

112. BiblioCon 2024 in Hamburg

Auf Wolke 7 oder doch nur Wolke 4?

**Katalogisierung mit FOLIO im GBV live –
Ein Bericht nach 8 Monaten Online-Update**

Felix Hemme, Antje Niemann, VZG Göttingen

04.06.2024



Auf Wolke 7 oder doch nur Wolke 4?

Wolke 7

„Wer auf Wolke 7 schwebt [...], der hat das höchste Glück erreicht.“ [1]

Wolke 4

„Lass uns die Wolke 4 bitte nie mehr verlassen,
Weil wir auf Wolke 7 viel zu viel verpassen.
Ich war da schon einmal, bin zu tief gefallen.

Lieber Wolke 4 mit dir als unten wieder ganz allein.“ Song von Philipp Dittberner

„Es ist okay - es geht halt nicht besser.“ - „Man kann nicht immer auf Wolke 7 sein – bei all dem Stress“ [2]

Wolke?

Wir sind nicht aus allen Wolken gefallen!

[1] <https://www.geo.de/geolino/redewendungen/3988-rtkl-redewendung-auf-wolke-7-schweben>
[2] <https://oe3.orf.at/stories/2728578/>



Wo stehen wir?

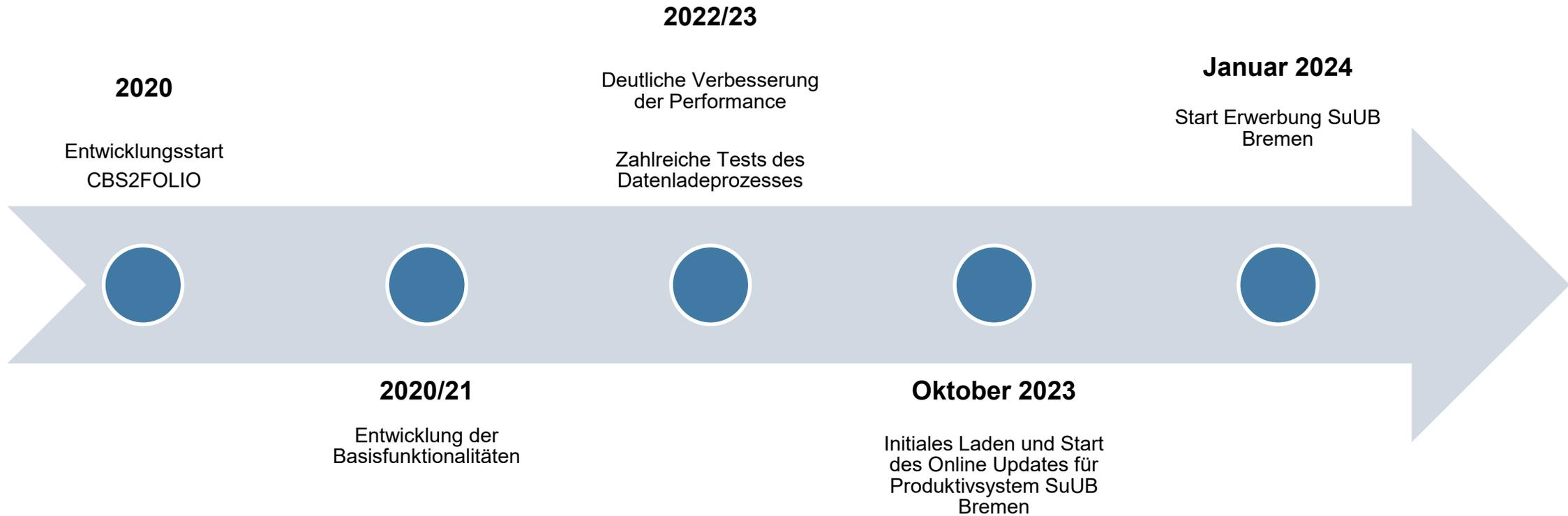
- Über 200 GBV-Bibliotheken katalogisieren im K10plus-Verbundkatalog (CBS)
- FOLIO als zukünftiges System u. a. für Erwerbungs- und Ausleihvorgänge
 - Dazu werden Titel- und Exemplardaten aus dem CBS benötigt, und zwar alle „alten“ Daten und neue in Echtzeit und hochperformant
- CBS2FOLIO
 - Schnittstelle zur Überführung relevanter Titel- und Exemplardaten aus dem CBS zu FOLIO
 - Entwicklungsprojekt der VZG seit 2020 zusammen mit Index Data
 - Ansprechpartner:in in der VZG: Felix Hemme, Antje Niemann
- Seit Oktober 2023 erste Bibliothek (SuUB Bremen) live mit CBS2FOLIO



Worum geht es in unserem Vortrag?

