

Research Data Management & Preservation using PURE and Archivematica

Alan Morrison
Research Data Support Officer
University of Strathclyde

Research Data Deposit Policy

When [Research data](#) is deposited and assessed, [Research Data Support Staff](#) will potentially create and store three different versions of the [dataset](#): [original](#), [distribution](#) and [preservation](#) copies.

Very diverse types of electronic data are produced at University of Strathclyde and it is not possible for comprehensive digital curation activities to be performed on all data types nor is it sensible to restrict the types of data that can be produced and/or stored. Therefore while all data types will be able to be deposited in Pure, a selective policy with regard to preferred file formats and their corresponding level of support will operate.

- Open Policy to accommodate a STEM university with an unknown quantity of data & potential file formats
- OAIS as a model
- Sliding scale of prioritisation & support depending on data format
 - Preferred, Acceptable, Problematic

Preservation & Curation systems

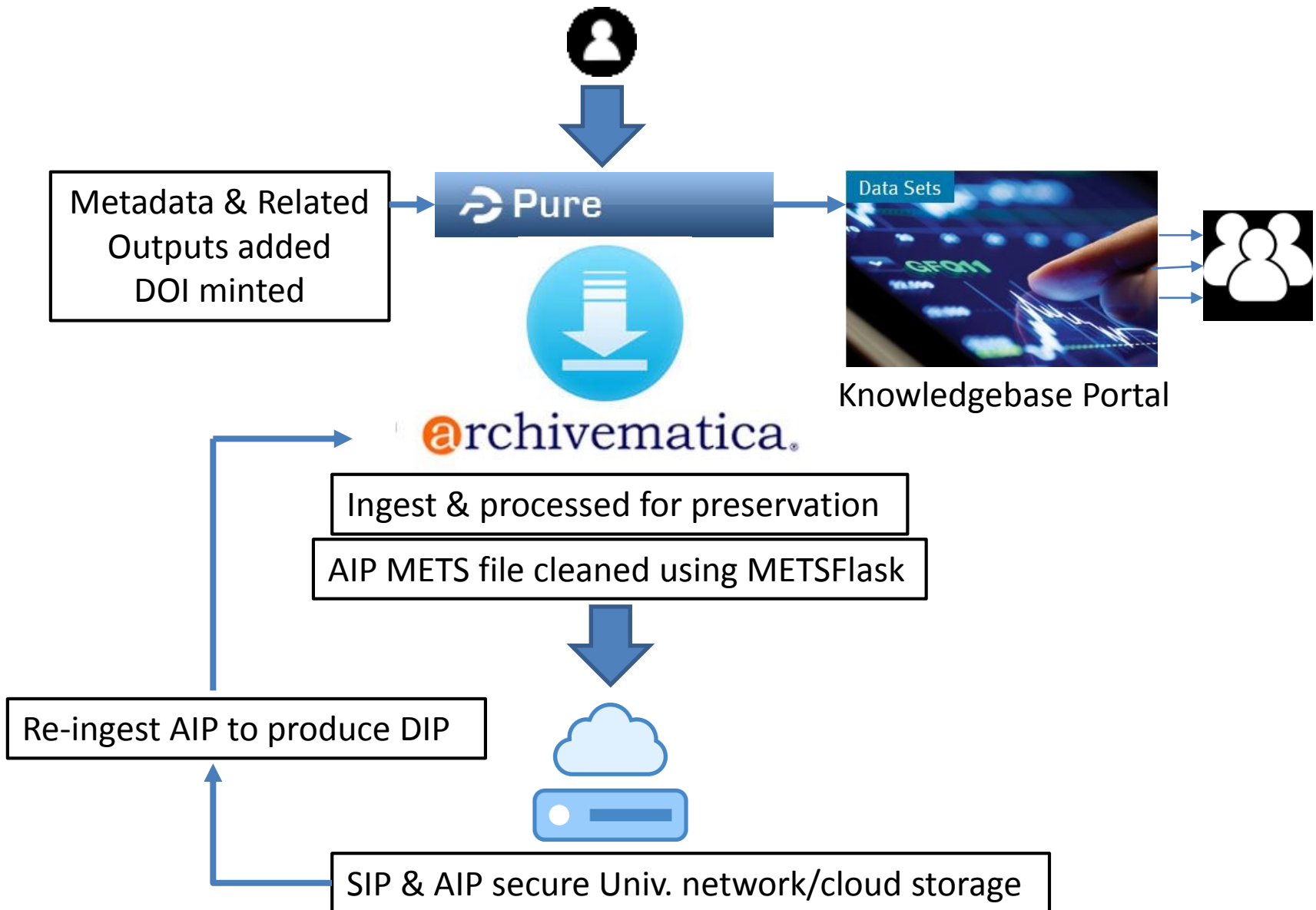
Pure:

- Existing proprietary CRIS @ Strathclyde
- Strathprints ePrints for research outputs
- Data storage, metadata, DOI & related content
- UKPure-Datasets Working Group
- Datasets no longer a road mapped priority?

