folio

future of libraries is open

A Look at the Harvester Admin App

WOLFcon 2024

Felix Hemme, Antje Niemann

September 24, 2024

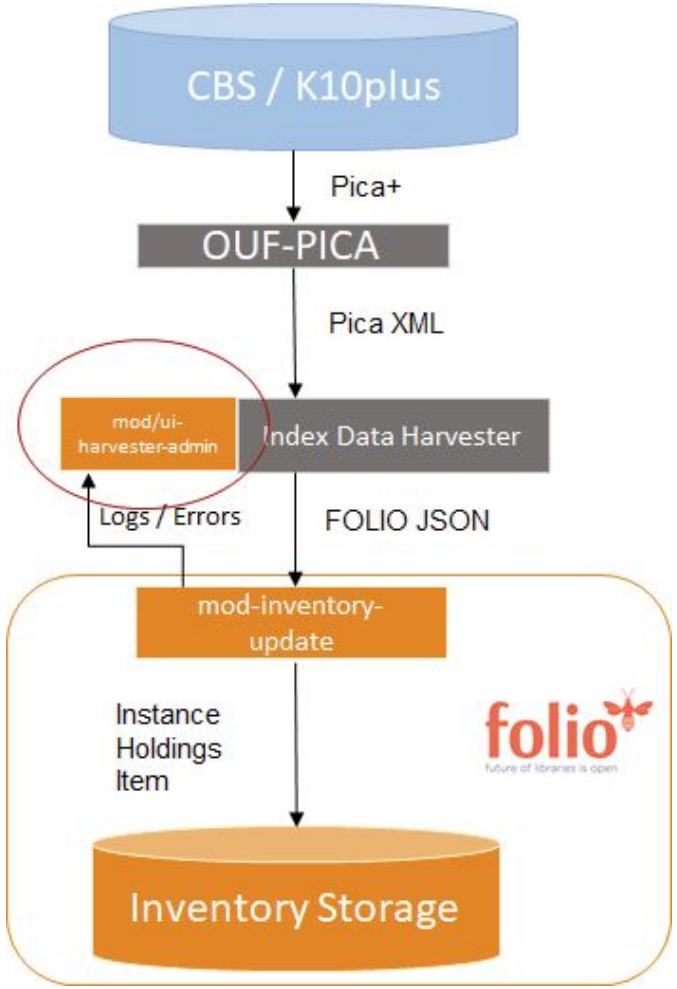
GBV Cataloging Practice in a Nutshell

- OCLC CBS (K10plus) as shared cataloging database for GBV and SWB network
  - ~ 500 German libraries, ~ 80 million title records, ~ 230 million ownerships, ~ 15 million authority records
- Links between title and authority records and between title records
- 80 % shared cataloging (only adding holdings at existing bibs)
- CBS cataloging outside of FOLIO
- Record format PICA (not MARC 21)

Data Management in Heterogeneous Systems

- Authoritative system
  - CBS, with some exceptions in FOLIO holdings and items
  - CBS also for Interlibrary Loan and as metadata source for Discovery systems
- Data to be transmitted to FOLIO
  - Selected bib data
  - With holdings/items
  - Without authorities
- Transition from CBS to FOLIO
  - One-time initial load of data as batch process
  - Real-time update for live operations

