

# **Workflow Automation of Translation Projects**

*Paul Kaeser*

People in many different sectors of the economy are talking about workflow systems, although the subject has only been discussed occasionally in the field of technical documentation. Concrete applications and solutions have been rare thus far. The following case study illustrates the possibilities and the, as yet unexploited, potential.

The preparation of technical documentation, particularly when publishing in several languages or on a large scale, requires the full cooperation of several specialists. The interfaces between these specialists are still handled manually for the most part, thus causing losses at the interfaces which diminish overall performance.

One of the central themes for the STAR Group's software development departments is to achieve the transparency of processes in the field of documentation through standardization and automation of the interfaces.

STAR is well-known to many as the producer of the TRANSIT translation tool, but in addition to this, STAR is working on a series of standard and specialized software products, all of which aim to optimize the entire process of technical documentation up to and including multilingual and multimedia publication. SGML has an important role to play here.

Exploiting the efficiency of the specialized standard tools while simultaneously controlling the subprocesses through a host control system and total integration were additional challenges.

In most cases, preparing and publishing technical documentation involves intensive team work. For this reason, technical documentation is fundamentally ideally suited to workflow solutions and these solutions generally become a requirement as a consequence of the range of disciplines which need to be handled (writing and editing, graphics, translation into a series of different languages, multilingual DTP etc). In practice, however, they are also required because of the large volume of the documentation, tight deadlines and the variety of versions required.

Commercially available applications, specifically in the form of authoring or translation tools, are usually only aimed at enhancing the productivity of a single job or process and until now, too little attention has been paid to the harmonious and conflict-free cooperation between working groups, or in modern jargon, to groupware solutions.

In the past few years, some specialized applications have made it possible to improve productivity markedly, but it is for this precise reason that the costs generated at the interfaces have risen sharply as a percentage of the whole, which is why they have become the focus of much interest. Added to this is the fact that, in many cases, team members now often work in several different locations rather than just in one place and may even be spread over several different countries.

## **The automation of the translation process**

The possibilities which already exist will become clear from the optimization of the translation and foreign language publication processes using a purpose-built translation workflow system. The potential to be extracted in this situation is less to do with the translation itself and much more to do with handling and interfaces in the translation environment. The translation itself will already have been substantially optimized using TRANSIT, a commercially available standard translation tool. The translation environment demands a certain amount of administrative and data processing work and these activities can, in certain cases, exceed the actual effort demanded by the translation itself.

Although the translation process and the related handling processes are of primary interest at this stage, this subprocess must be incorporated into the documentation process in a meaningful and conflict-free manner. Upstream and downstream processes must be stable, ie they must handle the appropriate input and output interfaces reliably, if a subprocess is to be automated.

## **The translation process starts with the author**

The truism which states that the method used to prepare documentation affects the efficiency of subsequent processes has by now become sufficiently well-known within the documentation community (although whether this knowledge acted on in every case is another question).

Apart from the linguistic and terminological requirements on documentation to make it suitable for translation, elements such as standardized document structures and correctness in formatting also play a part in the automation of the translation process. For example, tables and indented paragraphs created with the tab key are frequently found, even today.

Put another way, documents should be structured in such a way that as few functions are used as possible, better still no functions at all. Texts outside automatic pagination are graphically controlled using paragraph formats. In this way, a translated text can subsequently be formatted automatically by the DTP system with practically no intervention. At the same time, this procedure also has the side-effect of significantly increasing the ease with which the source text can be maintained.

## **Translation Memory Systems**

STAR began development of the TRANSIT Translation System in the mid eighties. The concept of an integrated translation environment, and in particular the principle of Translation Memory, was barely taken seriously at all in the industry. Today there are virtually no medium or large-scale translation projects planned without the use of this technology.

A first step towards the integration of TRANSIT into the documentation process was the development of interfaces to all popular DTP systems (Interleaf, Framemaker, MS-Word, Quark Xpress, Pagemaker etc.) and to data formats (SGML, HTML etc).

