SEVENTH FRAMEWORK PROGRAMME
THEME 3
Information and communication Technologies

PANACEA Project

Grant Agreement no.:   248064

Platform for **A**utomatic, **N**ormalized **A**nnotation and
Cost-**E**ffective **A**cquisition
of Language Resources for Human Language Technologies

D2.4

Platform Software, Project Tools + Resources Licensing Policy and Exploitation Plan

Dissemination Level:	Public
Delivery Date:	June 20 2012
Status – Version:	Final
Author(s) and Affiliation:	Victoria Arranz, Khalid CHOUKRI Olivier Hamon (ELDA), Núria Bel (UPF), Prodromos Tsiavos (Legal advisor)

This document is part of technical documentation generated in the PANACEA Project, **P**latform for **A**utomatic, **N**ormalized **A**nnotation and **C**ost-**E**ffective **A**cquisition (Grant Agreement no. 248064).

Please send feedback and questions on this document to: iulatrl@upf.edu

TRL Group (Tecnologies dels Recursos Lingüístics), Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra (IULA-UPF)

Table of Content

List of figures

1   Executive Summary

This report elaborates on the exploitation of the PANACEA project assets. These assets have been clustered into a few items (a) the PANACEA Factory/Platform, (b) the web services integrated within the platform, (c) the associated workflows to manage the sequencing of web services (d) the tools developed during the project and last but not least (e) the data sets, i.e. Language Resources (LRs) produced mainly for Machine translation/localization, but not only, within the project, exploiting the platform, web services, and the workflows.

In this report, we will describe the four items, referring to the related reports whenever necessary, focusing herein on their exploitation aspects. Two major aspects will be addressed: (i) the technical (and hence market) value and (ii) the implied legal issues, both to ensure that our exploitation is done in a cleared legal framework as well as to ensure that such assets can be licensed to third parties under a clean, easy to understand and implement, licensing schema.

From a technical point of view, the consortium has worked out the arguments about the added value of its achievements and has also addressed the crucial legal aspects involved behind the proposed paradigm that should not be neglected as far as exploitation is concerned.

The current deliverable explores possibilities to exploit the project's assets and conducts an analysis of the legal framework in which such an exploitation could take place[1]. In order to do so, we also describe the potential users and their needs, leading to the design of our exploitation plan.

The business model envisaged in this exploitation plan is based mostly on exploiting the Platform to produce new resources on demand, design workflows on demand, connect to third parties new web services, and also to make available the platform for internal use by commercial organizations.

The consortium agreed to ensure that the assets of the PANACEA project, described in this report, are maintained for at least 24 months after the end of the project. This includes a commitment for keeping the whole platform in place and running (the platform including Registry, myExperiment currently located at ELDA, the workflow engine and the existing workflows and related web services, currently located in UPF, CNR-ILC, ILSP, UC, DCU, LT and ELDA). Such agreement will help better assess the future exploitation after the implementation of what is described in this report.

The Language Resources developed within the project will be exploited as part of the ELRA catalogue but also using the META-SHARE infrastructure (in addition to the UPF e-repository). Such agreement will help better tune the exploitation plan as described in this report through this "sustainability" period.

2   Introduction

---

[1] The legal analysis, appended herein, goes beyond the simple analysis and draws a number of conclusions/recommendations that could be used by the community when dealing with similar technical tasks.

PANACEA project has focused on the development of a factory of LRs that automates the stages involved in the acquisition, production, updating and maintenance of LRs required by MT systems, and by other based on Language Technologies (LT) applications. This automation is meant to cut down costs significantly, in terms of time and human effort. Such reductions are the only way to guarantee a continuous supply of LRs that MT and other Language Technologies may demand in a multilingual Europe. In order to address this objective, PANACEA has worked in the following areas:

1) The development of a platform, designed as a dedicated workflow manager, for the composition of a number of processes for LRs production based on combinations of different web services.

2) The automatic production of LRs for MT and other LT by the use of advanced components for the acquisition and normalization of corpora, monolingual and parallel corpora, the alignment of parallel corpora; the derivation of bilingual dictionaries out of aligned corpora; and the production of monolingual rich information lexica using corpus based automatic methods.

This report elaborates on the exploitation of the PANACEA project assets. These assets have been clustered into a few items (a) the PANACEA Factory/Platform, (b) the web services integrated within the platform, (c) the associated workflows to manage the sequencing of web services (d) the tools developed during the project and last but not least (e) the data sets, i.e. Language Resources (LRs) produced mainly for Machine translation/localization, but not only, within the project, exploiting the platform, web services, and the workflows.

In this report, we will describe the five items, from a user perspective. We will briefly elaborate on the technical (and hence market) value and the implied legal issues.

3 PANACEA achievements and assets

PANACEA has developed a LR production factory that manages sets of workflows that target the production of Language Resources (LRs) through the exploitation of web services. The PANACEA project achieved a number of objectives based on the set-up factory. Among these objectives, we can list:

1. The PANACEA Platform

2. Exploitation of various (third party or PANACEA partners) tools accessible via the platform set up as web services, running either on a PANACEA Partner server (current scenario) or on a third party server (future possibility).

