

## Technology

# Exchanging translated HTML data (almost) painlessly



*Kevin Donovan*

Kevin Donovan has worked on localization projects for over 14 years.

At some point, it has probably happened to you. Someone asked you to have a lot of HTML formatted text translated. This is not in itself usually too big of a deal. After all, SDL Trados and so many other translation tools can handle HTML tags without a hitch.

But how do you handle the exchange of newly translated HTML formatted text that happens to sit in SQL Server and not in a Windows or MacOS folder? And what if the translated text and associated formatting needs to be reviewed, modified and reviewed again before final approval by the customer? To top it off, what if the main reviewer has a horror of HTML tags and needs to review all the text in what-you-see-is-what-you-get (WYSIWYG) format? In addition, imagine that the customer wants to review the translated text and the original English in the same tool.

Do these kinds of constraints sound farfetched? They probably don't to anyone who has worked in localization for a while. They are, in fact, exactly the conditions that one of our customers demanded on a recent project. How

we managed to do this without ruffling too many feathers illustrates how translated data can be successfully exchanged across very different systems, organizations and countries.

### *Why store HTML in a database?*

You might wonder how HTML text — 60,000 words of it — came to be stored in a database management system. Well, it is a longish kind of story. Our company is in the behavior modification business. One of our products is a web portal. The customers buying this web portal, and a related mobile app, are typically organizations that want to help their employees or their own customers to break a bad habit such as smoking or acquire a healthy habit such as exercising.

Patients or employees who are trying to change some aspect of their behavior log in to this web portal to learn how to change their behavior and ultimately improve their lives. Depending on the particular behavior and the degree of progress that they are making, different texts that offer encouragement are displayed to them. There are over 300 separate feedback texts related to the different focus areas and stages of behavior change.

## Technology

We did not create these texts and the underlying behavior change methodology by ourselves. In partnership with ProChange Behavior Systems, Inc., a pioneer in behavior modification, we used strategies that included the texts mentioned above for use in the health and wellness field. These strategies complemented our own expertise in disease management and health coaching. Our collaboration, it is fair to say, has been mutually beneficial.

One of our biggest customers, which is based in Australia, agreed with this assessment. This organization did, however, want to be able to modify these different ProChange feedback texts in order to make them more Australian whenever they felt this was appropriate. In the initial product release, our Australian customer had to make an official request for any text changes through the defect tracking system. This meant that even modest modifications to the feedback text had to wait for the next product release.

To serve our Australian customer, we created an administration utility that allows a user to modify feedback texts whenever appropriate. This tool was, in fact, a miniature editor that accessed the same database as the web portal. Figure 1 provides a glimpse of this editor as it appears when the Design or WYSIWYG view is selected.

This ProChange Management utility assured that none of the underlying HTML tags would be deleted or corrupted and so the feedback text would remain displayable in a web browser. In addition, this module provided a way to guarantee that only authorized users would be able to modify the texts and that a record would be kept of where and how changes were made.

To provide these features, however, it was necessary to move the HTML formatted feedback texts from stand-alone files to the SQL Server database that the patient portal accessed. This form of storage prevented any kind of

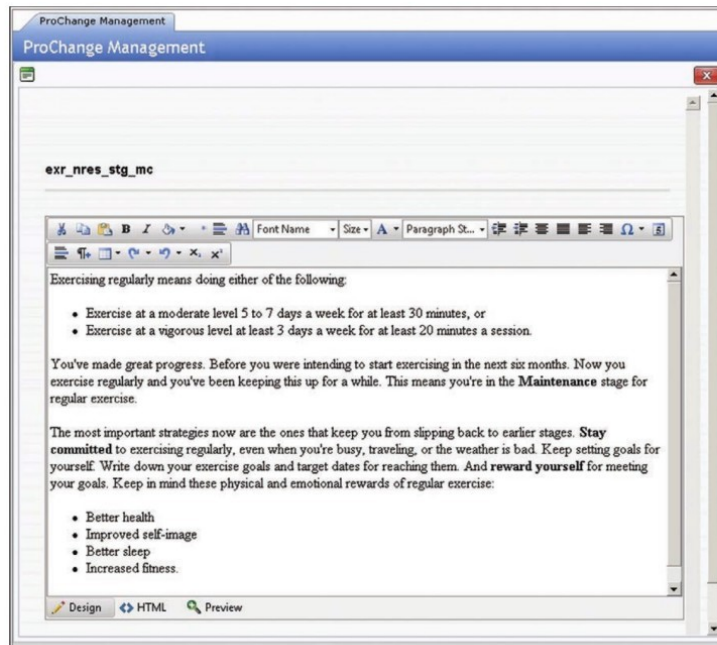


Figure 1: Feedback text in WYSIWYG mode.

change to the feedback text other than that performed by the ProChange Management utility.

