The GOKb as a Bridge between Provider Data and FOLIO Agreements.

WOLFcon 2022
Daniel A. Rupp (hbz) & Johann Rolschewski (StaBi Berlin)
2022 August 31
Content

❖ Short introduction
❖ Data flows in the German library landscape
❖ Ingesting e-resource data into FOLIO
❖ KBART and the challenge of e-resource data
❖ A Conflict of Interest – provider data vs. unique bibliographic reference
Operating organizations

❖ ZDB
  ➢ Zeitschriftendatenbank (national library network for serial publications - mainly journals, newspapers, and databases)
  ➢ Chair of GOKB steering committee
  ➢ Editorial office

❖ hbz
  ➢ Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (regional library network)
  ➢ Product owner
  ➢ Developer resources

❖ VZG
  ➢ Verbundzentrale des GBV (regional library network)
  ➢ Technical lead
  ➢ Developer resources
Goals

❖ Global exchange platform for metadata of electronic resources
❖ Accurate up to date information for access
❖ Providing data for automated processes
❖ Openness
  ➢ All data published with an open license
  ➢ Cooperative management through participating libraries
  ➢ Publishers can provide data
Cooperative

- Collaborative management of resources
- Organized in “curatory groups” (consortia, networks, libraries, content providers)
- Support with workflow management tools
Content

❖ Short introduction
❖ **Data flows in the German library landscape**
  ❖ Ingesting e-resource data into FOLIO
  ❖ KBART and the challenge of e-resource data
  ❖ A Conflict of Interest – provider data vs. unique bibliographic reference
Library dataflows in Germany

- EZB Electronic Journals Library
  - ZDB-ID
  - Holdings
  - Bibliographic metadata, EKI

- EBP eBook Pool
  - ZDB-ID, core bibliographic metadata
  - EKI, core bibliographic metadata

- ZDB Journal database
  - Holdings
  - Title list, package ID, ZDB-ID, full bibliographic metadata

- GOKb.org Global Open Knowledgebase
  - Package data, URL
  - EKI, core bibliographic metadata
  - Data download, package metadata, package content (e-resources)

- FOLIO LSP
  - Package metadata, ZDB-ID, package ID, FOLIO-ID, holdings, items

- K10plus Union Catalogue
  - Inventory, OUF, individual bibliogr. metadata, holdings, items
  - Individual bibliogr. metadata, holdings, ZDB-ID, FOLIO-ID

- LAS:ceR
  - License administration system for e-resources
  - Add consortia data (licenses, subscriptions)

- Discovery System
  - License and access data
Content

- Short introduction
- Data flows in the German library landscape

Ingesting e-resource data into FOLIO

- KBART and the challenge of e-resource data
- A Conflict of Interest – provider data vs. unique bibliographic reference
KBART upload into ERMS

Direct import of provider data

Knowledge Bases

Provider ➔ KBART ➔ ERMS

Provider ➔ Knowledge Base ➔ Package ➔ ERMS
Global Open Knowledgebase

The Source and Reference fields are used together as a unique identifier for packages. If a package with the given Source + Reference already exists it will be updated, otherwise a new package will be created. If the Provider entered does not exist, a new one will be created in the local knowledgebase in the Agreements app.

**Package name**
Emerald

**Package source**
emeraldgrouppublishing.com

**Package reference**
ZDB-55-EM1

**Package provider**
Emerald Publishing

**Trust this job as a source of title instance metadata**

Drag & drop to upload

or choose file

Maximum file size: 200 MB

File name
EmeraldPublishing_Global_EmeraldManagement 120_2021-12-20.txt

Uploaded
8/31/2022 6:48 AM
Emerald Management 120

Provider
Emerald Group Publishing Limited

Source
GOKa

Reference
Emerald_Emerald_Management_120

Agreements for this package

E-resources in package

Notes
Emerald Management 120

Provider: Emerald Group Publishing Limited
Source: GOB
Reference: Emerald_Emerald_Management_120

Agreements for this package
E-resources in package
Notes

Show and Tell: Demo VZG & ZBW
Felix Hemme • Jana Freytag • Antje Niemann • Jana Hentschke
Content

- Short introduction
- Data flows in the German library landscape
- Ingesting e-resource data into FOLIO

- KBART and the challenge of e-resource data
  - A Conflict of Interest – provider data vs. unique bibliographic reference
KBART and the Challenge of e-Resource data

- GOKb moved away from provider specific input formats, processed via OpenRefine, and relies solely on KBART
- Looking forward to KBART Phase III

Provider KBART

KBART records
View and download the latest knowledge bases and related tools (KBART) records for current subscriptions.

