



e·infra
austria

AP 4 Institutionelle Repositorien

Der Ablageprozess von Forschungsdaten und was von Zenodo gelernt werden kann

Thomas Haselwanter , Heike Thöricht 

Universität Innsbruck

September 2019



Der Inhalt dieser Veröffentlichung steht unter einer Creative Commons
Namensnennung 4.0 Lizenz (<http://creativecommons.org/licenses/by/4.0/>)

[DOI 10.25651/1.2019.0006](https://doi.org/10.25651/1.2019.0006)

Über e-Infrastructures Austria Plus

Das Projekt „e-Infrastructures Austria Plus“ (2017-2019) ist ein vom österreichischen Bundesministerium für Bildung, Wissenschaft und Forschung (<https://bildung.bmbwf.gv.at/>) gefördertes Projekt von neun österreichischen Universitäten. Ziel des Projekts ist der koordinierte Aufbau eines österreichischen Netzwerks zur Einrichtung und Weiterentwicklung gemeinsamer e-Infrastrukturen durch Bündelung von Ressourcen und vorhandenem Wissen.

Einleitung

Die Entwicklung in Richtung Open Science hinterlässt deutliche Spuren im aktuellen Forschungsprozess. Dies hat Auswirkungen für Forschende und Institutionen. Die Abbildung des Forschungsprozess, die Lokalisation der Stakeholder in dem Prozess und die Evaluierung der Bedürfnissen aufgrund der neuen Rahmenbedingungen können wesentlich zur Unterstützung der Forschenden beitragen und eine gemeinsame Weiterentwicklung von Universität und Forschenden fördern.

Ziel

Der Fokus des Arbeitspakets „Institutionelle Repositorien“ liegt auf der Erarbeitung eines musterhaften Vorgehens zur Auswahl eines institutionellen Repositoriums für Forschungsdaten. Durch die Darstellung eines detaillierten Ablageprozesses von Forschungsdaten in das Repositorium Zenodo lassen sich Empfehlungen für den Aufbau und Betrieb eines eigenen Repositoriums geben.

Der Ablageprozess von Forschungsdaten bei Zenodo

Die Wahl des geeigneten Repositoriums

Generell wird folgende Reihenfolge bei der Wahl eines geeigneten Repositoriums für die eigenen Forschungsdaten empfohlen:

- 1.) globales fachspezifisches Repositorium
- 2.) nationales fachspezifisches Repositorium
- 3.) institutionelles Repositorium

In diesem Beispiel wird das Repositorium *Zenodo* zur Beschreibung der Datenablage herangezogen, das für jede Person zugänglich ist und in dem jede Person die eigenen Daten ablegen kann.

Über Zenodo

Zenodo ist ein Ergebnis eines OPENAIRE Projekts, das von der Europäischen Kommission in Auftrag gegeben wurde, um ein Sammelregister für EU-finanzierte Forschung bereitzustellen. Das Repositorium wurde 2013 gelauncht.¹

„Zenodo, a CERN service, is an open dependable home for the long-tail of science, enabling researchers to share and preserve any research outputs in any size, any format and from any science.“²

Auf Zenodo kann jede Person ihre Daten in sämtlichen Formate und in diversen Status zu allen möglichen Themen speichern. Die maximale Dateigröße pro Datensatz beträgt derzeit 50 GB. Bei Bedarf kann hier eine höhere Datengröße beantragt werden.

