PROGRESS REPORT

§3 Project Progress

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

Periodic report: 2nd

Period covered: **from** 01/04/2011 **to** 31/03/2012

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)

Executive Abstract

In this report, we describe the work carried out over the second 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 period. For each work package, we present the main goals, highlights from the period, milestones and deliverables achieved, and plans for the final year. 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 plans for the coming period.

An overall highlight of the project has been the general progression in work from the first year in key areas. Data has been provided for new languages, machine translation (MT) systems have been built and translation memory (TM) resources have been created. The integration of MT and TM has been significantly developed in this period and comprehensive evaluations of translation quality, translation adequacy, and software usability have also been carried out. These activities have played an important part in the production of two user interfaces – IPTranslator and ParonPro – which form the basis of our exploitation strategy.

Another key highlight from the period has been the interaction between the consortium and users. Meetings have been held on a bimonthly basis with our partners at WON who have assisted us greatly in a number of areas including the preparation of usability experiments and the distribution of our user survey. Increased efforts have also been made to engage the wider intellectual property community and this is well reflected in our dissemination activities.

Solid foundations have been put in place to facilitate commercialisation of the software and services being developed in PLuTO. The launch of IPTranslator.com has helped to create a group of real users who we can exploit to learn what makes them tick and how to iterate our products to find product-market fit.

Our plans for the final year will be heavily driven by the feedback from users and our analysis of the behaviour and habits when using our services. From a technical perspective, a focus will be placed on improving translation quality, both of MT and the integrated system.

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.

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.

These sub-tasks include the addition of further languages pairs, improvement of translation quality through tighter integration of TMs and MT, together with detailed evaluations of translation quality. Additionally, the development of all project related software is driven by interaction with users – both within the consortium and externally – through the various dissemination activities and by the initial findings from research on potential channels for commercial exploitation.

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 (EN <-> DE, ES);
- The creation of TM resources for the above language pairs;
- The provision of relevant patent corpora to support the aforementioned tasks;
- The continued development of web applications and interfaces to support the potential channels to market identified by the work package on exploitation;
- A full evaluation of all relevant components including machine translation quality and adequacy, impact of integrating TMs on translation quality , and usability of the web applications;
- Provision of results of the user survey on the needs and requirements of patent professionals;
- Continued dissemination of the project including building awareness and attending industry events;
- Further exploration of commercialisation options based on increased interaction with users and development of more mature software solutions.

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 first period. Below is a summary of those recommendations along with an indication of how they have been taken into account for this period.

- Assess the viability of a “MT only” solution
  - ⇒ **Action:** Preparation of a feasibility study on this aspect, incorporated into D9.1 Exploitation Plan at M18 and deployment of IPTranslator which positions MT as the main selling point.
- Develop evaluation methods to assess the quality and impact of TMs
  - ⇒ **Action:** A new deliverable D7.9 reporting on the impact of TMs in the integrated translation system was produced at M24.
- Revise the Description of Work and reallocate resources to reflect changes in the consortium and project focus
  - ⇒ **Action:** Revised DoW was submitted to the European Commission at M16.
- Reassess project risks given the revised plans
  - ⇒ **Action:** Revised risk assessment document was submitted as a deliverable at M18.
- Increase involvement with users and consider options for bringing in an additional partner
  - ⇒ **Action:** User engagement increased significantly in year 2 whereby consortium members met with representatives from WON and more than 100 external individuals signed up to interact with the consortium.

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.

Mi2.3	(EN, PT, FR, DE) patent data available	☑
Mi2.4	(EN, PT, FR, DE, ES) patent data available	☑
D2.2	Data Corpora and Standards, v2	☑

Table 1 Milestones and deliverables due between M12 and M24

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. To this end, data has been provided for English—German and English—Spanish from two distinct sources: the IRF's MAREC corpus and the Alexandria data collection (see Deliverable 2.1 for a more detailed description of these collections).

Furthermore, following the first year review it was recommended that the consortium be given greater flexibility with regards to language selection in order to be able to react to market demands. To that end, the consortium has exploited this recommendation by sourcing data for two additional strategic languages: Japanese and Chinese. This was motivated by the findings of Deliverable 7.2 – the results of the user survey – which suggested that these were the top two languages for which translation was required by patent professionals. These findings were validated through numerous discussions with users at IP events.

The Chinese data was acquired through the IRF while the Japanese data was acquired through previous participation in the NTCIR[1] Patent Mining Task. More information can be found on these data sets in Deliverable 2.2 Data Corpora and Standards v2.

[1] http://research.nii.ac.jp/ntcir/index-en.html

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 previously in Deliverable 1.1a. These are *family-id*, *lang*, *kind*, and *alignment*. More information on the mark-up and metadata for the new data sets is provided in Deliverable 2.2.

T2.2 Selection Engine

This task has been made obsolete by the revisions to the Description of Work.

T2.3 Data Acquisition

The consortium has gone above and beyond its obligations for the period and acquired parallel patent data for four language pairs over the course of the second year of the project: English—German (DE), Spanish (ES), Japanese (JP), and Chinese (ZH). The motivation behind this effort was to react to the demands of users in by increasing translation coverage in order to attract increased usage of our early software releases. The data for these four language pairs has come from multiple sources.

Firstly, the EN—DE data was provided as part of the MAREC collection. This is a unified collection of patent documents provided by the IRF that is represented in XML format following a standard document type definition (DTD). Data for the EN—ES pair was extracted from the Alexandria collection which contains thousands of comparable patent documents in XML format. The EN—ZH corpus, obtained from the IRF, was created manually by searching for English language documents in the same family as an original Chinese document and using various techniques to align corresponding sections at the sentence level. Finally, the EN—JP from the NTCIR is a collection of aligned patent abstracts from the Japanese and US patent offices.

In terms of data acquisition for year three, our plan is to target two more of the top five languages requested by users: Russian and Korean. Our first port of call will be the Alexandria corpus where we will attempt to extract comparable sections in the same way we did for Spanish. Following this, we will consider turning our attention to the languages of patent offices that are becoming increasingly IP active, for instance Italy, Taiwan, and Turkey.

Further details on the content of the respective corpora, their metadata, and the standards to which they comply can be found in Deliverable 2.2.

2.1.4 Use of Resources

Beneficiary	PMs yr1	PMs yr2	PMs Total	PMs Planned 3yr
DCU	0.4	0.8	1.2	8.0
ESTeam	2.0	2.7	4.7	4.8
IRF	4.2	-	4.2	4.2
Cross Language	0	0	0	3
WON	0	0	0	0
Total	6.6	3.5	10.1	20

2.1.5 Summary

We have presented the patent corpora which have been compiled to date for use in the major technical components of the PLuTO system. Additional data has been acquired beyond the two required language pairs to cater for user demand. Plans have been put in place for data acquisition in the final year of the project.

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 period are shown in Table 2.

Mi3.2	Web application and user interface v2	☑
D3.2	Web application and user interface v2	☑

Table 2 Milestones and deliverables due between M12 and M24

2.2.2 Progress Highlights

Through our research in WPs 8 and 9 into potential opportunities for commercialisation of patent translation web services, we identified a number of distinct user scenarios: on-demand MT for patent searchers, and a patent researcher's workbench (see Section 2.8. for more information). In order to test these hypotheses independently, we developed two distinct user interfaces hereafter known as **IPTranslator** and **ParonPro** respectively.

