

Handreichungen zur Repositorienauswahl bei der Ablage von Forschungsdaten

Judith Köbler

Medizinische Universität Innsbruck



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

Einleitung

Seit längerem ist ein größerer Veränderungsprozess in der wissenschaftlichen Kommunikation im Gange: Open Science. Neben Open Access spielt auch Open Data als Teil der „Open Science Strategie“ eine immer wichtigere Rolle. Zusammen mit anderen Elementen wie Open Peer Review oder Open Methods soll der Schaffensprozess damit transparenter und partizipativer werden. Der diesbezügliche Transformationsprozess ist allerdings noch nicht abgeschlossen.

Ziel

Vor allem Fördergeber verlangen jedoch zunehmend die Ablage derjenigen Forschungsdaten, die einer Publikation zugrunde liegen, in einem öffentlich zugänglichen Repository. Der Ablageprozess in Bezug auf Forschungsdaten ist dabei nicht immer einfach. Diese Handreichungen sollen eine praktische Anleitung für Forscher*innen und Forscher bieten.

Handreichungen

Vorgaben der Fördergeber

Sollten Sie im Anschreiben des Fördergebers oder in den jeweiligen Förderrichtlinien konkrete Vorgaben zur Repositorienauswahl haben, sind diese zwingend einzuhalten.

Vorgaben der Herausgeber

Manche Verlage oder Journals geben auf ihren Webseiten Empfehlungen zu Repositorien, empfehlen die Vorabablage der Forschungsdaten in einem öffentlichen Repository (zB für die Reviewer) oder verlangen die Zurverfügungstellung aller oder ausgewählter Forschungsdaten als Supplement. Leider stimmen diese Vorgaben nicht immer mit den Vorgaben eines Fördergebers überein bzw. lassen sich die Anforderungen an Open Access für Referierte Publikationen nicht immer mit den Anforderungen an Open Access für Forschungsdaten in einer einzigen Datenablage erfüllen. Sollten Sie bei geplanten Publikationen konkrete Vorgaben des Verlags zur Ablage von Forschungsdaten erhalten, ist zunächst sinnvoll zu prüfen, ob diese mit den Vorgaben des Fördergebers vereinbar sind.

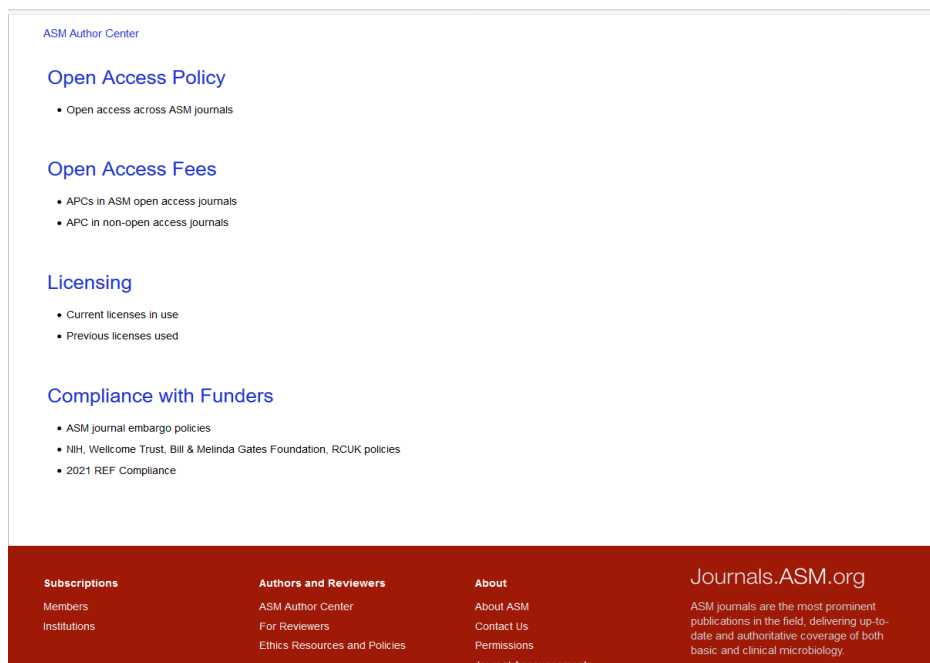


Abb. 1: AMS Journals „Compliance with Funders“

Lässt sich kein Repository finden, das allen Anforderungen genügt, ist ein anderes, separates Veröffentlichungsmedium für die Daten zu wählen, wobei hier wiederum die Vorgaben des Verlags zu berücksichtigen sind. Es kann in diesem Bereich also zu Zielkonflikten und Doppelungen kommen.

