

# PROGRESS REPORT

## §3 Project Progress

**Grant Agreement number:** 250416  
**Project acronym:** PLUTO  
**Project title:** Patent Language Translations Online  
**Project type:** Pilot B

---

**Periodic report:** 3rd  
**Period covered:** from 01/04/2012 to 31/03/2013

---

**Project coordinator name, title and organisation:** Dr. Páraic Sheridan, DCU

**Tel:** +353-1-7006706

**Fax:** +353-1-7006702

**E-mail:** psheridan@computing.dcu.ie

**Project website address:** <http://www.pluto-patenttranslation.eu>

**Authors:**

John Tinsley (DCU)

Páraic Sheridan (DCU)

Laura Grehan (DCU)

## Executive Abstract

In this report, we describe the work carried out over the third and final reporting period of the PLuTO project. We present the overall and periodic objectives for the project, emphasising work carried out based on the recommendations made by the project reviewers during the first and second year review meetings. For each work package, we present the main goals, highlights from the period, milestones and deliverables achieved, and some concluding remarks on the overall outcomes in the context of the project. In a separate section, we give an overview of the coordination activities and items which do not necessarily fall under specific tasks outlined in the work plan. We conclude by presenting post-project plans for the exploitation of the technology.

A large proportion of work during the final year of the project has focussed setting the stage for on-going development and commercialisation of the outputs beyond the end of the community support. From a technical perspective, we have progressed on the integration of translation memory and machine translation to the point that we can say there are gains in translation quality to be had when previously unseen ‘user’ translation memories are incorporated into the process. In addition to this, rather than developing machine translation systems for new language pairs, we have focused on addressing specific pain points in the existing systems as identified during evaluations. A number of the techniques developed and evaluated over the period have lead to improved translation output.

These translation technologies continue to be delivered to end-users via our two user interfaces – IPTranslator and ParonPro – which have been further developed and enhanced in the third period. Our evaluation cycle has also completed to include an assessment of translator productivity gains using our technology. This sits alongside our automatic evaluation, human adequacy judgement, usability testing, error analysis, and benchmarking, to provide a comprehensive picture of the overall health of our translation services.

From a non-technical perspective, the consortium has continued to engage with the WON user group who have again proved a source of invaluable guidance and ideas, as well as putting us in touch with relevant people in the industry. We have also continued to keep ourselves visible and build our profile through various dissemination activities.

Finally, in terms of commercialisation and exploitation, we have clearly defined the market opportunities for translation technology in the intellectual property space. We have identified, and in some cases implemented, a number of products which can be built on top of the technology to offer significant value to various sets of users. Some of these prospects have been validated and relationships have already been established with various potential customers and in some cases collaborative work has already commenced. On-going exploration will be supported through subsequent state investment rewarded to DCU and under the umbrella of ESTeam’s existing business.

In summation, over the course of the three years of the project we have developed state-of-the-art patent translation software and supporting architecture which is now in a position to compete for a segment of the language technology market. We have laid stable foundations upon which both the technology, expertise, and relations developed during the course of the project can be built upon and exploited to what we are optimistic will be a successful conclusion.

<b>EXECUTIVE ABSTRACT .....</b>	<b>2</b>
<b>1 PROJECT OBJECTIVES FOR THE PERIOD .....</b>	<b>5</b>
1.1 RECOMMENDATIONS FROM THE PRECEDING PERIOD .....	6
1.2 NOTE ON DELIVERABLES .....	6
<b>2 WORK PROGRESS AND ACHIEVEMENTS DURING THE PERIOD .....</b>	<b>7</b>
2.1 WP2 DATA ACQUISITION, SELECTION, AND INTEGRATION .....	7
2.1.1 Objectives.....	7
2.1.2 Progress Highlights .....	7
2.1.3 Tasks .....	7
2.1.4 Use of Resources .....	8
2.1.5 Summary.....	8
2.2 WP3 WEB APPLICATION AND USER INTERFACE .....	8
2.2.1 Objectives.....	8
2.2.2 Progress Highlights .....	8
2.2.3 Tasks .....	9
2.2.4 Use of Resources .....	9
2.2.5 Summary.....	10
2.3 WP4 TRANSLATION MEMORY .....	10
2.3.1 Objectives.....	10
2.3.2 Progress Highlights .....	10
2.3.3 Tasks .....	11
2.3.4 Use of Resources .....	12
2.3.5 Summary.....	12
2.4 WP5 MACHINE TRANSLATION.....	12
2.4.1 Objectives.....	12
2.4.2 Progress Highlights .....	12
2.4.3 Tasks .....	13
2.4.4 Use of Resources .....	14
2.4.5 Summary.....	14
2.5 WP6 SYSTEM INTEGRATION .....	14
2.5.1 Objectives.....	14
2.5.2 Progress Highlights .....	15
2.5.3 Tasks .....	15
2.5.4 Use of Resources .....	16
2.5.5 Summary.....	16
WP7 EVALUATION AND QUALITY ASSURANCE .....	16
2.5.6 Objectives.....	16
2.5.7 Progress Highlights .....	16
2.5.8 Tasks .....	17
2.5.9 Use of Resources .....	18
2.5.10 Summary.....	18
2.6 WP8 DISSEMINATION.....	19
2.6.1 Objectives.....	19
2.6.2 Progress Highlights .....	19
2.6.3 Tasks .....	19
2.6.4 Use of Resources .....	20
2.6.5 Summary.....	20
2.7 WP9 EXPLOITATION AND STANDARDISATION .....	21
2.7.1 Objectives.....	21
2.7.2 Progress Highlights .....	21
2.7.3 Use of Resources .....	23
2.7.4 Summary.....	23
<b>3 DELIVERABLES AND MILESTONES TABLES .....</b>	<b>24</b>

<b>4</b>	<b>PROJECT MANAGEMENT.....</b>	<b>25</b>
4.1	MANAGEMENT TASKS AND ACHIEVEMENTS.....	25
4.1.1	<i>Commercial Development</i> .....	25
4.1.2	<i>User Engagement</i> .....	25
4.1.3	<i>Advisory Board</i> .....	26
4.1.4	<i>Recruitment</i> .....	26
4.2	PROJECT MEETINGS.....	26
4.2.1	<i>Second Year Review Meeting – Ghent – May 15<sup>th</sup> 2012</i> .....	26
4.2.2	<i>User Group Meeting – Venlo – June 15<sup>th</sup> 2012</i> .....	26
4.2.3	<i>Management Meeting – Hamburg – November 6-7<sup>th</sup> 2012</i> .....	27
4.2.4	<i>Evaluation Meeting – Ghent – ? 2012</i> .....	27
4.2.5	<i>User Group and Technical Meeting – Ghent – 15/16<sup>th</sup> January 2013</i> .....	27
4.2.6	<i>AGM – Paris – April 6-7<sup>th</sup> 2012</i> .....	27
	DISSEMINATION .....	27
4.3	ACTIVITIES .....	27
	<i>Dissemination Materials</i> .....	28
4.3.1	<i>and Social Media</i> .....	28
<b>5</b>	<b>SUMMARY.....</b>	<b>28</b>

# 1 Project Objectives for the Period

The overall aim of PLuTO is to develop professional web-based solutions for patent translation through the integration of machine translation and translation memory technologies. These solutions should meet real world commercial needs and be exploited by the consortium come the end of the project. Iterative improvements will be made to the early stage prototypes and concepts over the course of the project, guided by the outcome of evaluations carried out by Cross Language and interactive feedback from the WON user group and other entities engaged with the consortium.

