

Beate Rusch, Dr. Hildegard Schäffler

Open-Access-Transformation mit DeepGreen: Was wir wollen, schon erreicht haben und noch anstreben

Ein sportlicher Vortrag zum DeepGreen-Abschluss-Workshop
14. November 2017, Technische Universität Berlin

DeepGreen - Projektsteckbrief

- ▶ Projektziel: Vertragskonforme, automatisierte Datenablieferung (Metadaten + Volltexte) von Verlagen an Repositorien
- ▶ Start: 01.01.2016
- ▶ Aktuelle Förderphase: 2 Jahre bis 31.12.2017
- ▶ Drittmittelförderung durch die DFG, Wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS) – Ausschreibung Open-Access-Transformation
- ▶ Projektbeteiligte:
 - Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV, Projektleitung)
 - Bibliotheksverbund Bayern (BVB)
 - Bayerische Staatsbibliothek (BSB)
 - Friedrich-Alexander-Universität Erlangen-Nürnberg, Universitätsbibliothek (FAU)
 - Technische Universität Berlin, Universitätsbibliothek (TU Berlin)
 - Helmholtz Open Science Koordinationsbüro am Deutschen GeoForschungszentrum GFZ

Inhalt

Die Ausgangslage

Die Aufgaben

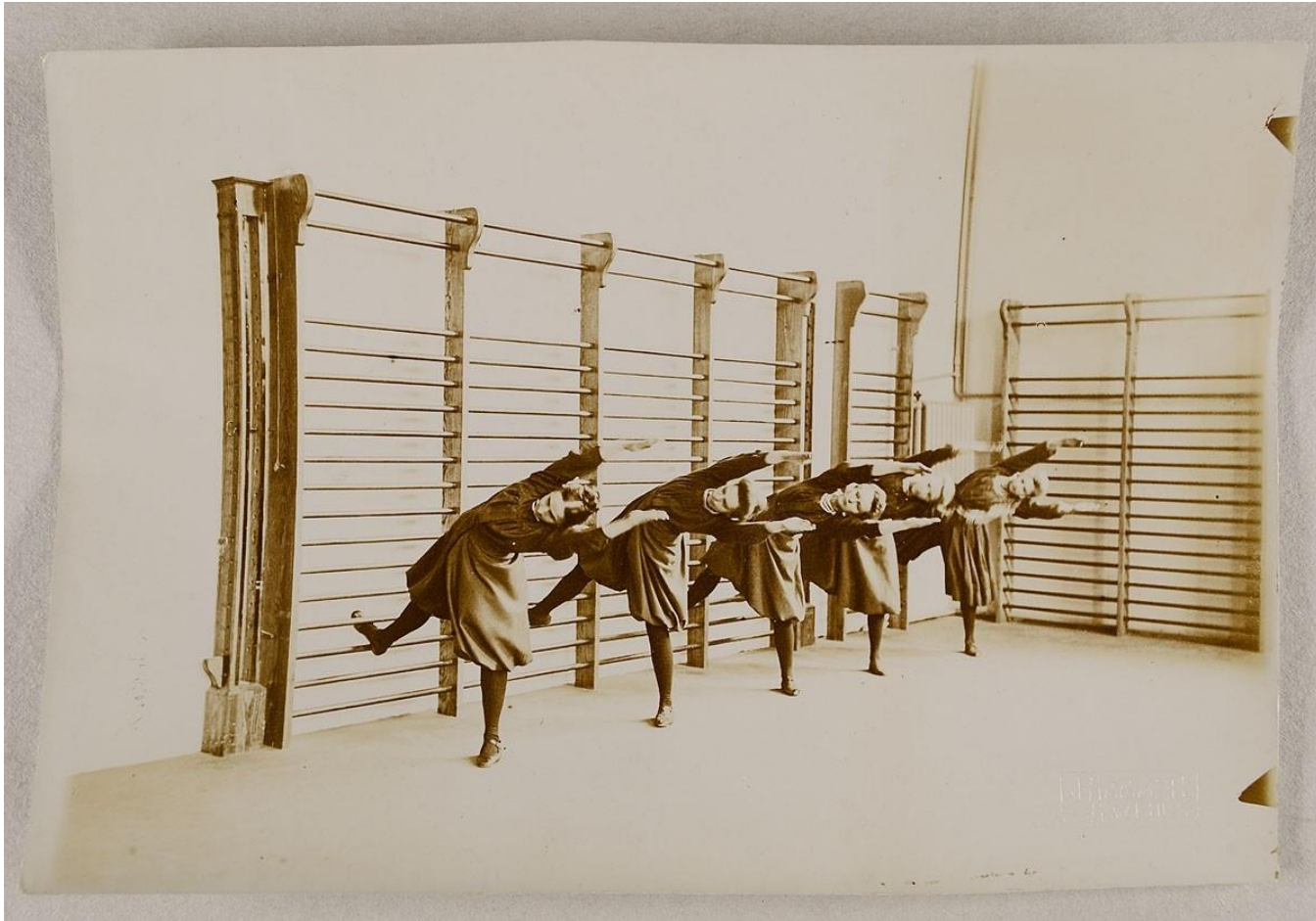
Die Technische Lösung

Erste Ergebnisse

Perspektiven

Zusammenfassung

Die Ausgangslage



Die Akteure

- ▶ Bibliotheken pflegen institutionelle Repositorien mit dem Ziel, alle (Open-Access-)Veröffentlichungen der Mitarbeiter/innen der eigenen Institution nachzuweisen.
- ▶ Wissenschaftler/innen möchten die Zweitveröffentlichung, aber keine „zweite Arbeit“.
- ▶ Bibliotheken kennen nicht alle Veröffentlichungen der Mitarbeiter/innen ihrer Institution und müssen die (bekannten) Daten einzeln herunterladen.
- ▶ Verlage haben Daten und (idealerweise) Angaben zur Affiliation der Autoren/innen.

Gleichzeitig ...

- ▶ entwickelt sich eine neue Publikationskultur und wächst der Druck im Sinne einer Open-Access-Transformation,
