

14 – AIP Update: Dateien in bereits langzeitarchivierten Objekten austauschen

Was ist ein AIP?

Laut dem [OAIS Modell](#) der Langzeitarchivierung handelt es sich beim AIP um das „Archival Information Package“. Dies bezeichnet das zu erhaltende Objekt, wenn es mit seinen Metadaten bereits im permanenten Speicher des Langzeitarchivs angekommen ist.

Idealerweise werden Änderungen am AIP nur innerhalb des Langzeitarchivs vorgenommen, um die Lesbarkeit zu erhalten, beispielsweise durch Formatmigration. Das entspricht aber nicht der Arbeitsrealität von vielen von uns. Oftmals wird seitens der Datenproduzent*innen nachträglich noch eine Änderung oder Aktualisierung gefordert – oder es werden gar die Veröffentlichungsrechte zurückgezogen.

Fälle des AIP Update aus dem Arbeitsalltag

Im Langzeitarchiv der ZBW, das auf Rosetta der Firma Ex Libris basiert, sind AIP Updates eher die Regel als die Ausnahme. Vor allem von den Datenproduzent*innen für das Publikationsarchiv [EconStor](#) kommen die Änderungswünsche. In der Regel werden die Objekte bereits in der Folge nach der Veröffentlichung auf dem Publikationsserver EconStor im digitalen Langzeitarchiv gespeichert. Wenn es zu nachträglichen Änderungswünschen seitens der Datenproduzent*innen kommt, wird daher stets ein AIP Update in Rosetta notwendig.

Folgende vier Uses Cases sind zu unterscheiden:

- Datenproduzent*in zieht Veröffentlichungsrechte zurück
- Datenproduzent*in reicht zusätzliche Datei zum Datensatz ein
- Datenproduzent*in tauscht Datei aus
- Der Identifier ändert sich (in unserem Fall die PICA Produktionsnummer)

Zwischen August 2018 und November 2020 wurde in fast 400 Fällen das Veröffentlichungsrecht zurückgezogen und fast 300 Mal hat ein*e Datenproduzent*in die Datei ausgetauscht. Der Workflow zum Ändern der Access Rights wurde von uns daher schon vor Jahren automatisiert. Nun wurde es notwendig, auch den Workflow zum Austausch der Datei zu automatisieren.

Datei in bereits archivierten Objekten ersetzen

Mithilfe der Rosetta Webservices hat der Kollege der Gruppe Langzeitarchivierung, der für die Entwicklung von Schnittstellen zu Rosetta zuständig ist, ein Tool entwickelt, das nun den automatisierten Dateiaustausch in Rosetta ermöglicht. Wichtig ist vor allem, dass dies auch „in bulk“ mithilfe einer beliebig langen CSV-Datei geschehen kann, solange die CSV-Datei entsprechend normalisiert vom Team Langzeitarchivierung aufbereitet wird.

Folgende Informationen werden hierfür benötigt:

Feld Name	Beschreibung
Handle	Dient als Identifier
Alter Dateiname	
Neuer Dateiname	
Status	Wird vom Programm ausgefüllt, so steht „tested“ für erfolgreichen Test und „done“ für abgeschlossenen Austausch
Bemerkungen LZA	
SRU key	
Update Error	
IE POD	Rosetta identifier, der vom Programm ermittelt wird

Screenshot der CSV-Datei:

