

Aufbau und Betrieb eines institutionellen Repositoriums für Forschungsdaten anhand des Datenrepositories OpARA

DINI/nestor-Workshop:
Bedarfserhebungen – Grundlage für passgenaue Infrastrukturen?

30.01.2019

Falkenbrunnen 46b

Raum 242

Tel. +49 351 - 463 – 42330

Christian Löschen (christian.loeschen@tu-dresden.de)



Was ist OpARA?

- OpARA: **Open Access Repository and Archive**
- **Gemeinsames institutionelles Repotorium** der TU Dresden sowie TU Bergakademie Freiberg für digitale Forschungsdaten:
 - **Archivierung** und
 - **Publikation** digitaler Forschungsdaten
- Entwicklung und Betrieb:
 - ZIH (TU Dresden)
 - URZ (TU Bergakademie Freiberg)
- Start des Dienstes: Februar 2018
- Webseite: <https://opara.zih.tu-dresden.de>

- DFG: Empfehlungen zur Sicherung guter wissenschaftlicher Praxis
 - **Empfehlung 7: Sicherung und Aufbewahrung von Primärdaten**

Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, zehn Jahre lang aufbewahrt werden.

- TU Dresden: Richtlinien zur Sicherung guter Wissenschaftlicher Praxis
 - **§ 5: Sicherung und Aufbewahrung von Primärdaten**
 - (1) Primärdaten als Grundlagen für Veröffentlichungen sind auf haltbaren und gesicherten Trägern in der Lehr- und Forschungseinheit, in der sie entstanden sind, aufzubewahren. [...]
 - (2) In der Regel müssen die Primärdaten für zehn Jahre zugänglich bleiben.
[...]

OpARA – Der Anlass

- Bereitstellung eines Initiativbudgets für die **Entwicklung hochschulübergreifender Informationsinfrastrukturen** des Sächs. Ministerium für Wissenschaft und Kunst (SMWK)
- Projektlaufzeit 07/2014 - 12/2016
- Partner:
 - TU Dresden
 - TU Bergakademie Freiberg
 - Sächs. Landes- und Universitätsbibliothek SLUB (assoz.)
- Vornehmliches Ziel:
 - Einfache Lösung zum Archivieren digitaler Forschungsdaten (gemäß guter wiss. Praxis)
- Weiterhin:
 - Publikation von FD
 - Beratungen/Schulungen zum Datenmgmt.

Umfrage und Interviews

- Zum Umgang mit digitalen Forschungsdaten, der Archivierung von dFD, und zur Akzeptanz von Open Access
- Ziel auch: Schaffung von Awareness, Publicity für unsere Aktivitäten
- Umfrage an ~ 450 Professuren geschickt, zusätzlich einige Interviews
 - Return > 10 % (vollständig ausgefüllte)
- Ergebnisse in etwa wie erwartet:
 - FDM in allen erdenklichen Umsetzungsstadien
 - Alle Variationen von Daten (-typen)
 - Ablehnung oder Unkenntnis von FD-Publikation (OA) bis zu voller Zustimmung
 - Anerkennung der Notwendigkeit der Archivierung von FD
 - Frage nach dem Kosten (Aufwand)/Nutzen
 - Tlw. sehr spezielle Anforderungen
 - Lokaler Dienst eher vertrauenswürdig

Evaluation von Repository-Softwares

● Unsere Quintessenz:

- Möglichst niederschwelliges Angebot
- Möglichst keine Kosten → Geschäftsmodell
- Publikation nicht immer gewünscht oder möglich
- Vorerst Ausklammern von speziellen Anwendungsfällen

● Anforderungskatalog

- Authn, Authz, Schnittstellen, Ingest (-Workflow), Metadatenhaltung, Metadatenstandards, Datenformate, Datenhaltung, Discovery, Dissemination, Persistente IDs, Support, Mandantenfähigkeit
- DSpace, Fedora, Islandora, DSpace-CRIS, Lockss, Eprints, ...

