

# Anwendung von OPUS für Forschungsdaten – das Beispiel eines außeruniversitären Forschungsinstituts

Wolfgang Peters-Kottig, Kooperativer Bibliotheksverbund Berlin-Brandenburg

## Zusammenfassung:

OPUS wurde als Dokumentenserver konzipiert und ursprünglich nicht für den Nachweis von Forschungsdaten angelegt. Mit vergleichsweise geringem Aufwand können dennoch Forschungsdaten als Supplementary Material zu Zeitschriftenbeiträgen publiziert werden. Auch reine Forschungsdatenpublikationen können via OPUS veröffentlicht werden, selbst wenn die gegenwärtig auf Textpublikationen zugeschnittene Metadatenerfassung weniger gut für die Beschreibung reiner Forschungsdaten geeignet ist. Ein Unterschied zu anderen Systemen und Anwendungsszenarien besteht in der Beschränkung, großvolumige Daten nicht innerhalb von OPUS selbst vorzuhalten. Der für das institutionelle Repository des Zuse-Instituts Berlin konzipierte pragmatische Lösungsansatz ist in OPUS ohne zusätzliche Programmierung umsetzbar und skaliert für beliebig große Forschungsdatensätze.

## Summary:

Primarily, OPUS has been designed as a document server and originally was not intended to serve as a repository for research data. However, it is possible to publish research data as supplementary material to a journal article with comparatively simple means. Independent data sets may also be published via OPUS, although the metadata description in OPUS, which has been customized for text publications, is less suited for the description of research data. What makes it different from other systems, however, is the limitation of storage capacity. Large data volumes should not be stored in OPUS. The pragmatic approach designed for the institutional repository of the Zuse Institute Berlin can be implemented in OPUS without additional programming, and is well suited for data sets of any size.

**Zitierfähiger Link (DOI):** <https://doi.org/10.5282/o-bib/2018H2S79-91>

**Autorenidentifikation:** Peters-Kottig, Wolfgang: GND 128736712; ORCID: <http://orcid.org/0000-0003-4486-2422>

**Schlagwörter:** Repositorien; Forschungsdaten; Publikationsprozess

## 1. Einleitung

In Deutschland ist OPUS derzeit die am weitesten verbreitete Software für den Betrieb von Repositorien an Hochschulen und Kultureinrichtungen, die eigene Publikationen im Open Access in einem instituts- oder fachbezogenen Repository zur Verfügung stellen wollen.<sup>1</sup> OPUS 4 ist eine vergleichsweise einfach zu administrierende Software, die auch von vielen kleinen Einrichtungen

<sup>1</sup> Laut Directory of Open Access Repositories (zuletzt geprüft am 19.06.2018, [www.openoar.org/](http://www.openoar.org/)) werden mit Stand Februar 2018 von 203 gelisteten Repositories in Deutschland 81 mit OPUS betrieben. Die Statistik der KOBV-Zentrale zählt rund 100 OPUS-Installationen in Deutschland (zuletzt geprüft am 19.06.2018, <http://www.kobv.de/entwicklung/software/opus-4/referenzen/>).

mit beschränkteren Ressourcen eingesetzt wird.<sup>2</sup> Das institutionelle Repositorium des Konrad-Zuse-Zentrum für Informationstechnik Berlin, kurz Zuse-Institut Berlin oder ZIB, wird ebenfalls mit der Software OPUS betrieben. Seitdem das Thema Forschungsdatenmanagement auch im außeruniversitären Bereich immer größere Bedeutung erlangt und die Relevanz von Forschungsdatenpublikationen zunehmend wahrgenommen wird, gibt es am ZIB Planungen, den Forschenden die Möglichkeit zur Veröffentlichung der eigenen Daten zu geben. Von der Geschäftsleitung des ZIB wurde der Beschluss gefasst, die Möglichkeiten einer Forschungsdatenpublikation in OPUS zu evaluieren und Umsetzungswege zu suchen. Dazu wurde eine gemeinsame Arbeitsgruppe von KOBV (Kooperativer Bibliotheksverbund Berlin-Brandenburg), der ZIB-Abteilung IT & Data Services und interessierten Wissenschaftler/innen des ZIB gegründet. Da die Weiterentwicklung von OPUS 4 seit 2010 maßgeblich von der Verbundzentrale des KOBV am ZIB verantwortet wird, bot sich an, die Planung und eine testweise Implementierung von Funktionalitäten zusammen mit dem OPUS-Team in der haus-eigenen ZIB-OPUS-Instanz durchzuführen.<sup>3</sup> Erste Ergebnisse der Arbeiten und ein Ausblick auf das Thema digitale Langzeitarchivierung werden im Folgenden dargestellt.

## 2. OPUS als Dokumentenserver des ZIB

Das ZIB ist ein außeruniversitäres Forschungsinstitut für angewandte Mathematik und Informatik in Trägerschaft des Landes Berlin. Die Forschungsergebnisse der vielfältigen Kooperationsprojekte werden im ZIB-Repositorium vor allem in Form von Preprints und Postprints der wissenschaftlichen Fachbeiträge auf Open-Access-Basis veröffentlicht. Daneben gehören Reports und technische Manuskripte sowie Abschlussarbeiten der kooperierenden Berliner Universitäten zu den Inhalten des Dokumentenservers (Abb. 1).

