

# DSpace über die DSpace-API nutzen

Ein Werkstattbericht



DSpace Anwendertreffen in Bamberg 11.04.2019

Roland Keck

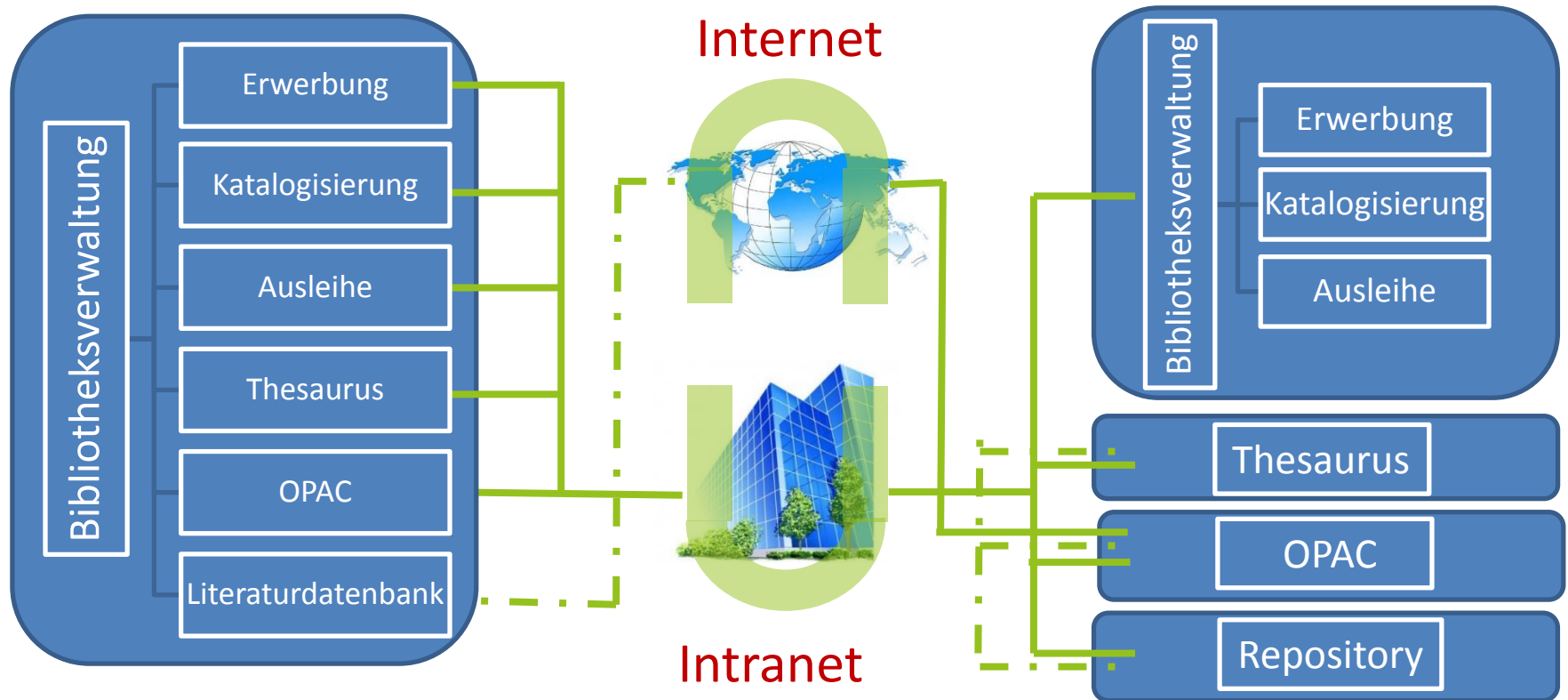
## Das Bundesinstitut für Berufsbildung (BIBB) ...

... ist DAS anerkannte Kompetenzzentrum zur Erforschung und Weiterentwicklung der beruflichen Aus- und Weiterbildung in Deutschland. Es führt seine Aufgaben im Rahmen der Bildungspolitik der Bundesregierung durch.