IPTranslator

The IPTranslator interface allows the end-user – in this case a patent searcher using web-based search tools to find relevant documents – to access patent translation on-the-fly by means of a web browser plugin. Following translation, the user can access further features including bilingual keyword extraction, segment highlighting, and a post-editing interface. A very first version of this interface was presented at the year one review meeting. It has since been made available to number of users as private beta in October 2011 and was subsequently released publically at IPTranslator.com in March 2012.

ParonPro

The ParonPro interface has evolved from the first PLuTO prototype presented in Deliverable 6.1 with the aim of providing patent professionals with a range of tools to support their everyday needs such as search and translation management as well as access to translation services. This interface also provides the entry point to the integrated translation system combining TMs and MT. ParonPro will be released for close beta testing in early-June 2012.

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 and statistics/logs.

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 landing page at IPTranslator.com gives the user the user three important options: creating a new account, login, and download the browser plugin. The translation interface can be accessed directly by logging in order by requesting a web translation through the plugin. Should the user login directly, the translation services can be accessed by uploading a file for translation or by typing or copy/pasting text.

Once logged into the interface, the user can access various menu items available, such as 'My Account', 'My Terminology', and 'My History'. These features and more, including detailed information about functionality, are described further in Milestone Report 3.2.

ParonPro

The ParonPro interface represents a slicker, more user friendly version of the prototype of Deliverable 6.1. It is designed as a collaborative environment in which patent professionals can import, organise, and share relevant data with colleagues.

Users can import patent data from search tools, collect them into sets, and share them via an email interface. Should translation of a particular document be required, the user has access to the PLuTO translation systems also (which on the back-end are the same systems used by IPTranslator). More details on the feature set, existing and proposed, are given in Milestone Report 3.2.

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 yr1	PMs yr2	PMs Total	PMs Planned 3yr
DCU	1.0	1.8	2.8	10.0
ESTeam	15.3	14.9	30.2	40.5
IRF	2.5	-	2.5	2.5
Cross Language	0	0	0	12
WON	0	0	0	1
Total	18.8	16.7	35.5	66

2.2.5 Summary

Two distinct user interfaces and applications have been developed to support the commercial objectives of the project. IPTranslator positions translation as the main selling point, whereas ParonPro positions translation as an added-value feature. Both interfaces exploit the same translation technology on the back-end.

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 M12 and

Mi4.3	TM resources for EN-DE	☑
Mi4.4	TM resources for EN-ES	☑
D4.2	TM resources for 4 language pairs	☑

Table 3 Milestones and deliverables due between M12 and M24

2.3.2 Progress Highlights

Significant efforts were made in this period to improve the quality of the TMs from year. This required investigation into methods for overcoming the lack of natural repetition in patents. Improving "quality" in the TMs essentially boils down to creating better alignments between bilingual pairs. In order to do this, a number of modules were developments specifically, including a formula tagger and a text segmentor. Applying these prior to alignment gave rise to an absolute improvement of 28% (from 5% to 32%) for the TMs hit rate leverage.

These techniques were applied to rebuild the TMs from year 1 for EN—PT and EN—FR. Additionally, new TMs for EN—DE and EN—ES were created based on the new data supplied as described in section 2.1.

2.3.3 Tasks

T4.1 Data Management

The low TM leverage achieved in year 1 was caused by various attributes of patents which made them less suitable for matching, such as non-repetitiveness and the use of named-entities and formulae amongst others. In order to better prepare the raw data for alignment, a number of processes were developed.

Firstly, a formula tagger was developed to *a priori* identify entities like chemical formulae and "neutralise" them in order to reduce variability and increase the likelihood of getting TM matches. Unlike the statistical approach to named-entity recognition for MT (described below in section 2.4), the formula tagger works using heuristic rules derived from manual inspection of the data. These rules use a combination of dictionary lookup, matching of special characters and more to identify relevant segments.

Additionally, text segmentation was improved by loosening the restrictions on sub-segment segmenters in order to find a better balance between how much the text is split and the levels of repetition. A list of segmenters was extracted through analysis of patent text with a view to being improved via "Dynamic Segmentation" in the future (see Deliverable 6.2).

T4.2 Structuring TM domains according to patent data domains

Similar to the data for Portuguese and French from year 1, the German data was structured according to the International Patent Classification (IPC) system as described in Deliverable 4.2. Further distinctions were made on the data according to whether segments originated in patent abstracts, claims, descriptions. This is relevant as the quantity of text (including sentences) in a document can have an effect on how difficult the task of alignment is.

For the Spanish data it was not possible to structure the data according to the IPC code as this data was not available from the original corpus.

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.

Equivalent portions of text are aligned and this alignment is improved by using the formula tagger to treat these entities as wildcards. A by-product of this approach is the creation of a bilingual resource of translated formulae which is added to the TM independently. Alignments with a score above a predefined threshold[2] of 80% are kept while those falling below this target are discarded.

T4.4 Data loading and quality control

The 80% threshold was introduced to control the quality of accepted alignments on one level. However, a manual analysis would allow for the identification of error patterns which could be systematically corrected using rules. While an exhaustive manual analysis in this regard is unrealistic, a manual spot-checking is planned for year 3 in order to further tune the alignments and threshold.

As mentioned previously, a 28% absolute improvement in TM hits was achieved on a test set of 8,000 sentences for English—French. Full details on this evaluation are described in Deliverable 4.2 while the correlation of this improvement in leverage with translation quality in the integrated system is described in Deliverable 7.9.

2.3.4 Use of Resources

Beneficiary	PMs yr1	PMs yr2	PMs Total	PMs Planned 3yr
DCU	0.3	0.0	0.3	18.0
ESTeam	12.5	13	25.5	41.0
IRF	0	-	0	0
Cross Language	0	0	0	0

[2] The derivation of this score is described in Deliverable 4.1

WON	0	0	0	0
Total	12.8	13	25.8	59

2.3.5 Summary

A number of techniques have been developed to improve the quality of translation memory resources for patents, including formula tagging and text segmentation. These improvements have been retrospectively applied to the EN—PT and EN—FR TMs from year 1, while new TMs have been created in this way for EN--DE and EN—ES. These changes have improve the TM leverage for patents by 28% (absolute) and evaluations have taken place to investigate how this improvement correlates with translation quality.

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 our investigations and findings of year 1 to further optimise the systems for patent translation. Deliverables and milestones falling due in the period are shown in Table 5.

Mi5.3	MT engine for EN-DE	☑
Mi5.4	MT engine for EN-ES	☑
D5.2	MT engine for 4 language pairs	☑

Table 4 Milestones and deliverables due between M12 and M24

2.4.2 Progress Highlights

In the context of this work package, the main achievement has been the development of Machine Translation engines for the two prescribed language pairs: English—German and English—Spanish. In addition to this, a further two MT systems have been developed for English—Japanese and English—Chinese. This was done in reaction to the findings of our user survey, described fully in Deliverable 7.2, which suggested that MT technology was most in demand for these two languages, particularly given the growing influence of Asian patents on the intellectual property (IP) landscape.