¹ <http://about.zenodo.org/> (Stand: 16. September 2019)

² <https://github.com/Zenodo/zenodo> (Stand: 16. September 2019)

Auch mit dem Übertragen der Daten in das Repository bleiben die Eigentumsrechte bei den UrheberInnen. Englisch wird bei Texten bevorzugt, jedoch sind alle Sprachen möglich. Eine Vergabe von Lizenzen für die eigenen Forschungsdaten ist unumgänglich. Diese regeln den Re-Use und Use der Objekte. Die eingetragenen Metadaten stehen unter der CCo Lizenz.³

Die hochgeladenen Daten werden unverändert in das Repository übernommen.⁴ Die Betreibende von Zenodo sind für den Inhalt der Daten nicht verantwortlich. Die Daten werden versioniert, in Budapest gespiegelt und jede Nacht gebackupt. Für den Fall, dass Zenodo geschlossen wird, wird bestmöglich versucht, die Daten in ein geeignetes alternatives Repository zu übertragen.⁵

Zenodo wird nach dem Best Effort Prinzip betrieben und die FAIR Prinzipien werden empfohlen.⁶

Zenodo, das auf der Open Source Software Invenio basiert, ist zudem auch als Code auf Github verfügbar, mit dem man ein eigenes Repository aufbauen kann.

„Zenodo is a small layer on top of Invenio, a free software suite enabling you to run your own digital library or document repository on the web.“⁷

Bei der Einrichtung eines eigenen Repositoriums ist zu beachten, das Zenodo Branding nicht zu benutzen, sondern sein eigenes, da Zenodo (R) und das Zenodo Logo durch CERN geschützt sind und von der GPLv2 Lizenz nicht gedeckt sind.⁸

Nachteilig an Zenodo könnte die mögliche Kostenpflichtigkeit sein, die zukünftig entstehen kann. Bislang heißt es in den Terms of Use:

„Inhalte können von Personen ohne direkten Zugriff auf ein organisiertes Rechenzentrum kostenlos hochgeladen werden.“⁹

³ <http://about.zenodo.org/policies/> (Stand: 16. September 2019)

⁴ <http://about.zenodo.org/terms/> (Stand: 16. September 2019)

⁵ <http://about.zenodo.org/policies/> (Stand: 16. September 2019)

⁶ <https://about.zenodo.org/terms/> (Stand: 16. September 2019)

⁷ <https://github.com/Zenodo/zenodo> (Stand: 16. September 2019)

⁸ <https://zenodo.readthedocs.io/en/latest/> (Stand: 16. September 2019)

⁹ <https://about.zenodo.org/terms/> (Stand: 16. September 2019)

Der Prozess

Einloggen

Nach Eingabe von zenodo.org im Browserverlauf zeigt sich folgendes Bild:

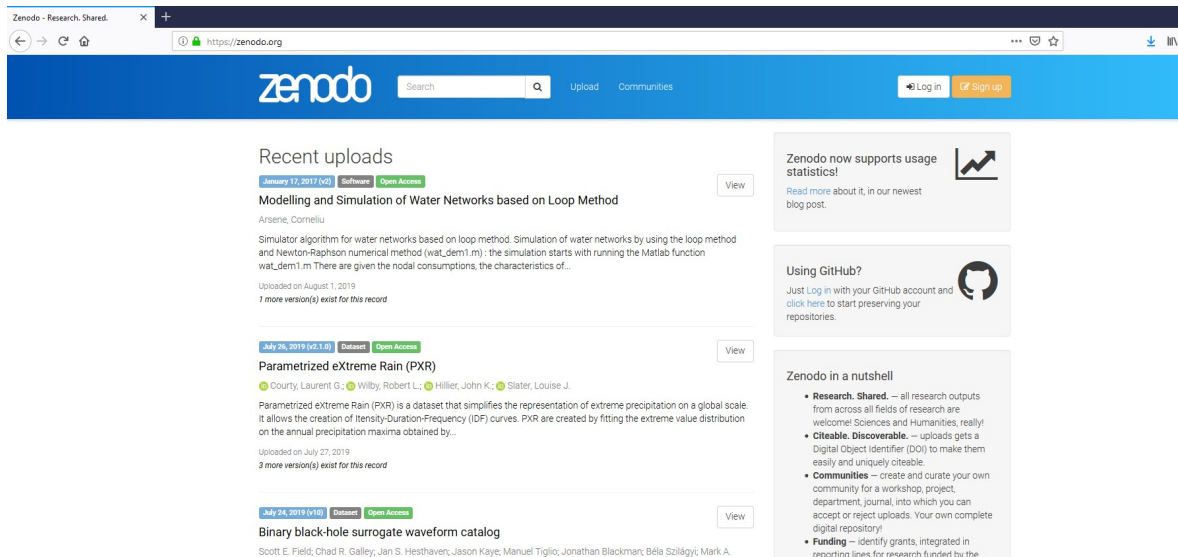


Abb. 1: Landingpage von zenodo.org

Die Startseite zeigt die neuesten Uploads von Open Access Daten und ermöglicht die Suche nach Begriffen und das Einloggen in das System. Beim erstmaligen Nutzen ist es notwendig, eine Emailadresse, einen Benutzernamen und ein Passwort zu hinterlegen. Mit dem Erhalt der Bestätigungsemail lässt sich die Emailadresse verifizieren. Direkt danach lässt sich Zenodo nutzen. Alternativ lässt es sich auch mit einem ORCID-oder GitHub-Konto einloggen.

Upload von Daten

Dank der einfachen Handhabung von Zenodo ist es direkt nach dem Einloggen möglich die eigenen Daten mittels Drag & Drop bzw. Uploadfunktion hochzuladen. Es lassen sich einzelne Daten oder mehrere Daten gleichzeitig hochladen.

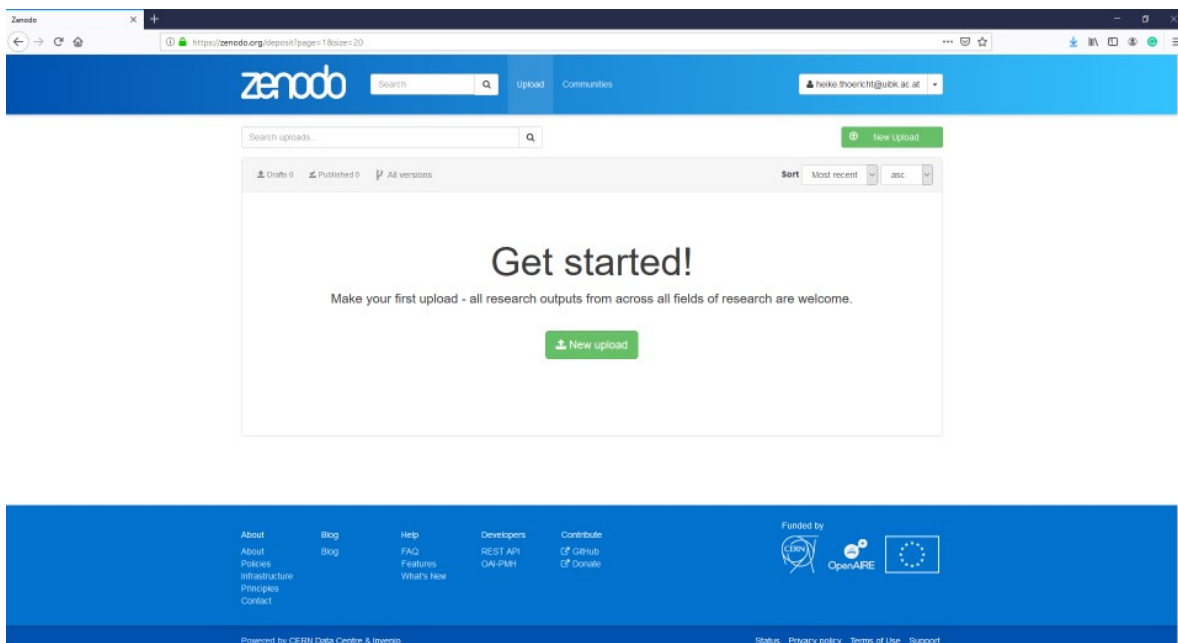


Abb. 2: Uploadfunktion bei zenodo.org

Nach dem Upload der Daten lässt sich der Datentyp bestimmen. Bei der Auswahl „Publication“ oder „Image“ lässt sich der Typ noch konkretisieren.