3. A set of workflows, designed to chain the various web services so as to produce useful and innovative LRs for the MT sector.

4. The set of language resources produced within the project and using the platform (e.g. monolingual, bilingual textual corpora, etc.), based on partners' existing background as well as web sources.

These led to valuable assets that we are planning to exploit.

3.1 The PANACEA Platform / Factory

The PANACEA platform is defined as an interoperability space, that is, it offers a number of features that ensure a successful integration into workflows that makes it possible to: (i) interrelate different web services by exchanging a "travelling object"[2], that is output data of one

---

[2] A travelling object is the data, in different stages of the processing flow, that is output of a processing component and input of another processing component.

web service and input data of another web service, and (ii) substituting a web service by another one that performs the same operation without re-programming the workflow. The specifications are: Travelling Object definitions and Common Interface definitions, respectively.

The concept of the platform is depicted in Figure 1. This figure provides a definition of the PANACEA platform structure but also all the items that we have to be careful about when talking about licensing them from/to third parties. This structure already points out the kind of parameters to take into consideration (e.g. tools, data, web services, workflows, workflow editor and engine, catalogues, output LRs, users, etc.) and the stages where such analysis should be considered. The different elements were analyzed, as indicated within Figure 1 that provides the main components of the platform and the production processes. From this figure, we derived Figure 2, which is providing us with all the check-points where deep analysis was required to tackle questions and worries that the platform users may run into (points here derived from discussions with the actual project partners and potential platform users).
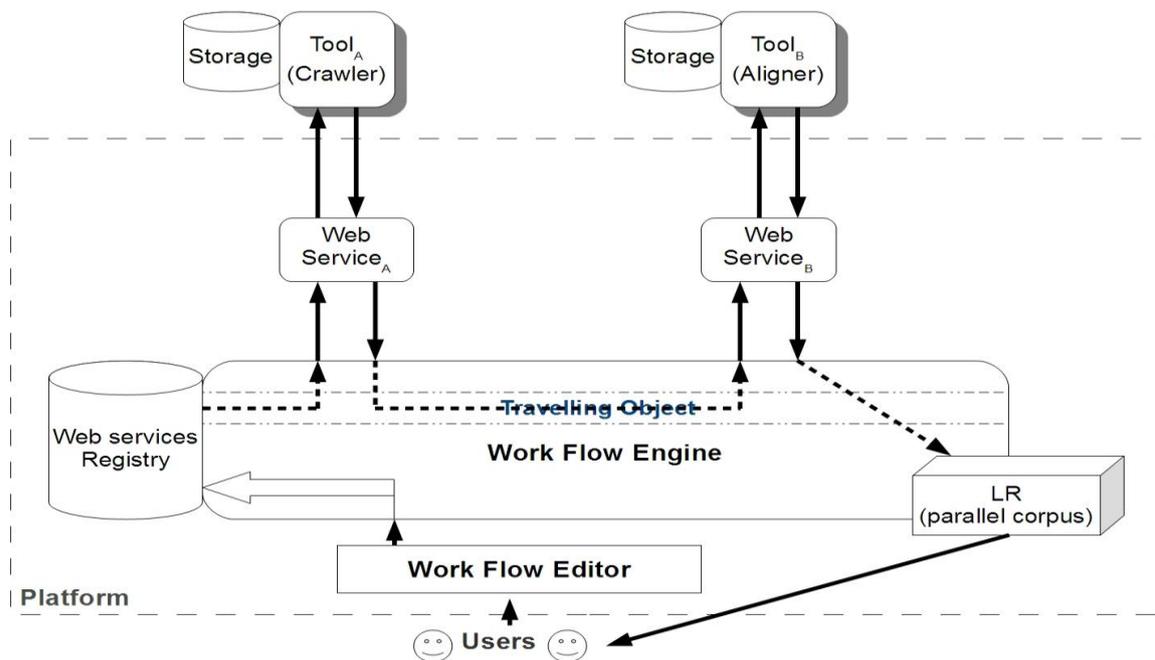

