PROGRESS REPORT

Grant Agreement number: 250467

Project acronym: ATLAS

Project title: Applied Technology for Language-Aided CMS

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

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

Period covered: from 01.03.2010 to 31.08.2010

Project coordinator name, title and organisation:

Anelia Belogay, CEO, Diman Karagiozov, CTO,

Tetracom Interactive Solutions

Tel: 35924950444

Fax: 35924950443

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

Project website address: www.atlasproject.eu

DECLARATION BY THE PROJECT COORDINATOR

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

- The attached periodic report represents an accurate description of the work carried out in this project for this reporting period;
- The project (tick as appropriate):

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

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

Name and position of Coordinator: Anelia Belogay, CEO

Date: ...../ ...../.....

Signature: .....

PUBLISHABLE SUMMARY

Introduction

The advent of the Web revolutionized the way in which content is manipulated and delivered. As a result, digital content in various languages has become widely available on the Internet and its sheer volume and language diversity have presented an opportunity for embracing new methods and tools for content creation and distribution. Although significant improvements have been made lately in the field of web content management, there is still a growing demand for online content services that incorporate language-based technology. Mechanisms such as automatic annotation of important words, phrases and names, text 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.

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.

Description of work performed since the beginning of the project and main results so far

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

- 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.

- The first three 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.

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

- A set of use cases to be used by user groups to evaluate the ATLAS platform and the online services was prepared.
- The existing software requirements were revised and updated in order to meet the objectives set for this project. The work on the software requirements and design specification documents also started.
- A specification of the linguistic framework for the language tools to be integrated into the ATLAS platform was drafted, agreed upon and finalized.
- The development of the i-Publisher visualization layer has begun as well as the work on the automatic text categorization tool.
- A prototype of a categorization tool in English was implemented.
- A system (based on a standardized framework) that allows the seamless integration of various types of language tools was prepared.
- Tasks on the adaptation of existing tools and the building of language processing chains for text annotation (a separate processing chain is implemented for each target language) started.
- Improvement of existing translation models needed for the addition of machine translation of text excerpts to the ATLAS platform and the online services has begun as well as the preparation of text corpora for summarization, categorization, and machine translation.

Work carried out on dissemination for this period includes:

- Project information was disseminated through various channels – project web site, newsletters, and distribution of printed materials (brochures.)
- The project was presented on several national and international forums.

Expected final results

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

Main expected final results:

- The software solutions built during the project reveal the true value, capabilities and power of several existing tools for web content management, multilingual versioning, and natural language processing by combining them in an innovative manner and offering the end results to the general public at no cost.

- With i-Librarian and i-Publisher users can easily create, manage and publish multilingual content without installing and maintaining a standalone system. Nevertheless, they retain full control over their content regardless of whether it is in their private workspace, shared

or published. EUDocLib provides easy and intuitive access to a vast collection of EU law documents.

- The ATLAS platform is designed with extensibility in mind, which allows for easy addition of tools for currently unsupported languages as well as new tools for already supported languages.

- Furthermore, ATLAS significantly reduces the time and efforts for content authoring and editing because it automatically categorizes, summarizes, annotates and translates documents regardless of their language and format. The software platform enables i-Librarian users to find the most essential texts from large document collections by displaying text summaries and extracted important phrases, words and names.

- Finally, ATLAS improves content navigation by interlinking content items based on text annotations and by automatically placing the content items in appropriate subject categories.

Potential impact

The project brings together advanced technologies for multilingual web content management and text mining (such as 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 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 Plus will be established at the beginning of the project. Europeana will be approached by the

end of the first year, when the consortium will be able to demonstrate the potential value of ATLAS to the European digital library.

