Question <u>Mikael Toivonen</u> · Aug 17, 2017

## Charset problem: UTF-8 file downloaded via sFTP displays incorrectly

My issue is that I don't know why Ensemble processes incoming data in the wrong encoding.

Update: Issue was solved, I think, see my answer below.

our customer uploads UTF-8 encoded CSV files to our sFTP server. When I view the files on the server side, the scandinavian characters Ä and Ö (A and O with two dots on top of themselves, respectively) display correctly in the source file. When Ensemble downloads the files using a FTP Service, the characters display incorrectly in Ensemble.

I am unable to pinpoint the reason for this behavior and I was hoping this is easily solvable.

Edit: two additional details that I did not remember to mention but Gertjan brought up:

- In the inbound FTP adapter's settings I can't set "UTF-8" as the CHARSET. When connecting to a SFTP server (this is done by setting the "SSL Configuration" value in the service's "Additional Settings" category to "!SFTP"), the "CHARSET" field in the FTP service's "Additional settings" category cannot be "UTF-8". It causes an error message ("SFTP does not support ascii"). If you chooose "Binary", it works, but displays the scandinavian characters incorrectly.
- 2. In the RecordMap's settings I have set "UTF-8" as encoding.

Here is a sample of the original file:

ITEM\_CATEGORY | ITEM\_CAT2 | BR06002 VERISUONEN KANNATINNAUHA | SYDÄN/VERISUONI LEIKKAUSTARV. |

Here is the data that the Ensemble operation received:

```
<?xml version="1.0" ?>
   <!-- type: x.x.x.Record id: 1930 -->
   <Record xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns:s="http://www.w3.org/2001/XMLSchema">
        <ITEMCATEGORY>BR06002 VERISUONEN KANNATINNAUHA</ITEMCATEGORY>
        <ITEMCAT2>SYDÃN/VERISUONI LEIKKAUSTARV.</ITEMCAT2>
```

The character " 'A' with two dots on it " is displayed as " 'A' with a wave on it "

## #Ensemble #FTP

## Source

URL: https://community.intersystems.com/post/charset-problem-utf-8-file-downloaded-sftp-displays-incorrectly