



# 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)

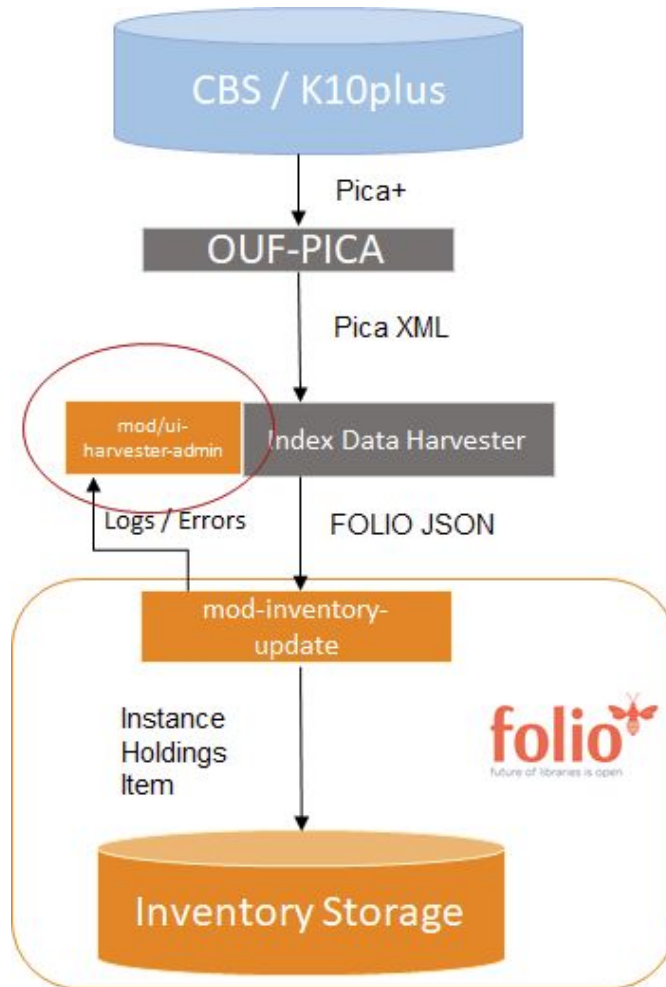


# Data Management in Heterogeneous Systems

- 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** | Agreements | Bulk edit | Check in | Check out | Circulation log | Courses | Dashboard | Data export | Data import | eHoldings | Apps | Zentrale

Harvestables | Jobs | Failed records

Harvestables  
1 record found

Name	Status	Records	Last harvest finished	Enabled?	Job class	ID	Message (processed/loaded/deleted/failed)
bremen_test	OK	2	2:58:50 PM, August 16, 2024	No	XML bulk	8063472 3119884 2	Instances:2/2/0(0)/0 Holdings:4/4/0/0 Items:4/4/0/0

End of list



# Harvester admin App > Jobs

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

Harvester admin

Agreements Bulk edit Check in Check out Circulation log Courses Dashboard Data export Data import eHoldings Apps ? Zentrale