Use

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

Target groups

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

Table 1: Target groups

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

PROJECT PROGRESS

1. Project objectives for the period

The project objectives for the first reporting period are listed below in their respective subgroups:

- Management:
  - Establish an appropriate management and coordination framework for the project and prepare the first project report to be submitted to the EC
  - Monitor project resources, control the quality and ensure the timely delivery of the project report and deliverables relevant for the period
  - Organize the first three project meetings and establish the channels ensuring good management and technical communication

- Technical:
  - Prepare a set of use cases that will be used by user groups to evaluate the ATLAS platform and the online services
  - Revise and update software requirements, and start working on the software requirements and design specification documents
  - Formalize the linguistic framework for the language tools to be integrated into the ATLAS platform
  - Start working on the i-Publisher visualization layer
  - Begin the work on the automatic text categorization tool
  - Prototype of a categorization tool in English
  - Preparation of a system (based on a standardized framework) that allows the seamless integration of various types of language tools
  - Start adapting existing tools and building language processing chains for text annotation (a separate processing chain is implemented for each target language.) Monitor the quality of annotation of noun phrases, words and name entities.
  - Start improving existing translation models needed for the addition of machine translation of text excerpts to the ATLAS platform and the online services.
  - Begin the preparation of text corpora for summarization, categorization, and machine translation

- Dissemination:
  - Disseminate project information through various channels – project website, newsletters, distribution of printed materials (brochures)

- Present the project on both national and international forums

These objectives were fully achieved as illustrated in the following sections of this document.

2. Work progress and achievements during the period

Work package 2 – Software Specification, Implementation and Deployment

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 look 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

- A server was rented, which is used for testing the i-Publisher prototype.
- Tetracom improved its text mining engine, and it is now based on the UIMA framework. ActiveMQ (and respectively the Java Message Service) is used for communication between different components in the engine. Furthermore Tetracom created an environment for testing the integration of language tools within the UIMA framework.
- Several of the English language tools were improved, including the stemmer, noun phrase extractor and named entity recognizer. Furthermore, a lemmatizer, syntactic parser and word sense disambiguation tool were added to the engine.
- Tetracom built a computer cluster that is used for testing various software components including the text mining engine, the distributed databases HBase and GridSQL, and the statistical machine translation system Moses.
- The i-Publisher User Interface prototype 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.

T2.9 – User guides

Tetracom started preparing the user guide for i-Publisher.

Work package 3 – Automatic Categorization

T3.1 – Prototype and integration

- Two monolingual English corpora were prepared that will be used to test automatic categorization algorithms and tools. The first corpus is based on a collection of documents publicly available on the website of the International Relations and Security Network (ISN). The second corpus is based on the English documents in the EUR-Lex database. A set of activity diagrams was also prepared to better illustrate the various workflows related to automatic text categorization.
- Implementation of a categorization tool prototype has started by the introduction of three classification algorithms together with initial runtime behavior experiments. Different categorization approaches were compared and tested on raw text and as a result a concept was defined. Further experiments will be necessary, once the language analysis components for more languages are available. In the first phase a Naive Bayesian classifier and an Entropy-based classifier will be made available, that use initial uni- and bi-gram models over text tokens. A concept for initial multilingual text classification without the use of full text machine translation has been discussed and is planned for the next version of the classification tools.

T3.2 – Fine-tuning of the categorization tool to Croatian

The categorization approaches have been tested on various text types. For the Naive Bayesian and the Entropy-based classifier we have reached some preliminary setting and tuning.

Work package 4 – Language Processing Chains

T4.1 – Implementation of LPCs for the project languages

The partners working on language processing chains selected tools that are not only efficient and effective but that would also fit well within the UIMA framework. Each partner prepared a document with a brief description of the requirements, inputs and outputs of their tools. These documents will serve as a basis for the preparation of a plan for integrating the language tools into the text mining engine. Finally, the partners started adjusting the language tools in accordance with the linguistic framework specification. The ICS PAS and IBL made the prototype integration of most of the Polish and Bulgarian tools with WebCASDebugger, a visual test interface for running and displaying UIMA annotations on different linguistic levels.