### Solve one problem, create another

While the new utility was a boon to the Australian customer, it set the stage for a big problem with a different customer in Brazil. Exporting the HTML formatted text out of the database, having it translated, reviewed, approved and then reloaded into the database was an operation that no one had considered when implementing this WYSIWYG feature.

As discussed in the September 2015 issue of *MultiLingual*, we developed a tool to export translatable text from the database to XML files. This tool replaced an earlier utility that exported translatable strings to Excel files. Translating HTML text inside Excel cells was a daunting task and was one of the factors that led to the creation of the new XML utility.

While this tool did, in fact, address the import and export of translated data successfully, it left another one to resolve — preserving HTML formatting during customer review. We searched the internet for any kind of information on this peculiar situation and, alas, no help was to be found. We were stranded, metaphorically, on an unknown shore. It seemed that no one but us had ever had to deal with HTML text that was embedded in XML. And certainly no one had ever written anything on the review and modification of translated HTML formatted text that had been embedded in XML.

### Terra incognita

It might help to understand this challenge better if you look at an example taken from the XML file. Figure 2 shows how the English source text and Portuguese translation were originally defined. The English text is enclosed between the `<Origin>` and `</Origin>` tags, and the Portuguese



## Technology

```
<Record ID="707" FileName="exr_nres_stg_mc" TypeName="ProchaskaMobile">
  <Column Name="Content_Text">
    <Origin><p>You've made great progress. You've moved to the <b> Maintenance</b> stage
    for regular exercise. </p> <p> The most important strategies now are the ones that keep you
    from slipping back to earlier stages. <b> Stay committed </b>, even when you're busy,
    traveling, or the weather is bad. Keep setting goals for yourself and target dates for reaching them. And
    <b> reward yourself</b> for meeting your goals. Keep in mind these rewards of regular
    exercise: </p> <ul> <li> Better health</li> <li> Improved self-image</li>
    <li> Better sleep</li> </ul> <p> People who are successful at
    exercising regularly continue to use these strategies for a lifetime.</p> </Origin>
    <Translation>Você está no estágio da Manutenção do exercício regular.<p> <p>Na Manutenção,
    as estratégias mais importantes são as que ajudam a prevenir recaídas para um estágio anterior. Permaneça
    comprometido em se exercitar regularmente, mesmo quando estiver ocupado, viajando, ou o dia estiver
    chuvoso ou frio. Recompense a si mesmo. Escreva seu objetivo num papel e cole-o em um lugar onde possa
    vê-lo todos os dias.<p> <p>Lembre-se dos benefícios físicos e emocionais que você ganha com
    o exercício físico:<p> <ul style="margin-left: 40px; margin-left: ;"> <li>
    Aumento da auto-estima </li> <li>Melhora do sono</li> <li>Melhora da
    forma física</li> <li>Melhora da saúde</li> </ul> <p>Pessoas que se
    exercitam regularmente continuam usando essas estratégias a vida toda.<p> </p>&nbsp;</p>
    </Translation>
  </Column>
</Record>
```

