

# **Gemeinsamer Verbündeindex**

## **Sachstand und (Erfahrungs-)bericht von drei unterschiedlichen Frontends**

**Bibliothekskongress 2019  
in Leipzig**

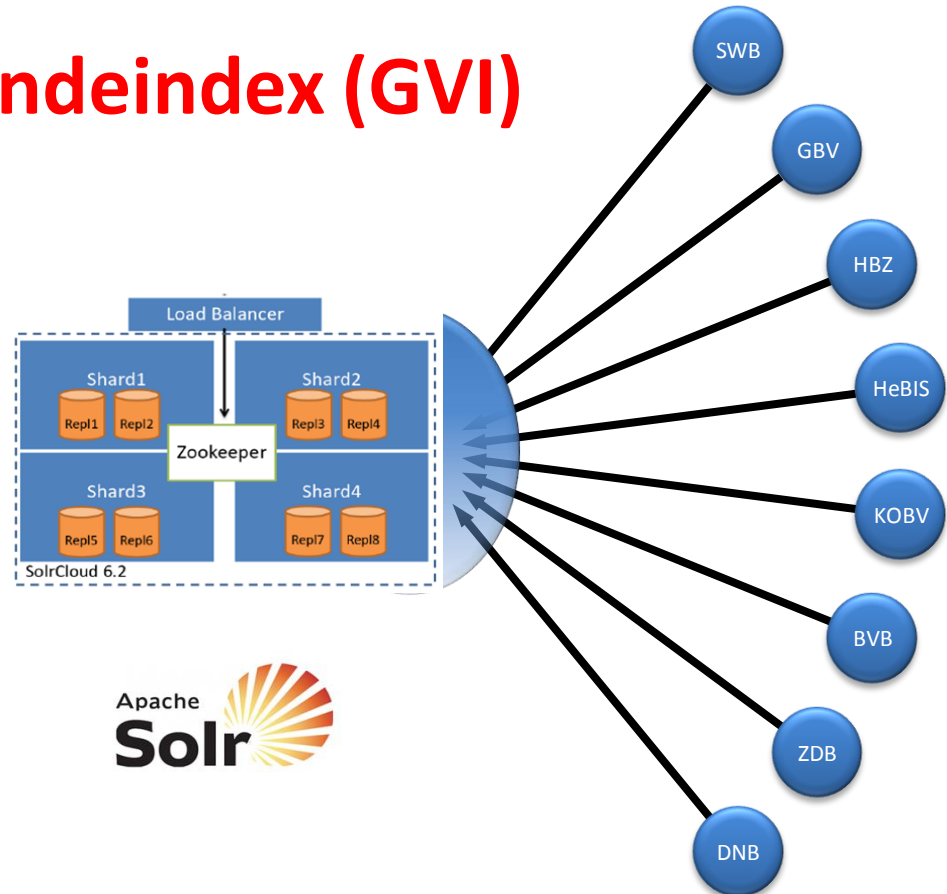
Stefan Winkler, BSZ  
Stefan Lohrum, KOBV  
Uwe Reh, HeBIS

- Einführung GVI
- Deduplizierung / Gruppierung / Matchkeys
- Einsatz des GVI in den Verbänden
- Schnittstellen, Nachnutzung / Ausblick

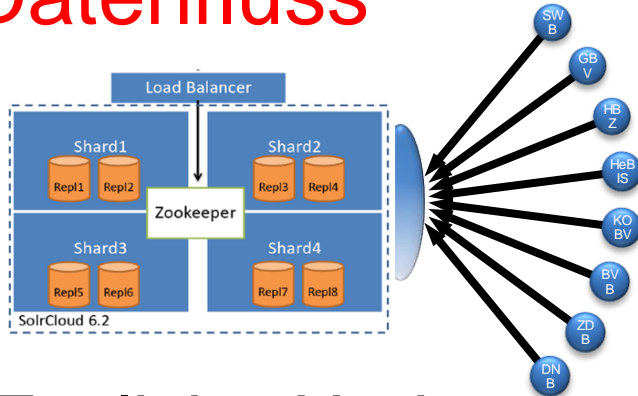


# Gemeinsamer Verbändeindex (GVI)

- 170 Mio. Titel in einem zentralen Index
- Tägliche Updates



# Datenfluss



Tägliche Updates  
Standard-Prozess

# Prozesssteuerung

The screenshot shows the Jenkins web interface. The top navigation bar includes the Jenkins logo, a search bar, and the user name 'Stefan Lohrum'. The main content area displays a table of jobs with columns for status, name, last success, last failure, and last duration.

