Tip Sheet

RSA Archer GRC Platform

Encoding in Data Feed Source File

September 2013

How to include HTML and XML characters in Data Feed source file.

Introduction

All data coming into the RSA Archer GRC Platform is sanitized to ensure that HTML security vulnerabilities (such as cross-site scripting, JavaScript, and others) are excluded from the content. This sanitation occurs whether data in the Platform is entered in the user interface (UI) or imported through a data feed. There is one difference between the UI and Data Feed. The UI knows what information in the content is markup and what information is not. The Data Feed Manager has no way to make that determination.

Content Entered Through the UI

To expand on that idea, let’s take the example of the following string: *<b>Text is bold</b>*. This string can be entered in the UI in three different ways,

- Directly into the Text Area field exactly as shown in the example
- In the HTML Editor exactly as shown in the example.
- Highlighted in the Text Area field and formatted with the Bold button in the toolbar.

When a user enters the string in the Text Area field as *<b>Text is bold</b>*, the content is stored in the database as *&lt;b&gt;Text is bold&lt;/b&gt;*. The *<b>* and *</b>* are considered part of the content and are encoded so the browser will interpret the characters as literal characters instead of markup. When displayed in the browser, the content is displayed as *<b>Text is bold</b>*. When a user enters the string in either of last two ways, the content is stored in the database as *<b>Text is bold</b>*. The *<b>* and *</b>* are considered markup and the string is displayed as **Text is bold.**

Content Imported from a Data Feed

Because a data feed does not have any way to determine whether the characters are part of the actual content or part of markup, the content needs to be properly encoded in the source file. So if the content should really be *<b>Text is bold</b>*, it must be encoded in the source file like this *&lt;b&gt;Text is bold&lt;/b&gt;*. Data Feed Manager saves the content in the manner it exists in the source file. If it was not encoded, the content is stored in the database as *<b>Text is bold</b>*, resulting in it being rendered in the browser as **Text is bold.**

Why does content like *Is 28 really < 35?* need to be encoded? While it may be clear to you that the less than sign (<) is not markup, it is not clear to the Platform. Part of the Platforms sanitization process cleans up improperly formatted HTML. Most browsers attempt to display HTML even when it is not formatted properly. The problem with the browser trying to be helpful is that it can allow vulnerabilities to pass through to the HTML. The Platform sanitizer attempts to fix up the markup before sanitizing it to ensure vulnerabilities do not slip through. When the sanitizer identifies that there is an opening markup character but no closing markup character, a closing markup tag is created automatically (and a few more things are added to make it valid markup). Therefore, content comprising characters that could be interpreted as markup characters must be encoded when they are not intended to be used as markup characters. A string such as *Is 28 really* < *35* must be specified in the source file as *Is 28 really &lt; 35*.

In the cases where you may not have control over the source file, for example content is coming from a third-party web site, you can create a Date Feed calculated field and use a function called *HtmlEncode([fieldname])*. Data Feed calculated fields are created on the Source tab of the Data Feed Manager. The fieldname is the name of the field to be encoded. The target application field can be mapped to the calculated field and the value will be properly encoded before the content is stored in the database.

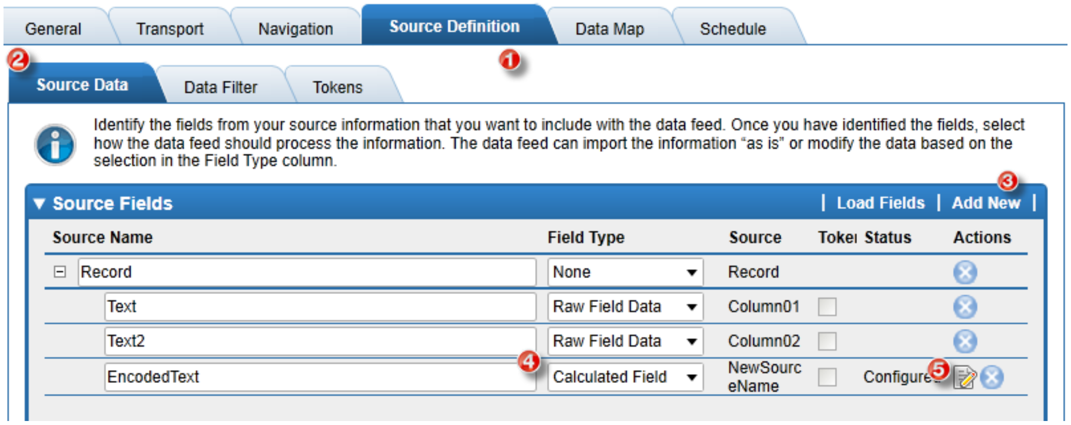
The following example provides the instructions for creating and mapping a calculated field. Note that Figure 1 illustrates steps 1 through 5:

1. Click the Source Definition tab.

2. Click the Source Data tab.

3. Click Add New.

4. In Field Type, select Calculated Field. The Source Name can also be changed to a more meaningful value if desired.

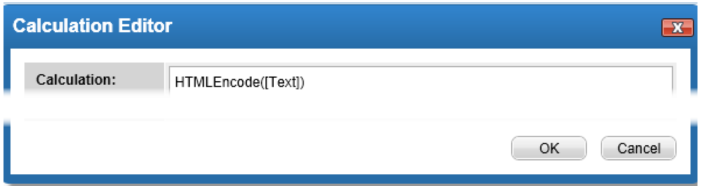


Identify the fields from your source information that you want to include with the data feed. Once you have identified the fields, select how the data feed should process the information. The data feed can import the information "as is" or modify the data based on the selection in the Field Type column.

Source Name	Field Type	Source	Toke	Status	Actions
⊟ Record	None	Record			⊗
Text	Raw Field Data	Column01	□		⊗
Text2	Raw Field Data	Column02	□		⊗
EncodedText	Calculated Field	NewSourceName	□	Configure	⊗

Figure 1

5. Click the Edit icon. The Calculation Editor is displayed, as shown in Figure 2.

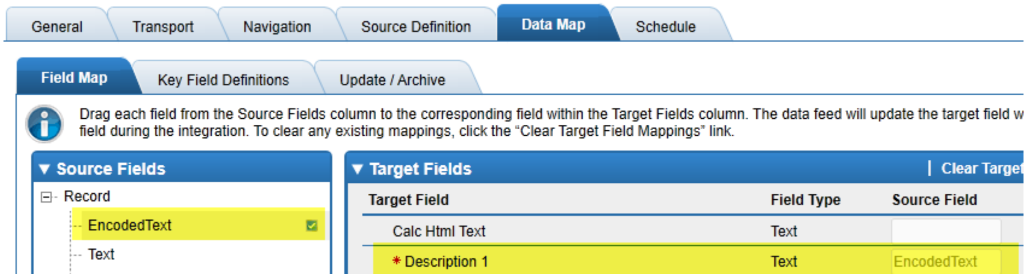


Calculation: HTMLEncode([Text])

Figure 2

6. In Calculation, enter HTMLEncode([Text]), as shown in Figure 2, and click OK

7. Map the calculated field to the desired field in the target application, as shown in Figure 3.



Drag each field from the Source Fields column to the corresponding field within the Target Fields column. The data feed will update the target field w field during the integration. To clear any existing mappings, click the "Clear Target Field Mappings" link.

Target Field	Field Type	Source Field
Calc Html Text	Text	
* Description 1	Text	EncodedText

Figure 3

The following is a list of the characters that must be encoded:

Character	Encoded Value	Description
“	&quot;	double quotation mark
&	&amp;	ampersand
‘	&pos;	apostrophe
<	&lt;	less-than sign
>	&gt;	greater-than sign

Support and Service

Customer Support Information	www.emc.com/support/rsa/index.htm
Customer Support E-mail	archersupport@rsa.com
RSA Archer Community	https://community.emc.com/community/connect/grc_ecosystem/rsa_archer
RSA Archer Exchange	https://community.emc.com/community/connect/grc_ecosystem/rsa_archer_exchange
RSA Solution Gallery	https://gallery.emc.com/community/marketplace/
RSA SecurCare Online	https://knowledge.rsasecurity.com/cleartrust/ct_logon.asp?CTAuthMode=BASIC&language=en&CT_ORIG_URL=https%3A%2F%2Fknowledge.rsasecurity.com%3A443%2F&ct_orig_uri=%2F

The Community enables collaboration among GRC clients, partners, and product experts. Members actively share ideas, vote for product enhancements, and discuss trends that help guide RSA Archer product roadmap.

The Exchange is an online marketplace dedicated to supporting eGRC initiatives. The Exchange brings together on-demand applications along with service, content, and integration providers to fuel the success of RSA Archer clients.

The RSA Solution Gallery provides information about third-party hardware and software products that have been certified to work with RSA products. The gallery includes Secured by RSA Implementation Guides with step-by-step instructions and other information about interoperation of RSA products with these third-party products.

The RSA SecurCare Online provides unlimited access to a wealth of resources on the Web, 24 hours a day. The secure system provides Members access to a support knowledgebase, to download current platform patches and bug fixes, to sign up for notifications, to manage your support cases and more.

Trademarks