KBART (Knowledge Bases and Related Tools) is an industry standard used by Knowledge Bases and Libraries to find out what eBook and Journal content they licensed and how to link to it. This is an integral part of the Discovery layer. With KBART, a library's Discovery Service knows which books and journals a library has access to from a publisher. This allows the library's users to search the content the library purchased and when found click through to the content hosted on the publisher's platform. We provide (per the KBART Phase II Recommended Practices), KBART lists for each package we license, and for consortia licenses, so customers can select their licensed content in their Discovery Service.

Springer Nature KBART Lists

Enter keywords, authors, DOI etc.
KBART and the Challenge of e-Resource data

❖ unique identification of providers and packages
KBART and the Challenge of e-Resource data

❖ unique identification of providers and packages

Libraries, Archives, Museums and related Organisations

Listed are all addresses of organizations with an ISIL or Sigel.

Emerald Management 120 eJournal Collection [ZDB-55-EM1]
KBART and the Challenge of e-Resource data

❖ unique
KBART and the Challenge of e-Resource data

- unique identification of providers and packages
- processing of large packages (> 10,000 titles)
KBART and the Challenge of e-Resource data

❖ unique identification of providers and packages
❖ processing of large packages (> 10,000 titles)
❖ "non standard" KBART files
### KBART and the Challenge of e-Resource data

- unique identification of providers and packages
- processing of large packages (> 10,000 titles)
- “non standard” KBART files
- different “cultures” of identifier namespaces

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 publication_title</td>
<td>print_identifier</td>
<td>online_identifier</td>
<td>title_id</td>
</tr>
<tr>
<td>2 Aslib Journal of Information Management</td>
<td>2050-3806</td>
<td>2050-3814</td>
<td>ajim</td>
</tr>
<tr>
<td>3 Asian Libraries</td>
<td>1017-6748</td>
<td>2054-5541</td>
<td>al</td>
</tr>
<tr>
<td>4 Aslib Proceedings</td>
<td>0001-253X</td>
<td>1758-3748</td>
<td>ap</td>
</tr>
<tr>
<td>5 The Bottom Line</td>
<td>0888-045X</td>
<td>2054-1724</td>
<td>bl</td>
</tr>
<tr>
<td>6 Collection Building</td>
<td>0160-4953</td>
<td>2054-5592</td>
<td>cb</td>
</tr>
<tr>
<td>7 Collection and Curation</td>
<td></td>
<td>2514-9326</td>
<td>cc</td>
</tr>
<tr>
<td>8 Digital Library Perspectives</td>
<td>2059-5816</td>
<td>2059-5824</td>
<td>dlp</td>
</tr>
<tr>
<td>9 Data Technologies and Applications</td>
<td></td>
<td>2514-9288</td>
<td>dta</td>
</tr>
</tbody>
</table>
KBART and the Challenge of e-Resource data

- unique identification of providers and packages
- processing of large packages (> 10,000 titles)
- “non standard” KBART files
- different “cultures” of identifier namespaces
KBART and the Challenge of e-Resource data

- unique identification of providers and packages
- processing of large packages (> 10,000 titles)
- “non standard” KBART files
- different “cultures” of identifier namespaces
- automated ingest and update
KBART and the Challenge of e-Resource data

❖ unique identification of providers and packages
❖ processing of large packages (> 10,000 titles)
❖ “non standard” KBART files
❖ different “cultures” of identifier namespaces

Settings for Automated Import and Update

URL

Update Cycle
Daily

Namespace for title_id values
Emerald

Last Run
2022-08-31 10:18:17

Activated

Import Now
KBART and the Challenge of e-Resource data

❖ unique identification of providers and packages
❖ processing of large packages (> 10,000 titles)
❖ “non standard” KBART files
❖ different “cultures” of identifier namespaces

Settings for Automated Import and Update

URL:

Update Cycle:
Daily

Namespace for title_id values:
Emerald

Last Run:
2022-08-31 10:18:17

Activated

Import Now
KBART and the Challenge of e-Resource data

- unique identification of providers and packages
- processing of large packages (> 10,000 titles)
- “non standard” KBART files
- different “cultures” of identifier namespaces
- automated ingest and update
- KBART file editing
KBART and the Challenge of e-Resource data

- unique identification of providers and packages
- processing of large packages (> 10,000 titles)
- "non standard" KBART files
- different "cultures" of identifier namespaces
- automated ingest and update
- KBART file editing
Content

- Short introduction
- Data flows in the German library landscape
- Ingesting e-resource data into FOLIO
- KBART and the challenge of e-resource data

- A Conflict of Interest – provider data vs. unique bibliographic reference
GOKb’s main goal

Representation of **provider data** of e-resources, organized in **packages**, and optimization for the **whole data flow**.
Conflict of interests

ERMS
- Data representation as close to original file as possible
- With the aim of provider communication

Fundamental Dilemma
- Missing or incorrect IDs
- No cross-package identification of titles
- Differences in representation of identical data

Catalogue / Discovery
- Uniqueness
- Identification
- Bibliographic reference
Global Open Knowledgebase

Reference Title
Unique & enriched Record

Package Title (TIPP)
Structured representation of KBART data

Enrichment (& stub record creation)

Catalogue Discovery

Future of libraries is open
Reference Titles

Multiple *Package Titles* (TIPPs) can be linked to a single *Reference Title* with additional bibliographic, descriptive metadata.
Enrichment

Via ZDB (German journal database) and additional sources

- Identifiers
- Bibliographically correct title name
- Title history
- (DDC subject…)

Automated stub creation planned
Review Requests

manual checks of
❖ Inconsistencies
❖ Mismatches
❖ Functional KBART errors
❖ Enrichment problems

Review – Namespace Conflict

Date Created: 2022-05-16 12:58:16
Status: Open

Component (Package Title)
Geriatric Orthopaedic Surgery & Rehabilitation

Action taken

Cause
A package title eISSN matched an existing reference title print-ISSN or vice versa.

Other Components
Geriatric Orthopaedic Surgery & Rehabilitation

Link new component

ToDo
Check if the reference title has a print-ISSN erroneously marked as eISSN or vice versa and correct those identifiers if necessary.
Check the package title identifiers and the KBART record. Correct them if necessary.
Review Requests

manual checks of
- Inconsistencies
- Mismatches
- Functional KBART errors
- Enrichment problems
Global Open Knowledgebase

Enrichment (& stub record creation)

Reference Title
Unique & enriched Bibliographic Record

Package Title (TIPP)
Structured representation of KBART data

Catalogue Discovery

GOKb Global Open Knowledgebase

Additional Sources?

Provider

KBART
Conclusion

❖ The GOKb …
❖ Closes the functional gap between provider data and bibliographically enriched data
❖ Provides data normalization, correction and enrichment
❖ Serves as identity broker for e-resources
❖ Serves as source system at the beginning of data flows
❖ Focuses heavily on identifiers and interfaces
❖ FOLIO ERM as main user
   □ excellent cooperation
Conclusion

❖ The GOKb …
❖ Closes the functional gap between provider data and bibliographically enriched data
❖ Provides data normalization, correction and enrichment
❖ Serves as identity broker for e-resources
❖ Serves as source system at the beginning of data flows
❖ Focuses heavily on identifiers and interfaces
❖ FOLIO ERM as main user
   □ excellent cooperation

Collect and Distribute: Utilizing the APIs of GOKb
Moritz Horn
Conclusion

❖ The GOKb …
❖ Closes the functional gap between provider data and bibliographically enriched data
❖ Provides data normalization, correction and enrichment
❖ Serves as identity broker for e-resources
❖ Serves as source system at the beginning of data flows
❖ Focuses heavily on identifiers and interfaces
❖ FOLIO ERM as main user
   □ excellent cooperation
Further information

• GOKb Website:
  https://gokb.org/

• Technical Wiki:
  https://github.com/openlibraryenvironment/gokb/wiki

• KBART:
  https://www.niso.org/standards-committees/kbart/kbart-frequently-asked-questions

Show and Tell: GOKb demo
  Moritz Horn
Contact information

Daniel Rupp
rupp@hbz-nrw.de

Johann Rolschewski
Johann.Rolschewski@sbb.spk-berlin.de

GOKb
info@gokb.org