| S | W | Name ↓                     | Letzter Erfolg          | Letzter Fehlschlag    | Letzte Dauer          |
|---|---|----------------------------|-------------------------|-----------------------|-----------------------|
| ● | ☁ | Full GVI Cache Data Export | 14 Tage - #28           | 20 Tage - #26         | 2 Tage 5 Stunden      |
| ● | ☀ | Get Solr Index from BSZ    | 1 Jahr 2 Monate - #9    | 1 Jahr 2 Monate - #10 | 4 Stunden 4 Minuten   |
| ● | ☀ | GVI Cache Data Export      | 2 Monate 16 Tage - #787 | Unbekannt             | 2 Minuten 15 Sekunden |
| ● | ☀ | GVI Data Transfer to BSZ   | 2 Monate 16 Tage - #581 | Unbekannt             | 3.3 Sekunden          |
| ● | ☀ | gvi_get_bsz_backup         | 1 Jahr 2 Monate - #14   | 1 Jahr 2 Monate - #8  | 4 Stunden 35 Minuten  |

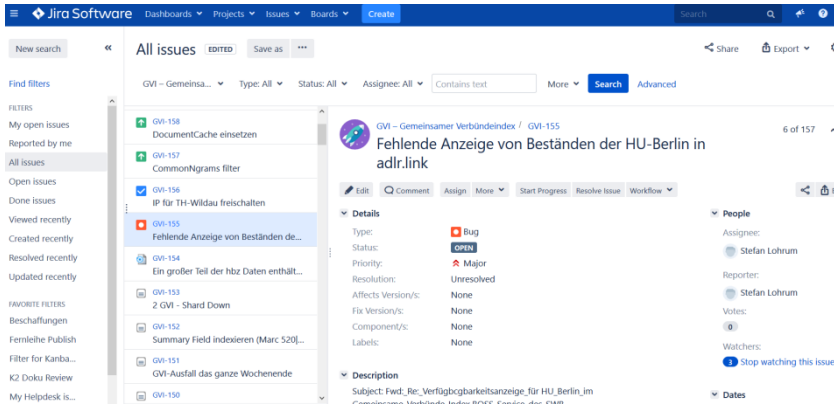
# Monitoring

<https://status.kobv.de/>



# Betriebsorganisation

## Ticketsystem



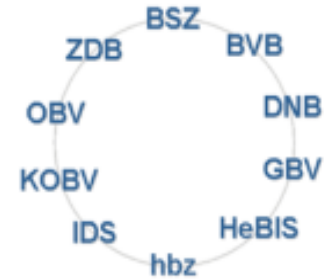
## Team

| Name                 | Jira/Confluence      | E-Mail                               | Verbund |
|----------------------|----------------------|--------------------------------------|---------|
| Alexey Sharenkov     | asharenk             | sharenkov@zib.de                     | KOBV    |
| Bettina Sunckel      | bettina.sunckel      | bsunckel@hebis.uni-frankfurt.de      | HEBIS   |
| Christian Strohmaier | christian.strohmaier | Christian.Strohmaier@bsb-muenchen.de | BVB     |
| Christine Baron      | christine.baron      | baron@hbz-nrw.de                     | HBZ     |
| Cornelius Amzar      | cornelius.amzar      | cornelius.amzar@bsz-bw.de            | BSZ     |
| Elmar Schackmann     | elmarschackmann      | schackmann@hbz-nrw.de                | HBZ     |
