

Forschungsdaten & Open Access – so publizieren Sie Ihre Daten

Skript zum Open Access Talk vom 09. Februar 2021 von Frauke Ziedorn

Folie 5

Sie forschen an einem Projekt und haben auch schon einige Ergebnisse in Papern veröffentlicht. Jetzt stehen Sie vor einem Haufen Daten und fragen sich: „Was hiervon soll ich denn noch veröffentlichen?“ Denn oftmals entstehen so viele Daten, dass es unmöglich ist, diese alle zu veröffentlichen. Dies erfordert schließlich nicht nur viel Speicherplatz, der über mindestens 10 Jahre zur Verfügung stehen muss, sondern es kostet auch Geld. Also müssen Sie eine Auswahl treffen, welche Daten sollen veröffentlicht, welche dauerhaft archiviert werden und welche Daten können auch gelöscht werden.

Bedenken Sie dabei: Nicht alle Daten müssen aufgehoben werden. Entscheiden Sie bewusst und begründet, was Sie aufheben und was nicht.

Bei der Entscheidung sollten Sie sich folgende Fragen stellen:

Gibt es rechtliche Gründe, die Daten aufzubewahren, oder Richtlinien, die dies vorgeben?

Sind die Daten einzigartig oder von besonderem wissenschaftlichem oder historischem Wert und daher von besonderem Belang?

Oder können die Daten relativ einfach reproduziert werden? Dann reicht es womöglich, nur die Art der Datenerhebung zu dokumentieren.

Sind die Daten die Grundlage für Forschungsergebnisse sollten sie auch aufbewahrt werden, um die Nachvollziehbarkeit der Ergebnisse zu gewährleisten.

Schließlich sollte auch immer das Kosten-Nutzen-Verhältnis abgewogen werden, denn eine dauerhafte Archivierung oder Publikation kostet auch Geld.

Und zuletzt auch: Ist eine Nachnutzung durch andere wahrscheinlich? Denn wenn es kein plausibles Szenario gibt, in dem sich jemand anderes für die Daten interessiert, müssen sie auch nicht weiter aufgehoben werden.

Detaillierte Informationen zum Auswahlprozess finden Sie auch auf den Seiten des Digital Curation Centers – kurz DCC. Das DCC ist sowieso eine sehr gute Quelle für alle Belange des Forschungsdatenmanagements. Ich kann Ihnen nur empfehlen, sich durch deren „How to“-Guides einmal durchzuklicken.

Folie 6

Sie haben nun eine Auswahl getroffen, welche Daten Sie langfristig ausbewahren und für die Nachnutzung zur Verfügung stellen wollen. Egal ob für die Archivierung oder die Publikation sollten die Daten den Standards ihres Faches entsprechend und das allerwichtigste: gut dokumentiert sein!

Was nun sollte man über die Daten niederschreiben?

Beschreiben Sie, warum und wie die Daten entstanden sind.

Wie sind die Daten aufgebaut und wie gehören sie zusammen?

Wurden die Daten bearbeitet oder verändert?

Gibt es noch andere Versionen der Daten?

Und auch ganz wichtig: Wie kann auf die Daten zugegriffen werden? Denn es ist nicht immer nötig, dass alle Daten komplett frei im Internet stehen. Es ist durchaus möglich, die Daten nur Personen mit berechtigtem Interesse zugänglich zu machen. In jedem Fall sollten aber die Zugangsmodalitäten transparent bekannt gemacht werden.

Diese Aufzählung kommt übrigens von forschungsdaten.info – eine deutsche Seite, die sich ganz dem Forschungsdatenmanagement verschrieben hat. Auch hier finden Sie alle Informationen rund ums FDM, und sogar zum FDM in verschiedenen Fächern oder Bundesländern.

Folie 7

Aber wenn es um die Aufbereitung von Forschungsdaten für eine Veröffentlichung geht, darf eines nicht fehlen: Die FAIR Prinzipien.

FAIR ist ein Akronym und steht für Findable also Auffindbar, Accesible also Zugänglich, Interoperable also technisch dialogfähig und Re-Usable also nachnutzbar.

Was steckt hinter diesen Begriffen?

Folie 8

Jetzt haben Sie Ihre Daten ausgewählt, aufbereitet und gründlich beschrieben. Wo nun können Sie sie veröffentlichen? Schauen Sie am besten zuerst nach, ob Sie ein geeignetes Fachrepositorium Ihrer Disziplin finden. Dies können Sie am besten über die Plattform re3data.org tun. Hier sind inzwischen mehr als zweieinhalb tausend Repositorien aus aller Welt verzeichnet. Allerdings handelt es sich in der Mehrzahl um Repositorien, in die nur Angehörige bestimmter Institutionen oder Projekte Daten *hochladen* können. Beim Durchsehen der Ergebnisse können Sie anhand von kleinen

Symbolen schnell einschätzen, welche Anforderungen ein Repository erfüllt. Auf der re3data-Webseite erhalten Sie nähere Informationen, wenn Sie mit der Maus über die Icons fahren. Am einfachsten ist es aber, wenn Sie direkt die Filtereinstellungen Ihren Wünschen entsprechend setzen.