Software-Basis für das Repositorium ist das Release OPUS 4.6 mit einzelnen ZIB-spezifischen Anpassungen in der Administration und Benutzeroberfläche. So werden nicht nur die OPUS-„Standardfelder“ mit Metadaten zur inhaltlichen Erschließung befüllt.<sup>4</sup> Zusätzlich werden ergänzende Informationen gesammelt, die eine Einordnung von Dokumenten nach weiteren Kriterien erlauben, beispielsweise danach, ob die betreffende Publikation ein Peer-Review-Verfahren durchlaufen hat. Im Publikationsprozess erfolgt ein automatischer Export der Metadaten in die jeweilige persönliche Publikationsliste auf der ZIB-Webseite.<sup>5</sup>

Der administrative Betrieb von OPUS und die Beratung der Nutzer/innen zählen wie in vergleichbaren Einrichtungen zu den Aufgaben der Institutsbibliothek, wobei das Hosting der Anwendung durch die

2 OPUS 4 ist eine Open-Source-Software unter der GNU General Public License, zuletzt geprüft am 19.06.2018, <https://opus4.kobv.de>. Siehe auch das OPUS-Handbuch unter, zuletzt geprüft am 19.06.2018, <http://www.opus-repository.org/userdoc/>.

3 Zuletzt geprüft am 19.06.2018, <https://opus4.kobv.de/opus4-zib/>

4 Liste der 50 Standard-Metadatenfelder für Dokumente in OPUS: zuletzt geprüft am 19.06.2018, <http://www.opus-repository.org/userdoc/reference/fields.html>. Diese Metadatenfelder basieren auf der DINI-Empfehlung von 2010: DINI Arbeitsgruppe Elektronisches Publizieren, Deutsche Nationalbibliothek und Bibliotheksservice-Zentrum Baden-Württemberg, *Gemeinsames Vokabular für Publikations- und Dokumenttypen*, 2010, <http://doi.org/10.18452/1492>.

5 Beispiel einer automatisch aus den OPUS-Einträgen gespeisten Publikationsliste im ZIB (unter Publications): zuletzt geprüft am 19.06.2018, <http://www.zib.de/members/baum>.

KOBV-Zentrale übernommen wird. Hierbei kann die technische Infrastruktur des ZIB genutzt werden.<sup>6</sup> Neue Dokumente können nach Login in OPUS in Eigenregie von den Wissenschaftler/inne/n hochgeladen werden. Vor der Freigabe zur Veröffentlichung erfolgt eine Sichtprüfung der eingegebenen Metadaten durch die Institutsbibliothek und es werden Hinweise zur Lizenzvergabe gegeben, z.B. die Empfehlung, Creative-Commons-Lizenzen (bevorzugt CC-BY oder CC0) einzusetzen.



Dokumenttyp	
ZIB-Report	(1529)
Artikel	(1318)
Konferenzbeitrag	(932)
Masterarbeit	(220)
Buchkapitel	(150)
Dissertation	(119)
Sonstiges	(80)
Buch	
(Monographie)	(60)
Beitrag	
Sammelband	(57)
Bachelorarbeit	(48)
ZIB-Jahresbericht	(32)
Poster	(32)
Habilitation	(15)
Bericht	(7)
Konferenzband	(6)
Forschungsdaten	(1)

Abb. 1: Dokumenttypen der Publikationen in der OPUS-Instanz des ZIB. Die Zahlen geben die Anzahl der zugehörigen Dokumente an (Stand Februar 2018).

### 3. Forschungsdatenmanagement am ZIB

Am ZIB werden große Mengen von Forschungsdaten mit unterschiedlicher Komplexität in einem großen Fächerspektrum mit wissenschaftlichen und industriellen Kooperationspartnern erzeugt. Schwerpunkte liegen in den Lebens-, Natur- und Ingenieurwissenschaften mit der Modellierung, Simulation und Optimierung etwa von Verkehrs- und Versorgungsnetzen und in der Digitalen Medizin. Neben den Herausforderungen der reinen Datenhaltung wird hier die Etablierung eines expliziten Forschungsdatenmanagements (FDM) immer bedeutender.<sup>7</sup> Bei der Identifizierung der konkreten Herausforderungen für das ZIB ist der Organisationsstatus als verhältnismäßig kleine

<sup>6</sup> Das ZIB beherbergt einen Supercomputer mit entsprechend leistungsfähiger Hardware-Ausstattung.

<sup>7</sup> Einen guten Überblick zum aktuellen Stand des Forschungsdatenmanagements in Deutschland geben die Konferenzbeiträge zu den E-Science-Tagen 2017: Jonas Kratzke und Vincent Heuveline, Hrsg., *E-Science-Tage 2017: Forschungsdaten managen* (Heidelberg: heiBOOKS, 2017), <http://doi.org/10.11588/heibooks.285.377>

außeruniversitäre Forschungseinrichtung wichtig, die zudem keiner Wissenschaftsorganisation angehört.<sup>8</sup> Die Abteilungen innerhalb des ZIB kooperieren über sogenannte Brückenprojekte auch disziplinübergreifend. In einem dieser Brückenprojekte haben unter dem Titel RDM@ZIB vier Wissenschaftler/innen aus drei Abteilungen in Eigeninitiative eine Ad-hoc-Arbeitsgruppe gegründet, die Vorschläge für ein institutionelles FDM entwickelt und umsetzt.<sup>9</sup>

Als eine wichtige Herausforderung und zentrale Chance in Bezug auf ein institutionelles FDM wurde im Rahmen zweier interner Workshops mit Arbeitsgruppenleitern die Datenpublikation identifiziert. Dabei wurde deutlich, dass aufgrund zu geringer Anreize zwar weiterhin wenig Eigeninitiative der Forscherinnen und Forscher im Haus existiert, Daten zu publizieren. Gleichzeitig wäre aber eine Mehrzahl durchaus bereit, eigene Daten zu veröffentlichen unter der Maßgabe, dass Unterstützung im Institut erfolgt und möglichst wenig zusätzliche Arbeitszeit investiert werden muss.