| Gerald Steilen       | gerald.steilen       | Gerald.Steilen@gbv.de                | GBV     |
| Matthias Groß        | magross              | gross@bsb-muenchen.de                | BVB     |
| Heiko Jansen         | heiko.jansen         | jansen@hbz-nrw.de                    | HBZ     |
| Stefan Lohrum        | slohrum              | lohrum@zib.de                        | KOBV    |
| Stefan Winkler       | stefan.winkler       | stefan.winkler@bsz-bw.de             | BSZ     |
| Till Kinstler        | till.kinstler        | kinstler@gbv.de                      | GBV     |
| Thomas Kirchhoff     | thomas.kirchhoff     | thomas.kirchhoff@bsz-bw.de           | BSZ     |
| Uwe Reh              | uwe.reh              | reh@hebis.uni-frankfurt.de           | HEBIS   |
| Uwe Risch            | uwe.risch            | risch@hebis.uni-frankfurt.de         | HEBIS   |
| Viktoria Gerlach     | vschuber             | gerlach@zib.de                       | KOBV    |
| Volker Conradt       | volker.conradt       | volker.conradt@bsz-bw.de             | BSZ     |

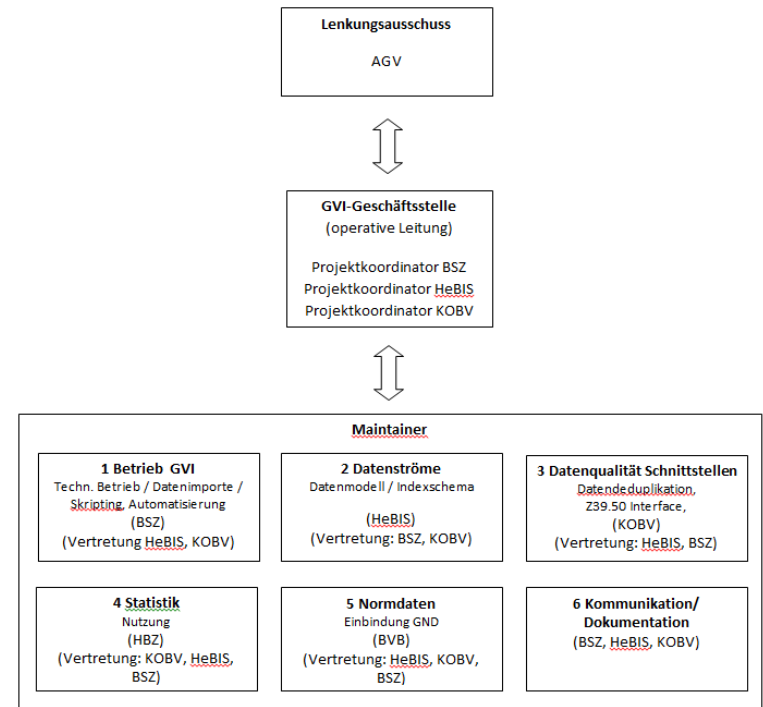
## Dokumentation im Wiki



# Betriebsvereinbarung der AG der Verbundsysteme



- Datenlieferungen
- Support
- Geschäftsstelle
- Weiterentwicklung
- Technische Realisierung und Betriebsgrundlagen
- Zugriff und Zugang



# Dubletten

# Dubletten

**HARRY  
POTTER**



# Dubletten

## HARRY POTTER

Sechs Verbände + DNB

# Dubletten

**HARRY  
POTTER**

Sechs Verbände + DNB



<https://gratisography.com/photo/scared-man/>

## Unterstützung der Deduplizierung im GVI

- [https://lucene.apache.org/solr/guide/7\\_1/result-grouping.html](https://lucene.apache.org/solr/guide/7_1/result-grouping.html)
- Suchfelder für die Aufnahme von Cluster-Kennungen externer Projekte zur Deduplizierung.