Metadata Pipeline and Software Components



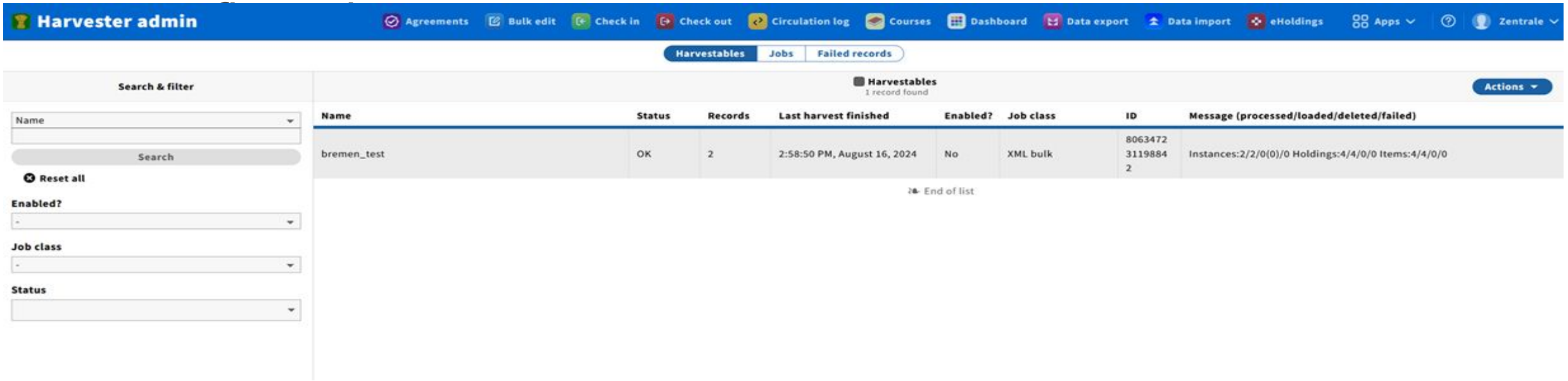
- Index Data Harvester
  - Reads data from CBS/OUF-PICA and transforms to FOLIO JSON
- mod-harvester-admin
  - Okapi service that can be put in front of the Index Data Harvester
  - Provides FOLIO based access to control the Index Data Harvester
- ui-harvester-admin
  - Provides an FOLIO/JSON based interface to the configuration database that FOLIO clients (like a Stripes UI) can use
- mod-inventory-update
  - Okapi service that can be put in front of mod-inventory-storage
  - Populates inventory storage with instances, holdings, and items

Harvester admin App

- Developed by Index Data on behalf of VZG
- Initial goal: Storage of job runs and failed records directly in FOLIO (PostgreSQL) including access to this data for system librarians
- Now also provides access to the configuration of the process if user has appropriate permissions
- Not yet part of `folio-complete`, but simple installation via Docker Container

