012, IPIPAN, Warsaw, Poland
 - CESAR road-show – June, 7, 2012, Bratislava, Slovakia
 - CESAR road-show – October 29, 2012, Belgrade, Serbia
 - CESAR road-show – November 30, 2012, Zagreb, Croatia
 - CESAR road-show – January 18, 2012, Budapest, Hungary
- ATLAS project was presented at CESAR project meeting, 26 June 2011, Budapest, Hungary by ICS PAS and DCL – IBL with the purpose to create synergies between ATLAS project by exchange of ideas and exploring possible ways of partnerships.

D-SPIN - a German infrastructure for Language Resources and Tools:

<http://weblicht.sfs.uni-tuebingen.de/englisch/index.shtml>

The project D-SPIN provides the basis for a stable and sustainable infrastructure of language resources and language technologies, serving above all empirical research in humanities and social sciences.

CLARIN project: <http://www.clarin.eu/>

The CLARIN project is a large-scale pan-European collaborative effort to create, coordinate and make language resources and technology available and readily useable for the whole European Humanities (and Social Sciences) community.

- Exchange of ideas and distribution of project leaflets by ICS-PAS at D-Spin project meeting, 29 June 2010, Giessen, Germany
- CLARIN/D -SPIN workshop on linguistic Web services, 16-17 November 2010, Freudenstadt, Germany

PROMIS Lingua project: <http://www.promislingua.eu/>

The overall aim of the PROMISLingua Pilot project is the translation, localisation and implementation of the existing PROMIS® online service (at the moment available in English, German and Italian) in additional six languages and markets (Spanish, French, Portuguese, Greek, Romanian, Hungarian). The project aims at shortening the time-to-market of PROMIS® and delivering a cost-efficient and easy-to-use Internet based service for Safety, Health, Environment and Quality (SHE-Q) management via multilingual translation and innovative support services to SMEs.

- ATLAS project was presented by ATLANTIS at a Kick-off meeting of the PROMIS Lingua international project in 13-15 April 2011, Luxembourg among business consultants, linguistic, and IT experts, as well as a European SME umbrella organisation.

EUROPE Innova KIS, <http://www.europe-innova.eu/>

The KIS-IP is a European initiative funded under Europe INNOVA, with the aim to accelerate the take-up of services innovations in Europe. The initiative focuses on the development and testing of new or better innovation support mechanisms for innovative small and medium sized enterprises (SMEs), in particular in technological and industrial fields. The KIS-IP brings together public and private partners from different countries willing to cooperate in developing new forms of support for innovation, taking into account the specific needs of service companies as well as the potential role of services innovation in support of societal needs.

- Atlantis made in initial presentation of the project and distribution of EN flyer at KIS Partnering Forum 2010, 15-16 March 2010, Rome, Italy and at Europe Innova KIS Partnering Forum, 27-28 January 2011, Warsaw, Poland among IT experts, stakeholders and experts from the financing, business consulting.
- ATLANTIS made a presentation about the project and several bilateral meetings were held at the Joint event of the Europe INNOVA Annual Partnering Event and EPISIS INNO-Net, 14-16 June 2010 Copenhagen, Denmark

OpenScout: Skill based scouting of open user-generated and community-improved content for management education and training

<http://learn.openscout.net/>

The project aims at providing an education service in the internet that enables users to easily find, access, use and exchange open content for management education and training.

ShareTEC: SHARing digital REsources in the Teaching Education Community

<http://portal.share-tec.eu/>

Share.TEC is devoted to fostering a stronger digital culture in the TE field and to supporting the development of a Europe-wide perspective among those working in and with the TE community. To do this, Share. TEC is developing an online platform which will help practitioners across Europe search for, learn about and exchange resources of various kinds, and will support the sharing of experience about the use of those resources.

SmartBook, Bulgarian national project

<http://dse.fmi.uni-sofia.bg/SmartBook/index.htm>

This project aims to further develop technologies that will facilitate the creation and use of a new generation of 'smart' books: e-books those are evolving, highly interactive, customisable, adaptable, intelligent, and furnished with a rich set of author and reader support services.

- ITD organized Workshop: Digital Business Perspectives for the Publishers and Librarians on 26 September 2011 in Sofia to discuss the possibilities provided by Share.TEC portal and ATLAS project to the publishers and librarians. The ATLAS project and i-Publisher, i-Librarian are presented and demonstrated among representatives from the libraries, universities, publishing houses and wide public. ITD and ATLAS project are co-organizers of the EDE 2011
- ITD organized Workshop: The Future of e-Book and e-Libraries Workshop, 28 October 2010, Sofia, Bulgaria. ATLAS, Share.TEC, Smartbook and Open Scout projects were presented. The participants were: researchers, e-book companies, Sofia University Library. City Library, National Centre for Information and Documentation, ASTEA Solution and StockPodium (technological companies).

OpenAIRE: Open Access Infrastructure for Research in Europe.

OpenAIRE establishes underlying structures for researchers to support them through a European Helpdesk System. EuDML (The European Digital Mathematics Library) which brings together the dispersed European heritage of digital mathematical literature in virtual collections.

- Presentation of ATLAS project by ITD at National Information Day - Open Access to Scientific Information, 22 October 2010, Sofia Bulgaria among researchers, representatives of Sofia City Library, Varna Regional Library "Pencho Slaveykov" and technological companies

MUMIA Network - Multilingual and Multifaceted Interactive Information Access (funded by EU COST Action IC1002), <http://www.mumia-network.eu/>

The main mission of MUMIA is to launch and establish a cooperation framework of researchers working in fields related to Multilingual and Multifaceted Interactive Information Access (MUMIA) as well as patent professionals and practitioners to coordinate the collaboration between them, foster research and technology transfer in the broad areas of Machine Translation, Information Retrieval and Multifaceted Interactive Information Access, aiming to facilitate the development of next generation search systems.

- Presentation of ATLAS project and exchange of ideas by ITD at MUMIA Network meeting, 22 September 2011, Amsterdam, Netherlands

DYLAN: Language dynamics and management of diversity, FP6 IP Project No: 028702, <http://www.dylan-project.org/>

The project aimed to provide scientific backing to the concept of multilingual repertoires as resources that can be put to use in a variety of professional, political and educational contexts.

- Presentation of ATLAS project and exchange of ideas by ITD at Dylan Final Conference, 22 September 2011. Participants: EC policy makers, Universities, Companies and NGOs in the field of Language Dynamics and Management of Diversity

ELLIOT: Experiential Living Lab for the Internet of Things, <http://www.elliott-project.eu/>

The project aims to develop an Internet Of Things (IOT) experiential platform where users/citizens are directly involved in co-creating, exploring and experimenting new ideas, concepts and technological artefacts related to IOT applications and services. ELLIOT will allow studying the potential impact of IOT and the Future Internet in the context of the Open User Centred Innovation paradigm and of the Living Lab approach.