Europe PMC etwa löst diesen Konflikt für dort eingestellte Publikationen, in dem sowohl die Supplemental Files als auch die Forschungsdaten in externen Repositorien verlinkt werden¹.

¹ Europe PMC, Where is the primary data?,

https://erc.europa.eu/sites/default/files/content/pages/pdf/Europe_PMC_as_a_tool_for_researchers.pdf last accessed 2021-02-20.

Auswahl eines Repositoriums für FWF-Projekte

Für vom FWF geförderte Projekte gilt die Open Access Policy² des FWF. Sie ist zwingend einzuhalten. Hierzu wird auch ein Monitoring betrieben. Die Erfüllung der Open Access Verpflichtungen ist im Endbericht nachzuweisen.

Der FWF macht insbesondere spezielle Vorgaben für den Umgang mit Forschungsdaten³:

Für Forschungsdaten ist vorgesehen, dass institutionelle, disziplinspezifische oder disziplinübergreifende Repositorien gewählt werden können. Die Medizinische Universität Innsbruck etwa verfügt derzeit über kein eigenes institutionelles Repository.

Der FWF setzt bei einem Repository mindestens voraus:

- Listung in <https://www.re3data.org/>
- Persistenter Identifikator für Datensätze

Der FWF empfiehlt:

- <https://zenodo.org/>
- <https://datadryad.org/stash>
- <https://osf.io/>
- Zertifizierte Repositorien zB mit CoreTrustSeal (<https://www.coretrustseal.org/>)

Die drei empfohlenen Repositorien werden im Folgenden kurz vorgestellt.

Praktisches Vorgehen

Als Ausgangspunkt empfiehlt sich in jedem Fall die Webseite des Repositorienportals <https://www.re3data.org/>, auch wenn Sie bereits ein spezielles Repository bevorzugen. Re3data listet die wesentlichen Kriterien der dort hinterlegten Repositorien auf und macht sie damit gut vergleichbar. Dort gibt es auch verschiedene Suchmöglichkeiten. Es kann eine einfache Suche durchgeführt werden, zB wenn Sie den Namen eines Repositoriums schon wissen.

² Diese finden Sie unter <https://www.fwf.ac.at/de/forschungsfoerderung/open-access-policy> (last accessed 2021-02-02).

³ FWF, Open Access für Forschungsdaten, <https://www.fwf.ac.at/de/forschungsfoerderung/open-access-policy/open-access-fuer-forschungsdaten> (last accessed 2021-02-02).

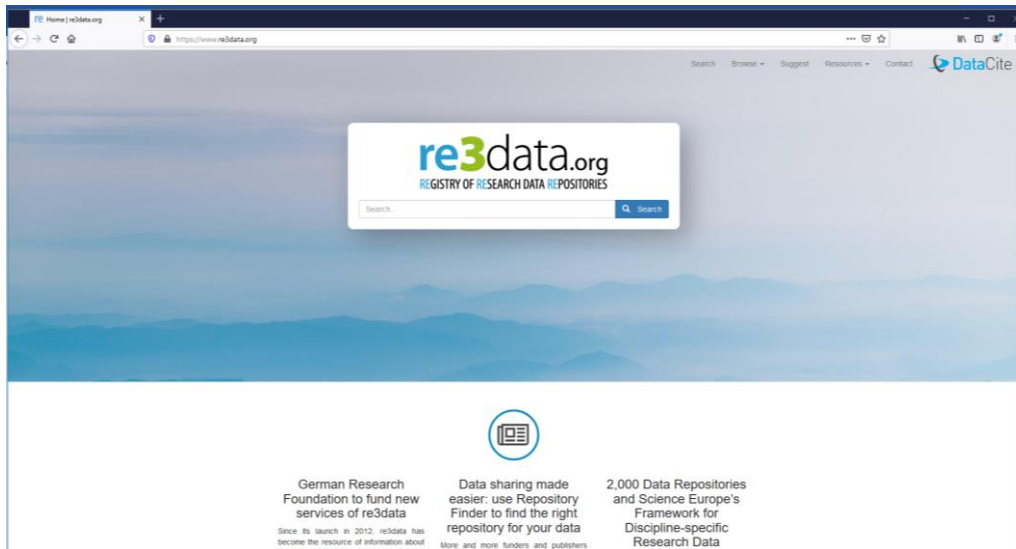


Abb. 2: Suchschlitz re3data

Es kann auch eine facettierte Suche durchgeführt werden, wenn Sie nach bestimmten Kriterien suchen wollen (zB ob ein Repositoryum zertifiziert ist).