The EN—DE and EN—ES have been exposed for integration with the TMs while these and the EN—JP system have been released publically through the IPTranslator interface. The baseline EN—ZH system has just recently been built and performance is not yet sufficient to make it publically available.

A number of other techniques have been developed any deployed to help improve translation quality for specific language pairs and also for patent text. For example, a statistical named-entity recogniser has been developed to help identify certain constructs and handle them separately. Additionally, methods for truecasing, compound splitting and joining, preordering of input, and language identification have been developed and are discussed further in the next sections.

2.4.3 Tasks

T5.1 Adapting existing MT technology to the patent domain

The experiments from year 1 helped us to find a stable setup for our MT engines, including the combination of language and translation models, the application of processes such as sentence splitting, and the architecture for deploying the systems as a web service.

We subsequently sought to identify further ways in which we could improve translation quality as relates specifically to patents. In order to do this, we focussed on the feedback from the human evaluations carried out as part of Deliverable 7.6 which, amongst other issues, highlighted systematic error with specific entities such as chemical formula and similarly complex constructs.

The problem with such formulae was often that they were split into multiple tokens during the pre-processing stages when in fact they should have been treated as single tokens. To overcome this issue, we trained a named-entity recogniser using a manually annotated corpus of examples to mark up such instances. By doing this, we can treat them separately from the other tokens in a sentence and insure that the integrity of the construct is maintained.

In addition to this, as a support tool for our translation systems and the applications described further in Deliverables 6.2 and Milestone 3.2, we built a patent-specific language recogniser. It is a n-gram-based statistical model, trained using the full monolingual patent corpora described in Deliverable 2.2, which identifies the language of the input text in order to send it to the relevant translation server thus fully automating the engine selection process[5].

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 some of the techniques we have applied to improve translation quality for specific language pairs.

Truecasing

Case can be an issue for MT as the same token with different case, e.g. “phone” and “Phone” are seen as two distinct words and thus the statistics generated over the data are less reliable. Our previous approach to handling case was to simply lowercase all training data and input and reintroduce case as a post-process using statistical models training using conditional random fields. This was relatively straightforward as it was mainly sentence initial characters are proper names that needed case reintroduced.

However, this task is more complicated for a language like German where all nouns are capitalised. In this case, we employ truecasing whereby we only lowercase words which do not typically have upper case characters, e.g. sentence-initial words. All other words with upper case characters are kept as is for training. Case is the reintroduced where necessary during post-processing using a small number of rules.

[5] As described elsewhere, users of the applications will have already set their default target language.

Compound Splitting

As with casing, compounding is an issue for MT as there is an almost endless number of ways in which words can be compounded which means that most of them will not have been seen previously by the translation system. In order to handle this, compounds are identified in the training data and split into their constituent parts. This is done by first selecting the compounds, i.e. those tokens which do not appear in the original German vocabulary, splitting them into all possible variations, and selecting the most likely split based on unigram frequency of the individual segments.

Following translation into German, some compounds are recreated. However, this task is not as crucial as it does not affect the grammaticality or readability of the output.

Input Pre-ordering

With certain language pairs, just as English—Japanese and English—German the word order can be substantially different. For instance, in both Japanese and German we can have verb final sentences. This long distance movement of terms is problematic for MT as it is difficult to capture using statistical approaches.

To counteract this, we have begun preliminary investigations into so-called pre-ordering of the input text to be translated in order to make it align better with the target language. This involves using dependency information to move words around in the source text, e.g. moving a verb to the end of an English sentence prior to translation into German. Early results using this technique are promising for English—Japanese but further experimentation is required before be deploy it in our online systems.

T5.3 Integration between MT and TM

As described in Deliverables 5.1 and 6.1, 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. For example, the payload now includes word alignment information between the source and target text which is used during the TM alignment phase to better identify corresponding pairs. Incidentally, this information is also used in the IPTranslator interface to facilitate the segment highlighting feature (presented in Milestone 3.2).

2.4.4 Use of Resources

Beneficiary	PMs yr1	PMs yr2	PMs Total	PMs Planned 3yr
DCU	22.1	25.3	47.4	41.0
ESTeam	6.5	6	12.5	17.6
IRF	0	-	0	0
Cross Language	0	0	0	0
WON	0	0	0	0
Total	28.6	31.3	59.9	58.6

2.4.5 Summary

New machine translation systems have been developed for four additional languages pairs, three of which have been deployed publically and two of which have been made available for TM integration. A number of approaches have been described to improve the specific processing of patent text while further language (pair) specific techniques have been employed to improve overall translation quality. Full details on the work described in this section are given in Deliverable 5.2 while evaluations of translation quality can be found in Deliverable 7.6.

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.

Mi6.3	Integrated System v1	☑
D6.2	Second prototype	☑

Table 5 Milestones and deliverables due between M12 and M24

2.5.2 Progress Highlights

To a large extent, highlights in this work package reflect main achievements in the project as a whole in this period; that is to say the consortium's positive response and agility when it comes to change, be it due to circumstances beyond our control or a reaction to new findings.

In the first instance this can be seen in the redoubled efforts on TM and MT integration following the assessment of the project progress at the first review. We now have a scenario in which the improved TMs exploit the MT system in a novel way based on different levels of text segmentation and alignment, with plenty of scope for improvement in year 3.

Secondly, having attracted the attention of users in need of on-demand patent translation and based on the reviewer's assessment that the feasibility of an 'MT-only' scenario should be investigated, we have produced the IPTranslator tool; a production level translation service which has been released publically and is undergoing constant development using lean techniques based on extensive user feedback.

Finally, with the revision of the work plan and the removal of focus on the search aspect, the year 1 prototype was reimagined and redeveloped into what is now the ParonPro interface. This application serves as a workbench and research tool to cater for the collaborative needs of patent professionals. Amongst the added functionality in this application is a translation option which includes the integrate MT/TM framework.

Both IPTranslator and ParonPro applications are online, available for testing, and will be presented at the year 2 review. Test cases and a user experience report have also been produced for these systems at M24 in the context of this work package.

2.5.3 Tasks

For both tasks in this work package, integration analysis and generation, there are some elements specific to the IPTranslator application, some specific to ParonPro, and some overlapping aspects. In the following, we will identify these where appropriate.

T6.1 Integration Requirements Analysis

Much of the work on assessing the various integration requirements was done in year 1. This includes aspects such as general technical architecture, dependents, I/O formats, etc. However, over the course of year 2, we identified a number of features which would require additional functionality which needed to need assessed accordingly. For the TM/MT integration, a more efficient way of calling the MT system was required along with supplementary information from the MT API such as word alignment data. This would afford the TMs with requisite information to improve alignment information and carry out dynamic segmentation, while at the same time providing IPTranslator information to allow for translation editing and segment highlighting. Furthermore, both IPTranslator and ParonPro required methods through which users could “import” the patent documents they need translated.

Once identified, this functionality was implemented as touched upon below and described in detail in Deliverable 6.2.