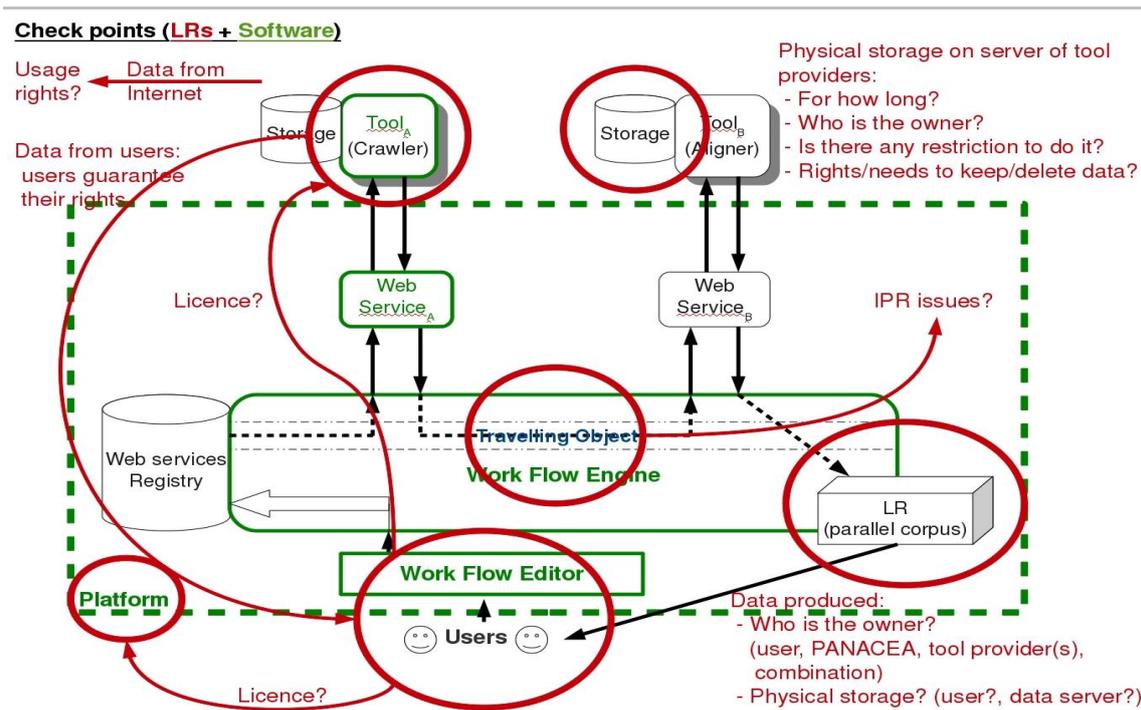**Figure 1: PANACEA platform. Starting point for the analysis**

Figure 2: "Check-points" concerns to be handled within the platform

The first thing worth noting is that PANACEA is using third party tools or applications, in particular, the platform backbone is based on Taverna[3], a workflow editor and engine, and a number of other related applications: Soaplab[4], myExperiment[5] and BioCatalogue[6].

The workflow editor and engine, Taverna, is offered under a GNU Lesser General Public License v2.1 (LGPL v2.1) and therefore already grants rights for derivative works. In particular, it allows that derivatives, that are made with but contain no code of the library, can be released as non-open. Particular workflows are considered a derivative work, which can be copied and distributed as well.

PANACEA also uses Soaplab, a tool that can automatically generate and deploy different Web Services interfaces and it is of main interest for integrating new web services. Soaplab is made available under the Apache License, Version 2.0[7] that neither has restrictions for being used in a context such as the PANACEA platform.

Furthermore, PANACEA uses the collaborative space called myExperiment, a tool that allows to share workflows (and other Research Objects) and to build communities and BioCatalogue*, "a centralised registry of curated Life Science Web Services. It allows you to easily discover, register, annotate, monitor and use Web Services."* Both applications are usable under a BSD licence, which imposes minimal restrictions on the redistribution and use of covered software.

---

[3] "Taverna is an open source and domain-independent Workflow Management System, a suite of tools used to design and execute scientific workflows and aid in silico experimentation" (http://www.Taverna.org.uk/).

[4] http://soaplab.sourceforge.net/soaplab2/

[5] http://www.myexperiment.org/

[6] http://www.biocatalogue.org/

[7] http://www.apache.org/licenses/LICENSE-2.0

These components, described in more details in deliverable D3.4, are used within the project in a way that ensures they would be usable by third parties given their respective licences.

Other tools implemented third parties and deployed as web services, are licensed under various particular licenses, mostly GPL v3 or BSD, and none of them has restrictions for its use in PANACEA platform. All PANACEA partners tools used as web services within PANACEA are offered by their owners as public services, accessible to all too. Usage conditions are mentioned at the PANACEA registry for every web service. A complete list of web services and usage conditions is at annex 2 of this document.

The exploitable part herein is the whole distributed platform. Sustainability plans have been agreed among the partners, and the commitment to maintain and support the platform (and services included) for at least two years, has been made by all consortium members. Users will be able to exploit the various web services in workflows and include new ones to produce useful outputs (e.g. Resources, performance measures, added-value annotations, etc.).

❖ **<u>Summary and Plan:</u>**

After analysis, it turned out that all the web services can be freely used by the research community for research purposes, and most of them with no restrictions. Usage conditions have been added to every web services, as well as licensing conditions of the tools running behind for some of them to reassure the user. A list is given in annex 2.

