

**Leitlinien zur Publikation von
Forschungsdaten im PUBLISSO –
Fachrepositorium
Lebenswissenschaften („Data
Policy“) – Stand: Oktober 2019**

Hintergrund

Forschungsdaten werden mit großem Aufwand produziert, liegen in vielfältiger Form vor, sind disziplinspezifisch und gehören daher zu den wertvollen Ressourcen von wissenschaftlichen Einrichtungen. Forschungsdaten untermauern wissenschaftliche Aussagen. Das Veröffentlichen von Forschungsdaten ermöglicht deren Zitierfähigkeit und Nachnutzbarkeit, steigert ihre Zitierhäufigkeit und leistet der Überprüfung wissenschaftlicher Ergebnisse Vorschub. Zudem fördert es die Entwicklung neuer wissenschaftlicher Erkenntnisse.

PUBLISSO unterstützt im Sinne der „Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen“ sowie der „Regeln zur Sicherung guter wissenschaftlichen Praxis“ alle Initiativen zur Weiterentwicklung und Umsetzung eines qualitätsbewussten Forschungsdatenmanagements mit dem Ziel der verbesserten Wissenschaftskommunikation und Sichtbarkeit der lebenswissenschaftlichen Forschung.

Definition von digitalen Forschungsdaten

Unter Forschungsdaten sind „[...] digitale und elektronisch speicherbare Daten zu verstehen, die im Zuge eines wissenschaftlichen Vorhabens z.B. durch Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen“ (Deutsche Forschungsgemeinschaft). Es handelt sich also um sämtliche Produkte und Medien, die im Rahmen von wissenschaftlicher Forschung entstehen und/oder die in ihrem Kontext verwendet werden. Sie können im Zusammenhang mit wissenschaftlichen Textpublikationen, aber auch eigenständig veröffentlicht werden.

Ein integraler Bestandteil von Forschungsdaten sind deren Metadaten sowie ggf. begleitende Materialien, wie Datensatz-Dokumentation, Methodenbeschreibungen, Projektberichte, Berechnungscodes, Fragebögen etc. Das Gesamtdatenpaket (Forschungsdaten, Metadaten, Begleitmaterial) wird als „digitales Objekt“ bezeichnet.

Lebenszyklus von Forschungsdaten

Forschungsdaten durchlaufen einen Lebenszyklus, der von der Projektplanung, Primärdatenerhebung, über die Anreicherung der Forschungsdaten mit Metadaten, ihre Qualitätssicherung, Veröffentlichung, Speicherung, Nachnutzung durch andere Forschende, kuratierte Archivierung über viele Jahre bis hin zu deren eventueller Löschung reicht. Bei diesen Phasen werden jeweils verschiedene Akteure an unterschiedlichen Orten (z.B. Forschungsinstitute, Rechenzentren, Bibliotheken, Archive) einbezogen. Um die Qualität der Forschungsdaten im Verlauf ihres Lebenszyklus zu sichern, ist es wichtig, bereits vor der Datenerhebung einen Datenmanagementplan (DMP) zu erstellen, der die verschiedenen Lebensphasen und involvierten Akteure reflektiert und dokumentiert. ZB MED unterstützt Forschende im Umgang mit DMP und bietet einen Beratungsservice zur projektbezogenen Begleitung an.

Gegenstand

ZB MED bietet mit dem PUBLISSO – Fachrepositorium Lebenswissenschaften (FRL) einen Service zur Publikation und zum dauerhaften Speichern von Texten und Forschungsdaten an. ZB MED verpflichtet sich, die technischen Voraussetzungen dafür zu schaffen und diese aufrecht zu erhalten. Zu diesem Zweck kooperiert ZB MED aktuell mit dem Hochschulbibliothekszentrum NRW (hbz).

Das Fachrepositorium zielt insbesondere darauf ab, das dauerhafte Sichern wertvoller Forschungsdaten zu unterstützen sowie – im umfassenderen Sinne – deren Auffindbarkeit, Zugänglichkeit, Interoperabilität und Nachnutzbarkeit zu fördern (vgl. [FAIR-Data-Prinzipien](#)). Dafür werden die Forschungsdaten mit strukturierten Metadaten beschrieben und mit einem persistenten Identifikator (DOI) versehen, wodurch die Datenpublikationen eindeutig referenzierbar und dauerhaft zitierbar werden. Dieses wird dadurch gewährleistet, dass ZB MED DataCite-Mitglied und Vergabestelle für DOI in den Lebenswissenschaften ist.