Figure 2: Raw XML data with embedded HTML.

```
<Record ID="707" FileName="exr_nres_stg_mc" TypeName="ProchaskaMobile">
  <Column Name="Content_Text">
    <Origin>
      <p>You've made great progress. You've moved to the <b> Maintenance</b> stage for regular
      exercise. </p>
      <p>The most important strategies now are the ones that keep you from slipping back to earlier stages.
      <b> Stay committed </b>, even when you're busy, traveling, or the weather is bad. Keep setting
      goals for yourself and target dates for reaching them. And <b> reward yourself</b> for meeting your
      goals. Keep in mind these rewards of regular exercise: </p>
      <ul>
        <li> Better health</li>
        <li> Improved self-image</li>
        <li> Better sleep</li>
      </ul>
      <p>People who are successful at exercising regularly continue to use these strategies for a lifetime.</p>
    </Origin>
    <Translation>
      <p>Você está no Estágio da Manutenção do exercício regular.</p>
      <p>Nesse estágio, as estratégias mais importantes são as que ajudam a prevenir recaídas para um estágio
      anterior.</p>
      <p>Permaneça comprometido em se exercitar regularmente, mesmo quando estiver ocupado, viajando, ou o
      dia estiver chuvoso ou frio. </p>
      <p>Recompense a si mesmo. </p>
      <p>Escreva seu objetivo num papel e cole-o em um lugar onde possa vê-lo todos os dias.</p>
      <p>Lembre-se dos benefícios físicos e emocionais que você ganha com o exercício físico:</p>
      <ul style="margin-left: 40px; margin-left: ;">
        <li> Aumento da autoestima </li>
        <li> Melhora na qualidade do sono</li>
        <li> Melhora da forma física</li>
        <li> Melhora da saúde</li>
      </ul>
      <p>Pessoas que se exercitam regularmente continuam usando essas estratégias a vida toda.</p>
    </Translation>
  </Column>
</Record>
```

Figure 3: Formatted data after processing.

Tools and Services Showcase		
 <b>MEDILINGUA TRANSLATIONS</b> <h3>Medical Translations</h3> <p>MediLingua is one of Europe's few companies specializing in medical translation. We provide all European languages and the major languages of Asia and Africa as well as the usual translation-related services.</p> <p>Our 450-plus translators have a combined medical and language background.</p> <p>We work for manufacturers of medical devices, instruments, in-vitro diagnostics and software; pharmaceutical companies; medical publishers; national and international medical organizations; and medical journals.</p> <p>Call or e-mail Simon Andriesen or visit our website for more information.</p> <p><b>MediLingua BV</b> Leiden, The Netherlands simon.andriesen@medilingua.com www.medilingua.com</p>	 <h3>Individual Approach. Customized Solutions.</h3> <p>Ciklopea is an award-winning high-tech language solutions partner with more than a decade of experience in technical, medical, pharmaceutical, legal, marketing, business and general translation and localization.</p> <p>With a network of regional production offices and professionally trained, specialized and highly experienced native translators and management team, Ciklopea is the right choice for all translation and localization projects in SEE languages.</p> <p>Ciklopea is certified in accordance with ISO 9001:2008, EN 15038 and ISO 27001:2013.</p> <p><b>Ciklopea d.o.o.</b> Zagreb/Rijeka, Croatia • Belgrade, Serbia info@ciklopea.com • www.ciklopea.com</p>	  <p>THE RIGHT PEOPLE MAKE A WORLD OF DIFFERENCE</p> <h3>Executive Search and Brokerage Services</h3> <p>Finding the right person to direct your global business or figuring out how and when to sell your business... is not easy. We can help.</p> <p>Anzu Global provides:</p> <ul style="list-style-type: none"> <li>• Executive Search to find the right person to direct your globalization launch</li> <li>• Brokerage Services to find the right company to sell to once you reach your goals</li> <li>• Localization Staffing to support your ongoing translation, testing and localization management needs</li> </ul> <p>Anzu Global – the right people make a world of difference.</p> <p><b>Anzu Global</b> Stow, Massachusetts USA mklinger@anzuglobal.com www.anzuglobal.com</p>
 <h3>How to Be World Savvy</h3> <p>International Marketing Communications Market Research Marketing</p> <p><b>JFA Marketing</b> Minneapolis, Minnesota USA john@jfamarketing.com www.jfamarkets.com</p>	 <h3>Professional Translation and Localization Service</h3> <p>WISE-CONCETTI LTD (Vnlocalize) is the most renowned localization supplier in Southeast Asia. We don't just want to be the most reliable language service supplier but a professional company that provides our clients with real added value and the highest level of satisfaction.</p> <p>Why choose us for localization and translation service?</p> <ul style="list-style-type: none"> <li>• More than 80 linguists in 4 offices: Vietnam, Thailand, Indonesia, Malaysia</li> <li>• Audited company with ISO 9001:2008 certification</li> <li>• Full localization service: Translation – Engineering – Testing</li> </ul> <p><b>Wise-Concetti</b> Vietnam, Indonesia, Malaysia, Thailand michael@vnlocalize.com • www.vnlocalize.com</p>	  <h3>Europe's No. 1 Greek Localizer</h3> <p>Since 1986, EuroGreek has been providing high-quality, turnkey solutions, encompassing a whole range of client needs, for the following language combinations:</p> <ul style="list-style-type: none"> <li>• English into Greek</li> <li>• Greek into English</li> <li>• German into Greek</li> <li>• French into Greek</li> </ul> <p>All EuroGreek's work is produced in our Athens production center and covers most subjects:</p> <ul style="list-style-type: none"> <li>• Technical</li> <li>• Medical/Pharmaceutical</li> <li>• IT/Telecommunications</li> <li>• Economics/Legal</li> </ul> <p>All EuroGreek's work is fully guaranteed for quality and on-time delivery.</p> <p><b>EuroGreek Translations Limited</b> London, UK • Athens, Greece production@eurogreek.gr • www.eurogreek.com</p>