<u>It is the plan of the consortium individual partners to exploit the platform as described above to produce new resources on demand. The business model behind the exploitation of the platform is part of section 5.</u>

3.2 Web services

As indicated above, PANACEA web services are interfaces to running tools provided by the consortium partners that allow processing some data input and produce some data as output. A typical example of web-service is the language identifier. Such web service uses an algorithm to identify the language of a file in a given format (e.g. HTML/XML/PDF/) and is running on a partner server. Another web-service implements the BLEU scoring/metric algorithm. It takes as input translated data and human-translated texts (references) and outputs a BLEU Score.

It is important to understand how one can exploit such web services and for what purpose. In addition, one needs to understand the legal issues are involved. It is not always the case that web service use is governed by clear use conditions and terms, so one need to interpret these or their absence with respect to the usages but also the rights on the outcomes. This analysis has been carried out and recommendations are given with respect to the legal status of the input, output, and the web service itself.

The web services are the core part of the assets of PANACEA. Most of the NLP processing tools can be deployed as web services. More than 150 services are registered within PANACEA factory and are usable for a large set of processing (Annex 2 gives a list of available web services at the end of the project).

Regarding the usage of an application deployed as a web service, we would like to emphasize the strong relationship between the web service provider and the user. Indeed, whether a web service provider is the owner of the application or not, he must guarantee that the usage of the application(s) provided respects the usage rights of the application owner and that all IPR issues have been cleared between them. In particular, the web service provider must follow the specifications in the application license (i.e. attributions). For that purpose, we assume that the web service providers have cleared legal issues as required. They should indicate the terms-

conditions of use of the web-service. Figure 3 below illustrates how this information is being provided within the platform.
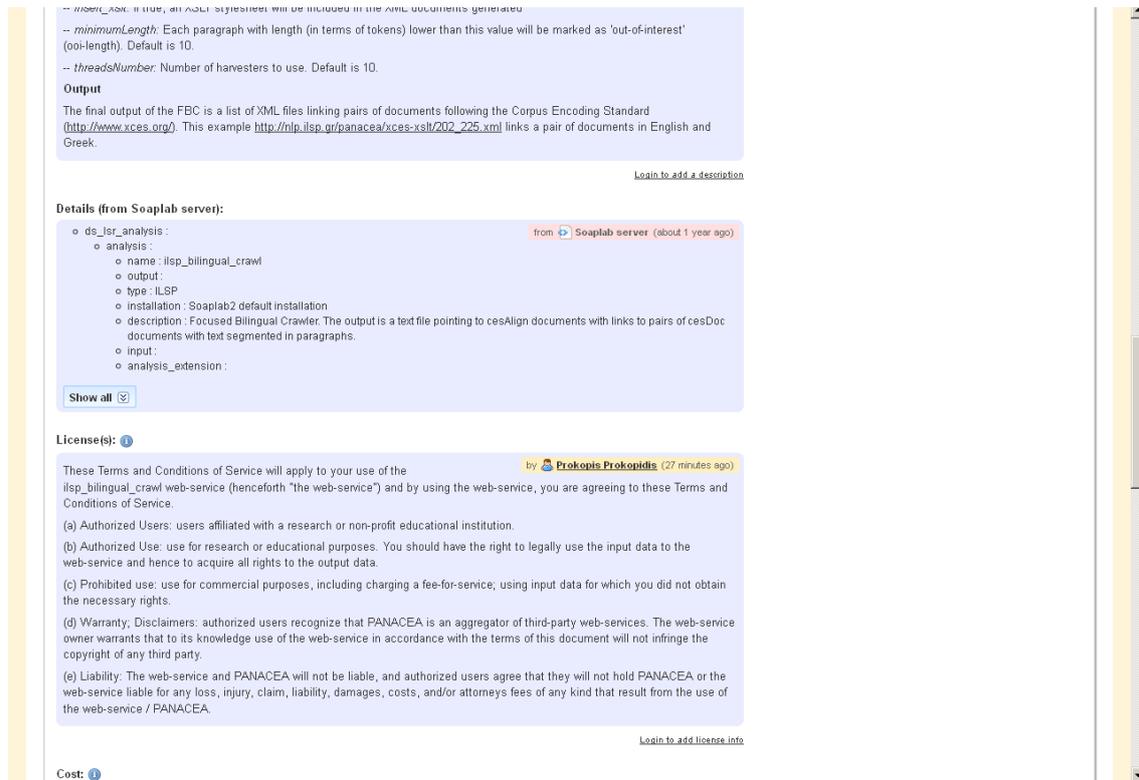

Figure 3: Sample of licensing information for web services

In addition to the practical application-usage point of view, one further disclaimer has been put into place. This states the fair use of the platform and delimits the number of processes that can be submitted in parallel. The exact details are as follows:

Fair Share Policy on Parallel Process Running

Users are kindly asked not to submit more than 3 processes/requests in parallel. This is part of the fair share policy implemented so as to allow all users to benefit from the web services offered by the PANACEA platform. If this policy is not complied with in a way that prevents other users from using the web services, users concerned may be prevented from submitting processes/requests, their exceeding processes may be killed and they may be black-listed for future use.

