

Arbeitsanleitung zu Open Research Data

1. Einleitung

In den letzten Jahren hat in der Schweiz das Bewusstsein für die Wichtigkeit der Veröffentlichung von Forschungsdaten zugenommen. In diesem Zusammenhang haben sich die wichtigsten Wissenschaftsorganisationen und Forschungsförderer – das SBF1, swissuniversities und der SNF – zur Forderung und Förderung von offenen Forschungsdaten bekannt.¹ Angesichts dieser Entwicklungen hat die ZHAW ihr Commitment zur Open Research Data ebenfalls zum Ausdruck gebracht. In der [F&E Policy der ZHAW](#) (Ziffer 3.2) wird folgendes festgehalten:

Grundsätzlich werden alle F&E-Leistungen der ZHAW vorbehältlich vertraulicher Informationen [...] in geeigneter Form zur Verbreitung in der Öffentlichkeit bzw. wissenschaftlichen Gemeinschaft freigegeben.

Ziel dieser Arbeitsanleitung ist es, die Forschenden der ZHAW auf dem Weg zu Open Research Data zu unterstützen. Die Arbeitsanleitung enthält grundsätzliche und fachübergreifende Empfehlungen und Hilfestellungen. Fachspezifische Besonderheiten werden nur bedingt berücksichtigt.

2. Grundsätze von Open Research Data

2.1 Was ist Open Research Data?

Unter «Open Research Data» (ORD) versteht man allgemein Forschungsdaten, die für weitere Nutzung und Weiterverarbeitung öffentlich zugänglich gemacht wurden. Dabei meint der Begriff «Open» mehr als die reine Datenpublikation. Er umfasst gleichsam eine breite Palette von Workflows – wie z.B. die technische Aufbereitung von Daten, ihre Anonymisierung und Dokumentation – mit dem Ziel, den Zugang zu Daten gemäss den [FAIR-Prinzipien](#) zu ermöglichen, d.h. Daten auffindbar (**F**indable), zugänglich (**A**ccessible), interoperabel (**I**nteroperable) und nachnutzbar (**R**e-usable) zu teilen:

- **F:** «Findable» - Die Daten werden mit standardisierten Metadaten und eindeutigen Identifikatoren versehen sowie in einer durchsuchbaren Ressource registriert, damit sie sowohl von Menschen als auch von Maschinen leicht gefunden werden können.
- **A:** «Accessible» - Auf (Meta-) Daten kann zugegriffen werden.
- **I:** «Interoperable» - (Meta-) Daten werden mit unterschiedlichen Computersystemen kompatibel, u. A., weil sie in offenen Dateiformaten vorliegen.
- **R:** «Reusable» - Metadaten und ergänzende Dokumentation liefern ausreichend aussagekräftige Kontextinformationen, die erlauben, Daten richtig zu interpretieren und somit auch nachnutzen.

¹ Siehe dazu [Nationale Schweizer Strategie für Open Research Data](#), die im Auftrag des Staatssekretariats für Bildung, Forschung und Innovation SBF1 von swissuniversities erarbeitet und im Jahr 2021 verabschiedet wurde. Siehe auch die [Grundsatzklärung des SNF zu Open Research Data](#).

2.2 Wozu dient Open Research Data?

Eines der Ziele von ORD besteht in der Gewährleistung der Transparenz und der Nachvollziehbarkeit von Forschungsergebnissen. Ein weiterer Mehrwert von ORD zeigt sich in ihrer möglichen Sekundärnutzung, was unnötige Doppelarbeit bei der Datenerhebung verhindert, die häufig mit hohen personellen, zeitlichen und finanziellen Ressourcen verbunden ist. Es ist daher ratsam, vor jeder Datenerhebung zu überprüfen, ob identische oder vergleichbare Daten bereits erhoben und zugänglich gemacht wurden. In Forschungsprozessen sollte deshalb nicht nur eine Literatur-, sondern auch eine Datenrecherche durchgeführt werden.

