

Morpho - Bug #3454

Fixing docid conflicts in metacat or local file

07/16/2008 03:51 PM - Callie Bowdish

Status:	Resolved	Start date:	07/16/2008
Priority:	Normal	Due date:	
Assignee:	Jing Tao	% Done:	0%
Category:	morpho - general	Estimated time:	0.00 hour
Target version:	1.7	Spent time:	0.00 hour
Bugzilla-Id:	3454		
Description			
<p>Having Morpho profiles information loaded on different computers can cause file conflicts. One way to prevent this is by having each profile have a different identifier prefix. This does not require that a user have a different network account or even different profile name. However the part that should be unique is the Identifier prefix which is used as the scope for the document id. The following text can be added to the Data Package Identification part of creating a profile. It is stored in the profile under the scope element.</p> <p>Text to replace the current text:</p> <p>Enter a short identifier prefix. All data packages you create under this profile will bear this prefix. To prevent data package name conflicts it is recommended that you create a different prefix for each computer you load Morpho on.</p> <hr/> <p>Please note that if the naming conventions, such as global lisd names, for data packages change the use of this to prevent file conflicts may become obsolete.</p> <hr/> <p>Currently bug 3120 describes a section of the profile creation form that is not displaying correctly. It may be convenient to fix both of these as the same time.</p>			
Related issues:			
Is duplicate of Morpho - Bug #2309: File name conflict resolution when using ...		Resolved	12/06/2005
Is duplicate of Morpho - Bug #2360: Document Identifier Conflict Resolution		Resolved	02/16/2006

History

#1 - 07/16/2008 04:25 PM - Matt Jones

I think people should be able to choose whatever prefix they want. They shouldn't have to change prefixes just to change computers. That Morpho does not properly (and seamlessly) handle identifier conflicts is the true bug, and the one that should be fixed.

#2 - 07/17/2008 09:33 AM - Callie Bowdish

Perhaps the priority needs to be fixing the conflict problem, yet I think that adding the text to the profile install could help with some of the headaches around profile problems. When someone loads Morpho on a new computer they already have to create a profile again.

I've tried to think of a way to map out the problem areas with using the same scope on different computers and with different Metacat servers. I haven't had much luck with looking at the numerous angles to it all.

One of the tricky parts is with the data tables. When people save a data package to the network with one table and then add another table, would a dialog box have to pop up telling the user that their first table already exists, and do they want to replace it. I have found the most problems with the data tables. Having a dialog box come up with the EML file saying it already exists seems to work sometimes. I have not seen a dialog box come up with a data table (same id and version number) that already exists on Metacat.

With finding a solution, I think there are many scenarios to be mapped out.

Note that the potential phrase to be added states, that "it is recommended" rather than have to change.

To prevent data package name conflicts it is recommended that you create a different prefix for each computer you load Morpho on.

I agree that adding the phrase needs some thought before it is done. Another thing I think that needs to be done with the profile creation is having some restraints on the scope name. I think spaces and special characters can cause problems. It might be helpful to include some kind of constraint when people fill in the scope field.

#3 - 07/18/2008 03:05 PM - Jing Tao

In order to avoid docids conflict, I added some code to keep lastid have newest value.

1) Now morpho will compare the last id in remote metacat, local file system and profile when morpho starts. The biggest one will be the new value in the profile. It didn't check local file system in previous version.

2) Now morpho will check the last id in remote metacat when use change the metacat url.

3) When you start a morpho at a offline computer, morpho will get lastid by comparing lastid in local and profile. However, when the computer gets network connection, morpho will check remote metacat and update lastid in profile if necessary.

Those measures are already done and I think it can make sure morpho wouldn't create a conflicted docid when it creates a new package.

However, if people work around on couple computers or different metacats, it may still cause problem:

Here is one scenario:

User opened an eml doc jones.2.1 with data file jones.3.1 from local system. He edited jones.2.1 and want to save it in both local and remote metacat as jones.2.2. However, jones.2.2, joens.3.1 already exist in remote metacat, which created from another morpho in another computer. What we should do?

- 1) Give user's options to choose replacing the exist id by new id (last id + 1) or updating revision number?
- 2) Give user no options, but an alter. Then update revision number.

#4 - 07/25/2008 12:38 PM - Callie Bowdish

I did some testing on the changes that Jing added to Morpho to help with id collisions.