In the event of an exceptional need to use the platform in a manner not covered by this disclaimer, users are kindly advised to address the contact point of the web service(s) required so as to study the possibility of establishing an exceptional usage for those web services.

The list of current web services is given on http://registry.elda.org/services and consists of more than 150 services, deployed by the consortium on the basis of consortium-owned tools or tools from third parties. (A list of web services and license information is in Annex 2).

❖ **Summary and Plan:**

It is the plan of the consortium (individual partners) to exploit the web services within the PANACEA platform to process/produce new resources on demand. Such perimeter could

be extended at a later stage. The business model behind the exploitation of the platform is part of section 5.

3.3 Language Resources (LRs):

The PANACEA platform is designed to produce LRs. It has been put in operation to produce large set of useful language resources specially for Machine Translation purposes, but not only.

Resources produced are in the Resources Catalogue (D2.5) and available at the PANACEA web site. The following list sums up the type and languages covered.

- • Monolingual Corpora (EL, EN, ES, FR, IT)

- • Monolingual Corpora N-grams (EL, EN, ES, FR, IT)

- • Monolingual Dependency Parsed Corpora (EL, ES, IT)

- • Bilingual Aligned Parallel Corpora (EN-FR EN-EL)

- • Bilingual Glossaries (EL-EN, FR-EN, DE-EN)

- • Monolingual rich lexica (EN, ES, IT)

However, and as stressed above, it is the intention of the consortium to continue the exploitation of the platform for similar productions.

The platform allows the collection and/or processing of LRs. Two main types of data services may be requested by a user: either exploit some web services to collect data (data coming from the Internet, for instance, when using web services for crawling), or exploit some services to process material which is already "owned" by the user (or rather, "in the user hands"). In either case, the usage of data to produce other data might be restricted due to legal constraints.

The paradigmatic case would be web crawling for deriving any type of Language Resource. This is an important facet of the PANACEA Platform exploitation and we would like to elaborate more on the various issues involved.

Data crawling may be defined as the act of collecting different forms of information from the public Internet in an automatic fashion, which is then stored and processed in different ways. Crawling could be the initial step for different PANACEA web service and workflow operations. Text would be automatically compiled from different, automatically selected web sites. The first implication is that some of the crawled data may be protected e.g. for copyright considerations. However, not all texts in the web are protected. For instance, works made by the public sector agencies remain outside the protection of copyright law for most European legislation jurisdictions. The second implication has to do with what are the acts that are going to be performed upon the data, and then to assess two factors: (a) the degree to which results of such acts fall within the acts restricted by laws and (b) the extent to which results of such acts are visible enough to expose a PANACEA user to the risk of legal action. PANACEA has got expert advice about these two aspects. The main rationales behind this and the conclusions are important to our exploitation plan and hence are fully appended to this report. A brief summary of the recommendations is given herein:

(a) In the case of web crawling for LT purposes, the acts indeed include copying and processing of the texts and also the creation of derivative works with the aim of being used and/or communicated to third parties. These acts require separate permission of the

right holders, unless this is clearly indicated in a legal statement or the site indicates being offered under some permissive licenses such as Creative Commons.

(b) Though the act of web crawling is part of the daily operation of a web site and could be covered by an implicit license, the owner of a web site has technological means to prevent the site to be indexed or copied. Thus, if they do not exist, it might imply that the web site owner wishes it to be copied. To rely on the implied license might be considered a rather prudent attitude even though the goal of crawling for LT purposes may not have been the intention of the web site owner. However, legal advice suggests to reduce the risks crawling for LT, by doing the following:

    (i)   only crawl sites where bots are allowed

    (ii)   include a public notice stating that their works only derive from web sites that do not prohibit crawling;

    (iii)   include a notice and a takedown procedure indicating under which circumstances the material will be taken down, what the decision making procedure is and an e-mail address where relevant complaints could be addressed.

With regard to **sources provided by the users of the PANACEA platform**, they must guarantee that rights to use them has been granted by the right-owners (explicitly or because of the licensing schema). This is established as such within the **Terms of Use** of the platform. It is the sole responsibility of the input provider to check and ensure that (s)he has the right to use the input data (s)he uses within the platform. All these points and other related ones are duly indicated within the PANACEA platform to avoid any misunderstanding. The project aims at providing users with clear statements so as to know how to handle every scenario.

For PANACEA project data sets, i.e. the materials that have been produced for testing of different web services and which indeed have been derived from Internet crawling, there has been a laborious procedure for contacting crawled web site right holders and obtaining the relevant rights (ELDA has been granted with appropriate rights in written). In addition, a disclaimer notice, providing contact details for web site owners that want their content to be deleted, has been produced. The disclaimer states:

DATA DISTRIBUTION DISCLAIMER: The right to use the sentences contained in this data set has been granted by their copyright holders. This usage is exclusive for research purposes and no profit can be made out of it. We are grateful to all sources for their kind and generous contribution. For further information on these sources, please see: Acknowledgements. We tried our best to remove any content that holds copyright issues from the data released. However, if you find any such content please contact us and we will remove it.