T4.3 – ATLAS LPC support extension

The existing text mining engine was modified and extended in order to allow the seamless integration of various language tools. The improved text extraction and language detection components allow documents to be easily preprocessed and forwarded to one of several language-specific processing engines. Further testing is required to ensure that the text mining engine is ready to operate in a production environment.

Work package 5 - Text Summarization

The Consortium considered a necessity to start working on components of this working package before the initially planned start in PM 10. The actual work done refers to preparation of the anaphora resolution component in Romanian, done by UAIC that will be later used by all partners.

T5.1 – Implementation of text summarization tools

- UAIC updated RARE (a general anaphora resolution framework) to be used in text summarization.
- RARE English and German models were done (work done before schedule).
- update of RARE documentation, uploaded on the ALTAS server for the use of all partners

Work package 6 – Machine Translation and Cross-Lingual Search

DFKI restricted considerably their involvement in the project due to internal management and technical issues. The Coordinator redistributed the initially planned work among the other partners and ensured that the DFKI team will transfer some specific knowledge needed for the successful implementation of the tasks. Tetracom and University of Hamburg replace the DFKI in this working package. This working package has started earlier than planned in order for all partners to gain knowledge and experience with the machine translation tools.

T6.1 – Fine-tuning of existing translation models

Tetracom installed Moses and the other Machine translation tools on the test environment and experiments were initiated both with existing and new language models. Partners concentrated on understanding the parameters and language models involved in the Moses system in order to be

able to train the best possible models for each language pair. As a result a conclusion was drawn, namely that the choice of models and parameters is highly connected with language features. Therefore a language customisation is absolutely necessary to be implemented within the ATLAS system.

ICS PAS, Tetracom and IBL DCL compiled a corpus of 100 titles commonly available in all project languages to aid training translation software. The total word count in the corpus exceeds 11M. Moreover, a set of existing ready-to-use parallel corpora have been identified.

Work package 7 - Testing and User Evaluation

WP7 has just started (in July 2010) with the focus being on analyzing the scope and objective of this WP and defining, in practical terms, the methodology to be followed.

Scope & objectives

Ensure the acceptable quality of the final ATLAS system, by assessing the technical performance and the level of fulfillment of the system specifications, as well as assessing the level of user expectations fulfillment.

The scope of main activities includes:

- Design and development of a User Evaluation Plan
- Design of a Technical Test Plan
- Setup and maintenance of a User Group
- Perform Technical & User Acceptance Evaluation
- Analysis of evaluation findings; respond with system improvements.

T7.1: User Evaluation and feedback adjustments

The Task has not officially started yet. However, the methodology for the work has been already elaborated, as outlined below:

- Definition of a User Acceptance Evaluation Plan, with predefined qualitative, usability and user satisfaction indicators.
- Preparation of one or more user questionnaires.
- Involvement of users, through local user groups (1 per country).
- Collection of user evaluation data and a subsequent analysis of that data; improvement of identified shortcomings.

The task will start in November 2010, with the first priority being to prepare the User Acceptance Evaluation Plan. Preliminary work on establishing local user groups will start in September 2010.

T7.2: System Testing (component, integrated system)

T7.2 has just started (in July). The methodology is finalized and centered on:

- The definition of a Technical Test Plan, including technical indicators (their threshold values and the pass/fail criteria), as well as simple confirm indicators. It will also include the Test Cases needed for the unambiguous measuring of the indicators' values.
- The design of an XLS structure to facilitate the whole indicators' measurement and recording exercise.

Apart from the definition of the methodology to be followed, the partners have also made the decision to base the technical indicators on the KPIs (Key Performance Indicators) currently being defined in WP2. The partners have decided to consider indicators at two main levels: a) system wide, b) module specific indicators. The following system wide indicators have been defined so far (work is ongoing; has not been finalized yet):

- MTBF (mean time between failures)
- System uptime
- System response per main functionality.

