

Julia Alexandra Goltz-Fellgiebel (KOBV)

Markus Putnings (FAU)

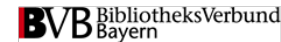
# Open-Access-Transformation mit DeepGreen: Gemeinsam den (grünen) Schatz heben

107. Deutscher Bibliothekartag 2018, Berlin

14. Juni 2018, Session: Open Access – Zukunftsfähig?

# Projektsteckbrief

- ▶ Projektziel: Vertragskonforme, automatisierte Datenablieferung (Metadaten + Volltexte) von Verlagen an Repositorien statt manueller Selbstarchivierung
- ▶ Erste Förderphase: 01.01.2016 – 31.12.2017
- ▶ Zweite Förderphase bewilligt für 2 Jahre: voraussichtlich 01.08.2018 - 31.07.2020
- ▶ Drittmittelförderung durch die DFG, Wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS) – Ausschreibung Open-Access-Transformation
- ▶ Projektbeteiligte:





# Ausgangslage



# Ausgangslage

- ▶ Bibliotheken pflegen institutionelle Repositorien mit dem Ziel, alle Veröffentlichungen der Mitarbeitenden der eigenen Institution nachzuweisen.
- ▶ Forschende möchten die Zweitveröffentlichung, aber keine „zweite Arbeit“.
- ▶ Bibliotheken kennen nicht alle Veröffentlichungen der Mitarbeitenden ihrer Institution und müssen die (bekannten) Daten einzeln recherchieren und herunterladen.
- ▶ Verlage haben Daten und (idealerweise) Angaben zur Affiliation der Autoren/Autorinnen. Sie unterstützen die Zweitveröffentlichung.  
(teils, nominell in Policies und Verträgen, real mittels Deposits über SWORD)



# Gleichzeitig ...

- ▶ entwickelt sich eine neue Publikationskultur und wächst der Druck im Sinne einer Open-Access-Transformation (z.B. unterschiedliche Open-Access-Strategien in Bund, Ländern und Forschungseinrichtungen),
- ▶ werden/wurden auf bundesweiter Ebene DFG-geförderte Allianz-Lizenzen und Nationallizenzen mit weitgehenden grünen Open-Access-Rechten verhandelt,
- ▶ die aber real kaum genutzt/umgesetzt werden,
  - ... da dem/der Autor/Autorin nicht zwingend bekannt und
  - ... die Pflege für den/die Autor/Autorin bzw. die Bibliothek sehr ressourcenintensiv ist.

# Aufgaben



# Aufgaben

- ▶ Verlage müssen motiviert werden, Metadaten und Volltexte (PDF) zu liefern.
  - ✓ Karger und SAGE Publications waren von Beginn an Projektpartner
  - ✓ Erhaltene Testdaten von der European Mathematical Society, DeGruyter, Oxford University Press, Royal Society of Chemistry
  - ✓ Potenzielle weitere Partner über JISC und div. Verlagsverhandlungen
- ▶ Technische Implementierung muss
  - Voraussetzungen der Verlage berücksichtigen,
  - Inhalte der Lizenzverträge (Informationen sind über die EZB abrufbar) abbilden,
  - lokale Voraussetzungen der Bibliotheken/Repositorien abbilden (z. B. standardisierte Importformate, Schnittstellen, Protokolle).

# Technische Lösung: Keep it simple!

- ▶ Die Lösung: ein weitgehend automatisierter Workflow über eine zentrale Datendrehscheibe unter Einbeziehung der Verlage!
  - Nachnutzung bestehender Konzepte und verfügbarer Software
  - Benutzerzentrierter Service: Abwicklung über Konten mit Profil-Angaben
  - Daten werden gemäß (Verlags-)Kontenprofil angenommen, gegen Kontenprofil von Bibliothek abgeglichen und entsprechend bereitgestellt
  - Design als **Reines Push-Forward-System** mit **Notifications**
  - Kein Dark Archive: Datenspeicherung erfolgt mit beschränkter Verweildauer



# Technischer Steckbrief

- ▶ Programmierung auf Basis des [Jisc-Publications Event Router](#) (V1, 2016)
- ▶ *Python Framework Flask v0.9*
- ▶ Interne Datenverarbeitung über ElasticSearch
- ▶ Unterstützte Standards
  - Datenannahme: sFTP, SWORD v2, Metadatenformat: DTD NLM/NISO JATS, **DTD RSC**
  - Datenabgabe: OAI-PMH, SWORD v2, Metadatenformat / Export: NLM/NISO JATS, RIOXX, DC, **METS/MODS, DTD RSC, OPUS-XML, ESciDoc**
- ▶ Softwareentwicklung: <https://github.com/OA-DeepGreen>, Lizenz: [Apache, v2.0](#)

