

**DIANA JANK
ANDREAS TAES
PIET VERLEYSEN**

Computerised Workflow in the Translation Service

and its transportability to translation centres in the
associated countries

1 Introduction

This paper contains a description of the current computer environment and working methods of the European Commission's Translation Service. A step-by-step account of the workflow is followed by a series of factsheets describing the principal tools used by managers and translators to manage the workflow, share resources and promote greater consistency of terminology in EU texts. Each factsheet includes the name of a person to contact for further information.

A brief summary examines the relevance of these tools and procedures to the translation units in the Associated Countries in the light of their primary tasks of coordinating the translation of the *acquis* into their national language, maintaining an accurate and up-to-date record of those translations and helping consistent and harmonised terminology.

2 Computerised workflow in the Translation Service

2.1 General architecture of the information systems

The information systems of the Translation Service are organised around three main categories of data:

- a *administrative data* used to monitor the progress of translation requests (Winsuivi, LTD) and to manage human and documentary resources;

- b *documents* in various stages of preparation (held in working directories on NT servers) and archived (originals and final translations in read-only format on SdTVista), and general reference documents available in other databases (CELEX, SdTNet, EuropaPlus, Europa, etc);
- c *linguistic resources* in the form of terminology (EURODICAUTOM and MULTITERM), translation memories (EURAMIS) and dictionaries for machine translation (SYSTRAN) or general use (CD-ROM).

2.2 *Workflow and tools*

2.2.1 Action by the requesting department

The requesting department decides which documents need to be translated, into which target languages, and at which stage of the drafting process. It also assesses the quality of translation it requires. If it needs a rough draft for discussion, or simply to understand the general tenor of an original text drafted outside the Commission, it may e-mail the document directly for machine translation (SYSTRAN), or request a rapidly post-edited machine translation.

If it wants a high-quality translation it allocates a translation request number and submits its request (Poetry¹) to the Translation Service, attaching the original document for translation and any reference documents.

2.2.2 Receipt of translation request by the planning office

The translation request is registered (WinSuivi) and the source document and any reference documents are placed on the document server (SdTVista). A certain amount of preparatory work - preprocessing, terminology searching, documentary support - may be done at this stage.

2.2.3 Translation

When the work of translation is started in the relevant target-language unit, progress is recorded in LTD and WinSuivi. If the unit head decides to send the document for freelance translation, this is recorded in the TREFLE system. Translators have a number of tools at their disposal (Translator's Workbench, EURAMIS, EURODICAUTOM, MULTITERM, CELEX, various

¹ Processing of Electronic Translation Request.

CD-ROM, etc) to help with the task of translation itself. The majority of documents are translated and revised in the same unit.

2.2.4 Logging out the translation

The finished translation is archived (SdTVista), logged out (WinSuivi), and e-mailed to the requester. Freelance translations are logged out in TREFLE. New linguistic data (terminology and translation segments) may be dispatched to EURAMIS. If the requesting department subsequently makes changes to the translation, it is desirable for a copy of the final version to be sent to the Translation Service via Poetry, so that the final text can be included in SdTVista and EURAMIS.

2.2.5 Publication

This stage of the document production process is being increasingly rationalised. Specific systems are used for the Bulletin and General Report (tags) and the Budget (SEI-BUD²). All these procedures confirm that the management of structured documents greatly facilitates the work of document creation, translation and publication.

2.2.6 Distribution

At present the main instruments for electronically distributing full-text documents produced by the institutions are the Europa Web server (carrying CELEX and other documentary databases) and CD-ROM (*Official Journal*, L and C series). The EUDOR³ database contains the electronic archive of the *Official Journal* in image format.

3 Factsheets

3.1 *WinSuivi and Local Translation Database (LTD)*

² Système Editorial Informatisé pour le Budget.

³ <http://www.eudor.com/>.

3.1.1 Purpose

To allow tracking and workflow management of all translation requests.
To allow production of statistical information.

3.1.2 Content

All translation requests and their status. The data include details of the translation request (requesting department, year, number, version, part, source and target languages, title of source document), the author, the translator, various dates and organisational data.

3.1.3 Technical environment

Oracle database server (version 7) on SNI Unix (Pyramid) with PowerBuilder (version 4/5) user interface. ODBC link with MS-Access local database on Pentium PCs (Windows 3.1 and NT).

