

# Alma Stresstest –

wie man 25 Mio. Datensätze in einer Network Zone aktualisiert

Michaela Sieber

---

BSZ  GBV

---

# Hintergrund – K10plus

Fusion der Verbunddatenbanken

von GBV und SWB

Eine gemeinsame Datenbank

mit mehr als 100 Mio. Titeldaten

Alma Network Zone (CBDZ)

Enthält bis 15.3.2019 nur SWB-Daten



# Hintergrund

- Zusammenlegung der Verbunddatenbanken von SWB und GBV
  - Änderung des Primären Identifiers = K10plus-PPN
    - MARC 035 (**DE-627**).....
  - In der CBDZ – alle Daten des SWB mit SWB-PPN
    - MARC 035 (DE-576).....
- Für laufende Updates aus K10plus wird die K10plus-PPN als eindeutiges Match-Kriterium benötigt

# Der Plan

- Alle 25 Millionen SWB-Daten in der CBDZ neu laden
  - Abzug über alle Titel in K10plus mit SWB-PPN
  - Update der bereits existierenden Daten in der CBDZ über SWB-PPN = Match auf MARC 035 (DE-576)
  - Nach Möglichkeit vollautomatisiert, Vermeidung von manuellem Eingreifen

→ Verwendung von Import-Profil in Kombination mit Webhooks

# Fragen

- Wie groß darf eine Datei sein?
- Wie viele Datensätze dürfen in einer Datei geladen werden?
- Wird das Mandantensystem beeinträchtigt durch die Massenänderung von Titeldaten in der Network Zone?
- Wie lange wird der Ladeprozess insgesamt dauern?

# Die Umsetzung - Vorbereitungen

- MARCxml-Dateien mit jeweils ca. 3GB
- Empfehlung Ex Libris: max. 1 Mio. Titel pro Datei
- Separates Importprofil für K10plus-Update welches durch API getriggert wird
- Webhook in Alma konfigurieren und Endpoint im BSZ programmieren

# Die Umsetzung - Vorbereitungen

- Importprofil

Profile Details			
Profile name	Update BSZ data for k10plus		
Profile description	updates SWB records (DE-576) via SWB-PPNs		
Originating system	The Library Service Center for Baden-Wuerttemberg	File name patterns	-
Import Protocol	FTP	Cross walk	No
Physical source format	XML		
Source format	MARC21 Bibliographic	Target format	MARC21 Bibliographic
Status	Inactive		

Scheduling	
Files to import	New
Scheduler status	Inactive
Scheduler	Not scheduled

# Die Umsetzung - Vorbereitungen

- Webhook-Konfiguration

Alma Configuration

< Integration Profile Cancel Save

Webhooks i ▼

Code WEBHOOKS  
Integration Type Webhooks  
Managed in Network No

General Information Actions Contact Info

WEBHOOK CONFIGURATION

Status **Pending activation** Activate

Webhook listener URL \*

Secret \*

Message type \*  JSON  XML

SUBSCRIPTIONS

Job Finish  

Notifications  Configure notification types

Pressing this button will send a challenge request to the configured URL and validate the response. The profile must still be saved to retain this activation. For more information, please see the Webhooks documentation.

# Die Umsetzung - Vorbereitungen

- 35 MARCxml-Dateien mit jeweils ca. 3GB

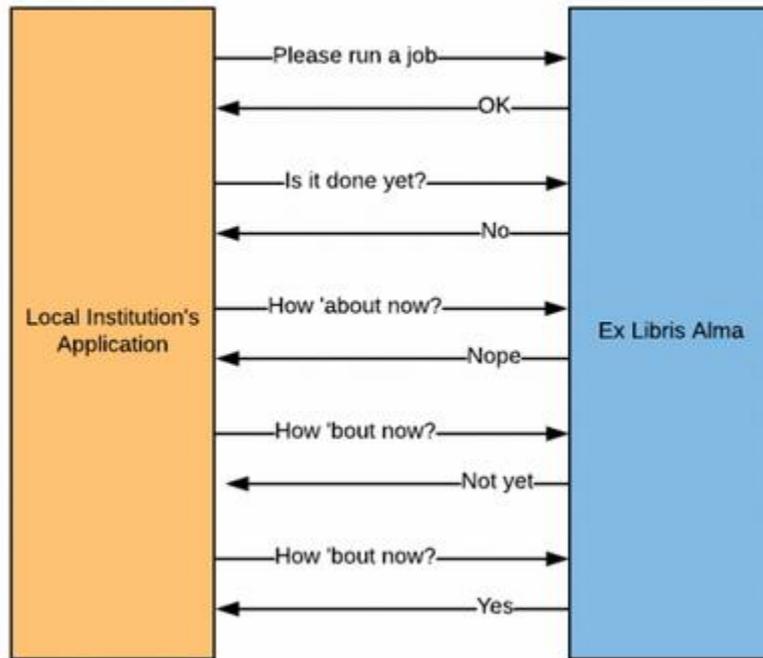
Datensätze	Datei	Status
463.419	001	geladen
547.586	002	geladen
501.465	003	geladen
471.195	004	geladen
446.850	005	geladen
450.098	006	geladen
436.963	007	geladen
<b>410.901</b>	008	geladen
490.668	009	geladen
793.969	010	geladen
843.127	011	geladen
844.193	012	geladen
859.421	013	geladen
889.524	014	geladen
895.885	015	geladen
905.871	016	geladen
898.900	017	geladen
931.738	018	geladen
900.219	019	geladen

