

**ECP 2006 DILI 510049**

**ENRICH**

# Personalised Translation Interface delivery report

|                            |  |
|----------------------------|--|
| <b>Deliverable number</b>  | <i>D6.2</i>                                |
| <b>Dissemination level</b> | <i>Public</i>                              |
| <b>Delivery date</b>       | <i>20 December 2008</i>                    |
| <b>Status</b>              | <i>Final</i>                               |
| <b>Author(s)</b>           | <i>Elsa Zipstein, SYSTRAN S.A., France</i> |



***eContentplus***

This project is funded under the *eContentplus* programme<sup>1</sup>,  
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

---

<sup>1</sup> OJ L 79, 24.3.2005, p. 1.

### Document Version Control

| Version | Date             | Change Made (and if appropriate reason for change) | Name and Institution of Commentator(s) or Author(s) |
|---------|------------------|--|---|
| 1.0     | 30 November 2008 | First Draft  | EZ, SYS   |
| 1.0     | December 2009    | Comments   | TP, AIP<br>NK, IMI<br>HS, DSP                       |
| 1.1     | 17 December 2008 | Final version                                      | EZ, SYS   |
|         |                  |  |   |

### Document Review

| Reviewer | Institution | Date and result of Review                       |
|----------|-------------|---|
| TP       | AIP         | 18/12/2008<br>Accepted as technical coordinator |
| ZU       | NKP         | 19/12/2008<br>Accepted for submission to the EC |

### Document Signature/Approval:

Before the table of contents each document is to contain an approval signoff form.

| Approved By (signature) | Date |
|-------------------------|------|
|                         |      |

| Accepted by at European Commission (signature) | Date |
|--|------|
|  |      |

## **1 Executive Summary**

The Work package 6 aims to the integration of a tailored multilingual module via a Enrich user friendly sophisticated access.

Based on SYSTRAN's machine translation technology, this module will provide also terminology extraction and machine translation customization tools for the construction and retrieval of personalised metadata within the aim to create new multilingual digital documents and multilingual ontologies in Czech, Polish, Spanish, Portuguese, German, Italian, English, French, Danish, Hungarian, Russian and Serbo-Croatian.

One of the first objectives of that workpackage is to tailor the system so as to generate & provide personalisation resources for the translation of specific fields of the manuscripts description and/or documentation. In this deliverable we present the translation tools and linguistic resources integrated to the Enrich personalized translation interface.



**CONTENT**

**1 EXECUTIVE SUMMARY ..... 3**

**2 INTRODUCTION ..... 5**

**3 ENRICH PERSONALISED TRANSLATION INTERFACE OVERVIEW ..... 8**

**4 ENRICH PERSONALISED TRANSLATION INTERFACE TOOLS ..... 9**

**4 CONCLUSIONS ..... 13**

## 2 Introduction

Content partners have provided us with feedback regarding the usability of a machine translation web service for the automatic translation of manuscripts description and the integration of the multilingual module in third applications.

This document aims to present the personalised translation interface tailored by SYSTRAN for Enrich needs accessible at <http://enrich.systran.fr/>.

Enrich Personalised Translation Interface is a comprehensive solution able to meet the full range of translation needs on the project's scale. Supporting unlimited Enrich user access and millions of translations per day, the solution is deployed easily at the servers of several content partners. It translates all types of documents and files formats, and also seamlessly integrates with Enrich application hosted at Manuscriptorium providing machine translation for manuscripts description personalized according corpus analysis conclusions and validation tests outcome.

Efficient access to the solution's rich features is made available through a Web-based interface letting Enrich users to instantly translate texts, emails, Web pages, RSS feeds, and a wide variety of file formats while preserving the original document layout in the translations. Common to both interfaces is a dictionary search tool that provides Enrich users with instant access to alternative meanings of selected terms from several built-in dictionaries as well as Enrich dictionaries.

Enrich users can improve and control quality translations with Enrich user-defined profiles made of up unique options that are applied to specific translation tasks calling the first version of SYSTRAN Translation Stylesheet developed in already to provide parameterized translation according the metadata structure upon content partners approval and validation tests.

Manuscriptorium added easily translation functionalities to its applications through the open APIs in Enrich Machine Translation platform. The Administration tools allow for easy maintenance, activity reporting, and managing Enrich user permissions.