- Seit Oktober 2023 erste Bibliothek (SuUB Bremen) live mit CBS2FOLIO
 - Bewährte Praktiken für eine reibungslose Implementierung
 - Was erwartet unsere Bibliotheken?
 - Welche Arbeiten übernimmt die Verbundzentrale des GBV (VZG)?
 - Welche Aufwände entstehen vor Ort?
 - Herausforderungen und Lösungen bzgl. initialem Laden der Bestandsdaten und Einrichtung des Online-Updates
 - Bilanz zur Verlässlichkeit der Prozesse nach über einem halben Jahr
- Vorstellung der CBS2FOLIO-Weiterentwicklungen der letzten 12 Monate

CBS2FOLIO





Vorarbeiten für das initiale Laden der Daten

VZG

Gesamtabzug aus CBS

Selektion der Konvolute aus dem LBS

Servereinrichtung mit aktuellsten CBS2FOLIO-Komponenten

XSLT Mapping-Tabellen → GitHub-Repository

Laden der Referenzwerte → GitHub-Repositories

Index (OpenSearch)

Bibliothek

Datenbereinigung u. Auskunft über eigene Daten

- Doppelte Barcodes bereinigen
- Verdeckter Bestand vh.
- ...

Lokale Einstellungen

- Standorte → FOLIO (Browser) und XSLT-Stylesheet ([locations2uuid-ilmXX.xsl](#))
- Theken → FOLIO (Browser)
- Rechtemanagement für Katalog-Daten → FOLIO (Browser)



Initiales Laden der Daten durch VZG

- Konfiguration, Anstoßen und Überwachen des Prozesses
- Gründliche Auswertung der Log-Dateien für SuUB Bremen
- Probleme und Fehler (z. B. Codierung der Umlaute) beim Bremer Load konnten wir zeitnah korrigieren und als *“Erkenntnisgewinne”* verbuchen
- Weitere Verbesserung der Performance in 2024 (9 Stunden für die Daten der SUB Göttingen, insg. > 25 Mio FOLIO-Records)
- Automatisierter Prozess zur Erstellung der Konvolute in FOLIO
- Automatisierung und Generalisierung des Datenladeprozesses und der Konfiguration der CBS2FOLIO-Komponenten

Fehlermeldungen

1. Doppelter Barcode

```
lower(jsonb ->> 'barcode':text) value already exists in table ite
```

Bei der **Fehlerbehebung** durch die Bibliothek muss beachtet werden, dass der Barcode

2. Kein Titel vorhanden

```
must not be null
```

Erscheint, wenn kein Titel (4000) vh. ist (z. B. bei Av-Sätzen) und wird jetzt durch XSLT **Fehlerbehebung**: Liste der PPN erstellen und Touch Title (Felix hat failed-records.json)

3. Kein Item angelegt

```
Holdings record does not exist
```

→ bei allen unten aufgeführten Titeln mit dieser Fehlermeldung wurde **kein Item angelegt** und **entsprechend auch keine Fehlermeldung mehr ausgegeben**. → **Beobachten**

Fehlerbehebung: Liste der PPN erstellen und Touch Title (Felix hat failed-records.json)

138324298: kein Barcode im CBS

135623006:

138324344: c-Stufe → **erneutes Anfassern der Titelaufnahme bringt ein intellectua**

138324298: CBS sieht alles gut aus

139894314:

141520027: c-Stufe

141520043: kein item, obwohl es einen Barcode hat. Das Buch wäre auch nicht ausleihen

141520035: Barcode nicht gespeichert wie auch 8100

4. Parellele Zugriffe (optimistic locking)

```
ERROR: Cannot update record 93d9b2cb-db59-49df-9632-6e9c39fb4253 b
```

→ Auf einen schnellen Blick sehen die Titel gut aus.