## Unterstützung der Deduplizierung im GVI

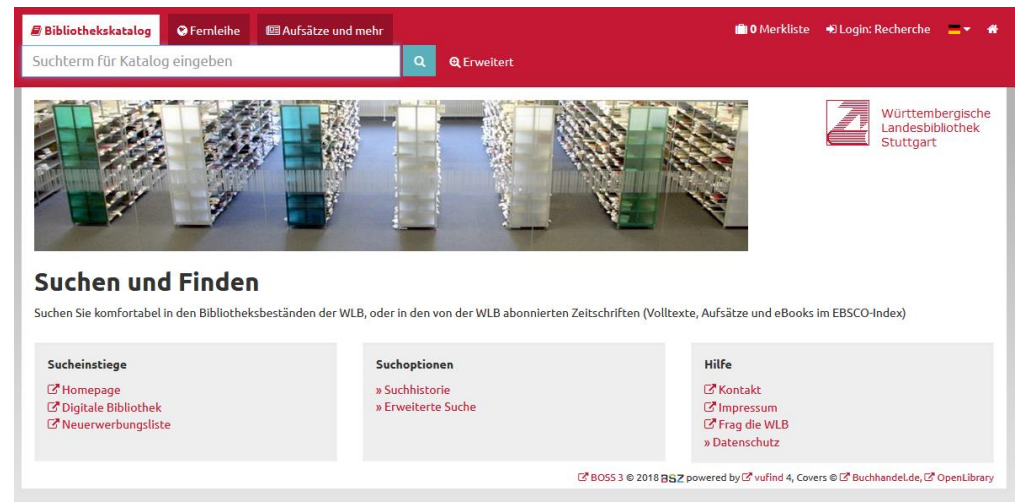
- [https://lucene.apache.org/solr/guide/7\\_1/result-grouping.html](https://lucene.apache.org/solr/guide/7_1/result-grouping.html)
- Suchfelder für die Aufnahme von Cluster-Kennungen externer Projekte zur Deduplizierung.
- Eigene „Matchkey“ Felder für die Anzeige und für Kandidatenlisten.

# Einsatz des GVI im SWB

- **BSZ**
  - Betrieb der GVI-SolrCloud
  - Fernleihportal (32 Bibliotheken)
  - BOSS (25 Bibliotheken, Discovery & Fernleihe)
  - DESI (Anschaffungsvorschläge für 20 Bibliotheken)
- **asTec**
  - aDIS/BMS (35 Lokalsysteme, Fernleihe)
- **PiciBird**
  - Hybrid Bookshelf (KIM Konstanz, HS Aalen)
- **UB Freiburg**
  - KatalogPlus (5 UBs, Fernleihe)
  - REDI (Fernleihe, im Test)
- **SLUB Dresden**
  - Datenmanagement
- **UB Mannheim**
  - Linked Open Data
- **UB Tübingen**
- **FH Furtwangen**
  - Literatursuche

## BSZ - One Stop Search (BOSS)

- Discovery System des BSZ
- Open Source (VuFind)
- Literaturrecherche und -beschaffung im SWB
- Diverse Suchräume:
  - Artikelindices (EDS, Summon, Primo)
  - GVI-Nutzung für
    - Katalogsuche
    - Fernleihe
  - eBook-Indices (PDA)
  - K10Plus-Zentral
  - Finc (demnächst)



# Dubletten in der Fernleihe

The screenshot shows a search results page for 'java profi'. The search criteria are 'java profi' and 'Erweitert'. The results are sorted by 'Relevanz' and show 10 hits out of 204. The first six results are duplicates of the same book: 'Der Weg zum Java-Profi: Konzepte und Techniken für die professionelle Java-Entwicklung' by Inden, Michael (Q, GND-ID), published by Heidelberg dpunkt.verlag in 2018. The entries are numbered 1 through 6. Each entry includes a book cover, a title, author, publisher, and edition information. The 6th entry is marked as an 'E-Book'.

