Exchanging Translated HTML Data Painlessly (Almost)

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 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 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? Probably not 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?

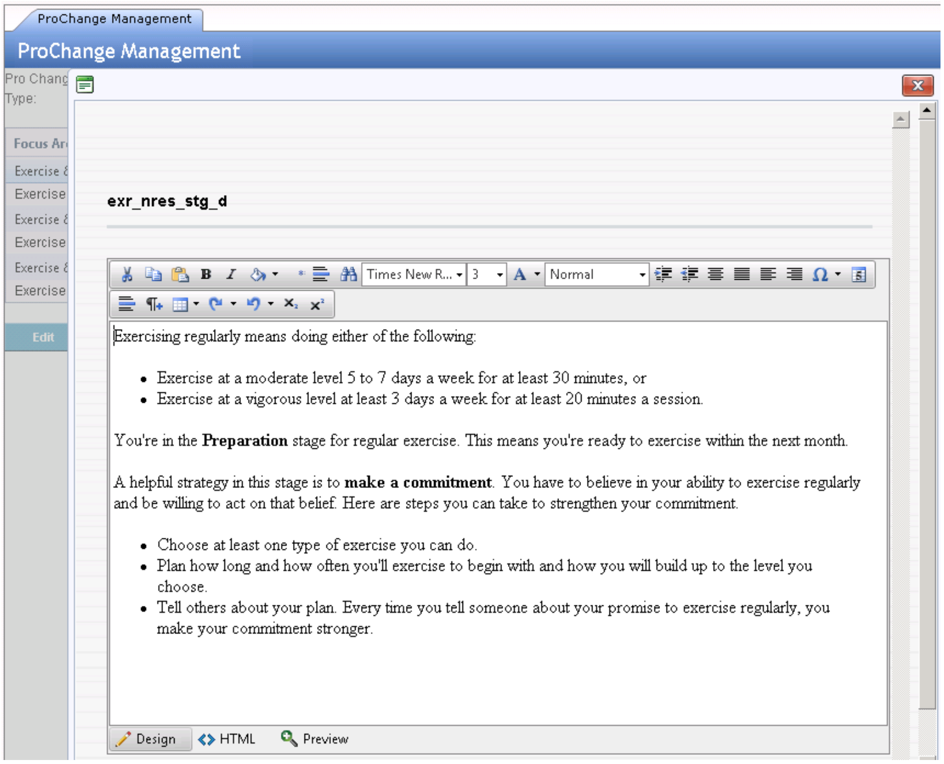
You might wonder how HTML text – 60000 words of it – came to be stored in a DBMS. Well, it is a longish kind of story. Our company is in the behavior modification business. One of our products is a web portal. The customers who buy this web portal - and a related mobile app - are typically organizations 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 login 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.

We did not create these texts and the underlying behavior change methodology by ourselves. In partnership with Pro-Change 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 – based in Australia – agreed with this assessment. This organization did however, want to be able to modify these different Pro-Change *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.



Exercising regularly means doing either of the following:

- Exercise at a moderate level 5 to 7 days a week for at least 30 minutes, or
- Exercise at a vigorous level at least 3 days a week for at least 20 minutes a session.

You're in the **Preparation** stage for regular exercise. This means you're ready to exercise within the next month.

A helpful strategy in this stage is to **make a commitment**. You have to believe in your ability to exercise regularly and be willing to act on that belief. Here are steps you can take to strengthen your commitment.

- Choose at least one type of exercise you can do.
- Plan how long and how often you'll exercise to begin with and how you will build up to the level you choose.
- Tell others about your plan. Every time you tell someone about your promise to exercise regularly, you make your commitment stronger.

Figure 1 –Feedback Text in WYSIWYG Mode[*]

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 standalone files to the SQL Server database that the patient portal accessed. This form of storage prevented any kind of 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 ‘Health Care Localization Process Case Study’ in the September, 2015 issue of *Multilingual*, we developed a tool to export translatable text from the database to XML files. This tool