- Presentation of MLeCel Living Lab and ATLAS by ITD at ELLIOT Summer School "Open Innovation & User eXperience Design & Assessment" - 12th Sep. 2012, Milano, Italy. Presentation: <http://livinglab.itd-bg.eu/sites/default/files/ELLIOT-Virtech-Milano-Sep12-2012.pdf>

9. Project Impact Measurement

9.1 Technological and social impact

The project brings together advanced technologies for multilingual web content management and text mining (such as automated 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, facilitates the spread of the project achievements.

The tables below presents the ATLAS achievements contributing to the technological and social impact described in the DoW.

Table 2 Technological Impact

Planned	Achievements	Extent
Integration of text mining tools into content management systems	The ATLAS framework employs technologically and linguistically diverse natural language processing (NLP) tools in a platform, based on UIMA2. The UIMA pluggable component architecture and software framework are designed to analyse content and to structure it. The ATLAS core annotation schema, as a uniform representation model, normalizes and harmonizes the heterogeneous nature of the NLP tools. The language processing tools are integrated in a language processing chain (LPC), so the output of a given NLP tools is used as an input for the next tool in the chain.	100%
Integration of text mining services	The ATLAS platform integrates automatic: categorization; summarisation; text annotation; extraction of important phrases, words and names and translates documents regardless of their language and format. The ATLAS platform is designed with extensibility in mind, which allows for easy addition of tools and services for currently unsupported languages as well as new tools for already supported languages.	100%
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	Machine Translation: <ul style="list-style-type: none"> Working version fo the Example-based MT engine Training of SMT models for : <ul style="list-style-type: none"> SL <--> English SL < ---> German based on JRC-Acquis Communautaire, EU-constitution and Europarl. Setting up a framework for training models for 	100%

<p>language pairs have strong relevance for the Central- and East-European commercial space.</p>	<p>domain adaptation that implied:</p> <ul style="list-style-type: none"> - setting -up the methodology, including a strategy for low resourced language pairs - definition of domains for which models will be trained - collection of training - data (on-going) - proof-of-concept prototype for financial domain. <p>Crosslingual retrieval:</p> <ul style="list-style-type: none"> • Standardization of the interface between the ATLAS-platform and the crosslingual search engine. • Implementation of a first prototype of the crosslingual search engine. 	
<p>Contribution to the development of text processing chains for languages, which lack resources at present</p>	<ul style="list-style-type: none"> • New, previously non-existing implementations of important tools have been created (e.g. Spejd-based multiword noun phrase lemmatizer for Polish). • Improved versions of resources have been created (e.g. Nerf – Polish named entity recognizer). • Better, more efficient implementation of existing tools (e.g. Pantera – a Brill tagger for Polish or Spejd – a partial parser, used for NP extraction). • Server versions of tools have been prepared to improve overall efficiency by pre-loading of language models (e.g. grammar rules pre-compiled Spejd for Polish) or enabling concurrent execution (e.g. simultaneous processing of input file split into parts for Pantera and Spejd). 	100%
<p>Adherence to and promotion of existing and future web standards</p>	<p>Language Processing Chains have been successfully used to promote the UIMA framework.</p>	100%
<p>Practical and economically viable solutions for nearly-automatic provision of multilingual online content and services for some EU languages</p>	<p>For Publishing houses; Media and PR agencies:</p> <ul style="list-style-type: none"> - News Service Demo, http://newsdemo.atlasproject.eu/ - Video materials demo version, http://videodemo.atlasproject.eu/ - Press clipping service in Bulgarian and English, http://mediaimage.tetracom.com/ - Plagiarism service – TextMatch, http://textmatch.eu/ 	100%

Table 3 Social Impact

Planned	Achievements	Extent
Facilitate exchange of information and knowledge	<p>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.</p> <p>While metadata identifies formal characteristics of a document like author, date and format, the automatic annotations represent its content. ATLAS processes multilingual data pools, analyses and streamlines them with semantic annotations. As a result the reader gains access to a layer of information which, though important, lies hidden from view:</p> <ul style="list-style-type: none"> • the reader spots information-dense entities like personal names, geographical locations, organisations, common phrases, short summary even before reading the text • the reader can search through these entities and reach all relevant content • the reader gains access to similar documents, pages and information recommended automatically by ATLAS instead of editors. <p>A new categorisation tree, containing 80 categories grouped in 6 top-level groups, enhances the users in cataloging their documents automatically. Categorisation models for all project languages have been developed.</p> <ul style="list-style-type: none"> • A new categorisation tool provides the users with the functionality to split their documents into groups. • The multiple file upload enhances the users while uploading documents and as a result the user can upload all their documents into i-librarian at one go. • The “public” library contains a collection of more than 22'000 books in English from Project Gutenberg. The books are processed, analysed and as a result enriched with revealing details like the most commonly used phrases, name entities, suggestion for similar books, short summary and are categorized automatically, according the library catalogue. 	100%

Simplify authoring, management and exploitation of heterogeneous multilingual content	I-Publisher Simple Mode. The complexity of i-Publisher, a powerful instrument for creating complex web sites, made it difficult for the group of inexperienced test users to work with it. As it is a wider and an important target group, The Consortium decided to extend the web based tool i-Publisher with simplified layer for non experienced users. The users choose from ready-to-be-used websites or they create websites with predefined visualisation (themes) still having the possibility to define the content and its structure. As none of the existing now online services, some of the ready-to use websites provide the functionalities related with the linguistic framework so that the user can benefit from its applications like automatic annotations, automatic translation. The interface of the Simple Mode is localised in all project languages.	100%
Address the needs of a large number of people belonging to different target user groups – individuals and organizations	<p>For all individuals (librarians, researchers, IT experts, students, publishers, etc.) and organisations (NGOs, Libraries; Publishing houses; Media and PR agencies; International organisations; Corporate clients):</p> <ul style="list-style-type: none"> - i-Publisher Live demo: http://i-publisher.atlasproject.eu/atlas/i-publisher/demo - i-Librarian website, http://www.i-librarian.eu/ - EUDocLib website, http://eudoclib.atlasproject.eu/ <p>For Publishing houses; Media and PR agencies:</p> <ul style="list-style-type: none"> - News Service Demo, http://newsdemo.atlasproject.eu/ - Video materials demo version, http://videodemo.atlasproject.eu/ - Press clipping service in Bulgarian and English, http://mediaimage.tetra.com/ - Plagiarism service – TextMatch, http://textmatch.eu/ <p>For NGOs and international organizations:</p> <ul style="list-style-type: none"> - UNESCO Website Demo, http://unesco.atlasproject.eu/ - UNESCO Chair Website Demo, http://a2.atlasproject.eu:9090/unesco - World Health Organization Website Demo, http://asset.atlasproject.eu/who - Icon-painters from Bundovtzi family website, http://www.atlasproject.eu/asset_demo/icons/bg/index.html 	100%
Cross the language barrier Facilitate culture exchange	The linguistic platform ASSET employs Natural Language Processing tools, a categorisation tool, a summarisation tool and a machine translation engine. The produced services uses ASSET which facilitate the	100%