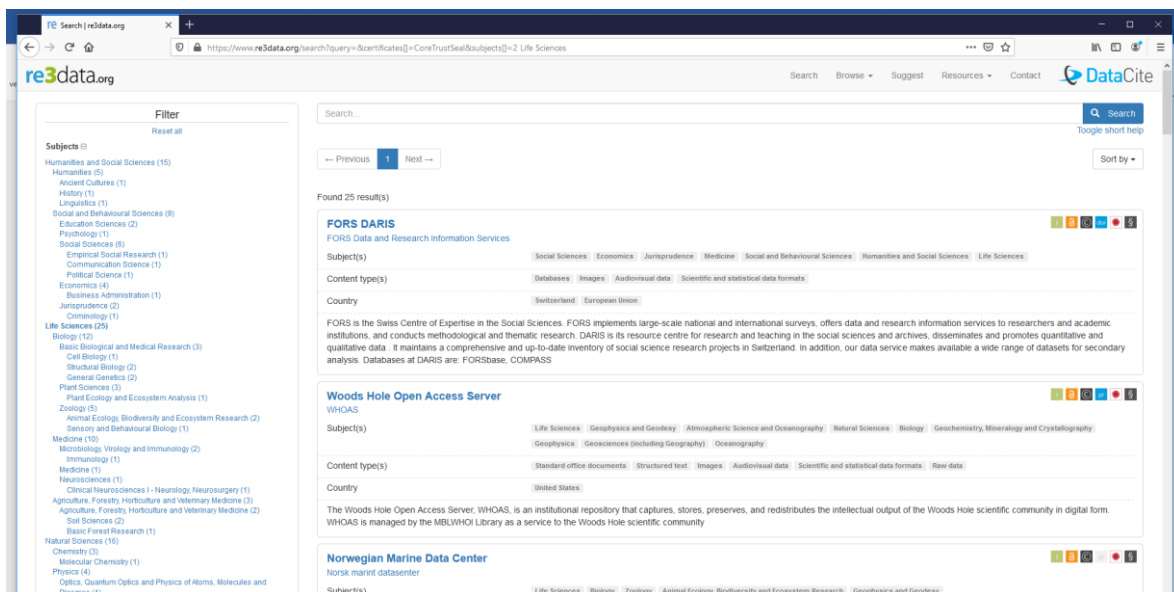


Abb. 3: Facettierte Suche bei re3data

Es kann eine spezifische Suche nach Disziplin, Land oder Inhalt durchgeführt werden. Dazu verwenden Sie bitten den Button „Browse“. Es erscheint eine graphische Darstellung (zB Kreissegmente, Karte). Sie können auch auf „Text“ umstellen, dann erscheint eine Auswahlliste. Durch Anklicken des jeweiligen Feldes können Sie die Auswahl immer mehr eingrenzen. Wenn das von Ihnen bevorzugte Fachgebiet nicht exakt abgebildet wird, können Sie Ihre Forschung evtl. einem benachbarten oder übergeordneten Fachgebiet zuweisen.

