

WO PUBLIZIERE ICH MEINE DATEN?

27 OCTOBER 2021

TORSTEN BRONGER

ZENTRALBIBLIOTHEK, FORSCHUNGSZENTRUM JÜLICH

ZIEL DER DATENVERÖFFENTLICHUNG

- Pflichterfüllung gegenüber Förderer und/oder Arbeitgeber.
- Zukünftig besteht kein Aufwand mehr, wenn die Forschungsdaten nachgefragt werden.

Für beides ist es nahezu egal, wo man veröffentlicht. Nicht jedoch für die folgenden Ziele:

- Erhöhte Transparenz und Sichtbarkeit der eigenen Forschung.
- Daten können von potentiellen Kooperationspartnern gefunden werden.
- Daten können wie Textpublikationen eindeutig zitiert werden \leadsto Reputation.
- Vermeidung doppelter Erhebungen und damit unnötiger Kosten (Zeit und Geld).
- Vorbereitet sein auf steigende Bedeutung von Datenpublikationen bei Evaluierungen.

⇒ Auffindbarkeit und Nachnutzbarkeit maximieren!

AUS UNSEREN LEITLINIEN

Aus den [Leitlinien für Forschungsdaten des FZJ](https://go.fzj.de/datenpolicy) (<https://go.fzj.de/datenpolicy>) (<https://go.fzj.de/datenpolicy>):

„Forschungsdaten sollten unmittelbar nach ihrer Erstverwertung [...] durch die für sie verantwortlichen Wissenschaftler öffentlich zugänglich gemacht werden, sofern dem keine erheblichen Gründe entgegenstehen.“

Was ist die Erstverwertung?

→ Typischerweise eine Textpublikation (oder eine Serie von Textpublikationen) in einer Fachzeitschrift, die auf diesen Daten basiert. Umfang und Dauer der Erstverwertung kann in einem DMP festgelegt werden.

BEGRIFFE

Rohdaten

- die eigentlichen Rohdaten – gerne auch aufbereitet, d.h. Baseline abgezogen etc.
- wissenschaftliche Metadaten wie Probenname, Experiment-Temperatur, Name des Experimentators, ...

Metadaten

- Zeitstempel
- Kontaktdaten des Ansprechpartners
- Schlagworte
- Lizenz
- Institut, POF-Topic
- evtl. Link auf die Rohdaten

WAS HEISST „VERÖFFENTLICHEN“?

Ein Datensatz ist veröffentlicht, wenn die *Metadaten* in einem Repository, das Zugriff für mindestens 10 Jahre garantiert, per PID (typischerweise DOI) weltweit abrufbar sind.

Insbesondere heißt „veröffentlichen“ nicht zwangsläufig, daß die *Rohdaten* weltweit abrufbar sind. Ein möglicher Weg für andere Forscher ist:

1. Forscher findet interessante Forschungsdaten über Suchmaschine.
2. Forscher klickt auf DOI.
3. Forscher landet auf Webseite mit den Metadaten, u.a. einer Kontakt-Email-Adresse.
4. Dahin schickt Forscher eine Anfrage.
5. Das Institut, daß die Daten erzeugt hat, prüft die Anfrage.
6. Das Institut gibt die Rohdaten für diesen Forscher frei, u.U. mit Hinweisen, wie sie verwendet werden können (und wie nicht).

UNTERSCHIEDE ROHDATEN ↔ METADATEN

Normalerweise werden Roh- und Metadaten an derselben Stelle veröffentlicht. Es kann aber gute Gründe geben, das nicht zu machen:

- Rohdaten sind zu groß, um irgendwo hochzuladen.
- Rohdaten sollen im Institut (oder zumindest in dessen Einzugsbereich) bleiben.
- Rohdaten sind bereits irgendwo anders hochgeladen.

Ganz generell sollte man hinterfragen, warum Rohdaten überhaupt durch kopieren auf ein Repository verdoppelt werden sollen.

FAIR-PRINZIPIEN

Findable

- Daten sind mit Metadaten beschrieben.
- Die Metadaten enthalten Link zu Daten.

Accessible

- Daten sind über standardisiertes Protokoll erreichbar (mit Authentifizierung, wenn nötig).

Interoperable

- Metadaten nutzen Standards/standardisierte Vokabularien.

Reusable

- Metadaten beschreiben die Daten ausreichend und nach fachlichen Standards.
- Metadaten legen über eine Lizenz (vorzugsweise CC-0 oder CC-BY) fest, welche Nutzung erlaubt ist.