In terms of non-technical objectives, continued dissemination and exploitation activities are necessary in order to attract attention to the project and to build awareness and foster relationships with potential users. Additionally, the consortium must establish potentially viable channels for commercialisation of the proposed work and ensure that technical developments are in line with this.

The majority of the technical groundwork was laid in the first two years of the project to support the achievement of these goals. The primary objective of year three is to build upon this work to ensure that the technology and opportunities are in a position to be competitive in the market beyond the end of the community support. This involves commercial exploration and assessment of the landscape as well and maintenance of the presence of the (outputs of the) project in the Intellectual Property (IP) and language technology communities.

Significant foundations were laid in the first year of the project to support the achievement of these overreaching goals. The primary objective of year two is to build upon these foundations to convert the initial prototype of Deliverable 6.1 into a more production grade software solution(s) in order to allow for real world testing. This objective is supported by a number of sub-tasks which, upon integration, comprise the development of the overall solution along with assessments of its viability.

A number of sub-tasks were outlined for year three to support the overreaching objectives. These tasks tie up a number of loose ends in terms of the work described in the various work packages, and complete the development and evaluation cycles of the various system components. More specifically, the objectives of the consortium for the period described in this report amount to:

- The development of MT engines for 2 additional language pairs;
- The creation of TM resources for the same;
- The provision of relevant patent corpora to support the aforementioned tasks;
- The production of a final integrated system with a comprehensive functionality report;
- Delivery of a stable, robust, and accessible interface through which the applications can be used;
- Continued evaluation of the translation quality with particular emphasis on the impact of TM integration;
- Provision of results of a final user survey on the opportunities for the PLuTO technology from an end-user perspective;
- Continued dissemination of the project including building awareness and attending industry events;

- Delivery of a final exploitation plan outlining concrete ways in which the PLuTO technology can be used beyond the end of the project.

### ***1.1 Recommendations from the preceding period***

In addition to these objectives, a number of recommendations were made to the consortium based on the technical review of the second period. Below is a summary of those recommendations along with an indication of how they have been taken into account for this period.

- Reconsider how efforts on TM and TM/MT technology is directed
  - ⇒ **Action:** Repeated previous evaluation on a new language pair (EN-DE) to rule out the high quality of the MT system as a factor;
- Develop a fully specified product specification
  - ⇒ **Action:** Functionality reports and user manuals have been prepared including pitch decks describing the products and their value propositions
- Develop a fully specified dissemination plan
  - ⇒ **Action:** Extensive go-to-market and feasibility studies have been prepared and used to secure additional state investment.
- Further refine TM and MT evaluation
  - ⇒ **Action:** Carried out evaluations using previously unseen ‘user’ TM to assess the impact on translation quality and extended MT evaluation to include and assessment of translator productivity gains

### ***1.2 Note on Deliverables***

A number of deliverables for this period have been consolidated in order to reduce unnecessary repetition. For example, in Work Package 5 Machine Translation D4.3 ‘MT Systems for 2 language pairs’ and D4.4 ‘Final Report on MT’ have been consolidated to include a report on the activities in the final year as well as a retroactive look at MT over the three years. A similar scenario occurs in Work Package 4 while Deliverables 9.2 and 9.3 have been combined as they are essentially addressing the same topic. Finally, for Work Package 3, the individual documents on various components of the user interface have been consolidated into two reports which essentially comprise ‘user manuals’ for the IPTranslator and ParonPro interfaces. Backend functionality and architecture for these offerings is described in the context of Work Package 6.

## 2 Work Progress and Achievements during the Period

In this section, we describe each work package in detail outlining the global objectives, progress made during the period, followed by specific details on the individual tasks set out in the Description of Work. Work package 1 – Management – is excluded here as it is treated as a standalone topic in section 4.

### 2.1 WP2 Data Acquisition, Selection, and Integration

#### 2.1.1 Objectives

The global objectives of this work package are to ensure the constant availability of patent data to the consortium for the purpose of training MT engines and producing TM resources. New data should be made available at the approximate rate of 2 language pairs per project year. Deliverables and milestones falling due in the period are shown in Table 1.

Mi2.5	(EN, PT, FR, DE, ES, JP) patent data available	✓
Mi2.6	(EN, PT, FR, DE, ES, JP, ZH) patent data available	✓

Table 1 Milestones and deliverables due between M24 and M36

#### 2.1.2 Progress Highlights

As stated, the key objective of this work package is to provide data across a number of language pairs to be used for MT training and TM building. Although the description of work indicated that data would be made available periodically as it was sourced for a particular language, in practice the majority of the data was available immediately as it was essentially coming from two data sets - MAREC and Alexandria – which the IRF has access to from the beginning of the project.

While data was acquired for new languages from alternative sources (Portuguese, Chinese, Japanese), again this data was at hand from early in the project. Given this, the work required under this work package was limited in this period to preparation and mark-up of the data for different experiments.

#### 2.1.3 Tasks

##### T2.1 Meta-data definition

A metadata definition has been agreed across the partners as the format of the data is integral to the key components, namely the MT engine (input/output formats), and the integrated TM/MT system.

The important fields in the mark-up remain consistent with those described in previous deliverables (Deliverable 1.1a, Deliverable 2.2).

##### T2.2 Selection Engine

This task was discontinued following M12.

## T2.3 Data Acquisition

As discussed in Deliverable 1.1b, by the end of the second period, WP 2 had delivered all of the data it was obliged to provide over the course of the project. As will be discussed in subsequent reports and presentations for this period, for this period the consortium decided to take a “quality over quantity” approach and focus on improving the quality of translation technology for existing languages rather than source data for further new languages. To that end, no new data was acquired in year three.

? Monika data?

? Chinese dictionary?

### 2.1.4 Use of Resources

Beneficiary	PMs to date	PMs yr3	PMs Total	PMs Overall
DCU	1.2			8.0
ESTeam	2.7	0.3	3.0	4.8
IRF	4.2	-	4.2	4.2
Cross Language	0			3
WON	0		0	0
<b>Total</b>	10.1			20

### 2.1.5 Summary

We have presented and summarised the data acquisition activities in the project over the course of the project. Data was acquired for all the relevant language pairs (and new languages as we adapted to feedback) ahead of time allowing the consortium to focus efforts on improving the quality of the language technology.

## 2.2 WP3 Web Application and User Interface

### 2.2.1 Objectives

This main goal of this work package is to design and implement the front-end user interface of the PLuTO service(s). The web application will provide the end-user a gateway to access translation services and other functionality on the back-end. Deliverables and milestones falling due in the final period are shown in Table 2.