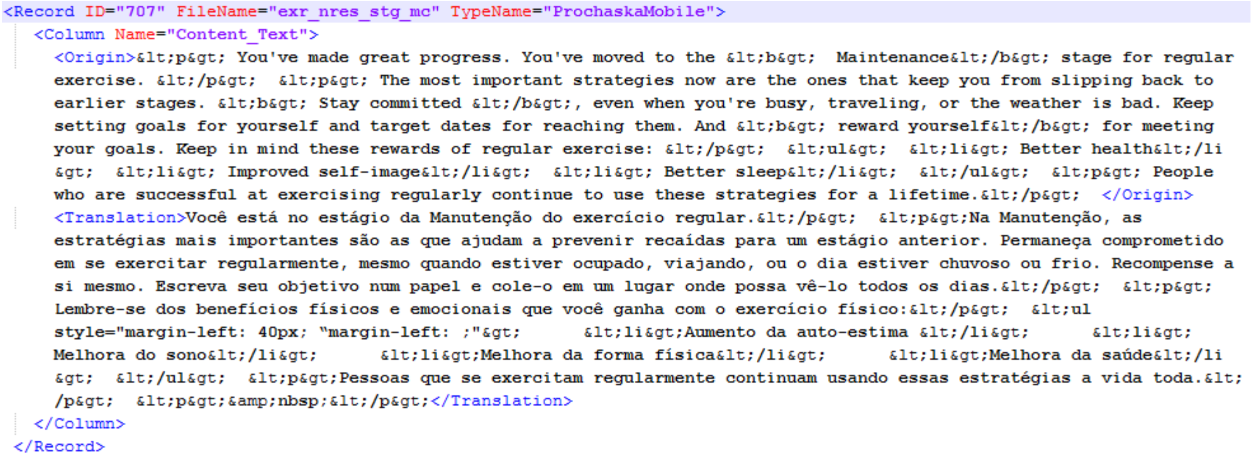
[*] Figures 1 through 4 contain copyrighted material belonging © 2015 Pro-Change Behavior Systems, Inc. All Rights reserved.

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 the peculiar situation where we had found ourselves. Alas, no help was to be found and we were stranded, metaphorically, on an unknown shore. It seemed that no one but ourselves 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 text between the `<Translation>` and `</Translation>` tags.



```
<Record ID="707" FileName="exr_nres_stg_mc" TypeName="ProchaskaMobile">
  <Column Name="Content_Text">
    <Origin>&lt;p&gt; You've made great progress. You've moved to the &lt;b&gt;  Maintenance&lt;/b&gt; stage for regular
    exercise. &lt;/p&gt;  &lt;p&gt; The most important strategies now are the ones that keep you from slipping back to
    earlier stages. &lt;b&gt; Stay committed &lt;/b&gt;, even when you're busy, traveling, or the weather is bad. Keep
    setting goals for yourself and target dates for reaching them. And &lt;b&gt; reward yourself&lt;/b&gt; for meeting
    your goals. Keep in mind these rewards of regular exercise: &lt;/p&gt;  &lt;ul&gt;  &lt;li&gt; Better health&lt;/li
    &gt;  &lt;li&gt; Improved self-image&lt;/li&gt;  &lt;li&gt; Better sleep&lt;/li&gt;  &lt;/ul&gt;  &lt;p&gt; People
    who are successful at exercising regularly continue to use these strategies for a lifetime.&lt;/p&gt;  </Origin>
    <Translation>Você está no estágio da Manutenção do exercício regular.&lt;/p&gt;  &lt;p&gt;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.&lt;/p&gt;  &lt;p&gt;
    Lembre-se dos benefícios físicos e emocionais que você ganha com o exercício físico:&lt;/p&gt;  &lt;ul
    style="margin-left: 40px; "margin-left: ;"&gt;      &lt;li&gt;Aumento da auto-estima &lt;/li&gt;      &lt;li&gt;
    Melhora do sono&lt;/li&gt;      &lt;li&gt;Melhora da forma física&lt;/li&gt;      &lt;li&gt;Melhora da saúde&lt;/li
    &gt;  &lt;/ul&gt;  &lt;p&gt;Pessoas que se exercitam regularmente continuam usando essas estratégias a vida toda.&lt;
    /p&gt;  &lt;p&gt;&amp;nbsp;&lt;/p&gt;</Translation>
  </Column>
</Record>
```

Figure 2 Raw XML Data With Embedded HTML

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 like this. Making changes to the raw XML file was scant improvement over editing an Excel cell that contained HTML text. Even if the customer was 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 i.e. 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 a problem.