914.372	020	geladen
<b>934.364</b>	021	geladen
826.379	022	geladen
852.168	023	geladen
908.956	024	geladen
899.061	025	geladen
820.383	026	geladen
696.433	027	geladen
706.879	028	geladen
677.369	029	geladen
813.137	030	geladen
886.647	031	geladen
791.532	032	geladen
672.522	033	geladen
447.526	034	geladen
485.844	035	geladen
<b>25.315.564</b>		

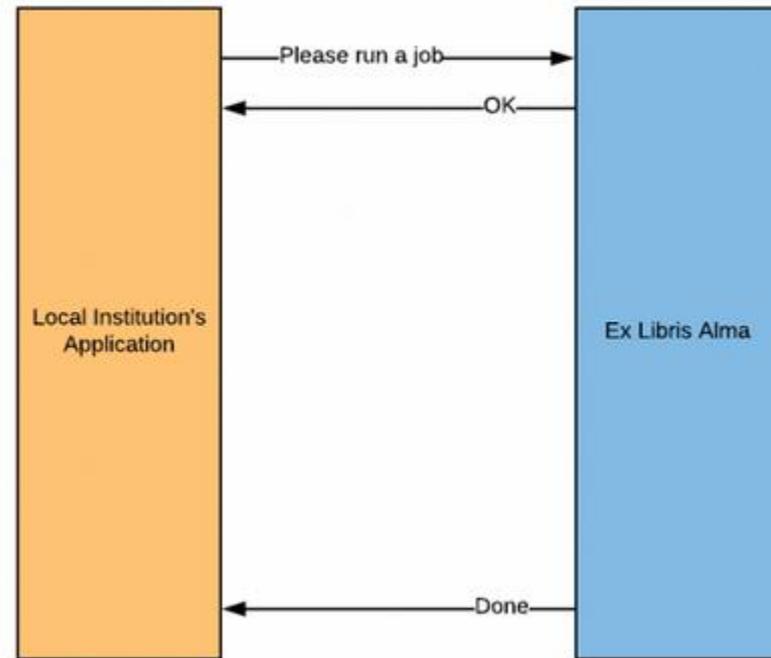
# Die Umsetzung – Exkurs Webhooks

*“Webhooks are “user-defined HTTP callbacks.” They are usually triggered by some event... When that event occurs, the source site makes an HTTP request to the URI configured for the webhook.”*