	<p>cross of the language barrier:</p> <p>TextMatch recognizes the language of the document using a two-stage language detection system. Specific language tokenizers, lemmatizers and other analyzers are utilized for <i>English, Bulgarian, German, French and Russian</i>. For languages that are currently not supported, a language independent comparison algorithm is used.</p> <p>i-Librarian offers the intelligent way to browse, categorize and analyze your electronic books and documents in <i>English, Bulgarian, German, Greek, Polish and Romanian</i>.</p> <p>News Demo service categorizes and analyzes the news in English, Bulgarian and Greek.</p> <p>All services using ASSET could be extended applying other languages.</p>	
<p>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.</p>	<p>Liaison with the Europeana and EuroMatrix Plus was established in order to foster language diversity in content creation and distribution.</p> <p>Tools and models developed within EuroMatrix are available by using related input data specification.</p> <p>The input between the MT-Serverland and ATLAS is established and integrated into the ATLAS-MT engine.</p> <p>A bridge to Euromatrix is integrated in the workflow of the ATLAS-engine.</p>	

9.2 User Interaction in the Living Lab

Effective interaction and networking with users and ATLAS communities was performed through the following MLeCel Living lab activities:

- Identification of prospective users and user groups, user needs analysis;
- Gathering user requirements related to i-Librarian and i-Publisher in a production setting and in a real-life environment;
- Ongoing user evaluation of the ATLAS services;
- Conduct experiments with i-Publisher in the MLeCeL LL for building a multilingual web site for SULSIT and for integration of the linguistic platform and the classification tool in the Electronic Library of Sofia University;
- Test of the i-Librarian service and the advanced functionalities of the i-Publisher service in order to evaluate the flexibility of the i-Publisher for creating and customizing a rich content driven web site;

- Conducting workshops, expert meetings, educational and training activities with local partners (developers, universities, SMEs, NGOs) and co-operating with other Living Labs.

The table below presents the indicators, measuring the progress the Living Lab activities and achievements within the Living Lab towards networking and interaction with target communities:

Table 4 Living Lab activities contribution

Indicators	Planned	Achieved	Comments
User interviews	> 20	112	<p>15 (Dec 2011) Report including users' feedback from Dec. 2011 workshops is available here: http://livinglab.itd-bg.eu/sites/default/files/Results_Dec2011.pdf</p> <p>17 (Febr. – March 2012) Results are at: http://livinglab.itd-bg.eu/sites/default/files/Results_March%202012.pdf</p> <p>80 users tested (30 provided general feedback) (Nov. – Dec. 2012), Users feedback: http://livinglab.itd-bg.eu/content/3rd-atlas-testing-round-living-lab-experts-feedback</p>
User suggestions	> 30	38	Collected from all the events described
Feedback from idea generation workshops with lead-users	3	3	<p>Feedback from joint idea generation workshops with lead-users, organized by ITD and TetraCom were presented in relation with Two experiments.</p> <p>During joint workshop on 03/05 in the Living Lab, ideas will be collected also from potential users and experts.</p>
Feedback from idea generation workshops with experts	2	4	<p><i>UNESCO Interfaculty Chair website:</i> 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</p> <p><i>Integration with DSpace:</i> Feedback from developer from Sofia University: http://livinglab.itd-bg.eu/sites/default/files/Scenarios%20for%20Integration%20with%20Dspace.pdf http://livinglab.itd-bg.eu/</p> <p><i>Demo version for video materials:</i> A Living Lab activity: a response to an end-user request for showing similar video objects for the Bulgarian National Television:</p>

			http://videodemo.atlasproject.eu/ <i>Icon-painters from Bundovtsi family</i>
Online suggestion box	ongoing	Available	To collect suggestions from the casual users in the database
Feedback from workshops with users and prospective users	4	8	Description of Workshops is on: http://livinglab.itd-bg.eu/content/atlas-test-workshops-mlecel-II Feedback is collected through questionnaires, available on the website.
Feedback from prototype testing	2	6	Feedback was collected by questionnaires during prototype testing on evaluation workshops: http://livinglab.itd-bg.eu/content/atlas-test-workshops-mlecel-II Feedback of 3 experts from 3 rd round testing: http://livinglab.itd-bg.eu/content/3rd-atlas-testing-round-living-lab-experts-feedback
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	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	2	3	Received feedback from education and training sessions at university level, published at: http://livinglab.itd-bg.eu/content/atlas-test-workshops-mlecel-II in SULSIT and SU
Feedback from training workshops with developers	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	2	2	Received feedback from education and training sessions at university level, published at: http://livinglab.itd-bg.eu/content/atlas-test-workshops-mlecel-II

Catalyzing research activities based on ATLAS	at least 1 PhD study and 10 MSc studies	15	All studies related to different research problems related to ATLAS project. 3 PhD students and 12 MSc students
---	---	----	--

9.3 Impact of dissemination activities

Contribution of the dissemination activities performed by the ATLAS partners to the expected impacts by knowledge dissemination across organizations, target users and communities is presented in the following table:

Table 5 Dissemination activities

Dissemination activities	Achievements
Production of dissemination materials	1 general ATLAS brochure 2 marketing brochures for i-Publisher and i-Librarian Partners` brochures and leaflets 10 other publications to the general public Dissemination kit available
Doing research in the field	31 articles and papers, published in scientific and technical journals and conference proceedings to the Research Community
Development of demonstration cases for potential customers	13 demonstration websites - the main tool in the ATLAS exploitation strategy for targeting customers
Organising events to the Research Community	3 organised large events for Research community 10 posters, introduced at conferences and exhibitions for better presentation of the developed services, used technologies, present and further benefits, and widening the scope of the potential users.
Organising events to the Developers and Users Community /Working with specific target groups through organised user workshops	8 evaluation workshops with 112 users carried out in the Living Lab Collected feedback from users and experts, available on the LL website 38 user suggestions, collected though the LL website Feedback of 3 experts from 3rd round testing 7 organised ATLAS events within Living Lab targeting users and customers 4 pilot experiments carried within the Living Lab
Organising events to the potential customers	8 meetings with potential customers in Bulgaria and abroad
Giving presentations to other events	More than 62 presentation and demonstration made by partners at international events