- Allein schon aus dem Zusammenspielen aller Verbundkataloge bedingt
- Doppelung bei Zeitschriften durch die ZDB-Daten
- Doppelungen durch Nachnutzung von GBV/BVB-Daten im KOBV-Abzug
- Teilweise auch Dubletten in einem Katalog (sollte nicht sein)
- Das gleiche eBook mehrfach aber von verschiedenen Anbietern erfasst?!

# Dubletten-Gruppierung im Fernleihportal / BOSS

Button zum Auf/Zuklappen der „Dubletten“

Anzahl der „Dubletten“

Checkbox „Treffer gruppieren“

Suche: java profi / Erweiterte Suche bearbeiten

Treffer 1 - 10 von 86 für Suche java profi, Suchdauer: 0,05s

Sortieren Relevanz

Treffer gruppieren

Suche einschränken

1 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverlag, 2016 Auflage: 4., überarbeitete und aktualisierte Auflage **Dubletten 4**

2 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunkt 2017 Auflage: 4., aktualisierte Auflage **Dubletten 2**

3 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung. Aktuell zu Java 9.** von **Inden, Michael** Veröffentlicht: Erscheinungsort nicht ermittelbar dpunkt.verlag 2017; Weisbaden dtvbw GmbH **Dubletten 2**

4 **Der Java-Profi: Persistenzlösungen und REST-Services : Datenaustauschformate, Datenbankanwicklung und verteilte Anwendungen** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverlag, 2016 Auflage: 1. Auflage **Dubletten 2**

5 **Der Java-Profi: Persistenzlösungen und REST-Services : Datenaustauschformate, Datenbankanwicklung und verteilte Anwendungen** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverlag, 2016 Auflage: 1. Auflage **Dubletten 2**

6 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverl., 2015 Auflage: 3., aktualisierte u. überarb. Auf. **Dubletten 7**

Masterrecords aus dem SWB

Suche: java profi / Erweiterte Suche bearbeiten

Treffer 1 - 10 von 86 für Suche java profi, Suchdauer: 0,11s

Sortieren Relevanz

Treffer gruppieren

Suche einschränken

1 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverlag, 2016 Auflage: 4., überarbeitete und aktualisierte Auflage **Dubletten 4**

2 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverlag 2018 Auflage: 4., überarbeitete und aktualisierte Auflage **Dubletten 2**

3 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverlag 2018 Auflage: 4., überarbeitete und aktualisierte Auflage **Dubletten 2**

4 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverlag 2018 Auflage: 4., überarbeitete und aktualisierte Auflage **Dubletten 2**

5 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverlag 2018 Auflage: 4., überarbeitete und aktualisierte Auflage **Dubletten 2**

6 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverlag 2018 Auflage: 4., überarbeitete und aktualisierte Auflage **Dubletten 2**

7 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverlag 2018 Auflage: 4., überarbeitete und aktualisierte Auflage **Dubletten 2**

8 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverlag 2018 Auflage: 4., überarbeitete und aktualisierte Auflage **Dubletten 2**

9 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverlag 2018 Auflage: 4., überarbeitete und aktualisierte Auflage **Dubletten 2**

10 **Der Weg zum Java-Profi : Konzepte und Techniken für die professionelle Java-Entwicklung** von **Inden, Michael (Q GND-ID)** Veröffentlicht: Heidelberg dpunktverlag 2018 Auflage: 4., überarbeitete und aktualisierte Auflage **Dubletten 2**