- ▶ 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/in nicht zwingend bekannt und
 - ▶ ... die Pflege für den/die Autor/in bzw. die Bibliothek sehr ressourcenintensiv ist.

Die Lösung: ein weitgehend automatisierter Workflow über eine zentrale Datendrehscheibe unter Einbeziehung der Verlage!

Die Aufgaben



Verlage motivieren

- ▶ Verlage müssen motiviert werden, Metadaten und Volltexte (PDF) zu liefern.
- ▶ Welche Daten (Verlagsproduktion; Vertragsinhalt; mit/ohne Embargo) können/wollen Verlage liefern?
- ▶ Welche Qualität haben die Affiliationsangaben in den Verlagsdaten?
- ▶ Welche Formate können Verlage liefern?
- ▶ Welche Ablieferschnittstellen können Verlage bedienen?
- ▶ Karger und SAGE Publications waren von Beginn an Projektpartner.
- ▶ Technische Implementierung muss jeweils spezifische Voraussetzungen der Verlage berücksichtigen.

Vertragsinhalte abbilden

- ▶ Vertragsinhalte müssen jeweils spezifisch abgebildet werden.
- ▶ Welche Bibliothek nimmt an welchem (Allianz-/Nationallizenz-)Vertrag teil?
- ▶ Welche Zeitschriftentitel umfasst der (Allianz-/Nationallizenz-)Vertrag?
- ▶ Welche Zeitschriftenjahrgänge umfasst der (Allianz-/Nationallizenz-)Vertrag?
- ▶ Welche Bibliothek nimmt mit welchem Fach- bzw. Jahrgangsmodule des (Allianz-/Nationallizenz-)Vertrags teil?
- ▶ Informationen liegen in der **Elektronischen Zeitschriftendatenbank (EZB)** vor.

Bibliotheken einbinden

- ▶ Lokale Voraussetzungen der Bibliotheken/Repositorien müssen abgebildet werden.
- ▶ Welche Autoren/innen gehören zu welcher Institution/Bibliothek?
- ▶ Oder: Wie wird die Affiliationszugehörigkeit erkannt, wenn die Vorauswahl nicht durch den Verlag erfolgt?
- ▶ Welche Formate können (Bibliotheks-)Repositorien lesen?
- ▶ Welche Schnittstellen/Standards können Bibliotheken/Repositorien bedienen?

Die Technische Lösung



Keep it simple, stupid! (KISS)