## Polling with APIs



## Webhooks

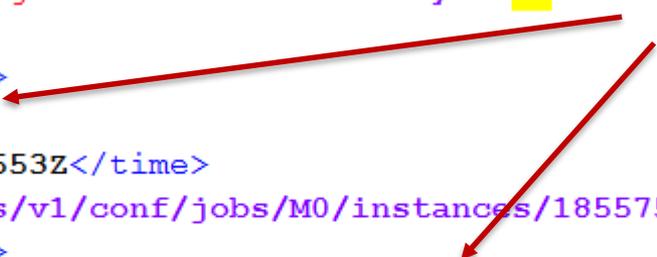


# Die Umsetzung – Exkurs Webhooks

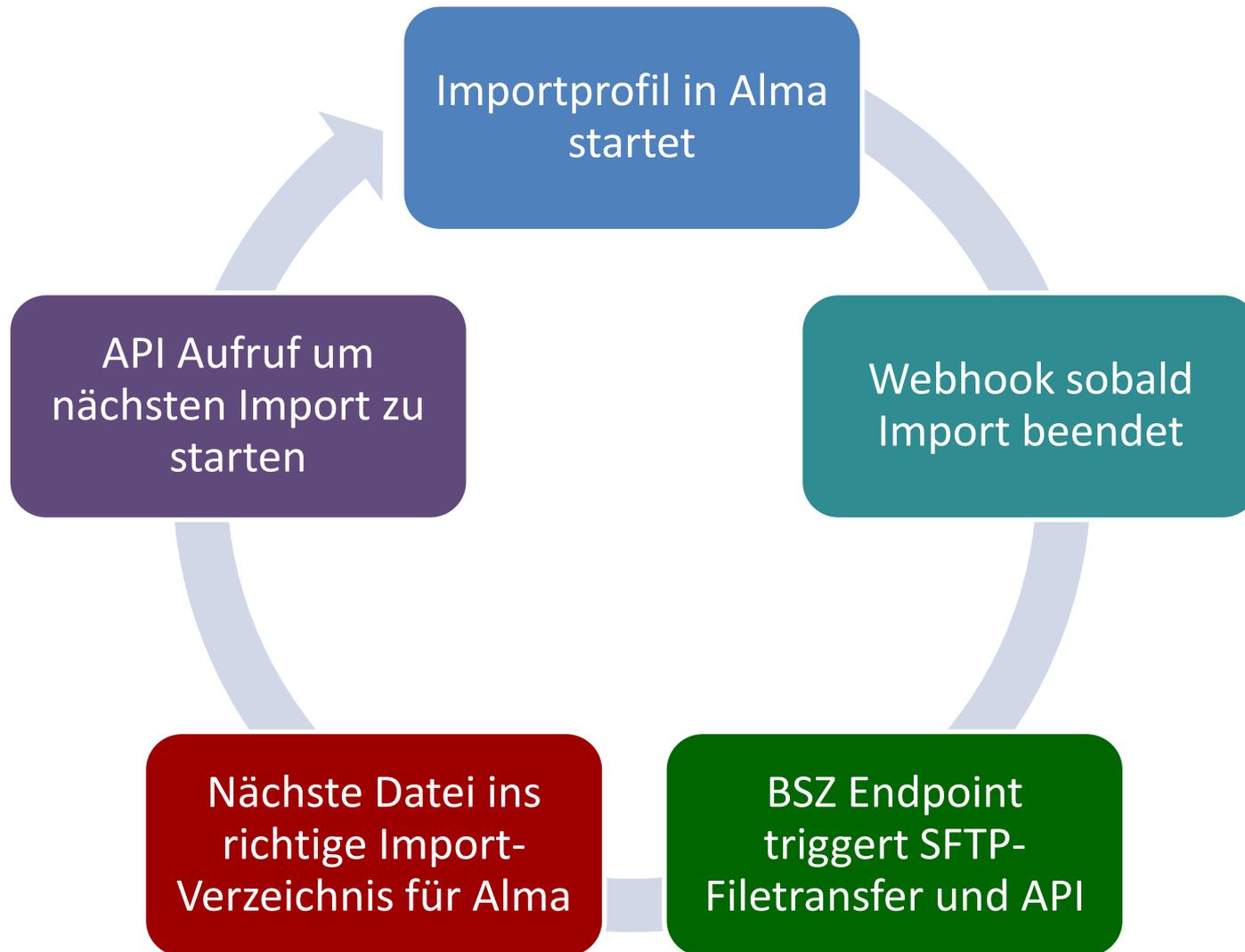
- Webhook meldet abgeschlossenen Import-Job