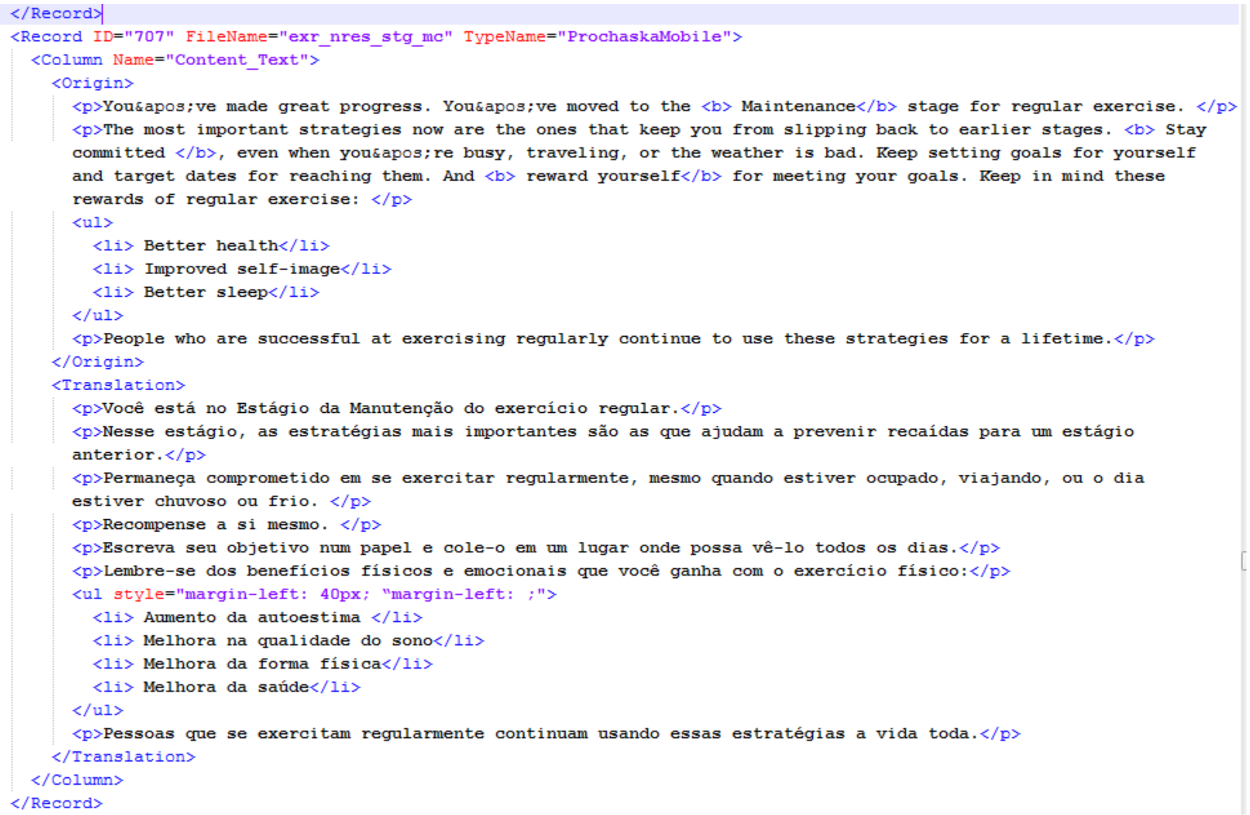
A Few Minor Glitches

Before we could begin using the Serna XML Editor with the customer, however, 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.