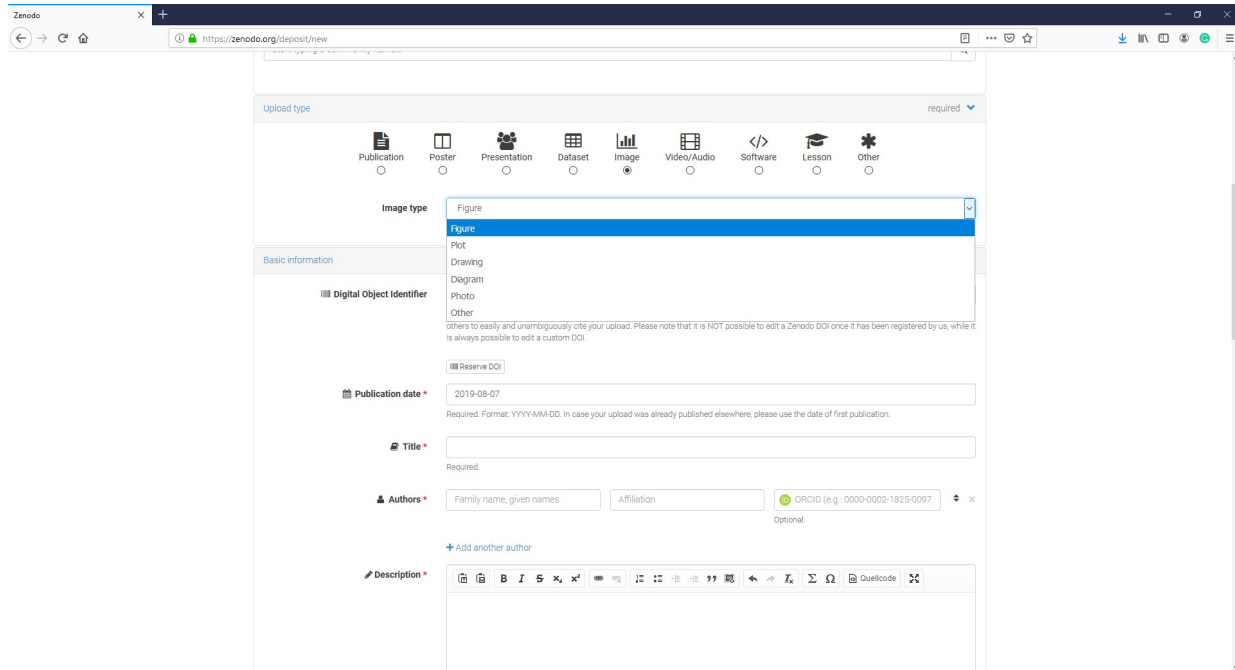


Abb. 3: Wahl des Imagetypen

Anschließend sind grundlegende Angaben wie Digital Object Identifier (DOI), Erstveröffentlichungsdatum, Titel, UrheberInnen und eine Beschreibung der Daten notwendig. Ohne vorliegenden DOI wird ein persistenter Identifikator durch das System vergeben. Vor dem Hochladen können sich Forschende einen DOI reservieren („reserve DOI“). Dieser wird vor dem Hochladen der Daten angezeigt und lässt sich so vorab in die Daten bei Bedarf eintragen. Für den Fall, dass mehrere Daten gleichzeitig hochgeladen werden, wird ein DOI für den Datensatz verwendet. Zu den UrheberInnen lassen sich auch ORCID iDs angeben, so dass eine automatische Verknüpfung zum bestehenden ORCID System erfolgen kann. Angaben zur Version, Sprache, Schlüsselwörter und zusätzliche Anmerkungen sind optional möglich.