Eine systematische Befragung aller AGs im ZIB ist noch nicht erfolgt, die Ergebnisse der Workshops und einzelner, unstrukturierter Interviews mit Arbeitsgruppenleitern geben aber einen Eindruck der bestehenden Probleme. Derzeit besteht eine Datenveröffentlichung am ZIB in Aufzeichnungen auf persönlichen oder Projekt-Homepages, die keine recherchierbare, standardisierte Metadatenbeschreibung, keine Versionierung und keine nachhaltige Bereitstellung erlauben. Für den Datenaustausch mit Dritten werden häufig auf Arbeitsgruppenebene Server betrieben und es werden (versteckte) Webseiten erzeugt, auf denen die Daten von den Partnern wieder abgeholt werden können. Hier fehlt es – abgesehen von hardware- und softwarebedingten Anforderungen – an Unterstützung bei der Metadatenbeschreibung. Dies betrifft sowohl deskriptive Metadaten im Fachkontext als auch technische und administrative Metadaten. Authentifizierungs-/Autorisierungsmechanismen und eindeutige Identifizierung von Objekten in allen Stationen des Datenlebenszyklus werden ebenfalls noch nicht standardmäßig eingesetzt. Die bisherige ZIB-Praxis bezogen auf nachhaltige Verfügbarkeit von Projektergebnissen als Forschungsdaten besteht überwiegend darin, nach Ende eines Forschungsprojekts oder dem Ausscheiden verantwortlicher Personen (insbesondere Doktorand/innen), unpublizierte Daten auf einen gesonderten Archivbereich des Dateisystems auszulagern. Die Übertragbarkeit dieser Bestandsaufnahme auf die Archivierungsbedarfe von originären Forschungsprojekten am ZIB ist aus technischer, organisatorischer und finanzieller Perspektive noch ungeklärt.

Im Ergebnis hat die AG RDM@ZIB zwei vorrangige Herausforderungen aus dem Status quo identifiziert:

- Beseitigung der Defizite bei Datenpublikationsmöglichkeiten, Erhöhung der Publikationsanreize und verbesserte Information über vorhandene Optionen.
- Standardisierung von Metadatenbeschreibungen und Schaffung geregelter Workflows zu Datentransfer und -analyse, Datenablage, Archivierung sowie Datenweitergabe.

---

8 Am ZIB sind rund 230 Personen beschäftigt.

9 Eine Initialzündung für den Start des institutionellen FDM war die ZIB-Beteiligung am DFG-geförderten Projekt EWIG; s. die im Projekt entstandene Broschüre: Roland Bertelmann et al., *Einstieg ins Forschungsdatenmanagement in den Geowissenschaften* (Potsdam, 2014), <http://doi.org/10.2312/lis.14.01>.

Die zweite Herausforderung ist deutlich komplexer und wird in diesem Kontext (mit Ausnahme der Archivierung) nicht näher beleuchtet, weil sie erst in einem späteren Schritt angegangen werden soll. Fraglich ist z.B., wo die Grenze zwischen Aufgaben des institutionellen FDM und Hardwarebereitstellung als Aufgabe der „IT-Infrastruktur“ bzw. dem Forschungsgegenstand des ZIB liegt.<sup>10</sup>

Die Datenpublikation mit OPUS wurde als wichtigste Aufgabe zuerst angegangen. Ein wichtiger begrenzender Faktor ist hier die Organisationsstruktur des ZIB als kleine Forschungseinrichtung mit geringen Personalkapazitäten für Service-Aufgaben. Deshalb war es wichtig, an einer praktikablen Lösung zu arbeiten, die in der ersten Annäherung möglichst ohne zusätzliche Entwicklungsarbeiten OPUS „datengerecht“ anpassen kann und nur geringe Änderungen im Workflow vorsieht. Noch nicht geklärt ist für die ZIB-Instanz die Frage, ob und in welchem Umfang eine weitergehende, vielleicht sogar fachliche Qualitätsprüfung vor Veröffentlichung erfolgen kann.<sup>11</sup> Ein Upload soll auch bei Forschungsdaten möglichst eigenständig durch die Beitragenden selbst erfolgen können, wie es bereits bei Textpublikationen in der derzeitigen OPUS-Instanz der Fall ist. Parallel wurde ein initialer Beratungsservice durch die AG aufgebaut, um die Interessentinnen und Interessenten bei allen Fragen zum Thema Data Publishing zu unterstützen – konkret auch bei der Erarbeitung von Datenmanagementplänen.

#### 4. Publikation von Forschungsdaten mit OPUS

Der vergleichsweise geringe administrative Aufwand wird im ZIB als großer Vorteil von OPUS gesehen. Aus den langjährigen Erfahrungen der KOBV-Verbundzentrale – nicht zuletzt im Austausch mit externen Repository-Betreibern – wurde in der AG der Schluss gezogen, dass komplexe Software-Alternativen wegen eines zu erwartenden hohen Konfigurations- und Wartungsaufwands nicht in Frage kommen würden. Gleichzeitig sollte ein Betrieb des Repositoriums weder physikalisch noch „ideell“ ausgelagert werden, so dass etwa Zenodo oder figshare als potentielle Publikationsorte ebenfalls ausschieden. Als Vorteil beim Einsatz von OPUS 4 kann gelten, dass die Forscherinnen und Forscher bereits bekannte Workflows nutzen können. Zudem kann lokal vorhandene Infrastruktur eingesetzt werden. Publikationen und Forschungsdaten können an einer zentralen Stelle im ZIB verknüpft und bereitgestellt werden. OPUS 4 besitzt zudem ein flexibles, auch lokal erweiterbares Metadatenschema.

Ideen und Umsetzungsvorschläge für die Verzeichnung von Forschungsdaten mit OPUS bestehen schon seit längerem, etwa bei den Entwicklungspartnern von OPUS, der UB Stuttgart und dem Bibliotheksservice-Zentrum Baden-Württemberg.<sup>12</sup> Basierend auf diesen Vorarbeiten wurden Gespräche

---

10 Die Bereitstellung einer Plattform für hochperformante Analyse und Verarbeitung von großvolumigen Daten in Echtzeit wird von der AG RDM@ZIB nicht zum Aufgabenbereich des institutionellen FDM gezählt. Die Metadatenbeschreibung in diesem Kontext aber sehr wohl – dafür gibt es jedoch noch kein tragfähiges Konzept am ZIB.