## Technology

**Record:**

**Column:**

**Origin:**

You're now in the **Maintenance** stage for regular exercise!

The most important strategies now are the ones that keep you from slipping back to earlier stages. **Stay committed**, even when you're busy, traveling, or the weather is bad. Keep setting goals for yourself and target dates for reaching them. And **reward yourself** for meeting your goal. Keep in mind these rewards of regular exercise:

- Better health
- Improved self-image
- Better sleep

People who are successful at exercising regularly continue to use these strategies for life.

**Translation:**

Você está no Estágio da Manutenção.

Nesse estágio, as estratégias mais importantes são as que podem prevenir recaídas para um estágio anterior.

Permaneça com o seu objetivo, mesmo quando estiver ocupado, viajando, ou o dia estiver chuvoso ou frio.

Estabeleça metas, escreva num papel e cole-o em um lugar onde possa ver todos os dias.

Recompense a si mesmo.

Lembre-se dos benefícios físicos e emocionais que você ganha com o exercício físico:

- Aumento da autoestima
- Melhora na qualidade do sono
- Melhora da forma física
- Melhora da saúde

Pessoas que se exercitam regularmente continuam usando essas estratégias a vida toda.

Figure 4: HTML formatted feedback text in WYSIWYG XML editor.

text between the `<Translation>` and `</Translation>` tags.

Even if our customer had not insisted on using a WYSIWYG editor for review, we could not have, in good conscience, proposed that anyone review translated text in a file such as this. Making changes to the raw XML file was scant improvement over editing an Excel cell that contained HTML text. Even if the customer had been willing to review the translations in this way, there would be the non-trivial risk that coding errors would be introduced in either the HTML or XML formatting.

The search was on, then, for a way to present the original and translated text in a way that both the XML and HTML tags were hidden from view. A

number of classic XML editors were downloaded. The criteria that we employed in evaluating them were as follows:

- ease of use
- low cost
- large file capacity
- WYSIWYG capability
- ability to mark text as approved

We quickly settled on the Serna XML Editor. It was relatively easy to use, very low cost (free), and it could load and store large files very easily. It was also a WYSIWYG editor that would not only hide the existing HTML tags but also enforce the proper creation of new ones. Unfortunately, it provided no means to indicate that a translated text element was approved by the customer,

but we hoped that this would not prove too big of a problem.

### A few minor glitches

Before we could begin using the Serna XML Editor with the customer, though, a few small issues had to be addressed. These were related to the proper display and processing of the translated data.

Not surprisingly, the biggest problem concerned HTML formatting. Serna supports XHTML formatted files very well, but neither the original English nor the translated feedback text was compliant. To be displayed properly in the editor, each text had to be re-encoded. In practice, this meant that for each `<p>` and `<li>` tag a corresponding `</p>` or `</li>` tag had to be added at the end of the particular HTML element. The fact that this was a one-time task was some consolation to the developer assigned to it.

In addition, in the raw XML file shown in Figure 2, you may have noticed that the HTML tags themselves are defined as codes such as `&lt;p&rt;` and `&lt;/p&rt;`. This was done automatically in order for the XML file to be parsed correctly on import. Before loading the XML file into the Serna Syntext Editor, however, it was necessary to swap `&lt;` for `<` and `&gt;` for `>`. This simple substitution was worked into a pre-Serna script.

Figure 3 shows how the data looks in a simple text editor after swapping in HTML tags.

After these tasks were complete, the original feedback text and its translation could be loaded into the Serna XML Editor, as shown in Figure 4.