Harvestables Jobs Failed records

Jobs  
38,894 records found

Actions

Harvestable name	Status	Records	Seconds	Started	Finished	Job class	Message (processed/loaded/deleted/failed)
bremen_test	OK	2 (2)	28	12:58:22 PM, August 16, 2024	12:58:50 PM, August 16, 2024	XML bulk	Instances:2/2/0(0)/0 Holdings:4/4/0/0 Items:4/4/0/0
bremen_test	OK	1 (1)	11	12:53:16 PM, August 16, 2024	12:53:27 PM, August 16, 2024	XML bulk	Instances:1/1/0(0)/0 Holdings:2/2/0/0 Items:2/2/0/0
bremen_test	OK	2 (2)	8	12:48:27 PM, August 16, 2024	12:48:35 PM, August 16, 2024	XML bulk	Instances:2/2/0(0)/0 Holdings:5/5/0/0 Items:5/5/0/0
bremen_test	OK	2 (2)	12	12:48:17 PM, August 16, 2024	12:48:29 PM, August 16, 2024	XML bulk	Instances:2/2/0(0)/0 Holdings:5/5/0/0 Items:5/5/0/0
bremen_test	OK	1 (1)	6	12:43:21 PM, August 16, 2024	12:43:27 PM, August 16, 2024	XML bulk	Instances:1/1/0(0)/0 Holdings:3/0/0/3 Items:3/0/0/3
bremen_test	OK	1 (1)	9	12:43:11 PM, August 16, 2024	12:43:20 PM, August 16, 2024	XML bulk	Instances:1/1/0(0)/0 Holdings:3/3/0/0 Items:3/3/0/0
bremen_test	OK	1 (1)	9	12:33:16 PM, August 16, 2024	12:33:25 PM, August 16, 2024	XML bulk	Instances:1/1/0(0)/0 Holdings:1/1/0/0 Items:1/1/0/0
bremen_test	OK	1 (1)	10	12:28:15 PM, August 16, 2024	12:28:25 PM, August 16, 2024	XML bulk	Instances:1/1/0(0)/0 Holdings:1/1/0/0 Items:1/1/0/0
bremen_test	OK	1 (1)	5	12:24:00 PM, August 16, 2024	12:24:05 PM, August 16, 2024	XML bulk	Instances:1/1/0(0)/0 Holdings:1/1/0/0 Items:1/1/0/0



# Harvester admin App > Failed records

- One entry per instance record, pre-filtered by date  $\geq$  yesterday

**Harvester admin** | Agreements | Bulk edit | Check in | Check out | Circulation log | Courses | Dashboard | Data export | Data import | eHoldings | Apps | Centrale

Harvestables | Jobs | **Failed records** (1 record found)

**Search & filter** | **Failed records** | Actions

Record number	Instance title	Errors	Time stamp
129082430	Le mouvement social / publ. par l'Association Le Mouvement Social. Avec le concours de l'Institut des Sciences Humaines et Sociales du Centre National de la Recherche Scientifique et avec la collaboration du Centre d'Histoire Sociale du XXe Siècle de l'Université Paris I	ERROR: Cannot update record ce031f66-6124-41e1-86e3-1c0709499843 because it has been changed (optimistic locking): Stored _version is 11, _version of request is 10 (23F09)	2024-08-

End of list

From: 08/15/2024 | To: MM/DD/YYYY





# Settings > Harvester admin

- Transformation steps: Store the XSLT stylesheets

The screenshot shows the 'Settings' page with the 'Harvester admin' section active. Under 'Transformation steps', a modal window is open for the step named 'bremen\_test locations2uuid-ils21 (XML -> XML)'. The modal displays the following details:

- Name:** bremen\_test locations2uuid-ils21
- Description:** -
- Enabled (This is unused):** X
- Type:** XmlTransformStep
- Input format:** XML
- Output format:** XML
- Script:**

