

# Metadaten für K10plus-Zentral

- [Übersicht in K10plus-Zentral enthaltene Bibliographische. Metadaten](#)
- [Grundsatz](#)
  - [1. Ablieferung von Metadaten](#)
  - [2. Bereitstellung Python-Code](#)
- [Planung](#)
- [Rechtlicher Rahmen](#)
- [Grundsätze für Metadaten](#)
- [Anmerkungen](#)
  - [Statistische Auswertung](#)
  - [Rechtmanagement](#)
- [Schema zur Lieferung von Zeitschriftenartikeln](#)

## Übersicht in K10plus-Zentral enthaltene Bibliographische. Metadaten

<https://verbundwiki.gbv.de/display/VZG/K10plus-Zentral>

### Grundsatz

Die Inhalte von K10plus-Zentral werden kooperativ, ähnlich wie ein Verbundkatalog, gepflegt. Die VZG besitzt nämlich keine eigenen Ressourcen, um alle Wünsche zu neuen Quellen Bibliotheken erfüllen zu können. Um Bibliotheken die Anlieferung von Metadaten zu Zeitschriftenartikel zu ermöglichen, richtete die VZG eine spezielle Prozessketten ein. Diese Prozesse werden AutoImport[1] genannt. AutoImport verarbeitet Daten in einem JSON-Format, das kompatibel zu den bestehenden Katalogisierungsrichtlinien durch ein Schema definiert ist.[2]

Bibliotheken bietet der AutoImport aktuell 2 Möglichkeiten:

#### **1. Ablieferung von Metadaten**

Bibliotheken liefern Metadaten im obigen Format an einen eine unidirektionale WebDAV-Schnittstelle (Hotfolder). Die Verarbeitung der Daten beginnt sobald Dateien in den Hotfolder gelangen. Update-Zyklen sind lediglich von der Frequenz der Lieferungen durch die Bibliothek abhängig.

#### **2. Bereitstellung Python-Code**

Bibliotheken stellen Python-Code für die VZG bereit, der Quelldaten lädt, valides JSON gemäß obigem Schema generiert und an einen Hotfolder sendet. Der Python-Code muss sich am Muster-Code der VZG orientieren, der sowohl als Jupyter-Code[3] und als auch in einem installierbares Python-Paket[4] zur Verfügung veröffentlicht wurde.

Bei beiden Möglichkeiten werden die Metadaten der Zeitschriftenartikel vollautomatisch mit dem vom BSZ entwickeltet AKET-Tool in PICA+ konvertiert, in spezifische CBS-Datenbanken importiert und schließlich über K10plus-Zentral nutzbar gemacht.

Bisher werden regelmäßig von der UB Braunschweigmehrere Quellen[5] für K10plus-Zentral und von einer Firma Daten für OLC[6] auf diesem Wege geliefert. Python-Code, der zukünftig von der VZG betrieben werden soll, ist aktuell in der UB Braunschweig in Arbeit. Die SUB Hamburg, prüft, ob sie Python-Code für die Konvertierung von Metadaten des Verlags Hogrefe erstellen wird.

AutoImport wird auch für VZG-interne Prozesse verwendet wird, w.z.B. für die Springer Artikel[8], die auch die Grundlage für OLC stellen.

[1] Siehe: Protokoll der Sitzung der AGDiscovery am 12.02.2020

[2] <https://github.com/gbv/articleformat>

[3] <https://github.com/gbv/vzg.jconv-example/blob/master/example.ipynb>

[4] <https://pypi.org/project/vzg.jconv>

[5] arXiv, biorXiv, chemrXiv, engrXiv, techrXiv, [preprints.org](https://preprints.org)

[6] Die Daten werden für die SUB Hamburg, die UB Kiel und das IAI erstellt.

