



It's a multicultural world...  
Are you ready?™

*"CMS? TMS? L10n? OMG!"*

## **Integrating Translation Efforts: Navigating the CMS Marketplace**

*By Sean Oliver*

### **Decisions, Decisions**



There are over 1,000 pieces of software that can be labeled a CMS, built on 40-odd content management frameworks, with a myriad of technologies (Java, SQL, .NET, PHP, etc.). Some can be very simple web-based systems, while others can be gargantuan server-based systems that are used to control and update nearly all of an organization's content, even social content. Some of them are proprietary and closed-source, some are free and/or open-sourced. Some systems support easy translation and localization, and some of them don't even support non-English characters. Making a decision that's right for your organization in terms of what CMS to choose is hard enough, but place the specter of translation and non-English content on top of that, coupled with teaching a completely new workflow to your employees, IT staff, translation vendor, or even transitioning from an existing CMS... and you have the potential for a less than ideal outcome, if your organization fails to plan strategically.

First of all, we should state the key function of CMS technology... except that it can vary, depending on who you talk to. Some experts say the main purpose is to maintain the *consistency* of content. Others say it's the *portability* of content, or ensuring that content only needs to be created once, and can be managed and distributed to virtually any platform. For many companies who don't currently have any kind of CMS, keeping marketing, sales, IT, training and SOP language all consistent is often difficult, relying on branding guides and manual updates. Keeping content consistent manually can be incredibly cumbersome, especially if your organization undergoes any kind of rebranding efforts. Additionally, no portability means re-creating articles, web pages, reports, blog entries, whitepapers, and other content if you need to move between formats or platforms.

Two other key functions of CMS technology are the separation of content and layout, and content search. With many CMSs, there are standard templates on top of which users create new or updated copy, and these updates will automatically propagate onto a webpage or into other internal databases. In theory, this should cut the need for designers to a fraction of their current work, as the entire text-placing step should be automated through the CMS (the layout is already done, and the CMS drops in the text) -- leaving only the creation of templates to be done by



It's a multicultural world...  
Are you ready?™

programmers/designers. Search speaks for itself; if there's no way to sort through your repository of content, what's the point of organizing it?

The best possible outcome for purchasing and implementing a CMS, is you cut your DTP and web design hours by a huge percentage, and only small pieces of content are sent over for translation instead of whole documents, saving you untold dollars on translation services. But even in the best case scenarios, you will face a learning curve of at least a few months, custom code will likely be required (or you'll need to search out the right plugins), and even with your best efforts, at least some of your internal stakeholders will likely not be 100% happy with the eventual functionality of the CMS, as it applies to their role.

The worst possible outcome is no one uses it. Or rather, you're unable to get buy-in from employees to use the new system themselves, rather than relying on developers to make changes to web pages for them. Or maybe it's so cumbersome, that it takes more than a year to figure it out... time that your front-line employees could spend on generating revenue, while letting your developers address content changes to your site. Or you spend \$1-\$20 million dollars trying out different expensive CMS tools, only to junk them all at the end. Or your CMS works great, but while it supports non-English characters, it doesn't allow CAT tools to work with the content inside the CMS.

Decisions you end up making about the CMS you go with are inevitably going to be based, at least in part, on decisions that you have little control over. For example, what systems were present in the past? How was your web page set up by your web designers? Do you have one Language Service Provider (LSP), an internal translation department, or a few different LSPs or freelancers that work for different business units? How tech savvy are your language service providers? How much control over the process do non-IT or non-executives have? Is the CMS going to be utilized organization-wide, or siloed within one business unit? Who decides what kind of content needs to be managed with a CMS?

A key thing to remember in selecting a CMS is that choosing the cheapest system which meets your needs *today* may cripple your organization in the long-run, as the types of content, amount of content, and ways content is utilized have the potential to increase at a rate that some of the cheaper CMS options simply can't accommodate. Also, if your CMS is initially siloed in one business unit or one type/format of content, and is shown to successfully drive down costs and drive up profits, the decision-makers in your organization are likely to require other business units or content types/formats to be stored in a similar manner. If you've purchased a CMS that doesn't have the capability to expand, or whose upgradability is suspect, then you may be in a world of pain when your needs inevitably change in the future. Also, if you only take away one thing from this white paper, realize that **if you don't have someone who is held accountable for the upkeep and management of the CMS, the chances of your implementation succeeding will be greatly reduced.**



It's a multicultural world...  
Are you ready?™

## Languages and TMS Technology

A significant challenge arises when trying to integrate translation processes, CAT Tools<sup>1</sup>, and TMS functionality. There are two great sites to help you make a CMS decision; a [comprehensive list of all CMS software](#) and a [checklist you can fill out and hand off](#) to a CMS vendor. It's slightly disconcerting, however, that each of these sites has a **single** variable for translation ("translation management" and "Web based translation management" respectively). The identification of languages and translation as ancillary, or the very last issue to be addressed, seems the norm in the [available literature](#) regarding CMS technology. Achieving seamless integration between content management and translation efforts is not as easy as checking a box after you've finished creating content to insert in a template. Many of the challenges arise from the difficulty of trying to integrate translation processes, CAT Tools, and TMS functionality.