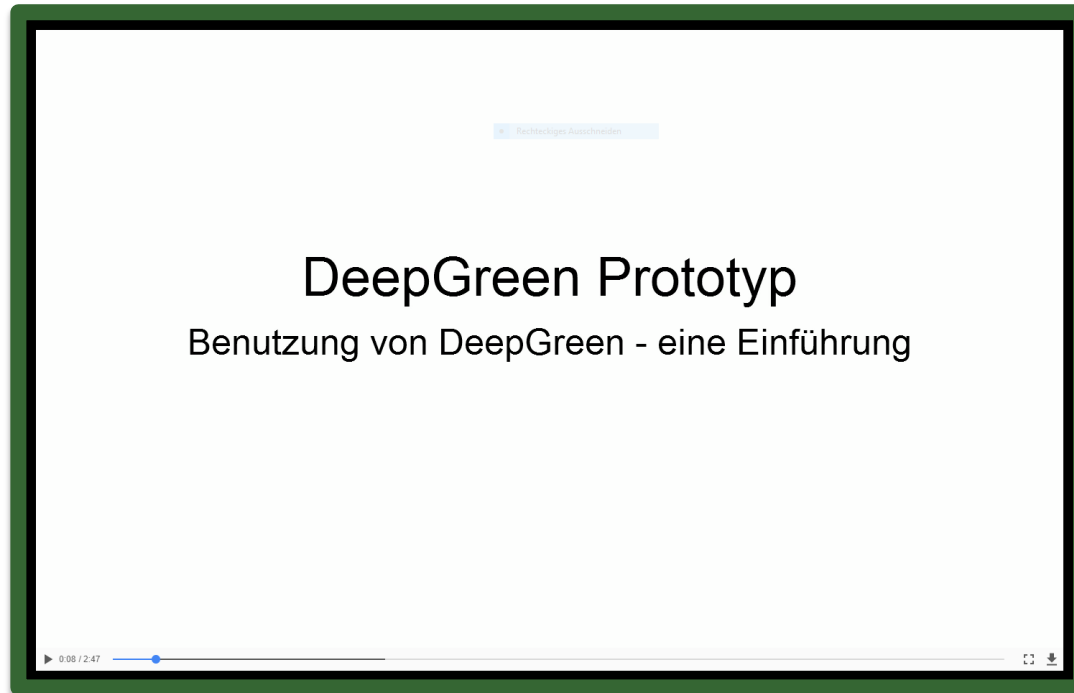
- ▶ Weitgehende Nachnutzung bestehender Konzepte und verfügbarer Software (→ Jisc, Basisversion des Publications Router)
- ▶ Benutzerzentrierter Service: Abwicklung über Konten mit Profil-Angaben
- ▶ Daten werden gemäß (Verlags-)Kontenprofil angenommen, gegen Kontenprofil von Bibliothek abgeglichen und entsprechend bereitgestellt.
- ▶ Design als **Reines Push-Forward-System** (*kein* Dubletten-Check!)
- ▶ Kein Dark Archive: Datenspeicherung erfolgt mit beschränkter Verweildauer (aktuell: 3 Monate)
- ▶ Idee: Service schickt **Notifications/Benachrichtigungen** über eingegangene Artikel, die dann im nächsten Schritt abgeholt werden können.

Der Kern – Die Zuordnungslogik

- ▶ Zweistufige Berechtigungslogik
- ▶ 1. Lizenzüberprüfung anhand der ISSN und des Publikationsjahrs (Jahrgang): Titellisten und Jahrgänge werden aus der EZB-Datenbank periodisch ermittelt
- ▶ 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)

Wie sieht das genau aus?



[Screencast](#)

Aus der Perspektive von
Verlagen (2,47 Minuten)

[Screencast](#)

Aus der Perspektive von
Bibliotheken (4,06 Minuten)

Erste Ergebnisse



Stabiler Testbetrieb

- ▶ Stabiler Testbetrieb (seit 01.05.2017)
- ▶ Akzeptable Durchsatzrate für prototypische Implementation
Lasttest: 21.636 Artikel (*nur Metadaten*) in 48h analysiert (→ ~10 docs/min)
- ▶ Testdaten wurden verarbeitet von: **European Mathematical Society, Karger, Oxford University Press, Royal Society of Chemistry, SAGE Publications**
- ▶ Verschiedene Repositorien wurden testweise erfolgreich angebunden:
OPUS 4 (FAU), **DSpace** (TU Berlin), **ESciDoc/Pubman** (GFZ)
- ▶ **Anbindung an die EZB** (Vertragsinformationen) funktioniert prototypisch.
- ▶ Artikel-Zuordnung erfolgt korrekt.
- ▶ In Vorbereitung für einen Betrieb wurden **Standard-Konten** für ca. 250 wissenschaftliche Einrichtungen in Deutschland eingerichtet.

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

Perspektiven: geplante 2. Förderphase 2018-2019



DeepGreen als Dienst etablieren

- ▶ DeepGreen vom Projektstadium in einen bundesweiten Betrieb überführen
- ▶ Nachhaltiges Geschäfts- und Betriebsmodell entwickeln
- ▶ 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

DeepGreen ausweiten auf neue Lizenzkontexte

- ▶ Konsortiallizenzen ohne explizite grüne Open-Access-Komponente aufnehmen
 - ▶ Liefern Verlage auch Post- und Preprints?
- ▶ FID-Lizenzen aufnehmen
 - ▶ FID-Lizenzierungsgrundsätze folgen vielfach den Allianz-Richtlinien.
- ▶ Goldene Open-Access-Artikel aufnehmen aus
 - ▶ reinen OA-Zeitschriften
 - ▶ Transformationsverträgen
 - ▶ DEAL-Verträgen ...