1603104666

130687952

(Felix hat failed-records.json und liefert Antje liste der Instance-HRIDs) ERLEDIGT

5. HRID missing or empty

```
HRID is missing or empty.
```

timestamp-20231020094407833

Automatisierung durch Scripting



```
# This Config file is intended for the automatic
# setup of the OUF pipeline for new tenants OR in
# preparation of initial loads.
#
# Please change all needed settings very carefully
# and then run the run.sh script.
#
#####
# What to do?
#
runHarvesterSetup="false"  ### Configure Harvestable and Pipelines
prepareInitialLoad="true"  ### Prepare Migration-Slots
runInitialLoad="true"      ### Initial Data load for tenant
                           ### OUF-Setup must already be done!

# tenant related settings
# -----
folioTenant='bremen'
folioTenantIln='21'
folioUsername='XXXXXXX'
folioPassword='XXXXXXX'

# okapi related settings
# -----
okapiProtocol='https'
okapiUrl='okapi.gbv.de'
okapiPort=''

# HARVESTABLE related settings
# -----
oufBatchSize='20'
oufServiceProvider='193.174.240.7:1093'
oufLogLevel='WARN'
oufPodName='ouf-bremen-prod'

# Initial Load related settings
# -----
ilNumberOfImportSlots=10
ilOufLogLevel='INFO'
ilOufBatchSize='500'
ilOufSplitSize='1000'
ilOufCbsDataFileName='Abz_21_Tit.001_20230910_165600.nrm.zst'  ### Data File in Nextcloud
```

Config-Datei

- 00-get-okapi-token.sh
- 01-create-storage-engine.sh
- 02-create-steps.sh
- 03-fill-steps-with-xslt.sh
- 04-create-transformation-pipeline.sh
- 05-create-harvestable.sh
- 06-create-initial-load-harvestables.sh
- 07-prepare_initial_load.sh
- 08-do-initial_load.sh
- 09-remove-il-harvestables.sh

Skripte

```
# Check if runHarvesterSetup is set and if it's true
if [[ "$runHarvesterSetup" == "true" ]]; then

    echo -e "\n\033[1mErstelle Storage Engine\033[0m\n"
    ./01-create-storage-engine.sh

    echo -e "\n\033[1mErstelle Steps\033[0m\n"
    ./02-create-steps.sh
fi

# This step seems to be common for both runHarvesterSetup and prepareInitialLoad
echo -e "\n\033[1mXSLTs werden in die steps eingefügt\033[0m\n"
./03-fill-steps-with-xslt.sh

if [[ "$runHarvesterSetup" == "true" ]]; then

    echo -e "\n\033[1mErstelle Transformation-Pipeline\033[0m\n"
    ./04-create-transformation-pipeline.sh

    echo -e "\n\033[1mErstelle Harvestable\033[0m\n"
    ./05-create-harvestable.sh

    echo -e "\n\033[1mReady. Check https://harvester.folio.gbv.de/harvester-administration\033[0m\n"

    Check if prepareInitialLoad is set and if it's true
    if [[ "$prepareInitialLoad" == "true" ]]; then

        echo -e "\n\033[1mErstelle Migrations-Harvestables\033[0m\n"
        ./06-create-initial-load-harvestables.sh

        echo -e "\n\033[1mBereite das initiale Laden vor.\033[0m\n"
        ./07-prepare_initial_load.sh
    fi

    # Check if runInitialLoad is set and if it's true
    if [[ "$runInitialLoad" == "true" ]]; then
        echo -e "\n\033[1mBeginne Initial load, starte Migrations-Harvestables.\033[0m\n"
        ./08-do-initial_load.sh
    fi
fi
```

run.sh ruft Skripte auf

VZG: Generalisierung/Automatisierung stets im Fokus



indexdata / cbs2folio-transformations

<> Code Issues 1 Pull requests Actions Wiki Security Insights

Files cbs2folio-transformations / gbv /

master

Go to file

- etc
- gbv
 - codes2uuid.xml
 - holdings-items.xml
 - locations2uuid-iln21.xml
 - locations2uuid-iln26.xml
 - locations2uuid-iln40.xml
 - locations2uuid-iln63.xml
 - locations2uuid-iln90.xml
 - pica2instance-new.xml
 - relationships.xml
- harvester
- hebis

```
148 ### POST locations2uuid
149
150 urlLocations2uuid=""
151
152 for url in "${urls[@]"; do
153     if [[ $url == *"locations2uuid"* ]]; then
154         urlLocations2uuid=$url
155         break
156     fi
157 done
158
159 if [ -z "$urlLocations2uuid" ]; then
160     echo "Keine URL gefunden, die 'locations2uuid' enthält."
161 else
162     echo -e "\nDownload von locations2uuid-iln$folioTenantIln.xml"
163     curl "https://raw.githubusercontent.com/indexdata/cbs2folio-transformations/master/gbv/locations2uuid-iln$folioTenantIln.xml" -o "./locations2uuid-iln$folioTenantIln.xml"
164
```