T6.3 Integration Prototype Generation

In order to expose word alignment information in the API, the requisite changes were implemented and MT systems were recompiled. Exposing this information caused the systems to require additional space in memory and thus steps were also taken to reduce this. This allowed the integrated system to only have to make a single call to the MT systems (and not multiple calls as was the case previously, which was inefficient) and exploit the alignment information to extract relevant segments.

For IPTranslator, the principal source of patent data for translation is from patent search tools, which are in most instances web-based. In order to allow the end user to send this data to the translation service, a browser plugin was implemented which identifies relevant patent text for a given search result in all major patent search tools[6], extracts this text from the HTML source, and sends it to the translation service. A full description of the functionality is described in Milestone 3.2.

ParonPro imports patent data from web-based search tools into the user’s account using bookmarklet technology. Similar to the browser plugin, but without the need for installation, the importer captures text from a website behind the scenes and stores in a database. The next time the user visits their ParonPro account, the imported data is available to manipulate.

2.5.4 Use of Resources

Beneficiary	PMs yr1	PMs yr2	PMs Total	PMs Planned 3yr
DCU	3.0	4.3	7.3	13.0
ESTeam	14.2	14.4	28.6	36.0

[6] The results of the user survey in Deliverable 7.2 helped us identify the main search tools used by patent professionals. We were made aware of other tools through interaction at IP events, both end users and search providers.

IRF	3.1	-	3.1	3.1
Cross Language	1.5	0	0	1.5
WON	0	0	0	0
Total	21.8	18.7	40.5	53.6

2.5.5 Summary

Two applications, IPTranslator and ParonPro, have been built on top of the translation services provided by the machine translation and integrated translation systems. A large portion of this back-end development was carried out in the course of year 1 and work in this period consolidated these efforts, analysed and implemented changes where appropriate to improve the delivery of the services, and iterated in reaction to various factors such as changes in the work plan, demands of end-users, and recommendations from the year 1 review. Full details of the work presented in this section are given in Deliverable 6.2.

2.6 WP7 Evaluation and Quality Assurance

2.6.1 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 – and user experience, but also by engaging WON and other users on the overall usability of the system taking the individual components into account. Deliverables and milestones falling due in the period are shown in Table 6.

D7.2	Publish report on results of user survey	☑
D7.6	Report on the intrinsic and extrinsic quality of MT	☑
D7.9	Report on the impact of TMs in the integrated translation system	☑

Table 6 Milestones and deliverables due between M12 and M24

2.6.2 Progress Highlights

The main highlights to date in the context of this work package have been the comprehensive evaluation of the English—French and English—Portuguese machine translation engines. In addition to evaluation using automatic metrics on the level of IPC class, translation quality has been evaluated by three expert evaluators for each language *direction*[7] in terms of adequacy, a blind ranking task to benchmark the systems against Google and Systran, and an error analysis on the output. On top of this, a usability test was designed and carried out with professional patent searchers to quantify how useful translation output was for the tasks required of such individuals. These evaluations are

[7] This constitutes four language pairs and, thus, 12 evaluators in total.

described in greater detail in Deliverable 7.6. Furthermore, preliminary evaluations have already been carried out for the English—German and English—Japanese MT systems. These results will be presented at the year 2 review.

Following the recreation of the TMs using the new techniques described previously, the impact of the TMs on baseline MT quality was also evaluated for French to English translation. This was done as a blind ranking experiment where the evaluator would select the best of two outputs; one being MT and the other the integrated system. The results of this evaluation are published in Deliverable 7.9.

The results of the user survey published as Deliverable 7.1 were collected and analysed. The findings provided invaluable information to the consortium on the needs of patent professionals such as the languages for which they require translation most, the search tools they use to retrieve patent documents, and the solutions they currently use to deal with patent translation, amongst others. These results are given in full in Deliverable 7.2.

Finally, in addition to having released the IPTranslator service publically and attracting feedback via that means, a user experience evaluation was also carried out on the interface and functionality. This served to not only identify issues like bugs and typos, but also design and layout features which could be improved. The results of this evaluation are actually published under work package 6's Milestone 6.3.

2.6.3 Tasks

T7.1 Usability and utility to patent searchers

The specific objective of usability evaluation is to ensure that the overall applications meet the needs of potential end users. MT usability evaluation is mainly user centred and takes into account use cases of translated text, which goes beyond the classical approach in MT evaluation. This includes a simulation of typical user tasks with translated text. Through consultation with the PLuTO working group at WON, as well as members of our advisory board, an experiment was designed to evaluate the usability of the MT output as relates to the job of a patent searcher.

To describe their job briefly, patent searchers (or attorneys or IP specialists) will typically have an invention at hand for which they will need to carry out some type of search, e.g. infringement, freedom to operate, etc., and make a judgement as to whether the results are relevant to them or not. Often times, many of the search results (patents) will be in a foreign language and thus the searcher will need translation in order for them to decide whether the patent needs to be looked at in greater detail.

The experiment was delivered to a number of users via an online interface. The results, presented in full in Deliverable 7.6, are generally positive where enough respondents were available. In other cases, the results were inconclusive.

In addition to the translation usability experiment, usability/user experience tests were carried out on the IPTranslator interface. These cross-browser tests checked a number of aspects of the web service including layout/design and functionality. The results, while broadly positive, highlighted a number of areas in which the overall experience of the user could be improved. Incidentally, a number of the findings correlated with some of the informal feedback we received from actual users. The results of this are published in Milestone Report 6.3.

T7.3 Translation Evaluation

A range of tests were carried out to evaluate the performance of the English—Portuguese and English—French machine translation (MT) systems submitted in Deliverable 5.1. 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. For the automatic evaluations, test sets 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.

Both automatic and human evaluations have shown that the PLuTO engines produce translations of a reasonable to good quality. The output of the PLuTO engines was preferred by all evaluators for all language pairs over that of Google Translate and Systran.

Further analysis revealed that there are quality differences across languages and IPC domains. Full details on these evaluations are given in Deliverable 7.6. At the review meeting for year 2, preliminary results will be presented for the English—German and English—Japanese language pairs where available.

2.6.4 Use of Resources

Beneficiary	PMs yr1	PMs yr2	PMs Total	PMs Planned 3yr
DCU	2.7	1.1	3.8	33.0
ESTeam	0.1	2	2.1	6.0
IRF	0.8	-	0.8	0.8
Cross Language	2.2	14.1	16.3	20.0
WON	0	1.0	1.0	1.0
Total	5.8	18.2	24	60.8

2.6.5 Summary

A comprehensive automatic a human evaluation for translation quality for English—French and English—Portuguese has been carried out which indicates the MT systems are performing to an acceptable level. The impact of translation memories on overall quality has also been assessed. Finally, there has been significant engagement with users to gather survey results to help us better define the level of service which needs to be provided by our applications.

2.7 *WP8 Dissemination*

2.7.1 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.