[7] <https://kxp.k10plus.de/DB=1.205/>

### Planung

| Produkt | Anbieter | Typ der Quelle (Volltext, bibliografisch), kurze beschreibung | Bezugskonditionen, rechtliche Informationen | Format und Umfang der Metadaten | Bearbeitungsstand |
|---------|----------|---|---|---------------------------------|-------------------|
|---------|----------|---|---|---------------------------------|-------------------|

|                         |              |   |  |   |  |
|-------------------------|--------------|---|--|---|--|
| de Gruyter              | dito         | bibliografische Nachweise   | E-Mail Kommunikation   | XML   | <ul style="list-style-type: none"> <li>regelmäßige Datenlieferung</li> <li>Umsetzung nach Pica+ fast fertig, gilt für Nationallizenzen also auch für laufende Lieferungen</li> </ul>   |
| Wiley                   | dito         | Metadaten, ggf. auch Volltexte  |  | Testdaten liegen vor  | <ul style="list-style-type: none"> <li>wurde von der ZBW Kiel/Hamburg verhandelt</li> <li>Vertrag von VZG Dez. 2018 unterzeichnet</li> </ul>   |
| Emerald                 |              |   |  |   | <ul style="list-style-type: none"> <li>wurde von der ZBW Kiel/Hamburg verhandelt</li> <li>Daten bis 11/2019 liegen vor. Spätere Daten defekt.</li> <li>Konvertierung mittels: <a href="https://github.com/gbv/vzg.jconv">https://github.com/gbv/vzg.jconv</a></li> </ul>   |
| Taylor & Francis        |              | Metadaten   |  |   | Gespräche mit Verlag.  |
| Sage                    |              | Metadaten   |  |   | Gespräche mit Verlag.  |
| China Academic Journals |              | ca. 40 Mio. Datensätze, Metadaten und Volltexte   |  |   | SBB/PK, Daten sind vorhanden, können bereitgestellt werden   |
| FIS Bildung             | DIPF         | bibliografisch  | frei   | Bibliografische Nachweise von hauptsächlich gedruckt vorliegender Literatur | <p>Es befindet sich noch in Klärung, ob eine Indexierung der Daten in K10plus auch dann möglich ist, wenn keine Statistiken zur Nutzung erzeugt werden können. Das Thema soll auf der nächsten Jahrestagung von FIS Bildung im Frühjahr 2018 besprochen werden. (Stand: 9.11.2017)</p> <p>Im BSZ wird FIS Bildung für einzelne PHs in der Discovery auf einem eigenen Reiter angeboten, z.B. PH Freiburg (<a href="#">Beispielsuche</a>)</p> <p>Die SuUB Bremen bekommt die Daten zur Nutzung in der E-LIB, importiert aber nur Teile des Gesamtpakets. Ein wesentlicher Aspekt ist hier auch die Erschließung des vor Ort vorhandenen Printbestandes. Die Nutzung ist sehr gut.</p> |
| SSRN                    | Elsevier     | 750.000 Dokumente   |  |   | ZBW Kiel und SuUB Bremen hat Interesse an den Daten. Derzeit jedoch zurückgestellt.  |
| Cross Ref               | dito         | bibliografische Nachweise in Ergänzung zu Verlagsdaten, siehe Dokumentation K10plus Zentral | kostenfrei über Rest-Schnittstelle<br><br>VZG bekommt Daten vom KOBV (Stand Dez. 2018) | JSON, sehr gut dokumentiert   | Prüfung abgeschlossen, Daten bei Bedarf<br><br>siehe <a href="#">Präsentation von Gerald Steilen aus der AG Discovery vom 10.5.</a>  |
| Hogrefe                 | dito         | Metadaten + Volltexte für K10plus-Zentral   |  |   | Daten werden regelmäßig an die VZG geliefert   |
| DIN-Normen              | SLUB Dresden |   | via UB Ilmenau   | SQL Dump, entspricht keinem bibliographischen Format                        |  |
| TIB AV-Portal           |              |   | via UB Ilmenau   |   |  |

## Rechtlicher Rahmen

Ein [Mustervertrag](#) englisch liegt vor. Außerdem kann für die Metadaten den Vorgaben der DFG gefolgt werden:

### Grundsätze für den Erwerb DFG-geförderter überregionaler Lizenzen

DFG-Vordruck 12.18 – 6/10

URL: <https://www.nationallizenzen.de/ueber-nationallizenzen/al-grundsaeetze>

Seite 7, Punkt 22

„Der Lizenznehmer oder ein von ihm beauftragter Dritter ist berechtigt, die Metadaten in jeder Weise zu nutzen, die geeignet ist, die Nutzung des lizenzierten Produkts bzw. der darin enthaltenen einzelnen Informationsobjekte durch autorisierte Nutzer zu ermöglichen, zu befördern, zu erleichtern und zu unterstützen. Insbesondere können zu diesem Zweck die Metadaten indexiert werden und ggf. mit Verknüpfungen, die einen direkten Zugang autorisierter Nutzer zum lizenzierten Produkt bzw. den darin enthaltenen einzelnen Informationsobjekten ermöglichen, in lokale Katalogsysteme, regionale oder überregionale Verbundkatalogsysteme, sowie andere Bibliotheksdienste und Informationssysteme Dritter (einschließlich, aber nicht beschränkt auf Suchmaschinen) eingebunden werden. Das Recht, die Metadaten in dieser Weise zu nutzen, steht allen autorisierten Einrichtungen zu. Ausgeschlossen ist die kommerzielle Nutzung der überlassenen Metadaten.“

## Grundsätze für Metadaten

## Grundsätze für den Erwerb DFG-geförderter überregionaler Lizenzen (Allianz-Lizenzen)

DFG-Vordruck 12.181 – 03/15

URL: [http://www.dfg.de/formulare/12\\_181/12\\_181\\_de.pdf](http://www.dfg.de/formulare/12_181/12_181_de.pdf)

Seite 6f., Punkt 8

"Für Metadaten gilt darüber hinaus:

- a) Sie erfüllen sämtliche Voraussetzungen für eine konsistente, automatisierte Verarbeitung. Sie sind zeitgleich zur Bereitstellung / Freischaltung des lizenzierten Produkts auszuliefern.
- b) Daten müssen in genormten Zeichensätzen (möglichst UTF-8) geliefert werden.
- c) Jeder Metadatensatz enthält eine eindeutige, unveränderbare Identifikationsnummer.
- d) Die Metadatenlieferung erfolgt über FTP (File Transfer Protocol), im OAI-PMH-Verfahren (Open Archives Initiative Protocol for Metadata Harvesting) oder im B2B (Business to Business)-Verfahren über eine zu vereinbarende Schnittstelle.
- e) Der Lizenznehmer oder ein von ihm beauftragter Dritter ist berechtigt, die Metadaten in jeder Weise zu nutzen, die geeignet ist, die Nutzung des lizenzierten Produkts bzw. der darin enthaltenen einzelnen Informationsobjekte durch autorisierte Nutzer zu ermöglichen, zu befördern, zu erleichtern und zu unterstützen. Insbesondere können zu diesem Zweck die Metadaten indexiert werden und ggf. mit Verknüpfungen, die einen direkten Zugang autorisierter Nutzer zum lizenzierten Produkt bzw. den darin enthaltenen einzelnen Informationsobjekten ermöglichen, in lokale Katalogsysteme, regionale oder überregionale Verbundkatalogsysteme, sowie andere Bibliotheksdienste und Informationssysteme Dritter (einschließlich, aber nicht beschränkt auf Suchmaschinen) eingebunden und mit diesen Systemen ohne Einschränkung als Linked Open Data freigestellt werden. Das Recht, die Metadaten in dieser Weise zu nutzen, steht allen autorisierten Einrichtungen zu.
- f) Datenelemente und Metadaten folgen den Vorgaben des Papiers Arbeitsgemeinschaft der Verbundsysteme / „Anforderungen der deutschsprachigen Verbundsysteme und der Deutschen Nationalbibliothek an Metadatenlieferungen zu E-Books und E-Book-Paketen 2013“. <http://www.dnb.de/SharedDocs/Downloads/DE/DNB/wir/agVerbundAnforderungenMetadatenEbooks2011.pdf>
- g) Folge- und Updatelieferungen sowie Löschungen werden behandelt wie im Papier Arbeitsgemeinschaft der Verbundsysteme / „Anforderungen der deutschsprachigen Verbundsysteme und der Deutschen Nationalbibliothek an Metadatenlieferungen zu E-Books und E-Book-Paketen beschrieben (<http://www.dnb.de/SharedDocs/Downloads/DE/DNB/wir/agVerbundAnforderungenMetadatenEbooks2011.pdf>): Folgelieferungen für neu hinzugekommene Titel und Auflagen werden mit deren Freischaltung über die üblichen Wege bereitgestellt. Werden verlagsseitig Änderungen oder Korrekturen an den Metadaten vorgenommen, erfolgt die erneute Lieferung der Datensätze in separaten Updatelieferungen. Können die Datenkorrekturen nicht separat, sondern nur zusammen mit neuen Titeln geliefert werden, sind die einzelnen Datensätze eindeutig als "neu" oder "korrigiert" zu kennzeichnen. Korrigierte Sätze sind mit der gleichen Identnummer zu liefern. Fallen Titel oder Auflagen verlagsseitig aus den E-Book-Angeboten heraus (z.B. wegen Zurückziehung der Veröffentlichungsrechte durch den Autor), erfolgt eine erneute Lieferung der betroffenen Metadaten in separaten Löschlieferungen. Können die Löschungen nicht separat, sondern nur zusammen mit neuen oder korrigierten Titeln geliefert werden, sind die einzelnen Datensätze eindeutig als "neu", "korrigiert" bzw. "gelöscht" zu kennzeichnen. (Die Kennzeichnung erfolgt in MARC 21 im Leader auf Position 5, in ONIX für Books im Element <Notification or update type code> gemäß Codeliste 1. Update-Lieferungen in ONIX for Books 2.1 müssen immer einen vollständigen ONIX-Datensatz umfassen.)"