Mi3.3	Final web application and user interface	✓
D3.3	Final web application and user interface (Database Schema, Web Service Interface, Business Utility Interface)	✓

Table 2 Milestones and deliverables due between M24 and M36

### 2.2.2 Progress Highlights

In Deliverable 1.1b, we presented our two distinct user interfaces: IPTranslator and ParonPro. Over the course of the final period, these interfaces have evolved based on our research in WP 9 and our findings and feedback from WON in WP 7 and through discussions within the intellectual property (IP) community in WP 8. This interaction with end users has proved invaluable in guiding the development of these interfaces and bringing them to a



level where they can be potentially released as market ready products. We have structured the final deliverables for this WP (Deliverables 3.3a and 3.3b) in “user manual” style describing the functionality (and in some cases the motivation behind this functionality) of the respective offerings.

### 2.2.3 Tasks

#### T3.1 Data Layer

The data access layer, or data layer, concerns all instances in which data is read or written when using PLuTO services. It is the component in the web application that connects the user interface to the various data repositories, e.g. patent documents, statistics/logs, user database, etc.

Both interfaces make use of databases, as described originally in Deliverable 3.1, to store the requisite user information such as login details, translation history, and various logging and statistics. The type of information stored includes uploaded data, source and target languages, number of words translated, and, depending on the configuration, the translated text.

#### T3.2 User Interface

The user interface concerns the means by which the end-user will interact with the system; essentially the web-based GUI.

#### IPTranslator

The IPTranslator user interface has been extended and refined on a number of levels including the user signup process, the translation workflow for various file types, and the ability to interact with the translation output. This functionality is supported by a number of developments on the back-end in WP5 and WP6 described further in the respective deliverables for those work packages. These user interface features and more are described in greater detail in Deliverable 3.3a.

#### ParonPro

By the end of the second project period, the idea behind the ParonPro interface was still a relatively new concept. Following extensive engagement with end users, this interface has been significantly extended to include a number of features and additions specifically identified by patent professionals as key functionality for such a workflow management tool.

This functionality includes, amongst others, the development of an API through which users can batch import patent data from a number of different sources, the ability to annotate searches and documents for the purpose of collaborative research, and the ability to export final search reports into custom templates. These features and more are described further in Deliverable 3.3b.

#### T3.3 Application Interface

The application interface addresses the definition of the web services through which the front end and back end of the applications communicate. The details of these services remain the same as those described previously in Deliverable 3.1.

### 2.2.4 Use of Resources

---

Beneficiary	PMs to date	PMs yr3	PMs Total	PMs Overall
-------------	-------------	---------	-----------	-------------

---

DCU	2.8			10.0
ESTeam	30.2	25.7	52.9	40.5
IRF	2.5	-	2.5	2.5
Cross Language	0			12
WON	1	1	2	1
<b>Total</b>	36.5			66

## 2.2.5 Summary

Two distinct user interfaces and applications have been developed to support the commercial objectives of the project. The IPTranslator interface has been released and in use by more than 200 patent professionals over the past 18 months. The ParonPro interface is now in alpha release and is currently being tested by a closed set of invited users.

## 2.3 WP4 Translation Memory

### 2.3.1 Objectives

The key aim of this work package is to create translation memory (TM) resources from the data provided in work package 2. This involves pre-processing of the raw data, structuring of the data based on the IPC system and alignment of multilingual segments. These resources are then to be exposed in a database to other components (web application, MT engine) via web services defined in work package 6 and implemented in work package 3. Deliverables and milestones falling due in the period are shown in Table 3 Milestones and deliverables due between M24 and

Mi4.5	TM resources for EN-JP	✓
Mi4.6	TM resources for EN-ZH	✓
D4.3/4	Final report on TM data resources	✓

**Table 3 Milestones and deliverables due between M24 and M36**

### 2.3.2 Progress Highlights

In previous periods, the majority of efforts in this work package were focused on developing TM resources for new languages. As results to date regarding the integration of TMs have not been favourable, more effort spent in this period improving existing processes and working on the concept of user TMs.

To this end, the final configuration for TM/MT has been developed with new statistical processes implemented for formula tagging and alignment filtering. Alignment dictionaries were enhanced all previous language pairs were reprocessed given the new processes. Additionally, the client-specific TM was prepared with segment and sub-segment alignments applied and manually verified. Finally, small “proof of concept’ TMs were developed for the Asian language pairs, Chinese and Japanese, to demonstrate that these languages are a further option when assessing post-project possibilities. A more detailed description of these activities is given in Deliverable 4.3/4.

### **2.3.3 Tasks**

#### **T4.1 Data Management**

In addition to the bottleneck of automatic text alignment quality, TM contributions are competing against the MT system that has been trained on the exact same aligned data. It was indicated previous reviews that it made sense to also investigate the potential of a combined TM-MT scenario in client-specific cases. In that case, the client-specific TM is of high quality (sentence, segment, sub-segment aligned pairs have been verified) and can be used in collaboration with the “general purpose” MT system trained on general patent data.

During this period, we have applied the completed TM-MT integration scenario in both the general case and a client-specific case. The initial implementation of the heuristic-based formula tagger was replaced with a statistical approach which achieved significantly faster predictions with better generalisation capability. A full set of evaluation and results for the new tagger are provided in Deliverable 4.3/4 while the impact of the client-based TM on translation quality is assessed in Deliverable 7.10.

#### **T4.2 Structuring TM domains according to patent data domains**

While data for German, French, and Portuguese was structured according to the IPC classification, this was not possible for the Spanish data as the requisite metadata was not available. A similar scenario was encountered in the Japanese and Chinese, although one might suspect there would have been data sparseness issues due to the relatively small size of the proof of concept TMs. It is conceivable that, in the future, the data could be automatically classified according to the IPC system and annotated as such.

#### **T4.3 Alignment**

The TMs have been aligned at the hierarchical levels: sentence, segment, and sub-segment. The notion of paragraph level alignment from year 1 has been dropped, as repetition at this level in patent documents is essentially non-existent.

During this period, work has focused on improving the alignment results but, most importantly, improving the reliability of the alignment results. For the former, we have enhanced the multilingual dictionaries which are referenced by the text aligner. For the latter, we have developed a statistical classifier for rejecting inaccurate align pairs, as discussed below.

#### **T4.4 Data loading and quality control**

Alignment results have been filtered automatically with the use of the new statistical classifier. This classifier has been trained on a sample of 1000 manually revised alignment pairs for each language pair and each of five align score regions (<20%, <40%, <60%, <80%, <100%) in order to have a balanced representation of the full set of alignments, even though the size of the training set is not adequate to achieve generalization and avoid a highly biased model.

While the results were not very encouraging using this approach, we expect the performance would increase substantially if the module was trained on a larger set of manually verified samples. Full details of the experiments leading to these results are given in Deliverable 4.3/4.

### 2.3.4 Use of Resources

Beneficiary	PMs to date	PMs yr3	PMs Total	PMs Overall
DCU	0.3			18.0
ESTeam	25.5	13.4	38.9	41.0
IRF	0	-	0	0
Cross Language	0	0	0	0
WON	0	0	0	0
<i>Total</i>	25.8			59

### 2.3.5 Summary

A number of existing processes have been implemented and new modules developed, designed to improve the quality of existing TMs. While results to date employing this processes in the general TM case, in this period they have been applied to a significantly greater effect in the client-specific TM case. This brings the TM efforts to an encouraging conclusion, not only in the context of patents, but with the possibility of even greater effects in domains for which TMs are already established.