● Entscheidung für „fertiges“ System mit großer Community → DSpace

● Mit Anpassungen an eigene Infrastruktur und Anforderungen

Geschäftsmodell

- Variante 1

- Wissenschaftler bezahlen pro Einreichung/Datenvolumen/Zeitraum

- Variante 2

- Einrichtungen bezahlen „Mitgliedsbeitrag“

- Variante 3

- Trägerschaft des Angebots durch das Land Sachsen (SMWK), keine weiteren Kosten für Einreichende oder Universitäten

- Entscheidung für V3:

- Wissenschaftler/Projekte haben keine zusätzlichen Kosten (Niederschwelligkeit!)
 - Wir haben keinen Abrechnungsoverhead:
 - Was als Berechnungsgrundlage nehmen?
 - Wie die tatsächlichen Kosten ermitteln?
 - Attraktiv für andere Einrichtungen
 - Trägerschaft durch das Land allerdings noch im Fluss

Was bietet OpARA?

- **Kostenfreie Archivierung** für zunächst **zehn Jahre** für Wissenschaftler der TUD und TUBAF
 - Länger für Open-Access
- **Optionale Publikation**
- Zweifache **Begutachtung** (fachlich + technisch)
- Vergabe von **DOI** zur persistenten Referenzierung der Datensätze
- Redundante, räumlich getrennte Archivierung der Daten
(2 x Dresden + 1 x Freiberg)
- **Authentifizierung** über die zentralen IDM der Universitäten
- Automatische **Autorisierung** des Nutzers an seinem Bereich der Universitätssstruktur
- **Rechtesystem** zur Vergabe von Administrator-, Gutachter-, Ersteller-, Einreicher- oder Leserechten
 - Vergabe von Zugriffsrechten für die eigenen (nicht veröffentlichten) Daten

Die Strukturen in OpARA

Universitätsstruktur:

- Bereiche
- Fakultäten
- Fachrichtungen
- Institute
- Professuren

Bereiche in OpARA

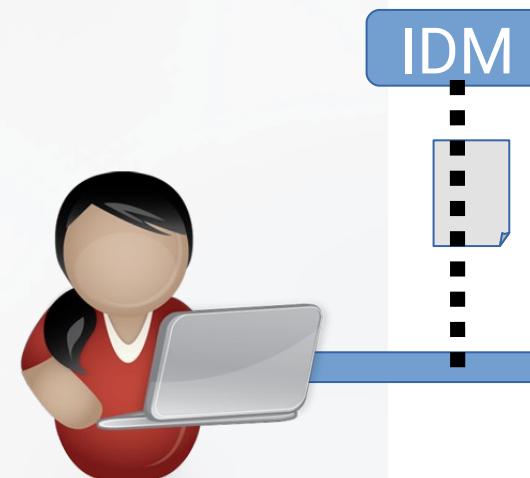
Wählen Sie einen Bereich, um die enthaltenen Sammlungen zu durchstöbern.

- Technische Universität Bergakademie Freiberg
- Technische Universität Dresden [2]
 - Bereiche [2]
 - Bereich Bau und Umwelt [1]
 - Bereich Ingenieurwissenschaften
 - Bereich Mathematik u. Naturwissenschaft. [1]
 - Fak. Mathematik und Naturwissenschaften [1]
 - L/V Mathematik und Naturwissenschaften
 - FR Mathematik
 - FR Physik
 - Inst. Angewandte Physik (IAP)
 - P. Didaktik der Physik
 - Inst. Theoretische Physik
 - P. Computational Physics
 - P. Theorie der kondensierten Materie

Die Strukturen in OpARA – Autorisierung beim Login

Universitätsstruktur:

- Bereiche
- Fakultäten
- Fachrichtungen
- Institute
- Professuren



Bereiche in OpARA

Wählen Sie einen Bereich, um die enthaltenen Sammlungen zu durchstöbern.

- + Technische Universität Bergakademie Freiberg
- Technische Universität Dresden [2]
 - Bereiche [2]
 - + Bereich Bau und Umwelt [1]
 - + Bereich Ingenieurwissenschaften
 - Bereich Mathematik u. Naturwissenschaft. [1]
 - Fak. Mathematik und Naturwissenschaften [1]
 - + L/V Mathematik und Naturwissenschaften
 - + FR Mathematik
 - FR Physik
 - + Inst. Angewandte Physik (IAP)
 - P. Didaktik der Physik
 - Inst. Theoretische Physik
 - P. Computational Physics
 - P. Theorie der kondensierten Materie

Die Strukturen in OpARA

An seinem Institutsbereich **durch den Nutzer zu erstellen**:

● Projekte

- stellen idR Forschungsprojekte, Dissertationen, Diplom-, Master-, Bachelorarbeiten dar
- beinhalten **Datensammlungen**, evtl. weitere **Unterprojekte**

