

Metacat - Bug #2566

need to set filename for download files

10/17/2006 08:17 AM - Matt Jones

Status:	Resolved	Start date:	10/17/2006
Priority:	Immediate	Due date:	
Assignee:	Shaun Walbridge	% Done:	0%
Category:	metacat	Estimated time:	0.00 hour
Target version:	1.9	Spent time:	0.00 hour
Bugzilla-Id:	2566		
Description			
<p>Currently metacat does not explicitly set the filename on downloaded files to anything sensible. Instead, it just uses a filename like 'metacat' for anything downloaded (except zip files, which are now named with a zip extension using the package id). The proper name for a data file should probably be the original entity name plus and appropriate extension based on the content type. A call like this would do the trick:</p> <pre>response.setHeader("Content-Disposition", "attachment; filename=" + docId + ".zip");</pre> <p>For XML files (including EML documents), its probably best to name the file using the pattern "namespace.id.rev.xml" or something similar. We also want to be sure we are setting the Content-type properly so browsers know how to deal with the downloaded file.</p>			
Related issues:			
Blocked by Metacat - Bug #3084: Need to be able to set filenames on data uploads			Resolved 01/22/2008

History

#1 - 11/09/2006 11:27 PM - Matt Jones

Partial fix checked in, definitely an improvement. Changed the filename that is used when data files are downloaded and when XML files are downloaded in XML format. Now, data files use the format "docid-docname". XML files use the format "docid.xml".

The major outstanding issue is that docname is not always set very meaningfully by clients. If the file uses a filename when uploading, then it will likely have an appropriate extension and our file will also share that extension. If the client did not upload an appropriate extension, our file will lack one. We could use the mime type to determine if the extension in docname, if any, is appropriate, and if not then replace it. We'll leave it like this and discuss with others.

#2 - 11/09/2006 11:39 PM - Matt Jones

Also note that some clients, for example Morpho, use pretty meaningless filenames when uploading data. When Morpho imports a file locally, it stores it based on its docid (e.g., a data file with docid jones.1.1 would be stored in the file '~/.morpho/profiles/jones/data/jones/1.1'. Morpho places the original filename as a suggestion in the ENtityName metadata field in EML, but the user is free to change that. When Morpho uploads the file to metacat, it does so using the docid (jones.1.1) and the filename (1.1), which is recorded in metacat as docname=1.1

As a consequence, when creating download filenames from morpho for the data file jones.1.1, the filename that metacat will set will be 'jones.1.1-1.1' which is only marginally useful. We probably want to change what Morpho does here when uploading data files.

#3 - 12/18/2008 06:15 PM - Shaun Walbridge

Further fixes applied in [r4685](#).

The FGDC special case still needs handling, pending discussion of the doctype used for FGDC documents and a decision that this matters (FGDC data are delivered using a special mechanism currently which provides all data within a zip archive).

The last step on this is to iterate over existing EML documents and find their dataset children. For each of these children, pull the objectName and replace the docname with this value for all documents where docname = accession.rev, which were created previously by Morpho or older versions of the Registry.

#4 - 12/24/2008 03:13 PM - Jing Tao

Data file from morpho is using object name as docname in metacat.

#5 - 01/06/2009 12:34 PM - Shaun Walbridge

<below is copied from a recent email discussing the bug>

Bug resolved, with two outstanding issues:

- The metadata docid is passed as a parameter from the download links, and embedded in the EML XSLT. The best way to solve this would be an additional column in the xml_documents table listing parent docid for BIN objects, so that we could easily query the list, and wouldn't need to use the hacked stylesheet approach, could determine the names when coming in from other pathways (e.g. direct download without GET parameters). But I didn't want to implement a major change with an imminent release, and this should be discussed before implementation.

- The FGDC special case still needs handling, pending discussion of the doctype used for FGDC documents and a decision that this matters (FGDC data are delivered using a special mechanism currently which provides all data within a zip archive). The work Chris Barteau did isn't currently in the main Metacat servlet, but uses different code for its FGDC handling, unused by the majority of the skins. The FGDC documents (currently numbering 11) are also currently receiving the `doctype` of `metadata`, which is certainly wrong and should be updated.

Issues solved:

1. When documents are created or updated, the objectName element should be copied into the `docname` field within the `xml_documents` table. This requires that clients send correct filename when delivering the data (so Metacat can correctly set `docname` using the registerDocument f'n). Jing recently updated Morpho to do the right thing here, as does the Perl Registry client. From now on, all new documents should have the correct docname.

2. Preexisting documents which were created by older releases of Morpho or uploaded in other ways (replication? Not sure how other systems generate `docname`). These can easily be seen by looking for documents which have generic docnames:
SELECT COUNT FROM xml_documents WHERE docname = docid || '.' || rev AND doctype = 'BIN';

To eliminate these misnomers, I searched for all binaries linked to from metadata documents:
SELECT nodedata, docid FROM xml_nodes WHERE nodetype = 'TEXT' AND nodedata LIKE 'ecogrid%'

These were then fed into a Perl script (src/perl/eml_get_objectnames.pl) which read the related metadata documents, determined the correct filename, and output an SQL script which can be used to update misnomers. This was tested and applied to KNB, so all previously misnamed files are now correct.

3. The download naming logic in Metacat was partially correct, set as `docid-docname` but lacked metadata docid. The name generation was rewritten to match the sensible names logic.

#6 - 03/27/2013 02:20 PM - Redmine Admin

Original Bugzilla ID was 2566