Der Publikationsservice wird gegenwärtig kostenlos angeboten. ZB MED behält sich das Recht vor, zukünftig zur Deckung der Betriebskosten bestimmte Dienstleistungen mit Gebühren zu versehen.

Datenformate und Standards

Wie bereits unter „Definition von Forschungsdaten“ angesprochen, sind die Forschungs- und Erhebungsmethoden, einschließlich verwendeter (Mess-)Geräte und Standards, durch deren Anwendung Forschungsdaten entstehen, vielfältig. Entsprechend vielfältig sind auch die entstehenden Datentypen (z.B. Bilder, Tabellen, Video- und Audioaufnahmen etc.). ZB MED erkennt deswegen an, dass auch die Datenformate in denen Forschungsdaten vorliegen, vielfältig sein können. Aus diesem Grund akzeptiert ZB MED zurzeit alle Datenformate, solange diese von anderen Forschenden nachgenutzt werden können und maschinenlesbar sind. Eine Dateigröße von 1 GB kann derzeit nicht überschritten werden.

In der der Publikation beigefügten Dokumentation zu den Forschungsdaten sollte angegeben werden, mit welcher Software die Datenformate geöffnet werden können und – zwecks Nachvollziehbarkeit der Forschungsergebnisse – welche Standards bei der Datenerhebung zur Anwendung kamen. Generell gilt, dass die verwendeten Formate und Standards den nationalen und internationalen Vorgaben der jeweiligen Fachdisziplin entsprechen sollten.

Die Nachnutzung der Forschungsdaten, d.h. die Möglichkeit der weiteren Arbeit mit den Daten, ist ein wichtiger Punkt bei der Publikation. Formate, die nur eine Darstellung, aber keine Bearbeitung ermöglichen, schränken potenzielle Szenarien neuer wissenschaftlicher Erkenntnisse ein. Zudem sollten proprietäre Formate nach Möglichkeit vermieden werden, da ihre zukünftige Nutzbarkeit selbst in Updates derselben Software nicht gesichert ist und die digitale Langzeitarchivierung erschweren. Nach derzeitigem Erkenntnisstand sind solche Datenformate empfehlenswert, die möglichst verlustfrei sind und beschreibende Elemente beinhalten, die das jeweilige Dokument transparent und eindeutig interpretierbar machen. Im besten Fall sind diese auch für Menschen lesbar, um auch dann eine Migration zu gewährleisten,

wenn die jeweilige Software nicht (mehr) vorliegt. Dieses wird aktuell für XML, SGML und TIFF sowie darauf aufbauende Formate angenommen.

Struktur der Daten

Der Aufbau von Forschungsdaten, z.B. ob diese zusammengefasst in einer Datei oder als mehrere Dateien vorliegen, hängt stark von den Daten selbst ab sowie davon, wie diese gewonnen wurden. ZB MED akzeptiert deswegen einzelne oder zusammengefügte Dateien, bietet aber auch die Möglichkeit, mehrere Dateien unter einem Metadatensatz zusammenzufassen. Darüber hinaus können auch mehrere Dateien mit jeweils eigenem DOI publiziert und ihre Zusammengehörigkeit untereinander in den Metadaten dargestellt werden.

Datensicherheit und Aufbewahrungsdauer

Derzeit wird eine Verfügbarkeit der im Fachrepositorium vorhandenen Dokumente für jeweils mindestens 10 Jahre vom Zeitpunkt der Publikation auf dem Repositorium-System an gewährleistet.