The importance and efficiency of the use of the platform by third parties may seriously be hindered by the legal constraints, in particular when it comes to **obtaining the authorization to use external data. A** case study has been carried out within the project based on the crawled data from the crawling technologies that have been integrated. This study has consisted in the following and has been published in Arranz, Victoria and Hamon, Olivier (2012). **On the Way to a Legal Sharing of Web Applications in NLP.** In Proceedings of LREC 2012, Istanbul, Turkey.

    (i)   analyzing quantitatively the full implications behind using the crawled data, in particular with the perspective of future massive data handling;

    (ii)   providing the means for users to do so themselves;

(iii)   describing the procedure and execution cost clearly.

This "feasibility" study clearly describes the steps to carry out to benefit from the platform and supports our plans for its exploitation as a Language Resource Production factory.

❖ **Summary and Plan:**

It is the plan of the consortium partners (and in particular ELDA) to exploit such resources immediately, incorporating them within its ELRA LR catalogue and within the newly established repository network (META-SHARE), for research purposes.

The business model behind the exploitation of the platform is part of section 5.

3.4 Workflows

PANACEA has delivered various workflows that implement the way web services are chained as sequences, with specific input and output (all are at PANACEA myExperiment). For instance, a workflow that consists of a sequence of two "calls" to web services and may take as input a set of URL and identify the language(s) in which they are written and hence give as out a table with pairs (URL, language). It could be chained with one that crawl all URLs that are in a given language.

Such work flow could be (identify-languages (list of URLs) + Crawl (URL, Language=Fr).

(1) (Input = set of URLs) ➔ Web services 1: identify languages ➔ (output (URLs, languages)

(2) (input (URLS, Languages) ➔ Web services 2: Crawl (URLs, Language=Fr) and store ➔ output (set of XML, HTML, PDF, … pages stored on a given address)

It is clear that more sophisticated workflows could be designed, requiring deep analysis and more creativity and hence triggering the application of copyright and other property rights.

Some of the workflows are illustrated by the diagrams given below. More details are given in the appendix elaborating on legal issues.
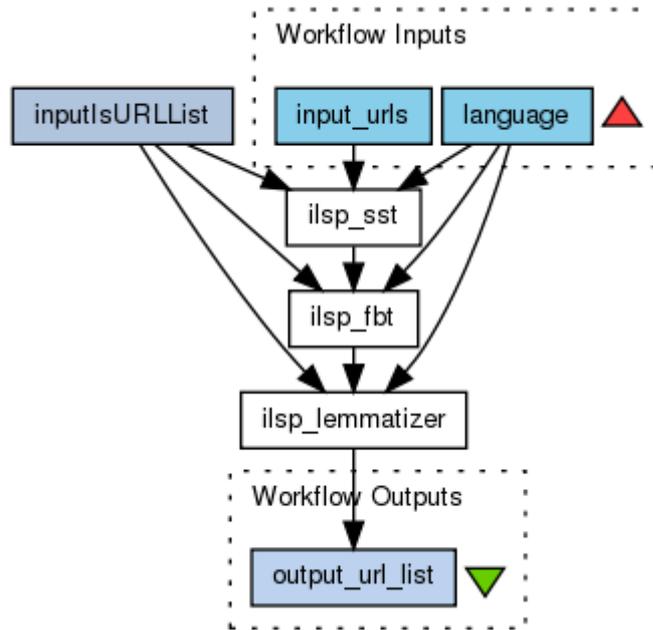
Figure 4: Workflow illustration (Sentence Splitting, PoS tagging and Lemmatization)

The consortium took seriously all aspects related to the ownership of workflows, both to ensure that the exploitation within the consortium is done in a cleared legal framework as well as to ensure that such assets can be licensed to third parties under clean, easy to understand and implementable licensing schema.

From a technical point of view, the consortium can easily work out the arguments about the added value of its achievements but it is also crucial that the legal aspects involved behind such a paradigm should not be neglected.

When dealing with workflows, data are not only stored on the server of the different web services, but are also "travelling" between web services. To guarantee the privacy of the data transferred from one web service to another, the transfer protocol must be secured so as to avoid any security bridge. Indeed, data going from one server to another (e.g. in the case of a workflow process) or from a client machine to a server (e.g. in the case of a single web service process) should be secure enough so as not to be corrupted or retrieved by a third user. In PANACEA, this process is secured by using SOAP[8] (Simple Object Access Protocol), which allows to reach a sufficient level of security since SOAP transports data using both SMTP and HTTP (and although not implemented in the current version, potentially HTTPS).

Workflows are, as explained above, original works produced by the partners and that are relying on a LGPL license of TAVERNA. The consortium decided to offer them freely and under a Creative Common license (most existing Workflows are now published with CC_BY_SA 3.0 licenses).