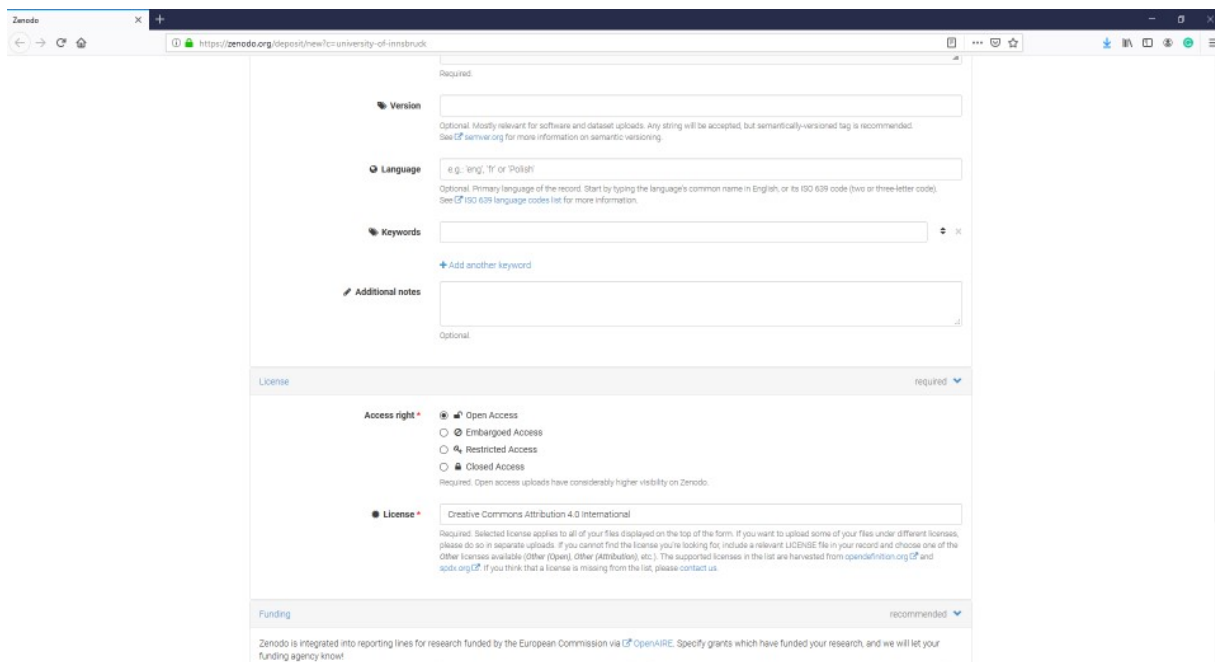


Abb. 4: Zugriffsrechte und Lizenzen

Verpflichtend sind im Anschluss Angaben zu Zugriffsrechten und die Wahl einer Lizenz. Mögliche Optionen sind Open Access (CC Attribution 4.0 International), Zugriffsrecht mit Embargo (mit der CC Attribution 4.0 International Lizenz nach dem Veröffentlichungszeitpunkt), eingeschränkter Zugriff und kein Zugriff. Bezüglich „Closed Access“ ist darauf hinzuweisen, dass Zenodo nicht geeignet für geheime oder vertrauliche Daten ist. „Closed Access“ bedeutet, dass andere UserInnen keinen Zugriff auf die hochgeladenen Daten haben. Allerdings werden die Daten unverschlüsselt gelagert und können von Betriebspersonal von Zenodo unter bestimmten Bedingungen angesehen werden.¹⁰

Auch Angaben zur Finanzierung lassen sich bei Zenodo eintragen. Nach Auswahl der FördergeberInnen und ersten Buchstaben zum Projekt werden Vorschläge seitens des Repositoriums angezeigt.

