Creation to Translation

By Shannon Zimmerman, Sajan

For quite some time, the translation industry has struggled with disparate tools and a limited ability to leverage content. Early technologies such as Translation Memory (TM) provided linguists with the ability to become more efficient. They would use these tools for purposes of consistency as well as efficiency, and, as important to the buyer, cost reduction.

The challenge that existed then still exists today, in that these assets are not readily available or accessible to content authors. In the rare cases where the data is available to the authors, the integrity and quality of the data often leaves much to be desired. This does not mean that TM does not create some value; it simply means that there are significant opportunities for improvement.

"The emphasis to achieve maximum value must be on two areas:

- *the organization of the data.*
- *the availability of the data to the user."*

If ever you doubted this last statement, look at the myriad of companies developing technologies to more effectively mine, consolidate, and organize TMs, all to gain greater value from the data. This is done to compensate for a deficiency that exists. This is done as the voice of the market demands greater value and deeper integration of assets.

We hear about integration with Content Management Systems (CMS), but are they really integrated? This concept parallels other content-related demands observed in the business and technology sectors. We have all watched the evolution of more mainstream content solutions. At one time, electronic storage of content was valuable. Next, the ability to mine within those documents or files became a significant improvement. Finally, the ability to intelligently access data on-demand became the market requirement. Not to be forgotten, the market also demands automated workflow wrapped around all of this.

The intent of this article is not to condemn TM or any of the existing technologies. In fact, anyone who appreciates technology can appreciate what TM has done for the language translation industry. My objective is to elicit some new thinking around how we in the industry view these assets and how we access and obtain value from these assets. I also add that the following solution description is not hypothetical or theoretical, but actual.

The emphasis to achieve maximum value must be on two areas:

- The organization of the data.
- The availability of the data to the user.

Data

The extent of the value derived is directly proportionate to the intelligent organization of the multilingual content. We learned from a recent issue of CSN Magazine about what was referred to as an innovative approach to improve contextual accuracy within TM files. This approach should be applauded, as it does address a key missing piece of functionality that has historically plagued TM.

> ***"If used during the source language creation process, we have seen an average of 15-20 percent translation cost reduction simply due to source content reuse."***

However, it is still predicated on the management and organization of decentralized data files, unless you wish to invest significantly into software. With no common or central source by which to synchronize your multilingual content, data integrity will be an ongoing challenge. Simply stated, you will have multiple targets for a given source; you will return target segments that don't exactly match the context of what the source demands.

For data to be most useful, it must be well organized. Most buyers will agree with this statement but will immediately conclude that it will involve significant upfront investments in time and cost to establish the data structure and maintain it over time. Therefore, the recommended solution is to utilize a technology solution that manages data as it moves through the content ecosystem and stores it in a manner which actually purifies it so as to avoid the typical integrity, redundancy, and accuracy issues.

This is accomplished when your workflow application is directly integrated into a segment level, centralized data repository. Segments are indexed and stored contextually to achieve precise contextual accuracy. There are many additional benefits that can be derived from this model, but suffice it to say that, if you use a centralized, highly organized data model initially, you can avoid the pitfalls associated with decentralization.

One cannot assume that this approach to storing your data leaves it resident on a server. Doing so would keep it separate from those who need it most. Additionally, it may be opportune to consider other uses for your source and target data. A robust repository can be used to serve multiple business purposes. For example, this valuable data asset is also available to those who are creating new content. If used during the source language creation process, we have seen an average of 15-20 percent translation cost reduction simply due to source content reuse.

Delivery

If the data is now in a centralized, cleaner and better organized state, the focus must shift to delivery of that data to the users that demand it. Technology plays a significant part in how this can be done effectively. Data must be exposed and available for user consumption in an intelligent manner. It is not only important to have access to the data, but to also have the ability to access that data when it is most opportune and advantageous to the user.

Improved access can be accomplished when desktop users passes data demands to the central server in an interactive manner, while working in their desired desktop applications. Functioning like a grammar check, similar source segments can be presented to the user for consideration. Only the relevant data is passed to the user, passed via Web Services to the desktop application.

This can be done during the authoring process. Writers can be presented with segments that are relevant to the subjects they are authoring. This is done in a non-invasive manner and serves to establish a more consistent source document. The user does not have to manually mine for data—it is automated. This can also compliment existing CMS repositories. Connected to a CMS repository, source language content is presented to establish greater consistency.

