

EML - Bug #5308

Data Manager Library: storageType content should be stored and used

02/15/2011 07:57 AM - Duane Costa

Status:	In Progress	Start date:	02/15/2011
Priority:	Normal	Due date:	
Assignee:	Duane Costa	% Done:	0%
Category:	datamanager	Estimated time:	0.00 hour
Target version:	DataManager 1.1.0	Spent time:	0.00 hour
Bugzilla-Id:	5308		

Description

'storageType' is an optional, repeatable element within the EML 'attribute' element. In addition to the documentation available in the EML normative documents, several old bug tickets describe the rationale behind this element: [#484](#), [#544](#), [#599](#).

When the Data Manager Library parses EML attributes, it does not record any 'storageType' content that may be present. This means that the hints that may have been provided by the metadata provider pertaining to how the attribute should be stored optimally (say, in a relational database table), are completely ignored by the Data Manager Library, which instead relies entirely on the 'measurementScale' content for this purpose.

To cite a specific example of how 'storageType' content can be helpful, the document knb-lter-gce.1.9 (<http://metacat.lternet.edu/knb/metacat/knb-lter-gce.1.9>) contains three attributes for year, month, and day, respectively. Each of the attributes has storageType set to 'integer' and measurementScale set to 'dateTime'. When loading the data table into a relational database, the Data Manager Library sets the corresponding database fields to type 'timestamp' (in Postgres), having no knowledge that the storage type "hint" was to set the fields to type integer ('int4' in Postgres). The result is that in the original data table entity, the fields appear like this:

2000 8 26

while in the relational database, they appear like this:

```
year          |          month          |          day
-----+-----+-----
2000-01-01 00:00:00 | 0001-08-01 00:00:00 BC | 0001-01-26 00:00:00 BC
```

It's clear that in this particular case, the Data Manager Library could have used the storageType hint to select a more appropriate data type for these attributes.

The goal of this task is to:

1. Enhance the EML parsing phase of the Data Manager Library, so that it parses and stores all storageType elements that are provided for an attribute.
2. Enhance the data loading phase of the Data Manager Library, so that it uses storageType content, if provided, to make a more informed decision about which data type to define for the attribute. This may involve the need for heuristics to determine which data type is most appropriate under a given set of circumstances, particularly in cases where more than one storageType element is provided for an attribute.

History

#1 - 03/27/2013 02:29 PM - Redmine Admin

Original Bugzilla ID was 5308