

# TRANSACCOUNT PROJECT MLIS - 5016 MANAGEMENT REPORT No 2, as of December 31, 2000

---

**General comment:** Till end of June 2001 Eurosources was mainly responsible for managing the project and delivering reports, cost statements etc. to the CEC. This was agreed at the beginning of the project and realised as a subcontract with Alpnet Technology GmbH. Due to the bankruptcy of Eurosources with end of May 2001 no documents have been delivered so far to the CEC, esp. with regard to period 2 of the project. The decision is now that this will be done again by the main project co-ordinator – Alpnet Technology GmbH. Unfortunately several documents are kept at Eurosources side and cannot be accessed anymore.

## Content

I - CHANGES RELATING TO TRANSACCOUNT'S PARTNERS .....	2
II - TRANSACCOUNT'S PARTNERS ACTIVITIES AS AT December 31, 2000 .....	2
EU (EUROSOURCES SA).....	2
EP (Alpnet Technology GmbH).....	2
SY (SYSTRAN SA) .....	3
AM (Amyot - Exco).....	4
III - TRANSACCOUNT'S PROJECT PROGRESS AS AT December 31, 2000 .....	5
1 - Deliverables produced .....	5
2 - Deliverables to be produced .....	5
3 - Information and dissemination .....	6
4 - Cost Statements .....	6

## I - CHANGES RELATING TO TRANSACCOUNT'S PARTNERS

- a) **EU - Eurosources:** Eurosources went bankrupt and terminated its business with end of May 2001. No information is now available any more from Eurosources, esp. no information about Eurosources progress made in the 2 milestone
- b) **EP - Electronic Publishing Partners:** The company has a new business name: Alpnet Technology GmbH. This is reflected in the 1.st amendment to the contract.
- c) **AM - Amyot Exco:** Amyot Exco expressed its wish to leave the project and de-facto left the project end of December 2000. Because of that several deliverables are missing from their side. This deliverables will be brought into the project by a new partner.

## II - TRANSACCOUNT'S PARTNERS ACTIVITIES AS AT December 31, 2000

This section describes the activities performed by the partners in TransAccount project, for the period from July 1, 2000 to December 31, 2000. Under each partner's activities, we have mentioned the WP(s) in which those activities have been performed together with the corresponding deliverables.

### EU (EUROSOURCES SA) WP1 - WP5

Due to the bankruptcy of Eurosources no information is available.

### EP (Alpnet Technology GmbH) WP 1 - WP5

- a) Adaptation of the web server for TransAccount ([www.transaccount.org](http://www.transaccount.org))
- b) Installation of a TransAccount Globelix server on a Solaris machine accessible for the partners in the project. This machine is intended to serve as reference implementation till a Windows based system is available. The machine can also be used to create the TM database. It gives access both to an advanced alignment tool and an importer into a translation memory database. Test suites can be run by supplying new documents and comparing team against the TransAccount database.
- c) Developing a Machine Translation Interface between TransAccount Globelix and Systran
- d) Porting of server kernel to Windows NT to allow maximum access. This is based on Perl and the Apache Web server. It allows the user either to run it as a standalone version or

as a service via the intra or internet. This includes also access to the TransAccount form parser for pre-processing financial accounts.

- e) Starting researching and developing a form based account parser
- f) Start of developing a TransAccount specific GUI based on Globelix
- g) Evaluation of the specifications of the other partners

## **SY (SYSTRAN SA) WP3 - WP4**

### **Personalization of translation system.**

On the basis of frequency analysis done on Milestone 1, the following customization protocol has been applied concerning the textual part of balance sheet report.

- > frame for integration of important specialized terminology
- > grammatical specialization : ie deep financial structure description
- > stylistic customization

### **I) Specialized terminology :**

According to coverage and frequency analysis, the following kind of terminology needs to be introduced in system in order to reach a correct level of description/understanding.

This terminology includes :

- a) general technical finance (part built on milestone I), and connex terminology (company related information, currency description, verb for representing variation of values (that are : sum, nb of employees, expansion rate)
- b) Balance sheet specialized terms (probably a minimal set of 10 000 entries will be needed)
- c) Database of entity's name (markets, companies, bank, investment group...)

A database of type c) has been extracted from textual resource, and the a) dictionary relative to previous work has been coded.