- locations2uuid-iln63.xml
- locations2uuid-iln90.xml
- pica2instance-new.xml
- relationships.xml



Zugriff auf Einstellungen über Harvester-Admin-App

Harvester Admin eManagement Massenbearbeitungen Apps Online

Harvest-Jobs Jobs Fehlgeschlagene Datensätze

Suchen & filtern

Name

Suchen

Alles zurücksetzen

Aktiviert?

Jobklasse

Status

Einstellungen 1 Datensatz gefunden

Einstellungen Harvester Admin Transformation-Steps

bremen

- Ausleihe
- Bestellungen
- Datenexport
- Datenimport
- eHoldings
- eManagement
- Entwickler
- Erwerbungssteams
- eUsage
- Fernmagazin
- Finanzen
- GOBI-Anbindung
- Harvester Admin**

Storage-Engines

Transformations-Pipelines

Transformation-Steps

- bremen codes2uuid (XML → XML)
- bremen holdings-items (XML → XML)
- bremen Instance XML to JSON (XML → JSON)
- bremen locations2uuid-ilm21 (XML → XML)**
- bremen pica2instance-new (XML → XML)
- bremen relationships (XML → XML)

bremen locations2uuid-ilm21 (XML → XML)

```
<permanentLocationId>
  <xsl:choose>
    <!-- TBs -->
    <xsl:when test="$lower='tbwirtschaft'">0d2c3271-917b-4719-938f-e6597dae8991</xsl:when>
    <xsl:when test="$lower='tbwirtschaftmag'">0d2c3271-917b-4719-938f-e6597dae8991</xsl:when>
    <xsl:when test="$lower='tbnavtik'">0d2c3271-917b-4719-938f-e6597dae8991</xsl:when>
    <xsl:when test="$lower='tbsozialw'">182fdb6b-6bd1-4a27-bbbc-867c2aa28f91</xsl:when>
    <xsl:when test="$lower='tbtechnik'">182fdb6b-6bd1-4a27-bbbc-867c2aa28f91</xsl:when>
    <xsl:when test="$lower='tbbhv'">cac66b47-aa65-4403-90e8-d838c1f5ec0e</xsl:when>
    <xsl:when test="$lower='tbkunst'">61d7524b-0508-4e33-a07f-5b71cc9bda4e</xsl:when>
    <xsl:when test="$lower='tbmusik'">08819597-e7fb-421a-98bd-5bf4f89e84b4</xsl:when>
    <!-- Uni/BBs -->
    <xsl:when test="$lower='juridicum'">bcbddc8d-806c-430f-8b04-d514dcb620ce</xsl:when>
    <xsl:when test="$lower='bbnatnw1'">09545336-c48d-4fb9-9e57-752f9b085ebf</xsl:when>
    <xsl:when test="$lower='bbwiwi'">efd3e3e5-b9be-4bfb-8400-b3cbea1dd86a</xsl:when>
    <!-- Uni/Zentrale -->
    <xsl:when test="$lower='magazin'">ab302f5d-221d-4b33-9efb-60fde83d3a87</xsl:when>
    <xsl:when test="$lower='magazinturme01'">55c2115c-cf87-4229-b63f-2b98814d7330</xsl:when>
    <xsl:when test="$lower='magazinturme02'">eb2fb342-0326-43fd-902c-0b7d43529e40</xsl:when>
    <xsl:when test="$lower='magazinturme03'">a43d8a7f-5b50-4e96-982f-eba9020a2b88</xsl:when>
    <xsl:when test="$lower='magazinturme04'">3ed45e06-5fbc-42f1-9fc6-0834a7b5c5d0</xsl:when>
    <xsl:when test="$lower='mediathek'">5f97bf6c-8e94-4ff6-b693-89dba2557153</xsl:when>
    <xsl:when test="$lower='magazinzs'">abddeb0b-102a-4303-a607-c6e31d7d5a14</xsl:when>
    <xsl:when test="$lower='e01'">08a95ad2-865b-4a36-8e30-7ae235ea4362</xsl:when>
    <xsl:when test="$lower='e02'">9fca61eb-af8a-4702-801e-8b143e71a55c</xsl:when>
    <xsl:when test="$lower='e03'">a49313b4-479e-466f-b894-9a66fac2ff42</xsl:when>
    <xsl:when test="$lower='e04'">9df56980-10ed-4a8e-809c-dae115ca11e4</xsl:when>
    <xsl:when test="$lower='zsesesaal'">96b8ec76-2438-4347-abd0-f39430137042</xsl:when>
    <xsl:when test="starts-with($lower,'mf')">5a70506f-8c0d-47ed-ad9a-726196a7f739</xsl:when>
    <!-- Rara -->
    <xsl:when test="$lower='handschrls'">d1819465-4ed0-4af8-8e87-cba2a3c27080</xsl:when>
    <xsl:when test="$lower='tiflis'">d1819465-4ed0-4af8-8e87-cba2a3c27080</xsl:when>
    <xsl:when test="$lower='legion'">d1819465-4ed0-4af8-8e87-cba2a3c27080</xsl:when>
```