Following the lines of the Description of Work, ENRICH, Annex 1, the main objectives of the task 6.2 is described in the Work package 6.

### Work package Description

|                              |   |                    |          |                  |           |
|------------------------------|---|--------------------|----------|------------------|-----------|
| <b>Work package number :</b> | <b>6</b>  | <b>Start date:</b> | <b>3</b> | <b>End date:</b> | <b>24</b> |
| <b>Work package title:</b>   | <b>Multilingual and Enrich user friendly sophisticated access</b> |                    |          |                  |           |

### Objectives

This work package aims to the integration of a multilingual module via a Enrich user friendly sophisticated access:  
 multilingual search application, multilingual forums, and multilingual ontology editor.  
 Based on SYSTRAN's machine translation technology, this module will provide also terminology extraction and machine translation customization tools for the construction and retrieval of personalised metadata within the aim to create new multilingual digital documents and multilingual ontologies in Czech, Polish, Spanish, Portuguese, German, Italian, English, French, Danish, Hungarian, Russian and Serbo-Croatian.

### Description of work

**Work package leader: SYS****Task 6.1 Multilingual access development (m0-m12)****Task leader: SYS****Task participants: NKP, AIP**

The project will provide two types of multilingual access:

- via the API integration in the data retrieval interface associated or independent of a multilingual search.
- a dedicated translation interface where ENRICH expert Enrich users can fine-tune dynamically the machine translation tools thanks to adapted linguistic tools for terminology extraction and translation post-editing and customization. The parameters and resources constructed will be automatically taken into account by the API in the access presented above

**Task 6.2 Translation Stylesheet design and use (m3-m24)****Task leader: SYS****Task participants: NKP, AIP, CCP, NFC, NLF, IMI, ULV, SAM, CSH, DSP, NLI, BNE, BUTE, ULW**

Activities:

- analysis of heterogeneity of metadata regarding machine translation.
- implementation of STS exploiting metadata information.
- cross-language validation of STS, optimization of translation parameters.

As far as the Metadata translation module implementation is concerned SYSTRAN will provide a fully customized Translation Stylesheet.

SYSTRAN Translation Stylesheets (STS) use XSLT to drive and control the machine translation of XML documents (native XML document formats or XML representations — such as XLIFF — of other kinds of document formats). STS will provide a simple way to indicate which part of the document text is to be translated, and will enable the fine-tuning of translation, especially by using the structure of the document to help disambiguate natural language semantics and determine proper context. Thanks to STS machine translation is considered as part of the authoring and publishing process: source documents can be annotated with natural language mark-up produced by the author, a mark-up which will be processed by STS to improve the quality of translation, the gateway to the automatic publishing of a multilingual website from a monolingual (annotated) source. The mechanism is implemented through XSLT extension functions for consulting and for setting linguistics options in the translation engine. SYSTRAN will deliver this xslt file in order to fine-tune the system according to the ENRICH xml data elements.

**Task 6.3 VICODI implementation (m6-m24)****Task leader: SYS****Task participants: NKP, AIP,**

- definition and homogenization of initial ontology applicable for this project
- specification of Enrich user-friendly web-interface for visualization of multilingual ontology - special interface for modification
- implementation of the web-interface

Based on previous experience in the visualization and contextualization of digital content (IST project VICODI) SYSTRAN technology has been implemented for the construction of multilingual ontologies. The Research Center for Information Technologies (FZI) constructed multilingual ontologies available under GNU Free Documentation License (FDL) thanks to the EU-funded IST project Vicodi <http://www.vicodi.org/> ). Enrich will implement and use VICODI ontologies for the contextualization of the digital content.

**(Inter-) Dependencies, milestones<sup>1</sup> and expected result**

Based on the ENRICH Corpus Analysis (Month 6) SYSTRAN will build ENRICH Translation Stylesheet. After a quality assessment procedure and based on the evaluation results (Month 20) SYSTRAN will proceed to the finalisation of ENRICH Translation Stylesheet (Month 24). The WP depends mainly on the results of WP3, but is also interrelated with WP4. The feedback is necessary from WP7.