11 Wie von Van Tuyl und Whitmire (2016) empfohlen, sollten „Repository Manager“ zumindest prüfen können, ob hinreichend fachspezifische Metadaten vorhanden und die Forschungsdaten grundsätzlich nachnutzbar sind. Vgl. Steven Van Tuyl und Amanda L. Whitmire, „Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia,” *PLoS ONE* 11, Nr. 2 (2016), e0147942: 14, <https://doi.org/10.1371/journal.pone.0147942>.

12 Matthias Schulze, „Virtuelle Forschungsumgebungen und Forschungsdaten für Lehre und Forschung: Informationsinfrastrukturen für die (Natur-)Wissenschaften,” in *Semantic Web & Linked Data : Elemente zukünftiger*

mit dem OPUS-Team beim KOBV geführt, um technische Grundlagen und funktionale Anforderungen zu klären und eine Testumsetzung zu implementieren. Die folgende Definition der Funktionalität von Datenpublikationen diene als Grundlage für die Implementierung in der ZIB-OPUS-Instanz:

**a) Publikation als Supplementary Material.** In OPUS wird bei einer Textpublikation auf zusätzlich vorhandenes Supplementary Material verwiesen. Dazu steht ein neues Metadatenfeld „URL der Forschungsdaten“ zur Verfügung sowie ein neues Metadatenfeld „Bemerkungen zu Forschungsdaten“ mit der Möglichkeit, die verlinkte Ressource zu beschreiben. Das Feld „URL der Forschungsdaten“ verweist entweder zu einer externen Quelle, wie einem Forschungsdaten-Repository/einer Verlagsseite/einem Cloudspeicher, oder zu einem eigenständigen Forschungsdaten-Eintrag in OPUS mit eigener Landingpage, wenn die Forschungsdaten auf einem ZIB-Server vorgehalten werden (und über einen persistenten Identifier, z.B. DOI verfügen). Auf dieser zusätzlichen Landingpage ist ein Enrichment-Feld „Data Download“ vorhanden, das bei Anklicken den eigentlichen Download ermöglicht.

**b) Eigenständige Forschungsdatenpublikation.** Wenn keine zugehörige Textpublikation in OPUS vorhanden ist, kann auch ein eigenständiges Dokument „Forschungsdaten“ angelegt werden. Die Forschungsdaten werden entweder außerhalb des ZIB bei einer beliebigen Quelle oder wie unter a) genannt auf einem eigenen ZIB-Server vorgehalten. Im Feld „URL der Forschungsdaten“ wird dann entweder auf die externe Ressource verlinkt oder unter „Data Download“ das Herunterladen vom ZIB-Server ermöglicht. Eine Beschreibung der Ressource ist im Feld „Bemerkungen zu Forschungsdaten“ möglich. Mit dem Feld „URL“ ist auch ein Rückverweis auf z.B. später erschienene Textpublikationen zum Forschungsdatensatz möglich.

Der erste Arbeitsschritt bestand in der Festlegung notwendiger Erweiterungen zum Metadatenchema von OPUS. Anhand von Supplementary Material zu einer aktuellen Publikation zweier ZIB-Mitarbeiter wurden auf Basis des DataCite-Metadatenchemas und im Abgleich mit weiteren Datenschemata wie dem RADAR Metadata Schema die Felder identifiziert, für die kein Mapping von vorhandenen Standardfeldern in OPUS sinnvoll oder möglich ist.<sup>13</sup> Für diese Felder können in OPUS vergleichsweise einfach (ohne Programmieraufwand) sogenannte Enrichment-Felder in der Administrationsschicht angelegt werden. Diese Möglichkeit gilt für alle OPUS 4-Betreiber. Abb. 2 listet für die ZIB-Instanz vorläufig neu definierte Enrichment-Felder auf. Hier gilt es abzuwägen, welche Felder wirklich generisch für alle Fachdisziplinen gelten und welche eher fachspezifisch sind. Weil OPUS beim Anlegen von Enrichments sehr flexibel ist, wurde vorläufig entschieden, bei Bedarf

*Informationsinfrastrukturen; Proceedings der 1. DGI-Konferenz, 62. Jahrestagung der DGI, Frankfurt am Main, 7.-9. Oktober 2010, Tagungen der DGI 14, hrsg. Marlies Ockenfeld (Frankfurt am Main, 2010), 172, <http://dx.doi.org/10.18419/opus-6344>.*

Einen aktuellen Umsetzungsvorschlag des BSZ zur Forschungsdatenpublikation in OPUS hat Friederike Gerland auf dem OPUS-Workshop beim Bibliothekskongress 2017 präsentiert: Friederike Gerland, „Forschungsdatenmanagement in OPUS 4“ (Vortrag auf dem 106. Deutschen Bibliothekartag in Frankfurt am Main 2017), Vortragsfolien, <http://nbn-resolving.de/urn:nbn:de:0290-opus4-30859>.

- 13 DataCite Metadata Working Group, *DataCite Metadata Schema Documentation for the Publication and Citation of Research Data*. Version 4.1 (2017), <http://doi.org/10.5438/0014>.  
RADAR Projektteam, *Dokumentation des deskriptiven RADAR Metadatenchemas*. Version 0.9 (FIZ Karlsruhe, 2017), zuletzt geprüft am 14.05.2018, [https://www.radar-service.eu/sites/default/files/RADAR\\_Metadaten\\_Dokumentation\\_v09.pdf](https://www.radar-service.eu/sites/default/files/RADAR_Metadaten_Dokumentation_v09.pdf)

auch sehr spezifische Erweiterungen zuzulassen.<sup>14</sup> Gleichzeitig wurde ein neuer Dokumenttyp „Forschungsdaten“ (Research Data) in der OPUS-Administration angelegt, der das Browsen in der Instanz bezogen auf Forschungsdaten ermöglicht (s. auch Abb. 1).