Sharing experience within the communities	Online video tutorials, guides and technical documentation about i-Publisher and i-Librarian about 100 performed dissemination activities (18 exhibitions, 40 conferences and 40 workshops, seminars, meetings and lectures...)
Communicating within the network	ATLAS project website and Living Lab website for interaction with target users
Passing on the project information in the partners` institutions though seminars and lectures	13 organised seminars within partners` institutions with students and researchers
Exchanging idea and conducting discussions and interaction with other projects and networks in the field	Synergy with 10 relevant projects (META-NET, Cesar, D-SPIN, CLARIN, PROMIS Lingua, OpenScout , ShareTEC, SmartBook , DYLAN, ELLIOT) and 3 networks (MUMIA, EUROPE Innova KIS and OpenAIRE) by organisation of joint events, presentations, demonstrations, exchange of ideas and discussion of future exploitation activities
Generating new projects	1 new project SmartCulture (approved)
Communicated ATLAS results with policy makers	2 presentations at UNESCO Chairs and Bulgarian policy makers: http://livinglab.itd-bg.eu/content/events

Annex 1: Dissemination activities within ATLAS duration

Performed dissemination activities for the period: 01.03.2012-28.02.2013 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
Springer article	Summarizing Short Texts Through a Discourse-Centered Approach in a Multilingual Context	2013	UAIC, ICS PAS, UHH, DCL-IBL, ATLANTIS, TETRACOM	Springer article	Daniel Anechitei, Dan Cristea, Ioannidis Dimosthenis, Eugen Ignat, Diman Karagiozov, Svetla Koeva, Mateusz Kopeć, Cristina Vertan (2013, to appear). Summarizing Short Texts Through a Discourse-Centered Approach in a Multilingual Context. In Neustein, A., Markowitz, J.A. (eds.), Where Humans Meet Machines: Innovative Solutions to Knotty Natural Language Problems. Springer Verlag, Heidelberg/New York.	Linguists, computational linguists, IT experts
Peter Lang article	i-Publisher, i-Librarian and EUDocLib - linguistic services for the Web	December 2012	TETRACOM, ICS PAS,UHH, DCL – IBL, ATLANTIS UAIC, ITD	Article on the linguistic services in ATLAS	i-Publisher, i-Librarian and EUDocLib - linguistic services for the Web by Anelia Belogay, Damir Ćavar, Dan Cristea, Diman Karagiozov, Svetla Koeva, Roumen Nikolov, Maciej Ogrodniczuk, Adam Przepiórkowski, Polivos Raxis and Cristina Vertan. P. Pęzik (ed.): Corpus Data across Languages and Disciplines, Lodz Studies in Language Series, vol. 28, pp. 203-212, Peter Lang. ISBN 978-3-631-62547-7.	Linguists, representatives of social sciences and humanities
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	international ICT industry No of people reached:

Deliverable 8.3

					Asset: http://www.cebit.de/product/asset-adds-value-to-your-content/291715/C913812	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 Karagoyozov, Svetla Koeva, Cristina Vertan, Adam Przepiórkowski, Dan Cristea, Plovios 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 and article	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	Article at LREC 2012 Workshop and poster presentation	Anelia Belogay Diman Karagoyozov 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	Infotheca Open Access journal	June, 2012	TETRACOM ,ICS	Journal Article	i-Librarian – Free online library for	Linguists,

Deliverable 8.3

Journal		Belgrade, Serbia	PAS, UHH, DCL – IBL, ATLANTIS, UAIC		European citizens by Diman Karagiozov, Anelia Belogay, Dan Cristea, Maciej Ogrodniczuk, Polivos 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/ .	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+
Conference	8th International Conference Formal Approaches to South-Slavic and Balkan Languages	September 19-21, 2012, Dubrovnik, Croatia.	DCL – IBL	Article	Koeva, Svetla, Ivelina Stoyanova. Toward automatic generation of Parallel corpora. - In Proceedings from the 8th International Conference Formal Approaches to South-Slavic and Balkan Languages. September 19-21, 2012, Dubrovnik, Croatia.	Linguists, computational linguists
Workshop	Joint Workshop on Exploiting	23-27 April, 2012	DCL – IBL	Article	Koeva, Svetla. ATLAS - Human	Linguists,

Deliverable 8.3

	Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra),	Avignon, France			Language Technologies integrated within a Multilingual Web Content Management System. – In Proceedings of the Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra), Avignon: Association for Computational Linguistics, 2012, pp. 72-76. ISBN: 978-1-937284-19-0 . http://dl.acm.org/citation.cfm?id=2387966&CFID=250398417&CFTOKEN=38583902	computational linguists
Forum	EUREKA Venture Forum	18-19 October, Istanbul, Turkey	ATLANTIS, Thessaloniki, Greece	Presentation and discussions for commercial exploitation of ATLAS	ATLANTIS, Athens, Greece	30-40 European entrepreneurs, 40-50 international investors
Organised exploitation meeting	Business exploitation events with representatives from the Hellenic Mobile Cluster, Athens Technical University, Greek Public Properties Company (ETAD), big private corporate groups	19-21 November 2012, Athens, Greece	ATLANTIS	Presentation of i-Librarian and i-Publisher	Project brochure, project presentation, live presentation and demonstration of the 2 ATLAS services in particular business cases	HAMAC Cluster, Theta SA, NTUA, Newsphone group, UPstram SA, ETAD
Pan-European Forum	2 nd 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
Seminar	NLP Seminar, University of Warsaw	3 January 2012 Warsaw, Poland	ICS PAS, Poland	Presentation of WP3-related task	Latent Dirichlet Allocation – algorithm realisation by Kacper	Students 15-20 participants