2.3 Welche Forschungsdaten soll/kann/muss/darf ich veröffentlichen?

Die Veröffentlichung von allen Forschungsdaten, die im Rahmen eines Projekts erhoben werden, ist nicht immer möglich und in unterschiedlichem Mass sinnvoll. Bei der Entscheidung über die Datenpublikation sind sogenannte Soll-, Kann-, Muss- und Darf-Kriterien grundsätzlich zu beachten:

- Forschungsdaten **sollen** grundsätzlich veröffentlicht werden, wenn sie für die Nachvollziehbarkeit der Ergebnisse einer wissenschaftlichen Publikation notwendig sind und/oder wenn sie das Potenzial für eine Sekundärnutzung haben.
- Forschungsdaten **können** veröffentlicht werden, sofern keine technisch oder finanziell unverhältnismässige Hürden vorliegen. Einige Forschungsförderer übernehmen bestimmte Kosten für die Zugänglichmachung von Forschungsdaten (z.B. bietet hierfür der SNF eine Unterstützung mit bis zu 10,000 CHF an).² Generell gilt die Empfehlung, dass der Aufwand der Veröffentlichung immer vom potenziellen Impact der Daten abhängig gemacht wird.
- Forschungsdaten **müssen** veröffentlicht werden, wenn eine Verpflichtung besteht (z.B. seitens Forschungsförderers). So erwartet z.B. der SNF von allen durch ihn unterstützten Forschenden, «dass sie zumindest alle mit einer Veröffentlichung zusammenhängenden Daten teilen»³ und dass sie eine Begründung angeben, wenn Daten nicht veröffentlicht werden.⁴
- Forschungsdaten **dürfen** veröffentlicht werden, wenn keine gesetzlichen Einschränkungen (z.B. aufgrund von datenschutzrechtlichen (siehe Ziffer 3.2.3) oder forschungsethischen Regelungen) oder vertraglichen Restriktionen (festgehalten z.B. in Kooperationsvereinbarungen, Förderverträgen oder Geheimhaltungsvereinbarungen (Non Disclosure Agreements)) vorliegen.

3. Workflows für Open Research Data

3.1 Planung

Um potenzielle Herausforderungen auf dem Weg zu Open Research Data frühzeitig zu erkennen und auf diese rechtzeitig reagieren zu können, hat sich das Erstellen von Datenmanagementplänen (DMP) als Best Practice etabliert. Ein DMP dient dazu, den Umgang mit Forschungsdaten, die im Rahmen eines Projekts anfallen, strukturiert und entlang ihres gesamten [Lebenszyklus](#) zu beschreiben. Der DMP ist somit ein Dokument, das einen Überblick darüber

² Für weitere Details betreffend die SNF-Finanzierung der Zugänglichmachung von Forschungsdaten siehe die [FAQ «An welchen Kosten für Open Research Data leistet der SNF einen Beitrag?»](#)

³ Siehe dazu die [FAQ «Welche Daten müssen geteilt werden?»](#)

⁴ Siehe dazu die [FAQ «Was passiert, wenn ich meine Daten nicht teilen kann?»](#)

gibt, wie die Forschungsdaten während des Projekts erhoben, verarbeitet und gespeichert werden und wie sie nach dem Projektende geteilt und archiviert werden.

Verschiedene Förderorganisation – z.B. der SNF und die EU – verlangen die Einreichung eines DMP nach der Bewilligung der Projektförderung. Zum Erstellen eines DMP für SNF-Projekte steht den ZHAW-Forschenden [dieses Template](#) zur Verfügung. Das Template enthält eine Anleitung zum Erstellen des DMP, Hilfetexte, Empfehlungen zur Datenaufbereitung sowie einige Beispielantworten.

3.2 Aufbereitung zu einer Veröffentlichung

Forschungsdaten müssen in nachnutzbaren Dateiformaten, mit nachvollziehbarer Dokumentation und unter Berücksichtigung des Datenschutzes (ggf. Anonymisierung) zur Publikation aufbereitet werden.