Enrichmentfeld	Benennung en.	Benennung dt.
EnrichmentScientificNote	Notes to research data	Bemerkungen zu den Forschungsdaten
EnrichmentScientificName	Scientific name	Wissenschaftlicher Name
EnrichmentScientificDateCollected	Date data collected	Datum der Datenerhebung
EnrichmentScientificDateCreated	Date data created	Datum der Datenerstellung
EnrichmentScientificGeolocation	Geolocation	Ort der Datenerhebung
EnrichmentScientificResourceTypeGeneral	ResourceTypeGeneral	ResourceTypeGeneral
EnrichmentRelatedIdentifier	Related identifier	Related Identifier
EnrichmentScientificDataDownload	Research data download	URL zum DataDownload
EnrichmentSoftwareDescription	Software description	Software Description

Abb. 2: Neu definierte Metadatenfelder (Enrichments) für Forschungsdaten in OPUS.

Bei der Verwendung von OPUS für Forschungsdatenpublikationen gibt es in der aktuellen Version folgende Einschränkungen bzw. Herausforderungen.

- Aufgrund der Architektur von OPUS ist es nicht sinnvoll, große Datenvolumina innerhalb von OPUS zu verwalten. Der Upload mehrerer 100 GB – wie bei Forschungsdaten im ZIB durchaus gängig – ist nur auf Umwegen möglich weil die Administration größtenteils über eine Web-Oberfläche erfolgt.
- Eine Auslagerung der eigentlichen Download-Daten auf einen externen Server (außerhalb der OPUS-Verwaltung) ist sinnvoll, um Performance-Probleme mit großvolumigen Forschungsdaten zu vermeiden. In diesem Fall kann aber das Rollen- und Rechtesystem von OPUS nicht genutzt werden. Diese Funktionalität sollte von einem eigenständigen Authentifizierungssystem abgedeckt werden.
- Eine Vorschau ist nicht möglich. Forschungsdaten können nicht direkt im OPUS-Repository geöffnet oder gestreamt werden. Stattdessen muss immer die gesamte Datei lokal heruntergeladen werden.
- Für bestimmte Datentypen, insbesondere Software, ist es wichtig, dass auf alle Versionen verwiesen werden kann (Vorläuferversion, Nachfolgeversion, Landingpage zum Haupteintrag für die Software). OPUS bietet aktuell keine detaillierte Versionierung für publizierte Objekte.

<sup>14</sup> Das Feld „Scientific name“ ist beispielsweise sehr spezifisch. Es bezeichnet in dem zugrundeliegenden Datensatz aus der Bioinformatik den taxonomischen Namen einer Tierart.

- Wissenschaftler/innen wollen zu Recht die arbeitsintensive, manuelle Mehrfacheingabe von Metadaten wie Autor/Principal Investigator, Titel, Förderkennzeichen u.ä., in Forschungsinformationssystemen, Repositories, Projektmanagementsoftware etc. vermeiden. Dieser Mehraufwand lässt sich derzeit mangels Schnittstellen zwischen heterogenen Systemen kaum vermeiden.

Die genannten Herausforderungen wurden als unkritisch für den gewünschten Einsatz von OPUS erachtet oder können umgangen werden. Die Performance-Einschränkung bei der Verwaltung großer Datenmengen wird vermieden, wenn Forschungsdatenobjekte (Dateien) grundsätzlich außerhalb von OPUS in einer eigenen Infrastruktur verwaltet werden. Auf der Landingpage (Frontdoor) in OPUS selbst wird nur ein Download-Link hinterlegt, der auf einen unabhängigen Downloadserver innerhalb des ZIB verweist (Abb. 3). Hilfreich ist hier die spezifische Organisation des ZIB, welches ein eigenes Rechenzentrum inklusive Hardware-Infrastruktur in Form der Abteilung IT & Data Services beherbergt. Somit stellte sich vor Ort nicht das Problem der Vertrauenswürdigkeit und Nachhaltigkeit eines extern betriebenen Downloadservers.<sup>15</sup>

Alternativ kann an dieser Schnittstelle auch ein separates Rollen- und Rechtemanagement vorgeschaltet werden, das im einfachsten Fall eine Passwortabfrage aufruft oder an ein Verfahren zur verteilten Authentifizierung und Autorisierung wie Shibboleth angebunden werden kann.<sup>16</sup> Eine Einschränkung des Zugangs im Open Access gilt es allerdings sorgfältig abzuwägen. Das ZIB verfolgt eine Open-Access-Strategie, die sich etwa in der langjährigen Praxis niederschlägt, (Zweit-)Veröffentlichungen von ZIB-Angestellten als sogenannte ZIB-Reports im Open Access herauszugeben.<sup>17</sup> Es wird angestrebt, dass auch Forschungsdaten möglichst nicht mit einem Embargo oder anderen Restriktionen versehen werden.<sup>18</sup> Gleichzeitig ist anzumerken, dass ein dem Download vorgeschaltetes Rechtemanagement auch zur Vereinfachung eines Peer-Review-Verfahrens beitragen kann. Gutachtern einer Datenpublikation lässt sich im Rahmen des Review-Prozesses auf diesem Weg flexibel der Zugriff auf Daten im Repository freischalten.

Die oben genannte eingeschränkte Möglichkeit zur Versionsangabe wird in Kauf genommen weil Aktualisierungen mit Verweisen auf Vorgänger- und Nachfolgeversionen durch die OPUS-Administration manuell in Enrichment-Feldern eingetragen werden können. Dies ist zwar umständlich, aber jedenfalls hinreichend aussagekräftig (eine Lösung für Versionenverwaltung befindet sich in der Entwicklungsplanung für zukünftige OPUS-Releases). In der Metadatenanzeige auf der Landingpage zu dem Ende 2017 veröffentlichten Testdatensatz sind einige noch nicht umgesetzte kleinere Probleme sichtbar, die v.a. der von den Wissenschaftlern erwünschten kurzfristigen Veröffentlichung geschuldet

---

15 Für OPUS-Betreiber, die nicht auf entsprechende institutionelle Infrastrukturen zurückgreifen können, kommen ggf. Cloudspeicher-Services als Lösung in Frage.