ORTE DER DATENVERÖFFENTLICHUNG

Kategorien der Wege einer Datenveröffentlichung:

1. Datenjournal
2. Repository
3. das Journal, bei dem auch die Textpublikation veröffentlicht wird
4. Jülich DATA

DATENJOURNAL

- Artikel-basierte Beschreibungen von Daten und ihrem Kontext.
- Fokus liegt auf ausführlicher Beschreibung der Daten, nicht auf deren Interpretation.
- Manche Datenjournale führen zur Qualitätssicherung Review-Verfahren durch.
- Überblick zu existierenden Datenjournalen findet sich bei forschungsdaten.org
(www.forschungsdaten.org/index.php/Data_Journals).
- Die Daten werden in Journal-eigenen oder externen Repositorien archiviert.

Meine Einschätzung: Das wird meist nicht der Plan A sein.

REPOSITORY

Datenbanken, in denen Objekte archiviert, dokumentiert und publiziert werden können.

Ein Repository ermöglicht es,

- Daten (Metadaten und Rohdaten) langfristig sicher zu archivieren.
- Daten über persistente Identifikatoren (PIDs, z.B. eine DOI) dauerhaft zu referenzieren.
- Daten zu teilen.
- selbst nach Daten zu suchen.

REPOSITORY

Repositorien lassen sich gliedern in

- medienspezifische Repositorien
- disziplinspezifische Repositorien
- Fachübergreifende (generische) Repositorien
- institutionelle Repositorien

DISZIPLINSPEZIFISCHE REPOSITORIEN

- speziell für eine Fachrichtung ausgelegt
- zentraler Ort, um nach Daten einer speziellen Fachdisziplin zu suchen
- gute Sichtbarkeit in der Fach-Community
- enthalten fachspezifische Metadatenstandards zur Datenbeschreibung
- oft mit besonderen Services ausgestattet, z.B. spezifische Werkzeuge zur Recherche und Visualisierung

⇒ Verwenden Sie es, wenn es eines für Sie gibt!

Beispiel: PANGAEA

- Qualitätskontrolle durch fachliche Datenkuratoren
- Fachspezifische Metadaten (Parameter, räumliche Abdeckung, usw.)

GENERISCHE REPOSITORIEN

Beispiel für ein öffentliches generisches Repository: [Zenodo](https://zenodo.org) (<https://zenodo.org>).

- offen für alle Forschungsgebiete und alle Arten von Forschungsdaten (z.B. Journalartikel, Poster, Präsentationen, Graphiken, Video-/Audiofiles)
- kostenfreier Dienst, Dateigröße bis zu 50 Gigabyte!
- DOI-Vergabe
- Integration mit GitHub
- jeder legt selbst fest, wie frei zugänglich seine Daten sein sollen
- alle Metadaten stehen unter einer CC-Lizenz

EIN REPOSITORY FINDEN

Empfehlung: Suchen auf dem Repositorien-Verzeichnis [re3data](https://re3data.org) (<https://re3data.org>). Das enthält mehrere Tausend Repositorien.

Gute Filtermöglichkeiten, zum Beispiel nach

- Fachdisziplin
- Datentypen
- Lizenzen
- Persistenten Identifikatoren
- Nutzungsbedingungen
- Datenzugriff

WIE KOMMEN DIE DATEN INS REPOSITORY?

Per Web-Formular: Hochladen der Daten über einen Internetbrowser, Beschreibung per Metadatenformular

→ Geeignet für das gelegentliche Hochladen von kleinen und mittleren Datenmengen

Per REST-API: Automatisiertes Einspielen von Daten und Metadaten

→ Geeignet für häufiges Hochladen oder größere Datenmengen

Per Client-Software: Eigene Software zum Hochladen und Beschreiben der Daten

→ Sicherstellen des Uploads, Wiederaufnahme bei Unterbrechung, Überprüfung der Metadaten

REPOSITORY DER EINRICHTUNG

In den letzten Jahren haben sich die meisten Forschungseinrichtungen ein eigenes Repository für Forschungsdaten zugelegt.

Diese machen fast die Hälfte aller DOI-vergebenden Repos aus.

Ob diese Zersiedlung und Verdoppelung von Diensten eine gute Idee war, mag bezweifelt werden ...

