

EML - Bug #5618

odd characters cause html display of eml to fail

06/05/2012 03:31 PM - gastil gastil

| | | | |
|------------------------|------------------|------------------------|------------|
| Status: | Resolved | Start date: | 06/05/2012 |
| Priority: | Normal | Due date: | |
| Assignee: | Margaret O'Brien | % Done: | 0% |
| Category: | utilities | Estimated time: | 0.00 hour |
| Target version: | Unspecified | Spent time: | 0.00 hour |
| Bugzilla-Id: | 5618 | | |

Description

While examining the HTML skins for EML, we found this example of a test doc caused the display to have an error (blank page):

<https://demo2.test.dataone.org/knb/metacat/knb-lter-knz.2.4/default>

and once the odd characters are removed, it displays normally:

<https://demo2.test.dataone.org/knb/metacat/knb-lter-knz.2.5/default>

These odd characters were found by

```
cat knb-lter-knz.2.4_mgb.xml | tr -d '\000\011\013\177' > foo
sort foo | uniq
```

â
Â°
Âµ

They are found on lines 39, 48, 577, 680, and 682 of knb-lter-knz.2.4 in the attached eml file. (This not likely to match the doc of the same pkg Id in the lter metacat, as that one is probably cleaned up.)

Note that showDataset can display that eml w/o removing the odd characters. They just appear as weird characters.

ie

<http://mcr-dev.lterinternet.edu/cgi-bin/showDataset.cgi?docid=knb-lter-knz.2.3>

where mcr-dev currently points to lava but that is likely to change.

History

#2 - 06/05/2012 03:43 PM - gastil gastil

To determine whether this bug is important, I used the pathQuery results from 30 April 2012 of all LTER eml abstracts. Of the 6826 eml docs, 240 of them contain some odd characters.

```
cat resultSet | tr -d '\000\011\013\177' | grep -v "^$" odd_chars | wc -l
240
```

So this affects less than 4 percent of the LTER eml docs in the LTER Metacat.

#3 - 06/06/2012 11:38 AM - gastil gastil

To clarify this bug, see knb-lter-bug.4103.2 versus knb-lter-bug.4103.3 in demo2.

The only difference is the presence of higher-order ascii characters in the abstract. In revision 3 they are commented-out.

#4 - 06/06/2012 07:50 PM - gastil gastil

Another example of this is the difference between

<https://demo2.test.dataone.org/knb/metacat/knb-test-nrs.569.3/default>

which does display

versus

<https://demo2.test.dataone.org/knb/metacat/knb-test-nrs.569.2/default>

which does not display, but instead fails with a "white-page".

The ONLY diff between revisions 2 and 3 is the pacakgeld and Bancroft's office

vs

Bancroft & B's office

in the abstract.

(revision 1 had denyFirst.)

#5 - 06/07/2012 12:24 AM - ben leinfelder

These are unfortunate errors, but are most likely do to copy-and-paste from a [MS Word] document into the metadata file.
I've added better error reporting during the transform so there's now an indication that rendering was not possible when the character is encountered.

#6 - 03/27/2013 02:31 PM - Redmine Admin

Original Bugzilla ID was 5618

Files

| | | | |
|--------------------------|---------|------------|---------------|
| knb-lter-knz.2.4_mgb.xml | 32.8 KB | 06/05/2012 | gastil gastil |
|--------------------------|---------|------------|---------------|