16 Maßgeblich ist hier der weitverbreitete Dienst des Deutschen Forschungsnetzes DFN-AAI: zuletzt geprüft am 19.06.2018, <https://www.aai.dfn.de/>.

17 Auch die im Oktober 2017 verabschiedete Open-Access-Leitlinie des KOBV am ZIB spiegelt diesen Anspruch wider: Kooperativer Bibliotheksverbund Berlin-Brandenburg, *Open Access als strategische Leitlinie des KOBV*, ZIB-Report 17-54, 2017. <http://nbn-resolving.de/urn:nbn:de:0297-zib-65303>.

18 Aus forschungsethischen und/oder datenschutzrechtlichen Gründen kann eine Zugriffsbeschränkung nicht immer vermieden werden, etwa im Fall der am ZIB etablierten Verarbeitung medizinischer Forschungsdaten.




waren (Abb. 3 bzw. der Originaldatensatz unter <https://doi.org/10.12752/4.DKN.1.0>). Vorteilhaft ist wiederum die derzeit in Entwicklung befindliche automatische DOI-Vergabe in OPUS.<sup>19</sup> Bei der Vorbereitung der Veröffentlichung des Testdatensatzes mussten die Meldung an DataCite und das Eintragen des DOI noch manuell erfolgen.


Der Link auf die zugehörige Veröffentlichung ist noch nicht direkt anklickbar und rechts oben auf der Landingpage ist noch kein Download-Icon angezeigt, sondern der Dateidownload erfolgt über den Link im Feld „Research data download“, was etwas unübersichtlich erscheint.


Die grundlegende Aufgabe, eine Funktionalität in OPUS umzusetzen, mit der eine zitierfähige, persistente Forschungsdatenpublikation mit geringem personellem Aufwand von geschätzt weniger als einem Personenmonat möglich ist, war aus Sicht des ZIB aber sehr erfolgreich. Mehrere ZIB-Arbeitsgruppen haben Anfang 2018 bereits Interesse an einer kurzfristigen Datenpublikation geäußert. Der Ansatz soll nach dem Prinzip „Learning by Doing“ laufend optimiert und ausgebaut werden.<sup>20</sup> Für Einrichtungen, die OPUS 4 ebenfalls für ihr institutionelles Repository einsetzen, dürfte der Lösungsweg des ZIB für die Publikation von Forschungsdaten mit vergleichbar geringem Aufwand umsetzbar sein.

<sup>19</sup> Die Integration dieser Funktionalität ist für eines der kommenden OPUS 4-Releases noch in 2018 vorgesehen.

<sup>20</sup> Beispielsweise ergaben sich wertvolle Hinweise für das OPUS-Team für die sich derzeit in Entwicklung befindliche automatisierte DOI-Vergabe. Beim Mapping der OPUS-Metadaten auf das DataCite MD-Schema sollte etwa der Feldinhalt von „GeoLocation“, der bei Forschungsdaten den Ort der Original-Datenerhebung bezeichnet, nicht mit dem Verlagsort belegt werden.

 Automated Segmentation of Complex Patterns in Biological Tissues: Lessons from Stingray Tessellated Cartilage (Supplementary Material)



 [David Knötel](#), [Ronald Seidel](#), [Paul Zaslansky](#), [Steffen Prohaska](#), [Mason N. Dean](#), [Daniel Baum](#)

 Supplementary data to reproduce and understand key results from the related publication, including original image data and processed data. In particular, sections from hyomandibulae harvested from specimens of round stingray *Urobatis halleri*, donated from another study (DOI: 10.1002/etc.2564). Specimens were from sub-adults/adults collected by beach seine from collection sites in San Diego and Seal Beach, California, USA. The hyomandibulae were mounted in clay, sealed in ethanol-humidified plastic tubes and scanned with a Skyscan 1172 desktop  $\mu$ CT scanner (Bruker  $\mu$ CT, Kontich, Belgium) in association with another study (DOI: 10.1111/joa.12508). Scans for all samples were performed with voxel sizes of 4.89  $\mu$ m at 59 kV source voltage and 167  $\mu$ A source current, over 360° sample 120 rotation. For our segmentations, the datasets were resampled to a voxel size of 9.78  $\mu$ m to reduce the size of the images and speed up processing. In addition, the processed data that was generated with the visualization software Amira with techniques described in the related publication based on the mentioned specimens.


EXPORT METADATA

[BibTeX](#) [RIS](#) [XML](#)

ADDITIONAL SERVICES

[Send a mail to the author of this document](#)

