

# PROGRESS REPORT

## PUBLISHABLE SUMMARY

Grant Agreement number: 270915

Project acronym: BOLOGNA

Project title: Bologna Translation Service

Project type:  Pilot A  Pilot B  TN  BPN

---

Periodic report: 1<sup>st</sup>  2<sup>nd</sup>  3<sup>rd</sup>  4<sup>th</sup>

Period covered: from March 2011 to February 2012

---

Project coordinator name, title and organisation:

Heidi Depraetere  
Director  
CrossLang

Tel: +32 498 90 65 91

Fax: +32 9 335 22 01

E-mail: [heidi.depraetere@crosslang.com](mailto:heidi.depraetere@crosslang.com)

Project website address: <http://www.bologna-translation.eu>

## Project Objectives

There is a continuing increasing need for educational institutes to provide course syllabi documentation and other educational information in English. Access to translated course syllabi and degree programmes plays a crucial role in the degree to which universities effectively attract foreign students and, more importantly, has an impact on international profiling. To present all education information in English is a major challenge for most higher education institutes. Figures and trends show that investment in traditional human translation services is prohibitive, consequently course materials and degree programmes are often provided in the local language only.

The Bologna project aims at providing a solution to this problem by offering a low-cost, web-based, high-quality machine translation (MT) service, the Bologna Translation Service (BTS). The project will cover 9 languages (Chinese, Dutch, English, Finnish, French, German, Portuguese, Spanish and Turkish). Existing statistical as well as rule-based solutions will be tailored to the educational domain and offered through the translation portal. The service will be accessible from a web-interface and from the working environments of the universities, typically from the database where course syllabi and degree programmes are produced and edited.

## Project Progress

Key areas in the first year include building machine translation systems for all languages in the project, developing the web application that will serve as the front-end for end users and connecting the machine translation systems to the web application.

## Translation Systems

In the course of the project, different types of systems with different levels of quality will be built. The best performing systems will be complemented with language- and technology-specific automatic post-editing modules to further refine the translation output.

Progress to date includes baseline SMT systems for all language pairs which will serve as a benchmark against which future progress will be measured and the creation of RBMT systems for language pairs Spanish—English and Portuguese—English. For a limited set of language pairs (French—English, Spanish—English, and German—English) we have also already built advanced SMT systems trained with in-domain data. For Spanish—English, the only language pair for which we have both an advanced SMT system and a customized RBMT system ready at this time, system combination experiments have been performed in an attempt to further boost the translation quality. Automatic evaluation scores for the advanced SMT systems and combined system for Spanish—English are presented in the table below.

Language Pair	Advanced and Combined		
	BLEU	METEOR	TER
German—English (advanced)	43.4	35.0	48.4
French—English (advanced)	51.1	42.9	32.9
Spanish—English (advanced)	34.8	37.7	45.1
Spanish—English (combined)	35.9	37.6	44.4

**Table 1: Automatic Evaluation Scores (Advanced and Combined Systems)**

We have also demonstrated the effectiveness of the APE component in the first year of the project by building a proof-of-concept system for German—English. For this language pair we built a provisional advanced system with the in-domain data that was available at the time and derived automatic post-editing rules from the output of that system. The outcome of this experiment has shown that further quality improvements can be expected from automatic post-editing in the second year of the project, when we will apply this methodology to all language pairs.

## Web Application

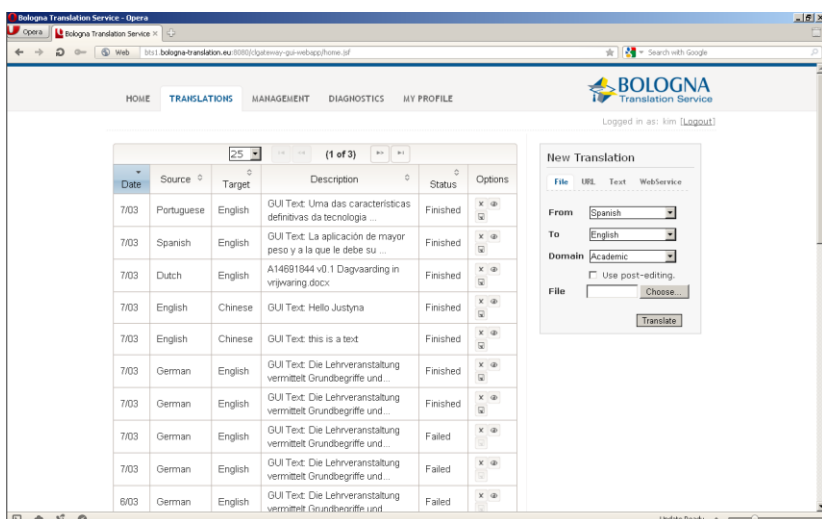


Figure 1: The Bologna Translation Service Portal

In parallel with the building of the translation systems we have been developing the web application that will serve as the front-end that users will interact with. A first beta release of the Bologna Translation Service (version 0.3) is now available. Focus of the development to date has been the front-end that the end-user will interact with (the User and Translation Dashboards) allowing us to demonstrate the service and give user group members early access to the service in view of soliciting feedback that can be incorporated during the second project year.

Basic components of the Management & Diagnostics Dashboards have also already been developed and will be further developed together with the Accounting Dashboard during the second year of the project.

Figure 2 provides an overview of all components to be developed. Green lines indicate components and connections that have already been implemented. Red lines indicate planned components and connections.