Deliverable 8.3

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 <i>No of people reached: 200</i>
Publication	ATLAS @ Biznes raport – a publication about innovative projects added to Polish daily newspaper Dziennik – Gazeta Prawna	31 January 2012, Warsaw, Poland	ICS PAS, Poland	Project information	A paragraph on ATLAS in a broader text "UE supports processing of Polish" (in Polish).	General public
Seminar	NLP Seminar, University of Warsaw	20 March 2012, Warsaw, Poland	ICS PAS, Poland	Presentation of the WP3 topic	Topic extraction by Kacper Chwiałkowski.	IT students 15 participants
Seminar	NLP Seminar, University of Warsaw	27 March 2012, Warsaw, Poland	ICS PAS, Poland	Presentation of the WP3 topic	Latent Dirichlet Allocation in text categorization by Kacper Chwiałkowski.	Students 15-20 participants
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	Ogrodniczuk, Maciej and Przepiórkowski, Adam. Polish Language Processing Chains for Multilingual Information Systems. G. Bouma et al. (ed.): NLDB 2012, LNCS 7337, pp. 152-157. Springer, Heidelberg.	Scientific and business participants <i>No of people reached: 100</i>
Project village	Human Language Technology Days 2012	27-28 September 2012, Warsaw, Poland	ICS PAS, Poland	Event related to CESAR ICT PSP project	ATLAS exhibition booth.	Linguists, computational linguists, IT experts 150-200 participants
Official Polish record of ICT projects	Polish Potential and Contribution to EU Success in ICT	October 2012, Poland	ICS PAS, Poland	Open publication	ATLAS project description, objectives, list of participants; pp. 200-201. See http://www.kpk.gov.pl/pliki/12357/pc- net.pdf	General public
Synergy	Synergy with CESAR	November 2012, Poland	ICS PAS, Poland	Dissemination of resources	Resources prepared for ATLAS (parallel corpora, corpus of summaries) made available for CESAR to be included into META-SHARE open access facility.	General public

Deliverable 8.3

Conference	The 8th Conference on Resources and Tools for Processing Romanian Language - ConSLR-2012	27 April, 2012 Bucharest, Romania	UAIC, Romania	ATLAS presentation, showcase of the summarization chain	Dan Cristea, Eugen Ignat și Daniel Anechitei: Proiectul ATLAS (in Romanian)	Linguists, computational linguists, IT experts No of people reached: 50
Workshop	Language Technologies in Romanian and Diaspora Research & Development, Bucharest	26-27 September, 2012, Bucharest	UAIC, Romania	Article ATLAS presentation, showcase of the summarization chain	Ignat, E. and Anechitei, D.: "Multilingual summarisation system based on analysing the discourse structure" in Language Technologies in Romanian and Diaspora Research & Development. Edited by Forăscu, C., Ioniță, A., Tufiş, D., Cristea, D. ISBN 978-973-703-813-5, Pp 34-35, Bucharest, 26-27 September, 2012.	Computational linguists, NLP engineers No of people reached: 70
International Exhibition	International Exhibition of Software Bucharest The Binary Conferences are a series of events that aim to present specific IT&C solutions for the SMB sector.	7-8 November 2012, Romania	UAIC, Romania	Demonstration	Demonstration http://binary.aries.ro/en/	General public, engineers, representatives of social sciences No of people reached: 100
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 Activity for generating ideas	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	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,	Computational linguists, industry people, mass media No of people reached:

Deliverable 8.3

research and industry					"Alexandru Ion Cuza" University Editing House, ISSN 2285-0929, pages 26-27.	80-100
Demonstration	Exposing the services included in ATLAS on the NLP-Group@UAIC-FII website	2011-2012	UAIC, Romania	Demonstration	http://nlptools.info.uaic.ro/WebClauseSplitterRo/ http://nlptools.infoiasi.ro/WebDiscourseParser/	NLP researchers
Article in Journal	Open Access Journal Translation: Computation, Corpora, Cognition (http://www.t-c3.org)	February 2013	UHH, Germany	Journal Article	Cristina Vertan, Melania Duma, Integration of Machine Translation in On-line Multilingual Applications – Domain Adaptation	Computational Linguists
Seminar	University of Winterthur, Winterthur, Switzerland	February 2013	UHH, Germany	Powerpoint Presentation and demonstration	Cristina Vertan, Walther v. Hahn, Embedding Language Technology Tools in Web Applications - Focus on Machine Translation and Domain Adaptation	Professional Translators
Conference	EACL 2012	23-27 April, 2012 Avignon, France	UHH, Germany	Paper	Cristina Vertan, "Two approaches for integrating translation and retrieval in real applications"	Linguists, computational linguists, IT experts
Conference	LREC 2012	21-27 May 2012 Istanbul, Turkey	UHH, Germany	Paper and workshop organization	Cristina Vertan, Walther v. Hahn, Monica Gavrila, Same domain different discourse style-A case study on Language Resources for data-driven Machine Translation	Computational 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	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

Deliverable 8.3

Conference	Diaspora in Cercetarea Stiintifica si Invatamantul Superior din Romania	26-27 September 2012, Bucharest, Romania	UHH, Germany	Powerpoint Presentation and Abstract in Workshop Booklet	Cristina Vertan, Integrarea unui sistem de traducere automată în platforme multilinguale exemplificată în cazul perechii de limbi Română-Germană	Linguists, computational linguists, IT experts
Book	Cambridge Scholars Press,	June 2012	UHH, Germany	Edited Book with collection of articles including contributions from ATLAS partners	Cristina Vertan, Walther v. Hahn Multilingual Processing in Eastern and Southern EU Languages: Low-Resourced Technologies and Translation	NLP Researchers, linguists
Invited Talk & demonstration	University of Buenos Aires, Argentina	17 November 2012, Argentina	UHH, Germany	Powerpoint Presentation and demonstration	Cristina Vertan, Walther v. Hahn, Embedding Language Technology Tools in Web Applications - Challenges, current limitations, and further perspectives	NLP Researchers and Students
Invited Talk & demonstration	University of Cordoba, Argentina,	27 November 2012, Argentina	UHH, Germany	Powerpoint Presentation and demonstration	Cristina Vertan, Walther v. Hahn, Embedding Language Technology Tools in Web Applications - Challenges, current limitations, and further perspectives	NLP Researchers and Students
Seminar & Presentation	University of Santiago, Santiago de Chile, Chile	29 November 2012, Chile	UHH, Germany	Powerpoint Presentation and demonstration	Cristina Vertan, Walther v. Hahn, Embedding Language Technology Tools in Web Applications - Challenges, current limitations, and further perspectives	NLP Researchers and Students
Organised Workshop	Workshop on Multilingual Digital Repositories and Services	3 May 2012, Sofia	ITDF	Programme, presentations, demonstrations, pictures	Programme, presentations, demonstrations, pictures: http://livinglab.itd-bg.eu/content/events	Librarians, researchers, potential users, TV, media, DSLL
Workshops	ELLIOT Summer School "Open	12 September	ITDF	Presentation of MLeCel Living	Presentation of MLeCel Living Lab:	Researchers and

Deliverable 8.3

	Innovation & User eXperience Design & Assessment"	2012, Milano, Italy.		Lab.	http://livinglab.itd-bg.eu/content/events	business
Organised Workshops	Workshop on the Process of creation and development in Living Labs	19 September 2012, Veliko Tarnovo.	ITDF	The workshop presents the Living Labs idea, perspectives, and few Living Labs. The workshop is organised within Digital presentation and preservation of cultural and scientific heritage Conference.	http://livinglab.itd-bg.eu/content/events	Librarians, researchers, potential users
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 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
Organised Workshop	Platform ATLAS: Demonstration	16 October 2012, Sofia	ITDF	Programme, presentations, demonstrations	Programme and presentation: http://livinglab.itd-bg.eu/content/events	Librarians, researchers, users