Masterrecord aus dem SWB

Werk mit gleichem Matchkey

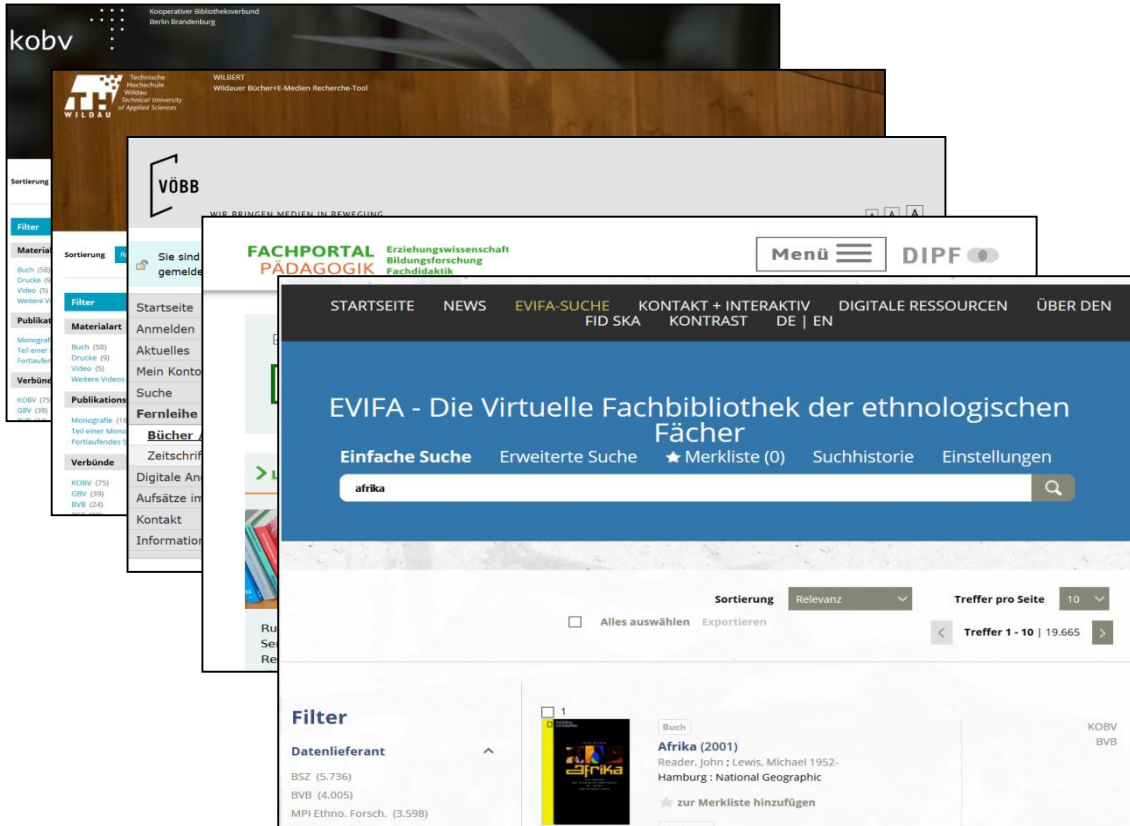
Werk mit verschiedenen Matchkeys



## Bewertung des Groupings

- Benutzer lieben es
- Auf den ersten Blick kaum Fehler
- Gruppierung ist klare Usability-Verbesserung
- Verfahren ist robust und schnell
- Jeder neue Record wird sofort mitgruppiert
- Leicht zu tunen (schnelle turnarounds)
- Fazit: Zur Nachnutzung empfohlen!

# Einsatz des GVI im KOBV



## Discovery Service ALBERT

- KOBV-Portal
- TH-Wildau
- EUV Frankfurt

## aDIS/BMS

- VÖBB-Portal

## Sisis InfoGuite

- FH-Potsdam

## Fachportale / FIDs

- Fachportal Pädagogik
- EVIFA

# Einsatz des GVI im KOBV

## Hintergrunddienst für die Fernleihe

- => Nachrecherche
- => Ausgehend von einem Treffer Bestand in den anderen Verbänden finden

## Datenbereitstellung für

- => Fachportale anhand fachlicher Kriterien (Klassifikationen)
- => Culturegraph

## Einsatz des GVI in HeBIS

- <https://www.portal-hebis.de>

## Einsatz des GVI im BVB

- Gateway Bayern
- Zentraler Fernleihserver (Gruppierung)

