# Protokoll der Sitzung der AG Discovery am 12.02.2020

**Beginn:** 14:00
**Ende:** 16:10
Protokoll: Jan Frederik Maas

Teilnehmer/Innen:

- Blenkle, Martin (SuUB Bremen)
- Conradt, Volker (BSZ)
- Genat, Berrit (TIB)
- Goldschmidt, Oliver (TUB Hamburg)
- Kaun, Matthias (SBB/PK)
- Maas, Jan Frederik (SUB Hamburg)
- Pianos, Tamara (ZBW)
- Simanowski, Jörg (UB Rostock)
- Steilen, Gerald (VZG)

Entschuldigt: Diedrichs, Reiner (VZG)

TOP 0 Begrüßung, Organisatorisches

Wahl Sprecher/In der AG Discovery: Tamara Pianos wird einstimmig für die nächsten drei Jahre als Sprecherin wiedergewählt.

Besetzung der AG: Aktuell liegen keine Bewerbungen für Teilnahme an der AG vor. Eine Erweiterung der AG ist mittlerweile grundsätzlich möglich. Auf der Webseite der AG soll auf die offenen Positionen verwiesen werden. Die AG sieht 12 Mitglieder als Obergrenze an.

Ausrichtung der AG: Die AG sieht ihren Tätigkeitsschwerpunkt im Umgang mit den Daten, die den Discoverysystemen zugrunde liegen (z.B. bibliografische Metadaten), da diese die gemeinsame Grundlage von Discoverysystemen darstellen.

TOP 1: Stand der Metadatenlieferungen: Gerald Steilen beantwortet Detailfragen zu dem Autoimport-Feature der VZG. Jeder Datenlieferant bekommt einen eigenen Zugang, Prozesse können so nachvollzogen werden. Grundsätzlich sind auch sehr große Lieferungen (>1 Mio. Datensätze) möglich. Datenlieferungen werden initial geprüft. Die Prozesskette ist nach Einrichtung vollautomatisch.

Folgend die inhaltliche Zusammenfassung der Diskussion durch Herrn Steilen:

*Bibliographische Metadaten für K10plus-Zentral*

*Die VZG besitzt keine eigenen Ressourcen, um Wünsche zu neuen Quellen, wie es die Aufstellung „Metadaten für K10plus-Zentral“[1] suggeriert, Bibliotheken erfüllen zu können. Um Bibliotheken die Anlieferung von Metadaten zu Zeitschriftenartikel zu*

[1] https://verbundwiki.gbv.de/pages/viewpage.action?pageId=50364431

*ermöglichen, richtete die VZG im Jahr 2020 eine neue Prozesskette ein. Diese neue Prozesskette wird AutoImport[2] genannt. AutoImport verarbeitet Daten in einem JSON-Format, das kompatibel zu den bestehenden Katalogisierungsrichtlinien und durch ein Schema definiert ist.[3]*

*Bibliotheken bietet der AutoImport aktuell zwei Möglichkeiten:*

*1. Ablieferung Metadaten*
*Bibliotheken liefern Metadaten im obigen Format an 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-Notebook[4] und als auch als installierbares Python-Paket[5] zur Verfügung veröffentlicht wurde.*

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

*Bisher werden regelmäßig Daten von der UB Braunschweig aus mehreren Quellen[6] für K10plus-Zentral und von der Firma AGI für OLC[7] auf dem ersten 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 zukünftig Python-Code für die Konvertierung von Metadaten des Verlags Hogrefe erstellen kann.*

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

Eine Diskussion und Klärung der Datenquellen und Abhängigkeiten sowie möglicher Datenhaltungskonzepte ist sinnvoll, soll aber in einer späteren Sitzung der AG und/oder in anderen Gremien fortgeführt werden.

TOP 2: Kennzeichnung der elektronischen Bestände / Bestandsnachweise elektronischer Bestände: Es liegen keine aktuell keine neuen Erkenntnisse oder Informationen vor. Herr Steilen spricht sich dafür aus, das Thema zusammen mit Datenhaltungskonzepten zu diskutieren.

---

[2] Siehe: Protokoll der Sitzung der AG Discovery am 12.02.2020
[3] https://github.com/gbv/articleformat
[4] https://github.com/gbv/vzg.jconv-example/blob/master/example.ipynb
[5] https://pypi.org/project/vzg.jconv
[6] arXiv, biorXiv, chemrXiv, engrXiv, techrXiv, preprints.org
[7] Die Daten werden für die SUB Hmaburg, die UB Kiel und das IAI erstellt.
[8] https://kxp.k10plus.de/DB=1.205/

Bestandsnachweise einzelner Bibliotheken an gedruckten sowie an elektronischen Medien in der bibliographischen Suchmaschine K10Plus-Zentral stammen bisher aus der Verbunddatenbank K10plus und aus OLC.

Ausgangspunkt und Basis für jede ERM-Lösung, die grundsätzlich dazu geeignet wäre Bestandsnachweise einzelner Bibliotheken für eRessourcen zu generieren, bildet eine Wissensbasis (Knowledge Base), in der alle spezifischen Informationen zu verfügbaren elektronischen Ressourcen (Titel, Pakete, Anbieter, Links) verwaltet werden und über Schnittstellen und automatisierte Prozesse für nachgelagerte Nachweis- und Verwaltungssysteme verfügbar sind.