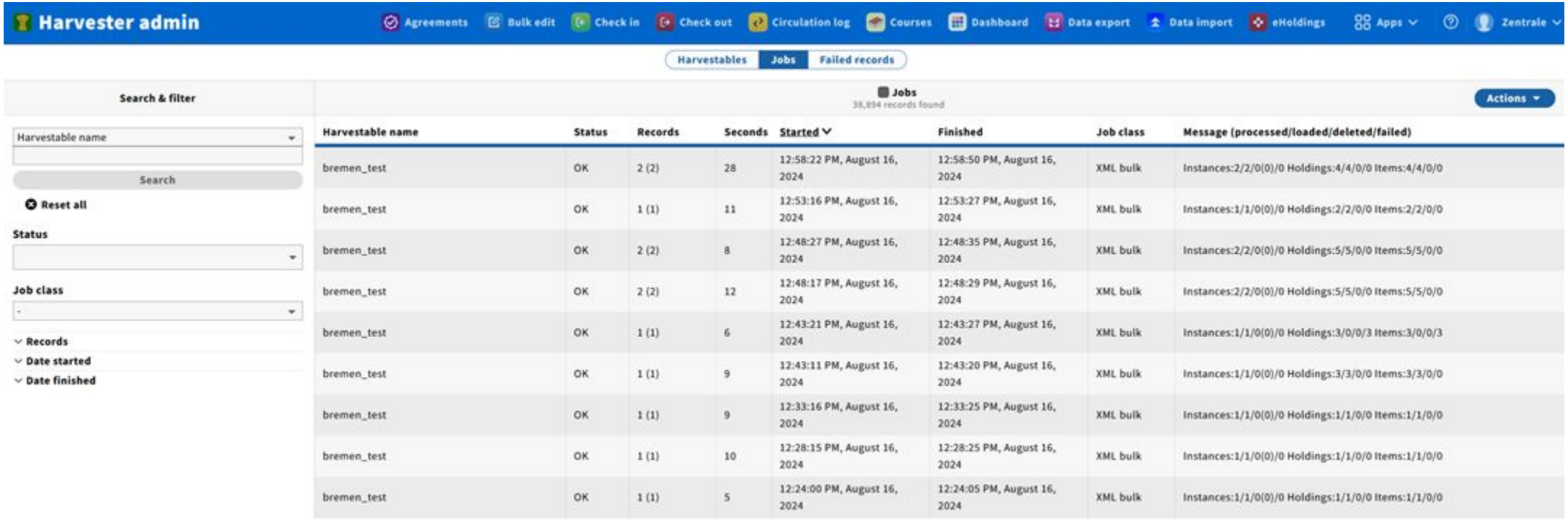
Harvester admin App > Harvestables

- A harvestable is a process responsible for the transition of the metadata (instances, holdings, items) to FOLIO Inventory
- One harvestable per tenant (at GBV), and potentially one harvestable per data provider
- Access to the configuration by clicking on the process



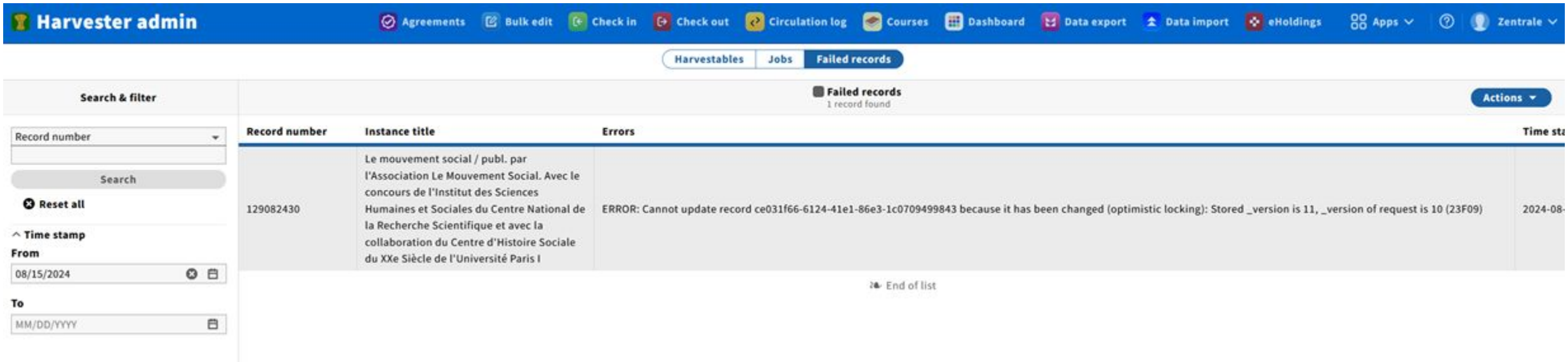
Harvester admin App > Jobs

- One entry per job run of the harvestable, sorted by start date in descending order