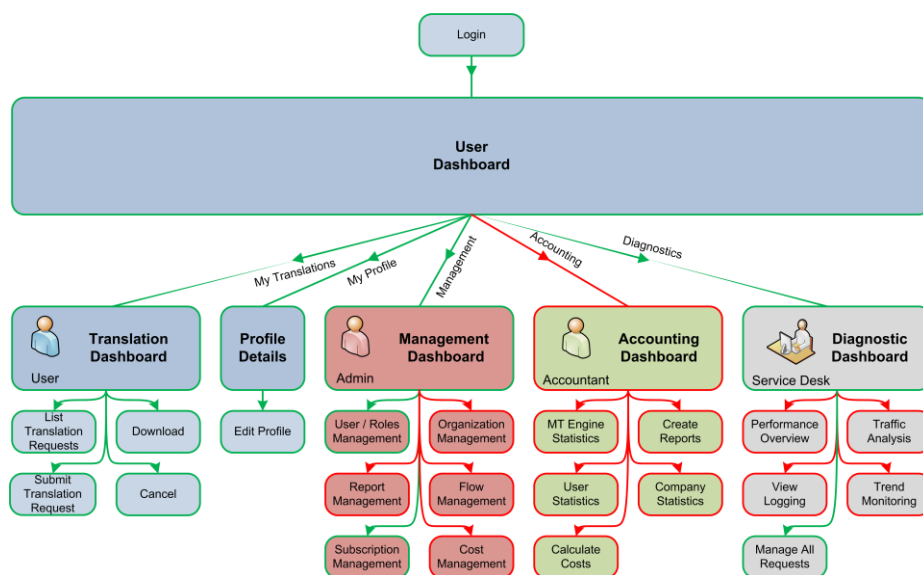


Figure 2: Development Status of BTS Web Application Components

## Integration

During the first year of the project, significant progress has been made on integrating the machine translation back-end systems with the web application front-end, i.e. the CrossLang Gateway and the Convertus Syllabus Translator. Furthermore, all SMT engines that have been built during the first project year and the proof-of-concept Automatic Post-editing module has been integrated into the platform.

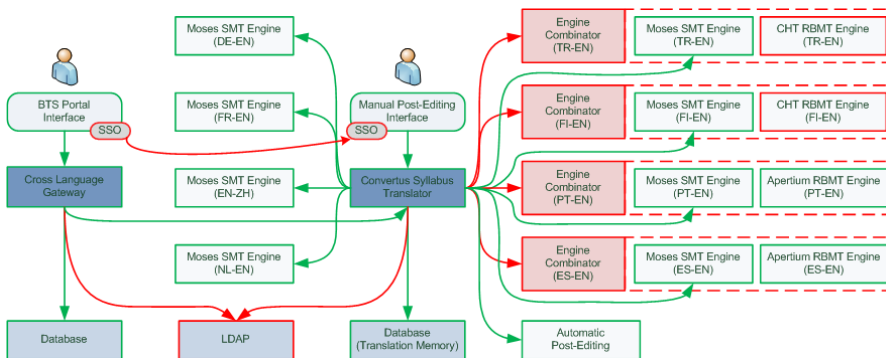


Figure 3 provides an overview of all components to be integrated into BTS. Green lines indicate components and connections that have already been integrated. Red lines indicate planned components and connections.

Figure 3: Integration Status of BTS Components

## Other Highlights

### Dissemination

Apart from defining a dissemination strategy and creating a dissemination plan Bologna has already been presented at a number of conferences across Europe and has had two peer-reviewed conference papers accepted. The events attended through Year 1 are listed in the table below.

Date	Event	Location
May 2011	EAMT Conference (European Association for Machine Translation)	Leuven, Belgium
June 2011	METANET-FORUM	Budapest, Hungary
June 2011	Localization World	Barcelona, Spain
September 2011	EAIE Conference (European Association for International Education)	Copenhagen, Denmark
November 2011	Taal in Bedrijf	Rotterdam, Holland
November 2011	Translating and the Computer	London, UK
March 2012	INTED2012	Valencia, Spain

Table 2: Overview of Attended Conferences

Collaterals including a project fact sheet, a flyer, a poster and a number of press releases have been produced to help raise awareness of the Bologna project. The project website ([www.bologna-translation.eu](http://www.bologna-translation.eu)), which has been live since March 2011, is Bologna's main dissemination channel.

A number of dissemination activities have already been planned for Year 2 including invited presentations at LREC and at a MosesCore workshop hosted by TAUS, and participation in the 24<sup>th</sup> Annual EAIE Conference.

## Project Meetings

The Bologna consortium has held a number of on-site meetings at various stages to address technical issues and discuss progress. These meetings have proven to be very productive and have helped speed up progress as well as encourage cooperation between partners. The pictures below provide an impression of the Year 1 meetings.



*Kick-off Meeting, March 2011, Gent, Belgium*



*Status Meeting, June 2011, Uppsala, Sweden*



*Advisory Board Meeting, February 2011, Gent, Belgium*

## Advisory Board

The consortium is proud to have three highly experienced specialists serve on our advisory board.

- Dr. Fred Hollowood, Symantec, Ireland
- Prof. Björn Gambäck, Swedish Institute of Computer Science, Sweden
- Dr. Lucia Specia, University of Sheffield, UK

The project consortium held its first Advisory Board meeting in Gent with participation of Dr. Fred Hollowood and Dr. Lucia Specia who gave invaluable feedback and advice.

## Outlook

During the second year of the project we will continue to develop and refine the MT engines with a keen focus on finalising the integrated architecture as well as evaluation. The BTS platform will be opened to interested parties for testing purposes. Dissemination activities will focus on expanding our user group and collect feedback from our prospective users.