These developments have since become indispensable parts of the daily routine in the industry.

The next step towards integration is extensive automation of processes, so that no more time is wasted on pre-editing and post-editing work. This has been achieved by making TRANSIT fully capable of integration and by adding server capabilities to it. This means that TRANSIT can take commands from other (control) applications, process them and send the results back to the control application.

The specialized work, the translation-relevant actions, are handled by the TRANSIT standard application.

The concept of the Translation Workflow System developed by explained below.

### **Mode of Operation**

A daemon on the server system (MS-Windows NT 4.0 in this concrete case) continuously monitors specified directories. As soon as an authoring system saves data to one of these monitored directories, a job file also supplied by the authoring system is read. The information a human “contractor” would also require to complete a job is stored in the job file. The job is processed by the system fully automatically.

The workflow system now automatically passes on appropriate commands to the TRANSIT Translation Memory System which is also installed on the server. These are, in particular, the following:

- Conversion, segmentation and automatic pretranslation using data identified by the workflow system as the most suitable reference material. Creation of control data for the translator’s client computer.
- Once the data has been prepared for the translator in this way, it is compressed into an archive file. The workflow system now searches in the database for the translator best suited for this job. Criteria include, for example, type and volume of documentation, language combination and the translator’s current availability. The archive file is now sent to the translator via FTP (with the options of network, modem ISDN or Internet).
- TRANSIT Light (the translator version of TRANSIT) is installed at the client (translator) side along with a workflow system client. The client reads the data from the server, installs the translation job automatically and informs the translator on screen.
- The translator now only has to click on the job-number and TRANSIT loads itself automatically with the parameters required for this particular job and the information to be translated. The translator now “only” has to translate and check his or her translation.

The return route is similar, once again fully automated, as soon as the translator approves the translation for dispatch.

The workflow system on the server now sends the same data to one or more revisers (depending on the specified process, eg language reviser, technical reviser). Certain automatic checks (eg consistency, code page, SGML/HTML structure, translation completeness etc) can be incorporated at this point before the translation is returned to the authoring system or customer. These checks enable the authoring system to paginate the foreign language data directly.

All handling, both at the server and client end, is fully automatic. The system tracks each job until it is returned, sets and monitors deadlines, sends reminders for missed deadlines etc. All actions and translation volumes are continually logged recording precisely who did or approved what and when, an important element in QA systems.

The information can be exported in Excel format at the end of the month or at other times, sorted by translator and/or customer. This means that invoicing can also be automatic, for instance. Messages are sent to the system administrator in the event of any problems, allowing for appropriate intervention, or discussion with the translator concerned.

All translations completed using this system are automatically added to the reference pool (ie Translation Memory management is automatic) so that, once completed, translations can be automatically reused by workflow server at any time.

### **The ability to integrate as an important criterion for tool evaluation**

The suitability of the tools to be integrated is critical in ensuring that a system of this type can function correctly.

The capability of integration ought to become the principal criterion in the evaluation of such tools for professional environments.

When would a workflow environment make sense? In principle, it makes sense to automate any time the stability of a complete process or authorship can be influenced. This is in the interest of overall productivity and is not, therefore, merely an end in itself. In practice, there are two typical scenarios which demonstrate the system's potential particularly clearly:

- A large volume of small documents which need to be regularly translated into or out of several languages to a deadline (eg service bulletins)
- Larger scale documentation which is prepared and maintained in modular fashion (eg systems documentation). This makes "Simship" (the simultaneous publication of all language versions) a relatively stress-free possibility for the translator, even when publications are on a particularly large-scale. This case is typical for CD-ROM or website publication.

The objectives are identical in both these cases:

- Reduction of costs by avoiding administrative and handling overheads
- Shorter throughput times
- Prevention of errors by cutting out the human factor in routine procedures
- Reliable monitoring of the status and quality of work being carried out

What are the prospects for a wider distribution of the integrated publication chain? A strong trend towards publishing in several formats simultaneously can already be detected. At least five different media are current today: paper, microfilm, online help, CD and the Web.