Online Update für kontinuierlichen Metadatenfluss

- Einrichtung nach initialem Laden durch VZG
- Systemverwalter:innen in den Bibliotheken können Stand im Browser (OUF API) jederzeit einsehen und zurücksetzen (schnell wieder auf Stand)
- Zugriff auf sehr detaillierte Logs
- Ausfallsicherheit und Monitoring seit Inbetriebnahme von Bremen-Prod weiter verbessert

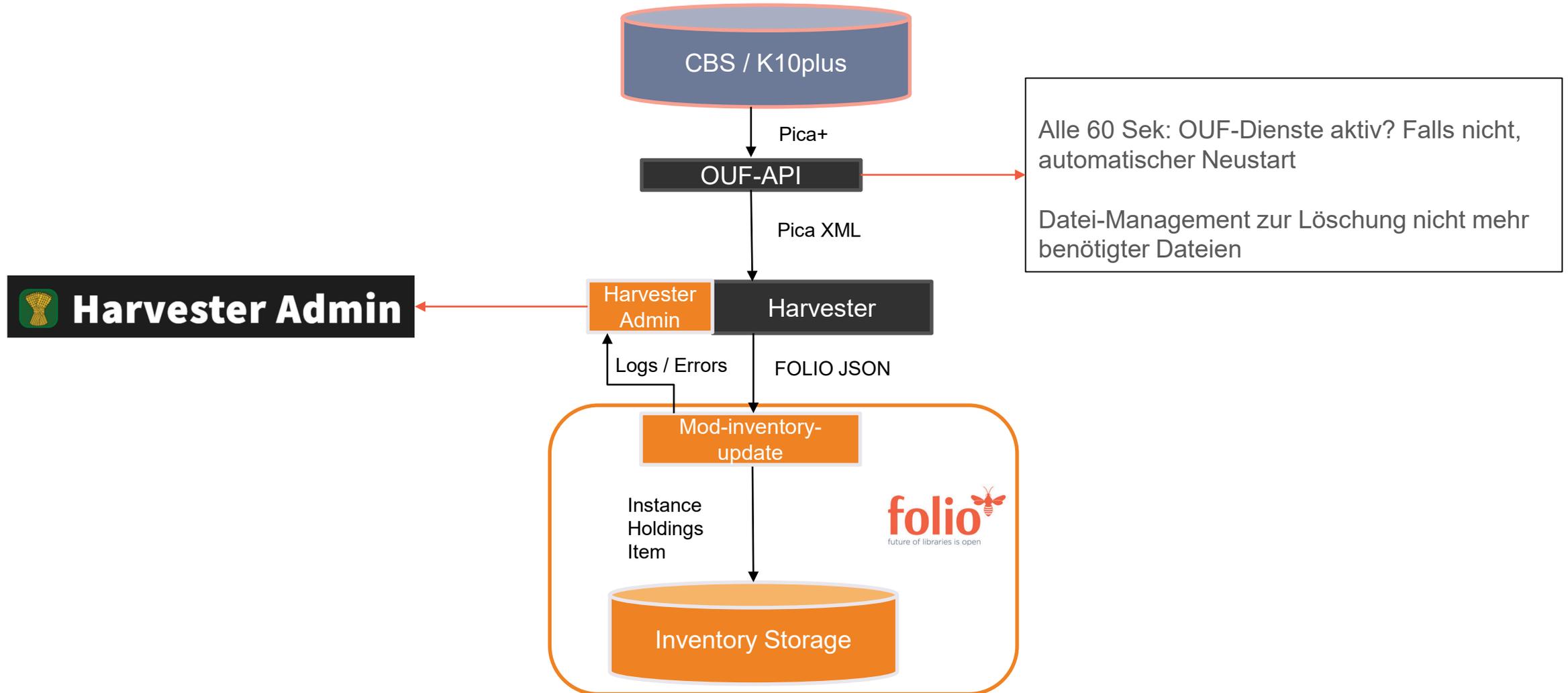
The screenshot shows the OUF API web interface. At the top, it displays 'OUF API' and 'Version 3.1.3 (2024-02-01)' with a 'Setup' button. Below this, there are radio buttons for 'ouf.ini' (selected) and 'other.ini'. A text area contains the following configuration for the UpdateSignal:

```
[UpdateSignal]
; Last Time is automatically updated each time a new Update Signal File
has been made accessible to the Update Process
; Starting point the next time that ouf update is started and begins
polling for update signals. All update signals which were received
; since this stored last update time will be retrieved
Last Time = 24-05-2024 18:23:10.496541
```