3.2.1 Nachnutzbare Dateiformate wählen

Nachnutzbare Dateiformate sind solche, die eine langfristige Archivierung und Sekundärnutzung von Forschungsdaten erlauben. Diese Formate sollen so weit wie möglich offen und standardisiert sein, um sicherzustellen, dass sie auch künftig von nicht-kommerziellen Software-Programmen geöffnet und verarbeitet werden können. *Idealerweise sollen Daten bereits bei der Erhebung in nachnutzbaren Formaten gespeichert werden, da durch eine spätere Konvertierung Risiken von Informationsverlusten entstehen.*

Hier sind einige Empfehlungen mit Beispielen für nachnutzbare Dateiformate:

- Empfehlungen der Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST):
 - [Katalog archivischer Dateiformate](#)
 - [Bewertungsmatrix zur Archivtauglichkeit von Dateiformaten](#)
- Empfehlungen der ETH Zürich: [Archivtaugliche Dateiformate](#)
- Empfehlungen von forschungsdaten.info: [Formate erhalten](#)

3.2.2 Dokumentation vorbereiten

Um Forschungsdaten korrekt zu interpretieren ist es oft notwendig, zusätzliche Kontextinformationen zu haben. Beispiele für solche Informationen sind Erklärungen zur Datenerhebungsmethode, Variablenbeschreibungen oder Informationen zur Anonymisierung der Daten. Im Sinne der FAIR-Prinzipien ist es äußerst wichtig, diese Kontextinformationen zu dokumentieren und zusammen mit Forschungsdaten zu veröffentlichen. Dazu gehören im Falle einer skriptbasierten Datenprozessierung auch die verwendeten Codes und deren Versionskontrolle. Idealerweise soll die Dokumentation bereits während des Projekts fortlaufend und regelmässig erfasst werden. Die spezifischen Kontextinformationen, die dokumentiert werden sollten, und die beste Methode, um dies zu tun, hängen von verschiedenen Faktoren ab, wie der Art und dem Inhalt der Daten. Hier sind einige Beispiele und Empfehlung hierfür:

- [forschungsdaten.info](#): Dokumentation – Warum, was und wie?
- [forschungsdaten-bildung.de](#): Forschungsdaten dokumentieren

3.2.3 Anonymisierung durchführen

Um den Datenschutz der betroffenen Personen bei der Veröffentlichung von Forschungsdaten zu gewährleisten, ist Anonymisierung ein wichtiger Schutzmechanismus. Anonymisierung im

Sinne der [DSGVO](#) beschreibt Verfahren, die jeglichen Personenbezug unwiderruflich aus den Daten tilgen – also eine Re-Identifikation des Einzelnen nahezu ausschliessen: Die betreffende Person darf nicht mehr mit „vertretbarem Aufwand“ identifizierbar sein.

Anonymisierte Daten gelten nicht als Personendaten. Sie fallen daher nicht (mehr) unter das Datenschutzrecht und dürfen somit grundsätzlich publiziert werden.

Zu unterscheiden sind in diesem Zusammenhang pseudonymisierte Daten. Bei pseudonymisierten Daten wird anstelle des Namens und der anderen identifizierenden Informationen ein bestimmter Schlüssel (Pseudonym) verwendet. Für pseudonymisierte Daten gelten weiterhin die Regeln des Datenschutzrechtes, da mit dem Pseudonymisierungsschlüssel der Personenbezug wieder hergestellt werden kann.

4. Veröffentlichung in einem Repositoryum

4.1 Welches Repositoryum eignet sich für die Veröffentlichung meiner Forschungsdaten?

Bei der Wahl eines Repositoryums sind verschiedene Aspekte zu berücksichtigen:

– **Schritt 1: Vorgaben der Forschungsförderer bzw. der Projektpartner beachten**

Häufig wird in Kooperationsverträgen oder in Förderrichtlinien festgelegt, wo Forschungsdaten nach dem Projektende veröffentlicht werden müssen oder dürfen. So erwartet z.B. der SNF, dass Forschungsdaten in Repositoryen publiziert werden, die nicht kommerziell ausgerichtet sind und die die FAIR-Prinzipien erfüllen.