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:output indent="yes" method="xml" version="1.0" encoding="UTF-8"/>
  <xsl:key name="original" match="record/original/item" use="@eps"/>
  <xsl:template match="@*" node()*>
    <xsl:copy>
      <xsl:apply-templates select="@*" node()*>
    </xsl:copy>
  </xsl:template>
  <!-- Map locations -->
  <xsl:template match="permanentLocationId">
    <xsl:variable name="electronicholding" select="substring(/datafield[@tag='002@']/
subfield[@code='0',1,1])"/>
    <xsl:variable name="lower" select="translate(,ABCDEFGHIJKLMNOPQRSTUVWXYZUVWXYZ-,
abcdefghijklmnopqrstuvwxyz)"/>
    <xsl:variable name="bridepn" select="preceding-sibling::brid"/>
    <xsl:variable name="i" select="key('original', $bridepn)"/>
    <xsl:variable name="xselectioncode" select="substring(SU/datafield[@tag='208@']/
subfield[@code='b',1,3])"/>
    <permanentLocationId>
      <xsl:choose>
        <!-- TBS -->
```



# Settings > Harvester admin

- Transformation pipelines: Connect the transformation steps in sequence

The screenshot shows the 'Settings' page with the 'Harvester admin' section selected. Under 'Harvester admin', 'Transformation pipelines' is highlighted. A modal window titled 'Transformation pipelines' shows details for a pipeline named 'bremen\_test'. The modal includes a 'Name' field with the value 'bremen\_test', a 'Description' field with the value 'created via MHA API', an 'Enabled' checkbox that is checked, and a 'Parallel' checkbox that is unchecked. Below these fields is a table titled 'Transformation steps' with the following data:

#	Name	In	Out
1	bremen_test pica2instance-new	XML	XML
2	bremen_test relationships	XML	XML
3	bremen_test holdings-items	XML	XML
4	bremen_test locations2uuid-illn21	XML	XML
5	bremen_test codes2uuid	XML	XML
6	bremen_test instance XML to JSON	XML	JSON



# Settings > Harvester admin

- Storage engines: Contains configurations for FOLIO access

The screenshot shows the Folio Settings interface. The top navigation bar includes various settings categories like Agreements, Bulk edit, Check in, Check out, Circulation log, Courses, Dashboard, Data export, Data Import, eHoldings, ERM comparisons, Apps, and Zentrale. The main content area is divided into three panes:

- Settings:** A sidebar menu with options like Acquisition units, Agreements, Calendar, Circulation, Courses, Data export, Data import, Developer, eHoldings, eUsage, Finance, GOBI integration, Harvester admin (selected), Inventory, Invoices, Licenses, and Local KB admin.
- Harvester admin:** A sub-menu with options for Storage engines (selected), Transformation pipelines, and Transformation steps.
- Storage engines:** A list of storage engines, with "Batch Upsert bremen\_test" selected. A "+ New" button is visible.

The details for the "Batch Upsert bremen\_test" storage engine are shown in a modal window:

- Name:** Batch Upsert bremen\_test
- Description:** created by iLOADER-Script via MHA API
- Enabled:**
- URL:** http://okapi.okapi.svc.cluster.local:9130/
- JSON configuration:**