## 2.4 WP5 Machine Translation

### 2.4.1 Objectives

The principal goal of the Machine Translation (MT) work package is to build MT engines for the language pairs being addressed in the project using the MaTrEx system. Additionally, in order to achieve optimal performance, we build upon the findings from the evaluations in WP7 to further optimise the systems for patent translation. Deliverables and milestones falling due in the period are shown in Table 4.

Mi5.5	MT engine for EN-ES	✓
Mi5.6	MT engine for EN-ZH	✓
D5.3/4	Report on MT Engines for each language pair, final	✓

Table 4 Milestones and deliverables due between M24 and M36

### 2.4.2 Progress Highlights

By the end of the second period we had delivered the required number of systems for the entire project. While it would have been possible for us to build systems for additional languages, the principal feedback we received from users and advisors was that we had built MT systems for the most crucial, in-demand languages (having modified our work plan to do so) and that our subsequent efforts should be concentrated on improving the quality of these systems, the main motivation being that translation quality will be the ultimate metric on which our service will be judged from a commercial perspective.

To that end, we identified specific areas in our various engines which could benefit from improvements, specifically for our German and Asian (Japanese and Chinese) language systems. We tried a number of approaches to address lexical coverage and structural

divergence issues which, while perhaps not being completely production ready, have yielded promising results with significant prospects for actively improving the quality of the systems going forward.

### **2.4.3 Tasks**

#### **T5.1 Adapting existing MT technology to the patent domain**

A number of the pain points in our existing MT systems are directly related to the fact that we are operating in the patent domain. One such pain point is the presence of untranslated words in the MT output which occurs when there are words in the input text which have not previously been seen during training (so-called out-of-vocabulary, or OOV, words).

This issue is particularly prevalent for patents as they are a rich source of highly technical terminology and neologisms. It is exacerbated for Asian languages as these technical terms can cause issues during pre-processing and segmentation (particularly for Japanese where the where the different writing systems come into play) where errors in tokenisation can also lead to words being left untranslated.

We explored two solutions to this problem for Japanese; one which addresses the problem prior to translation, and one which addresses it after translation. In the former, we developed a module to normalise the text to a single script to reduce the number of OOV characters. Using this approach, we achieved a 13% decrease in the OOV although questions remain as to the accuracy of the normalised words.

The second approach involved leveraging on fact many of the OOV words are written in the katakana script which is essentially a character/phoneme based version of English. We developed a machine transliteration system which took words left untranslated by the MT system and transliterated them into roman script. This approach was more effective and is close to being ready to be deployed in production systems. A full description of these approaches and their evaluation is presented in Deliverable 5.3.

#### **T5.2 Language Specific Processing**

In the Description of Work, this task describes “Integration between SMT/EBMT and RBMT”. However, as there is nothing new to report in terms of those approaches (as we have settled on a stable MT configuration), we will use this section to discuss another technique we have applied which is not directly related to the patent domain.

In Deliverable 5.3 last year, we mentioned that we had begun preliminary investigations into the pre-ordering of the input text to be translated in order to make it align better with the target language. We have carried out significant work on this approach for German-English using dependency parses to analyse the input text and identify the clauses which will give the highest gain when reordered. To date we have applied this approach to the input at runtime and the results are promised. Once we have defined the best approach to dynamic reordering for a particular language, we can retrain our systems with preordered input to increase the robustness of results.

#### **T5.3 Integration between MT and TM**

As described in from WP5 and WP6 to daye, the machine translation engines have been deployed as web services to accommodate integration with the translation memories. The payload returned from the MT service has been enhanced to include additional information

which can be exploited to improve the quality of the TMs and the integrated system. In this period, we have exposed word alignment information and translation scores in our API, in addition to the development of probabilistic bilingual dictionaries as an external resources. Full details of these features are described in Deliverable 5.3/4 and their use is detailed in Deliverable 4.3/4.

#### 2.4.4 Use of Resources

Beneficiary	PMs to date	PMs yr3	PMs Total	PMs Overall
DCU	47.5			41.0
ESTeam	12.5	10.1	22.6	17.6
IRF	0	-	0	0
Cross Language	0	0	0	0
WON	0	0	0	0
<i>Total</i>	59.9			58.6

#### 2.4.5 Summary

We have discussed the new work which has taken place over the course of the third period of the project. This involved efforts to improve translation quality based on input reordering for languages with syntactically divergence structures. We also looked at ways to decrease the out of vocabulary rates for our Asian language MT systems by means of a number of different techniques such as script normalisation, machine transliteration, and vocabulary enhancement. Many of these approaches exhibited promise though not all are yet ready to be deployed in our production systems. Aside from translation quality, we looking at efforts we have made to improve the MT web services including the IPTranslator user interface and feature set, the client API, and the translation memory integration with ESTeam. Looking at this WP as a whole, MT technology has continue to advance and improve and we believe it will serve as an extremely solid foundation for a viable commercial offering.

### 2.5 WP6 System Integration

#### 2.5.1 Objectives

The System Integration work package is essentially in place to provide the technical framework in which the various components – the machine translation systems, the translation memories, and the various other interfaces and components – developed over the course of the project can exist and interact. From a functional perspective, it can be considered the ‘back-end’ to the ‘front-end’ user interface tasks carried out in work package 3. Deliverables and milestones falling due in the period are shown in Table 5.

Mi6.4	Integrated Systems, final	✓
D6.2	Final system	✓

**Table 5 Milestones and deliverables due between M24 and M36**

## 2.5.2 Progress Highlights

As before, the highlights in this work package reflect the main achievements in the project as a whole in this period; that is to say as opposed to there being significant novel developments, the architecture has been enhanced and improved to bring it to a point where it is robust enough to support the production level commercialisation activities proposed by the consortium members beyond the end of the project.

This can be seen in the fact that many of the development directly improve the end-user interaction with the services. From the perspective of ParonPro, the extension of the data import function to facilitate automatic import through the EPO's API allows users to easily integrate EP documents which often form the core of European searches. Additionally, the language identification feature feeds directly into the translation feature which is increasingly important as the EPO, in particular, continues to add non-EP collections to their database.

With regards to IPTranslator, user management has been completely revamped with a new database, authentication, and mail service. This significantly increases the signup experience and facilitates the customer relationship management aspect from our perspective. Additionally, we have added on-the-fly optical character recognition (OCR) functionality to support the translation of PDFs through the user interface. Translation of PDFs was one of the key pain points identified when speaking to patent information professionals.

Finally, extensions to the machine translation API have increase the functionality of the TM/MT integration scenario as well as providing a foundation for the 'business to business' opportunities for IPTranslator.

## 2.5.3 Tasks

There are four main tasks outlined in the context of this work package:

1. Integration Requirements Analysis
2. Integration Prototype Generation
3. Integrated System Development
4. Integration Testing

These tasks are not sequential in nature. Aside from task (2), which was more related to the development of initial prototypes, these tasks represent a product development cycle in which plans are outlined (1), implements (3), and tested (3). Following testing, further requirements are often identified which typically leads to a new cycled of development.