DeepGreen ausweiten für neue Abnehmer

- ▶ Fachrepositorien und Forschungsinformationssysteme prototypisch aufnehmen
- ▶ Welche Datenelemente/Verfahren können zur fachlichen Identifikation genutzt werden?
- ▶ Welche rechtlichen Zusammenhänge sind zu beachten?
- ▶ Welche technischen Spezifikationen, Schnittstellen und Workflows sind notwendig?

Zusammenfassung



DeepGreen – Projektmeilensteine im Überblick

2016

- ✓ Projektstart Januar 2016, [Publikation des Projektantrags](#)
- ✓ Vortrag auf dem Bibtag 2016 mit [Veröffentlichung des Textes](#) in einer OA-Zeitschrift
- ✓ Konzeption/Veröffentlichung [DeepGreen-Metadatenschema](#), Version 1.1
- ✓ Workshop mit ExpertInnen
- ✓ Datendrehscheibe Alphaversion freischalten inklusive einem Konzept für Rechteverwaltung

2017

- ✓ Datendrehscheibe Version 2.0 freischalten: Die Datenabgabe an die Repositorien erfolgt im passenden Importformat
- ✓ Datendrehscheibe Version 3.0 freischalten: Standard-Konten eingerichtet
- ✓ Kennzahlenermittlung mit einer Stichprobe von Publikationen aller Projektbeteiligten (Jahr 2015)
- ✓ Öffentlicher Abschlussworkshop 14.11.2017, Präsentation des Projektstandes und Perspektiven

DeepGreen – Der technische Steckbrief

- ▶ Programmierung auf Basis der ersten Version des Jisc-Publications Event Router
(= fork von <https://github.com/jiscPER>)
- ▶ *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**
- ▶ Software: <https://github.com/OA-DeepGreen>
- ▶ Lizenz: [Apache License, Version 2.0](#)

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

DeepGreen – Projektkonsortium

Kooperativer Bibliotheksverbund
Berlin-Brandenburg (KOBV) –
Projektleitung

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

Bayerische Staatsbibliothek (BSB)
Bibliotheksverbund Bayern (BVB)

- Dr. Klaus Ceynowa
- Dr. Hildegard Schäffler
- Dr. Ortwin Guhling
- Michael Kassube
- Matthias Groß

Technische Universität Berlin (TUB),
Universitätsbibliothek

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

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

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

Helmholtz Open Science
Koordinationsbüro am Deutschen
GeoForschungsZentrum (GFZ)

- Roland Bertelmann
- Heinz Pampel
- Kaja Scheliga
- Tobias Höhnow

BVB BibliotheksVerbund
Bayern

BSB Bayerische
StaatsBibliothek
Information in erster Linie

FAU FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG



kobv

HELMHOLTZ
| GEMEINSCHAFT
Open Science

WISSEN IM ZENTRUM
UNIVERSITÄTSBIBLIOTHEK
Technische Universität Berlin

Deepgreen – Alte und neue Verlagspartnerschaften

seit Projektbeginn

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

seit Sommer 2017

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

DeepGreen – Alte und neue Partnerschaften

... institutionelle Repositorien

- ▶ Projektbeteiligten FAU, GFZ und TU Berlin

... disziplinäre Repositorien

- ▶ DIPF Frankfurt am Main (peDOCs)
- ▶ Staatsbibliothek zu Berlin (<intR>²Dok)
- ▶ 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)

Vielen Dank für Ihre Aufmerksamkeit!

Kontakt zum Projektmanagement: info@oa-deepgreen.de

AnsprechpartnerInnen:

[Julia Alexandra Goltz-Fellgiebel](#) (Projektmanagement)

[Dr. Thomas Dierkes](#) (Entwicklung)

Projektwebseite mit Informationen zum Projekt: <https://deepgreen.kobv.de/>

Prototyp der DeepGreen-Datendrehscheibe: <http://www.oa-deepgreen.de/>

Impressum

- ▶ Die Präsentation steht unter einer CC BY-Lizenz: <https://creativecommons.org/licenses/by/4.0/deed.de>.
- ▶ Fotos: Heinrich Hamann, CCO (<https://creativecommons.org/publicdomain/zero/1.0/>). Datengeber: Museum für Kunst und Gewerbe Hamburg, Fundort: <https://www.europeana.eu/portal/en/explore/galleries/heinrich-hamann>
- ▶ Animated GIF, Folie 17: Emilie Trébuchet, "Paris plage", source material via Europeana <http://bit.ly/2z3t97d>, Trove <http://bit.ly/2xj5K71> & DigitalNZ <http://bit.ly/2iknFKC>