– **Schritt 2: Wenn keine spezifischen Vorgaben gemacht wurden...**

...wird allgemein die Wahl eines etablierten fachspezifischen Repositoryums empfohlen.⁵ Sollte es in Ihrem Fachbereich keine solchen Repositoryen geben, empfiehlt sich die Publikation in einem generischen Repositoryum (z.B. [Zenodo](#)). Für die Publikation von Codes und Software eignen sich am besten spezielle Code-Repositoryen wie [GitHub](#), [GitLab](#), [Bit-Bucket](#) oder [SourceForge](#).⁶

Es ist allerdings entscheidend, dass das gewählte Repositoryum eine breite Reichweite für veröffentlichte Forschungsdaten bietet. Die Veröffentlichung sollte somit grundsätzlich in Repositoryen erfolgen, die in Ihrer Forschungsgemeinschaft gut etabliert und bekannt sind. Falls Sie nicht sicher sind, welche Repositoryen dies sind, können Sie sich folgende Frage stellen: In welchen Repositoryen suche ich in der Regel nach Forschungsdaten für Sekundäranalysen? Es kann auch hilfreich sein, Kolleg:innen aus Ihrem Fachbereich nach Empfehlungen zu fragen oder zu recherchieren, wo andere Autor:innen, die zu einem ähnlichen Thema geforscht haben, ihre Forschungsdaten publiziert haben.

– **Schritt 3: Publierte Forschungsdaten in der ZHAW digitalcollection nachweisen**

Die ZHAW betreibt kein eigenes institutionelles Repositoryum für die Publikation von Forschungsdaten. Mitarbeitende sind aber aufgerufen, veröffentlichte Forschungsdaten in der [ZHAW digitalcollection](#) nachzuweisen. Hierfür wird ein Eintrag mit den beschreibenden Metadaten der Publikation angelegt (siehe [Beispiel](#)).

⁵ Als Hilfestellung bei der Wahl eines fachspezifischen Repositoryums bieten sich Bewertungskriterien für Datenrepositorien an, die im Rahmen des [DLCM-Projekts](#) zusammengestellt wurden (siehe dazu [«Data Life Cycle Management Pilot Projects and Implications for Research Data Management at Universities of Applied Sciences»](#), dort Tabellen I und II)

⁶ Für weitere Details zu Code-Repositoryen siehe [FDM in der Informatik](#), dort Seiten 134-136.

4.2 Überlegungen zum Zugang

Eine sorgfältige Überlegung darüber, welche Zugriffsmöglichkeiten für welche Nutzergruppe gewährt werden sollen, ist ebenfalls von Bedeutung:

- **Open Access:** Sofern keine rechtlichen oder vertraglichen Gründe entgegenstehen, ist es ratsam, den vollständigen Zugriff auf Ihre Forschungsdaten für alle Nutzenden zu ermöglichen. In diesem Fall soll die Sekundärnutzung der Daten idealerweise über eine standardisierte Lizenz – z.B. eine Creative-Commons-Lizenz oder eine Open-Source-Lizenz – geregelt werden (siehe dazu Ziffer 4.4).
- **Restricted Access:** In bestimmten Fällen ist es sinnvoll, den Zugriff auf Forschungsdaten einzuschränken, sodass nur Metadaten der Daten öffentlich zugänglich sind. Der vollständige Zugriff auf die Daten ist dagegen nur einer bestimmten Gruppe von Nutzenden gestattet. Hier ist ein Beispielfall:
 - In der Einwilligungserklärung (Informed Consent) zur Erhebung von personenbezogenen Daten wurde festgelegt, dass diese Daten nach einer vollständigen Anonymisierung nur für Forschungszwecke weitergegeben werden. Im Falle einer Open-Access-Publikation kann allerdings nicht eruiert werden, zu welchem Zweck die jeweiligen Datensätze heruntergeladen werden. Im Gegensatz dazu ermöglicht der Restricted Access, die Nutzungsbedingungen mit der anfragenden Person zu vereinbaren.
- **Embargo:** Ein zeitliches Embargo definiert einen Zeitraum, in welchem lediglich Metadaten von Forschungsdaten in einem Repository einsehbar sind, während die Daten selbst für alle Nutzenden unzugänglich bleiben. Nach dem Ablauf der definierten Sperrfrist werden Forschungsdaten vollständig verfügbar gemacht. Ein Embargo ist z.B. sinnvoll, wenn Forschungsdaten während des Peer-Review-Prozesses einer mit den Daten in Verbindung stehenden Publikation veröffentlicht werden sollen.

