Text markup: rules and clean up

Content

Introduction

Within supplied data is a limited markup permitted. The purpose is that the receiving party can export data that is structured and has an unequivocal markup and suitable to administer an own styling.

Table of expected markup per field

The following markup can be expected on our website and in our web export database for the fields (including their translations) listed in this table:

Description	HTML Tag	Alternative	Data example	Display example	Specification	Permitted at [1] Dashboard text	Promotion name and text	Specific [2] specification
line end	 		Line 1 Line 2	Line 1 Line 2	Yes	Yes	Yes	Yes
italic	<i>…</i>	…	Regular <i>italic</i>	Regular *italic*	Yes	Yes	Yes	Yes
bold	…	…	Regular bold	Regular **bold**	Yes	Yes	Yes	Yes
underlined	<u>…</u>		Regular <u>underlined</u>	Regular underlined	Yes	Yes	Yes	Yes
summary	… …		- item 1 - item 2	• item 1 • item 2	Yes	Yes	Yes	Yes
HTML entities	For example: & < ™		For example: Black&Decker Verbruik < 50 watt Torx™	For example: Black&Decker Verbruik < 50 watt Torx™	Yes	Yes	Yes	Yes
hyperlink [3]	…		EZ-base	EZ-base	No	Yes	Yes	Yes
Other HTML	Everything that is not specified above, like HTML attributes		style="color:#ff0000;"		No	No	No	Yes
Spaces, line ends and tabs at the beginning or end of the field					No	No	No	No
Other text-markup [4]					No	No	No	Yes

[1] More information on cleaning up these and other fields: click here.

[2] If a supplier desires a different markup in his specification field for the own web shop or catalogue, this can be imported with the help of Specific specifications. These specifications will be pushed to EZ-web almost uncleaned and can be exported by choosing at Advanced export, Export settings for the option: "All specifications".
Note: when specific specifications are edited with the editor, the editor will process the same cleaning up and the same restrictions apply as with a dashboard or promotional text.

[3] About hyperlinks:
- If the protocol is missing in the href attribute of a hyperlink (http:// or https://) https:// will automatically be affixed. If the protocol is determined this will not be applied.
- Hyperlinks will aways be provided with attribute: target="_blank"

[4] With other text markup we mean:
- Double spaces
- Control characters (special characters that, for example, determine the sequence of a text)
- Manual enters (through the charachter set set up enters instead of HTML line ends)
- Manual tabs
-   (non-breaking spaces)

Clean up

Where and when will be cleaned up

Unpermitted markup in data is being cleaned up:
- During the nightly processing, when synchronizing data to our web site and web-export database.
- When this is edited with an editor in EZ-catalog. For the extended explanation see 'Markup in editors'.

The default as being maintained by EZ-base at this moment is in some cases a meddling with how this was advised in the past, or as this was produced by our own systems; this is being intercepted by the automatic clean up.

Which fields are cleaned up

This table shows which of the cleansing described in the next section is applied to these fields.

For the following translatable fields not listed in this table, only 1f, 3a, 3b and 3c are applied:
Article ERP descriptions
Article kinds
Article names
Article subgroups
Article types
Classification levels
Feature
Feature valutes
Merged features
Template proof sheet
Units

For all other fields not listed here, no cleanup will take place.

What will be cleaned up

Unpermitted markup will be replaced of deleted. This is described in the paragraphs below.

1   Replacement of HTML tags and attributes by comparable markup

a.   Permitted HTML tags will be replaced by their standardized variety in lower case characters

In the cleaned up data HTML tags will always be layed out in lower case characters. This results in an easier processing of possible future text-editing. Besides that, spaces will be removed from tags and the <br> tag will be layed out without slash.

Example:

      <I>Italic text</ I>
<B >Bold text</ B>

Becomes:

      <i>Italic text</i>
<Bold text

b. Summary characters (ASCII) will be replaced by HTML unordered lists (
…
…)

The combination of certain characters with a line end is being interpreted as a summary if the character is preceded by a line end (<br>) or is placed at the beginning of the text. As list-item the part of the summary character is taken till the end of the next line end. The list-items will be -if possible- placed within the same unordered-list.

This applies tot he characters ASCII 7 (•), ASCII 250 (·), ASCII 45 (-)

(also known as alt 7, alt 45 and alt 250)

Example:

· item 1
· item 2

Becomes:

item 1
item 2

c. HTML paragraph tags (<p>) will be replaced by line ends (<br>)

With the paragraph tag the open as well as the close-tag will be replaced by a line end.

Example:

Text 1<p>Text 2</p>Text 3

Becomes:

Text 1
Text 2
Text 3

d. HTML Sub- and Superscript will be replaced by the equivalent character set

The interpretation of HTML sub- and superscript tags in for example the catalogue production kan be a problem. Furthermore sub- and superscript can also be used in fields that do not support HTML. Because sub- and superscript from character set and through HTML tag often differ in appearance, it is unwanted that both are used alongside each other. Taking that in consideration, it is chosen to not support the HTML sub- and super tags, but transpose the most common characters to the corresponding Unicode characters.