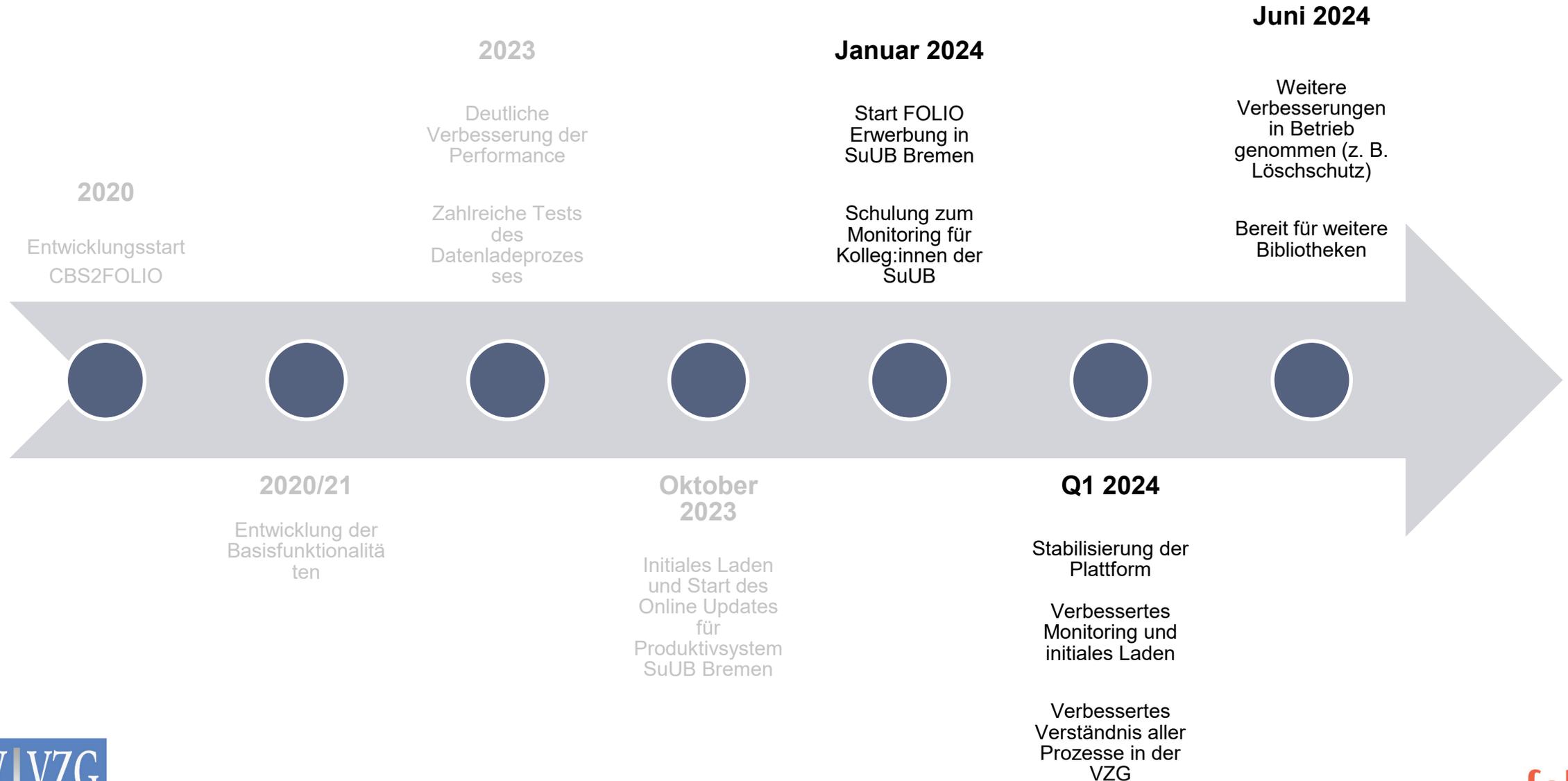
The 'Last Time' value is circled in red. Below the text area, there is a navigation bar with 'OUF API', 'Status', 'Settings', 'Env', 'Start', and 'Stop'. Underneath, there is a 'Timestamp (optional)' label and an input field with the example 'Ex. 16-06-2020 10:20:07'. At the bottom, there are 'Start' and 'Cancel' buttons.