3.1.4 Organisational aspects

The central database is managed by the planning departments and secretaries in the translation units. It is accessible to all Translation Service staff on request.

3.1.5 Contact

PIET VERLEYSEN
Tel: +352-4301-34356
Fax: +352-4301-33370
E-mail: jean-pierre.verleysen@sdt.cec.be

3.2 *SdTVista*

3.2.1 Purpose

To allow full-text storage and retrieval of source documents and translations in different formats.
To allow full-text searching for terms and expressions used in context (in all documents or subsets).

3.2.2 Content

Source documents, their translations and some general-purpose reference documents. Search engine allows search-and-display (bilingual) and copy-and-paste functions for all users. Documents selected can be downloaded to the user's workstation.

The database contains the document text and title, details of the translation request (requester, year, number, version, part, source and target languages), the author, the translator, and the subject.

3.2.3 Technical environment

Document Server: Fulcrum SearchServer (version 3.5) on SNI Unix (Pyramid), supplemented by various tools (VBasic) for automatic and user directed uploading of documents.

User interface for consultation: VBasic Client/Server interface for members of the Translation Service, Web (Netscape 3.01) interface for other users (requesting departments).

3.2.4 Organisational aspects

The central database is fed automatically by planning departments, translators and secretaries by means of integrated workflow procedures. It is freely accessible to all Translation Service staff. Requesting departments have access to their own documents only.

3.2.5 Contact

See 3.1.5 above.

3.3 *CELEX*⁴

3.3.1 Purpose

To provide lawyers, national administrations, consultancies and researchers with comprehensive and authoritative information on European Community law.

⁴ <http://europa.eu.int/celex/>.

3.3.2 Content

CELEX provides multilingual (11 official languages), full-text coverage of a wide range of legal acts - including the founding treaties, binding and non-binding legislation, opinions and resolutions issued by the EU Institutions and consultative bodies, and the case law of the European Court of Justice - and parliamentary questions and their answers, and gives references to preparatory documents. It features hypertext links to subsequent amending acts, earlier acts amended, and national legislation implementing Community directives.

Search strategies may be specific (eg a publication reference) or very broad. They may scan the entire contents of the database or a specified subset of documents. Searches may be based on keywords or classification headings and may be refined further, as required, by using multiple search criteria.

The database currently holds over 200 000 documents in each of its 11 official language versions.

3.3.3 Technical environment

The database runs on a Bull (GCOS8) computer in the Commission's Computing Centre in Luxembourg with the Mistral documentary software. There are two alternative client interfaces: the Mistral command language and a WEB interface.

A new version of CELEX WEB (SmartGateway V2) is under development (mid 1998). New features will include the highlighting of search terms in the documents found, the simultaneous display of two language versions of the same document with parallel scrolling, improved hyperlinks, and menus in all the Community languages.

3.3.4 Organisational aspects

The CELEX database is produced and managed by the Office for Official Publications of the European Communities (EUR-OP). Bibliographical data are updated once a week on Thursdays and textual information during the following week.

3.3.5 Contact

JACOBUS DOGGEN
Tel: +352-4301-42110
Fax: +352-4301-42580
E-mail: jacobus.doggen@opoce.cec.be

3.4 *SdTNet*

3.4.1 Purpose

To distribute on-line general and administrative information and reference documentation throughout the service.

3.4.2 Content

News and information about day-to-day life in the Translation Service (senior management decisions, news about projects, weekly newsletter, minutes of meetings, IT strategy, staff movements, training, new acquisitions in libraries, etc);
User guides, FAQs for databases, tools and work procedures;
Documentation (full text or selected information, eg titles of COM documents, Green papers, White papers);
Language resources subject to frequent updating (list of Member State governments, countries and currencies of the world);
Information can be accessed by alphabetical index, thematic menu, department responsible, or full-text search;
Links to Europaplus, Europa⁵ and other interesting Web sites.

3.4.3 Technical environment

Netscape 3.01 as web browser and Netscape Enterprise Server on UNIX mainframe with Verity Search 97 indexing tool. Tools for HTML editing and server management: Office 97, HoT MetaL*Pro and MS Frontpage.

⁵ <http://europa.eu.int/>.

3.4.4 Organisational aspects

Information providers (one for each organisational unit) manage their own Web pages and can instantly update the system. Overall strategy is discussed and decided by a small management team (webmaster).