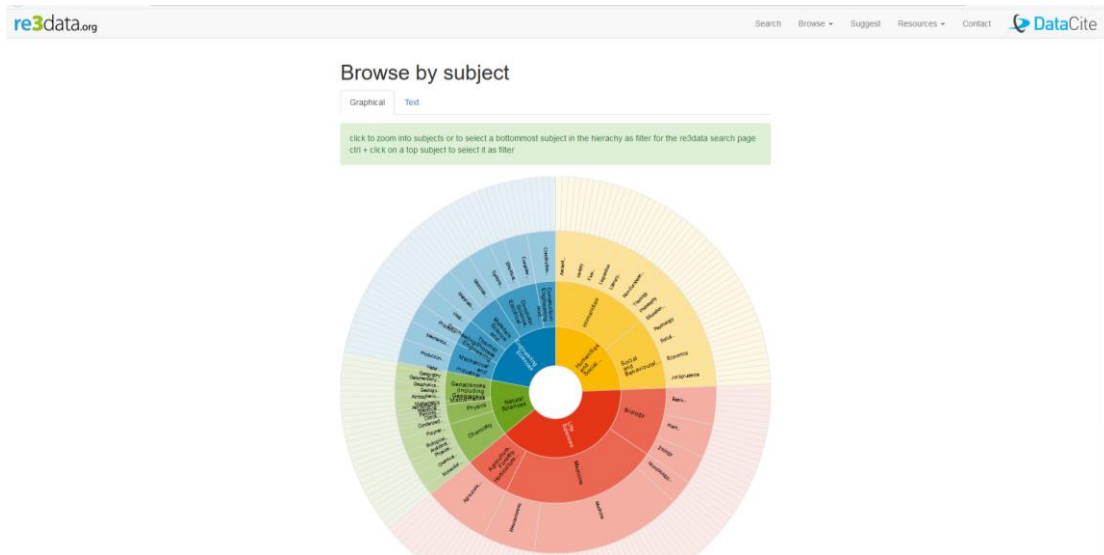


Abb. 4: „Browse by Subject“ bei re3data

Wenn Sie die Suche eingegrenzt haben, erscheint eine Übersicht mit vorgeschlagenen Repositorien. Dort erhalten Sie Informationen zu den enthaltenen Disziplinen, Datentypen oder dem Staat, in dem sich der Server befindet. Bevorzugt sollten Server gewählt werden, die sich innerhalb der EU befinden, insbesondere in Bezug auf sensible Gesundheitsdaten, deren Veröffentlichung ohnehin nur in anonymisierter oder pseudonymisierter Form erfolgen kann⁴.

Um die vorgeschlagenen Repositorien besser vergleichen zu können, zeigt re3data auf der rechten jeweils 6 Icons:

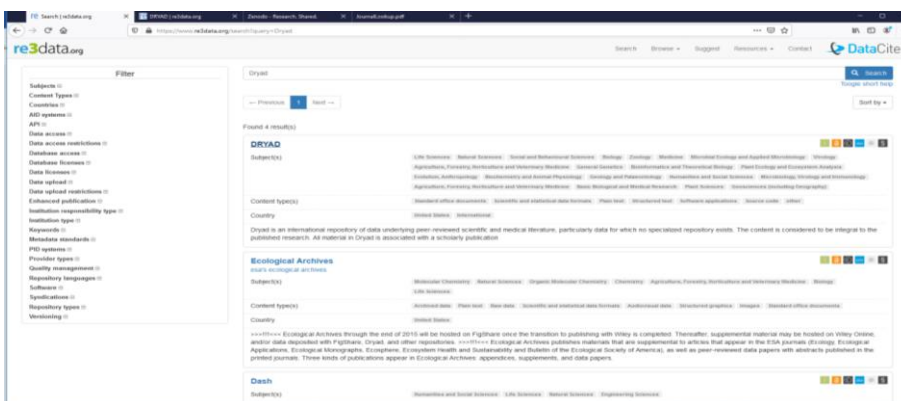


Abb. 5: Aktive Icons bei re3data

⁴ Da personenbezogene Daten unter anderem der Datenschutzgrundverordnung der EU unterliegen und daher nur in technisch sicherer und anonymisierter Form, falls nicht möglich in pseudonymisierter Form, weitergegeben werden sollen, kontaktieren Sie im Zweifelsfall bitte die Daten Clearing Stelle Ihrer Forschungseinrichtung oder eine ähnliche Stelle.

Fährt man mit der Computermaus über das Icon erhält man nähere Informationen. Für einen positiven Kurzcheck ist essentiell, dass folgende Icons zutreffend sind (dh nicht grau):

- ✓ Open access
- ✓ Terms of use and license
- ✓ Persistent Identifier

Hilfreich (zB bei der Erstellung eines Datenmanagementplans) sind auch Informationen über einen vorhandenen bzw. vom Repository verwendeten Metadatenstandard. Ob dies vorhanden ist, wird bei re3data wie folgt angegeben:

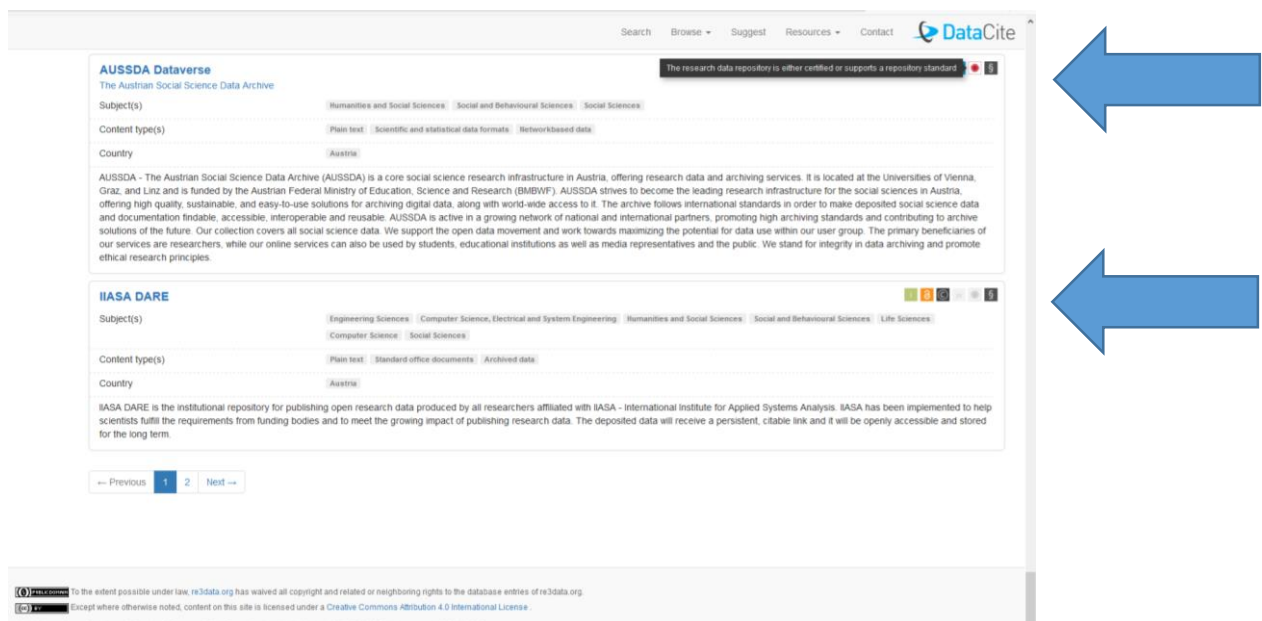


Abb. 6: Information zu Metadatenstandards nach re3data

Grundsatzempfehlung

Eine Grundsatzempfehlung seitens der eigenen Forschungseinrichtung ist umso schwieriger, je diverser das Forschungsfeld ist. Es können verschiedene Wege beschritten werden, um diese Forschung in Form von Daten nach außen sichtbar zu machen. Die Auswahl eines Datenrepositoriums hängt meiner Meinung nach wie die Frage nach dem richtigen Fachjournal für Publikationen – bei gegebener Erfüllung vorgeschriebener Kriterien wie Open Access - auch von der eigenen Persönlichkeit und dem derzeitigen Karrierestatus der Forscher*innen ab. Im Zuge der Altmetrics-Bewegung (<https://www.altmetric.com/>) könnten in Zukunft neben den Publikationen selbst auch die veröffentlichten Datensätze bzw. deren „clicks“ und „downloads“ eine zusätzliche Rolle spielen, sodass es mehr oder weniger wichtig sein kann, auch hierüber eine gewisse Sichtbarkeit in der Fachcommunity zu erzeugen.

Als Faustregel lässt sich folgende Reihenfolge empfehlen:

- Globales fachspezifisches Repository
- Nationales fachspezifisches Repository
- Institutionelles Repository.⁵

Sofern die eigene Forschungseinrichtung über ein institutionelles Repository verfügt, lassen sich die Forschungsdaten in der Regel dort einpflegen. Dies kann in Bezug auf die Handhabung und Verwaltung ein Vorteil, in Bezug auf die internationale Sichtbarkeit ein Nachteil sein, wenn das institutionelle Repository nicht über Schnittstellen in anderen Portalen sichtbar ist (zB <https://v2.sherpa.ac.uk/openoar/>). Falls Sie in einem Konsortium mit anderen Partnerorganisationen tätig sind, ist es mitunter auch möglich, die pseudonymisierten Daten in einem institutionellen Repository der Kolleg*innen abzulegen. Dies ist unter Umständen eine Frage, die in einem Datenmanagementplan (am besten vorab) geklärt werden könnte, gegebenenfalls in Verbindung mit einem Daten Clearing, wenn es sich um Nicht-EU bzw. EWR-Staaten handelt.

Zenodo

Ein Repository, das für die meisten Fälle geeignet sein wird, ist Zenodo. Es wird vom CERN in Zusammenarbeit mit OPEN AIRE (<https://www.openaire.eu/>) betrieben und von der Europäischen Union unterstützt. Zenodo entspricht den FAIR-Prinzipien⁶ und ist für Forscher*innen kostenlos. Der Upload ist mit drag & drop recht einfach. Für große Datenmengen kann angefragt werden, ob das Datenlimit hinaufgesetzt werden kann. Durch die „communities“ können die Forschungsdaten einer spezifischen Fachrichtung zugeordnet werden. Das Anlegen von communities ist dabei sowohl für Fachrichtungen, aber auch für Institutionen, Journals (zB British Journal of Medicine and Medical Research, International Journal of Medical, Medicine and Health Sciences) oder gemeinsame Initiativen mehrere Einrichtungen (zB LORY- Lucerne Open Repository) möglich. Diese communities bedürfen jedoch einer Kuratierung, dh sie müssen gepflegt werden.