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
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/gsc12011/?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	29 September	TETRACOM, ICS	Poster at the main	ATLAS — A Robust Multilingual	Computational

Deliverable 8.3

	of the German Society for Computational Linguistics and Language Technology (GSCL 2011)	2011, Hamburg, Germany	PAS, UHH, DCL – IBL	conference	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	linguists No of people reached: 50
Demo session	ATLAS @ Demo session at the 5th Language & Technology Conference (LTC 2011): Human Language Technologies as a Challenge for Computer Science and Linguistics	25-27 November 2011, Poznan, Poland	TETRACOM, ICS PAS, UHH, DCL – IBL, ATLANTIS, UAIC	Article in the proceedings, software and resource demonstration, information about the project, roll-up exhibition.	Language Processing Chains in ATLAS by Anelia Belogay, Dan Cristea, Eugen Ignat, Diman Karagiozov, Svetla Koeva, Maciej Ogrodniczuk, Adam Przepiorkowski, Polivios Raxis and Cristina Vertan: demonstrating i-Publisher, i-Librarian, EUDocLib and PLDocLib	IT scientists, computational linguists No of people reached: 100
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/1bf6f91-081c-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 <i>No of people reached: available in the Internet</i>
Conference	8th International Conference Practical Applications in Language and Computers (PALC 2011)	13-15 April 2011, Łódź, Poland	ICS PAS, Poland	Conference article Distribution of project leaflets More information: http://palc.ia.uni.lodz.pl	Maciej Ogrodniczuk (2011): "i-Publisher, i-Librarian and EUDocLib - Linguistic services for the Web", PALC 2011	Linguists, representatives of social sciences and humanities <i>No of people reached: 50</i>
Seminar	Linguistic tasks in ATLAS project (with particular focus on lemmatization of	18 April 2011 Warsaw, Poland	ICS PAS, Poland	ATLAS presentation, showcase of the multiword	A presentation by Adam Przepiorkowski, Maciej Ogrodniczuk	Computational linguists, NLP

Deliverable 8.3

	multiword units) @ Natural Language Processing Seminar			prototype	and Łukasz Degórski	engineers No of people reached: 30
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 Distribution of project leaflets	Degórski, Łukasz. Towards the Lemmatisation of Polish Nominal Syntactic Groups Using a Shallow Grammar. P. Bouvry et al. (ed.): SIIS 2011, LNCS 7053, pp. 370-378. Springer, Heidelberg. More information: http://iis.ipipan.waw.pl	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

Deliverable 8.3

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 @ INFISO.E1 consultation meeting	6-7 December 2011, Paris, France	ICS PAS	Distribution of new i-Publisher leaflets.	The INFISO.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	Language Technologies for Digital Humanities and Cultural Heritage Workshop associated with the RANLP 2011 Conference	C. Vertan, M. Gavrilă (2011): Using manual and parallel aligned corpora for Machine Translation Services within an Online Content Management System, Workshop on "Annotation and Exploitation of Parallel Corpora", in conjunction with RANLP 2011.	IT scientists, linguists No of people reached: 100
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

Deliverable 8.3

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, Universities, 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 from 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

Deliverable 8.3

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	Distribution of project brochures	Developers, users Living Labs representatives and

Deliverable 8.3

				APOLLON and IBBT, a workshop dedicated to issues of IPR for Living Labs		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. <i>No of people reached: 30</i>
European Congress	11th Congress of the European Business Angels Network (EBAN). <i>Theme: "Syndication and co-investment: partnerships and vision for the future"</i>	12-13 May 2011, Warsaw, Poland	Atlantis, Greece	Discussions on funding possibilities for ATLAS takup. Distribution of EN flyer.	Flyer and brochure	Business Angels from all over Europe. <i>No of people reached: 100</i>
Business development meeting	Corporate business day (Atlantis group of companies) with the participation of managers from all over Greece and Cyprus	10 February 2011 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 2011, Hamburg	Atlantis, Greece	Paper presentation on 26 September 2011	Paper titled "Towards the integration of monolingual NLP frameworks for multilingual applications: A case study for the Greek Language within the ATLAS project framework".	Computational Linguistics and Language Technology providers and researchers
Performed dissemination activities for the period: 01.03.2010-28.02.2011 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
Partnering forum	KIS Partnering Forum 2010	15-16 March 2010	Atlantis, Greece	Initial presentation of the project. Distribution of EN flyer.	Flyer	Business experts, IT experts <i>No of people reached: 60</i>

Deliverable 8.3

		Rome, Italy				
Partnering event	Joint event of the Europe INNOVA Annual Partnering Event and EPISIS INNO-Net	14-16 June 2010 Copenhagen, Denmark	Atlantis, Greece	80 EN flyers were distributed. ATLANTIS made a presentation about the project and several bilateral meetings were held.	Flyers and PPT New contact information	Business experts, IT experts <i>No of people reached:190</i>
International business and investment event	MOBIP 2010 and Investment in Mobile and IT Services	10-16 June 2010 Valencia, Spain	Atlantis, Greece	Distributed project material and exchange ideas about exploitation prospects.	New contacts	Stakeholders and experts from the financing, business consulting, IT, etc. sectors <i>No of people reached:128</i>
Partnering forum	INNOCAP - Experiences of Greece in Innovation Management	24 November 2010, Mersin, Turkey	Atlantis, Greece	Distribution of 60 EN flyers.	Flyers, New contact information	Stakeholders and experts from the financing, business consulting. <i>No of people reached:60</i>
Conference	Corallia Conference for the mi-clusters.job	20 December 2010, Athens, Greece	Atlantis, Greece	ATLANTIS distributed project 20 flyers in English. Discussions took place about the possibility to use ATLAS for online publishing of documents from microelectronics projects of the Greek mi-cluster.	Flyers	Business experts, IT experts <i>No of people reached: 15-20</i>
Conference	Conference on “measures to support enterprises in implementing investments in digital solutions”.	17 January 2011, Drama, Greece	Atlantis, Greece	Short presentation of the project, along with the distribution of 60 EN flyers. Investigation of the financial environment that the support program Digital Convergence offers in relation to ATLAS.	Flyers and PPT	Stakeholders and experts from the financing, business consulting, IT experts, etc. <i>No of people reached: 60</i>
Partnering forum	Europe Innova KIS Partnering Forum	27-28 January 2011, Warsaw, Poland	Atlantis, Greece	Distribution of 100 EN flyers.	Flyers	Stakeholders and experts from the financing, business consulting, IT, etc. sectors <i>No of people reached: about</i>