❖ **Summary and Plan:**

---

[8] http://www.w3.org/2002/07/soap-translation/soap12-part0.html

It is the plan of the consortium (in particular ELDA) to exploit such workflows to design new processing capabilities for potential users. In many cases, users do not care about the tools used to implement some particular processing. But in other cases this may happen, for instance we have seen users requiring some piece of textual corpora to be annotated by several taggers (multiple annotations), followed by the use a web-service executing a "majority-vote" to select the best tag of a given word. In this case, PANACEA workflow engine can be used for this purpose.

Several members of the consortium intend to offer services of this kind to the community. The business model behind the exploitation of the platform is part of section 5.

3.5 Tools developed in PANACEA

As already mentioned, new tools have been developed during the project and are also an asset of the project. In Table 1, we describe the tools that have been developed in the project (therefore foreground material as described in the Consortium Agreement), agreed licenses, and associated web service conditions. All of them will be exploited by their respective owners (no problems arise from two partners being involved in the same tool).

Name	Description	Other software involved	Program license	Web service Terms & conditions of use
ILSP Focused Crawler (ILSP-FC)	Research prototype for acquiring domain-specific monolingual and bilingual corpora	It depends on open-source libraries: BIXO for web mining and CASCADING for building data-processing workflows.	open-source Java project released under the GNU License, Version 3.0.	Use for Research purpose
DCU Bilingual dictionary generator P2G	Bilingual dictionary creation from factored phrase tables that include part of speech tagged text for EL-EN and FR-EN language pairs		GPLv3	Use for Research purpose
UPF LSF classifiers	They perform the classification of a list of nouns as belonging or not belonging to the given class, for instance EVENTIVE nouns.	They depend (pipeline) on open-source tools: Corpus Work-Bench CQP and Weka, both are licensed under GPL 3.0.	released under GPL v3 license	No usage restrictions on the web service.
UC Inductive subcategorization frame acquisition	It takes parsed text as input and produces a subcategorization		Available under GNU General Public License	No usage restrictions on the web service.

	frame lexicon			
CNR LMF_ML_MERGER	Multi-level merger for LMF lexica		GPLv3.0	usage free for Research and Commercial
CNR lmf_merger	Merges to LMF lexicons for (syntactic) subcategorisation information,		GPLv3.0	usage free for Research and Commercial
CNR Extractor_MW	A language independent tool that implements statistics-based methods for the acquisition of multi-word expressions		GPLv3.0	usage free for Research and Commercial
CNR SCF_Extractor_lang_indip	A language independent tool that implements and inductive method for the acquisition of verb subcategorisation frames		GPLv3.0	free for Research and Commercial
CNR SCF_Extractor_IT	Implements an inductive method for the acquisition of verb subcategorisation frames for Italian		GPLv3.0	free for Research and Commercial
LT-BiLex-Extract	Prototype of a bilingual lexicon extractor from phrase tables	--	Apache 2[9]	CC-BY-SA-NC
LT-Decomposer	Prototype for decomposing German compounds into their parts	--.	CC-BY-SA-NC	CC-BY-SA-NC
LT-Tagger	Prototype of a rule-based POS tagger for German	--	CC-BY-SA-NC	CC-BY-SA-NC

❖ **Summary and Plan:**

Tools developed in the project mainly have open licenses, although some of them will be restricted to research use only because of used software licences.

---

[9] The LT-BiLex-Extract component is already released in the ACCURAT toolbox framework (under the name LT-P2G).

4   Target User Communities

When looking for target users of whatever product to be "put in the market", one of the first things to do is analyse and define who may be such users and what their needs are. The analysis has to consider whether these needs match what we can offer, or whether we may even envisage expanding our offer, or just customizing it, to reach a larger market. Bearing all this in mind, the profiles of the potential user communities of a platform like PANACEA, and the other assets described in previous sections have been sketched. This analysis capitalizes on the Consortium's knowledge of the area, together with our experience acquired/enforced all along the project. This analysis is closely based on the outcomes of a number of events where different user types have been approached and have been granted the opportunity to have their saying on what PANACEA can do for them.

Generally speaking, three large user groups or communities have been defined:

1. Scientific community;

2. Technology development business community;

3. Translation and localization community (even if this can be seen as part of 2.).

All three communities have been approached at different stages of the project, as was planned, given that each community has different requirements. All these requirements have been translated into expectations that had a direct impact on the level of development for the platform. For instance, as can be understood, we cannot approach an industrial user with a not-very-developed prototype. Both his/her interest and feedback will be very limited.

4.1   Scientific Community

The first user group covers the R&D community, comprising researchers (and NLP students) interested in either applying the technology within PANACEA for their research or profiting from the data that may be produced within the project or by the platform at a later stage. Their objective is scientific and generally non-profit. Due to budget reasons (many of them doing research in the framework of their studies or research projects), they generally look for freely available data and software, or as low-cost as possible. Even if this is not financially viable for the PANACEA platform sustainability as a whole, it also provides some positive return in terms of:

-   exchanges regarding further improvements and development;

-   contributions from their own tools (being also deployed as web services).