Mi8.2	Dissemination and Training Plan	☑
D8.2	Report on Dissemination Activities	☑

Table 7 Milestones and deliverables due between M12 and M24

2.7.2 Progress Highlights

The main highlight of the dissemination work package has been the increased engagement with users over the course of the period and subsequent activities initiated because of this. We decided to focus on attending intellectual property (IP) related events in order to build awareness of the project among potential user/customer communities. This lead to more than 100 people signing up via our website to be kept up to date with project related goings on. We also built ties at these events with potential collaborators and customers on the business side, e.g. Thomson Reuters and Minesoft, and on the end-user side, e.g. Panasonic and Unilever.

In addition to travelling to events, a number of initiatives were undertaken to build interest and awareness in the project. While the project website is still maintained and updated, the IPTransator.com interface also provides information, including a blog, which keeps users up to date on features and related news. We also maintain a social media presence through our Twitter account @plutopatents as well as many of our personal accounts. Finally, we have received some external press for our activities over the year in both print and online media which has served to boost our profile.

It should be noted that all of these achievements came following the reallocation of the dissemination workload, which was previously mainly the IRF's responsibility, amongst the remaining partners.

2.7.3 Tasks

T8.1 Training

In order to make potential users aware of the software we have been developing, as well as how to use it, we have had exhibition booths at a number of events including the European Patent Office's Patent Information Conference (EPOPIC) and the International Patent Information Conference and Exhibition (IPI-Confex). Here we have given hands on demonstration of our browser plugin tool at prototype, beta, and full release stages.

We have used various online tools to further facilitate our education of users. We have used the customer relations managements system Campaign Monitor to send out newsletters to people who have signed up for information. Our blog at http://plutopatents.wordpress.com contains (and will continue to be updated with) short instructional videos on various features, while the landing page at http://www.iptranslator.com contains an introduction video as well as a slideshow tour of the IPTranslator tool and its features. An older version of the video is also available at http://www.youtube.com/watch?v=3tBbVhtY1OU.

T8.2 Dissemination

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 events such as the WIPO Symposium of Intellectual Authorities, the INTA conference, the IPWare Summit, the WON AGM, and the Patent Information User Group (PIUG) Conference (in addition to the EPOPIC and IPI-Confex). Our presence at these functions has been supplemented with more professional marketing materials such as pull-up banners, leaflets, and other handouts. Electronic versions of these will be included in an appendix to this report where possible.

This more professional appearance is also reflected in our presentation of the IPTranslator service as a commercial product. We found there was little uptake by users of our beta versions with reduced feature sets as they did not sufficiently meet their needs, e.g. support their browser or a particular language for which they required translation. There was somewhat of a consensus that users would keep track of the project and eventually return once a product was ready. This was insufficient for us, however, as we need a constant feedback loop (either directly or through analytics) to assess what features are attracting attention and to know when to iterate or pivot when we have built something that is not needed.

We have also received press for our project activities in the past year. Our attendance at the EPOPIC, incidentally held in Kilkenny, Ireland, was featured in the Irish newspaper the Sunday Business Post (http://www.pluto-patenttranslation.eu/?q=node/82) while the same story was also covered by leading Irish technology website Silicon Republic (http://www.siliconrepublic.com/start-ups/item/24094-patent-translation-system). Finally, German scientific magazine DUZ also published a feature on the project in their July issue.

2.7.4 Use of Resources

Beneficiary	PMs yr1	PMs yr2	PMs Total	PMs Planned 3yr
DCU	3.0	13.5	18.5	26.4
ESTeam	0.4	0.2	0.6	5.0
IRF	1.6	-	1.6	1.6
Cross Language	0	0.9	0.9	6.0
WON	0	1.0	1.0	2.0
Total	5.0	15.6	20.6	41

2.7.5 Summary

Dissemination activities, to build up serious user and customer interest in support of our exploitation and commercialisation strategies, have been ramped up significantly in this period. We have built up a user base of over 100 individuals and have interacted with numerous potential collaborators. Our marketing and public appearance has taken on a more professional feel in order to attract actual users and afford us a better opportunity to ultimately covert them to paying customers.

2.8 WP9 Exploitation and Standardisation

2.8.1 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. Additional, 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.

D9.1	Exploitation Plan, v1 (incl. MT only feasibility)	☑

Table 8 Milestones and deliverables due between M12 and M24

2.8.2 Progress Highlights

The main highlight in this work package to date has been the validation from users that there is a potential market for patent translation and related services. Following the success of our collaboration with the EPO, the next signs of validation came in the report on the user survey results in Deliverable 7.2 in which 65% of respondents said they would be willing to pay for an adequate MT solution. Following significant interest in the project at various events, we then launched the IPTranslator.com product in order to capitalise on the momentum we had generated.

We have carried out an analysis of the competitive landscape for patent translation and translation services in general as well as a SWOT[8] analysis of our own capabilities. Aside from the MT only scenario (IPTranslator) aimed at patent searchers who require instant gist translations, we looked at additional channels to market such as business to business models with patent search providers and support tools for patent translators (as seen in ParonPro).

T9.1 Market Observation

Building up our early assessments presented in Milestone 9.1, our subsequent observation of market trends and behaviour has suggested to us there is a clear gap for a dedicated patent translation offering. For on-the-fly translation, Google Translate is widely used by individual end-users as well as being incorporated into the offerings of search vendors such as PatBase (Minesoft), Thomson Reuters, and Questel Orbit.

However, there remain significant concerns with regards to data security when it comes to Google (to the extent that we have spoken to users who are prohibited from using Google in-house). Furthermore, we have demonstrated improved translation quality over Google Translate. In this regard, we have already had high-level discussions with a number of parties with regards to directly incorporating our MT engines into their offerings as we did previously with the EPO.

T9.2 Exploitation and IPR Strategy

Going into the final year of the project, we are going to explore three distinct opportunities we have identified to bring the fruits of the project to market, some of which are more advanced than others at this juncture.

[8] Strengths, Weaknesses, Opportunities, and Threats

We will continue to evolve the channel on machine translation services for patent searchers, as implemented in IPTranslator.com. Our plan is to continue to iterate the software to build and solidify our current user base before beginning the process of converting them to paid customers.

We will look deeper into the ParonPro interface which we envisage as a potentially being a patent researchers workbench with a number of services sold as an add-on. One such service is the integrated TM/MT platform for translation production (beyond the level of gisting). We may also consider applying the techniques developed here to additional domains where the integration may be even more effective.

More details on our exploitation plans can be found in Milestone 9.1, which was submitted at M18, while updated information on activities since then will be presented during the review.

2.8.3 Use of Resources

Beneficiary	**PMs yr1**	**PMs yr2**	**PMs Total**	**PMs Planned 3yr**
DCU	0.4	2.4	2.8	4.8
ESTeam	2.1	8.2	10.3	16.0
IRF	0.2	-	0.2	0.2
Cross Language	0	0	0	3.0
WON	0	0	0	0
Total	2.7	10.6	13.3	24

2.8.4 Summary

We have identified a number of potential channels to market which have been validated through a number of channels including our user group, external users, and our early adopter success with the EPO. While Google Translate remains the biggest threat to any potential wider uptake of solutions developed within the project, we believe there is sufficient scope and opportunity to adapt our technology to resolve issues that still remain in the offerings of Google and other competitors.