Kennzeichnend für die Arbeit des BIBB sind:

- enge Zusammenarbeit mit Bund, Ländern und Sozialpartnern
- Förderung von Innovationen in der nationalen und internationalen Berufsbildung
- Entwicklung neuer, praxisorientierte Lösungsvorschläge für die berufliche Aus- und Weiterbildung
- durch eigene wissenschaftliche Forschung zur Berufsbildungsforschung beitragen

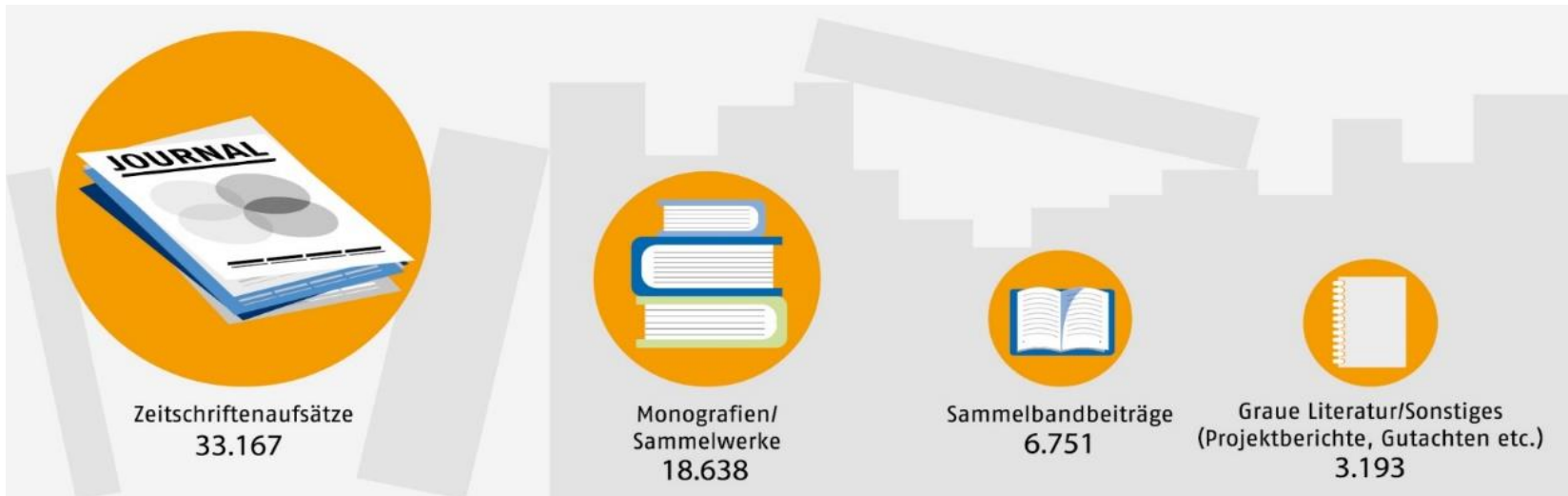
# Ausgangslage und Zielsetzung



# Von der Literaturdatenbank Berufliche Bildung (LDBB) zum VET Repository

- einzige Literaturdatenbank zur Berufsbildungsforschung in Deutschland
- verbindet Aktualität der Nachweise mit hochwertiger Auswertung
- Umfang: ca. 62.000 Datensätze ab Erscheinungsjahr 1988; Zugang: ca. 1.000/Jahr
- Anteil Online-Publikationen (Open Access): 43% ab 2009
- Hoher Anteil an hierarchisch strukturierten Informationen

# Ausgangsbasis: die LDBB-Daten



## Besonderheiten

- Mehrstufige Hierarchien
- mehrbändige Werke
- Strukturierungen

Spezielle Datenelemente, die für die Produktion von Bibliografien und anderen Produkten erforderlich sind!

# Wie kommen die LDBB-Daten ins Repository (DSpace)?

- Manuell?

Hmmm!?!?!

- AIP (Archival Information Package)?

`/bin/dspace packager -s -a -t AIP -e <eperson> -p <parent-handle> <file-path>`

- Simple Archive Format?

```
archive_directory/
item_001
  Dublin_Core.xml
  metadata_1prefix.xml
  contents
  collections
file_1.doc
file_2.pdf
item_001
  Dublin_Core.xml
  contents
file_1.png
...
```

archive\_directory/  
item\_001  
Dublin\_Core.xml -- qualified Dublin Core metadata for metadata fields belonging to the 'dc' schema.  
metadata\_1prefix.xml -- metadata in another schema. The prefix is the name of the schema as registered with the metadata registry.  
contents -- text file containing one line per filename.  
collections -- text file that contains the handles of the collections the item will belong to. Optional. Each handle in a row. Collection in first line will be the owning collection.  
file\_1.doc -- file to be added as bitstreams to the item.  
file\_2.pdf  
item\_001  
Dublin\_Core.xml  
contents  
file\_1.png  
...