 **Metadaten**


<b>Author:</b>	<a href="#">David Knötel</a> , <a href="#">Ronald Seidel</a> , <a href="#">Paul Zaslansky</a> , <a href="#">Steffen Prohaska</a> , <a href="#">Mason N. Dean</a> , <a href="#">Daniel Baum</a>
<b>Document Type:</b>	Research data
<b>ResourceTypeGeneral:</b>	Dataset
<b>Year of first publication:</b>	2017
<b>Notes to research data :</b>	Supplementary data to reproduce and understand key results from the related publication, including original image data and processed data.
<b>Date data created:</b>	2017
<b>Geolocation:</b>	San Diego and Seal Beach, California, USA
<b>Scientific name :</b>	<i>Urobatis halleri</i>
<b>Related identifier:</b>	<a href="https://doi.org/10.1371/journal.pone.0188018">https://doi.org/10.1371/journal.pone.0188018</a>
<b>Software description :</b>	The data was processed with the visualization software Amira. See <a href="http://www.zib.de/software/tesseract-segmentation">http://www.zib.de/software/tesseract-segmentation</a> for information about Amira and how to download the Amira extension package created for this publication. See <a href="https://github.com/zibamira/tesseract-segmentation.git">https://github.com/zibamira/tesseract-segmentation.git</a> for the source code of the Amira extension package.
<b>Research data download:</b>	<a href="http://www.zib.de/ext-data/2017_Knoetel_PLOSONE-Supplementary-Material.zip">http://www.zib.de/ext-data/2017_Knoetel_PLOSONE-Supplementary-Material.zip</a>
<b>DOI:</b>	<a href="https://doi.org/10.12752/4.DKN.1.0">https://doi.org/10.12752/4.DKN.1.0</a>
<b>Licence (German):</b>	 <a href="#">Creative Commons - Namensnennung</a>

Abb. 3: Landingpage (in OPUS „Frontdoor“) zum ersten als Supplementary Material veröffentlichten Forschungsdatensatz im ZIB.

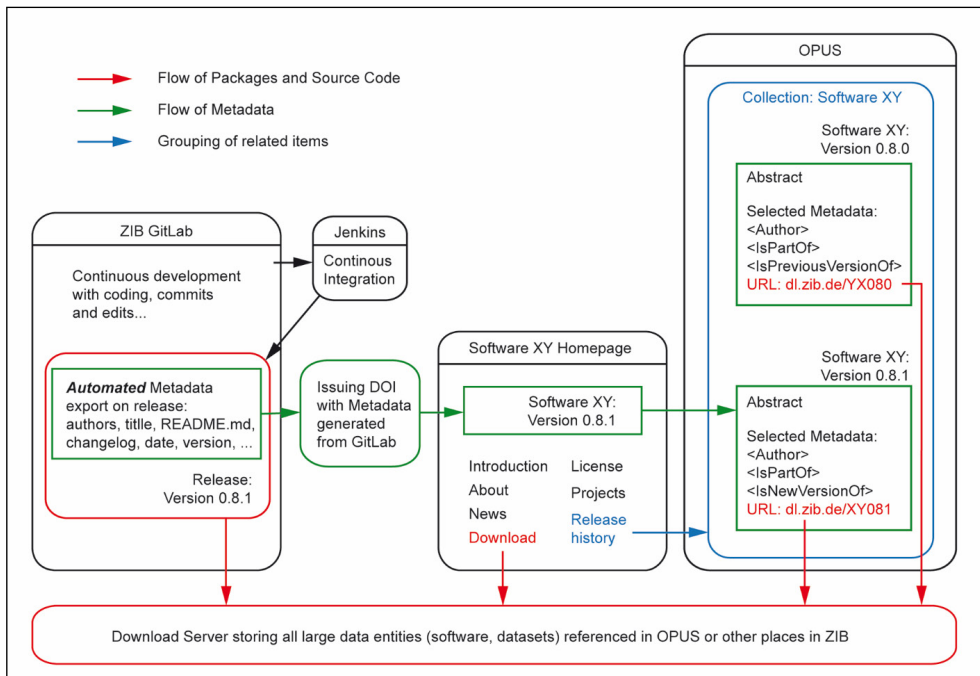


Abb. 4: Architekturkonzept bei Veröffentlichung von Software in OPUS.

## 5. Ausblick: Softwarepublikation und Langzeitarchivierung

Weil im ZIB mathematische Software eine wichtige Ressource und ein Forschungsprodukt darstellt und zukünftig als Forschungsdatum auch in OPUS publizierbar sein soll, gibt es erste Ideen für die Verknüpfung der Software-Entwicklung mit dem Veröffentlichungsworkflow (Abb. 4). Dabei soll ebenfalls der oben skizzierte Ansatz verfolgt werden, große Datenpakete nicht innerhalb von OPUS selbst zu verwalten, sondern diese auf eine getrennte Infrastruktur (innerhalb des ZIB) auszulagern. Der adressierte Server muss aber sicherstellen, dass keine nachträglichen Änderungen einmal publizierter Daten möglich sind, sondern Versionen separat publiziert werden. Kernaufgaben werden nach diesem Konzept die automatische Metadatenübernahme aus der Entwicklungsumgebung (GitLab) nach OPUS und die Abbildung der verschiedenen Versionen einer Software in OPUS sein. Dabei soll auch das Problem der mangelnden Zuschreibung einer Werkautorschaft für Software bearbeitet werden. In den ZIB-AGs, die Software entwickeln, herrscht große Personalfuktuation und die Projekte dauern häufig sehr lange. Dies gestaltet die angemessene Attribuierung komplizierter als bei herkömmlichen Text- und Forschungsdatenpublikationen.<sup>21</sup> Zudem wird Software in der Mathematik und den Ingenieur- und Naturwissenschaften bislang häufig gar nicht oder nicht

21 Laura Soito und Lorraine J. Hwang, „Citations for Software: Providing Identification, Access and Recognition for Research Software.“ *International Journal of Digital Curation* 11, Nr. 2 (2016): 53, <http://doi.org/10.2218/ijdc.v11i2.390>.

separat zitiert und die Zitationen sind diffus, d.h. Software wird auf unterschiedliche Weise genannt (Publikationen, Manuals, Webseiten, etc.). Für dieses disziplinübergreifende Problem der heterogenen Zitationspraxis von Software sollen in Kooperation mit swMATH<sup>22</sup> Lösungsvorschläge für die Mathematik erarbeitet werden.