3 Deliverables and Milestones Tables

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	Comments
7.2	First report on survey's results	7	CL	R	P	30/09/11	Yes	14/10/11	Lead taken by CL
7.6	First report on the intrinsic and extrinsic quality of MT	7	CL	R	P	30/09/11	Yes	14/10/11	Lead taken by CL
8.2	Dissemination Activities	8	DCU	P,D	P	30/09/11	Yes	14/10/11	Lead taken by DCU
9.2	Exploitation Plan	9	EST	R	CO	30/09/11	Yes	14/10/11	Extended to include new section on feasibility of MT-only scenario
1.1b	Annual Project report	1	DCU	R	P	31/03/12	Yes	30/04/12	Due 60 days after the end of the reporting period. Delivered in advance of the site review.
2.2	Data Corpora and Standards v2	2	EST/DCU	O	CO	31/03/12	Yes	30/04/12	Lead taken by EST/DCU
4.2	TMs for 4 languages	4	EST	O	CO	31/03/12	Yes	30/04/12	
5.2	MT for 4 languages	5	DCU	R	RE	31/03/12	Yes	30/04/12	
6.2	Second Prototype	6	EST	D	RE	31/03/12	Yes	30/04/12	
7.9	Report on the impact of TMs in the integrated system	7	CL	R	P	31/03/12	Yes	30/04/12	New deliverable added to the revised DoW

Table 2. Milestones

Milestone no.	Milestone name	Due achievement date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
2.3	DE patent data available	30/06/11	Yes	30/06/11	
2.4	ES patent data available	31/12/11	Yes	31/12/11	
3.2	Second web application and user interface available	31/03/12	Yes	31/03/12	
4.3	TM resources for EN-DE	30/09/11	Yes		
4.4	TM resources for EN-ES	31/03/12	Yes	31/03/12	
5.3	MT System for EN-DE		Yes		
5.4	MT System for EN-ES	31/03/12	Yes	31/03/12	
6.3	Integrated System v1	31/03/12	Yes	31/03/12	
7.2	First report on survey’s results published	30/09/11	Yes	30/09/11	
8.2	Prototype dissemination and training plan	30/09/11	Yes	30/09/11	

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 User Engagement

Interaction with users was limited in the first year of the project as the scheduled activities in the period were more focused on the development of technical areas. As our translation systems came online – with positive early feedback on quality part-driven through the EPO collaboration – and our software began to take shape, the benefits to be gained from increasing our engagement with users became more apparent.

To this end, we initiated some meetings with our partners at WON shortly after our first AGM (as described in Deliverable 1.1a). This has developed into a very strong working relationship over the second period and has extended beyond WON to include a much bigger user base from the wider IP community. In the following, we will describe our level of engagement with our partners at WON, the extended WON membership, and other users.

WON Core Working Group

From the entire WON organisation, there is a working group of between 5—9 individuals with whom we work very closely on project related activities. This has been facilitated greatly by CrossLang who, given their proximity to the Netherlands and the fact that they are a native Dutch speaking group, have coordinated communication between the working group and the consortium.

Since the first year review, we have held 5 dedicated meetings with the working group. At the meetings, we discussed a range of topics including software functionality, exploitation possibilities, and methods for evaluation. Prior to these meetings, we would often release a new version of the browser plugin tool for the members to test and provide feedback on in person. They also help us design, prepare, and disseminate the usability experiment described in Deliverable 7.6. A summary of these meetings is given below:

Date	Location	Notes
12/05/2011	Eindhoven	First meeting with full working group
20/09/2011	Eindhoven	
08/12/2011	Gent	Representatives from DCU travelled to Belgium to attend this meeting in person.
01/03/2012	Eindhoven	The users had been testing the new IPTranslator interface which was launched the following week.
12/04/2012	Utrecht	This meeting took place following a presentation at the WON AGM.

WON Community

Through our engagement with the PLuTO working group at one, we were afforded access to the entire membership base. In order to maximise the benefit of these users, and not to exhaust such a useful resource, we carefully selected the opportunities we would take to seek their input.

To this end, we first sent the user survey of Deliverable 7.1 to the members and received 84 responses. This was very important as it enabled us to make key strategic decisions such as which languages to develop translation systems for and which search tools we should support. In the survey, the users were asked if they would like to be kept up to date with the developments in the project to which the majority replied positively. This has allowed us to maintain direct contact with interested parties. We have since met a number of these individuals and some of the events described in section 4.3 and built upon our initial connections.

The second instance in which we engaged the WON members was during our search for participants in the usability experiment described in Deliverable 7.6. In total, we managed to get 11 people to take part in our French—English task. This figure is quite good considering the fact that the experiment was quite time consuming (approx. 2 hours).

Going forward, we plan to interact with these users more in terms of how me might improve and optimise our IPTranslator and ParonPro services through testing and in-situ evaluations.

Intellectual Property Community

In following our policy of attending more IP related events, as set out in Deliverable 8.2, we have had the opportunity to connect with “real” end-users in the community. Following the EPO Conference in October 2011, more than 90 users signed up through the project website to be kept informed of project activities.

Communication with these users has been maintained using the customer relations management tool Campaign Monitor and newsletters have been sent out before and after major events. The challenge going forward is to convert these interested parties into actual users of the services and ultimately covert them into customers.

4.1.2 IPTranslator Launch

As mentioned in section 2.7, there was a need to change the presentation of our software in order to acquire more serious users. To that end, the IPTranslator.com service was launched in March 2012. To date, the service has been run with a lean start-up mentality whereby frequent small changes (iterations) are made to the interface and feature set are based on intrinsic and extrinsic feedback from users.

In addition to discussing the service with users at events, we run web analytics tools on the website in order to observe users’ behaviour and interaction. This informs us which how many people visit the site and download the plugin, which features are most popular, at which point visitors are dropping off, and more. We can use this information to improve ease of use and provide more helpful information and instructions where needed.

In the month following the launch at the IPI Confex in March 2012, IPTranslator.com had over 150 unique visitors and more than 700 translation requests were processed. We intend to follow much the same approached with the ParonPro launch later this summer.

4.1.3 Reallocation of Project Resources

One task carried out in the second period which should not be underestimated in terms of administrative effort was the reallocation of the IRF's resources amongst the remaining partners and the revision of the description of work to reflect the new circumstances.

While the consortium had essentially been operating under the new assumptions since the announcement of the IRF's departure, the exact details of the new plan still needed to be formalised. A final version of the DoW was agreed amongst the partners at a management meeting in June 2011 (see section 4.2.2) and formalities completed with the commission by early 2012. The main changes in the new DoW are summarised below:

- Search was dropped and resources reallocated to allow for extra translation work;
- Cross Language's role was increased to include more evaluation and interface work. They also took the lead of WP7;
- New deliverables were added for evaluation of the translation memories and to assess the feasibility of MT as a service;
- DCU took over dissemination while exploitation resources were increased for all partners to reflect the more commercial focus of the project.

4.1.4 Risk Management

The risk management plan submitted in M6 as Deliverable 1.0b was revised to reflect the new circumstances in the project. This primarily focused on installing contingencies should the consortium suffer the withdrawal of a further partner which would render its composition invalid under the terms of the grant agreement. This revised plan was resubmitted at M18.