3.4.5 Contact

JOSEP BONET

Tel: +352-4301-34759

Fax: +352-4301-33915

E-mail: josep.bonet-heras@sdt.cec.be

3.5 *EURODICAUTOM*⁶

3.5.1 Purpose

To record EU and specialist terminology in a central location to which EU translators and administrators and freelance translators have online access.

3.5.2 Content

Entries (> 1 200 000) with the following structure:

- entry number and administrative information (created by, creation date, changed by, change date),
- entry attributes (subject code⁷),
- index fields (headwords and synonyms in any language),
- text fields (reference, definition, note).

3.5.3 Technical environment

The database is housed on a large-capacity mainframe (Siemens BS2000) located in the Commission's Computing Centre in Luxembourg. Migration to another platform (Fulcrum Searchtools) is under investigation. The client interface is either Web or terminal emulation (VT200).

⁶ <http://www2.echo.lu/edic/>.

⁷ Lench code.

3.5.4 Organisational aspects

EURODICAUTOM is owned by the European Commission. New data are entered by Commission terminologists (extracted from the Official Journal, specialist journals, daily press), translators (translation-oriented), and outside contractors. Entries are updated and consolidated by specialist suppliers under contract. EURODICAUTOM is made available to external users by ECHO.

3.5.5 Contact

HUBERT PAESMANS
Tel: +352-4301-32551
Fax: + 352-4301-34309
E-mail :hubert.paesmans@sdt.cec.be

3.6 MULTITERM

3.6.1 Purpose

To allow translators and terminologists to record, manage and retrieve terminology at local level (trouble-free uploading to EURODICAUTOM is possible using a Word macro), promoting consistency through the use of a standard format for terminology sharing at an early stage.

3.6.2 Content

Terms, phrases and abbreviations are recorded in a standard entry structure comprising the following fields:

- entry number and administrative information, generated automatically,
- entry attributes (subject code⁸, project code⁹),
- index fields (headwords and synonyms in any EU language),
- free text fields (reference, definition, note) for each index field.

8 Pick list of subject codes based on "Lenoch code", also used for EURODICAUTOM.

9 Pick list of project codes corresponding to "types of document" (eg regulation, directive, call for tenders, funding decision).

3.6.3 Technical environment

MULTITERM '95PLUS software (network version) is installed on the user's PC (standard configuration for all SdT staff). Both Windows 3.1 and NT 4.0 are supported. Databases are normally located on a NT server, allowing multi-user read/write access for up to 100 users. Unlimited guest access.

3.6.4 Organisational aspects

Translators record new terminology in their unit database, to which all translators have guest access. Terminologists use MULTITERM for recording and managing new multilingual collections before uploading to EURODICAUTOM. Ephemeral MULTITERM databases can be created for data downloaded from EURODICAUTOM via EURAMIS for use with TWB (automatic lookup).

3.6.5 Contact

See 3.5.5 above.

3.7 EURAMIS

3.7.1 Purpose

To allow the creation and sharing of linguistic resources.

3.7.2 Content

Central translation memory with maintenance, update and retrieval facilities,
Central sentence alignment¹⁰ tool,
Local sentence alignment editor.

¹⁰ Sentence alignment means the parallel segmentation of two different language versions of the same document to constitute a translation memory for use with the Translator's Workbench or other tools.

3.7.3 Technical environment

Client/server based system. The server is a Unix mainframe computer run by the Commission's Computing Centre in Luxembourg. The local client has been developed in Visual C++ and consists of input screens running on a Windows computer. Communication between client and server is by e-mail; the standard Windows DLL "mapi.dll" is used as an interface between the client and the mail system.

3.7.4 Organisational principles

Secretaries submit alignment requests, edit the automatically generated alignment, and send the resulting data files for central storage. Ideally a centralised team should align all language versions at the same time, thereby saving time and effort.

The translator retrieves the memory file relevant to the translation request in hand and imports it into his local memory before starting work with Translator's Workbench. Links have been developed to other translation tools, including machine translation, the terminological database (EURODICAUTOM), and the multilingual legislative database (CELEX).

After translation and revision, the document in the target language is cleaned up and the local translation memory is exported for central storage.