	A	B	C	D	E	F	G	H
1	Handle	fileNameNew	fileNameOld	state	notesDP	SRUKey	UpdateError	IE-PID
2	10419/165913	ifosd-v70-2017	ifosd-v70-2017	Tested		IE.dc.identifier.dcterms.URI		IE3478860
3	10419/163878	ifosd-v56-2003	ifosd-v56-2003	Tested		IE.dc.identifier.dcterms.URI		IE3472746
4	10419/222851	1702045005.pc	1702045005.pc	Tested		IE.dc.identifier.dcterms.URI		IE9725849
5	10419/165911	ifosd-v70-2017	ifosd-v70-2017	Tested		IE.dc.identifier.dcterms.URI		IE3478854
6	10419/165935	ifosd-v70-2017	ifosd-v70-2017	Tested		IE.dc.identifier.dcterms.URI		IE3478926
7	10419/175135	ifosd-v70-2017	ifosd-v70-2017	Tested		IE.dc.identifier.dcterms.URI		IE3479082
8	10419/181078	ifosd-v71-2018	ifosd-v71-2018	Tested		IE.dc.identifier.dcterms.URI		IE3524283
9	10419/181088	ifosd-v71-2018	ifosd-v71-2018	Tested		IE.dc.identifier.dcterms.URI		IE3524313

Diese Informationen können bereits von den Kolleg*innen der Plattform EconStor an das Team Langzeitarchivierung gegeben werden, so dass mögliche Nachbearbeitungen der CSV-Liste in der Regel nicht notwendig werden.

Screenshot des Tools:

LZA Tools

- Dashboard
- Repositories
- Exports
- Update**
 - Overview
 - Update single File
 - Update Files
 - Update Logs
 - Update Metadata

type to search

Update Files

Upload a CSV File for Updating files in Rosetta

CSV File
 Keine Datei ausgewählt.
 Required

Repository

Global Reason

Run test
☐

„Run test“ bedeutet, dass zunächst vom Programm geprüft wird, ob alle Objekte im permanenten Speicher des Archivs gefunden werden können und der alte Dateiname im Objekt vorhanden ist. Das Ersetzen der Dateien funktioniert folgendermaßen: Das Programm findet das Objekt in der Repräsentationsplattform (hier: EconStor) anhand des Handles via der OAI-Schnittstelle. Es lädt die neue Datei herunter.

Mittels des Handles und des alten Dateinamens findet das Programm das AIP in Rosetta. Es erlangt den Rosetta-internen Identifier, den IE PID, und trägt diesen in die CSV ein. Mittels der IE PID und des alten Dateinamens findet es die Identifier innerhalb unseres Rosetta-Langzeitarchivs für die Datei (File PID) und für die Repräsentation (Representation PID) via SRU. Nun sind alle drei Parameter verfügbar:

- File PID
- Representation PID
- IE PID

Rosetta stellt diverse Schnittstellen zur Verfügung, um von außen mit dem Rosetta-Archiv zu kommunizieren und Änderungen vorzunehmen. Um Objekte in Rosetta zu suchen gibt es eine [SRU](#) Schnittstelle, die Informationen über alle im [Solr](#) Index befindlichen Objekte bereitstellt. Da sich nur ein kleiner Teil der Informationen im Index befindet, ist vor allem die gefundene IE PID interessant, die den Einstieg in tiefergehende Suchen ermöglicht. Dafür dienen dann verschiedene [SOAP](#) Web-Services, die Rosetta zur Verfügung stellt. Diese stellen Funktionalitäten bereit, die einerseits detailliertere Informationen zu AIPs hervorbringen und andererseits aber auch Änderungen im Langzeitarchiv ermöglichen. Dabei werden natürlich die vorgenommenen Änderungen dokumentiert. Mittels der über die Web Services erhaltenen IDs von IE, Representation und File kann die Datei dann über einen weiteren Web Service ausgetauscht werden.

Dieser Workflow wird nun in unseren Arbeitsalltag integriert, so dass das mühsame, händische Ersetzen von Dateien im Permanent unseres Langzeitarchivs von nun an der Vergangenheit angehört.

Yvonne Tunnat

Projektmanagement Langzeitarchivierung

ZBW –Leibniz-Informationszentrum Wirtschaft

Kontakt: y.tunnat@zbw.eu , Webseite: www.zbw.eu



Weitere Kurzartikel aus der Reihe „nestor Thema“ finden Sie auf www.langzeitarchivierung.de - der Webseite von [nestor – Kompetenznetzwerk Langzeitarchivierung](#).