DataONE R Client - Task #848

Story # 928 (Closed): R Client needs to support multiple data formats

modify R client to use system and science metadata for table parsing

2010-09-14 18:56 - Matthew Jones

Status:	Closed	**Start date:**	2010-10-07
Priority:	High	**Due date:**	
Assignee:	Rob Nahf	**% Done:**	100%
Category:		**Estimated time:**	0.00 hour
Target version:			
Milestone:	None	**Story Points:**	

Description
Right now, the R client assumes the data is a CSV file with a single header row. This assumption will frequently be untrue. Need to modify the client to support the following: 1) Parse the system metadata to examine the object format 2) Parse the science metadata to examine the physical structure of the dataset -- for example, number of header rows, column names, etc. 3) Use info from 1 + 2 to choose an appropriate method and parameters for calling read.table

History

#1 - 2010-10-06 20:09 - Matthew Jones

- Target version deleted (CCI-0.6)

- Milestone set to CCI-0.6

- Parent task set to #928

#2 - 2010-10-07 12:35 - Dave Vieglais

- Start date set to 2010-10-07

- Tracker changed from Bug to Task

#3 - 2013-01-02 23:58 - Rob Nahf

- Milestone changed from CCI-0.6 to None

- Assignee changed from Matthew Jones to Rob Nahf

- Status changed from New to In Progress

built out the generic asDataFrame with a method in DataPackage that locates and uses the science metadata object to get table parsing instructions. Use the dataFormat in combination with DataTableDescriber abstract class to find the appropriate science metadata parser that will provide the parsing instructions.

Built an EML specific parser, based on EML 2.1.1, and with the assumption that the data table parsing elements can apply back to EML 2.0.

Need to split the EMLParser into it's own package

#4 - 2014-09-10 20:27 - Matthew Jones

- Status changed from In Progress to Closed

- translation missing: en.field_remaining_hours set to 0.0

Moved issue to github: https://github.com/DataONEorg/rdataone/issues/31