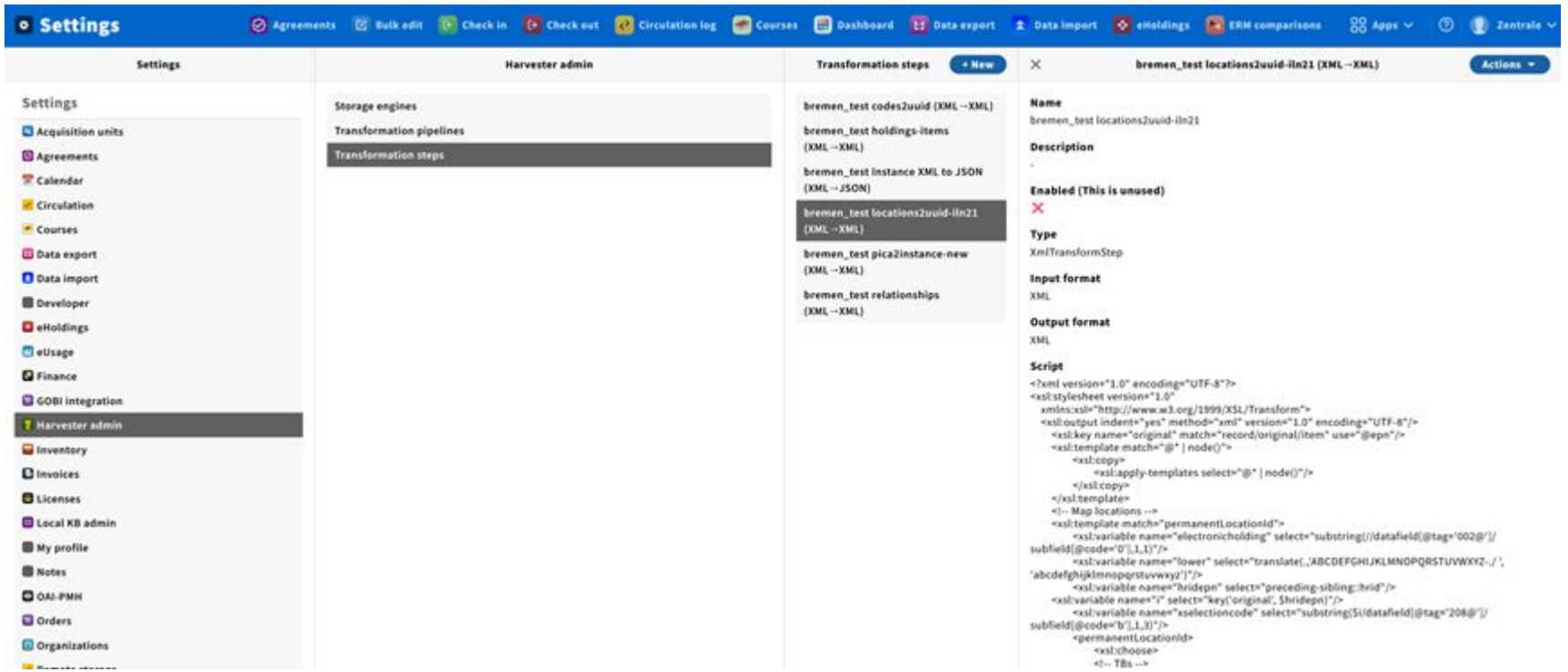
Harvester admin App > Failed records

- One entry per instance record, pre-filtered by date >= yesterday



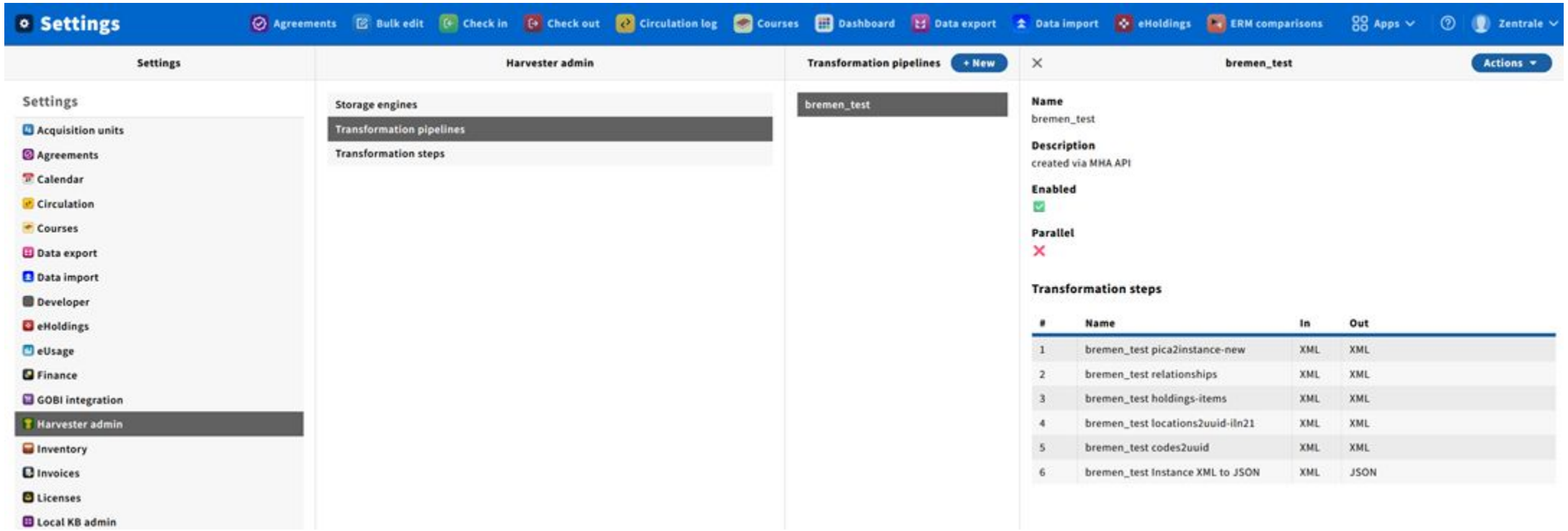
Settings > Harvester admin

- Transformation steps: Store the XSLT stylesheets



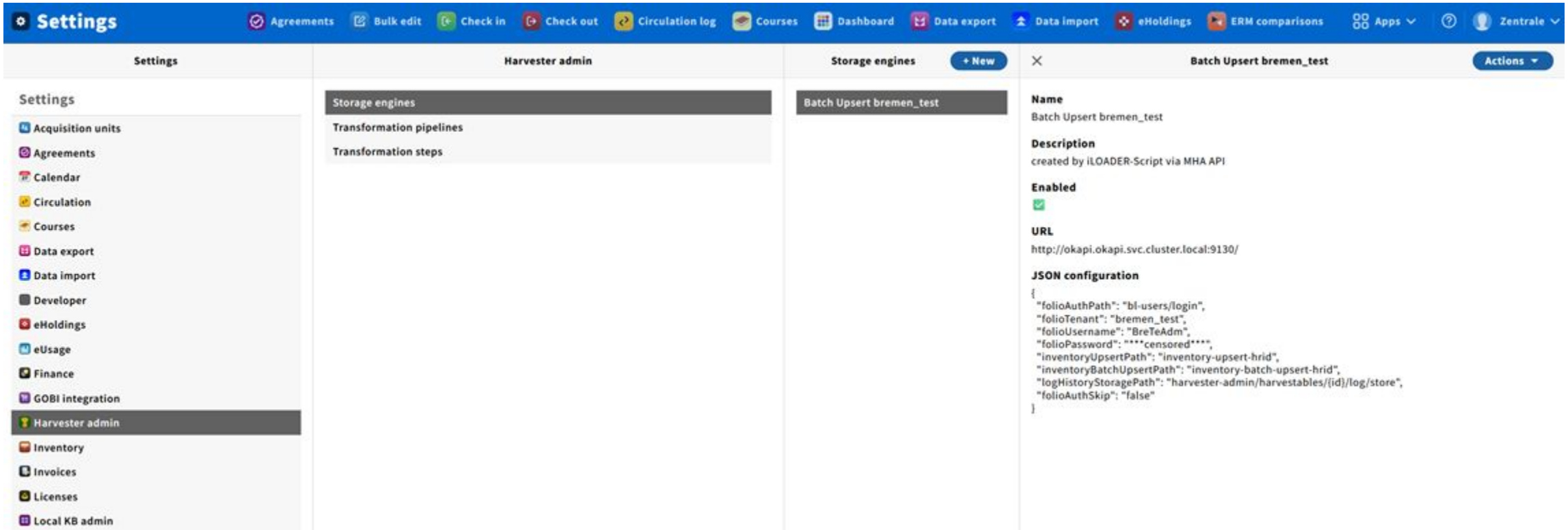
Settings > Harvester admin

- Transformation pipelines: Connect the transformation steps in sequence



Settings > Harvester admin

- Storage engines: Contains configurations for FOLIO access