Highlight clearly significant results

- Software specification;
- Graphical user interface of i-Publisher, based on ZK and Eclipse Equinox technologies;
- UIMA-compatible prototypes of most of the annotation tools from the chain taking English and Polish as sample languages. The results will be presented at the closest project meeting planned for September 2010.
- UIMA Asynchronous scale-out environment allowing horizontal scale of the different Language processing chains.
- Project web site and internal collaboration area;

Work package 8 - Dissemination and Exploitation

T8.2 Detailed dissemination plan

The first draft of the detailed dissemination, which describes the plan for the dissemination of the objectives and results of the ATLAS Project plan, was prepared. It intends to raise awareness and interest on the developed technologies and solutions among the target groups such as the users, the scientific community, the Governmental Institutions, the involved industry and the general public. The major focus is to ensure that the project's research and outcomes are widely disseminated to the appropriate target communities, at appropriate times, via appropriate methods. It includes identification of target communities for dissemination, channels, tools and activities to implement the dissemination strategy.

The Dissemination wiki was developed for collecting dissemination and exploitation plans and results as a communication and dissemination tool. Dissemination Wiki is available at: http://wiki.it.fmi.uni-sofia.bg/. The structure of this Wiki is following the structure of the Dissemination WP:

- Sustainability plan preparation - this section includes all the project partners' individual sustainability plans;
- Primary outreach materials: this section includes all dissemination materials produced during the project to be used in the dissemination activities;
- Reports on dissemination of outreach materials: This section includes periodic reports (say on each three months) regarding all implemented activities;
- Interaction with target communities: This section includes periodic reports (say on each three months) regarding the implemented activities;
- Project impact measurement - this section includes any impact measurement results achieved (Project web site monitoring, Statistical analysis, User feedback analysis, etc.)
- Nationwide dissemination events - planning and implementation of related events (workshops, conferences, seminars, software exhibitions, etc.)
- Trans-European dissemination events - planning and implementation of related events.
- Calendar – the section contains the Calendar with all relevant (at project level) dissemination activities.

T8.5 – Dissemination of primary materials

The first dissemination activities include presentations of the ATLAS projects and dissemination of primary materials on local and international conferences, workshops and meetings.

A short article about the project appeared in the Newsletter of the University of Hamburg (http://www.uni-hamburg.de/newsletter/EU-foerdert-cross-linguale-Suchmaschine-der-Arbeitsstelle-Computerphilologie-.html) in German. A longer version in the University Newspaper was prepared and will appear in the following months. (Cristina Vertan)

Dissemination activities:

The dissemination activities, performed by the partners in the first 6 months of the project, are presented in the following table:

Partner	Event name	Place	Type	Month	Audience	Type of activity
Atlantis, Greece	KIS Partnering Forum 2010	Rome, Italy	Partnering forum	March 2010	Business experts, IT experts	Initial presentation of the project. Distribution of EN flyer.
Atlantis, Greece	Joint event of the Europe INNOVA Annual Partnering Event and EPISIS INNO-Net	Copenhagen, Denmark	Partnering event	June 2010	Business experts, IT experts	80 EN flyers were distributed. ATLANTIS made a presentation about the project and several bilateral meetings were held.
Atlantis, Greece	MOBIP 2010 and Investment in Mobile and IT Services	Valencia, Spain	International business and investment event	June 2010	Stake holders and experts from the financing, business consulting, IT, etc. sectors	Distribute project material and exchange ideas about exploitation prospects.
Institute for Bulgarian Language, Bulgaria	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	University of Oslo		24 June 2010	Researchers	Working meeting with coordinator of the project Assoc. Prof. Atle Gronn and Assoc. Prof. K. Ra Hauge and distribution of the ATLAS brochure Number of participants/recipients/users involved etc. 4
Institute for Bulgarian Language, Bulgaria	Presentation of research projects (ATLAS focused) of the Department of computational linguistics ant the IBL - BAS, given by Svetla Koeva	University of Oslo		28 June 2010	Researchers at Norsk Ordbok 2014	Presentation and distribution of the ATLAS brochure Number of participants/recipients/users involved etc.: 21
Instytut Podstaw Informatyki Polskiej Akademii Nauk, Poland	LREC 2010	Valetta, Malta	Conference	17-23 May 2010	Linguists	Initial presentation of the project