- Auszug:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<notification>
  <id>185575835580003231</id>
  <action>JOB_END</action>
  <time>2019-03-12T10:38:02.553Z</time>
  <job_instance link="/almaws/v1/conf/jobs/M0/instances/185575835580003231">
    <id>185575835580003231</id>
    <name>Metadata Import: Update BSZ data for k10plus</name>
    <submitted_by desc="Admin, BSZ">admin</submitted_by>
    <submit_time>2019-03-12T10:37:55.127Z</submit_time>
    <start_time>2019-03-12T10:37:55.294Z</start_time>
    <end_time>2019-03-12T10:38:02.553Z</end_time>
    <progress>100</progress><status desc="Completed Successfully">COMPLETED_SUCCESS</status>
    <status_date>2019-03-12Z</status_date>
```



# Die Umsetzung - Kreislauf



# Die Umsetzung - Webhooks

- Einsatz von Webhooks ist effizient und zeitsparend

Monitor and View Imports

Run Import

Back

1 - 20 of 7,628

All



Refresh



Filter : All Status filter : All

Job ID	Profile Name	Profile Type	Status	User	Time Started	Time Ended	Files Processed
1 185982350750003231	Update BSZ data for k10plus	Repository	In Process	exl_api	2019-03-24 00:26:21	-	15%
2 185840893250003231	Update BSZ data for k10plus	Repository	Completed Successfully	exl_api	2019-03-23 21:16:01	2019-03-24 00:26:03	100%
3 185745324200003231	Update BSZ data for k10plus	Repository	Completed Successfully	admin	2019-03-23 14:27:10	2019-03-23 21:15:42	100%

# Die Umsetzung - Vorkehrungen

- MARC-Validation bei Massenänderungen an Titeldaten vorab abschalten

 Alma » [Resolve Import Issues](#) [Monitor and View Imports](#) [Monitor Jobs](#) [Manage Sets](#) [Manage Import Profiles](#)     

All titles ▾ Abrufzeichen erw ▾ 🏠 🔍 Advanced ▾

Monitor and View Imports Run Import

1 - 20 of 7,679 All ▾ 🔍 Refresh 🔄

Filter : All ▾ Status filter : All ▾

↕ Job ID	↕ Profile Name	↕ Profile Type	↕ Status	↕ User	↕ Time Started	▼ Time Ended	Files Processed
1 188468167130003231	Update BSZ data for k10plus	Repository	Manual Handling required (still processing files)	exL_api	2019-03-30 15:23:47	-	70% <span>⋮</span> 
2 188195433200003231	Update BSZ data for k10plus	Repository	Completed Successfully	exL_api	2019-03-29 13:08:52	2019-03-30 15:23:34	100% <span>⋮</span> 
3 187956039350003231	Update BSZ data for k10plus	Repository	Completed Successfully	exL_api	2019-03-29 00:39:35	2019-03-29 13:08:40	100% <span>⋮</span> 

# Die Umsetzung - Vorkehrungen

- MARC-Validation bei Massenänderungen an Titeldaten vorab abschalten

 Alma » [Resolve Import Issues](#) [Monitor and View Imports](#) [Monitor Jobs](#) [Manage Sets](#) [Manage Import Profiles](#)     

All titles ▾ Abrufzeichen erw ▾ 🏠 🔍 Advanced ▾

Monitor and View Imports Run Import

1 - 20 of 7,679 All ▾ 🔍 Refresh 📄

↕ Job ID	↕ Profile Name	↕ Profile Type	↕ Status	↕ User	↕ Time Started	↕ Time Ended	Files Processed
1 <a href="#">18819543320000323</a>	Update BSZ data for k10plus	Repository	Manual Handling Required	exl_api	2019-03-29 13:08:52	-	99% <span>⋮</span>  <span>⋮</span>
2 <a href="#">187956039350003231</a>	Update BSZ data for k10plus	Repository	Completed Successfully	exl_api	2019-03-29 00:39:35	2019-03-29 13:08:40	100% <span>⋮</span>  <span>⋮</span>
3 <a href="#">187786830480003231</a>	Update BSZ data for k10plus	Repository	Completed Successfully	exl_api	2019-03-28 17:38:59	2019-03-29 00:39:16	100% <span>⋮</span>  <span>⋮</span>
4 <a href="#">187540145130003231</a>	Update BSZ data for k10plus	Repository	Completed Successfully	exl_api	2019-03-27 23:24:26	2019-03-28 17:38:44	100% <span>⋮</span>  <span>⋮</span>

# Die Umsetzung - Vorkehrungen

- MARC-Validation bei Massenänderungen an Titeldaten vorab abschalten

The screenshot displays the Alma library system interface. At the top, there is a navigation bar with the Alma logo and several menu items: 'Resolve Import Issues', 'Monitor and View Imports', 'Monitor Jobs', 'Manage Sets', and 'Manage Import Profiles'. Below this, the 'Imported Record Details' page is shown, featuring a 'Marc Record' tab and a table of record details. To the right, an 'Error List' is displayed, showing two error messages for field 100. On the far right, there is a 'Files Processed' section with a progress bar and a 'Run Import' button.

**Imported Record Details**

Field	Value
LDR	01151cam a2200313 4500
001	302784292
003	DE-576
005	20180727144238.0