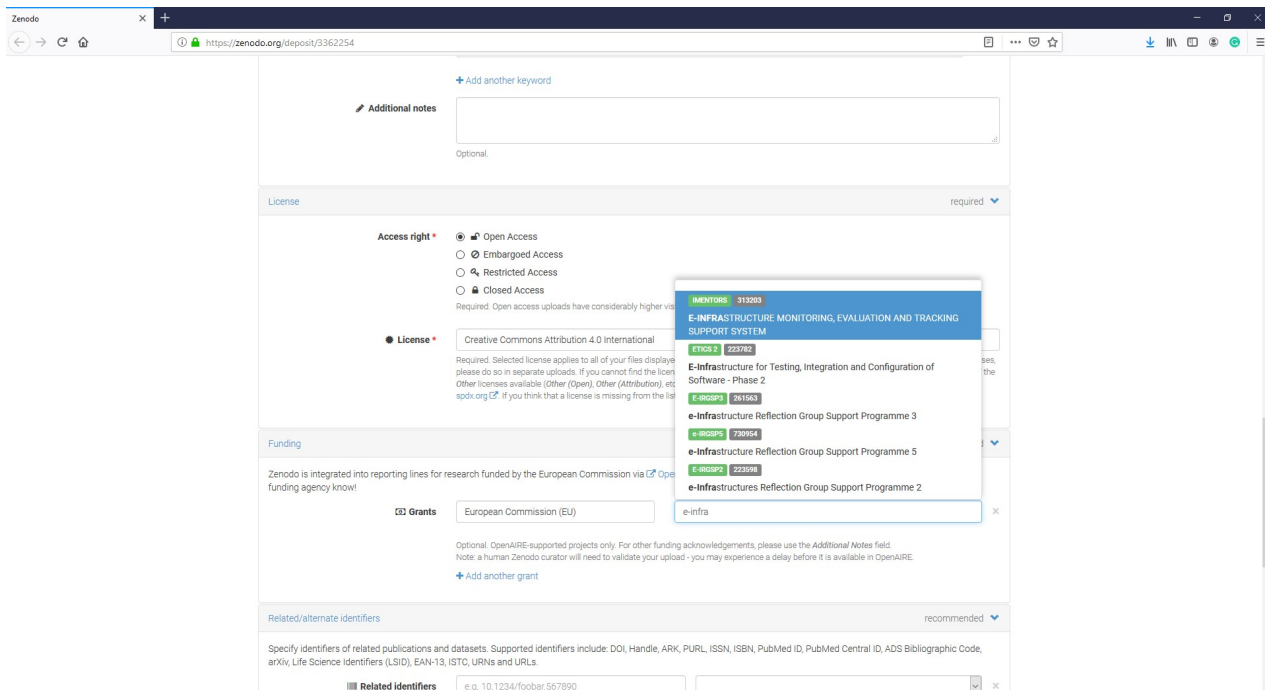


Abb. 5: Angaben zur Förderung

Weitere Informationen können Forschende angeben zu:

- Mitwirkende
- Referenzen (durch andere Identifikatoren)
- Journal
- Konferenz
- Buch
- Abschlussarbeit
- Ergänzung um spezifische Begriffe (aus der Taxonomie oder einem kontrolliertem Vokabular)

¹⁰ <https://about.zenodo.org/infrastructure/> (Stand: 16. September 2019)

Speichern und Veröffentlichen

Im Anschluss werden die Daten und die Eingaben mittels „Save“ gesichert. Mit „Publish“ lassen sich die Daten danach veröffentlichen. Zu beachten ist hier, dass einmal veröffentlichte Daten nicht mehr geändert werden können. Nichtsdestotrotz gibt es folgende Möglichkeiten, die Daten bei Bedarf zu ändern:

Wenn der Datensatz einen Zenodo-DOI enthält (Präfix: 10.5281), wird empfohlen, eine neue Version anzulegen. Dies kann bei folgenden Punkten sein notwendig: Änderungen des Datensatzes / der Software, Hinzufügung oder Entfernung von Dateien, neue Versionen von Datensätzen, Software oder Überarbeitungen der Papiere, einschließlich wichtiger Überarbeitungen sowie Tippfehler und Formatierungen / Formatierungen.