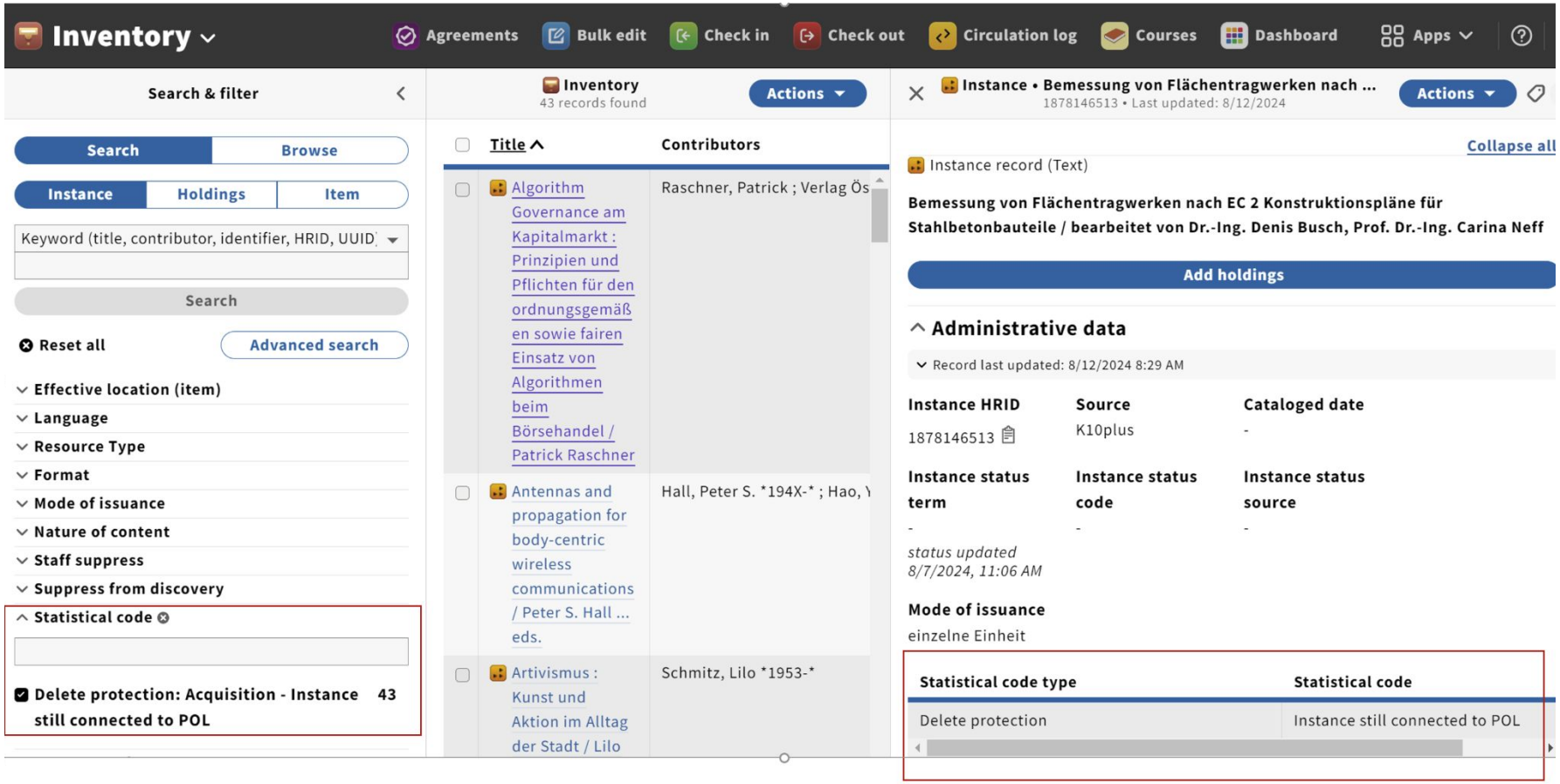


Harvester admin > Permissions

☐	Harvester admin (Settings): View storage engines, transformation pipelines and transformation steps	Berechtigung
☐	Harvester admin (Settings): View, edit storage engines, transformation pipelines and transformation steps	Berechtigung
☐	Harvester admin (Settings): View, edit, create storage engines, transformation pipelines and transformation steps	Berechtigung
☐	Harvester admin (Settings): View, edit, create, delete storage engines, transformation pipelines and transformation steps	Berechtigung
☐	Harvester admin: View harvestables	Berechtigung
☐	Harvester admin: View jobs and failed records	Berechtigung
☐	Harvester admin: View, edit harvestables, start/stop jobs	Berechtigung
☐	Harvester admin: View, edit, create harvestables, start/stop jobs	Berechtigung
☐	Harvester admin: View, edit, create, delete harvestables, start/stop jobs	Berechtigung
☐	Harvester-Verwaltung: Alle Berechtigungen	Berechtigung

