

# Connecting EPrints Repository Systems with the Archivematica Open Source Digital Preservation Workflow



17th International Open Repositories Conference, June 6th-9th in Denver, Colorado, USA

Tomasz Neugebauer, Concordia University  
[tomasz.neugebauer@concordia.ca](mailto:tomasz.neugebauer@concordia.ca)

Sarah Lake, Concordia University  
[sarah.lake@concordia.ca](mailto:sarah.lake@concordia.ca)

# Background



Spectrum Research Repository

archivematica®



# Digital Preservation through EPrints- Archivematica Integration

—————→

Tomasz Neugebauer	Justin Bradley	Justin Simpson
Concordia University	EPrints Services	Artefactual

**International Conference on Open Repositories 2018**  
Bozeman, Montana State University  
Tuesday, June 5<sup>th</sup>, 2018  
24x7 Presentation

<b>Digital preservation feature</b>	<b>EPrints</b>	<b>Archivematica</b>
OAIS-based	Yes (sort of)	Yes
Assign UUIDs	No	Yes
Create checksums	Yes (sort of)	Yes
Verify checksums	No	Yes
Record standards-based preservation metadata (e.g. PREMIS)	No	Yes
Identify file formats	Yes	Yes
Validate file formats	No	Yes
Perform normalization	Only for access	Yes

# Plugin Development

EPrintsArchivematica Plugin. GitHub.

<https://github.com/eprintsug/EPrintsArchivematica>

- Artefactual
- EPrints Services
- EPrints Community
- Concordia University Library
- Open Repositories Conferences (2018, 2022)
- Archivematica Camps (2018, 2021)

# Features - Beta phase

<1.2

Development beta phase

2018 - 2022

- modelling of archivematica export folder structure and transfer objects as a dataset in EPrints
- process\_transfers, create\_transfers, touch\_transfers BIN scripts
- basic configuration options (export location, trigger fields)

# Features - First Release

1.2.2

**First official release**

**February 2022**

- option for metadata only records
- logging of archivemata transfer objects (checksum mismatch)
- metadata.json and checksum file interoperable with Archivemata
- option for generating a checksum if one is missing in EPrints
- callback from Archivemata to set the UUID
- UTF-8 handling of file names