In the context of the third period, for (1) we constantly evaluated the on-going requirements for integration which lead to the various changes and improvements made over the period including, but not limited to: language detection, data import, API development, implementation of the OCR server, user account managements, and TM/MT integration. Task (3) reflects the development effort to put these features in place in the various back-and front-ends. Finally, in terms of (4), testing typically involves a first phase carried out by developers and a second phase in which the new features are released to beta users who provide feedback which feeds into the next development cycle.

## 2.5.4 Use of Resources

Beneficiary	PMs to date	PMs yr3	PMs Total	PMs Overall
DCU	7.3			13
ESTeam	28.6	13.7	42.3	36.0
IRF	3.1	-	3.1	3.1
Cross Language	1.5			1.5
WON	0	0	0	0
<i>Total</i>	50.5			53.6

## 2.5.5 Summary

The two applications, IPTranslator and ParonPro, which were originally built on top of the translation services provided by the machine translation and integrated translation scenarios, have been improved and enhanced to bring them to a point where they are capable to support the on-going commercialisation of the project outcomes following the end of the community support. Analysis and feedback on the respective service by users help to focus the software development on areas of maximum impact which would resonate with end-users when introduced or improved. Full technical details on the work described above are given in Deliverable 6.3.

## WP7 Evaluation and Quality Assurance

### 2.5.6 Objectives

The ultimate goal of the Evaluation work package is to ensure that the web applications we develop meet the needs of end-users, be it patent searchers, IP specialists, or other potential users. In order to ensure the required standards are met, the individual components of the system must undergo a thorough evaluation and quality assurance process throughout the duration of the project. This is carried out not only by evaluating translation quality – both MT only and integrated TM/MT – but also benchmarking the systems against existing tools and assessing the extrinsic value of the translation output through translator productivity evaluations. Deliverables and milestones falling due in the period are shown in Table 6.

Mi 7.3	Survey structure and content v2 available	✓
D7.3	Final report on survey's results	✓
D7.7	Final report on the intrinsic and extrinsic quality of MT	✓
D7.10	Final report on the impact of TMs in the integrated translation system	✓

**Table 6 Milestones and deliverables due between M24 and M36**

### 2.5.7 Progress Highlights

The main highlight for year three in the context of this work package has been the validation of the viability of the integrated TM/MT scenario in the patent domain. Previous evaluations as well as new evaluations carried out in this period have shown limited potential for the



integrated scenario when both the TMs and MT systems have been built on the same data. The most recent set of evaluations involved the addition of a previously unseen ‘client-specific’ TM into the integrated scenario. Despite the fact that this TM was relatively small (35,707 words) we nonetheless saw a significant increase in TM leverage to the point that TM/MT output was preferred to the MT only output in almost 60% of the evaluated cases. Full details of the experimental setup and results are provided in Deliverable 7.10.

The MT evaluation was rounded out by completing the extensive evaluations for the English—German, English—Japanese, English—Spanish, and English—Chinese systems. This set of languages evaluated in this period are considered more “difficult” for MT and while we did not expect the results to be as good as those presented in Deliverable 7.6, we were nonetheless pleased by the results achieved. Full details of the automatic scores and human evaluation, which includes benchmarking against competing systems, error analyses, and adequacy assessment, are given in Deliverable 7.7.

In addition to the MT evaluations repeated for new languages, this period also saw the first “productivity evaluations” in which we assessed the speed gains achieved by translators post-editing our MT output as opposed to translating from scratch. This evaluation is particularly relevant given the focus of our exploitation plans as described in Deliverable 9.2. Productivity tests were carried out for two translation directions: French->English, and Chinese->English. The results, again described fully in Deliverable 7.7, generally correlate with the results for the previous evaluations. Substantial gains were achieved for French->English where the quality of our systems is particularly high, while results were less conclusive from the Chinese->English system.

Finally, we carried out a smaller more detailed survey with the WON Pluto working group to help gather more intelligence and opinions with respect to our commercialisation plans. The responses from the survey are summarised in Deliverable 7.3.

## **2.5.8 Tasks**

### **T7.3 Translation Evaluation**

A range of tests were carried out to evaluate the performance of the English—German, English—Spanish, English—Japanese, and English—Chinese machine translation (MT) systems described originally in Deliverable 5.2. In addition to assessing the MT systems using automatic evaluation metrics such as BLEU and METEOR, a large-scale human evaluation was also carried out. MT system output is ranked from 1—5 based on the overall quality of translation, and the individual mistakes made were identified and classified in an error categorisation task.

On top of this standalone evaluation, the MT systems were also benchmarked against leading commercial systems across two MT paradigms: Google Translator for statistical MT and Systran (Enterprise) for rule-based MT. A comparative analysis was carried out using both the automatic and human evaluation techniques described above.

All evaluations were carried out using held-out test data randomly selected from our parallel patent corpora. Where possible, test sets for the automatic evaluation were segmented into sub-sets based on the IPC patent classification system. In doing this, the evaluation would indicate in which categories of patents (e.g. chemistry, engineering, etc.) the translation systems were performing better.

We saw a mix of results across the four language pairs including cases where the PLuTO engines performed well and were indisputably better across the board (English—German), cases where the results were inconclusive but demonstrated a clear need for improvement (English—Japanese), cases where the results were affected by the lack of available high-quality in-domain training data (English—Spanish), and cases where the results were mixed depending on the translation direction (English—Chinese).

Furthermore, the productivity evaluation involved assessing how much quicker a set of professional translators could translate a document by post-editing PLuTO MT output versus translating in from scratch. We saw an average productivity gain of 125% for French->English, while no conclusive gain was seen for Chinese->English. Full details from these evaluations are given in Deliverable 7.7.

#### **T7.4 Evaluation of Integrated TM and MT**

The specific objective of this task is to evaluate the impact of the translation memories on translation quality in the integrated scenario compared to the standalone machine translation quality. We carried out a series of experiments which provide a range of evaluation data detailing the performance on the German—English language pair. This pair was selected to complement the evaluations that had previously been performed for the French—English language pair.

In addition to assessing both systems using automatic evaluation metrics such as BLEU and METEOR, a human evaluation is carried out in two different evaluation tracks. By dividing evaluation into two tracks we are not only able to observe evaluate the impact of the updated TM/MT integration methodology in comparison to French—English evaluations, but are also able to evaluate the performance of TM/MT integration in a more customer-specific scenario, using different data sets for the training of the MT engine and the creation of TM.

While results were again less than promising for the generic integration scenario, the results were hugely encouraging for the customer-specific scenario. Full details and discussion on the evaluation methodology and results are provided in Deliverable 7.10.