## Einsatz des GVI im HBZ

- Fernleihportal in Planung

# Schnittstellen: SOLR/JSON

**http://.../solr/GVI/select?indent=on&q=title:De-Duplication&wt=json**

|   |   |
|---|---|
| <pre>{   "responseHeader":{     "zkConnected":true,     "status":0,     "QTime":65,     "params":{       "q":"title:De-Duplication",       "df":"allfields",       "indent":"on",       "echoParams":"all",       "spellcheck":"false",       "collection":"GVI",       "rows":"10",       "wt":"json"}},     "response":{"numFound":136,       "start":0,       "maxScore":22.661436,       "docs":[         {</pre> | <pre>"fullrecord":"&lt;?xml version=\"1.0\" encoding=\"UTF-8\"?&gt;   &lt;collection xmlns=\"http://www.loc.gov/MARC21/slim\"&gt;     &lt;record&gt;       &lt;leader&gt;01139cam a22003012c 4500&lt;/leader&gt;       &lt;controlfield tag=\"001\"&gt;291450660&lt;/controlfield&gt;       &lt;controlfield tag=\"003\"&gt;DE-603&lt;/controlfield&gt;       &lt;controlfield tag=\"005\"&gt;20170220181703.0&lt;/controlfield&gt;       &lt;controlfield tag=\"006\"&gt;m o u00 u&lt;/controlfield&gt;       &lt;controlfield tag=\"007\"&gt;cr nnn &lt;/controlfield&gt;       &lt;controlfield tag=\"008\"&gt;120314s2008 xx o u00 u eng c&lt;/controlfield&gt;       &lt;datafield tag=\"020\" ind1=\"\" ind2=\"\"&gt;         &lt;subfield code=\"a\"&gt;0738432180&lt;/subfield&gt;         &lt;subfield code=\"9\"&gt;0-7384-3218-0&lt;/subfield&gt;       &lt;/datafield&gt;       .....     &lt;/datafield&gt;   &lt;/record&gt; &lt;/collection&gt;\n",   "id":"(DE-603)291450660",   "_version_":1566479158434856960,   "recordtype":"gvimarc",   "index_date":"2017-05-04T15:15:39.973Z"},]   ... }}</pre> |
|---|---|

# Schnittstellen: SOLR/XML

<http://.../solr/GVI/select?indent=on&q=title:De-Duplication&wt=xml>

```
<?xml version="1.0" encoding="UTF-8"?>
<response>
<lst name="responseHeader">
<bool name="zkConnected">true</bool>
<int name="status">0</int>
<int name="QTime">9</int>
<lst name="params">
<str name="q">title:De-Duplication</str>
<str name="df">allfields</str>
<str name="indent">on</str>
<str name="echoParams">all</str>
<str name="spellcheck">>false</str>
<str name="collection">GVI</str>
<str name="rows">10</str>
<str name="wt">xml</str>
<str name="_">1495673969082</str></lst>
</lst>
```

```
<doc>
<str name="fullrecord">
&lt;?xml version="1.0" encoding="UTF-8"?&gt;;&lt;collection
  xmlns="http://www.loc.gov/MARC21/slim"&gt;
  &lt;record&gt;
  &lt;leader&gt;01139cam a22003012c 4500
  &lt;/leader&gt;&lt;controlfield tag="001,"
  &gt; ...
</str>

<str name="id">(DE-603)291450660</str>
<long name="_version_">1566479158434856960</long>
<str name="recordtype">gvimarc</str>
<date name="index_date">2017-05-04T15:15:39.973Z</date>
</doc>

<doc>
...
</result>
```