Translation Management Systems (TMS), a technology related to CMS, have become increasingly popular, with more and more marketplace offerings in recent years. The general purpose of most TMSs is to take some of the legwork and manual status updates out of the localization process by providing a full-service platform for the entire translation process -- including managing, buying, quoting, and sometimes retrieving completed translations. Many TMSs have built-in analysis tools, so that buyers can get an accurate word count (and quote) directly online, without waiting for their LSP to run an analysis on their end. Some allow you to maintain a cloud-based TM that's updated in real-time. The main compatibility issue that arises between some CMS and TMS technologies is that a lot of TMSs and CAT tools have their own specialized file formats, usually referred to as *bilingual* formats. In a bilingual format, the original "source" text will appear next to the translated "target" text, inside tags. Some examples of bilingual format file extensions are: .TTX, .TMX, .IIX, .MDF, .MTF, .MWF, .TMW, .BAK and .ITD. These are the formats that translators work in, and which have to be converted to the original file format requested by the client (Word, PPT, XML etc.) after translation. Most CMSs require specific tagged file formats, and so the translator's bilingual files have to be converted. Since tagged source files contain lots of code that a translator needs to leave intact, it's imperative that the translator, or agency, knows how to properly alter the settings of their TMS/CAT tools so that only the text is translated, and the tags (or HTML code) are preserved. It's still pretty common for small corruptions of the tags to occur, which is why manual review and testing are usually necessary. Some CMSs can function using Excel imports/exports, but it seems that most new CMS technology is moving towards tagged formats like XML, and towards uniform standards of content management architecture such as DITA.

---

<sup>1</sup> TMS systems always have CAT technology built-in. CAT tools, as referred to in this paper, refers to stand-alone software that's not housed inside a TMS.



It's a multicultural world...  
Are you ready?™

## Cost Savings

Compatibility/workflow issues aside, the cost-saving potential of a CMS in regards to translation can be huge. Even if your translation vendor uses a TM (Translation Memory) to reuse previously-translated segments, the process of running the new content through translation software such as SDL Trados™ every time you make a change to a document will result in charges for review. Most translation vendors have a minimum charge for each job, since most translators have a minimum charge as well. An ideal workflow with a CMS will require that translators/LSPs agree to forgo these minimum charges, and will be automatically notified when new, translatable content is introduced into a client's content/CMS/etc. (generally in a tagged format like XML). The benefit for the translators is not having to do any content management on their own (i.e. "Didn't we translate this last year?"). That way, instead of running an entire "updated" document through a translation memory, identifying the repetitions, and having the content reviewed, the CMS will send an alert about the *specific* content that's been updated, allowing your organization to manage translations at the paragraph level instead of at the document level.

## Gain Repetitions, but Lose Some Context

Using a CMS where text is exported and imported using a tagged format such as XML also helps in finding more repetitions, as some translation memories will not always identify the same text segment as a match if the characters are formatted differently. As XML/tagged text strings have no formatting, this issue is eliminated. That being said, the updated piece of content might be affected by existing content in the same document, or terminology across different sources, and the entire document may still require review by a linguist (or entire swaths of content might require update across multiple platforms). It's more difficult for a translator to translate tagged text strings than to look at a fully-formatted document, perhaps with graphics or other contextual elements visible. The ideal workflow for quality in translation involves a translator and a proofreader, and encompasses a full review of entire documents, every time. But as the available amount of information on the planet keeps multiplying exponentially, other, more cost- and time-saving solutions such as CMS/TMS technology have become a necessity, even if [very] slight reductions in quality are possible.

## What could go wrong?

You can read plenty of horror stories on translator forums where an organization's chosen CMS introduces compatibility issues with their translation vendor's workflow, and, for example, translation has to be done inside a web-based CMS, without the aid of CAT tools (and without direct access to the translation memory). There are CMSs that users can't simply update content using a spreadsheet, TTX formatting, etc. in multiple languages. What happens then? The translator is pretty much left with the option of receiving the text for translation, pasting it into a CAT-compatible format, running it through a CAT tool, translating it, and copy-pasting



It's a multicultural world...  
Are you ready?™

the results back into the appropriate fields of the web-based CMS. Most translators will charge for the extra work, which may gradually erode any cost savings you realized by implementing a CMS in the first place. Or worse yet, your web development team may end up constantly troubleshooting, and not reducing their workload at all.