Die im Datenlebenszyklus definierte Aufgabe der Archivierung hat im Sinn einer dauerhaften Nachnutzbarkeit eine der Publikation vergleichbare Bedeutung. In der ZIB-Instanz von OPUS soll die Archivierung als explizite, OAIS-konforme digitale Langzeitarchivierung an das Langzeitarchiv EWIG im ZIB ausgelagert werden.<sup>23</sup> Ein automatisierter Export aus OPUS nach EWIG erfolgt idealerweise in einem METS-Container; dabei kann der bereits bestehende Exportpfad mittels XMetaDisPlus als Basis für den Metadaten austausch dienen. In Abhängigkeit von der Entwicklungsroadmap des OPUS-Teams soll diese Funktionalität zukünftig in Kooperation mit der ZIB-AG „Digital Preservation“ entwickelt werden.

Für andere Repositorien kann es ebenfalls empfehlenswert sein, die eigentliche Forschungsdatenarchivierung unabhängig von der Online-Präsentation zu betreiben, weil dann kein komplexes „Alleskönner-System“ mit nur einer Software implementiert und gewartet werden muss. In OPUS selbst müssen durch die klare Trennung von Publikation und Langzeitarchivierung keine komplexen Qualitätssicherungswerkzeuge für die Datenprüfung<sup>24</sup> in den Veröffentlichungsworkflow implementiert werden und die Objekt- bzw. Metadatenverwaltung von OPUS lässt sich vollständig auf das eigentliche Veröffentlichen fokussieren.

## Literaturverzeichnis

- Bertelmann, Roland, Petra Gebauer, Tim Hasler, Ingo Kirchner, Wolfgang Peters-Kotting, Matthias Razum, Astrid Recker, Damian Ulbricht und Stephan van Gassel. *Einstieg ins Forschungsdatenmanagement in den Geowissenschaften*. Potsdam, 2014. <http://doi.org/10.2312/lis.14.01>.
- DataCite Metadata Working Group. *DataCite Metadata Schema Documentation for the Publication and Citation of Research Data*. Version 4.1, 2017. <http://doi.org/10.5438/0014>.

22 swMATH ist ein in Zusammenarbeit von ZIB, Forschungscampus MODAL und FIZ Karlsruhe betriebenes Informationssystem für mathematische Software, zuletzt geprüft am 19.06.2018, <https://www.swmath.org/>.

23 Der KOBV und das ebenfalls am ZIB angesiedelte Forschungs- und Kompetenzzentrum Digitalisierung Berlin (digiS) entwickeln und betreiben gemeinsam das digitale Langzeitarchiv EWIG; zuletzt geprüft am 19.06.2018, <http://www.kobv.de/services/archivierung/lza/>. Für technische Details zur Architektur von EWIG siehe: Marco Klindt und Kilian Amrhein, „One Core Preservation System for All your Data. No Exceptions!“, iPRES 2015 - Proceedings of the 12th International Conference on Preservation of Digital Objects, 101–108 (School of Information and Library Science, University of North Carolina at Chapel Hill, 2015), <http://hdl.handle.net/11353/10.429551>.

24 Solche Werkzeuge zur Dateiformatidentifizierung und -validierung sind Teil des Ingest-Prozesses des Langzeitarchivs.

- DINI, Arbeitsgruppe Elektronisches Publizieren, Deutsche Nationalbibliothek und Bibliothekservice-Zentrum Baden-Württemberg. *Gemeinsames Vokabular für Publikations- und Dokumenttypen*. 2010. <http://doi.org/10.18452/1492>.
- Gerland, Friederike. „Forschungsdatenmanagement in OPUS 4“. Vortrag auf dem 106. Deutschen Bibliothekartag in Frankfurt am Main 2017. Vortragsfolien. <http://nbn-resolving.de/urn:nbn:de:0290-opus4-30859>.
- Klindt, Marco und Kilian Amrhein. „One Core Preservation System for All your Data. No Exceptions!“, iPRES 2015 - Proceedings of the 12th International Conference on Preservation of Digital Objects, 101–108. School of Information and Library Science, University of North Carolina at Chapel Hill, 2015. <http://hdl.handle.net/11353/10.429551>.
- Kooperativer Bibliotheksverbund Berlin-Brandenburg. *Open Access als strategische Leitlinie des KOBV*. ZIB-Report 17–54, 2017. <http://nbn-resolving.de/urn:nbn:de:0297-zib-65303>.
- Kratzke, Jonas und Vincent Heuveline, Hrsg. *E-Science-Tage 2017: Forschungsdaten managen*. Heidelberg: heiBOOKS, 2017. <http://doi.org/10.11588/heibooks.285.377>.
- RADAR Projektteam. *Dokumentation des deskriptiven RADAR Metadatenschemas*. Version 0.9. FIZ Karlsruhe, 2017. Zuletzt geprüft am 14.05.2018. [https://www.radar-service.eu/sites/default/files/RADAR\\_Metadaten\\_Dokumentation\\_v09.pdf](https://www.radar-service.eu/sites/default/files/RADAR_Metadaten_Dokumentation_v09.pdf).
- Schulze, Matthias. „Virtuelle Forschungsumgebungen und Forschungsdaten für Lehre und Forschung: Informationsinfrastrukturen für die (Natur-)Wissenschaften“. In *Semantic Web & Linked Data : Elemente zukünftiger Informationsinfrastrukturen*; Proceedings der 1. DGI-Konferenz, 62. Jahrestagung der DGI, Frankfurt am Main, 7.–9. Oktober 2010. Tagungen der DGI 14, herausgegeben von Marlies Ockenfeld, 165–176. Frankfurt am Main, 2010. <http://dx.doi.org/10.18419/opus-6344>.
- Soito, Laura und Lorraine J. Hwang. „Citations for Software: Providing Identification, Access and Recognition for Research Software.“ *International Journal of Digital Curation* 11, Nr. 2 (2016): 48–63. <http://doi.org/10.2218/ijdc.v11i2.390>.
- Van Tuyl, Steven und Amanda L. Whitmire. „Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia.“ *PLoS ONE* 11, Nr. 2 (2016), e0147942. <https://doi.org/10.1371/journal.pone.0147942>.