Solution for Missing Information in Harvester admin App about Skipped Deletions

- Statistical code on Instance-, Holdings- and Item-level

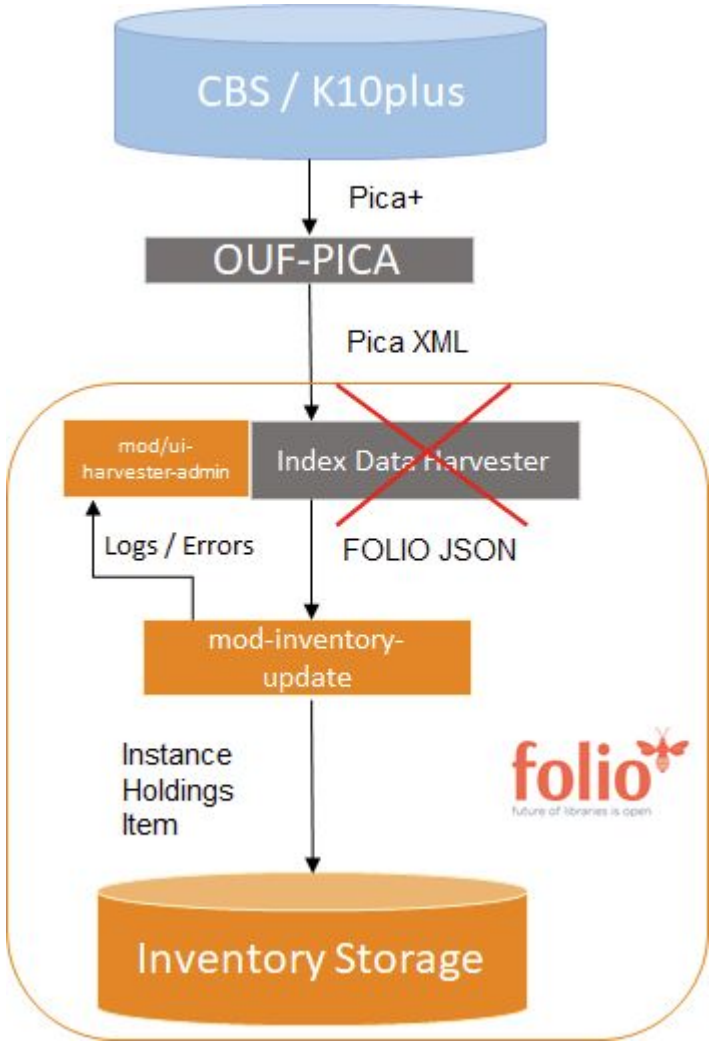


Solution for Missing Information in Harvester admin App about Skipped Deletions

- 4 different delete protection events
  - Acquisition - Instance still connected to POL
  - Holdings-Record cataloged in FOLIO Inventory
  - Item-Record cataloged in FOLIO Inventory
  - Circulation - Item status prevents deletion
- Documentation
  - https://github.com/folio-org/mod-inventory-update?tab=readme-ov-file#statistical-coding-of-delete-protection-events
  - Example: https://github.com/indexdata/cbs2folio-transformations/blob/master/pica2instance-new.xsl

Planned Development

- Automated deletions of logs of old jobs and failed records
- Minor enhancements in the UI
- Documentation in the FOLIO Wiki (work in progress)
- Refactor the legacy backend (Index Data Harvester) as native FOLIO module and part of mod-harvester-admin



folio

future of libraries is open

Thank you!

Antje Niemann (antje.niemann@gbv.de)
Felix Hemme (felix.hemme@gbv.de)