Monitoring des Online Updates



CBS2FOLIO

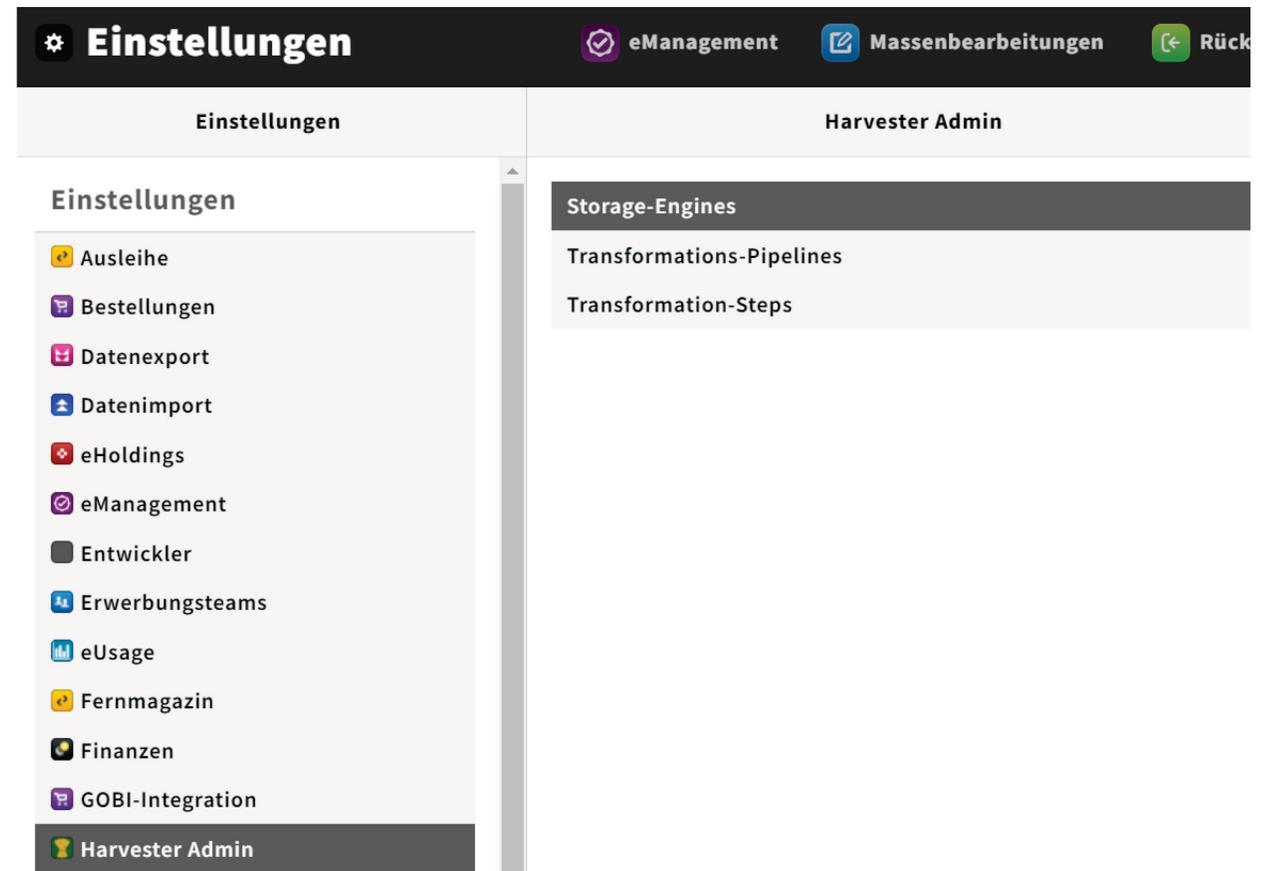




Wichtige Weiterentwicklungen der letzten 12 Monate

Harvester Admin bietet den Zugang zur Konfiguration von

- Transformation-Steps (Mappings/XSLT)
- Transformation-Pipelines (Reihenfolge der Transformation-Steps)
- Storage-Engines (Konfiguration von Zugangsdaten und APIs)
- Harvestables (parallele Initial Load Jobs und Online Update)





Wichtige Weiterentwicklungen der letzten 12 Monate

Harvester Admin ermöglicht das Monitoring und Zugriff auf Fehlermeldungen

Harvester Admin eManagement Massenbearbeitungen Rückgabe Ausleihe Ausleihprotokoll Apps Onlin

Harvest-Jobs Jobs **Fehlgeschlagene Datensätze**

Suchen & filtern Fehlgeschlagene Datensätze 6 Datensätze gefunden **Aktionen**

Datensatznummer	Instanz-HRID	Instanztitel	Fehler	Zeitstempel
311808174	311808174	Geschichte des Reisens und des Tourismus : ein Überblick von den Anfängen bis zur Gegenwart / Petra Krempien	lower(jsonb ->> 'barcode'::text) value already exists in table item: 46b\$003816311	2024-05-28T10:15:10
1613252617	1613252617	Zur Messung der Erlebnisqualität von Erholungs-Waldbeständen : eine experimentalpsychologische Analyse als Beitrag zur Umweltgestaltung / Hartmut Jacob	lower(jsonb ->> 'barcode'::text) value already exists in table item: 46\$0003046842	2024-05-28T14:49:46
615960863	615960863	Managing change in tourism : creating opportunities - overcoming obstacles / ed. by Peter Keller With contributions by Monika Bachinger ...	lower(jsonb ->> 'barcode'::text) value already exists in table item: 46b\$002933233	2024-05-28T16:53:43