**Deliverables**

- D 6.1 ENRICH Corpus Analysis report (Month 6), responsible partner: SYS
- D 6.2 Personalised Translation Interface delivery report (Month 12), responsible partner: SYS
- D 6.3 ENRICH Translation Stylesheet delivery report (Month 24), responsible partner: SYS
- D 6.4 Vicodi Ontologies implementation report (Month 24), responsible partner: SYS

### **3 Enrich Personalised Translation Interface Overview**

The translation tools included in Enrich Personalised Translation Interface can be used in conjunction with a wide variety of word processing, spreadsheet, email, internet browser, and other software applications. Specifically, the tools include:

- **Text Translation**
- **Web Translation**
- **File Translation**
- **RSS Translation**
- **Email Translation**
- **Dictionary Search**

With Enrich Machine Translation platform, translation processes, dictionaries and other translation resources are all stored and managed on a single server. This allows multiple Enrich users to draw on a common set of resources in performing their translation tasks, thus ensuring consistent Enrich-wide translation resources and standards.



## 4 Enrich Personalised Translation Interface Tools

Once installed, all users can access the Enrich Personalised Translation Interface Online Tools to perform numerous translation tasks. All access rights can be changed by your SYSTRAN administrator. The look-and-feel of the interface is based on the ENRICH url design.

### Translating Via Web Browser

Enrich users can employ any web browser (Internet Explorer, Firefox) to easily translate web pages, input text, and a large array of different file types (.txt, .rtf, .doc, .docx, .xlsx, .pptx, pdf, .or .html). Also, Enrich Personalised Translation Interface provides Enrich users with RSS feeds translation and email translation.

### Exploiting SYSTRAN Dictionary Resources

The web interface allows Enrich users to draw on various resources to greatly improve translation quality, including built-in SYSTRAN general and specialized Dictionaries, Enrich Dictionaries (UDs), Translation Memories (TMs), and Normalization Dictionaries (N Ds).

- Enrich user Dictionaries (UDs) consist of terms that are specific to content, as well as terms that are designated Do Not Translate (DNT) like those created from the Vicodi and Medieval Ontology instances.
- Translation Memories (TMs) are databases of pre-translated sentences. These sentences are matched to the source text during the translation process, and when appropriate they automatically replace previously translated sentences.
- Normalization Dictionaries (NDs) provide a way to regulate text both before and after translation, serving as a means for standardizing terminology and expanding abbreviations and acronyms. NDs can also be used to standardize spellings and to establish common terminology (e.g., instances in which two or more words are used for the same concept).

### Logging Into Enrich Translation Interface

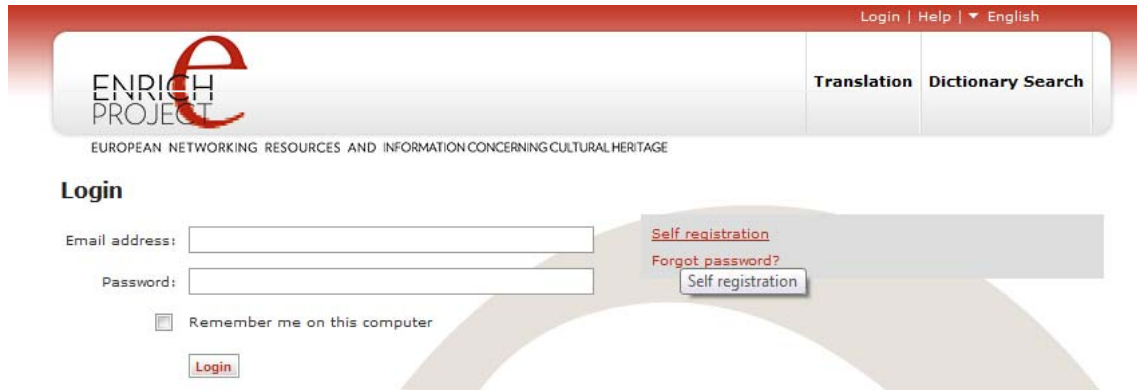
While it is not necessary to log into Enrich Personalised Translation Interface to make use of its online translation tools (with the exception of the **RSS Translation** tool, the use of which does require login), the functionality of the tools is restricted. For instance, without logging in, Enrich users cannot create or employ Translation Option Profiles.

Enrich users who are present in the system can quickly and easily log in at any time.