# Features - Second Release

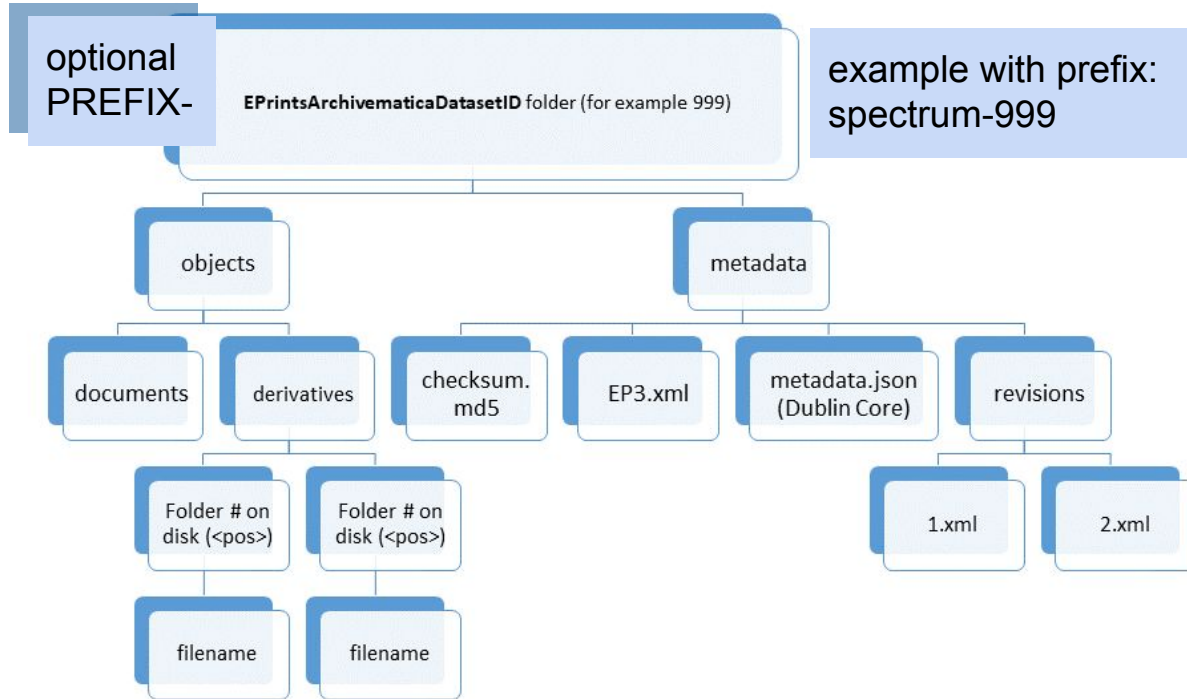
1.2.3

**Second release**

**May 2022**

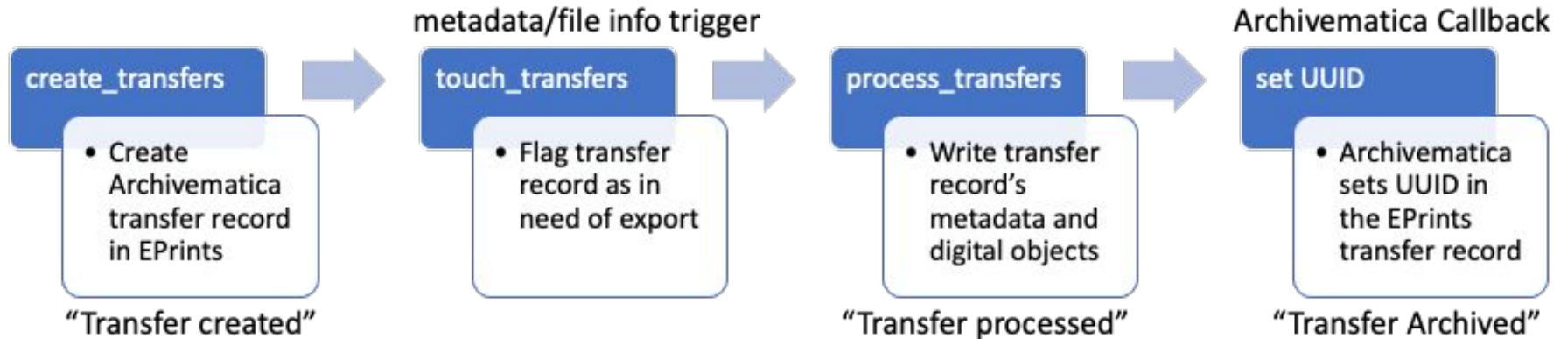
- improved touch\_transfers script
  - unset flag to bypass transfer
  - delete flag to delete transfer
- filename handling (apostrophe, colon, double-quote...)
- more detailed logging
- bug fixes in how transfer objects are indexed by EPrints
- custom prefix for top-level folder name of transfer folder

# Transfer Record Folder Structure





# Created, Processed, Archived



# Metadata & Checksums

Archival storage / Search

Transfer metadata

Keyword

Show files?

Search archival storage

Add new

## Browse archival storage

Total size: 144.4 GB | Files indexed: 425531

Create an AIC

Name	UUID	Size	Created
2830	b0def7f1-77dc-4a21-bc80-321825142aa7	10.9 MB	2022-02-02 15:46

Showing 1 to 1 of 1 entries

Select columns

# Archivemata configuration

- Automation Tools checks watched folder for new transfers and Archivemata processes with automated processing config

## artefactual/**automation- tools**

Tools to aid automation of Archivemata and AtoM.