007	tu
008	090210s2007 xx       00   ger c
035	__ a (DE-627)1372784292
035	__ a (DE-576)302784292
035	__ a (DE-599)BVBBV023011165
040	__ a DE-576  b ger  c DE-576  e rakwb
041	__ a ger  a eng
100	1_ a Dering, Florian
100	1_ a Eymold, Ursula
245	10  a Bier-Oktoberfest-Museum München  c Florian Dering; Ursula Eymold. [Hrsg.: Edith-Haberland-Wagner-Stiftung und Münchner

**Error List**

Field	Error Message
100	Multiple occurrences were found for non repeatable field 100
100 (2)	Multiple occurrences were found for non repeatable field 100

**Files Processed**

99%		...
100%		...
100%		...
100%		...

# Die Frage nach der Zeit



Completed  
Successfully

## Metadata Import: Update BSZ data for k10plus

<b>Process ID</b>	188195433200003231	<b>Started on</b>	29/03/2019 13:08:52 CET
<b>Finished on</b>	30/03/2019 15:23:34 CET	<b>Total run time</b>	26 Hours 14 Minutes 42 Seconds
<b>Status</b>	Completed Successfully	<b>Status date</b>	30/03/2019 15:23:34 CET
<b>Files Processed</b>	18635	<b>Records with exceptions</b>	0

### Counters

Total records processed **931738**

→ Die Ladedauer ist unabhängig von der Anzahl der in einer Datei enthaltenen Titel

# Die Frage nach der Zeit



Completed  
Successfully

## Metadata Import: Update BSZ data for k10plus

<b>Process ID</b>	188195433200003231	<b>Started on</b>	29/03/2019 13:08:52 CET
<b>Finished on</b>	30/03/2019 15:23:34 CET	<b>Total run time</b>	26 Hours 14 Minutes 42 Seconds
<b>Status</b>	Completed Successfully	<b>Status date</b>	30/03/2019 15:23:34 CET
<b>Files Processed</b>	18635	<b>Records with exceptions</b>	0

### Counters

Total records processed **931738**



Completed  
Successfully

## Metadata Import: Update BSZ data for k10plus

<b>Process ID</b>	188856964500003231	<b>Started on</b>	31/03/2019 08:47:59 CEST
<b>Finished on</b>	31/03/2019 16:14:47 CEST	<b>Total run time</b>	7 Hours 26 Minutes 48 Seconds
<b>Status</b>	Completed Successfully	<b>Status date</b>	31/03/2019 16:14:47 CEST
<b>Files Processed</b>	18688	<b>Records with exceptions</b>	0

### Counters

Total records processed **934364**

# Die Frage nach der Zeit



**Completed  
Successfully**

## Metadata Import: Update BSZ data for k10plus

<b>Process ID</b>	188195433200003231	<b>Started on</b>	29/03/2019 13:08:52 CET
<b>Finished on</b>	30/03/2019 15:23:34 CET	<b>Total run time</b>	26 Hours 14 Minutes 42 Seconds
<b>Status</b>	Completed Successfully	<b>Status date</b>	30/03/2019 15:23:34 CET
<b>Files Processed</b>	18635	<b>Records with exceptions</b>	0

### Counters

Total records processed **931738**



**Completed  
Successfully**

## Metadata Import: Update BSZ data for k10plus

<b>Process ID</b>	190450855420003231	<b>Started on</b>	03/04/2019 23:32:53 CEST
<b>Finished on</b>	05/04/2019 24:32:24 CEST	<b>Total run time</b>	24 Hours 59 Minutes 31 Seconds
<b>Status</b>	Completed Successfully	<b>Status date</b>	05/04/2019 24:32:24 CEST
<b>Files Processed</b>	9001	<b>Records with exceptions</b>	0

### Counters

Total records processed **450009**

# Die Frage nach der Zeit



## Metadata Import: Update BSZ data for k10plus

Completed  
Successfully

**Process ID** 191316549570003231  
**Finished on** 06/04/2019 19:30:08 CEST  
**Status** Completed Successfully  
**Files Processed** 17982

**Started on** 06/04/2019 13:37:03 CEST  
**Total run time** 5 Hours 53 Minutes 5 Seconds  
**Status date** 06/04/2019 19:30:08 CEST  
**Records with exceptions** 0

### Counters

Total records processed **899061**



## Metadata Import: Update BSZ data for k10plus

Completed  
Successfully

**Process ID** 190450855420003231  
**Finished on** 05/04/2019 24:32:24 CEST  
**Status** Completed Successfully  
**Files Processed** 9001

**Started on** 03/04/2019 23:32:53 CEST  
**Total run time** 24 Hours 59 Minutes 31 Seconds  
**Status date** 05/04/2019 24:32:24 CEST  
**Records with exceptions** 0

### Counters

Total records processed **450009**

# Fazit

- Positiv:
  - Automatisierung hat sehr gut geklappt
  - Keine nennenswerten Auswirkungen auf die drei angeschlossenen Institution Zones
- Negativ:
  - Konkrete Zeitplanung war nicht möglich
  - Ladezeit zu lange
  - Gesamtmenge aller Daten aus K10plus undenkbar

**Vielen Dank für Ihre Aufmerksamkeit !**

**Fragen?**

---

**BSZ  GBV**

Michaela.Sieber@bsz-bw.de

---