- REST-API

# Wie kommen die LDBB-Daten ins Repository (DSpace)?

- Manuell?

Hmmm!?!?!?

- AIP (Archival Information Package)?

```
/bin/dspace packager -s -a -t AIP -e <eperson> -p <parent-handle> <file-path>
```

- Simple Archive Format?

```
archive_directory/  
  item_000/  
    dublin_core.xml      -- qualified Dublin Core metadata for metadata fields belonging to the 'dc' schema.  
    metadata_[prefix].xml -- metadata in another schema. The prefix is the name of the schema as registered with the  
                           metadata registry.  
    contents             -- text file containing one line per filename.  
    collections         -- text file that contains the handles of the collections the item will belong to. Optional. Each  
                           handle in a row..Collection in first line will be the owning collection.  
    file_1.doc           -- files to be added as bitstreams to the item.  
    file_2.pdf  
  item_001/  
    dublin_core.xml  
    contents  
    file_1.png  
    ...
```

# Wie kommen die LDBB-Daten ins Repository (DSpace)?

- Manuell?

Hmmm!?!?!?

- AIP (Archival Information Package)?

`/bin/dspace packager -s -a -t AIP -e <eperson> -p <parent-handle> <file-path>`

- Simple Archive Format?

```
archive_directory/
item_001/
  dublin_core.xml
  metadata_prefix.xml
  comments
  collections
  file_1.doc
  file_2.pdf
  item_001/
  dublin_core.xml
  comments
  file_1.png
  ...
```

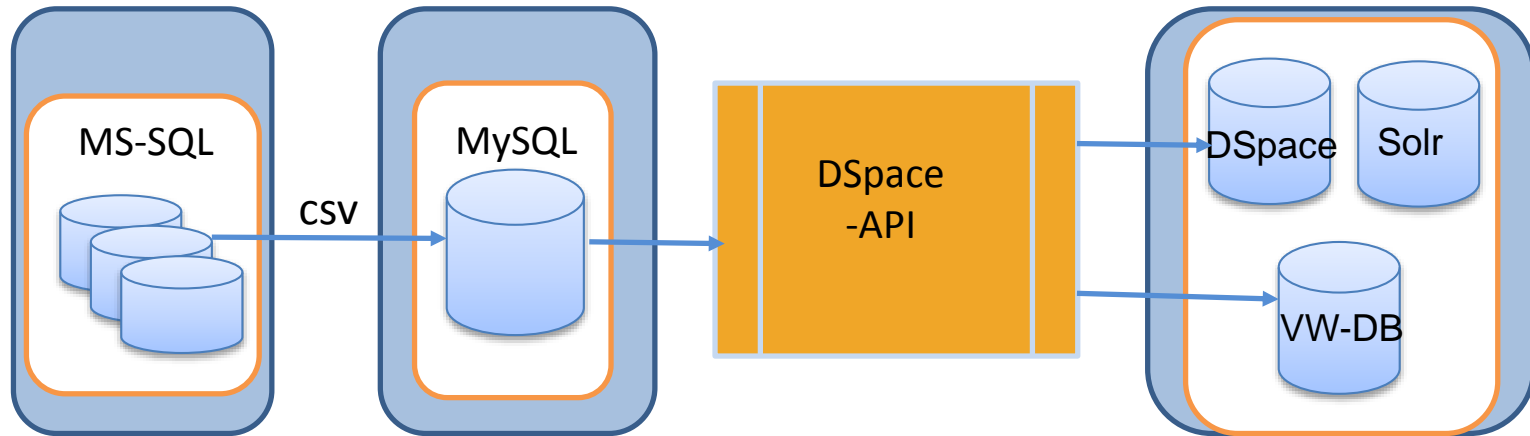
-- qualified Dublin Core metadata for metadata fields belonging to the 'dc' schema.  
-- metadata in another schema. The prefix is the name of the schema as registered with the metadata registry.  
-- test file containing one line per filename.  
-- test file that contains the handles of the collections the item will belong to. Optional. Each handle in a row. Collection in first line will be the owning collection.  
-- files to be added as keywords to the item.