Deliverable 8.3

Meeting	Discussion on the development and annotation of RuN - parallel corpus of many languages developing at the University of Oslo and parallel corpora needed for the ATLAS project	24 June 2010 University of Oslo, Norway	DCL – IBL, Bulgaria	Working meeting with coordinator of the project Assoc. Prof. Atle Gronn and Assoc. Prof. K. Ra Hauge and distribution of the ATLAS brochure	Distribution of the ATLAS brochure	100 Researchers <i>No of people reached: 4</i>
Workshop	Presentation of research projects (ATLAS focused) of the Department of computational linguistics and the IBL - BAS	28 June 2010 University of Oslo, Norway	DCL – IBL, Bulgaria Language, Bulgaria	Presentation and distribution of the ATLAS brochure	Presentation of the project and distribution of the ATLAS brochure	Researchers at Norsk Ordbok 2014 <i>No of people reached: 21</i>
Exhibition	Recent achievements of the BAS	20-30 September 2010, Bulgarian Parliament	DCL – IBL, Bulgaria	Distribution of the ATLAS brochures	Distribution of the ATLAS brochures	Bulgarian politicians
Exhibition	International Technical Fair	27 September - 02 October 2010, Plovdiv, Bulgaria	DCL – IBL, Bulgaria	Distribution of the ATLAS brochures	Distribution of the ATLAS brochures	International business representatives
Conference	FASSBL07: The Seventh International Conference Formal Approaches to South Slavic and Balkan Languages http://hmk.ffzg.hr/fassbl2010/	3-7 October 2010, Dubrovnik, Croatia	DCL – IBL, Bulgaria	Presentation of the ATLAS project; ATLAS poster; distribution of the ATLAS brochures	Presentation of the ATLAS project; ATLAS poster; distribution of the ATLAS brochures	International research community, representatives from other ICT PSP projects <i>No of people reached: 100</i>
Workshop	Language Resources in Teaching	18 January 2011	UHH, Germany	Presentation of ATLAS project and distribution of the brochure	Brochure	Researchers in Computational Linguistics, Linguists <i>No of people reached: 50</i>
Workshop	Future of Digital Editions	14 January 2011, Mainz, Germany	UHH, Germany	Presentation of ATLAS project and distribution of the brochure	Presentation of ATLAS project and distribution of the brochure	Researchers in digital humanities, Librarians <i>No of people reached: 70</i>
Workshop	Computing Center of the University of Hamburg	June 2010	UHH, Germany	Presentation of ATLAS project and distribution of the brochure	Presentation of ATLAS project and distribution of the brochure	Researchers and Linguists <i>No of people reached: 30</i>

Deliverable 8.3

Workshop	"Digital Humanities"	12-13 April 2010, Kolln Germany	UHH, Germany	Presentation of ATLAS project Distribution of the brochure	Presentation of ATLAS project and distribution of the brochure	Researchers, Linguists <i>No of people reached: 30</i>
Workshop	LREC 2010 Malta Workshop on "Exploitation of multilingual resources and tools for central and (south-) Eastern European languages	18-22 May, Malta	UHH, Germany	Presentation of ATLAS project Distribution of the brochure	Presentation of ATLAS project and distribution of the brochure	Linguists and researchers <i>No of people reached:80</i>
Article in Newsletter	EU fördert cross-linguale Suchmaschine der Arbeitsstelle Computerphilologie	April 2010, Hamburg, Germany	UHH, Germany	A short article about the project appeared in the Newsletter of the University of Hamburg in German. A longer version in the University Newspaper was prepared and will appear in the following months	Article: http://www.uni-hamburg.de/newsletter/EU-foerdert-cross-linguale-Suchmaschine-der-Arbeitsstelle-Computerphilologie-.html	Researchers, Linguists and general public <i>No of people reached: More than 1000</i>
Invited Talk University of Bucharest	Statistical and Rule-based approaches in machine Translation and Applications	16 October 2010, Bucharest, Romania	UHH, Germany	Presentation of ATLAS project. Distribution of the brochure	Brochure	Students and researchers in Computer Science <i>No of people reached: 50</i>
Conference	LREC 2010	17-23 May 2010 Valetta, Malta	ICS PAS, Poland	Initial presentation of the project	Presentation of the project	Linguists <i>No of people reached:50</i>
Employee meeting	Presentation of the project at ICS PAS	21 June 2010 Warsaw, Poland	ICS PAS, Poland	Presentation of the project at ICS PAS Presentation of the brochure	Brochure	IT scientists, linguists <i>No of people reached: 70</i>
Project meeting	NEKST project meeting	21 June 2010 Warsaw, Poland	ICS PAS, Poland	Exchange of ideas on categorization methods for Polish		IT scientists, computational linguists <i>No of people reached: 30</i>
Project meeting	D-Spin project meeting	29 June 2010, Giessen, Germany	ICS PAS, Poland	Distribution of project leaflets, exchange of ideas	Leaflet	Computational linguists <i>No of people reached: 15</i>

Deliverable 8.3

Conference	Digital Humanities 2010 http://dh2010.cch.kcl.ac.uk/	7-10 July 2010 London, UK	ICS PAS, Poland	Distribution of project leaflets	Leaflet	IT scientists, linguists, representatives of humanities <i>No of people reached: 100</i>
Conference	IceTAL 2010: 7th International Conference on Natural Language Processing	16-18 August 2010 Reykjavik, Iceland	ICS PAS, Poland	Distribution of project leaflets	Leaflet	IT scientists, linguists, representatives of humanities <i>No of people reached: 100</i>
Conference	COLING 2010 http://www.coling-2010.org	23-27 August 2010 Beijing, China	ICS PAS, Poland	Distribution of project leaflets	Leaflet	IT scientists, linguists, representatives of humanities <i>No of people reached: 1000</i>
Project meeting	CLARIN consortium meeting	18 October 2010 Vienna, Austria	ICS PAS, Poland	Distribution of project leaflets	Leaflet	Computational linguists, representatives of humanities <i>No of people reached: 45</i>
Conference	Computational Linguistics – Applications @ International Multiconference on Computer Science and Information Technology	18-20 October 2010 Wisła, Poland	ICS PAS, Poland	Distribution of project leaflets	Leaflet	IT scientists, linguists, representatives of humanities <i>No of people reached: 100</i>
Conference	Polish Digital Libraries 2010 Conference	18-22 October 2010 Poznań, Poland	ICS PAS, Poland	Distribution of project leaflets, poster session	Leaflet, Poster	Potential users: archivists, librarians, e-library users <i>No of people reached: 100</i>
Conference	SDH 2010: Supporting the Digital Humanities	19-20 October 2010 Vienna, Austria	ICS PAS, Poland	Distribution of project leaflets	Leaflet	Computational linguists, representatives of humanities <i>No of people reached: 50</i>
Conference	NEERI 2010: Networking Event for the European Research Infrastructures	21 October 2010 Vienna, Austria	ICS PAS, Poland	Distribution of project leaflets	Leaflet	Computational linguists, representatives of humanities <i>No of people reached: 50</i>
Workshop	CLARIN/D SPIN workshop on linguistic Web services	16-17 November 2010 Freudenstadt, Germany	ICS PAS, Poland	Distribution of project leaflets	Leaflet	Computational linguists, programmers <i>No of people reached: 35</i>
Publication of	Electronic Brochure: Program CIP-ICT	January 2011	ICS PAS, Poland	Presentation of ATLAS	http://www.kpk.gov.pl/pliki/11	Researchers, Linguists and

