Infrastructure - Task #7042

Story # 7053 (Closed): DataONE should preserve media type metadata when transferring objects between systems

Create an element for the character set encoding in the system metadata schema

2015-04-14 20:48 - Jing Tao

Status:	Closed		**Start date:**	2015-04-14
Priority:	Normal		**Due date:**	
Assignee:	Dave Vieglais		**% Done:**	100%
Category:	d1_schemas		**Estimated time:**	0.00 hour
Target version:	CCI-2.0.0			
Milestone:	None		**Story Points:**	
Product Version:	*			

Description

We need to create an optional element for the character set encoding the system metadata schema.

Please see related the issue https://redmine.dataone.org/issues/6391

Related issues:

Related to Infrastructure - Bug #6391: Science metadata files with different ...	**Closed**	**2014-11-04**	**2015-01-06**

Associated revisions

Revision 15450 - 2015-04-17 22:12 - Dave Vieglais

Added start of design for handling media type metadata. refs #7042

History

#1 - 2015-04-14 20:49 - Jing Tao

- Related to Bug #6391: Science metadata files with different checksums on CN and MN - encoding added

#2 - 2015-04-15 15:53 - Dave Vieglais

- % Done changed from 0 to 30

- Category set to d1_schemas

- Status changed from New to In Progress

This is necessary because the transfer encoding information for content is not being preserved, and is necessary to ensure reliable re-transmission of content. For example, an xml document that is not encoded in the assumed default character encoding will be resent using either no character encoding information or the default, which may be incorrect and lead to unexpected results.

#3 - 2015-04-17 22:14 - Dave Vieglais

This is a complex topic that requires broader changes to full support media type when transferring across multiple systems.

Discussion in: https://repository.dataone.org/documents/Projects/cicore/architecture/api-documentation/source/design/what_is_it.txt

#4 - 2015-04-17 22:19 - Dave Vieglais

- Parent task set to #7053

#5 - 2015-05-29 21:19 - Ben Leinfelder

I believe this is now covered by MediaTypeProperty element[s]?

#6 - 2015-07-08 18:36 - Chris Jones

while this is covered by the new MediaTypeProperty element which is part of the MediaType element used in an ObjectFormat, I think it may quite applicable. For instance, in the case of the "text/csv" media type, it has an optional "charset" parameter. When not used, it defaults to "UTF-8". However, if extend our text/csv object format to have a UTF-8 element, then we are defining ALL text/csv to be UTF-8 encoded, which certainly isn't the case. So, it seems that we DO need another element in the SystemMetadata type to cover this, no?

#7 - 2015-07-08 21:32 - Ben Leinfelder

- translation missing: en.field_remaining_hours set to 0.0

- Status changed from In Progress to Closed

- % Done changed from 30 to 100

v2.SystemMetadata.MediaType.MediaTypeProperty can be used to define the character encoding for individual objects without affecting the ObjectFormat.MediaType that may be used by other objects.