4.1.5 Advisory Board

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

Our good relationship with the advisory board continued in the second year of the project with even more interaction between the advisors and the partners. Partners met with Viggo Hansen at the EAMT Conference in Belgium, while PLuTO was invited to speak at Viggo's IPWare Summit in Italy. DCU was in constant contact with Fred Hollowood as he visits the CNGL office on a weekly basis, while other partners met with him at EAMT also. Stephen Adams was in attendance at most of the other IP related events attended by PLuTO participants including the PIUG Conference, the EPO Patent Information Conference, the WIPO Symposium, and the IPI Confex. These events are described in greater detail in section 4.3.

We should also note that Stephen Adams received a lifetime achievement award at the IPI Conference further highlighting his importance in this field and how beneficial it is to have him on our advisory board.

4.2 Project Meetings

4.2.1 First Year Review Meeting – Dublin – May 5th 2012

The first year review meeting was held at DCU 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 6; Cross Language 2; IRF 1; ESTeam 3

4.2.2 Management Meeting – Budapest – June 17th 2011

As senior members from all funded partners were in attendance at the METAFORUM in Budapest (see section 4.3) the opportunity was taken to hold a meeting in order to finalise the redistribution of the IRF's resources amongst the remaining partners. The revised Description of Work was subsequently finalised and submitted to the commission.

Attendees: ESTeam 2; DCU 3; Cross Language 2

4.2.3 General Assembly – Crete – September 29-30th 2011

A full two-day general assembly was hosted by ESTeam in Crete in order to review progress since the review and to finalise deliverables due at month 18. A significant portion of time was also devoted to planning in three areas: improving the integration of translation memories with MT, designing the evaluation for the new deliverable 7.9 on assessing the impact of the TMs on translation quality, and brainstorming strategies for exploitation.

Attendees: ESTeam 7; DCU 2; Cross Language 1

4.2.4 AGM – Berlin – April 18-19th 2012

The second PLuTO annual general meeting was held at ESTeam's headquarters in Berlin shortly after the official end of year two data and all partners were again represented. This time was used to review deliverables due at month 24 and establish more concrete strategies for year 2, particularly around exploitation.

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. Unfortunately, one member of the Advisory Board, Stephen Adams, was unable to travel to Berlin for health reasons but he nonetheless provided feedback on our written deliverable via email.

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

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. These activities are detailed in the following sections.

4.3.1 Intellectual Property Events

The IP related conferences attended by PLuTO representatives are summarised the in the tables below.

Event	International Trademark Association Conference (INTA)
Date/Loc	May 2011 – San Francisco, USA

Description	INTA is a large international conference with vendor exhibits and over 9,000 intellectual property professionals in attendance.
Attendees	Gudrun Magnusdottir, Jochen Hummel, Ingimar Andresson, Lambros Kranias (EST)
Justification	ESTeam attended the INTA conference with an exhibition booth at which PLuTO was demonstrated. Despite the fact that this event is primarily aimed at trademarks, the majority of attendees operate in the patent space also and this event serves as a valuable networking and customer relationship building exercise as companies with multilingual IP needs are heavily represented.

Event	Patent Information User Group Conference (PIUG)
Date/Loc	May 2011 – Cincinnati, USA
Description	The annual PIUG conference features world-renowned experts on patent information for technology research and planning, for legal organizations, and for overall corporate IP management.
Attendees	Aalt van de Kuilen (WON), Stephen Adams (Advisor)
Justification	PLuTO was presented by invitation at the PIUG Conference by the head of the WON PLuTO Working Group Aalt van de Kuilen. This was very appropriate as the PIUG is the world's largest patent user group and it was introduced to PLuTO for the first time by "one of their own" which gave the project much credibility.

Event	WIPO Symposium of Intellectual Property Authorities
Date/Loc	September 2011 – Geneva, Switzerland
Description	The World Intellectual Property Organisation's symposium of IP Authorities is an annual event at which heads of IP authorities, industry leaders and other stakeholders share ideas and experiences for improving services to be provided by IP authorities.
Attendees	John Tinsley (DCU), Stephen Adams (Advisor)
Justification	PLuTO was invited to present during a panel discussion entitled "How far can Machine Translation overcome Language Barriers?" Other members of the panel included the head of global databases at WIPO, the head of translation at WIPO, the head of business development and MT at Google, and a leading academic in the field of MT. Many early relationships with patent search vendors and patent office representatives were struck up at this event.

Event	IPWare Summit
Date/Loc	October 2011 – San Remo, Italy
Description	The IPWare Summit is an international conference for IP Professionals and vendors organised by PLuTO advisory board member Viggo Hansen.

Attendees	John Tinsley (DCU), Viggo Hansen (Advisor)
Justification	PLuTO was presented during a session on patent tools at the IPWare Summit. The intimate nature of this conference made it an extremely valuable experience for extracting knowledge from leading minds in the IP world as well as providing important networking opportunities.

Event	EPO Patent Information Conference (EPO PIC)
Date/Loc	October 2011 – Kilkenny, Ireland
Description	The EPO PIC is a popular conference held annually by the European Patent Office, former collaborators with PLuTO, at which users, vendors, and representatives of IP authorities attend.
Attendees	John Tinsley, Páraic Sheridan (DCU), Aalt van de Kuilen (WON), Stephen Adams (Advisor)
Justification	The opportunity was taken to exhibit PLuTO at the EPO PIC given its proximity to DCU. This event signalled the beta release of the browser plugin tool to users. It was a huge success with over 90 people signing up at the booth. We also saw some of the first signs of the commercial potential for such a translation offering given the reaction of people we spoke to. A number of connections were made at this event also when have continued to the present, including our good relationship with the people at Minesoft (PatBase). Finally, the Irish press became aware of the event following a press release DCU and PLuTO subsequently gained national exposure as detailed later in this section.

Event	IP Service World 2011
Date/Loc	November 2011 – Amsterdam, the Netherland
Description	IP Service World is an international conference and trade exhibition, with a strong focus on patents, attended by IP professionals and vendors.
Attendees	Jochen Hummel (EST)
Justification	At IP Service World, Jochen led and moderated a roundtable discussion on IP and multilinguality entitled “How to serve an increasing global customer base”. This turned out to be a great opportunity to have an intense exchange of practical experience with small groups of people representing a wide base of users.

Event	International Patent Information Conference and Exhibition (IPI Confex)
Date/Loc	March 2012 – Barcelona, Spain
Description	The IPI Confex is a conference and exhibition fair dedicated exclusively to patent professionals. It was heavily recommended to us by members of WON.
Funded Attendees	John Tinsley, Declan Groves (DCU), Aalt van de Kuilen (WON), Stephen Adams (Advisor)

Justification	PLuTO had an exhibition stand at the IPI-Confex at which the IPTranslator tool was officially launched. Significant interest was again shown in the tool at this event, particularly given the new branding, and this was reflected in the spike in usage during and in the weeks following the event.

4.3.2 Other Events

In addition to IP related events, consortium members have attended other events in the language technology sector in order to satisfy the requirements of the project as well as maintain the active profile of the various partners in this field. A summary of the relevant events is given below.