I think we are still having problems when data tables are added to data sets. Also on the SANParks skin there is a data upload feature that uses the logged in individual scope which we observed having problems when people used a fresh load of Morpho with the same profile.

Here is my report to Jing

The first time I tried two different profiles I did not work with a data table addition and did get the Warning! Problem Saving Data: Id already in use dialog box.

Here I think both computers had the same id at the beginning. Then I saved a data package on one. Then went to the other machine and saved it on the other - that is where I got the warning rather than it choosing the next id.

I did some more testing and it looks like there are still some problems with saving data packages using the same profile on two different computers when a data table is involved. Today I used the Head for Morpho on Linux and Mac machines.

First I created a data package which updated correctly from my Linux machine the last Id was 800 something and then when the data package was saved it saved as 927.1 which was good. Then after saving the data package on Linux machine I imported a table which was named 928.1 and save this to the network with the data package 927.2

But when I next used my Mac machine with the same profile to create a data package, saving it both local and to the network, it was saved it as 928.1. The metacat server has a file called bowdish.928.1 but it is a data table associated with 927.2. bowdish.928.1 eml file is only on my local computer. On metacat it is still 928.1 data table.

Yesterday

I created a data set on Windows bowdish.925.1 saved up to the Network and local.

Then I went to my Mac machine and save an eml file bowdish.926.1 to the Network and local.

Next I went back to my Windows machine and added a table to bowdish.925.1. the table was named bowdish.926.1. and the data package was upped to bowdish.925.3. If you click on the download file you do not get a table bowdish.926.1. Instead the eml file is opened. <http://dev.nceas.ucsb.edu/knb/metacat/bowdish.925.3/nceas> (click on the download link and you get an eml data set description instead of a data table to download)

#5 - 07/29/2008 01:50 PM - Jing Tao

- Bug 2360 has been marked as a duplicate of this bug. ***

#6 - 07/29/2008 01:51 PM - Jing Tao

- Bug 2309 has been marked as a duplicate of this bug. ***

#7 - 08/01/2008 01:52 PM - Jing Tao

I talked with ben about the id confliction issue today. We concluded that before writing file to metacat or local file system, we should check if the id existed or not. Here are some scenarios:

1. Save local. Check if local file system already has the id or not before writing file. If it does exists, docid will be increased silently to the max id when

morpho saves new file; or user will be asked to increase revision number or increase docid number when morpho updates file. Then morpho will write the file to local system with the new docid or new revision. The data package frame will only show the local icon.

2. Save metacat. Check if metacat already has the id or not before writing file to metacat. If metacat does have this id, docid will be increased silently to the max id when morpho saves new file; or user will be asked to increase revision number or increase docid number when morpho updates file. Then morpho will write the file to metacat with the new docid or new revision. The data package frame will only show the metacat icon.

3. Save both. Check if metacat and local system already have the id or not before writing file to metacat and file system. If they do have this id, docid will be increased silently to the max id when morpho saves new file; or user will be asked to increase revision number or increase docid number when morpho updates file. Then morpho will write the file to metacat and local system with the new docid or new revision. The data package frame will show both metacat and local icons.

4. Synchronize (from local to metacat). Check if metacat already has the id or not before writing file to metacat. If metacat already has, user will be asked to increase docid (max docid in both metacat and local system) or revision. Then write file to both metacat and local file system with the new docid or revision (so local system will have duplicate copy of this package). Data package frame will show both metacat and local icons.

5. Synchronize from metacat to local system. Check if local system already has the id or not before writing file to local system. If local system already has it, user will be asked to increase docid or revision. Then write file to local system. Data package frame will only show local icon.

Above rules will be applied to both data and metadata files.

#8 - 08/01/2008 02:51 PM - Jing Tao

By the way, i changed the order when morpho saves data and eml. Before it save eml first, then data tables. Now it is data table first, then eml. In this order, any docid change solving data table id confliction will be shown in eml documents.

#9 - 10/31/2008 11:35 AM - Jing Tao

Before saving metadata and data into metacat/local, morpho will detect docid conflict first. If there is a docid conflict, morpho will show a dialog to user to ask them if they want to increase revision or select a new id.

#10 - 08/12/2009 03:50 PM - Jing Tao

move to 1.7

#11 - 03/27/2013 02:23 PM - Redmine Admin

Original Bugzilla ID was 3454