In Ausnahmefällen ist es möglich, das Betriebspersonal von Zenodo bezüglich kleinerer Anpassungen wie Tippfehler, versehentliches Auslassen wichtiger Dateien oder Einbeziehung von versteckten oder vertraulichen Dateien zu kontaktieren. Änderungen sind nur möglich, wenn die Veröffentlichung nicht länger als eine Woche zurückliegt.

Falls es sich um keinen Zenodo-DOI handelt, ist das Betriebspersonal direkt zu kontaktieren.

Zur Kontaktierung bezüglich Anpassungen wird ein [Kontaktformular](#) bereitgestellt. Das Betriebspersonal von Zenodo behält sich die Option offen, die Änderungen nicht umzusetzen, falls es sich nicht um eine kleine Anpassung handelt.¹¹

Fazit

Zenodo wurde für EU-finanzierte Forschungsprojekte konzipiert und das Long Tail Repository bietet die Möglichkeit Forschungsergebnisse jeglicher Größe, jeglichen Formats und jeglichen Forschungsrichtung dort abzulegen und zu veröffentlichen. **Kostenfreiheit wird denjenigen zugestanden, die keinen Zugang zu einem Datenzentrum haben.** Dadurch kann ein **institutionelles Repository für Forschungseinrichtungen mit Datenzentren zukünftig eine wichtige Option werden, wenn Zenodo kostenpflichtig werden sollte.**

Der Ablaufprozess der Datenablage zeigt Ähnlichkeiten zu einer üblichen Datenablage. Die verpflichtende Angabe von DOIs und Lizenzen ist eine Neuerung, mit denen Forschende zunehmend konfrontiert werden. Während diese Themen für Publikationen den Forschenden zum Teil bekannt sind, herrscht **große Unklarheit bezüglich der Veröffentlichung von Forschungsdaten** (z. B. Lizenzen, Urheberschaft). Für die Ablage und Veröffentlichung der eigenen Forschungsdaten ist eine **Entscheidung für eine Lizenz bzw. Zugriffsrechte jedoch unumgänglich.** Aufgrund dessen wird eine **Schärfung der Kenntnisse der Forschenden zu Rechten, Lizenzen und Urheberschaft durch Beratungen und Schulungen empfohlen.**

Darüber hinaus lässt sich als potentielle **Betreibende** eines (institutionellen) Repositoriums Verschiedenes von Zenodo lernen:

¹¹ <https://help.zenodo.org/> (Stand: 16. September 2019)

Entscheidungen zur Offenheit

Während Zenodo das Repository für jede Person zur Verfügung stellt, ist eine solche Entscheidung zur Offenheit ebenfalls vom Repositorienbetrieb zu treffen:

- Wer darf auf veröffentlichte Daten zugreifen?
- Wer darf dort die eigenen Daten ablegen?
- Welche Mengen, Formate und Versionen von Daten darf man dort ablegen?

Weitere Aspekte zur Offenheit sind unter <https://www.plos.org/how-open-is-it> einsehbar. Generell wird empfohlen, ein Repository „as open as possible, as closed as necessary“ zu gestalten. Darüber hinaus wird empfohlen, entsprechende Use Cases für die verschiedenen Gruppen zu erstellen und auf Basis derer und der vorliegenden Infrastruktur Prozesse zur Einrichtung des Zugangs zu erstellen.