- REST-API

- Anforderungen an Hierarchien abbildbar (Zwischenspeicherung)
- Eigene Datenfelder
- Flexibilität
- Freie Auswahl Programmiersprache 😊



# Vorgehen



Produktion alt	Konvertierung	DSpace API	DSpace -Produktion
<ul style="list-style-type: none"> <li>Entladen relevante Tabellen (MAB-ähnliches Format)</li> </ul>	<ul style="list-style-type: none"> <li>Konvertierung nach Qualified DC(++)</li> </ul>	<ul style="list-style-type: none"> <li>Generiert Hierarchien on the fly</li> <li>Ingest der Metadaten</li> <li>Speist Verwaltungs-DB</li> </ul>	<ul style="list-style-type: none"> <li>VW-DB hält Konkordanz Altsystem &lt;-&gt; DSpace (ID und UUIDS)</li> <li>Relevant für Updates und nachträgliche Anreicherung mit Bitstreams</li> </ul>

# DSpace REST-API

- „Representational State Transfer“: Programmierparadigma für verteilte Systeme (Wikipedia)
  - Client-Server
  - Zustandslos
  - Caching
- „Befehle“ HTTP basiert
  - GET
  - POST
  - PUT
  - DELETE
- DSpace-API unterstützt gesamten DSpace-Funktionsumfang hinsichtlich Items, Bitstreams, Collections, Communities und Authentication

## Login => SeesionID

```
dspace_login($ch) {  
    $login_data = array( 'email'      => 'email@domain',  
                        'password'   => 'passwd');  
    curl_setopt($ch, CURLOPT_URL, 'https://dspace.de/rest/login');  
    curl_setopt($ch, CURLOPT_POST, 1);  
    curl_setopt($ch, CURLOPT_RETURNTRANSFER, 1);  
    curl_setopt($ch, CURLOPT_HEADER, 1);  
    curl_setopt($ch, CURLOPT_HTTPHEADER, array('Accept: application/json'));  
    curl_setopt($ch, CURLOPT_POSTFIELDS, http_build_query($login_data));  
    $result = curl_exec($ch);  
    if ($result === false)    return (0);  
    $sid = curl_multi_getcontent ($ch);  
    // find content of JSESSIONID  
    $pos = strpos($sid, "JSESSIONID");  
    $sid = substr($sid, $pos);  
    $pos = strpos($sid, ";");  
    $sid = substr($sid, 0, $pos);  
    return ($sid);  }
```