Partner	Event name	Place	Type	Month	Audience	Type of activity
Instytut Podstaw Informatyki Polskiej Akademii Nauk, Poland	Presentation of the project at ICS PAS	Warsaw, Poland	Employee meeting	21 June 2010	IT scientists, linguists	Presentation of the brochure
Instytut Podstaw Informatyki Polskiej Akademii Nauk, Poland	NEKST project meeting	Warsaw, Poland	Project meeting	21 June 2010	IT scientists, computational linguists	Exchange of ideas on categorization methods for Polish
Instytut Podstaw Informatyki Polskiej Akademii Nauk, Poland	D-Spin project meeting	Giessen, Germany	Project meeting	29 June 2010	Computational linguists	Distribution of project leaflets, exchange of ideas
Instytut Podstaw Informatyki Polskiej Akademii Nauk, Poland	Digital Humanities 2010	London, UK	Conference	7-10 July 2010	IT scientists, linguists, representatives of humanities	Distribution of project leaflets
Instytut Podstaw Informatyki Polskiej Akademii Nauk, Poland	IceTAL 2010: 7th International Conference on Natural Language Processing	Reykjavik, Iceland	Conference	16-18 August 2010	IT scientists, linguists, representatives of humanities	Distribution of project leaflets
Instytut Podstaw Informatyki Polskiej Akademii Nauk, Poland	COLING 2010	Beijing, China	Conference	23-27 August 2010	IT scientists, linguists, representatives of humanities	Distribution of project leaflets
Universitaet Hamburg, Germany	Workshop on “Digital Humanities”	University Kölln, Germany	Workshop	12-13 April 2010		Presentation of the ATLAS project

Partner	Event name	Place	Type	Month	Audience	Type of activity
Universitaet Hamburg, Germany	LREC 2010 Malta Workshop on “Exploitation of multilingual ressources and tools for central and (sout-) Eastern European languages	Malta	Workshop			Presentation of the ATLAS project
Universitaet Hamburg, Germany	Computing Center of the University of Hamburg	Hamburg		June 2010		Presentation of the ATLAS project

3. Deliverables and milestones tables

Deliverables (excluding the periodic and final reports)

TABLE 1. DELIVERABLES

Del. no.	Deliverable name	WP no.	Lead participant	Nature	Dissemination level	Due delivery date from Annex I	Delivered Yes/No	Actual / Forecast delivery date
D1.1	Revised detailed project plan	1	1	O	CO	M1	yes	M1
D8.1	Project web site	8	1	D	PU	M2	yes	M2

4. Project management

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

Deviation from work

- WP 5 - Summarization Tools

All partners start working on the tasks prior the initially planned start of the working package.

- WP6 - Machine Translation

DFKI restricted their involvement in the project to 2 PM. The amount of **47 184 euro** from the pre-finance funding of the European Commission was sent back to the Coordinator.

The new involvement of DFKI in Package 6: T 6.1 "Fine-tuning of existing translation models" will cover:
1. Consultancy work for the set up of the machine translation tools in the Atlas environment (this includes the final installation) – 3 weeks;
2. Consultancy work for the usage of the web service provided by "Euromatrix +" - 3 weeks;
3. Consultancy work on how to build new models with new test corpora in order to improve the translation – 2 weeks.

- WP 7 - Testing and User Evaluation

The deliverable D 7.1 planned to be delivered in PM 8 may have a delay of one month.