Die VZG betreibt mit der Global Open Knowledge base (GOKb) eine solche Wissensbasis. Im Zusammenspiel mit folio ERM wird die VZG Bibliotheken einen Service anbieten, der Bestandsnachweise für elektronischen Medien generiert. Diese Bestandsnachweise sind in K10Plus-Zentral zur Filterung geeignet.

TOP 3: DAIA/PAIA: Es liegen keine Neuigkeiten vor.

TOP 4: Update der UAG Formate: Kurzer Werkstattbericht von Jan Maas, Folien im Anhang.

Hinweis von Herrn Simanowski: Offenes Problem wo die Daten herkommen sollen, die nutzerfreundliche Formate-Facetten befüllen sollen.

Hinweis von Frau Genat: Für den Medientyp werden an der TIB 14 Pica-Felder ausgewertet.

Hinweis von Herrn Steilen: Für den K10Plus-Index wird mir Marc-Daten gearbeitet. Die Auswertung ist sehr komplex. Um eine Medientypfacette gut zu gestalten ist Arbeit auf zahlreichen Ebenen notwendig.
Genat: Maschinenlesbarkeit der Daten und Nachnutzung der Daten kommt bei Katalogisierung ggf. zu kurz.

Die AG Formate nimmt als nächstes Kontakt mit verschiedenen Personen auf um zu klären, ob wie die Anforderungen der AG erfüllt werden können.

Frage von Frau Genat: Wie oft wird Formatfacette genutzt? Antwort (Herr Steilen): Nutzung bei Lukida partiell sehr gut, vielen Anfragen aus Bibliotheken.

TOP 5: Austauschoptionen/Anwendertreffen: Es wurde ein Kommunikationskanal (Mailliste) zu DAIA/PAIA eingerichtet.

TOP 6: Handlungsfelder mit Bezug zum Thema Discovery

Die AG diskutiert Wünsche zum Thema Discovery, die an sie herangetragen werden:

Diskussion Anforderungen (UB Kiel):

*Relevanzsortierung*
*Insbesondere bei Suchen nach "Lehrbüchern", also z. B. Werke mit Begriffen wie Lehrbuch, Einführung, Introduction, etc. im Titel oder den Schlagworten treten*

*Probleme mit der Relevanzsortierung zu Tage, weil die Aktualität in diesen Fällen viel stärker gewichtet sein müsste.*

*Popularität ins Ranking einbeziehen*
*z.B. Zugriffsstatistiken der Bibliotheken bei elektronischen Inhalten und Ausleihzahlen bei gedruckten Beständen könnten regelmäßig aggregiert und akkumuliert ausgewertet und dann zur Anreicherung des Rankings genutzt werden.*

*Angebot verschiedener Ranking Presets, die von den einzelnen Bibliotheken in der Lukida Admin-Oberfläche für den Standort ausgewählt werden können. z.B. stärkeres Boosting von Aktualitätsmerkmalen angeboten werden, oder ein stärkeres Boosting von Verfügbarkeit (elektronisch-ausleihbar-präsent)*

*Etwas experimentell, aber vielleicht ganz interessant: query-basierte Beeinflussung der Relevanz z.B. auf Basis von bestimmten Suchbegriffen, die von der Bibliothek für ihr Discovery-System*

Die Ranking-Formel ist durch den Index fest (Okapi BM25) vorgegeben. Die Einstellung der Relevanzsortierung (Nutzung und Gewichtung der Felder, Boosting etc.) muss durch die lokalen Discoverysysteme vorgegeben und eingestellt werden. Lehrbücher, Aktualität etc. können z. B. über Boosting-Funktionen höher gerankt werden. Ranking-Presets sind spezifisch für das Discoverysystem und sollten daher in der jeweiligen Community (Lukida, VuFind, beluga core etc.) abgestimmt werden.

Die Popularität könnte mit indexiert werden, ist aber stark von der fachlichen Ausrichtung der Nutzer abhängig und somit von der Ausrichtung der Bibliothek. Somit sind lt. Ansicht der AG Discovery auch hier lokale Lösungen voraussichtlich ebenfalls zielführender.

Prüfung der Ziel- und Leistungsvereinbarungen der VZG für 2021: Die AG hat die vorgeschlagenen Ziel- und Leistungsvereinbarungen eingehend geprüft und schlägt keine Ergänzung vor.

TOP 7: LOC-DB: Infos zum Projekt (Volker Conradt):
Das Thema wird aufgrund von Zeitmangel in der nächsten Sitzung der AG ausführlich behandelt.

TOP 8: Sonstiges
Vorschlag von Herrn Steilen: Vorschläge für Themen für AG-Sitzungen sollen in einem Themenspeicher gesammelt werden. Die VZG prüft, ob ein nichtöffentlicher Bereich im vorhandenen Wiki eingerichtet werden kann.

Die nächste Sitzung der AG Discovery findet am 17.05.2021 um 14:30h-16:30h statt.