Archivematica:

- Free web & standards based open-source
- Customisable Micro-services
- archivematica-uk-group
- One of the preservation platforms for Jisc Research Data Shared Service

Curation workflow



Benefits

Pure

- Familiar to UoS researchers
- Metadata management & related content
- Customised front-end (Knowledgebase)

Archivematica

- Metadata harvesting & format identification
- Bulk processing of large datasets
- Open source/community development

Limitations

General

- No interoperability between the systems
- Duplication of data
- Too many manual processes

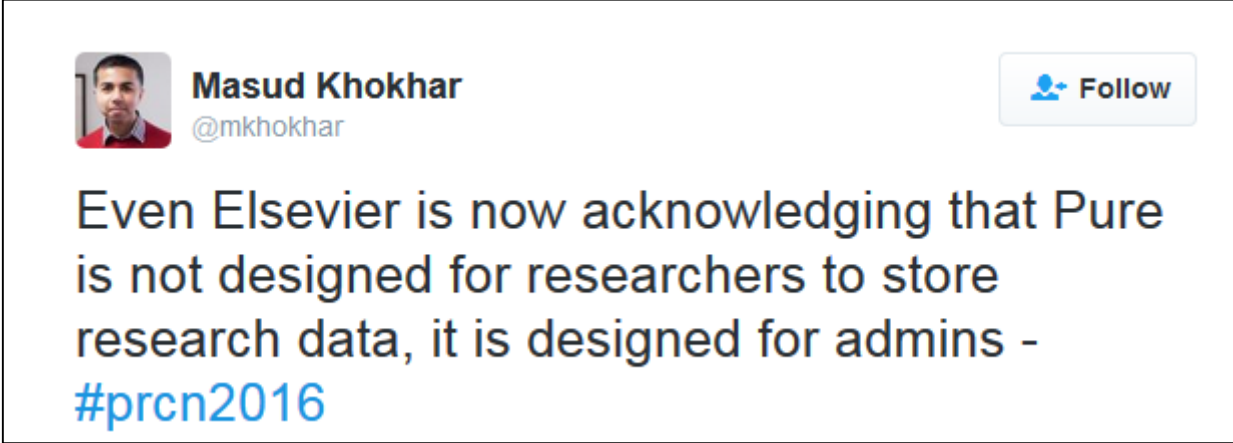
Pure

- Designed for reporting & storage not curation
- Functionality reliant on vendor priorities

Archivematica

- Only ca.40% of file formats recognized
- Improved file format registry with PRONOM needed
- Messy output
- Requires some technical support

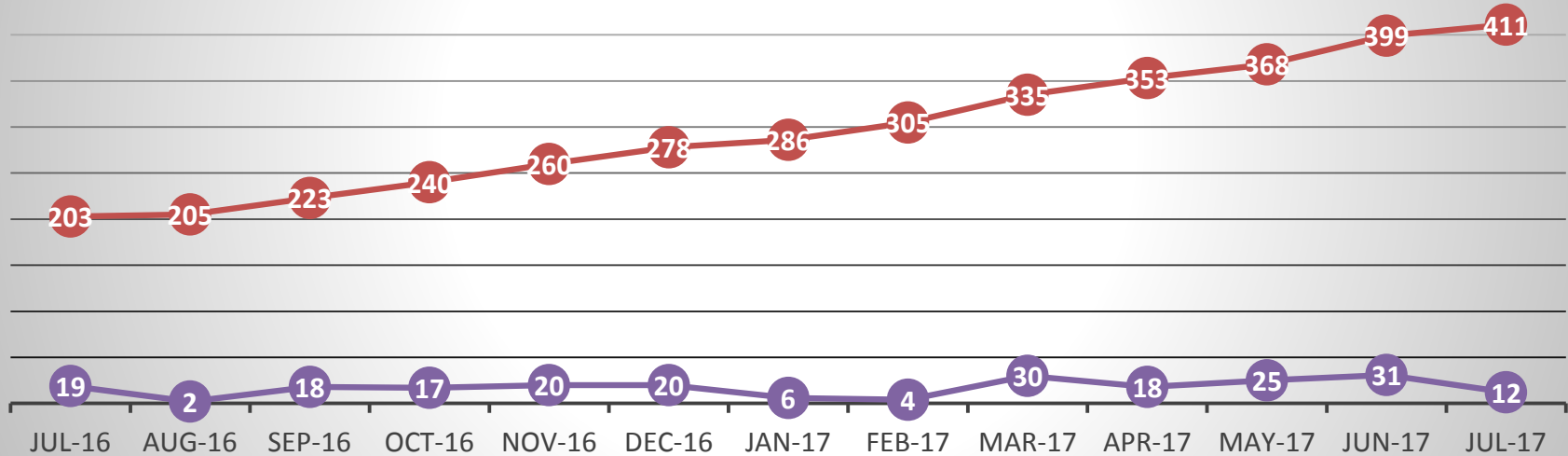
Future plans

- 

Even Elsevier is now acknowledging that Pure is not designed for researchers to store research data, it is designed for admins -
[#prcn2016](#)
- Assess processes using Data Seal of Approval criteria
- Improve pre-ingest routines & data management through FAIR advocacy & training

Deposits to date

Data Deposits Jul16 - Jul17





University of
Strathclyde
Glasgow