(Grün gekennzeichnet sind Erweiterungen von DeepGreen gegenüber der Basissoftware von Jisc)

# Die Zuordnungslogik und -implementierung



# Die Zuordnungslogik

## ▶ Zweistufige Berechtigungslogik:

1. Lizenzüberprüfung anhand der ISSN und des Publikationsjahrs (Jahrgang):  
Titellisten und Jahrgänge werden aus der EZB-Datenbank periodisch ermittelt  
→ hierfür prototypische **Anbindung an die EZB** (Vertragsinformationen)

2. Feststellung der berechtigten wiss. Einrichtungen durch Analyse der  
Affiliationsangaben in den Artikelmetadaten: Affiliationsangaben werden von  
der Einrichtung selbst gepflegt (z.B. Namensvarianten der Einrichtung)

## ▶ Jede Zuordnung/Zustellung wird protokolliert.

*(Grün gekennzeichnet sind Erweiterungen von DeepGreen  
gegenüber der Basissoftware von Jisc)*

# Prototypische Implementation

- ▶ Anbindung Verlage seit Oktober 2016
  - Kontenfunktionalität
  - Datenkonvertierung Verlagsformat auf DeepGreen-Internformat
- ▶ Anbindung Testrepositorien (OPUS 4, DSpace, ESciDoc) seit Mai 2017
  - Kontenfunktionalität
  - Einfügen spezifischer Parameter
  - Datenkonvertierung DeepGreen-Internformat auf Repositorium-Importformat
- ▶ In Vorbereitung für einen Testbetrieb wurden **Standard-Konten für ca. 250 wissenschaftliche Einrichtungen in Deutschland** eingerichtet.

*(Grün gekennzeichnet sind Erweiterungen von DeepGreen gegenüber der Basissoftware von Jisc)*



Source: Train Station Selzthal; Work of [Isiwal](#); [CC BY-SA 3.0](#)



# Zusammenfassung: Projektmeilensteine

## 2016

- ✓ Projektstart Januar 2016, [Publikation des Projektantrags](#)
- ✓ Vortrag auf dem DBT 2016 mit [Veröffentlichung des Textes](#) in der OA-Zeitschrift o-bib
- ✓ [Umfrage unter 14 Verlagen](#) 2016 – Erhebung des aktuellen technischen Status Quo der Allianz-Lizenzverlage
- ✓ Konzeption/Veröffentlichung eines [DeepGreen-Metadatenschemas](#), Version 1.1
- ✓ Workshop mit ExpertInnen aus Bibliotheken und Verlagen
- ✓ Freischaltung des Prototypen: Datendrehscheibe Alphaversion inkl. Konzept für Rechteverwaltung

# Zusammenfassung: Projektmeilensteine

## 2017

- ✓ Freischaltung Datendrehscheibe Version 2: Die Datenabgabe an die Testrepositorien erfolgt im passenden Importformat
- ✓ Freischaltung Datendrehscheibe Version 3.0: „Standard“-Konten eingerichtet
- ✓ [Artikel in ABI Technik](#) zur Umsetzung der Open-Access-Komponente aus den Allianz-Lizenzverträgen
- ✓ Öffentlicher Abschlussworkshop 14.11.2017, Präsentation des Projektstandes und Perspektiven
- ✓ Kennzahlenermittlung mit einer manuellen Stichprobe von allen Sage/Karger-Publikationen aller Projektbeteiligten (Jahr 2015) – Abgleich mit den Datenlieferungen beider Verlage



# Kennzahlenermittlung





# Kennzahlenermittlung

- ▶ Manuelle Erhebung über externe Datenbanken im Sommer 2017
- ▶ Gesamtjahrgangslieferung der Pilotverlage
  - Karger lieferte Ende 2017 die Jahrgänge 2015 und 2016, zeitgleich gibt es seit November 2017 zusätzlich aktuelle, wöchentliche Updates (jeweils Metadaten plus Publikation im pdf-Format).
  - Sage lieferte im Februar/März 2018 die Jahrgänge 2014 und 2015 (jeweils nur Metadaten).
- ▶ Vorläufiges qualitatives Ergebnis: Die manuell über externe Datenbanken erhobenen Werte (Anz. möglicher Artikel für Testrepositorien) decken sich zu ca. 95% mit den Werten (Anz. positiv zuzuordnende Artikel) der Verlagslieferungen.