#### **2.5.9 Use of Resources**

<b>Beneficiary</b>	<b>PMs to date</b>	<b>PMs yr3</b>	<b>PMs Total</b>	<b>PMs Overall</b>
DCU	3.8			33.0
ESTeam	2.1	0.0	2.1	6.0
<b>IRF</b>	<b>0.8</b>	<b>-</b>	<b>0.8</b>	<b>0.8</b>
Cross Language	16.3			20.0
WON	1.0			1.0
<i>Total</i>	24.0			60.8

#### **2.5.10 Summary**

A comprehensive suite of evaluations – both automatic and human – have been carried out for the 6 language pairs (12 language directions) developed over the 3 years of the project. In addition to these evaluations, more detailed evaluations for selected language pairs have been carried out in the context of the TM/MT integration scenario, translator productivity, and translation usability. We also engaged WON to take part in two user surveys that have helped guide the software developments and commercialisation efforts. All in all, these

evaluations have provided us with sufficient confirmation that our translation technology – which forms the core of any and all commercialisation efforts – is of sufficient quality to compete in the market while also identifying areas where we can continue to improve and grow.

## 2.6 WP8 Dissemination

### 2.6.1 Objectives

The main aim of the dissemination work package is to maintain the visibility of the project activities while also building and nurturing relationships with potential users (and user groups) and other initiatives. In addition to this, collaborations should be initiated with appropriate parties and a dissemination strategy designed to support the exploitation and commercialisation plans.

D8.3	Dissemination activities (booth or workshop) at least 3 MT conferences and 3 IP events	<input checked="" type="checkbox"/>
------	--	-------------------------------------

Table 7 Milestones and deliverables due between M12 and M24

### 2.6.2 Progress Highlights

The main highlight of the dissemination work package has been the on-going and further increased engagement with users in the context of all of the various commercialisation options discussed. As previously, we have continued to focus on intellectual property (IP) related events in order to maintain awareness among stakeholders in the industry. To that end, PLuTO members have attended 7 events over the course of the third period with a further two events already scheduled beyond the end of the project. While diminished because of its more academic nature, we have also maintained a presence in the MT/localisation community in order to continue absorbing ideas and feedback from a scientific perspective.

In addition to travelling to events, coordinating partner DCU has hosted a number of high-profile guests including government ministers from Ireland and the UK. All parties were provided with an extended demonstration of the PLuTO technology and services and these engagements were publicised through various channels.

### 2.6.3 Tasks

#### T8.1 Training

In order to familiarise users and potential customers with the functionality of the IPTranslator and ParonPro applications, we have made use of three channels of communication:

- *Exhibition Booths:* This high-intimacy approach is the most effective method of ‘training’ users on the various services. It provides us the opportunity to have a dialogue and get immediate feedback and impressions. We have attended a number of events at which we have presented the respective products in this context. The

obvious drawback of this approach is the cost associated with travelling and reservation of the exhibition space.

- *Presentations:* PLuTO partners have given a number of presentations at a range of events helping to create awareness in the project, educate users as to the functionality, and build a reputation as experts and thought leaders in the (patent) translation space. These presentations persist online and can be promoted in the future.
- *Online:* In addition to our “on the road” activities, we take advantage of online channels to educate users through videos, presentations, publications, white papers, and blog posts. This approach has the advantage of being relatively low cost while having a potentially wide reach.

Full details on activities related to training are provided in Deliverable 8.3.

## T8.2 Dissemination

Over the lifetime of the project, conference and workshop attendance has been a central element of PLuTO’s dissemination activities. With the addition of the activities in year three, our dissemination targets were far surpassed with a total of eight MT conferences, three IR conferences and twelve IP conferences attended (against targets of four, one and four, respectively). Consortium members have already confirmed attendance at a number of upcoming events. These include the WON AGM, which will be held in Utrecht, The Netherlands in April 2013 and the PIUG 2013 Annual Conference, which will be held in Alexandria, USA in April 2013.

As discussed in Deliverable 8.2, our priorities were to begin to focus more on IP events as our software solutions mature. To this end, we have attended a number of events, some for the second time including the EPO’s Patent Information Conference (in Hamburg), the IPI-ConfEx (in Seville), and the INTA Conference (in the USA). Once again, our presence at these functions has been supplemented with more professional marketing materials such as pull-up banners, leaflets, and other handouts.

### 2.6.4 Use of Resources

Beneficiary	PMs to date	PMs yr3	PMs Total	PMs Overall
DCU	18.5			26.4
ESTeam	0.6	0.7	1.3	5.0
IRF	1.6	-	1.6	1.6
Cross Language	0.9			6.0
WON	1.0	1.0	2.0	2.0
<i>Total</i>	20.6			41

### 2.6.5 Summary

Initially there was a focus on MT events and, as the PLuTO MT technology became more developed, the consortium placed a greater emphasis on IP events. The nature of dissemination activities also shifted from event attendance to more hands-on demonstrations at exhibition booths and presentations on PLuTO technology. Participation in these events has generated good visibility and strong brand recognition for PLuTO, which place the consortium in a solid position for further commercial exploitation of project outcomes.

## 2.7 WP9 Exploitation and Standardisation

### 2.7.1 Objectives

The work package on Exploitation and Standardisation of PLuTO is charged with keeping the consortium in touch with current market trends, in terms of both technical and commercial developments, in the area of translation service provision tools, particularly as relates to patents. Additionally, a strategy will be developed to exploit the results of the project via the most appropriate channels to market. Deliverables and milestones falling due in the period are shown in Table 8.


D9.2/3	Exploitation Plan, final	
--------	--------------------------	---

Table 8 Milestones and deliverables due between M24 and M36

### 2.7.2 Progress Highlights

The main highlight in this work package in the final period has been the in depth market evaluation leading to clear strategies for the exploitation of the results of the project. We have identified three distinct products to be commercialised beyond the end of the community support:

- **IPTranslator** – Patent-specific Machine Translation launched from DCU. Enterprise Ireland has already economically funded its first year in operation.
- **ParonPro** – Patent research workbench launched from ESTeam. It will continue to be marketed as an ESTeam product.
- **TM/MT** – Initial steps have been made towards a new product development in the field of translation services where combination of MT and TM is realised.

We have further identified three distinct target markets in which there is a commercial opportunity for these products:

- **Patent Search Professionals:** Individuals and teams working in multinationals and patent search firms (including attorneys) who carry out large volumes of patent searches are frequently require gisting translation. *This market is a target for the ParonPro and IPTranslator online products.*
- **Patent Tool Vendors:** Companies selling patent search and related tools to patent search professionals. There exists the potential to sell a combination of on-demand and bulk translation to these vendors. *This market is a business to business (B2B) target for IPTranslator.*
- **Language Service Providers:** LSPs selling (or open to selling) post-edited machine translation (PEMT) for patents and potentially other domains. *This market is a target for the TM/MT offering and IPTranslator.*

Full details on this research in provided in Deliverable 9.2.