⁵ T. Haselwanter/H. Thöricht, Der Ablageprozess von Forschungsdaten und was von Zenodo gelernt werden kann, Innsbruck, 2019, DOI [10.25651/1.2019.0006](https://doi.org/10.25651/1.2019.0006).

⁶ Näheres zu den FAIR Prinzipien s. TU Wien, <https://www.tuwien.at/forschung/fti-support/forschungsdaten/forschungsdatenmanagement/fair-prinzipien> (last accessed 2021- 02-02).

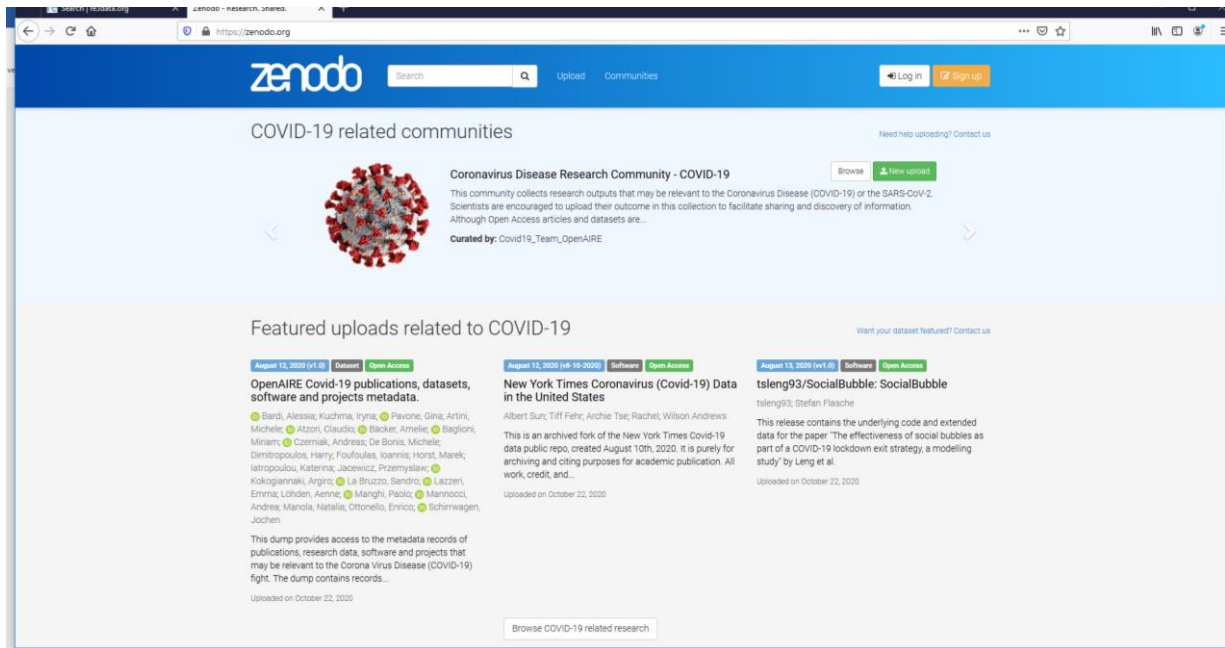


Abb. 6: Frontpage Zenodo

Dryad

Ein anderes, empfehlenswertes Repositorium ist Dryad. Es ist mitgliederbasiert und international. Es ist re3data mit einem ausführlichen Profil enthalten (zB die Fachdisziplinen).