In order to integrate terminology from the partners, the coding environment has been adapted to "financial term" (ie : to introduce an entry in a translation dictionary : this entry needs to be fully described on a syntactic/semantic and morphological point of view.

Internal tools (AllCoding) developed by Systran, allow to propose to linguists approval automatic coding for any entry. Nevertheless, the "generic" coding is not adapted to any kind of vocabulary (for example "open market" by default encodes as a location, and not as an abstract concept : this has a direct incidence on the translation of sentences as : on the open market (=au marché libre vs. sur le marché...).

### **II) Grammatical specialization :**

This category includes all linguistic pattern whose scope is too complex to be introduced in "dictionaries".

We have used for this purpose, new finite state representation. This representation is based on a very accurate monolingual description (English), built with a bootstrap principle, on the reference corpus extracted in Milestone I. The choice of this modern internal solution guarantees moreover the reusability of this description in other contexts (financial controlled language for instance) and the incrementability of the ressources even by a non Systran expert.

A new transfer module based on bilingual "alignment" theory has been designed for this purpose.

### **III) Stylistic customization :**

A study on stylistic customization of balance sheet writing is on progress. This part is necessary to guarantee that the choice of general grammatical structures, and transfer equivalent produce target document with appropriate style. Since additional terminology will imply some additional rules, no test stage of this module has been run so far.

### **AM (Amyot - Exco) WPO**

Due to the intention of Amyot leaving the project and de-facto leaving with end of December 2000) no activity can be recorded from Amyot.

### III - TRANSACCOUNT'S PROJECT PROGRESS AS AT December 31, 2000

#### 1 - Deliverables produced

The following deliverables have been due in this period.

##### Month 12:

- (a) Accounting Text Database (multilingual and aligned), datafile. This deliverable was in the hands of Amyot / Eurosources. No data file has been delivered so far from those partners. See also the comment at the beginning of the document.
- (b) Form based parser/generator for accounts for US and FR accounting systems, software. A report has been produced about the system architecture. The potation of relevant features of the Globelix software as the underlying software system has started and makes good progress. A version of the software was demonstrated at the review in June 2001.

#### 2 - Deliverables to be produced

The [www.transaccount.org](http://www.transaccount.org) web site is available containing all the deliverables existing. This is part of the management activities and updated by Alpnet. The site can be accessed by entering user name and password sent to all parties. The deliverables due in this progress report period are marked bold in the table below in chapter 2.

The following table shows the deliverables available.

No	Specification Document	Due to inMonths	Responsible	Deliverable Reference	Comments
WP 0/1	<a href="#">Specification of the user interface</a>	5	Amyot Exco	D0a	Version 1.0
WP 0/2	<a href="#">Content of Amyot Exco Accounting Database (Deliverable D0h)</a>	5	Amyot Exco	D0h	Version 1.0 - CD-ROM based data files
<b>WP 1/3</b>	<b>Accounting text database</b>	<b>12</b>	<b>Amyot Exco</b>	<b>D1a</b>	<b>not available</b>
WP 1/4	Multilingual terminological database	14	Eurosources	D1c	Geneter based data file
WP 1/5	<a href="#">Knowledge Base Design</a>	14	Eurosources	D1f	Version 1.0
<b>WP 2/6</b>	<b><a href="#">TransAccount Architecture Specification</a></b>	<b>12</b>	<b>Alpnet</b>	<b>D2a</b>	<b>Version 1.1</b>
<b>WP</b>	<b><a href="#">Form Based Parser</a></b>	<b>12</b>	<b>Alpnet</b>	<b>D2a</b>	<b>Version 1.2</b>

<b>2/8</b>					
WP 2/9	Accounting Thesauri for US and FR systems	14	Eurosources	D2c	not available
WP 2/10	<a href="#">Form Based Translator - User Interface Manual</a>	18	Alpnet	D2f	Version 1.3
WP 3/11	<a href="#">Sysran Technical Report</a>	(18)	Systran	D3.x	Version 1.0
WP 5/1	<a href="#">Detailed Workplan V1.0</a>	1	Eurosources	D5a	Initial version

### 3 - Information and dissemination

No special actions have been done in this period

### 4 - Cost Statements

The cost statements corresponding to the works and progress are attached to the present management report. They reflect the work done and the progress of the project.

Dr. Klemens Waldhör  
Alpnet Technology GmbH  
TransAccount Project Co-ordinator