eArchiving Conformance Report for the Generic Services Project OneClick

Document Control Information

Settings	Value
Document Title:	eArchiving Conformance Report for the Generic Services Project OneClick
Project Title:	eArchiving
Evaluators:	Karin Bredenberg, Carl Wilson.
Project Owner (PO):	Árpád Welker, DG CNECT
Doc. Version:	1.0
Sensitivity:	Limited
Date:	17/07/2023

Document Approver(s) and Reviewer(s):

NOTE: All Approvers are required. Records of each approver must be maintained. All Reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Action	Date
Árpád Welker	Project Owner	<Approve / Review>	

Statement of originality
eArchiving is funded by the European Union's Digital Europe Programme. It is operated by the E-ARK Consortium led by the Austrian Institute of Technology (AIT) under a service contract with the European Commission, contract number LC-01905904.

Conformance Result: OneClick is 100% E-ARK/eArchiving conformant.

Conformance Submission details:

A selection of 3 Information Packages (IPs) per type (SIP, AIP, DIP) were requested by the E-ARK/eArchiving Consortium, i.e. a total of 9 information packages (container packages in *.zip, *.tar, or *.tar.gz format) plus system log files (*.log or *.txt format) related to the creation of information packages.

9 IPs were supplied by OneClick together with system log files:

- 3 SIPs;
- 3 AIPs;
- 3 DIPs;

All 9 IPs validated correctly so OneClick is 100% E-ARK/eArchiving conformant.

Observations

- There are no non-conformances issues.
- The examples are clear.
- All of the supplied information packages were valid.
- There were some false validation negatives from the E-ARK validator due to a deprecated rule issue with the validator.
- The metadata files are valid and used in a minimal form.
- There were two different archival storage programs used, one being Archivematica and another one being Disec, and they are handling the IPs in two different ways.
- The examples provided rely heavily on an interface to show the different IPs in their original location. As they are they will not be able to be understood without a lot of unzipping and looking at the data files that are found in most cases in zipped files.
- The DC.xml files are not connected to a Dublin Core schema which means that validation is not possible.
- None of the descriptive metadata given in the DC.xml files provides information about the creator of the files or their reason to be in an IP.
- There are very few real log files, about half of the supplied files were output dumps from other programs.
- The Disec workflow performs a transformation of the package at each stage, SIP->AIP and AIP->DIP. Each time the previous IP is zipped, and a new IP is created, at the same time the metadata from the previous stage is simply copied to the new package, time stamps and all. When the DIP is delivered the user has to unpack the zip files to get to the original SIP. There is no package metadata that records these transformations, or why they were performed.
- The Disec package METS files for the AIP and DIP record the package XML metadata schema file MIME types as unknown “application/octet-stream” despite them having been identified as XML in the submitted SIP. Losing correct package metadata does subsequent users little service.
- The Disec AIP and DIP packages put the previous into a folder called submission when the E-ARK standards strongly recommend using the Representations folder.
- Too much of the data in the SIP METS files is created directly from a Python Jinja template which means details are hard coded, including the initials of the person submitting the package and their ID.
- The PIQL AIPs and DIPs are better than those produced by Disec.

Improvement Opportunities

- The OAIS-typing of the packages could use a review.
- The Dublin Core files could use a review and be updated to be following the standard.
- The use of PREMIS could be enhanced in the Disec solution so as to not just use it for describing an agent.