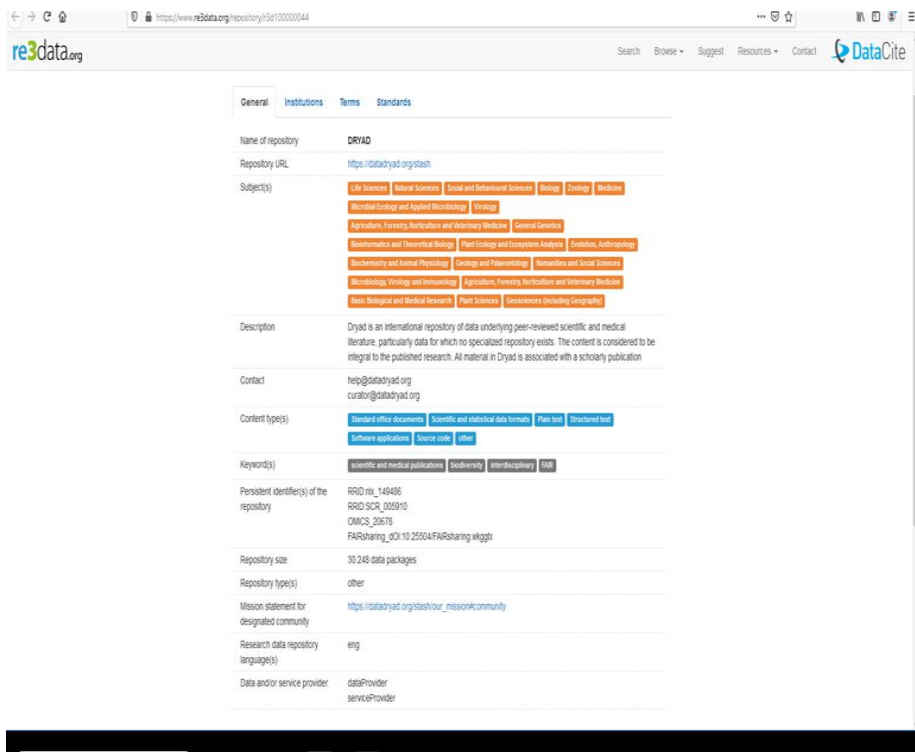


Abb. 7: re3data.org: DRYAD; editing status 2019-10-30; re3data.org - Registry of Research Data Repositories. <http://doi.org/10.17616/R34S33> last accessed: 2020-10-29.

Allerdings ist es möglich, dass die eigene Organisation (noch) nicht Mitglied ist. Es könnte aber eines der von Ihnen gewählten Journals Mitglied bei Dryad sein. Hierfür gibt es die Funktion:

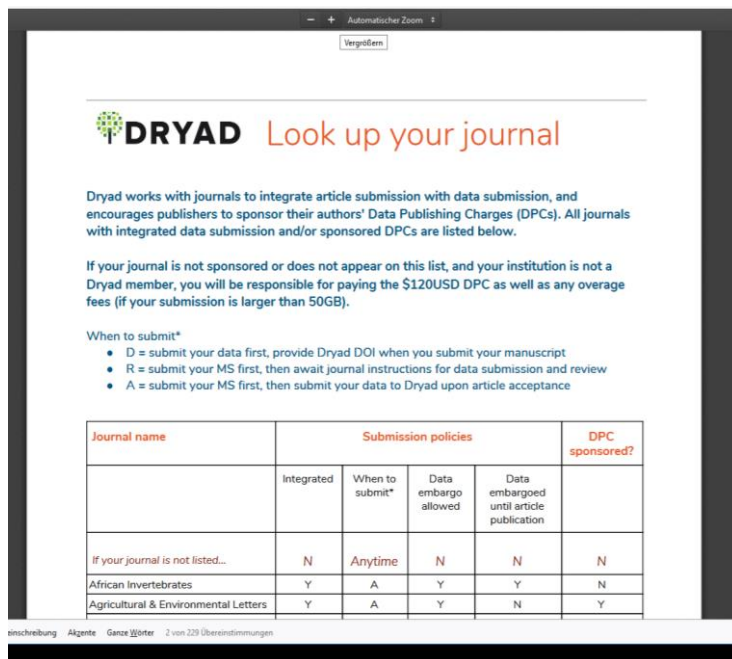


Abb. 8: Dryad Journals „Look up your journal“

Eine ausführliche Beschreibung der Dryad Dienste finden Sie hier: https://datadryad.org/stash/our_platform (zB Creative Commons licensing, DOI & ORCID⁷, DataCite Metadata Schema⁸). Grundsätzlich muss bedacht werden, dass die Daten auf Servern in den USA abgelegt werden und daher gegebenenfalls ein erhöhter Clearingaufwand zu betreiben sein wird.

Open Science Framework (OSF)

Ein weiteres, vom FWF empfohlenes Repositorium, ist das Open Science Framework. Es ist in re3data gelistet, interdisziplinär, offen, kostenlos, verwendet einen persistenten Identifikator, hat Nutzungsbedingungen, Lizenzen und verwendet eine Policy, ist aber nicht zertifiziert⁹.

Ausgewiesene Fachgebiete umfassen beispielsweise Lebenswissenschaften, Biologie, Medizin, Chemie und Mikrobiologie, aber auch Ökonomie oder Sozialwissenschaften.

⁷ ORCID ist ein persistenter Identifikator für Forscher*innen und wird von der gleichnamigen Organisation vergeben, s. <https://orcid.org/> (last accessed 2021-02-02).

⁸ Das Datenschema von DataCite ist mittlerweile in Version 4.3 veröffentlicht, <https://schema.datacite.org/meta/kernel-4.3/> last accessed 2021-02-02.

⁹ S. Datensatz in re3data: re3data.org: Open Science Framework; editing status 2020-02-11; re3data.org - Registry of Research Data Repositories. <http://doi.org/10.17616/R3N03T> last accessed: 2020-10-29.