#### T9.1 Market Observation

We have carried out in-depth market studies, including a supplementary external report supported by Enterprise Ireland. We have noted a number of trends which indicate that a market opportunity exists for patent translation and related services across the three target markets:

- Growing demand for specialist translation
- Increased patent filing in Asian languages
- Increased need for translation in knowledge intensive industries
- Higher rates paid for IP translation
- Multi-nationals invest in IP translation
- Low penetration of technology at Language Service Providers

As this is a niche market, there is not a huge competitive landscape and the unique selling propositions of the PLuTO technology will help the technology stand out. The propositions include:

- 1) **Patent-specific translation (TM and MT)**  
Patent language is a particular challenge for automated translation and many solutions cannot be applied 'as is'. PLuTO translation services have been adapted to the domain over 3 years and evaluation show the quality to be at worst, competitive, and at best, far superior to the alternatives.
- 2) **Quick Start**  
PLuTO is already trained to its domain – in contrast to other TM or MT based solutions that need training or customization, PLuTO is usable “out-of-the-box” which means the service can be provided at a lower cost.
- 3) **Integrated Hybrid Technology**  
PLuTO is unique in that it technologically directly integrates machine translation and translation memory. While others – if at all – just forward “no match” TM searches to the machine translation system, the ESTeam based integration re-assembles MT translated segments and post processes them to generate a best result than a simple either TM or MT result. This can be integrated into the platforms of LSPs and patent tool vendors.
- 4) **Security**  
PLuTO can provide assurances over data confidentiality and security that is not offering by competing solutions.

## **T9.2 Exploitation and IPR Strategy**

All offerings are rolled out as a Software-as-a-Service (SaaS) hosted online solution, serving individuals as well as organisational customers (so-called *tenants*). This approach, pioneered by Salesforce.com, negates our need for physical or direct distribution of our products. And incurs a relatively low cost for user provisioning. It also allows us to be more flexible with our pricing models. For the customer perspective, this model reduces their need to invest in infrastructure to support their business needs which essentially boils down to a more cost effective solution.

The creation of an SME was planned from the early inception of the PLuTO project. This is not the case at this stage and instead the partners in the consortium will exploit the foreground IP through a number of means. Enterprise Ireland has financed John Tinsley at DCU for one year for going to market with the machine translation technology that underpins IPTranslator. This will involve the incorporation of a company that will license the PLuTO IP from the university for the purposes of commercialisation. ParonPro will be marketing as a product of ESTeam who have the capacity to support the ongoing development required to bring it to market readiness. The two companies will then be in a

position to discuss a strategy for bringing the TM/MT solution to the market through licensing or other options.

### 2.7.3 Use of Resources

<b>Beneficiary</b>	<b>PMs to date</b>	<b>PMs yr3</b>	<b>PMs Total</b>	<b>PMs Overall</b>
DCU	2.8			4.8
ESTeam	10.3	5.1	15.4	16.0
<b>IRF</b>	<b>0.2</b>	<b>-</b>	<b>0.2</b>	<b>0.2</b>
Cross Language	0	0	0	3.0
WON	0	0	0	0
<i>Total</i>	13.3			24

### 2.7.4 Summary

We have identified a number of potential channels to market with a number of different products developed over the course of the last three years in PLuTO. We believe that these developments and plans constitute a success for the outcomes of the project. Regardless of the ultimate outcome of these activities, the foundations laid with support from the European Commission will continue to be developed and will have their chance to compete in the market. That being said, the consortium and individuals undertaking this work (including the Irish government who have continued to support it) are cautiously optimistic about their prospect.

### 3 Deliverables and Milestones Tables

Table 1. Deliverables										
Del. no.	Deliverable name	WP no.	Lead participant	Nature	Dissemination level	Achieved	Due delivery date from Annex I	Delivered Yes/No	Actual / Forecast delivery date	Comments
1.1c	Annual Project Report	1	DCU	R	P	Yes	31/03/13	Yes	03/04/13	
3.3	Final Web App and UI	3	EST	O	P	Yes	31/03/13	Yes	03/04/13	Consolidates D3.1, D3.2, & D3.3
4.3	Final TM Resources	4	EST	R	P	Yes	31/03/13	Yes	03/04/13	Consolidates D4.3 & D4.4
5.3	Final MT Engines	5	DCU	O	P	Yes	31/03/13	Yes	03/04/13	Consolidates D5.3 & D5.4
6.3	Final System	6	EST	D	RE	Yes	31/03/13	Yes	03/04/13	
7.3	Final Survey Report	7	CL	R	P	Yes	31/03/13	Yes	03/04/13	
7.7	Final MT Quality	7	DCU/CL	R	P	Yes	31/03/13	Yes	03/04/13	
7.10	Final TM Impact	7	EST/CL	R	P	Yes	31/03/13	Yes	03/04/13	
8.2	Dissemination Activity	8	DCU	O	P	Yes	31/03/13	Yes	03/04/13	
9.2	Exploitation Plan	9	EST	R	CO	Yes	31/03/13	Yes	03/04/13	Consolidates D9.3 & D9.3
Table 2. Milestones										
Milestone no.	Milestone name	Due achievement date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments					
2.5	ES patent data available	30/06/12	Yes	30/06/12						
2.6	ZH patent data available	31/12/12	Yes	31/12/12						
3.3	Final Web App and UI	31/03/13	Yes	03/04/13						
4.5	TM resources for EN-ZH	30/09/12	Yes	03/04/13						
4.6	TM resources for EN-JP	03/04/13	Yes	03/04/13						
5.5	MT System for EN-ZH	30/09/12	Yes	13/04/12						
5.6	MT System for EN-ES	31/03/13	Yes	13/04/12						
6.4	Integrated System final	31/03/13	Yes	03/04/13						
7.3	Survey structure v2	30/09/12	Yes	30/09/12						



## **4 Project Management**

### **4.1 Management Tasks and Achievements**

In this section, we describe work carried out over the course of the period of which the project coordinators were directly responsible. Additionally, we present tasks performed and other achievements that were not specifically outlined in the Description of Work.

#### **4.1.1 Commercial Development**

While the majority of the exploitation work in the project is detailed in Deliverable 9.2, a significant amount of “extra-curricular” commercial development work was carried out in year three. It was vital that all partners in the consortium were in agreement and in a position to understand and act upon the opportunities for the PLuTO technology beyond the end of the community support.

In July 2012, DCU received additional funding from the Irish Government agency Enterprise Ireland which is responsible for supporting new business development. This was used to commission a study on the feasibility of translation technology and the IPTranslator application in the IP community. The results of this study were two reports, a ‘Go To Market’ plan for IPTranslator and an initial ‘Sales and Lead Generation’ plan to guide the process.

Following this, in January 2013, project coordinator John Tinsley was awarded significant support from the same agency under their ‘Commercialisation Fund’ to create a spin-out company from the university, hire two full time employees for a period of 12 months, and licence the IP from DCU for commercial exploitation. This funding will kick-in on April 2 immediately following the end of the PLuTO project. To support this, DCU has hired a business development manager on a short-term basis to accelerate some specific opportunities in the lead up to the end of the project.

#### **4.1.2 User Engagement**

As we detail further in the upcoming section 4.2, engagement with WON has been on-going over the third year and it is expected that Aalt van de Kuilen will continue to work with the partners in an informal advisory role beyond the end of the project. The first step in this process will be a presentation at the WON AGM in Utrecht on April 18<sup>th</sup> immediately following the final review to wrap up the project for the wider WON community and introduce them to the follow-up developments.