Verantwortungen von UserInnen und Repositorienbetrieb

Die beiderseitigen Verantwortungen sind vorab zu klären, um die Kooperation von UserInnen und Repositorienbetrieb bestmöglich zu gestalten:

- Daten bleiben weiterhin im Besitz der UrheberInnen
- Keine inhaltliche Kontrolle der Daten seitens des Repositorienbetriebs
- Änderungen nur unter bestimmten Umständen möglich (kleine Anpassungen, innerhalb zeitlicher Frist) und die Entscheidung für die Anpassung obliegt dem Betriebspersonal des Repositoriums
- Keine Ablage von geheimen und vertraulichen Daten aufgrund des möglichen Zugriffs durch weitere Personen des Repositorienbetriebs

Usability¹²

Die Bedienungsfreundlichkeit ist ausschlaggebend, ob ein Repository genutzt wird. Deshalb werden folgende Umsetzungen beim eigenen Betrieb eines Repositoriums empfohlen:

- Gewohnte, intuitive Handhabung zum Einloggen und Hochladen der Daten
- Automatisierungen (z. B. Autovervollständigung, Auswahlmöglichkeiten)
- Übersichtlichkeit durch begrenzte Auswahl (z. B. an Lizenzen)
- Verknüpfung zu bestehenden Systemen (z. B. ORCID)
- Flexibilität für die UserInnen (z. B. 1 DOI für eine Datei oder für einen Datensatz)
- Default-Einstellungen (z. B. Voreinstellungen/-auswahl zur Unterstützung der FAIR-Prinzipien)
- Schlichtes Design mit Wiedererkennung

¹² User Experience war ebenfalls der Mittelpunkt der Überlegungen beim HRSM Projekt [Portfolio/ Showroom](#), bei dem eine Plattform auf Basis von User Stories, also tatsächliche Anwendungsszenarien von KünstlerInnen und WissenschaftlerInnen ins Leben gerufen wird. Die Plattform ermöglicht es, eigene Arbeiten anzukündigen, zu sichern und zu veröffentlichen und sich mit KollegInnen auszutauschen.

Kommunikation

Unabhängig von den Erfahrungen mit Zenodo wird empfohlen, Ressourcen für die Vermarktung eines Repositoriums zur Verfügung zu stellen. Nur wenn potentielle UserInnen davon wissen, kann dieses auch genutzt werden. Hier sollten verschiedene Kanäle zur Verbreitung herangezogen werden (z. B. Intranet, Veranstaltungen, Newsletter). Zudem sind folgende Informationen notwendig, um UserInnen, mögliche Anlaufstellen und Stakeholder (z. B. Bibliothek, Zentraler Informatikdienst) zu informieren. Diese Informationen sollten jedenfalls auf der Webseite des Repositoriums kommuniziert werden:

- Informationen zum Repositorium: About, Privacy Policy, Terms of Use, General Policies, Infrastructures, Principles, Roadmap, Contact
- Klare Kommunikation zur Urheberschaft der Daten und zum Umgang mit den Forschungsdaten seitens des Repositorienbetriebs
- Informationen zum Ablagesystem (z. B. hinsichtlich der Versionierung von Forschungsdaten)
- Empfehlung der FAIR Prinzipien
- Kontaktformular

Literatur

Public Library of Science (PLOS). HowOpenIsIt? <https://www.plos.org/how-open-is-it> [zuletzt abgerufen am 23. September 2019]

AutorInnen

Thomas Haselwanter hat an der Universität Innsbruck Angewandte Physik studiert und ist seit 2010 an der Universität Innsbruck im Zentralen Informatikdienst beschäftigt. Seit 2011 Abteilungsleiter für verschiedene Bereiche ist er derzeit für Web- und Informationssysteme zuständig und damit für viele der Web-Applikationen, die an der Universität Innsbruck in der Forschung eingesetzt werden. Er ist Projektleiter des HRSM Projekts „e-Infrastructures Austria plus“ und leitet im Projekt außerdem die Arbeitsgruppen Research Lifecycle und Institutionelle Repositorien.

Heike Thöricht, Diplom-Soziologin und Projektassistenz, unterstützt die Leitung des HRSM Projekts „e-Infrastructures Austria plus“ in den Bereichen Projektmanagement und Projektmarketing. Darüber hinaus arbeitet sie inhaltlich bei den Arbeitspaketen Research Lifecycle und Institutionelle Repositorien mit.