Work package 1– Project Management

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

T1.1 – Preparation of a revised detailed project plan

The partners responsible for the working packages provided a detailed description of the work within their package and the coordinator compiled a detailed project plan for the whole project. It has been revised by the consortium members and was published on the project website. The coordinator keeps the project plan updated. No delays occur during the reported period.

T1.2 – Commercial agreement preparation

A draft of a commercial agreement, which outlines the commercial interests of the consortium after the end of the project, will be prepared by the end of M7.

T1.3 – Technical communication

Technical communication within the consortium has been established and the information flow is dispersed through several channels:

- development mailing list. The Project CTO, the Work Package leaders, and the beneficiaries responsible for a certain technical task discuss the development process, the current tasks, their progress on a detailed level;
- monthly technical summary. All partners summarize their monthly work within the working packages and report them to the coordinator.

T1.4 – Project coordination and management

The project management activities took place from M1 to M6. A flow of information is organized as follows:

- management mailing list. The Project CEO and the partners staff executives discuss management topics in the Management mailing list;
- internal reports. An internal reporting system was established and a template for an internal report was approved. It is planned for the internal reports to be issued every three months. However 3rd PM internal report was inapplicable due to the lack of significant activities. The 6th PM internal reports were used as a base for compiling the first period project report to the European Commission.
- collaboration area. A project web site internal area is set to enable the partners to work on tasks and topics in protected environment. More details are given in the section Project web site.
- discussion forum. The discussion forum is part of the collaboration area and it is a place for the discussion of technical issues.

T1.5 – Project meetings

Two Project meetings and one Work Package meeting were held during the reported period. List of the meetings together with short overview is placed below. The meeting minutes together with all presentation of the participants for the meetings can be found on the project web site collaboration area.

Project Kick-off meeting
Place: Sofia
Date: 26-27.04.2010

Working meeting (non official)
Place: Malta
Date: 20.05.2010

Work Package 3 meeting
Place: Zadar
Date: 7-8.06.2010

Work Package 2, 3, 4, 5 and 6 meeting
Place: Vienna
Date: 19-20.06.2010

T1.6 – Work on contractual deliverables

 Resource Monitoring

- The coordinator received the EU grant for the project and distributed the funds among partners. Furthermore, the actual state of financial expenditures is constantly monitored through reports prepared by beneficiaries every three months.
- Quality control and work plan monitoring

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

- Reporting to the European Commission

The first project progress report is delivered by the Coordinator to the European Commission.

The project planning is described in T1.1

Project Web site

- Description

The ATLAS web site renders information about the ATLAS project, which aims at delivering a software platform for multilingual web content management unifying automatic annotation of important words,

phrases and names, text summarization and categorization, and computer-aided translation. The web site features two areas – public and restricted. The public area is used for dissemination purposes and the restricted serves area - as an internal collaboration tool for the partners.
The ATLAS web site was launched in April 2010 and includes the following information:

- Scope and objectives of the project
- Description of the consortium (with links to the web sites of partnering organizations)
- Expected results and their envisaged impact
- Who can benefit from the project
- Project news
- Information on project meetings
- Public deliverables and presentations
- Available dissemination material to download
- Links to other relevant projects and EU initiatives

- Goal

The ATLAS website is used as the main information repository of the project, as well as for dissemination of news. The portal also provides access to ATLAS communities and to the ATLAS platform and its repositories. The project web site also serves as a fast and effective communication tool for the purposes of the project consortium, including communication with the European Commission and other stakeholders, as well as planning and arranging project meetings.