● Datensammlungen

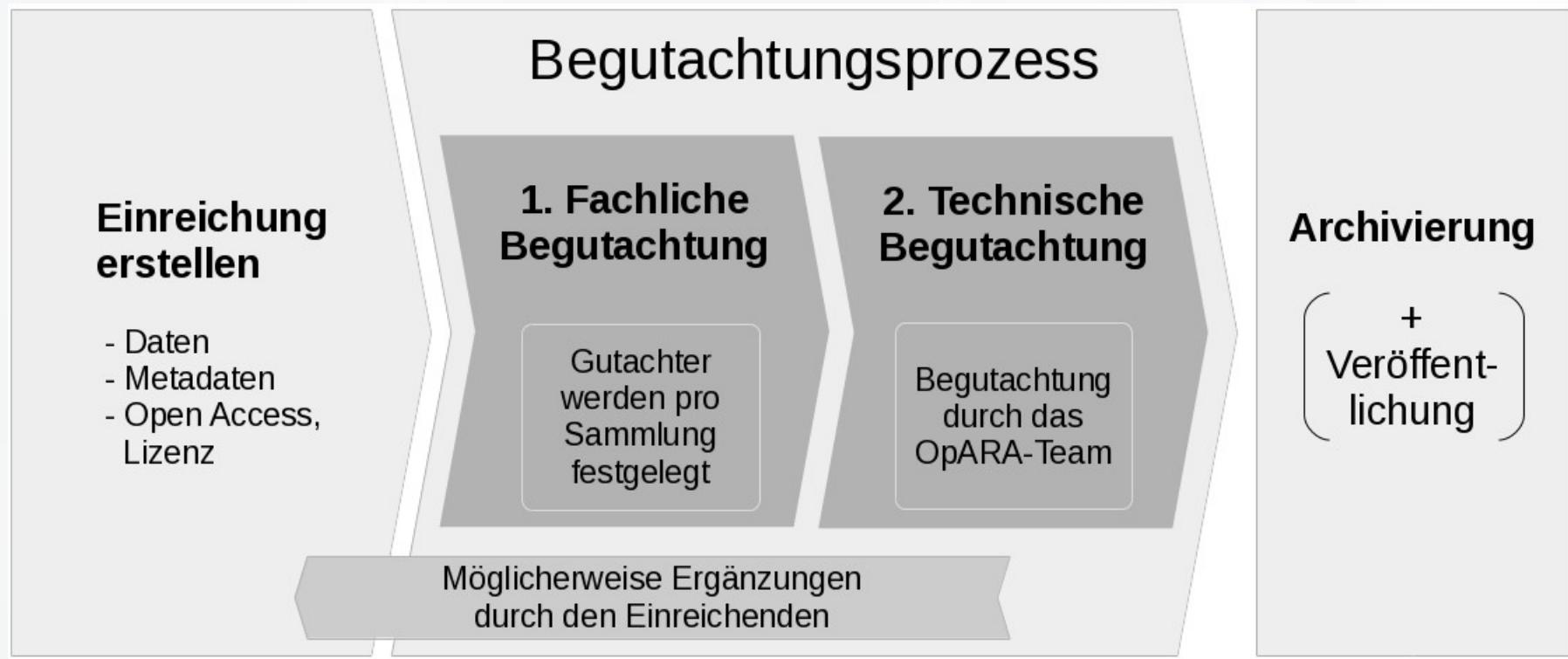
- Container für zusammengehörige **Datenpakete** innerhalb eines Projekts