1629007927	1629007927	Schlaraffenland, Bühne und Ventil : ein Plädoyer für den ganz normalen Bade- und Pauschal Tourismus / Walter Kiefl	lower(jsonb ->> 'barcode'::text) value already exists in table item: 46b\$001665898	2024-05-28T17:27:23
626107997	626107997	Tourismuslehre - ein Grundriss / Thomas Bieger	lower(jsonb ->> 'barcode'::text) value already exists in table item: 46b\$001060791	2024-05-28T17:34:04
246697881	246697881	GeoLex : Kurzlehrbuch und Geographie-Lexikon für den Tourismus / von Hans Bartl (Hrsg.)	lower(jsonb ->> 'barcode'::text) value already exists in table item: 46b\$002769018	2024-05-28T17:44:36



Wichtige Weiterentwicklungen der letzten 12 Monate

- Löschschutz für
 - in FOLIO erfasste Bestände und Exemplare
 - Instanzen, die mit einem Bestellsatz verknüpft sind
 - Exemplare im Ausleihprozess

MIU will avoid deleting items with following item statuses

^ Statistischer Code ✕

Löschschutz: Ausleihe - 1
Exemplarstatus
blockiert Löschung

- Awaiting delivery
- Awaiting pickup
- Checked out
- Aged to lost
- Claimed returned
- Declared lost
- Paged
- In transit



Ausblick CBS2FOLIO

- Dokumentation der Komponenten
- Fortsetzung der bedarfsgerechten Weiterentwicklung in Abstimmung mit den nachnutzenden Institutionen
- Inbetriebnahme weiterer Produktivsysteme

CBS2FOLIO

Wolke 7



GBV | VZG



antje.niemann@gbv.de
felix.hemme@gbv.de

Sie finden uns am gemeinsamen
Stand der Verbände (081).