## 2. Projektphase



# Oberstes Ziel der 2. Förderphase

- ▶ DeepGreen mit nachhaltigem Geschäfts- und Betriebsmodell innerhalb eines Jahres als Dienst implementieren (mit Konzentration auf Allianz-Lizenzen):
  - Beziehung zwischen Verlagen, Repositorienbetreibern und den Betreibern von DeepGreen rechtlich ausgestalten
  - Relevante Zahl von Verlagen mit regelmäßigen Datenlieferungen im Routinebetrieb gewinnen (Schritt 1: Allianz-Lizenzen)
  - Technische Infrastruktur konsolidieren und erweitern, z.B. Lieferroutinen mit Verlagen und Unterstützung weiterer Repositorien-Softwaretypen
  - Musterworkflows für Repositorien beispielhaft definieren und dokumentieren

# Perspektiven der 2. Förderphase

- ▶ DeepGreen ausweiten auf neue Lizenzkontexte:
  - Konsortiallizenzen ohne explizite grüne Open-Access-Komponente
  - FID-Lizenzen
  - *Goldene Open-Access-Artikel aus reinen OA-Zeitschriften, Transformationsverträge, DEAL-Verträge (Offsetting)*



# Perspektiven der 2. Förderphase

- ▶ DeepGreen ausweiten für neue Abnehmersysteme:
  - Fachrepositorien und Forschungsinformationssysteme prototypisch als Datenabnehmer einbinden
  - Eruierung von Datenelementen/Verfahren zur fachlichen Identifikation der zuzuordnenden Artikel
  - Klärung rechtlicher Zusammenhänge
  - Aufwandsanalyse für notwendige technische Spezifikationen, Schnittstellen und Workflows

# Verlagspartnerschaften

## seit Projektbeginn

- ▶ Karger AG, Basel, Schweiz
- ▶ SAGE Publications, London, UK

## seit Sommer 2017

- ▶ BMJ, London, UK
- ▶ DeGruyter, Berlin, Deutschland
- ▶ MDPI, Basel, Schweiz

- ▶ Austausch mit Jisc (UK)

*erste Projektphase:  
technische Aspekte, Workflows*

*zweite Projektphase:  
Workflows, Vereinbarungen mit  
Datenlieferanten/-abnehmern*

# Institutionelle Partnerschaften

## ... institutionelle Repositorien

- ▶ Projektbeteiligten FAU, GFZ und TU Berlin
- ▶ FU Berlin

## ... disziplinäre Repositorien

- ▶ DIPF Frankfurt am Main (peDOCs)
- ▶ Staatsbibliothek zu Berlin (<intR><sup>2</sup>Dok)
- ▶ ZBW Kiel (Econstor)

## ... Forschungsinformationsdienste

- ▶ GESIS Mannheim (FID SSOAR)
- ▶ SUB Göttingen (FID GEO)

## ... Forschungsinformationssysteme

- ▶ UB Bielefeld (PUB)
- ▶ TIB Hannover (PURE und VIVO)

## Weitere Kooperationen

- ▶ UB Regensburg (Projekt OA-EZB)

# DeepGreen – Projektkonsortium 2016 bis heute

## Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV) – Projektleitung

- Prof. Dr. Thorsten Koch
- Beate Rusch
- Julia Alexandra Goltz-Fellgiebel
- Dr. Thomas Dierkes
- Jens Schwidder
- Laura Baumann (bis 08/2017)

## Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Universitätsbibliothek

- Konstanze Söllner
- Markus Putnings
- Cornelia Hoffmann
- Regina Heidrich

## Bayerische Staatsbibliothek (BSB) / Bibliotheksverbund Bayern (BVB)

- Dr. Klaus Ceynowa
- Dr. Hildegard Schäffler
- Dr. Lilian Landes (2016)
- Dr. Ortwin Guhling
- Michael Kassube (bis 07/2018)
- Matthias Groß

## Helmholtz Open Science Koordinationsbüro am Deutschen GeoForschungsZentrum (GFZ)

- Roland Bertelmann
- Heinz Pampel
- Kaja Scheliga (bis 07/2018)
- Tobias Höhnow

## Technische Universität Berlin (TUB), Universitätsbibliothek

- Jürgen Christof
- Monika Kuberek
- Per Broman
- Pascal Becker
- Dagmar Schobert (bis 12/2017)
- Michaela Voigt (bis 12/2017)
- Marsa Haoua (seit April 2017)
- Melanie Janßen (seit Sept 2017)

# Vielen Dank für Ihre Aufmerksamkeit!

## Impressum

Die Präsentation steht unter einer CC BY-Lizenz:  
<https://creativecommons.org/licenses/by/4.0/deed.de>.

## Lizenzen

Sofern nicht anders angegeben stehen alle verwendeten Bilder unter der [Lizenz CCO](#).

## Weitere Informationen zum Projekt

Kontakt zum Projektmanagement: [info@oa-deepgreen.de](mailto:info@oa-deepgreen.de)  
Projektwebseite mit Informationen zum Projekt: <https://deepgreen.kobv.de/>  
Prototyp der DeepGreen-Datendrehscheibe: <http://www.oa-deepgreen.de/>