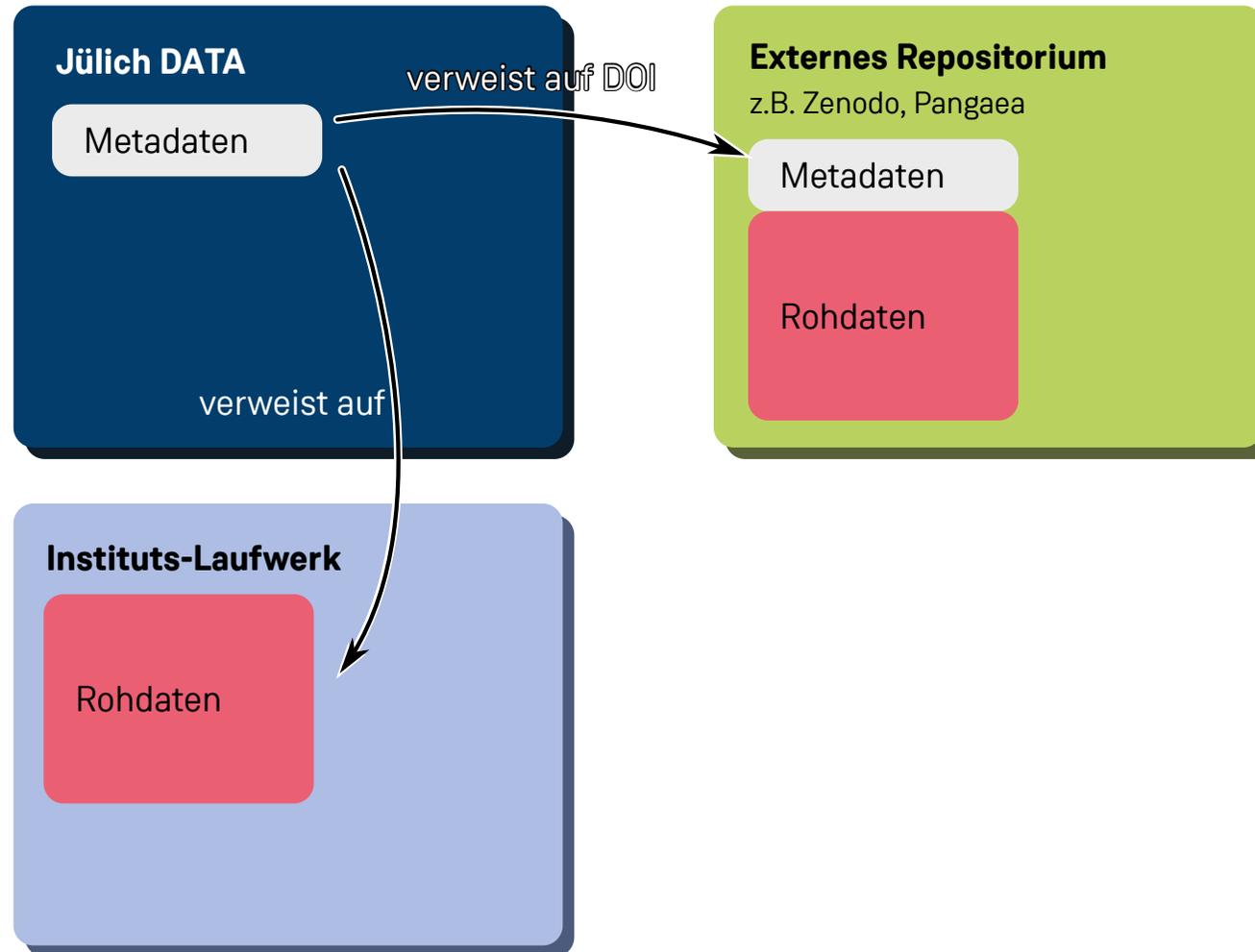
JÜLICH DATA

In Jülich haben wir das nur zur Hälfte nachgemacht: Es gibt Jülich DATA, aber dient dient in erster Linie dem *Nachweis* des Jülicher Daten-Outputs (z.B. um darüber Statistiken oder Evaluationen zu machen) und ist in dieser Hinsicht die Schwester von JuSER.

Wir empfehlen daher dringend, da nur Metadaten hochzuladen, die dann auf die Rohdaten zeigen. Das Zeigen kann über eine URL erfolgen, einen Verzeichnispfad auf dem Institutslaufwerk, oder eine textuelle Beschreibung. Es muß also nicht automatisch auflösen!