```
{
  "folioAuthPath": "bi-users/login",
  "folioTenant": "bremen_test",
  "folioUsername": "BreTeAdm",
  "folioPassword": "****censored****",
  "inventoryUpsertPath": "inventory-upsert-hrid",
  "inventoryBatchUpsertPath": "inventory-batch-upsert-hrid",
  "logHistoryStoragePath": "harvester-admin/harvestables/{id}/log/store",
  "folioAuthSkip": "false"
}
```



# Harvester admin > Permissions

<input type="checkbox"/>	Harvester admin (Settings): View storage engines, transformation pipelines and transformation steps	Berechtigung
<input type="checkbox"/>	Harvester admin (Settings): View, edit storage engines, transformation pipelines and transformation steps	Berechtigung
<input type="checkbox"/>	Harvester admin (Settings): View, edit, create storage engines, transformation pipelines and transformation steps	Berechtigung
<input type="checkbox"/>	Harvester admin (Settings): View, edit, create, delete storage engines, transformation pipelines and transformation steps	Berechtigung
<input type="checkbox"/>	Harvester admin: View harvestables	Berechtigung
<input type="checkbox"/>	Harvester admin: View jobs and failed records	Berechtigung
<input type="checkbox"/>	Harvester admin: View, edit harvestables, start/stop jobs	Berechtigung
<input type="checkbox"/>	Harvester admin: View, edit, create harvestables, start/stop jobs	Berechtigung
<input type="checkbox"/>	Harvester admin: View, edit, create, delete harvestables, start/stop jobs	Berechtigung
<input type="checkbox"/>	Harvester-Verwaltung: Alle Berechtigungen	Berechtigung

# Solution for Missing Information in Harvester admin App about Skipped Deletions



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

The screenshot shows the Harvester admin app interface. On the left, there is a search and filter sidebar with options like 'Instance', 'Holdings', and 'Item'. A search bar is present, and a 'Delete protection: Acquisition - Instance 43 still connected to POL' checkbox is checked. The main area displays a list of records with columns for 'Title' and 'Contributors'. The detailed view on the right shows the instance record for 'Bemessung von Flächentragwerken nach EC 2 Konstruktionspläne für Stahlbetonbauteile / bearbeitet von Dr.-Ing. Denis Busch, Prof. Dr.-Ing. Carina Neff'. It includes administrative data such as 'Instance HRID', 'Source', and 'Cataloged date'. A table at the bottom of the detailed view shows the 'Statistical code type' and 'Statistical code' for the instance.

Statistical code type	Statistical code
Delete protection	Instance still connected to POL

# Solution for Missing Information in Harvester admin App about Skipped Deletions

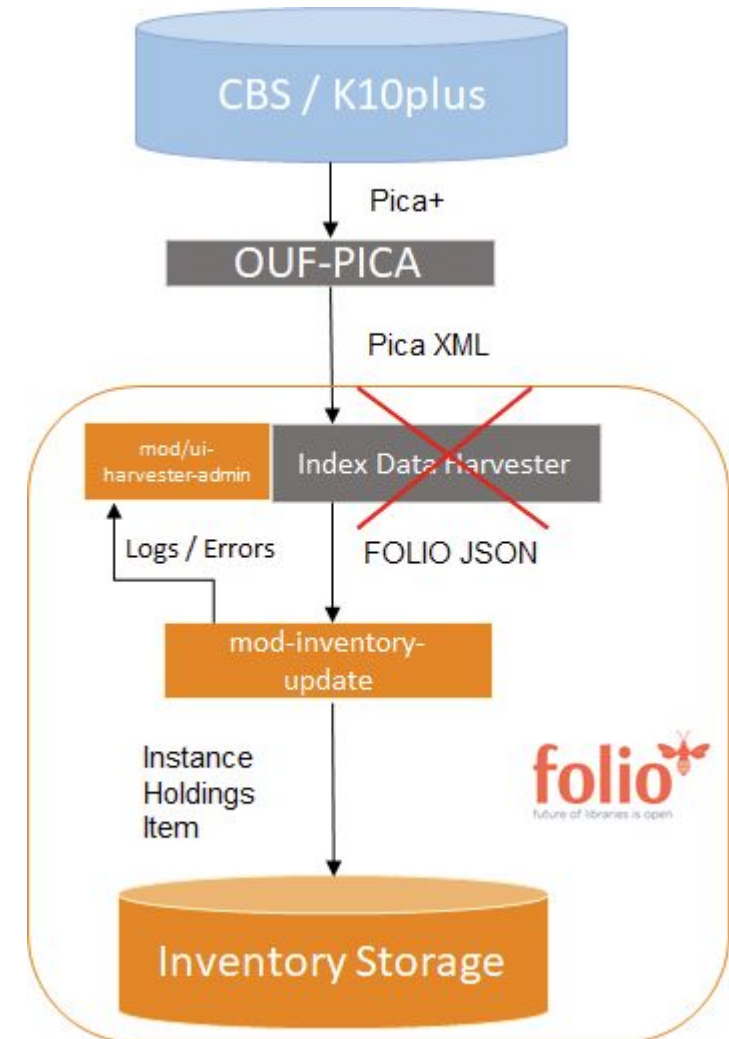


- 4 different delete protection events
  - Acquisition - Instance still connected to POL
  - Holdings-Record cataloged in FOLIO Inventory
  - Item-Record cataloged in FOLIO Inventory
  - Circulation - Item status prevents deletion
- Documentation
  - <https://github.com/folio-org/mod-inventory-update?tab=readme-ov-file#statistical-coding-of-delete-protection-eve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





Thank you!

Antje Niemann ([antje.niemann@gbv.de](mailto:antje.niemann@gbv.de))

Felix Hemme ([felix.hemme@gbv.de](mailto:felix.hemme@gbv.de))