# Schnittstellen: Z39.50 (Metaproxy)

```
Z> open ...../gvi
Connecting...OK.
Sent initrequest.
Connection accepted by v3 target.
ID : 81
Name : Metaproxy/YAZ
Version: 1.11.9/5.21.1 45b3c27d3df070
5d9e86f8af0651c359f5a31171
Options: search present
Elapsed: 0.000000

Z> find @attr 1=4 De-Duplication
Sent searchRequest.
Received SearchResponse.
Search was a success.
Number of hits: 136
records returned: 0
Elapsed: 0.062400

Z> show 1
Sent presentRequest (1+1).
Records: 1
[gvi]Record type: USmarc
01120cam a22002892c 4500
001 (DE-603)291450660
003 DE-603
```

```
005 20170220181703.0
006 m o u00 u
007 cr nnn
008 120314s2008 xx o u00 u eng c
020 $a 0738432180 $9 0-7384-3218-0
020 $a 9780738432182 $9 978-0-7384-3218-2
035 $a (DE-599)BVBBV037354916
035 $a (OCoLC)288593546
040 $a DE-603 $b ger $c DE-603 $d DE-603 $e rakwb
041 $a eng
245 00 $a Guide to data de-duplication $h [Elektronische Ressource]
$b the IBM system storage TS7650G protecTIER de-duplication
gateway
$c Babette Haeusser ... [et al.]
264 1 $a [United States?] $b IBM International Technical Support
Organization $c 2008
300 $a ix, 290 p S.
490 0 $a IBM redbooks
500 $a Title from title screen. - "SG24-7652-00."
530 $a Online-Ausg.:
533 $b Sebastopol, CA [u.a.] $c Safari Books Online
$e Online-Ressource $f Safari Tech Books Online $n [Online-Ausg.]
655 7 $a Online-Publikation $2 gnd
700 1 $a Haeusser, Babette
856 $u http://proquest.safaribooksonline.com/0738432180 $x Verlag
```

# Schnittstellen: SRU (Meataproxy)

**http://.../gvi?version=1.1&operation=searchRetrieve&query=bauhaus&startRecord=1&maximumRecords=10&recordSchema=dc&recordPacking=string&stylesheet=**

Record Schema: MARCXML, Dublin Core

```
<zs:searchRetrieveResponse><zs:version>1.1</zs:version>
<zs:numberOfRecords>38910</zs:numberOfRecords><zs:records>
<zs:record><zs:recordSchema>dc</zs:recordSchema><zs:recordPacking>string</zs:recordPacking>
<zs:recordData><?xml version="1.0"?>
<dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>Das Bauhaus dargestellt anhand der Sammlung des Bauhaus-Archivs</dc:title>
  <dc:creator> Jaeggi, Annemarie Sonstige oth </dc:creator>
  <dc:creator> Bauhaus-Archiv &lt;Berlin&gt; Sonstige oth </dc:creator>
  <dc:type>software, multimedia</dc:type>
  <dc:type> Bauhaus-Archiv gnd </dc:type>
  <dc:publisher>Berlin Bauhaus-Archiv</dc:publisher>
  <dc:date>[2009]</dc:date>
  <dc:language>und</dc:language>
  <dc:subject>Bauhaus</dc:subject>
</dc:dc>
</zs:recordData>
<zs:recordPosition>1</zs:recordPosition></zs:record> .....
```



# Nachnutzung

- Ist gerne gesehen!
- Freischaltung erfolgt zügig
- Kostenlos!
- Melden bei [gvi-info@agv-gvi.de](mailto:gvi-info@agv-gvi.de)
- Code ist Open Source bei github
  - <https://github.com/gemeinsamerverbuendeindex/>
  - <https://github.com/BSZBW/boss/tree/develop>
- Bugtracking-System der GVI-Entwickler
  - Mitarbeit erwünscht!
  - <https://tickets.zib.de/jira/projects/GVI/summary>

## Ausblick 2019

- Verbundmigration von SWB & GBV zu K10plus bis 1.4.
- Hardwareausbau (3 neue Server, Tests mit VMs)
- Verbessertes Monitoring
- Hochverfügbarkeit
  - Kleinere Shards, mehr Replicas
  - Verbessertes Caching
  - Verbesserte Proxy Architektur
  - Spiegel in KOBV und HeBIS
- Neue Daten (Österreich)
- Interface für Primo

## Vielen Dank für die Aufmerksamkeit!

- Kontakt:
  - Stefan Winkler (BSZ)
  - Cornelius Amzar (BSZ)
  - Thomas Kirchhof (BSZ)
  - Uwe Reh (HeBIS)
  - Stefan Lohrum (KOBV)