Metacat - Bug #4130

data tables in KNB display dataTable->physical->objectName

06/08/2009 11:56 AM - Oliver Soong

Status:	Resolved		**Start date:**	06/08/2009
Priority:	Normal		**Due date:**	
Assignee:	Michael Daigle		**% Done:**	0%
Category:	metacat		**Estimated time:**	0.00 hour
Target version:	1.9.2		**Spent time:**	0.00 hour
Bugzilla-Id:	4130			

Description

KNB names data tables according to the objectName element within the physical element. By contrast, Kepler's EML 2 Dataset actor displays the entityName. This is mostly confusing for a few dataTables for which the entityName is not the same as the objectName. For example look at judithk.609.27, entityName=Dailyrainl2005.txt and objectName=rainfall2005.txt. I'm not sure which what the best way to handle this is, but it seems more intuitive to refer to the dataTable by the entityName rather than the objectName of the physical container.

History

#1 - 02/25/2010 03:04 PM - Jim Regetz

Yes, this should be changed. The current behavior is especially problematic for data packages that separately describe multiple tables associated with a single physical object. For example, if three tables from an Access database are described, the registry listing of data tables repeatedly displays the common objectName (rather than distinct entityNames) for each table:

```
Data Tables, Images, and Other Entities:
    -----------------------------------------------
    Data Table: foo.mdb (View Metadata)
    Data Table: foo.mdb (View Metadata)
    Data Table: foo.mdb (View Metadata)
```

See e.g.:
http://knb.ecoinformatics.org/knb/metacat/nceas.904.11/nceas

Additional comments from Matt:

    I agree with your proposal. It should be an easy change. We did
    actually discuss this before when we made the stylesheet -- in the more
    typical situation, there is a 1:1 relation between entity and object, so
    we decided to list the name of the physical file that is being
    downloaded. The rationale was that when downloading, the objectName is
    going to be used to name the file on disk, so we wanted the display to
    correspond to the filename. In this case, where there is one object
    with three entities, it might be worthwhile to list both, something like:

    Data table: speciesList -- foo.mdb (View Metadata)
    Data table: surveyData -- foo.mdb (View Metadata)

#2 - 02/26/2010 02:35 PM - Michael Daigle

Try to get this into the 1.9.2 release if possible.

#3 - 03/12/2010 12:32 PM - Jim Regetz

It turns out showing both names looks a little odd when they are identical (which is usually the case), and potentially quite ugly when those names are long.

Two options:

1. Implement some conditional logic -- if both names are the same, only show it once, but if they differ, show both as suggested in the original bug report. Is this hard? Note that even if we do this, there might still be cases where the names differ but are both long and ugly to display together.

2. Only show the entityName. The original motivation for showing the objectName is that it becomes part of the filename of the downloaded object, whereas the entityName might be totally different. However, I don't think this is a big deal for the following reasons:

- There is plenty of precedent for this on the web - in fact, it's quite common that the actual filename of something you download won't be apparent before you download it.
- By design, the filename still has plenty of info to allow someone to figure out what dataset the file actually contains.
- If you click the "View Metadata" link, the object name appears right there near the top under the Physical Structure Description section.
- In most cases, it won't matter because the two names are the same.
- In cases like the DP I gave as an example, it won't matter because there are no tables to download anyway; AFAIK this will be true of all DPs produced by importing from MS Access in Morpho.

I'm leaning towards #2...

#4 - 03/12/2010 01:23 PM - Matt Jones

Jim,

I think you should talk this over with Mark -- he was the one who made the strong case for using objectName, and that the link text should reflect the filename that is downloaded. Mark felt pretty strongly about it and pushed for the current design. Getting his input before we decide to completely change the link text is worthwhile.

In my opinion, entityName is a lot lett predictable than objectName. In morpho, the user suppies the entityname as a free text field, whereas the objectName is taken from the original filename of the object. I think using entityName will end up with a lot of nonsensical text. But that's just an impression of mine. Doing a quick survey of the contents of entityName and objectName fields for the KNB datasets would probably be helpful here.

#5 - 03/16/2010 01:40 PM - Jim Regetz

(In reply to comment #4)

> In my opinion, entityName is a lot lett predictable than objectName. In
> morpho, the user suppies the entityname as a free text field, whereas the
> objectName is taken from the original filename of the object. I think using
> entityName will end up with a lot of nonsensical text. But that's just an
> impression of mine. Doing a quick survey of the contents of entityName and
> objectName fields for the KNB datasets would probably be helpful here.

An empirical contribution...

Shaun extracted a slew of entityNames and objectNames from the cached EML docs on KNB, excluding many highly regularized LTER docs but presumably including just about everything ever uploaded via the registry or Morpho. I then parsed the XML result set in R.

Out of about 5000 unique objects, ~1200 have an entityName that differs from its objectName. In many cases, the difference is inconsequential (e.g. one has .txt appended, the other doesn't). In cases where there are material differences, the entityName is generally more informative, hence arguably more useful to display in the registry view. For example, "Cover Data for 1957-1966" rather than "cover12.xls".

There **is** a tendency for the entityNames to be slightly longer; the difference is only 1 character on average, but about a dozen entityNames are over 80 characters whereas only one objectName is that long. But I'm really not seeing entityNames that I would call nonsensical, and they are less likely to be indecipherable than are the objectNames.

#6 - 03/17/2010 03:29 PM - Michael Daigle

Changed the xsl to only use the entityName

#7 - 03/27/2013 02:25 PM - Redmine Admin

Original Bugzilla ID was 4130