HTML tag	Subset [1]	Supporting characters [2] [3]	Remarks
[…]	Superscript - Latin 1	0,1,2,3,4,5,6,7,8,9,+,-,=,(,)	
[…]	Superscript - Additional. Latin characters 1	a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,r,s,t,u,v,w,x,y,z	'q' is not being supported
[…]	Superscript - Additional. Latin characters 2	A,B,D,E,G,H,I,J,K,L,M,N,O,P,R,T,U,V,W	multiple charachters are not being supported
[…]	Superscript - Greek characters	β,γ,δ,θ,φ,χ	
_…	Subscript - Latin 1	0,1,2,3,4,5,6,7,8,9,+,-,=,(,)	
_…	Subscript - Additional Latin characters	a,e,h,I,j,k,l,m,n,o,p,r,s,t,u,v,x	multiple charachters are not being supported
_…	Subscript - Greek characters	β,γ,ρ,φ,χ	

[1] not all characters can show all the supporting sub- and superscript characters. Also with the conversion of export files to a different encoding problems may occur. The "Latin-1" characters have a vast support.

[2] commas are not supported, these are meant to make a distinction between the characters.

[3] not supported characters are being kept as "regular" character during the clean up.
   Example:
      _{opqr}
   Becomes:
      $_{op}q_r$

e. Inline style attributes will be converted to HTML tags

This concerns the following values for the *style* attribute:

Style attribute value	Converted to	Example source value [1]	Example converted
font-weight: bold;	...	text	text
font-style: italic;	<i>...</i>	text 	<i>text</i>
text-decoration: underline;	<u>...</u>	text 	<u>text</u>

[1] As example the tag is used, but this conversion will be done on every tag. Multiple styles will be placed within each other.

Example of conversion of multiple styles at once:

span with italic bold and underlining

Becomes:

<i><u>span with italic bold and underlining</u></i>

f. Non-breaking space ( ) and control characters (ascii karakters 000-031) such as manual enters, manual tabs and control characters will be replaced by a space

Example 1:

This is a space

Becomes:

This is a space

Example 2:

Text   1

Text 2

Becomes:

Text 1 Text 2

2   Removing unpermitted HTML tags and attributes and non -HTML markup

a. HTML tags that are unpermitted will be removed

With this will be attempted to maintain the text within a tag.

Example:

Left text Center text Right text

Becomes:

Left text Center text Right text

b. HTML attributes that are unpermitted, within HTML Tags that are permitted, will be removed

Example:

<strong style="color:red;">Bold text

Becomes:

Bold text

3   Removal of other markup that is not permitted

a. More than one subsequential space will be removed

Example:

T e x t   1

Becomes:

T e x t 1

b. Spaces/line ends in front of or behind the field will be removed

Example:

   Text 1
Text 2

Becomes:

Text 1
Text 2

c. Multiple line ends will be removed

Before and after a summary   if more than one line end stands behind one another
else   if more than two line ends stand behind one another

Example:

text 1

text 2

text 3

text 4

Becomes:

text 1

text 2
text 3
text 4

d. Empty summaries (<ul>) and empty list elements (<li>) will be removed

Example:

text 1
text 2
text 3

Becomes:

text 1
text 2
text 3

e. _x00??_ will be replaced by space

The values are as follows:

_x000D_

_x001D_

_x001F_

_x005F_

If multiple spaces occur, they are deduplicated.

Example:

text 1 _x000D_ _x005F_ text 2

Becomes:

text 1 text 2

Markup in import file

If an import file meets the rules, it will have the same markup on our web site, in our web-export database and with an export from EZ-catalog.

Import files with a markup that do not meet the rules will not be rejected: the imported data will be saved unaltered in EZ-catalog. Because the clean up only takes place with nightly processing or in the editor (and not "blocked at the door"), problems with data from data sources that cannot meet the stated rules are solved as much as possible.

Note: if data differs from the set rules, this can result in unwanted clean up.

Note: this has as a consequence that the data as shown on our web site and in our web-export database can have a different markup than with an export from EZ-catalog.

Therefore the advise is: supply import files that meet the set rules as much as possible.

Markup in editors

Data with which HTML is permitted, can be edited in the rich text editor within EZ-catalog:

- Article groups: Specification (header/footer)
- Profile: Dashboard text
- Promotion: Commercial text
- Specifications: Header text and footer text

These text editors and the Formalize screen come close to the representation in EZ-web as much as possible.

The operation principle of the editors is being explained in the paragraphs below.