# Benutzte API-Calls I

## Authentifizierung

*dspace\_login(\$ch)*

*dspace\_logout(\$ch, \$sessionid)*

*dspace\_status(\$ch, \$sessionid)*

## Communities

*dspace\_get\_communities(\$ch, \$sessionid)*

*dspace\_create\_community(\$ch, \$sessionid, \$data, \$community)*

*dspace\_delete\_community(\$ch, \$sessionid, \$data)*

*dspace\_get\_topcommunities(\$ch, \$sessionid)*

## Collections und Subcollections

*dspace\_get\_collection(\$ch, \$sessionid, \$collectionid)*

*dspace\_create\_collection(\$ch, \$sessionid, \$data, \$community)*

*dspace\_get\_collections(\$ch, \$sessionid)*

*dspace\_delete\_collection(\$ch, \$sessionid, \$data)*

# Benutzte API-Calls II

## Items

*dspace\_create\_item(\$ch, \$sessionid, \$data, \$collection)*

*dspace\_update\_item(\$ch, \$sessionid, \$data, \$uuid)*

*dspace\_delete\_items(\$ch, \$sessionid, \$itemid)*

*dspace\_delete\_items\_metadata(\$ch, \$sessionid, \$itemid)*

*dspace\_get\_items\_from\_collection(\$ch, \$sessionid, \$collectionid)*

*dspace\_get\_items\_ids\_from\_collection(\$ch, \$sessionid, \$collectionid)*

*dspace\_get\_item\_metadata\_from\_id(\$ch, \$sessionid, \$itemid )*

*dspace\_get\_item\_bitstream\_from\_id(\$ch, \$sessionid, \$itemid)*

## Bitstreams

*dspace\_create\_bitstream (\$ch, \$sessionid, \$data, \$uuid)*

*dspace\_delete\_bitstream(\$ch, \$sessionid, \$bitstreamid)*

# Systemkonzept

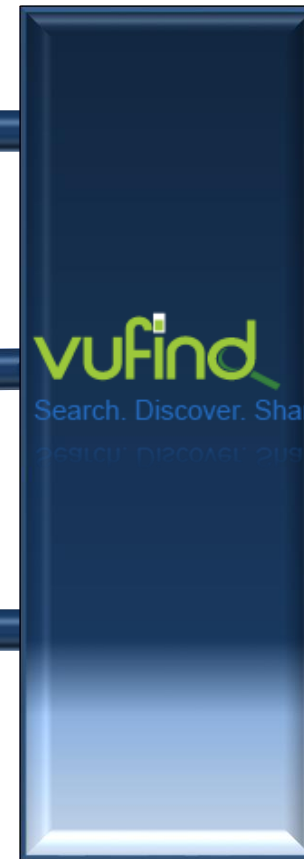
Bibliothek



VET  
Repository

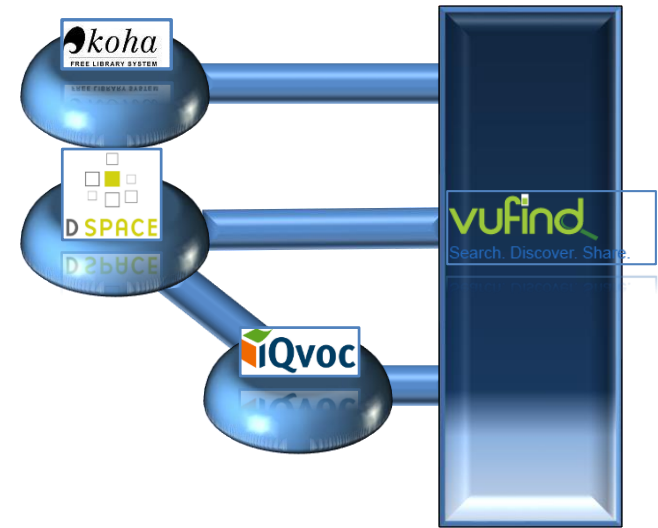


Thesaurus



# Datenversorgung DSpace / VuFind

- Datenbereitstellung via OAI
- Ausspielung im DIM-Format
- Konvertierung in VuFind
- Nachbearbeitung „fulltext“
  - Über DC-Feld „ID“ aus DIM-Daten
  - => Item: `dspace_find_item($sch, $sessionid, $id);`
  - => UUID: `dspace_get_item_bitstream_from_id($sch, $sessionid, $uuid)`
  - Zugriff auf
    - LICENCE
    - TEXT
    - ORIGINAL (Austausch des Links)





zbw [Alle Felder]  Nur Volltexte [Suche]

Erweiterte Suche »

Gesamtbestand VET Repository Bibliothekskatalog

Behalte die Filtereinstellungen.

Suche: zbw

Treffer 1 - 20 Treffer von 113 für Suche 'zbw', Suchdauer: 1,34s

# Trefferliste

Treffer pro Seite 20  
Sortieren Relevanz

Aktionen

Suche verfeinern

Volltext

### Bestand

VET Repository 113

### Format

Zeitschriftenaufsatz 102

Rezension 16

Interview 14

Vortrag 1

Gutachten/ Studie 1

### Weitere Schlagworte

Berufspädagogik 20

Berufsbildung 16

Berufsschule 13

1

Struktur und Qualität wissenschaftlicher Aufsätze in der ZBW : Inhaltsanalyse der ZBW-Jahrgänge 2000 bis 2008

Zu den Favoriten ★  
In die Zwischenablage +

von Tenberg, Ralf; Bach, Alexandra; Sterrenberg, Margret

Format	Bestand	Verfügbarkeit
Zeitschriftenaufsatz Volltext	VET REPOSITORY	Volltext 

Enthalten in Zeitschrift für Berufs- und Wirtschaftspädagogik (2009)  
"... diesem Hintergrund in den zurückliegenden neun Jahren in der ZBW veröffentlicht wurden, welcher zeitliche..."

Volltext



# Fazit



- **Stabil**
- **Performant**
- **Vollständig**



- Dokumentation nicht ganz komplett und eindeutig
- Fehlermeldungen wenig aussagekräftig
- Problem Update Metadaten

**Fragen?**

Antworten

gerne hier und jetzt

oder unter

[roland.keck@bibb.de](mailto:roland.keck@bibb.de)

02228/ 107 1082