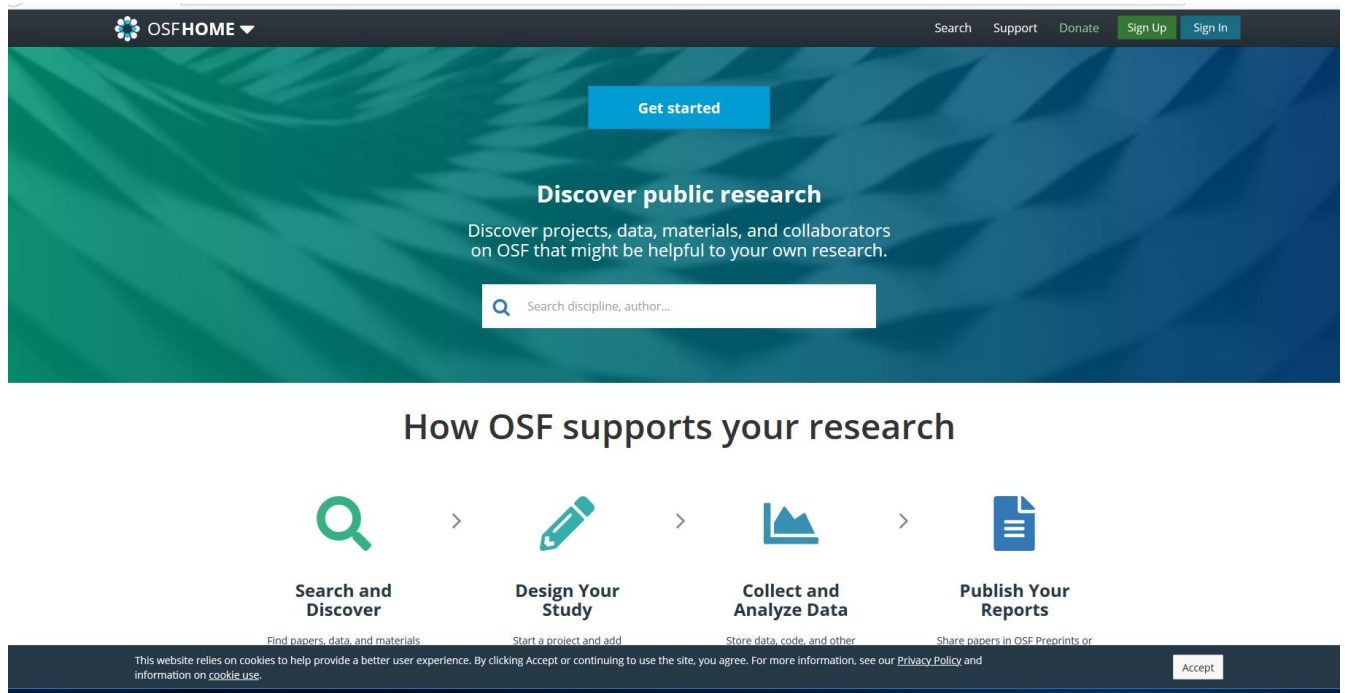


Abb. 9: Frontpage Open Science Framework

In OFS können ganze Projekte angelegt, dazu gehörende Dateien abgelegt und Beschreibungen verfasst werden.

Darüber hinaus versucht OSF DSGVO-Compliance herzustellen, ist aber zB nicht HIPAA¹⁰-konform. Als Speicherort für die Daten kann Frankfurt/Deutschland ausgewählt werden (siehe FAQ). OSF hat ein Dateilimit von 5 GB, kann aber durch Add-ons von Drittanbietern erweitert werden. Diese speichern dann aber gegebenenfalls nicht in Europa.

Ausblick

Die Umstellung des gesamten Forschungsprozesses auf Open Science wird noch einige Zeit in Anspruch nehmen. Ein elementarer Baustein dafür ist die Veröffentlichung der Forschungsdaten, die den Publikationen zugrunde liegen. Sie machen nicht nur das gefundene Ergebnis nachvollziehbar, sondern ermöglichen auch weiterführende Forschungsarbeiten. Ein solcher Transformationsprozess kann aber letztlich nur gelingen, wenn sich die Betroffenen diesen Veränderungsprozess zu eigen machen. Wichtig dafür ist unter anderem, den Umstieg durch vorhandene und neu zu schaffende Ressourcen und Strukturen so einfach wie möglich zu machen und die Forscher*innen so bestmöglich zu unterstützen.

¹⁰ US- Health Insurance Portability and Accountability Act of 1996 (HIPAA)
<https://www.cdc.gov/phlp/publications/topic/hipaa.html> (last accessed 2021-02-02).