

## EML - Bug #5427

### round-trip encoding of missing values uploaded then queried from a db table is lost

06/23/2011 02:37 PM - gastil gastil

<b>Status:</b>	New	<b>Start date:</b>	06/23/2011
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Jing Tao	<b>% Done:</b>	0%
<b>Category:</b>	datamanager	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	Unspecified	<b>Spent time:</b>	0.00 hour
<b>Bugzilla-Id:</b>	5427		

#### Description

Round-trip encoding of missing values in EML datasets uploaded to and queried from a database table is lost with current version of DML.

Short link to this doc is <http://goo.gl/2mq9T>

Might pertain to:

EML Data Manager Library

PASTA workflows

EML-parsed data delivered from a DML-loaded database

Might be a feature request for a future iteration,... Far in the future!

This is about handling missing values in data tables, whether they are stored in a database table as codes or as nulls, and how they are then coalesce()-ed in a VIEW of that table.

Background:

EML allows multiple missing value codes for the same data table column, as there may be more than one reason for missing a value. This is good. The EML Data Manager Library (DML) compares missing value codes with a string comparison, not a numeric comparison. So -9999.0 does not match -9999. A numeric column may have a text missing value code such as NaN or na in a column of float type.

Issue:

When the data is inserted into the database table, obviously the non-numeric string missing value codes cannot be inserted literally. They are inserted as nulls. I looked at that part of the DML code. Any datum which matches one of the missing value codes for its column gets inserted as a null.\*

Since any missing value code is collapsed into a simple null, then the original information about what kind of missing value code it was has been lost.

When querying that data table, either a VIEW specifically written for that table or the code constructing that query could use the EML to assign a missing value code to nulls using coalesce(), but only if there were only one missing value code per column. Where multiple codes exist, it would be wrong to just arbitrarily assign the first-listed code to all nulls of a column.

Proposed Solution:

Alternatively, the DML could store missing value codes, assigning numeric codes to replace non-numeric codes where necessary (a tricky feat since it implies knowledge of the range of valid values, which may not be specified in the EML.) Then a corresponding query would have to be stored as a VIEW, with a CASE wrapping that column to translate back to the original codes.

Details:

\*Notes relating to actual java code are below.

In the DML DatabaseAdapter class, in the method generateInsertSQL() gets three inputs: the attributeList, the tableName, and oneRowData. Each attribute value is compared as a literal string to the possible missing value codes for that attribute using the private method issMissingValue on line 523

In generateInsertSQL() line 267 if a value is a missing value then it jumps to the next attribute in the list. The insert statement then does not upload that column. This is equivalent to inserting a null into that column. Since the DML does not seem to put any NON NULL constraints on any columns, (let alone any constraints at all), that in itself does not generate an error. It does, however, mean that missing value codes are not stored in the database table and so if there are multiple codes for a column that information is lost.

This was looking at svn revision number 2195 of Duane's branch

[https://code.ecoinformatics.org/code/eml/branches/DATAMANAGER\\_QUALITY/src/org/ecoinformatics/datamanager/database/Datab](https://code.ecoinformatics.org/code/eml/branches/DATAMANAGER_QUALITY/src/org/ecoinformatics/datamanager/database/Datab)

[aseAdapter.java](#)

on 23June2011

---

## History

### #1 - 06/23/2011 03:43 PM - gastil gastil

I said something wrong: "Since the DML does not seem to put any NON NULL constraints on any columns, (let alone any constraints at all),..." That is not true. The DML does model constraints. I was referring to a particular application which uses the DML, but does not use the constraint parts of DML.

### #2 - 03/27/2013 02:30 PM - Redmine Admin

Original Bugzilla ID was 5427