System Development: Data Enrichment
Software Records with Citation Capture Pipeline

Taylor Jacovich and the ADS Team

ADS Users Group Meeting, 9-10 Nov. 2022
Software Records with Citation Capture Pipeline

Overview

- CitationCapture Pipeline handles ingest for certain software records
  - Registers Zenodo records with ADS
    - Two types of records
      - **Version**: point to a specific version of the software
        - `property:release`
      - **Software Source**: refer to the software as a whole
    - Forwards Zenodo records and Github urls to the Asclepias project
  - **Asclepias**: Collaboration between AAS, Zenodo, and ADS to create a tool for scientific software discovery
  - Only ingests from reference lists of other ADS records
    - We do not ingest every Zenodo software record
    - Currently have ~12700 records
- Curators can non-destructively modify any metadata
  - Curation persists across metadata updates
Citation Capture Records
Associated Works

• All software records are automatically matched with any existing versions in ADS
  • The version name will reflect whatever versioning scheme the maintainers use
Citation Capture Records

Metrics

- Aggregate metrics can now include CitationCapture records
- Metrics can now be generated for individual CitationCapture records
Citation Capture Records
Potential Extensions

• Search function to aggregate associated works

• Better handling of Software Source records.