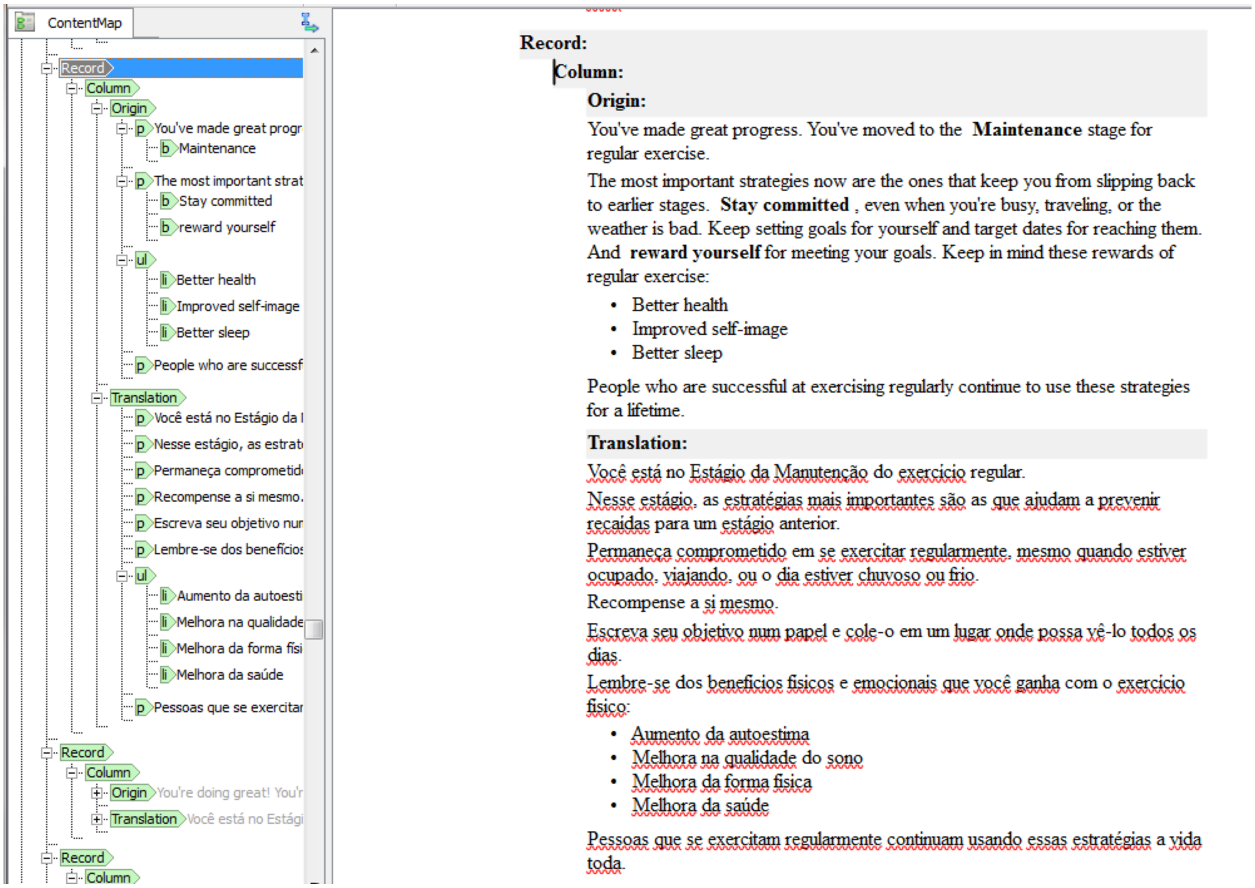
Figure 3 shows how the data looks in a simple text editor after swapping in HTML tags.



```
</Record>
<Record ID="707" FileName="exr_nres_stg_mc" TypeName="ProchaskaMobile">
  <Column Name="Content_Text">
    <Origin>
      <p>You&apos;ve made great progress. You&apos;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&apos;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

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.



Record:

Column:

Origin:

You've made great progress. You've moved to 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 goals. 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 a lifetime.

Translation:

Você está no Estágio da Manutenção do exercício regular.

Nesse estágio, 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.

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

After the changes were made, we had to upload the changed XML file into the DB. 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 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 DBMS and the customer.

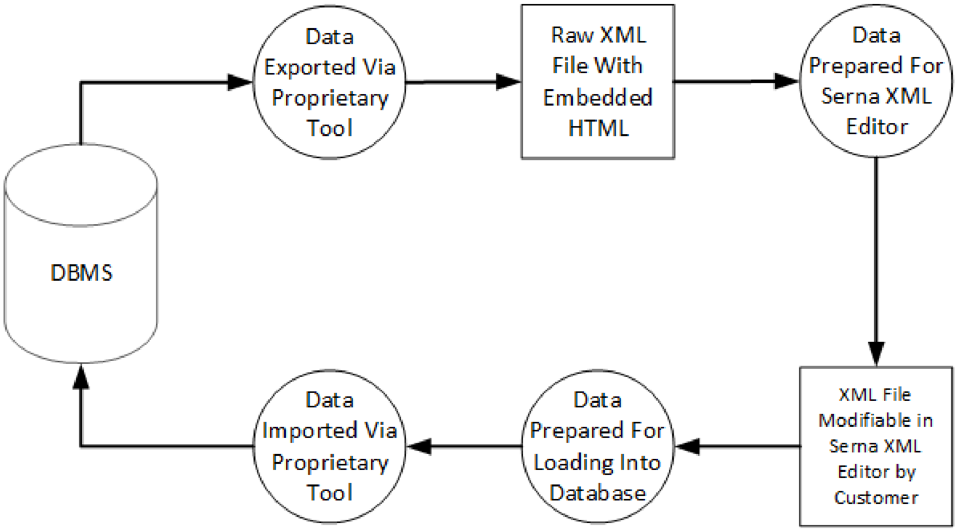

Data Exported Via Proprietary Tool
Raw XML File With Embedded HTML
Data Prepared For Serna XML Editor
DBMS
XML File Modifiable in Serna XML Editor by Customer
Data Prepared For Loading Into Database
Data Imported Via Proprietary Tool

Figure 5 Data Flow of Translated HTML Text

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 when 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