● Datenpakete

- beinhalten die **archivierten Daten und Metadaten** der Einreichungen

Der Einreichungsprozess

- I. Einreichung erstellen
- II. Begutachtungen
- III. Archivierung



● Quelle: opara.zih.tu-dresden.de/xmlui/manual

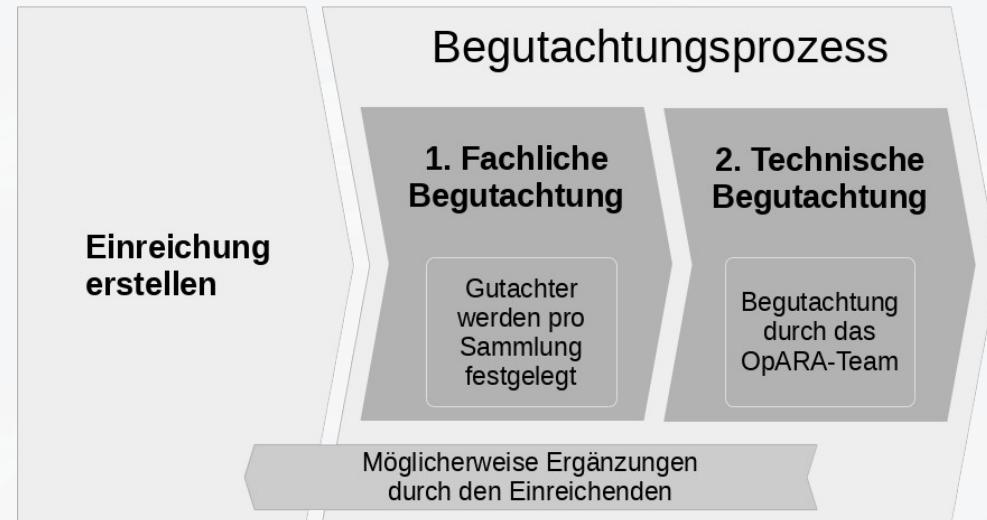
Der Einreichungsprozess I - Erstellen einer Einreichung

- Schritt 1: Auswahl von Open Access
- Schritt 2: Beschreibung der Einreichung
- Schritt 3: Angabe der beteiligten Personen
- Schritt 4: Beschreibung der Daten
- Schritt 5: Angabe der geografischen Bezüge
- Schritt 6: Auswahl bzw. Angabe der Lizenz
- Schritt 7: Daten-Upload
- Schritt 8: Kontrolle der Angaben

Der Einreichungsprozess II - Der Begutachtungsprozess

1. Fachliche Begutachtung

- durch einen Gutachter, ausgewählt für jede Datensammlung
- Kontrolle der inhaltlichen Richtigkeit und Vollständigkeit



2. Technische Begutachtung

- durch das OpARA-Team
- Datenformate, Dokumentation, Metadaten

- Bei notwendigen Ergänzungen jeweils Zurückverweisung an den Einreichenden

Der Einreichungsprozess III – Archivierung (& Publikation)

- Zusammenstellen des Archivpakets aus den Daten und Metadaten der Einreichung
- Vergabe des DOI
- Redundantes Sichern des Archivpakets in den Bandarchiven (2 x Dresden, 1 x Freiberg)
- Optionale Veröffentlichung (ggfs. nach Ablauf der Embargofrist)

Was läuft gut, ...

- Lokaler Betreiber, kurze Wege
- Nachfrage von Weiterbildungen und Beratungen
- Verbesserungsvorschläge während Begutachtungsprozess werden gut angenommen
- Weniger kritische Begutachtung für nichtpublizierte Daten möglich
- Sichtbarkeit der Datenpublikationen an der eigenen Professur

..., was nicht so sehr

- Zu wenig Doku vs. Zu viel Doku
- Benennung des fachlichen Gutachters
- Ausreichende Beschreibung des Datenpakets
(Datenpaket soll am Ende selbsterklärend sein vs. Bereits gemachte Angaben in der übergeordneten Struktur: Projekt, Datensammlung)
- Upload via Webformular nicht immer sinnvoll/machbar
(zusätzliches Tool auf der Todo-Liste)
- Zu wenig Personal atm (Features, Publicity, ...)
- Wartbarkeit unserer Modifikationen
(insb. wegen wesentlicher Umbauten während letztem (API) und nächstem (WebUI) DSpace Major Release)

Einbettung in FD-Infrastruktur

- Gemeinsame Kontaktstelle Forschungsdaten der SLUB und TU Dresden
 - Beratungen, Öffentlichkeitsarbeit
 - SLUB, ZIH (Dienste), SG 3.5 (Informationssicherheit, Datenschutz), Jur. Fak., Institut für Geistiges Eigentum, Technikrecht und Medienrecht IGETeM (Erläuterung rechtlicher Rahmenbedingungen)
 - Ab 01.01.2019: Zwei Teams zur individuellen techn. Umsetzung von FDM auf Anfrage von Instituten (jeweils 1 Wissenschaftler + 1 Fachinformatiker)
- Leitlinien für den Umgang mit Forschungsdaten an der TUD (13.08.2018)
- RDMO (ab 2019)
- MASI (ab 2019), metadatengetriebenes FDM
- Regelmäßige offene Weiterbildungsangebote, individuelle Weiterbildungstermine