Deliverable 8.3

the information brochure	PSP Przykłady polskiego uczestnictwa			Project	429/Broszura_wersja_elektroniczna.pdf	general public <i>No of people reached: more than 1000</i>
Meeting	Meeting Krzysztof Jassem and Maciej Piasecki teams	14 January 2011, Poznań, Poland	ICS PAS, Poland	Distribution of the initial information on the project among the computational linguists from associated centers. Distribution of project leaflets.	Leaflet	Two teams of computational linguists from Poznań / Wrocław <i>No of people reached: 10</i>
Lecture	ATLAS @ XML and content management lecture at the Warsaw University	17 January 2011, Warsaw, Poland	ICS PAS, Poland	Initial information about the project at Maciej Ogrodniczuk's lecture. Distribution of project leaflets, information on the project	Leaflet	IT students <i>No of people reached: 30</i>
Workshop	"Bring IT On" http://bringiton.info.uaic.ro/	10 December 2010, Romania	UAIC, Romania	Presentation of projects, poster session and distribution of the brochures within BringITon event.	Brochure	Researchers, Linguists, Investitures <i>No of people reached: 90</i>
Conference	"PROMISE"-Processing Romanian in Multilingual, Interoperational and Scalable Environments	29-31 March 2010	UAIC, Romania	Presentation of project	Presentation of project	Researchers, Linguists <i>No of people reached: 20</i>
Organized workshop	How Ideas Evolve	23 October 2009, Sofia, Bulgaria	ITD, Bulgaria	Initial presentation of the project	http://www.digitalspaces.info/events/2009-10-sme-week/how-ideas-evolve-en.html	Researchers, web design companies, web service providers, IT vendors in the field of Digital Content and i-Publishing. <i>No of people reached: 40</i>
Organized workshop	National Information Day - Open Access to Scientific Information	22 October 2010, Sofia Bulgaria	ITD, Bulgaria	Leaflet distribution and presentation	http://www.math.bas.bg/new/site/?call=USE~news;&action=extract&id=44	Researchers, representatives of Sofia City Library, Varna Regional Library "Pencho Slaveykov" and technological companies <i>No of people reached: 20</i>

Deliverable 8.3

Organized workshop	The Future of e-Book and e-Libraries	28 October 2010, Sofia, Bulgaria	ITD, Bulgaria Tetracom, Bulgaria	Leaflet distribution and ATLAS presentation	http://www-it.fmi.uni-sofia.bg/ede2010/programme_en.htm	Researchers, e-book companies, Sofia University Library, Sofia City Library, National Centre for Information and Documentation, ASTEA Solution and StockPodium (technological companies) <i>No of people reached: 20</i>
Organized workshop	The Electronic Libraries and the European Learning Area	16 December 2010, Sofia	ITD, Tetracom, Bulgaria	ATLAS project and services presentation	http://aub-bg.org/index_en.html	Librarians and researchers/ <i>No of people reached: 30</i>

Annex 2: List of dissemination materials produced

The table below presents the URL to the Dissemination materials, presented in the **Deliverable 8.2 Dissemination kit** and latest tools developed.

Dissemination materials	Available at:
ATLAS project website	http://www.atlasproject.eu/
ATLAS project brochures	http://www.atlasproject.eu/atlas/project/dissemination/en
i-Publisher and i-Librarian brochures	http://www.atlasproject.eu/atlas/project/dissemination/en
ATLAS posters	http://www.atlasproject.eu/atlas/project/news/details/en?ci=005926e9-4346-4cc3-ae4b-c742e07dbe9f
ATLAS presentation at CeBIT 2012	http://www.atlasproject.eu/atlas/project/news/details/en?ci=76e85104-10bc-43c4-9d34-23177ca25c88 http://www.cebit.de/product/atlas-build-your-website-at-no-cost/291713/C913812#
Linguistic platform Asset presentation at CeBIT 2012	http://www.cebit.de/product/asset-adds-value-to-your-content/291715/C913812
ATLAS video presentation at Webit 2011, Sofia	http://livinglab.itd-bg.eu/content/projects http://wiki.atlas.itd-bg.eu/images/e/e0/Atlas_Webit_short.wmv
ATLAS presentation in Polish at Success Stories website about Polish successful projects	http://en.kpk.gov.pl/index.php?option=com_sobi2&sobi2Task=sobi2Details&catid=56&sobi2Id=118&Itemid=142&lang=en
i-Publisher website	http://i-publisher.atlasproject.eu/
Technical documentation and guides of Atlas and i-Publisher	http://i-publisher.atlasproject.eu/atlas/documentation/chapter_details?ci=e11013fc-5017-4125-9528-6746c8771141
ATLAS and i-Publisher Getting Started video tutorial	http://i-publisher.atlasproject.eu/atlas/i-publisher/videos
i-Publisher Live demo	http://i-publisher.atlasproject.eu/atlas/i-publisher/demo
i-Librarian website	http://www.i-librarian.eu
EUDocLib website	http://eudoclib.atlasproject.eu/
PLDocLib website	http://www.atlasproject.eu/pl/
News Service Demo	http://newsdemo.atlasproject.eu/
UNESCO Website Demo	http://unesco.atlasproject.eu/
UNESCO Chair Website Demo	http://a2.atlasproject.eu:9090/unesco
World Health Organization Website Demo	http://asset.atlasproject.eu/who
Video materials demo version	http://videodemo.atlasproject.eu/
Press clipping service in Bulgarian and English	http://mediaimage.tetacom.com/
Plagiarism service	http://textmatch.eu/
Icon-painters from Bundovtsi family website	http://www.atlasproject.eu/asset_demo/icons/bg/index.html
MleCel Living Lab website	http://livinglab.itd-bg.eu/