- Access

The ATLAS web site can be accessed at http://www.atlasproject.eu. As already mentioned, there is a public area, which is accessible to all users and does not require registration. Non-registered users can see and browse all the content available in the public area. Registered users
(consortium members only), on the other hand, when logged-in can view and modify the content published in the internal area. What should be noted here is that consortium members have their personal workspaces and can chose the access rights for the content they contribute to the web

- Technical Details
  - The Knowledge Management System – a CMS developed by Tetracom IS – is used to set up the ATLAS web site.
  - The web site has a Subversion repository, which facilitates the exchange of project documentation and data.
  - Access to the internal area is provided through username and password after approval of a registration.
  - Common formats such as .doc, .odt and PDF are used for public project documentation to ensure accessibility.
  - During the project the web site will be continuously updated.
  - At the end of the project a DVD-based public showcase will be available for download on the project web site.

USE OF RESOURCES

Overview Person-Month Status (cumulative)

Workapackage	WP1		WP2		WP3		WP4		WP5		WP6		WP7		WP8		Total	
	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual total	Planned total
Tetracom	7.18	7.18	17.72	17.72	0	0	2.86	2.86	0	0	2.48	2.48	0.1	0.1	8.19	8.19	38.53	38.53
DFKI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Atlantis	0.7	0.5	0.5	0.67	0	0.6	2.25	1.47	0	0	0.25	0.12	0.25	1.25	0.75	0.83	4.7	5.44
IBL DCL	0.7	0.7	0.3	0.3	0	0	2	2	0.5	0.5	0	0	0	0	0.1	0.1	3.6	3.6
ICS PAS	0.24	0.24	0.38	0.38	0.21	0.21	3.79	3.79	0	0	0.31	0.31	0	0	0.48	0.48	5.41	5.41
UHH	1	1	1	1	0.5	0.5	1	1	0.5	0.5	2	2	0	0	1	1	7	7
UAIC	0.5	0.5	0	0	0.7	0.7	2.21	2.21	3.24	3.24	0.12	0.12	0.19	0.19	0.28	0.28	7.24	7.24
UniZD	0.47	0.47	2.32	2.32	3.52	3.52	1.20	1.20	0	0	2.96	2.96	0	0	0.34	0.34	10.81	10.81
ITD	0	0	7	7	0	0	0	0	0	0	0	0	0	0	1	1	8	8
TOTAL	10.79	10.59	22.22	22.39	4.93	5.53	15.31	14.53	4.24	4.24	8.12	7.99	0.54	1.54	11.14	11.22		

Actual = number of person months consumed from the beginning of the project to the end of this period
Planned = total effort planned for the project in the latest version of the description of work - annex I to the grant agreement.

Explanation of the use of the resources

TABLE 3.1 PERSONNEL, SUBCONTRACTING AND OTHER MAJOR COST ITEMS FOR BENEFICIARY 1 TETRACOM INTERACTIVE SOLUTIONS FOR THE PERIOD 01.03.2010-01.09.2010

Work Package	Item description	Amount	Explanations
	Personnel costs	80 867	
	Subcontracting	-	
	workshops, dissemination meetings	7 536	Project Meetings: Zadar, Vienna, conference Malta, summer school Zadar
	Meeting in Sofia	999	
	Deprecations	828	
	Hardware	2 512	
	Software	436	
	Server rent	247	
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT		**93 426**	

TABLE 3.2 PERSONNEL, SUBCONTRACTING AND OTHER MAJOR COST ITEMS FOR BENEFICIARY 2 DFKI FOR THE PERIOD 01.03.2010-01.09.2010

Work Package	Item description	Amount	Explanations
	Personnel costs	0	
	Subcontracting	0	
	Major cost item 'X'	0	
	Major cost item 'Y' …….		
	Remaining costs		
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT		0	

TABLE 3.3 PERSONNEL, SUBCONTRACTING AND OTHER MAJOR COST ITEMS FOR BENEFICIARY 3 - ATLANTIS FOR THE PERIOD 01.03.2010-01.09.2010

Work Package	Item description	Amount	Explanations
1, 2, 4, 6, 7, 8	Personnel costs	28.040	
	Subcontracting	-	
	workshops, dissemination meetings	1487.98	Project Meetings: Sofia, Vienna
	Deprecations	-	
	Hardware	-	
	Software	-	
	Server rent	-	
TOTAL DIRECT COSTS AS CLAIMED IN FINANCIAL STATEMENT		**29 527.98**	