We have also established relations with the Patent Information User Group, including president Susan Hantos. This has led to an invitation to present at their annual conference in Alexandria in May 2012. John Tinsley will travel under the banner of the Enterprise Ireland engagement to present a tutorial entitled “Understanding Machine Translation and the Challenge of Patents”.

#### **Commercial Relations**

In addition to our engagements with user groups, we have been networking extensively over the course of the third year of the project laying the foundations for a number of commercial engagements following the end of the project. In the context of end-user patent information specialists, we have had over 200 users subscribe to our online applications.

In the context of business to business relations, we are in advanced discussions with a number of patent service vendors with a view to engaging with them as service providers. We are also in touch with a number of patent offices with regards to upcoming tenders. Finally, in terms of the translation and language service provider (LSP) market, we are currently in the evaluation stage with a leading IP LSP with a viewing to on-going commercial collaboration.

### **4.1.3 Advisory Board**

**Members:** Fred Hollowood; Viggo Hansen; Greg Grefenstette; Stephen Adams.

Our good relationship with the advisory board continued through the final year of the project with even more interaction between the advisors and the partners. DCU was in constant contact with Fred Hollowood as he visits the CNGL office on a weekly basis, as well as the LRC Conference. Stephen Adams was in attendance at most of the other IP related events attended by PLuTO participants including the EPOPIC and the IPI ConfEx. Finally, the highly agreeable nature of the interaction with the advisory board is illustrated by the factor that Greg Grefenstette hosting our final AGM and AB meeting at his offices in Paris.

### **4.1.4 Recruitment**

The loss of key staff was highlighted in our risk management plan as one of the biggest threats to the project. This risk was heightened in the final year of the project, particularly for DCU, as the capacity to replace people would be hampered by the short-term nature of the engagement. This risk came to fruition when Alexandru Ceasusu, one only 3 full-time members of the project, moved on from DCU in October 2012. DCU reacted swiftly to replace Alexandru with two short-term interns to see out the MT development work in the project. This worked out well in the end as the replacements were tailored to suit some of the open issues remaining in WP5, e.g. a Japanese intern was hired to continue work on the transliteration experiments (see Deliverable 5.3).

## **4.2 Project Meetings**

### **4.2.1 Second Year Review Meeting – Ghent – May 15<sup>th</sup> 2012**

The second year review meeting was held at the CrossLang offices in the month following the AGM. A half-day preparatory meeting was held the day preceding the review and a further debriefing meeting was held the following day in order to make preliminary plans going forward.

**Attendees:** DCU 3; Cross Language 3; ESTeam 4; WON 1

### **4.2.2 User Group Meeting – Venlo – June 15<sup>th</sup> 2012**

A meeting with the WON working group was scheduled in the wake of the second year review meeting. Facilitated by WON member Ap van der Mark at the Océ offices in Venlo, the main topic for discussion at the meeting was to define a set of features for the ParonPro application. The resulting report from this meeting was used as a 'product roadmap' to support on-going developments.

**Attendees:** ESTeam 1; WON 2; Cross Language 2

### **4.2.3 Management Meeting – Hamburg – November 6-7<sup>th</sup> 2012**

Senior members from each partner were in attendance at the EPO's Patent Information Conference in Hamburg. A number of meetings were held over the course of the two days with discussions focusing predominantly on exploitation and commercialisation plans. Additionally, a number of logistical and administrative matters were dealt with.

**Attendees:** ESTeam 3; DCU 2; Cross Language 1; WON 1.

### **4.2.4 Evaluation Meeting – Ghent – ? 2012**

An extraordinary meeting was held towards the end of the year with specific focus on the evaluation of the integrated TM/MT scenario. As the main recommendations coming from the second year review related to this task, and the possibility to carry out customer-specific TM integration became greater, it was decided that an intensive working meeting would be the most effective way to address the issue.

**Attendees:** ESTeam 1; Cross Language 2.

### **4.2.5 User Group and Technical Meeting – Ghent – 15/16<sup>th</sup> January 2013**

A two-day meeting was held at the beginning of the year with dual purposes. Firstly, representative from each partner were in attendance to finalise plans for the completion of work and deliverables leading into the final AGM and review. Secondly, a day of user meetings were held with two external parties (a patent attorney and a representative from the Flanders Innovation Board) and Aalt van de Kuilen from WON to discuss commercialisation plans and our final user survey.

**Attendees:** ESTeam 1; Cross Language 2; DCU 1; WON 1.

### **4.2.6 AGM – Paris – April 6-7<sup>th</sup> 2012**

The third and final PLuTO annual general meeting was hosted by advisory board member Greg Greffenstette at the Exalead offices in Paris shortly before the official end of project. This time was used to review deliverables due at month 36 and use the advisory board as a sounding board for our commercialisation and exploitation plans.

Like last year, the second day was devoted to a dry-run of presentations for the actual review with the project's advisory board. This was again a very fruitful exercise during which the advisors expressed their pleasure at the progress of the project in the preceding year.

**Attendees:** ESTeam 3; DCU 2; Cross Language 1; Advisory Board 4

## **4.3 Dissemination Activities**

PLuTO members have attended a number of events over the course of the second year of the project. As mentioned in last year's annual report and again in Deliverable 8.2 on Dissemination Activities, we have stepped up our efforts to reach out to more users by attending more relevant events. These promotional efforts have been supported through various other channels such as physical and social media, websites and blogs, as well as customer relations management. Full details on the events attended are provided in

Deliverable 8.3 while in the next section we summarise some of the other activities related to dissemination.

### 4.3.1 Dissemination Materials and Social Media

In order to support the promotion of the project and present a professional look and feel, a number of marketing items have been designed and distributed. New items produced in year 3 include a first trifold leaflet for ParonPro and an improved effort for IPTranslator. Additionally, an infographic has been developed based on the results of the user survey first published in Deliverable 7.2. This is designed to demonstrate our capacity to provide knowledge and information to the IP community. Finally, in order to increase the visual impact of our exhibitions at events, we produced an IPTranslator branded tablecloth as shown in the image below.

Our social media presence has been enhanced with the addition of a dedicated @IPTranslator Twitter account and our combined tweets total almost 300 over the course of the last year which had drawn significant traffic to the sites. While the dedicated project website [www.pluto-patenttranslation.eu](http://www.pluto-patenttranslation.eu) is not updated as frequently as previously, this is counter-balanced by the frequent activity on the IPTranslator site, including the blog at [www.iptranslator.com/blog](http://www.iptranslator.com/blog) and the new ParonPro site [www.paronpro.com](http://www.paronpro.com).

## 5 Summary

We have documented the progress made in both technical and administrative aspects of the PLuTO project during the second reporting period. Key areas have been highlighted, such as the maturing of the application interfaces, the promising translation evaluation results particularly in terms of the TM/MT integration scenario, and the dissemination and exploitation developments.

Looking forward, the future looks bright in terms of the exploitation of the fruits of the project with partners already committed to on-going work and the potential for further collaboration between members of the consortium on related activities.

In conclusion, the partners are extremely pleased with the outcome of the project and our capacity to collaborate despite the number road blocks in our path along the way. We ultimately believe the fruits of PLuTO will continue to be seen in the IP and language technology sectors over the coming years.