I talked to a few people in charge of translations and translation management at global organizations, and some of them shared that in the past, they had purchased translation CAT tools themselves (SDL Trados™) in order to try to cut costs and manage the TM on their end... but it proved entirely too cumbersome. Managing content at the word/segment level is generally a level of granularity that's not required by most translation buyers (as it can be ridiculously time-consuming) but reducing the reliance on the translation vendor and managing content internally as much as possible can be beneficial. The only options besides purchasing CAT Tools yourself, or a CMS, is to manually mark-up documents (track changes, etc.) indicating which sections have been changed. But in our experience, something always falls through the cracks (i.e. updates to the source text post-translation), which generally require a review of the entire document anyway - especially if your content-creation workflow involves passing a document between multiple parties during multiple layers of review. Without a CMS, it can be quite difficult to manually keep track of the final approved versions of content without significant time investment in manually tracking it.

I talked to another party about a year ago, who had implemented a CMS that their IT department had selected for them (apparently without input from other departments). There was no way for the content creators to view how the placed translations would look. They would upload the new source (English) content into the CMS, and then notify the translation vendor that there was content that would need to be translated. The translation vendor would process the request, but in the interim, there was no way to work in real-time with the translations. The client had no way to check to see if the translation would actually fit into the pre-created templates until after it was done. Then, they would have to amend their original content or template, send it back for translation, rinse, lather, repeat.

I also talked to a translation coordinator who recounted an experience with trying to line up their new CMS with their translation efforts. They simply gave up after a year of having to go back to the developer to create new templates for them all the time, and constantly having to preview, re-load, and re-preview the final appearance of the changes they made. They spent more time trying to work with the CMS than it saved them in terms of layout. It wasn't a workable solution for them in the long run, both in terms of time and money spent.

Even if your CMS is set up to handle the most common languages, you can run into issues with languages you might not anticipate. Welsh, for example, has "ŵ" and "ŷ", which aren't available characters on some CMSs. You might not be able to use



It's a multicultural world...  
Are you ready?™

double-byte characters (most Asian languages). You might have some two-byte characters (Chinese, Japanese) supported, but not all (Korean, Arabic). Right-to-left (RTL) languages might not be supported. Knowing all possible languages in which you might have to manage content is advantageous... but not always possible. In today's rapidly-globalizing business world, your organization might suddenly need content in Turkish or Tagalog; needs that can be difficult to meet post-CMS implementation. If there's a chance that the content will be globally exported, it might be prudent to pick a CMS with as many language options as possible, or one that can be customized to expand the number of languages at a later date.

In reality, for many small/medium-sized businesses whose only real need is managing web content, you might be better off in the long run just having someone in your office learn the basics of your web design software, and have them design some templates for you. That way, you kill two birds with one stone by having an internal resource who can update your webpage, and avoiding having to require the entire office to learn how to interact with the CMS. It might be cheaper and easier in the long run than learning a whole CMS system, changing your entire content-creation workflow, and inevitably relying on your IT team to fix the unavoidable problems that arise. In my conversations with CMS users, one thing stood out: **organizations that had a devoted member of their IT staff whose sole function was the management and maintenance of the CMS could make the system work for their organization.** Those organizations that lacked the internal IT resources often ended up abandoning the CMS and resorting to manually updating and managing content. The big question to ask is whether the cost savings are worth the money and time investments for learning the new process, and whether you've got the resources and translators to be able to make it work.

### **Isn't There an App for That?**

Unfortunately, no. The issues you run into when trying to integrate a CMS with translation efforts are the same ones you run into when doing DTP, or subtitling, or localizing e-learning across several languages:

1. You don't know what the final product is going to look like until it's done, when the translator performs a QA review, and gives you corrections or feedback.
2. The person doing the actual DTP/subtitling/coding almost never has all the language skills for the content that their working in. You can imagine what the hourly rate would be for a graphic/web designer, who can also read Spanish, Japanese, Russian, Swahili and Hindi. In reality, they don't exist<sup>2</sup>.
3. The *source* (probably English) content developers and designers, whether they're internal or external, might have little to no understanding of the localization process, and might not even know to take simple steps like

---

<sup>2</sup> To be fair, one of our vendors has a Russian/Chinese/English speaking designer who we love working with, for obvious reasons.



It's a multicultural world...