Eine datengeberseitig veranlasste Löschung einmal eingestellter digitaler Objekte ist grundsätzlich nicht vorgesehen. In begründeten Ausnahmefällen (bspw. fehlerhafte Dateien, rechtliche Probleme) können digitale Objekte nach Rücksprache zurückgezogen werden. Die Metadaten werden weiterhin öffentlich nachgewiesen und der vergebene DOI löst weiterhin auf den entsprechenden Metadateneintrag auf. Dieser wird um einen Hinweis auf den (Teil-)Rückzug der Daten ergänzt.

Zugang zu Daten und Metadaten

Es können ausschließlich Forschungsdaten eingestellt werden, die für eine Veröffentlichung und wissenschaftliche Nachnutzung vorgesehen sind. Eine Vergabe einer Open-Data-Lizenz ist deswegen für die Publikation von Forschungsdaten im Fachrepositorium Lebenswissenschaften verpflichtend. Hierfür bietet ZB MED Lizenzen an, welche den Grundsätzen von Open Data entsprechen. Damit diese für die Allgemeinheit nachvollziehbar und die Rechte zur Nachnutzung klar definiert sind, verwendet ZB MED standardisierte Lizenzen von bekannten Organisationen (siehe Vertrag §1 Vertragsgegenstand).

Die Rechte des Urhebers werden nach den Regelungen des § 7 des mit ZB MED geschlossenen Vertrages übertragen. Zur Wahrung der Möglichkeit der Erstverwertung seitens der Datenproduzentinnen und -produzenten, insbesondere bei Forschungsdaten, die aus laufenden Projekten zur Verfügung gestellt werden, besteht die Möglichkeit eines Embargos. Dieses ist auf maximal zwei Jahre begrenzt. Nach Ablauf der festgelegten Embargofrist erfolgt eine Freischaltung der zuvor nicht öffentlich sichtbaren digitalen Objekte. Die beschreibenden Metadaten werden bereits während dieser Embargofrist öffentlich zugänglich gemacht und mit einem DOI versehen.

Datenschutz

Die einschlägigen Bestimmungen des Datenschutzes müssen vom Autor beachtet werden. Insbesondere müssen Datensätze, die besondere personenbezogene Informationen auf Individualebene enthalten (bspw. Patienteninformationen, Umfragedaten), so anonymisiert sein, dass eine Identifizierung von Individuen ausgeschlossen ist (faktische Anonymität). Personenbezogene Daten, die in den Forschungsdaten enthalten sind, können nur bei Nachweis einer Einwilligung durch die Betroffenen eingestellt werden.

ZB MED befürwortet einen möglichst uneingeschränkten und dabei datenschutzkonformen Zugang zu Forschungsdaten. Somit können nur Forschungsdaten eingestellt werden, die so beschaffen und beschrieben sind, dass eine wissenschaftliche Nachnutzung möglich ist. Das schließt auch die Vergabe einer adäquaten Nutzungslizenz ein.

Überprüfung der Inhalte

ZB MED überprüft alle eingesendeten Daten und Begleitmaterialien in Bezug auf die Dokumentation, die Aufbereitung und die technische Nachnutzbarkeit/Integrität. Dieses bedeutet im Einzelnen:

- ▶ Archivierbarkeit: Es ist technisch möglich, den Inhalt zu archivieren.
- ▶ Dokumentation: Der Inhalt ist mit Metadaten angereichert und hinreichend zur Nachnutzung dokumentiert.
- ▶ Technische Nutzbarkeit: Der Inhalt kann mit allgemein für das Fachgebiet üblicher Software geöffnet und gelesen werden; ggf. sind Codes/Software, die bei der Datenerhebung genutzt wurden, beigelegt.
- ▶ Aufbereitung (vorbehaltlich technischer Nutzbarkeit): Der Inhalt weist eine verständliche, d.h. nachvollziehbare Struktur auf.

Daten, welche bei der Überprüfung auffallen, werden zur Bearbeitung zurückgesendet. ZB MED behält sich das Recht vor, die Publikation ggf. auch abzulehnen.

Beratung

ZB MED bietet Beratung zu allen Fragestellungen rund um das Publizieren und Archivieren von Forschungsdaten an. Hiervon ausgenommen ist eine juristische Beratung im Sinne einer Rechtsdienstleistung. Sprechen Sie uns an oder nutzen Sie unsere [Beratungsseiten](#) zur Information.