

OGC-Slender Node - Story #7170

Evaluate the feasibility of extracting provenance information from the journal.txt document

2015-06-08 19:03 - Dave Vieglais

Status:	New	Start date:	
Priority:	Normal	Due date:	
Assignee:	Dave Vieglais	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:			
Story Points:			
Description			

History

#1 - 2015-06-08 19:05 - Dave Vieglais

- Assignee set to Dave Vieglais

Each entry on the FTP site contains a "journal.txt" document that basically describes the provenance of the package.

This is unfortunately in a very much human-ingestible form.

The goal of this activity is to determine what, if any, useful provenance information can be divined from the journal and associated information (e.g. manifest files).

e.g. journal file:

<ftp://ftp.nodc.noaa.gov/nodc/archive/arc0064/0118783/14.14/about/journal.txt>

#2 - 2015-06-08 21:28 - Dave Vieglais

Likely to be a laborious free text processing exercise. From John Relph:

No, the journal.txt is not available in any machine-readable format. We have talked about implementing a system to enable that, but currently the file is a mostly free-form file for the use of the human data content manager to record actions performed and other information.

We will, at some point soon, start populating lineage sections in our ISO metadata with information about the various revisions of the accessions.