Are you ready?™

using Unicode font or allowing room for text growth/shrinkage. These developers/designers are more likely to think that issues with the appearance of the translated text bleeding outside of boxes, or containing corrupted characters as beyond the scope of their role. A big part of a successful multilingual CMS rollout is getting buy-in from your non-linguistically-inclined staff to change their workflow to accommodate working with translated documents, tagged formats, and dubbed/subtitled video in addition to the challenges of a CMS.

## **Be Proactive**

Realize that the investment in new technology is always twofold: money *and* time. No matter what anyone promises you, even integrating a simple CMS takes time, and experimentation, and making mistakes, and learning from those mistakes as part of new workflows and procedures. It takes a lot of finding out what doesn't work, and reevaluating your workflows to fit the specific needs of your content creators, IT staff, and translation coordinators for optimal organizational performance. A CMS isn't going to magically create new content for you, any more than a TMS or CAT tools will magically translate your content. Also, as with any kind of tech vendor, don't expect the CMS vendor to be able to anticipate *all* the possible issues. Do your own research. Ask the vendor to provide you with contact information for people who've gone with the system you're considering. Find out what their issues were, how they resolved them, and whether they would do anything differently.

Demonstrate the functionality (or lack thereof) of your new systems to **all** your staff, not just the ones directly involved with the nuts and bolts of the CMS. Demonstrate both the cost- and time-savings, and show how the changed workflow will mean improved productivity, better consistency, and less time spent on formatting. If your source-content creators are better able to understand the trials and tribulations of your translation coordinators and IT staff, they might be able to make better decisions on how to create and amend content and templates at the *outset* of a project. Getting buy-in from not just senior management and leadership, but from front line employees and IT staff can be critical, as it's likely they will be the ones who have to work out the bugs. Also, introducing a feedback system for all involved parties at the start of your CMS rollout can help you identify issues quickly, no matter how far down the organizational ladder they occur.

## **In Conclusion**

CMS technology has revolutionized the ways that organizations create, use, and store content. Being able to instantly access previously-created content, move content around on different platforms, and more easily manage localized content can provide huge benefits. However, the decision should not be made lightly, as time and monetary investments will be steep. If you are going to roll out a CMS, make sure you make the decision about which platform to go with holistically, and



It's a multicultural world...

Are you ready?™

consider the concerns of all possible stakeholders as well as your current localization needs. Choose a solution that will be scalable when your organization expands, or when you expand to new non-home-country locations. Above all, make sure you have someone internally dedicated to the maintenance and upkeep of the CMS in some major capacity. Managing the technical aspects of a CMS isn't something you want to leave to external vendor, as you'll undoubtedly end up having to contract them for lots of work. I hope this was helpful as you consider the advantages and disadvantages of using a CMS for your multilingual content.



It's a multicultural world...  
Are you ready?™

## **Terms:**

**CMS:** Content Management System - Software or strategy for managing data and workflow among a large set of users. A CMS will often: organize and/or store data, separate content from layout, make data portable, control access to data, and maintain consistency of data across media. A list of available CMS software can be accessed [here](#).

**CCMS:** Component Content Management System - A CMS that manages content by segment/component/granular level, instead of at the document level. Organizes structured content through the use of tags (e.g. XML tags). Structured content is usually customer-facing (e.g., product usage, training, marketing, and support) and not as likely to be internal content such as reports, project information etc.

**DITA:** Darwinian Information Typing Architecture- A standard XML data model for authoring and publishing content. Content is created as a series of short searchable "topics", and the layout is indicated through DITA "maps".

**ECMS:** Enterprise Content Management System - Technologies encompassing knowledge management, records management, and/or management of information such as transaction information, policies and procedures, emails, etc. Different from CCMS in that it generally focuses on unstructured content that is more often utilized within an organization, instead of being utilized for client-facing content.

**WCMS:** Web Content Management System - A specialized CMS that manages creation, collaboration, and administration of web content. Intended for users without much expertise in web page design.

**TMS:** Translation Management System – Software that manages translation project information, organizes the translation process and workflow, and often is the platform from which all tasks in the translation process are managed and performed, including ordering/delivery and terminology management. Generally a system employed by translation providers, rather than translation buyers. However, many TMS technologies have the capacity to allow a buyer to purchase translation, or check project status, directly through a website or other TMS portal. Generally, a TMS will have CAT software built-in.

**TM:** Translation Memory - An electronic repository of an organization's previous translations, that are accessible and re-usable, generally created using a CAT tool.

**CAT Tool:** Computer Assisted Translation – Software that aids a translator in terminology consistency, reuse of previously translated content, and improved time-to-market.