Event	EAMT Conference (European Association for MT)
Date/Loc	May 2012 – Leuven, Belgium
Description	EAMT is an annual event and is the largest event dedicated to MT in Europe.
Attendees	John Tinsley, Páraic Sheridan, Alex Ceausu (DCU), Heidi Depraetere, Joeri van de Walle (CL), Viggo Hansen, Fred Hollowood (Advisors)
Justification	EAMT 2012 held a special FP7 showcase plenary session to highlight those projects exploiting MT technologies. PLuTO was introduced during a short oral presentation and then during a poster presentation by John. Alex Ceausu also presented the PLuTO paper on domain adaptation for patent MT during the plenary sessions.

Event	META-FORUM 2011
Date/Loc	June 2011 – Budapest, Hungary
Description	Language Technology Symposium led by PLuTO Collaborators METANET
Attendees	John Tinsley, Páraic Sheridan (DCU), Gudrun Magnusdottir, Jochen Hummel, Mihai Lupu (EST), Heidi Depraetere (CL)
Justification	METAFORUM was a symposium which gathered organisations working on language technology solutions. PLuTO had an exhibition booth at the METAFORUM where the pre-release browser plugin was demonstrated to professionals from across Europe.

Event	CNGL Technology Showcase
Date/Loc	November 2011 – Dublin, Ireland
Description	CNGL's annual public showcase of technology to government, academia, and industry representatives.
Attendees	John Tinsley, Páraic Sheridan, Alex Ceausu, Jian Zhang (DCU), Fred Hollowood (advisor)

Justification	PLuTO was again demonstrated at an exhibition booth during this event where the public and representatives from the localisation industry were invited. The event was opened by the Irish Minister for Enterprise following an invitation extended to him by DCU at the EPO Patent Information Conference.

Event	EACL Conference (European Association for Computational Linguistics)
Date/Loc	April 2012 – Avignon, France
Description	EACL is the largest European conference dedicated to computation linguistics and related disciplines.
Funded Attendees	John Tinsley (DCU)
Justification	John was invited to speak on behalf of the PLuTO project at a workshop on the integration of machine translation with information retrieval. The talk, entitled “Facilitating Patent Search with Machine Translation” was accompanied by a peer-reviewing extended abstract. At this meeting, members of the MOLTO project were engaged to collaborate on a supplementary evaluation of our translation engines.

4.3.3

4.3.3

4.3.3

4.3.3 Dissemination Materials and Social Media

4.3.3

In order to support the promotion of the project and, in the case of IPTranslator, present a more professional look and feel, a number of marketing items have been designed and distributed. These include two pull up banners which are used to increase visibility at exhibitions and trade fairs. An example of these banners on show can been see in the picture below. Trifold leaflets and more visually appealing consumables have also been produced and these are included as physical appendices to this report.

4.3.3

4.3.3

4.3.3

A social media strategy has also been developed to help promote the project and the IPTranslator brand. The project’s twitter account @plutopatents is updated on a daily basis reflected project activities where appropriate, but also stimulating discussion on new topics in the IP world. This is supported by the project blog[9] which we endeavour to update twice weekly with interesting topics for discussion. For example, one recent post which received good traffic was a comparison of Google Translate output from one year ago on the new Google Translate which has been enhanced for patents. We also run a “New Features” series which various aspects of IPTranslator are highlighted, often accompanied by an instructional video.

4.3.3

4.3.3

4.3.3

4.3.3

4.3.3

[9] http://plutopatents.wordpress.com

4.3.3

4.3.3

4.3.3

IPTranslator.com
PLuTO
Patent Language Translations Online
www.pluto-patenttranslation.eu

The project website at http://www.pluto-patenttranslation.eu has been maintained and updated. In order to give it a more dynamic feel, our Twitter feed has been embedded into the main page. IPTranslator is also supported by a standalone website (iptranslator.com) to give the brand a distinct feel. This site serves as the landing page for the IPTranslator service and is populated with an introductory video, a tour of the service and feature, along with other information.

4.3.4 Other related activities

In addition to conferences and the web, the consortium has been active and visible in other areas. As mentioned previously, our presence at the EPO PIC in Ireland attracted press within Ireland and we made sure to promote this exposure.

There was also some interaction between PLuTO and other EC funded projects. A collaborative agreement was signed with the METANET[10] project in July 2011, while we agreed to collaborate with the MOLTO[11] project by allowing them access to our web service in order for them to perform a comparative analysis with their translation systems. We fully intend to continue to build these relationships going forward.

5 Future Plans

Business Development

The second year of the project has already seen exploitation plans accelerated to capitalise on the interest from potential customer segments. This has been necessary from a strategic perspective, in terms of capitalising on the gap that appears to exist in the market, and from a practical perspective, given that financial support from the EC ends in early 2013.

This will be a driving activity in year three as we continue to explore different channels to market and test various hypotheses in order to find product-market fit. Early plans are already in place to launch ParonPro in early summer which will give us a greater insight as to

[10] http://www.meta-net.eu/
[11] http://www.molto-project.eu/

the best way to position our services in the market, i.e. translation plus other services vs. other services plus translation.

We will also remain flexible in how we might look to grow the business opportunities in terms of looking "outside the box" at options and potential developments not explicitly specified in our description of work.

Dissemination, Marketing, and Sales

As the business side of the project continues to develop, dissemination activities will naturally gravitate towards becoming a vehicle for product promotion and sales. To some extent, by exhibiting software at trade fairs we are already positioning our offerings in the commercial space in the eyes of users.

At such a time as we are in a position to begin monetising these services, sales will enter the equation in order to attract more business. If/when this happens, our dissemination activities will adapt accordingly while still maintaining their original objectives which are to promote the project, build and maintain awareness, and foster relationships with users (customers). We intend also to take greater advantage of our partnership with WON to publicise the business and act as champions for our solutions.

Translation Quality

To some extent, a recall-based approach was adopted in terms of the development of our translation offering, particularly the machine translation engines. That is to say, we focussed on actually bringing systems online for the most in demand languages rather than making sure that each system that was released was the best it possibly could be.

Now that these systems are in place, year three resources will be expending on improving the translation quality – the precision (as opposed to the recall) to continue the analogy. This also ties in with the planned future work in terms of integration of TM and MT where we still see a lot of potential for improvement, particularly for those languages where MT is not as strong.

6 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 increased involvement of users, both within the consortium and the wider intellectual property community, the improved integration between the machine translation and translation memory technologies, the comprehensive evaluations which have been carried out, and the accelerated exploitation activities including the launch of the IPTranslator product.

Looking forward, we will continue to try to improve translation quality within both the machine translation and integrated scenarios. Business development activities will increase and will be supported by advanced dissemination.

In conclusion, we are pleased with our progress over the period. We took on-board advice from many quarters, including the project reviewers at year one, our advisory board, and our users, and we believe this stands us in good stead going into the final year.

7 Bibliography

Tinsley, J., A. Ceausu, and J.Zhang. 2012. PLuTO: Automated Solutions for Patent Translation. (extended abstract) In *Proceedings of the Workshop on Exploiting Synergies between Information Retrieval and Machine Translation at EACL.* Avigon, France.