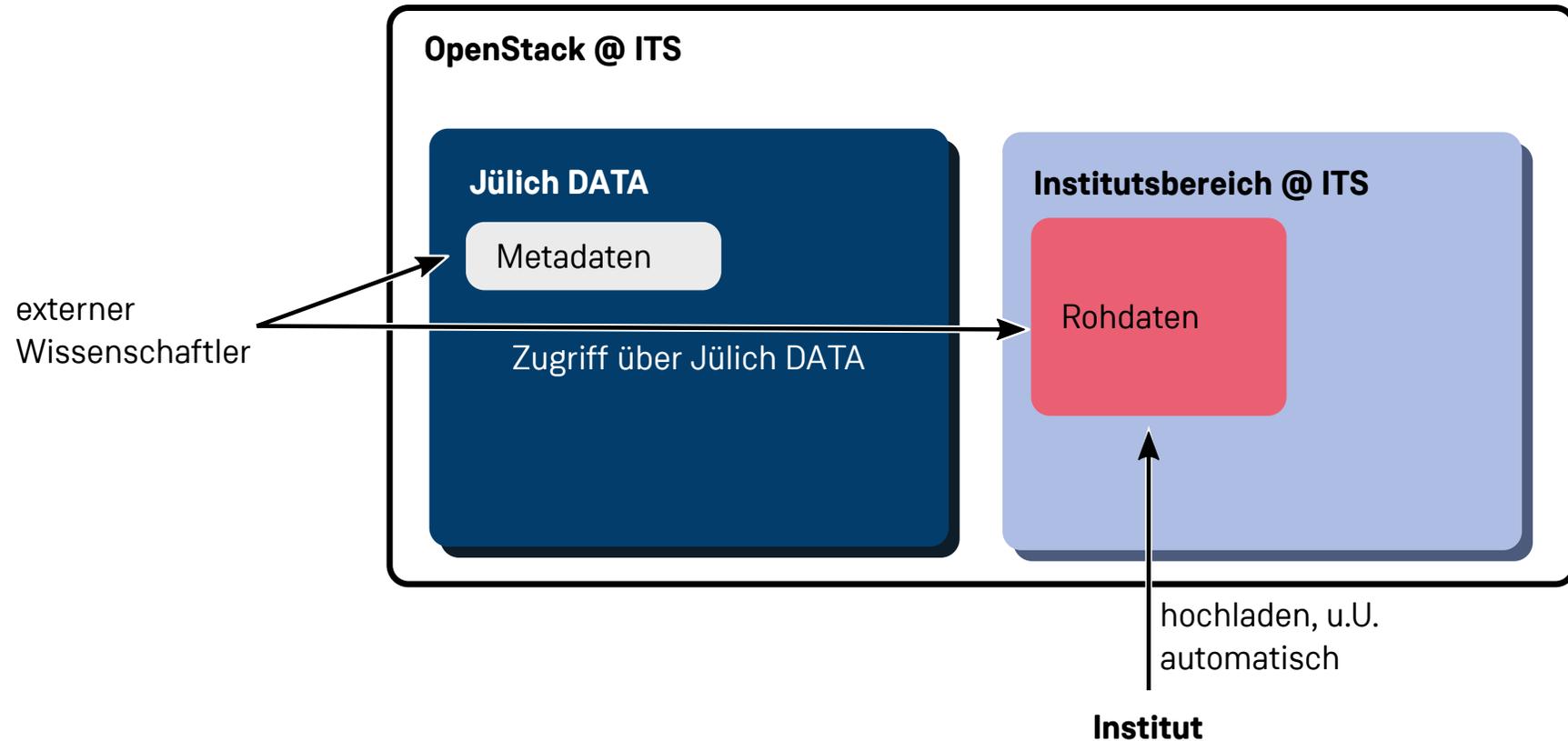
Technisch kann man allerdings auch Rohdaten hochladen. Die Gesamtkapazität ist bei Jülich DATA mit 10 TB regelrecht winzig.

JÜLICH DATA



VISION: INSTITUTSDATEN NAH AM REPOSITORY

Vorschlag für das Jülicher Repo-Modell:



THANK YOU

Torsten Bronger
Zentralbibliothek, Forschungszentrum Jülich
t.bronger@fz-juelich.de (mailto:t.bronger@fz-juelich.de)