## Self Registration

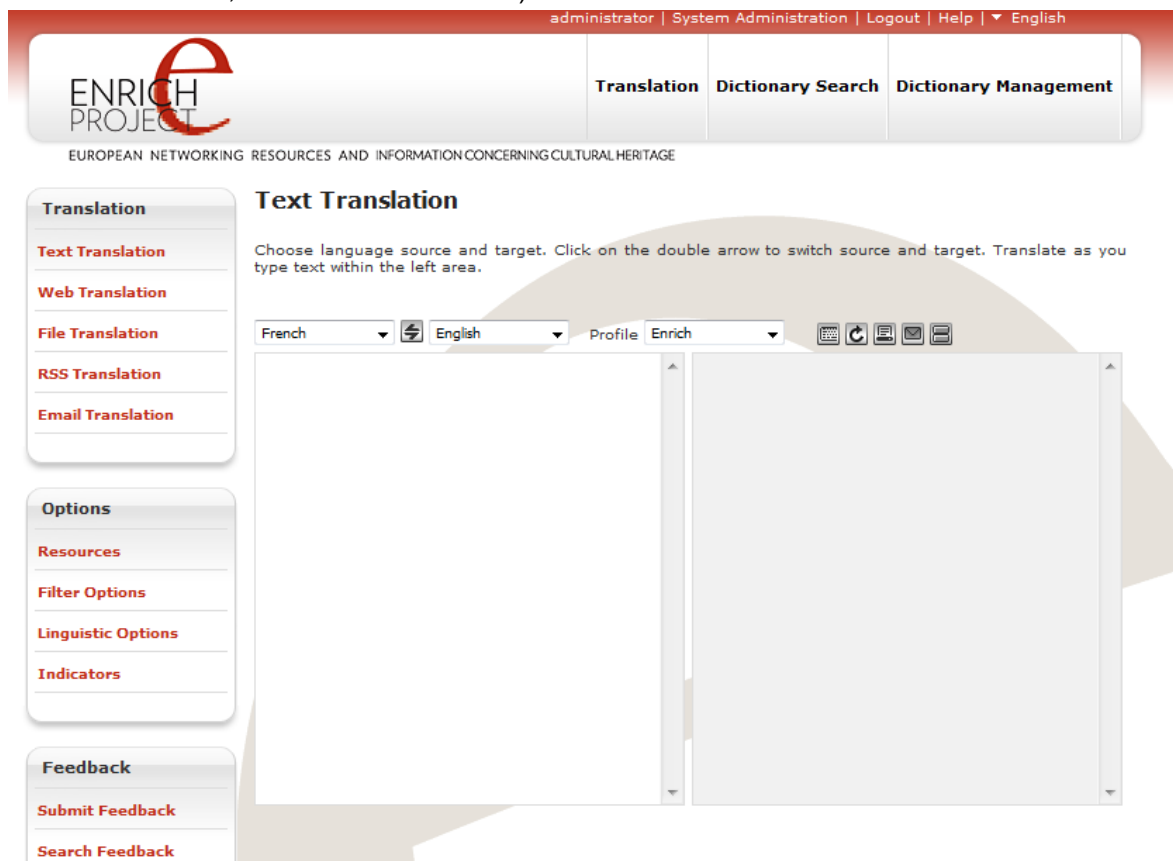
With Enrich Machine Translation platform’s new **Self registration** feature, Enrich users can effectively gain immediate access to the system and circumvent administrative wait time.



## Using the Translation Tools

The **Translation** menu bar link is the entry point to the Enrich Personalised Translation Interface toolset, with the exception of the **Dictionary Search** function. It is from this link that the translation tools are accessed, and from which all translations are configured and initiated.

By default, the **Text Translation** tool is displayed when the Translation service is first accessed, however Enrich users can easily open any of the other four (4) Translation tools via the ever-present left-hand **Translation** menu (**Web Translation**, **File Translation**, **RSS Translation**, and **Email Translation**).



## Translating Input Text

Use the **Text Translation** tool to translate any text that is entered into the tool's left-hand menu. The tool works in an active fashion, translating "as you type" and thus it is not necessary to initiate translation.

## Translating Web Pages

The **Web Translation** tool offers Enrich users a means for translating entire web pages while retaining the formatting and functionality of those pages. In addition, the tool features *fluid navigation*, carrying translation through to any web pages accessed via links in a translated page.

## Translating Files

The **File Translation** tool provides a quick and easy way to translate a wide variety of different file types, including .pdf, .txt, .rtf, .doc (Microsoft Word), .html, and the primary Microsoft Office 2007 file formats (.docx, .xlsx, and .pptx). The tool can be directed to deliver translated files via email as well, thus allowing Enrich users to launch numerous large document translations and receive notification of their completion.