 17  
Contributors

 15  
Issues

 37  
Stars

 30  
Forks



# AIP in Archivemata

Standard



Transfer type

Transfer name

Accession no.

Access system ID

Browse

Start transfer



Approve automatically

Transfer

UUID

Transfer start time



6880	84106466-7e1e-4080-9003-e51c15554a9e	2022-05-05 12:05	
▸ Microservice: Create SIP from Transfer			
▸ Microservice: Complete transfer			
▸ Microservice: Generate METS.xml document			
▸ Microservice: Examine contents			
▸ Microservice: Validation			
▸ Microservice: Parse external files			
▸ Microservice: Characterize and extract metadata			
▸ Microservice: Extract packages			
▸ Microservice: Identify file format			
▸ Microservice: Change transfer filenames			

# AIP in Archivemata

Archival storage / Search

Any ▾

Keyword ▾

Show files?

Search archival storage

Add new

## Browse archival storage

Total size: 144.4 GB | Files indexed: 425368

Create an AIC

Name	UUID	Size	Created	Status
15017	e80a7ad5-66c4-4794-ac57-822812a4878c	1.1 MB	2022-04-15 20:30	Stored
15004	1634b352-c39d-4f9a-bffd-d1890c578f99	1.3 MB	2022-04-15 19:51	Stored
15002	3bae69c3-c2b8-4143-b425-b01e173bbfef	13.8 MB	2022-04-15 19:46	Stored

# AIP in Archivematica

- A callback in Archivematica updates the item record in EPrints once an eprint AIP has been created and stored

Archivematica ID	Dataset	Object ID	Archivematica UUID	New Transfer Pending	Action	Timestamp	Result	Comment	Actions
2779	eprint	986367	7367d957-4840-44d3-9f58-61d4ef081d54	No	Transfer Created, Transfer Processed, Transfer Archived	23 January 2022 01:00:33 UTC, 2 February 2022 18:06:48 UTC, 2 February 2022 18:11:14 UTC	Success, Success, Success	created via trigger, processed, UUID: [7367d957-4840-44d3-9f58-61d4ef081d54]	 

# Future enhancements

- Configuring checksum mismatch error handling
- Translating processing logs into PREMIS metadata

```
<premis:event>
  <premis:eventIdentifier>
    <premis:eventIdentifierType>EPrintsArchivematica Event ID</premis:eventIdentifierType>
    <premis:eventIdentifierValue>EA-2</premis:eventIdentifierValue>
  </premis:eventIdentifier>
  <premis:eventType>fixity check</premis:eventType>
  <premis:eventDateTime>2019-07-04T22:46:07.773391+00:00</premis:eventDateTime>
  <premis:eventDetailInformation>
    <premis:eventDetail>checking that EPrints checksum matches</premis:eventDetail>
  </premis:eventDetailInformation>
  <premis:eventOutcomeInformation>
    <premis:eventOutcome>Pass</premis:eventOutcome>
    <premis:eventOutcomeDetail>
      <premis:eventOutcomeDetailNote>Missing checksum in EPrints, generated new checksum and stored
    </premis:eventOutcomeDetailNote>
    </premis:eventOutcomeDetail>
  </premis:eventOutcomeInformation>
  <premis:linkingAgentIdentifier>
    <premis:linkingAgentIdentifierType>local</premis:linkingAgentIdentifierType>
    <premis:linkingAgentIdentifierValue>EPrintsArchivematica</premis:linkingAgentIdentifierValue>
  </premis:linkingAgentIdentifier>
</premis:event>
```

# Future enhancements

## OCFL/spec

The Oxford Common File Layout (OCFL)  
specifications



12

Contributors

24

Issues

38

Stars

14

Forks





# Thank you!

Tomasz Neugebauer, Concordia University  
[tomasz.neugebauer@concordia.ca](mailto:tomasz.neugebauer@concordia.ca)

Sarah Lake, Concordia University  
[sarah.lake@concordia.ca](mailto:sarah.lake@concordia.ca)