Die Dauer einer Sperrfrist kann variieren. Grundsätzlich gilt die Regel: so früh wie möglich, so spät wie nötig. Wenn beispielsweise Daten zu einem Manuskript gehören, sollten sie idealerweise gleichzeitig mit diesem Manuskript publiziert werden.

Verschiedene Repositorien ermöglichen verschiedene Zugriffsoptionen auf Forschungsdaten. Wenn Sie sich für eine bestimmte Zugriffsoption entscheiden, müssen Sie auch sicherstellen, dass das gewählte Repository diese ermöglicht.

4.3 Wo finde ich ein passendes Repositoryum?

Über [re3data](#), eine Suchmaschine für Repositorien, können Sie gezielt nach einem geeigneten Repositoryum suchen. Dort stehen verschiedene Filteroptionen zur Verfügung, mit denen Sie Ihre Suche eingrenzen können, z.B. nach Fachbereichen oder bestimmten Zugriffsoptionen.

4.4 Unter welcher Lizenz soll ich meine Daten veröffentlichen?

Die ZHAW bekennt sich zu Open Science. Deshalb empfiehlt es sich, Forschungsdaten unter einer Creative-Commons-Lizenz (CC-Lizenz) zu veröffentlichen, solange keine rechtlichen oder vertraglichen Einschränkungen vorliegen.

- [Überblick über CC-Lizenzen und ihre Vorgaben](#)
- Das Tool «[License chooser](#)» unterstützt bei der Auswahl einer CC-Lizenz
- [Moodle-Kurs](#) der ZHAW «FAQ zu Urheberrecht, CC-Lizenzen und OER»



Die CC-Lizenzen eignen sich jedoch weniger für die Veröffentlichung von Code. In diesem Fall empfiehlt sich die Nutzung von [Open Source Lizenzen](#).

Beachten Sie, dass die Veröffentlichung unter einer der oben erwähnten Open-Science-Lizenz von der ZHAW als Rechteinhaberin genehmigt werden muss (üblicherweise ein Linienentscheid durch die vorgesetzte Person).

5. Support an der ZHAW

5.1 ZHAW Services Forschungsdaten (ZSF)

Die ZHAW Services Forschungsdaten (ZSF) bietet eine Unterstützung und Beratung zu allen Themen rund um Forschungsdaten. Sie besteht aus Mitarbeitenden der Hochschulbibliothek, dem Ressort F&E und der ICT, womit bereits bestehende Kompetenzen und Ressourcen gebündelt werden. Die Servicestelle bearbeitet individuelle Anfragen von Forschenden und nimmt auch gerne Rückmeldungen und Inputs entgegen.

– **Kontakt:** researchdata@zhaw.ch

5.2 Service Portal der ZHAW

Informationen zu bestehenden ZHAW-Services rund um Forschungsdaten (z.B. Tools, Schulungen, Checklisten, FAQ) finden Sie im [Service Portal der ZHAW](#).

6. Erlassinformationen

Die englische Übersetzung des Erlasses finden sie unter: [Z AA Instructions on ORD.pdf](#)

Betreff	Inhalt
Erlassverantwortliche:r	ZHAW Services Forschungsdaten
Beschlussinstanz	Leiter:in Research & Infrastructure, HSB
Themenzuordnung	6.07.00 Bibliothek
Publikationsart	Public