All this may very well be focused towards the maintenance of the platform, which will indeed contribute towards its sustainability.

4.2   Technology Development Business Community

The second user group covers the industrial technology development community. Members within this group may carry out research for technology development and improvement, but mostly with a commercial purpose in mind. In that sense, they are cautious with the information they exchange (which may be part of their business plan and thus should not be unveiled) and they require products whose use is compatible with their commercial purposes, that is, not limited for research, and that they can easily and lawfully use and/or integrate within their systems. Generally, they do not look for promising results (unless fully targeting their interests) but rather expect components/resources which will ensure an improvement somewhere: in performance, in productivity, all of it with an impact on their revenues.

4.3 Translation and Localisation Community

The third user group gathers a specific subgroup of the industrial community. It may or may not comprise technology developers, comprising for instance:

- Professional translators in need of either data (translation memories) or the means to produce such data;

- Translation and localisation agencies running large and numerous translation projects, incorporating some technological systems/services.

Even less than the industrial technology developers, these users do not want to hear about development and how things are built and improved. They want to know how PANACEA can ease their life from a purely platform user's point of view.

5 The various Business models

5.1 The Platform, web services, and the workflows

It is important to review all possible means to generate financial revenues from the assets of PANACEA.

As indicated, we do not expect to collect revenues from the research community. A particular scenario could be that for users to benefit from PANACEA integrated web services, they should provide e.g. another web service to the platform. This is a reasonable and acceptable way to enrich the platform and keep it alive, while extending potential services to other communities. This may be particularly interesting for those users willing to share their data or web services so as to benefit from either platform web services or from special conditions towards some platform's assets. The consortium will offer hosting possibilities on its partners' servers (e.g. ELDA, UPF) for such services.

A second scenario is to offer the platform as a possible infrastructure to "host" or connect services based on users' web services. It is envisaged that users may be interested to utilize the platform as a host for their tools. In this case, the following possibilities are suggested:

- For free: providers are allowed to supply services for free as long as they do not make any profit from it, that is, as long as they give it on a non-profit basis, even if just for research. Providers have to commit to the quality of the services offered as well as availability over time.

- For exchange: providers are allowed to exchange their applications (uploaded as web services) for the using of some platform's web services, as long as they do not make any financial profit through the platform.

- For a fee: companies may wish to allow the community to use some of its applications through the PANACEA and charge a fee for that. The platform manager will handle these payments and will assure that a percentage is kept for the platform as service provider.

A third scenario is the exploitation of the Platform by a third party who needs to use the platform and its workflows for internal purposes. Such case, with the corresponding support and assistance, can be charged for.

5.2 The Language Resources

As indicated, LRs developed within the project will be made freely available to give more visibility to the platform and hopefully generate more business with productions on demand, in particular for commercial organizations.

Note that current versions of some resources (mainly glossaries and lexica) have only been partially evaluated (cf. D7.4), and although potential error might be low, they cannot be supplied as a product with enough quality.

6 Summary of the exploitation of the PANACEA Assets

The Platform: will be exploited as such to produce R&D resources on demand. It could be also exploited through creative workflows and with raw data, obtained with the adequate rights, to produce resources for industry, in particular Machine Translation and Localization players, but not only. Such exploitation service will be rendered by the Consortium members under market conditions.

The Language Resources (see the exhaustive list on http://www.PANACEA-lr.eu/en/info-for-researchers/data-sets/): such resources will be made freely available for research purposes. Although, this will not generate revenues, it is expected that such use will make PANACEA (and the services) more visible and may be trigger more commercial projects for similar productions.

Web services: integrated into the final version of the platform will be exploited by the consortium (as individual partners), for instance to generate new resources on demand.

Workflows: The workflows are intrinsically associated with the Platform and hence will be part of the platform exploitation. They may be exploited as standalone pieces if the users have adopted similar environment (Taverna, etc.).

The tools developed during the life of the project (therefore foreground material as described in the Consortium Agreement) will be exploited by their owners, for instance to generate new resources on demand.

7 Conclusions

The business model envisaged in this exploitation plan is based mostly on exploiting the Platform to produce new resources on demand, design workflows on demand, connect to third parties new web services, and also to make available the platform for internal use by commercial organizations.

The consortium agreed to ensure that the assets of the PANACEA project, described in this report, are maintained for at least 24 months after the end of the project. This includes a commitment for keeping the whole platform in place and running (the platform including Registry, myExperiment currently located at ELDA, the workflow engine and the existing workflows and related web services, currently located in UPF, CNR-ILC, ILSP, UC, DCU, LT and ELDA). Such agreement will help better assess the future exploitation after the implementation of what is described in this report.

The Language Resources developed within the project will be exploited as part of the ELRA catalogue but also using the META-SHARE infrastructure (in addition to the UPF e-repository). Such agreement will help better tune the exploitation plan as described in this report through this "sustainability" period.