After the changes were made, we had to upload the changed XML file into the database. It was at this point when we encountered the last glitch. We had to find a way to swap in HTML codes to replace the HTML tags. This was not as straightforward a procedure as substituting `<` for `&lt;` for and `>` for `&gt;`; since doing this would

## Technology

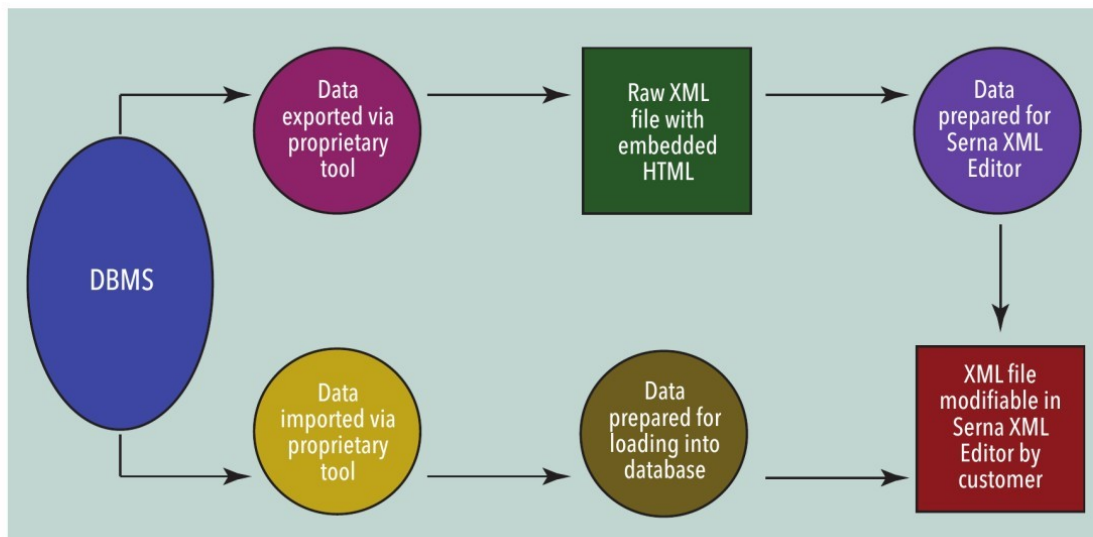


Figure 5: Data flow of translated HTML text.

corrupt the XML tags necessary to uploading the data. To get around this problem, the post-Serna substitution script looked only for instances of particular HTML tags and replaced them with the appropriate codes. After some trial and error, we were able to specify the correct substitutions and upload the customer-modified translations.

After we had tested and documented the new procedure thoroughly, we were ready to present it to the customer. Using the editor, a bilingual employee of that organization was able to make changes to the Portuguese text to make them more “Brazilian” while consulting the original English text. After receiving the file, we easily uploaded the modified file into SQL Server.

Figure 5 summarizes how we now exchange translated HTML data between the product database management system and the customer.

### Wrapping it up

While it is not likely that you will encounter exactly the same issue with translated data that we did, there are some things to keep in mind if you

are confronted with similar problems when moving translated data:

- don’t worry, a data exchange problem always looks bigger than it really is
- part of the solution might be available for free

■ a few small issues will invariably crop up — don’t panic; they have solutions

■ you can be happy with an outcome that meets only 80% of the initial requirements. [M]

- + Software localisation.
- + Web site localisation.
- + Technical and general translation.
- + Interpreting.
- + Third-party translation review.
- + Style guide creation.
- + Desktop publishing.
- + Linguistic advisory.
- + Terminology and document management.
- + Technical writing.
- + Multimedia translation.
- + Web site design, development and internationalisation.
- + Linguistic, typographic and style revising and review.
- + Video and audio tape transcription, including studio dubbing and voice-over.
- + Training on translation and localisation.

**HERMES,**  
THE SPANISH  
EXPERTS IN SPANISH  
TRANSLATION  
AND LOCALISATION:

PLEASED TO  
MEET YOU.



TRADUCCIONES Y SERVICIOS LINGÜÍSTICOS

Founded in 1991  
Código, 6, portal 2 - 3.ª I, Edificio Prisma,  
28230 Las Rozas, Madrid - SPAIN. Phone: (+34) 91 640 7640  
Parque Tecnológico de Andalucía Juan López Peñalver, 17, 3.ª, ofic. 6  
Edificio Centro de Empresas 29590 Campanillas, Málaga - SPAIN  
Phone: (+34) 952 020 525



Email: hermes@hermestrans.com www.hermestrans.com