Enrich users must login to Enrich Personalised Translation Interface to enable results emailing capability.

## Translating RSS Feeds

One of the two tools offered in Enrich Personalised Translation Interface is **RSS Translation**, a dynamic translation tool that removes the natural boundaries imposed by the source language of RSS feeds.

Enrich users employ the **RSS Translation** tool to generate custom RSS feeds which can then be added to RSS readers (commonly known also as feed readers or aggregators). Once added, these custom-generated feeds display the information correlating to each specific feed, automatically translated, thus allowing Enrich users to subscribe to and benefit from RSS feeds from websites that publish in a variety of languages.

## Translating Emails and File Attachments

The Email Translation tool is one of the two new tools offered in Enrich Personalised Translation Interface (RSS Translation is the other). It provides Enrich users with a means for translating any text contained in an email, as well file attachments of the following file types: .pdf, .txt, .rtf, .doc (Microsoft Word), .html, and the primary Microsoft Office 2007 file formats (.docx, .xlsx, and .pptx). If Email Translation is not the active translation tool, click Email Translation in the left-hand Translation menu.

## Using the Dictionary Search Tool

With the **Dictionary Search** tool Enrich users can search for the meaning of terms in any or all of the following dictionaries:

- SYSTRAN Dictionaries
- SYSTRAN Specialized Dictionaries
- Enrich user Dictionaries (UDs)
- Additional Dictionaries, depending on the selected language pair

## Using Translation Options

Translation Options enable Enrich users to fine-tune the manner in which Enrich Personalised Translation Interface performs translations.

From a wide variety of linguistic, grammatical, and formatting options, Enrich users who are logged into the system can construct option sets – or *Profiles* – which can then be applied to translation tasks. Before Enrich users can make use of the various Translation Options SYSTRAN created the Enrich translation profile integrating SYSTRAN Translation Stylesheet according to the Corpus Analysis outcome and validation tests as well as the name entities Do Not Translate dictionaries composed by Medieval ontology instances provided by NKP and Vicodi Ontologies;

Enrich users can perform an array of different actions on their Profiles from any **Options** screen, including:

- Changing Profile names
- Deleting Profiles
- Importing Profiles offered by other system Enrich users
- Export Profiles for use by other system Enrich users
- Configure Profile permission levels

### Using Profiles

Once created, the Enrich profile can be used within any of the Enrich Personalised Translation Interface translation tools via the **Profile** control.

## Using the Feedback Tool

Enrich Personalised Translation Interface enables Enrich users to submit feedback on translation quality, and thus raise issues and propose alternative translations. Once submitted, this feedback can be reviewed by a Dictionary Manager, who will thereafter be responsible for following its life cycle.

The Dictionary Manager determines the validity of submitted feedback. They can then quickly update the linguistic resources with proposed terminology entries and/or forward the feedback to SYSTRAN.

## Reviewing Feedback

The Search Feedback utility can be granted to Enrich users by a system administrator. Once in place, Enrich users access this tool via a **Search Feedback** command present in the lefthand Feedback menu.

The **Feedback Search Results** page will display, offering a list of feedback matches, as well as a number of tools that can be used to take action on the results selected.

## Acting on Feedback

A number of different actions can be performed on a feedback received, including:

- Exporting the selected feedback items
- Replay the translation of the selected feedback items
- Sending the feedback to SYSTRAN support
- Providing additional comments
- Changing the feedback status

Each feedback assumes **New** status at submittal. Using the **Status** radio button, this can be changed to **Open** (indicates that the feedback still requires analysis), **Accepted** (indicates the validation of the feedback), or **Invalid** (indicates rejection of the feedback). Once a feedback is validated and the resulting correction is available on the production server its status changes to **Validated**. All modifications will occur following submittal.

## 4 Conclusions

ENRICH personalized translation interface <http://enrich.systran.fr/> allows to Enrich users to benefit from personalized translation interface calling specific to manuscripts description linguistic customization resources. At the same time Enrich users have the possibility to interact automatically with the system updating the translation profiles in all language pairs updating on-the-fly the customization resources that are being applied via the API request from Manuscriptorium third applications for the best possible translation of Manuscripts description.