TABLE 3.4 PERSONNEL, SUBCONTRACTING AND OTHER MAJOR COST ITEMS FOR BENEFICIARY 4 IBL DCL FOR THE PERIOD 01.03.2010-01.09.2010

Item description	Amount	Explanations
Personnel costs	10444.2	Salaries were paid to the team leader, two programers, the accountant and for a technical support.
Subcontracting		
Travelling and hardware	5178.88	The amount was spent for the project meetings and some equipment.
Remaining costs		
TOTAL DIRECT COSTS AS CLAIMED IN FINANCIAL STATEMENT	**15 623.08**	

TABLE 3.5 PERSONNEL, SUBCONTRACTING AND OTHER MAJOR COST ITEMS FOR BENEFICIARY 5 – ICS PAS FOR THE PERIOD 01.03.2010-01.09.2010

Item description	Amount	Explanations
Personnel costs	19028.26	
Subcontracting	-	
workshops, dissemination meetings	2968.51	Project Meetings: Luxembourg, Vienna, Sofia
Deprecations	-	
Hardware	-	
Software	-	
Server rent	-	
TOTAL DIRECT COSTS AS CLAIMED IN FINANCIAL STATEMENT	**21 996.77**	

TABLE 3.6 PERSONNEL, SUBCONTRACTING AND OTHER MAJOR COST ITEMS FOR BENEFICIARY 6 – UNIVERSITY OF HAMBURG FOR THE PERIOD 01.03.2010-01.09.2010

Item description	Amount	Explanations
Personnel costs	42211.36	
Subcontracting	-	
Other specific costs	3009.64	
TOTAL DIRECT COSTS AS CLAIMED IN FINANCIAL STATEMENT	**45 221**	

TABLE 3.7 PERSONNEL, SUBCONTRACTING AND OTHER MAJOR COST ITEMS FOR BENEFICIARY 7 - UAIC FOR THE PERIOD 01.03.2010-01.09.2010

Item description	Amount	Explanations
Personnel costs	3950	
Subcontracting	-	
workshops, dissemination meetings	2033	Project Meetings: Sofia, Vienna
Deprecations	-	
Hardware	-	
Software	-	
Server rent	-	
TOTAL DIRECT COSTS AS CLAIMED IN FINANCIAL STATEMENT	**5 983**	

TABLE 3.8 PERSONNEL, SUBCONTRACTING AND OTHER MAJOR COST ITEMS FOR BENEFICIARY 8 UNIVERCITY OF ZADAR FOR THE PERIOD 01.03.2010-01.09.2010

Item description	Amount	Explanations
Personnel costs	26 812. 90	
Subcontracting	-	
workshops, dissemination meetings	2 011. 93	Project Meeting: Sofia
Bank fee	215	Money transfer fee
Deprecations	-	
Hardware	-	
Software	-	
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT	**29 039.83**	

TABLE 3.9 PERSONNEL, SUBCONTRACTING AND OTHER MAJOR COST ITEMS FOR BENEFICIARY 9 ITD FOR THE PERIOD 01.03.2010-01.09.2010

Work Package	Item description	Amount	Explanations
WP2	Personnel costs	9948.55	Payments to Roumen Nikolov - 2045,17 EUR, Krassen Stefanov - 1504,22 EUR, Marin Barzakov - 2472,68 EUR, Viktoriya Damyanova - 918,97 EUR, Ivan Koichev - 2151,72 EUR, Stanimira Yordanova - 427,90 EUR, Martin Hristov - 427,90 EUR
	Subcontracting		
	Equipment	2000.54	Purchase of a server (Deprecation taken into account)
	Major cost item 'Y' …….		
	Remaining costs		
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT		**11 949.09**	