## Anmerkungen

### Statistische Auswertung

Discovery Systeme funktionieren völlig anders, als herkömmliche Suchsysteme (z.B. OPAC, z39.50, ...) vor allem gibt es keine unmittelbar vergleichbaren Trefferlisten.

Man kann sich das vorstellen wie eine andere Art Lexikon: anstatt nach einer Anfrage zu den Stellen geführt zu werden, die exakt meiner Anfrage entsprechen, wird das ganze Lexikon umsortiert, und die Einträge stehen vorn, die gemäß meiner Anfrage die höchste Relevanz besitzen. Relevanz ist hier eine kompliziert ermittelte mathematische Größe, die mit der errechneten Relevanz korreliert.

Das heisst: egal, was gefragt wird, alle Datensätze werden immer mit umsortiert, und zwar immer alle, weil immer alle Indexeinträge mit umsortiert werden. Angesichts dessen lässt sich die Eigenschaft: "Datensatz gehört zur Treffermenge" schlicht nicht sinnvoll auswerten. Darüber können wir keine Angaben machen.

Außerdem lädt ein Browser immer vollständig die ersten (meist 50) Treffer der Rangliste, unabhängig davon, ob ein Nutzer die Treffer zu Gesicht bekommt. Denn auf diese Weise ist eine sehr schnelle und flüssige Navigation im Discovery-System möglich. Allein schon festzustellen welche Trefferanzahl ein Nutzer zu Gesicht bekommt, würde sowohl eine umfassende Analyse technischer Parameter des verwendeten Endgerätes (z.B. PC, Tablet, Smartphone) als auch des Nutzerverhaltens erfordern. Aus der ermittelten Trefferzahl müsste anschließend die Treffermenge der WTI zuzurechnenden Datensätze ermittelt werden. Abgesehen vom erheblichen technischen Aufwand, würde das eine umfassende Überwachung des Nutzerverhaltens bedeuten.

Daher müssen Statistiken, die einzelne Kollektionen (w.z.B WTI, JSTOR) betreffen von der Frontend-Anwendung (z.B. VuFind, Lukida) generiert werden.

### Rechtmanagement

Das Management der Zugriffsrechte erfolgt in der Regel auf der Ebene von Kollektionen (w.z.B WTI, JSTOR), ist aber prinzipiell bis zum einzelnen Datensatz steuerbar. Letzteres kam in der Praxis noch nicht vor und sollte auch vermieden werden, da dies zu Lasten der Performance gehen würde.

## Schema zur Lieferung von Zeitschriftenartikeln

Eine Definition eines einfachen und erweiterbaren JSON-Formates ist auf der Plattform GitHub veröffentlicht: <https://github.com/gbv/articleformat>

Ein JSON-Schema (<https://json-schema.org/>) beschreibt ein JSON-Format zur Lieferung von bibliographischen Metadaten zu Zeitschriftenartikeln an die Verbundzentrale des GBV (VZG). HTML-Dokumentation: [http://findex.gbv.de/articleformatdoc/schemas/article\\_schema.html](http://findex.gbv.de/articleformatdoc/schemas/article_schema.html) Ziel ist es, Bibliotheken ein einfaches Verfahren für die Lieferung von Daten zu ermöglichen, die automatisch validiert und direkt in eine CBS-Datenbank abgespeichert werden können.