3.7.5 Contact

DIMITRI THEOLOGITIS
Tel: +352-4301-33632
Fax: +352-4301-34069
E-mail: dimitri.theologitis@sdt.cec.be

3.8 SYSTRAN

3.8.1 Purpose

To provide end users with fast-response machine translation (MT):

EN > DE, EL, ES, FR, IT, NL, PT
FR > DE, ES, EN, IT, NL
ES > EN, FR
DE > EN, FR
EL > FR

in a quality adequate for information scanning or for use as a drafting or translation aid. The raw machine output can be rapidly post-edited or carefully polished to produce a final, high-quality translation, depending on the destination of the text.

3.8.2 Content

The dictionary structure is based on a mono-source/multi-target approach. In the translation process there is constant interaction between two dictionary types: a general dictionary containing individual words and an expression dictionary for contextual entries. The Commission's MT dictionaries have been built up over 20 years. Drawing on feedback from the various areas of activity of the institution, the performance and reliability of the dictionaries are now widely recognised.

3.8.3 Technical environment

The system runs on an Amdahl mainframe (IBM compatible) under the MVS operating system, capable of processing translations at a rate of 2 000 pages per hour. There are three different client interfaces linked with the e-mail system: the standard general purpose e-mail interface, a VBasic interface for Windows 3.1 or NT 4.0 (installed only in the Translation Service) and a Web interface (available to all European institutions). Access is routed through a Unix server connected to the internal e-mail network, allowing requests from all Commission departments in Brussels and Luxembourg or from any other European institution to be processed.

3.8.4 Organisational aspects

Every Commission official has access to the system simply by attaching a text to a mail message addressed to the EC-Systran server. Response time for the user averages about 15 minutes (including telecommunications time). The VBasic and Web client provide specific guidance to the users.

3.8.5 Contact

See 3.7.5 above.

3.9 *Translator's Workbench*

3.9.1 Purpose

To speed up translation by re-using repetitive sentences, and to promote consistency of terminology and style.

3.9.2 Content

A translation memory consists of a bilingual collection of text segments in the source and target languages, derived from previous original documents and their translations done with the same (or similar) tools or from the alignment¹¹ of previously translated documents. Each memory item also comprises system fields (creation date and author, change date and author, use date, usage counter) and user fields (translator, doc n°, doc type, domain, year).

The software can be used in conjunction with a MULTITERM database allowing automatic terminology look-up when no segment match is found in the translation memory.

3.9.3 Technical environment

Pentium PC, at least 1 Gb hard disk and 32 Mb RAM. Windows 3.1 or NT. Full benefit is obtained in a network where data can be stored and shared centrally. The software (Trados GmbH) is interfaced with WinWord. Data are transferred by importing and exporting text data files.

3.9.4 Organisational principles

Memory files are organised in collections according to subject, document type, requester and/or translator. Sharing of data can only be efficient if everyone uses the same database structure. A model has been created containing all subjects, document types and requesters. After translation the documents are cleaned up and the updated memory is exported and stored centrally.

¹¹ EURAMIS (in house) or WinAlign (market product).

3.9.5 Contact

See 3.7.5 above.

4 Transportability to translation centres in associated countries

A typical translation centre would be composed of 5 to 20 people (translators, terminologist/documentalist, administrative staff and IT support). The only essential difference between the computer environment and applications of a smallish translation centre and those of the EC Translation Service would be the number of users. The tools themselves would serve the same purpose.

- *General office requirements:* MS Office on PC (Windows 95 or NT)
- *Data transfer and communication:* local interconnection of PCs with a central server (Windows NT or Unix) for data sharing; e-mail communication between PCs internally and with the external world via the Internet.
- *Translation request management* in a central database (eg MS Access) in which all workflow operations are recorded.
- *Central document repository* with a unique document numbering scheme (eg CELEX numbers) and an indication of the stage of preparation (version number). Indexing of document repository by Fulcrum Searchserver or other Web technology.
- *Interactive translation with TWB*, shared translation memories on central server (Windows NT or Unix) or EURAMIS; document alignment with Winalign or *Euramis*.
- *Terminology management* with MULTITERM and one or more shared databases.
- *Distribution of information and documentation* on a local Web server, with possible access for freelance translators. Access to CELEX on CD-ROM or the Internet.

The usefulness of these tools and procedures in preparing the way for a smooth integration into the general structure of the Commission on accession cannot be overemphasized.

**PIET VERLEYSEN
ANDREAS TAES
DIANA JANK**