1   Cleaning of markup with saving and/or selecting another field

The rich text editors in EZ-catalog are set up to produce data with which the markup can be processed without further clean up. To make this possible they instantly clean up an edited text when the editor "looses focus": for example by clicking outside the editor or saving. If the text at the moment of focus loss does not meet the set rules, the markup will be corrected immediately. The correction is shown in the editor.

2   Confirmation by user of the by editor processed clean up

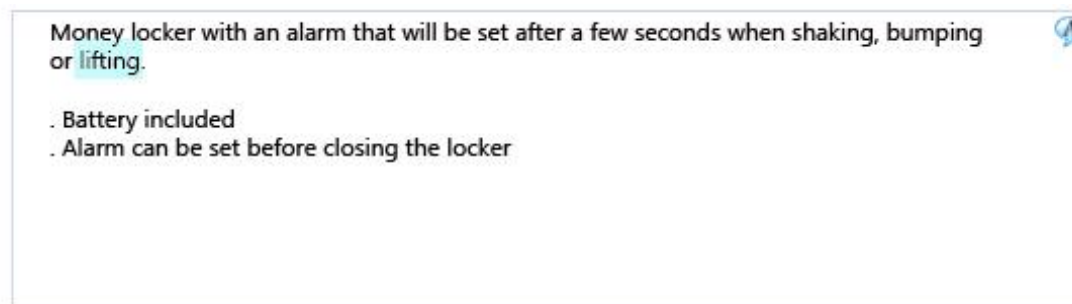2 Kinds of buttons are used in EZ-catalog to confirm an alteration.

If the editor has processed a clean up and user confirmation is required, the "Save" or "OK" button will remain clickable (see 1b and 2b). If the user tries to close the dialog differently there will be asked for a confirmation: after all, there are adjustments that have not yet been saved..

Example:

In the original specification the summary characters are layed out through alt 250 and <br>.
We change the word "lifting" into "shifting" and subsequently click "Save". In the screen, the content of the editor will be cleaned up and it will look as follows:

Specification	Money locker with an alarm that will be set after a few seconds when shaking, bumping or shifting. • Battery included • Alarm can be set before closing the locker

There will not be saved, the user can see if the cleaned up text meets up to expectations and can save afterwards.

When the altered specification will now be exported from EZ-catalog it will be with the cleaned up markup. (So: .. in stead of alt 250 and <br>)

Note: it can occur that the editor makes changes that are not instantly visible fort he user:
- *because they are too subtle.*
  For example removed double spaces, removed "invisible spaces" behind a piece of text, line ends that are removed behind the tekst.
- *because it concerns a technical change that visually has no effect.*
  Non-breaking-spaces that are replaced by normal spaces, alt 250 and <br> that are transformed to HTML inordered lists, HTML paragraph tags that are transformed in linebreaks (<br>), attributes that are removed.

Technically the difference between visible and unvisible changes is very hard to determine. Therefore it can occur that the user has to confirm an "unvisible" alteration.
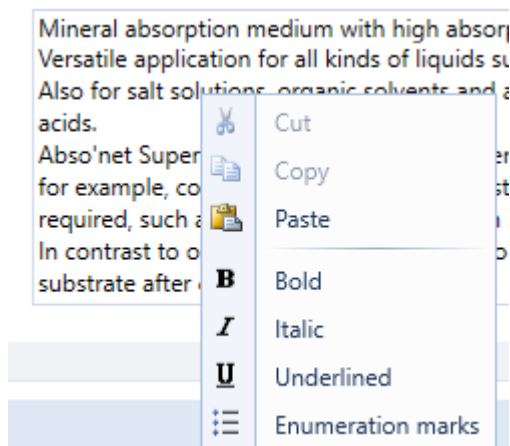
3 Clean up with pasting text in the text-editor

It's possible to paste text with markup from another program to the text-editor of EZ-catalog. Clean up regulations will be immediately processed to make sure that the pasted markup meets the current rules.

Note: in some cases, pasting markuped text to the editor can cause the cursor moves, characters disappear and the pasted text can be placed in an unintended place. Therefore, always check visiually before saving if the text and markup are as intended.

4   Inserting markup while working in the text-editor

With working in the editor, text can be provided with a certain styling by using the context-menu that appears by clicking with the right mouse button. This works in the ususal Windows-application way.

When a web address or an e-mail is entered in a text, it will automatically be changed to a hyperlink after pressing space or enter. If it is desired that the address is not changed in a hyperlink, the cange can be undone by placing the cursor directly behind the link and pressing backspace.

Through the editor it is not possible to alter the title attribute or markup of a hyperlink (to show a "pretty" link), but if the hyperlink is placed in the text through an import this will be remained.

5   Produce HTML entities

The editor converts certain reserved characters to HTML entities to prevent them from being interpreted as HTML:

Character	Entity
&	&
<	<
>	>
"	"

If the following tekst is entered in the specification field

then it will be saved as:

Storage temperature: > 7 degrees.