This type of intelligent integration presents significant value to the translation process. The same interaction from the desktop application to the central server can now present the user with preliminary leveraging results. A user has a clear understanding of the composition of the document before it is ever sent to a Language Service Provider (LSP). And if the desktop application is configured to be "server aware," a translation project can be initialized directly from the desktop.

For some time, the thought of having the best of both worlds seemed impossible. How can we achieve optimum data organization producing clean, high quality data in a model that has grown so highly decentralized? This is a valid question, and it's the reason we see so much effort within the industry to develop methods to correct things after the fact. Using a central system that weaves the process steps directly into the data achieves a high degree of control. While not addressed in this article,

> ***"Using a desktop tool that is intelligent and aware of the central system is precisely how buyers can achieve the best of both worlds."***

another key benefit of this model is the ability to capture real-time quality metrics throughout the process.

Using a desktop tool that is intelligent and aware of the central system is precisely how buyers can achieve the best of both worlds. The alternative, to use silo technology in a decentralized model, is too taxing and dilutes the overall value of technology. Centralized data and process management can coexist with a decentralized user base.

Start to Finish

To tie together the points made in this article, let's walk through a step-by-step example that illustrates how a typical user can leverage the described model to achieve greater value. Value is defined not only in financial terms, but also in terms of quality, consistency, and information exchange.

Technology prerequisite:

- Using a Web-based, central technology tool in a Software-as-a-Service (SaaS) model for process automation, user access and control, central data storage, and global collaboration. This is the common denominator or central hub for all activities.
- Using a desktop software tool that provides controlled access to both source language content and multilingual content. It can be connected to a central translation database or, with filters, pull directly from CMS repositories. It offers intelligent awareness to the central system.

A technical writer for Company X begins developing new documentation for the company's latest product release. Using any writing software, the writer launches the desktop software that establishes connection with the central server via Web Services.

Easily navigating through a visual representation of the data store, the writer activates content that is relevant to the documentation being developed. As the writing process continues, various source language segments are presented to the writer, each resulting in greater consistency with past content, offering the potential for translation savings later in the process.

Upon completion of the document, if desired, the writer can perform an evaluation of the document(s) to see what the leveraging percentages are. The source file can be transmitted into a workflow process for source content approval. Once ready, the file can be submitted directly into the translation process from the desktop. At any point, if the writer wishes to see the history of any segment, the writer simply chooses "segment details" to see a full history and version record.

Project status is readily available throughout the entire translation process. All process participants can easily connect via the central Web system to perform their functions. All tasks are visible and all data accessible—always. The data management function throughout the process is tightly controlled; a process participant may only perform functions that they are allowed to do

by the system administrator within their organization.

Upon completion, all files are posted and can be transmitted to the original requestor or to the destination of choice. Any data updates or versioning has been a natural part of the process and does not require additional TM maintenance of any kind. If quality metrics are desired, this data is generated and can be considered non-biased, as it is not coming from the LSP, but rather from the system. Regardless of which quality standard you subscribe to, capturing the quality elements naturally in the process offers unbiased results.

Results

The process reduced variability of the content itself by offering a means of achieving consistency during the authoring process—not to mention a much improved database. The centralized database with contextual indexing helps avoid the challenges associated with a decentralized TM approach. Automating the process steps and making them transparent to the user at the desktop, yet truly communicating with the central server, enabled a reduction in resource management time for both the client and LSP. Finally, by utilizing a SaaS model via the Web, all participants are loosely located, but tightly connected, resulting in greater efficiency and control.

By abstracting this advanced database and centralizing it, true integration into other peripheral systems, such as a CMS, is more easily accomplished. This type of solution can be very complementary to a CMS solution, if the two systems can truly exchange data. If it is nothing more than a file or package handoff, the value is diminished.

The days of disparate technologies and decentralized data are being challenged by the more educated language translation buyers. A cohesive solution does exist and can be delivered in a non-invasive manner. This should not be considered earth-shattering innovation. Instead, it is drawing from the more advanced technologies offered today and applying them in a no-nonsense manner. The conditioned thinking of this market is changing. If you challenge the traditional model, the results can be significant.

"Automating the process steps and making them transparent to the user at the desktop, yet truly communicating with the central server, enabled a reduction in resource management time for both the client and LSP."