Um selbst Daten zu veröffentlichen, achten Sie vor allem darauf, dass das Repository Ihrer Wahl sowohl einen Persistent Identifier – im Idealfall einen DOI - vergibt, als auch die Datennutzung eindeutig und rechtlich verbindlich regelt. Außerdem sollten Sie folgende Punkte bedenken:

- Wer darf überhaupt Daten *einstellen*?
- In welchen Formaten werden Daten und Metadaten akzeptiert? Wenn Sie sich schon vor Beginn der Datenerhebung nach geeigneten Repositorien umsehen, können Sie Ihre Daten und Metadaten direkt im richtigen Format erheben und ersparen sich so ein späteres Konvertieren.
- Welche Services bietet das Repository? Das geht nämlich oft über das reine Speichern und Veröffentlichen hinaus und kann zum Beispiel Dienstleistungen wie Qualitätsprüfungen oder das Konvertieren in aktuelle Dateiformate beinhalten.
- Was kostet das Einstellen von Daten? Das ist sehr unterschiedlich und hängt nicht nur von Speicherplatz und Speicherzeit, sondern gegebenenfalls auch von weiteren Dienstleistungen ab. Denken Sie daran, dass Sie solche Kosten bei vielen Drittmittelgebern mit beantragen können.

Finden Sie kein geeignetes fachspezifisches Repository schauen Sie bei Ihrer Uni oder Institution nach. Diese bieten oftmals sogar kostenfreie Repositorien für Ihre Mitarbeiter an.

Folie 9

Wenn Sie dann Ihre Daten veröffentlicht haben, verlinken Sie sie unbedingt mit dem dazugehörigen Artikel. Denn die beste Beschreibung ist immer noch ein ausführlicher wissenschaftlicher Artikel. Und so kann Ihre Arbeit auch als Ganzes viel besser zitiert werden.

Folie 10

Wenn Sie Daten veröffentlichen sollten Sie auch immer eine Lizenz mit angeben. Denn auch wenn Daten oft gar nicht die Schöpfungshöhe erreichen, um unter das Urheberrecht zu fallen, helfen Lizenzen, Unklarheiten zu beseitigen. So müssen Nachnutzer nicht Zeit damit verschwenden, zu recherchieren, ob und wie sie die Daten nun verwenden dürfen.

Eben weil Daten oftmals nicht unter das Urheberrecht fallen und weil es die Nachnutzung einfach am meisten erleichtert, wird oft die Verwendung der CC0-Lizenz empfohlen. Dies ermöglicht eine Nachnutzung ohne jegliche Restriktionen. Wie wichtig das ist, begreifen Sie, wenn Sie sich vorstellen, dass mitunter auch viele verschiedene Datensätze neu zusammengefügt werden. Dann ist es fast unmöglich, alle möglichen unterschiedlichen Lizenzen zu berücksichtigen.

Und bedenken Sie, dass die Nennung des Urhebers sowieso der guten wissenschaftlichen Praxis entspricht und allein schon durch die wissenschaftliche Gemeinschaft geregelt wird.

Die am meisten genutzten Lizenzen sind sicher die Creative Commons Lizenzen. Diese gibt es in unterschiedlichen Abstufungen – von CC0, d.h. keine Restriktionen, über CC-BY, d.h. die Urheber müssen genannt werden, bis zu starken Restriktionen wie non-commercial, also kommerzielle Nutzen ist verboten.

Folie 11

Wie gesagt sind Creative Commons Lizenzen die bekanntesten – der Vorteil ist, dass sie für alle möglichen Arten von Kreationen genutzt werden können: Texte, Bilder, Musik – alles Mögliche.

Doch manchmal ist es auch gut, speziell auf den eigenen Output zugeschnittene Lizenzen zu haben. Für Daten sind dies z.B. die Open Data Commons Lizenzen.

Die sind aus den Creative Commons Lizenzen entstanden, gehen jedoch stärker auf die Bedürfnisse von Datenerstellern und vor allem Datenbanken ein. Denn das Teilen von Datenbanken ist kompliziert, wenn man bedenkt, dass diese nicht nur als Ganzes weitergenutzt werden, sondern oft nur teilweise. Während diese Lizenzen auf den ersten Blick alle gleich aussehen, unterscheiden sie sich doch sehr in den Details des Lizenztextes.

Der Vollständigkeit halber erwähne ich hier auch noch die Datenlizenz Deutschland, die speziell für den deutschen Rechtsraum geschaffen wurde, aber eigentlich keinen wirklichen Mehrwert gegenüber den anderen Lizenzen hat.

Und weil bei der Diskussion um Forschungsdaten auch ganz oft Software mit einbezogen wird, hier auch noch der Hinweis auf choosealicense.com – eine sehr hilfreiche Webseite zum Aussuchen einer Lizenz für Software.

Folie 12

Jetzt denken Sie wahrscheinlich: Schön und gut das Ganze, aber wieso sollte ich mir die ganze Arbeit eigentlich machen?

Nun zum einen wäre da die „Peitsche“: die Vorgaben Ihrer Geldgeber und Ihrer Community. Es wird immer mehr gefordert und auch zum Standard, Daten zu veröffentlichen.

Aber es gibt auch das Zuckerbrot: Denn für jede Datenveröffentlichung können Sie auch Zitationen bekommen. Wenn Ihr Datensatz einen DOI über DataCite bekommt (und das ist bei vielen Repositorien Standard) können Zitierungen automatisch über die DataCite-Suche nachverfolgt werden. Und Veröffentlichungen und Zitierungen bringen Sie in Ihrer wissenschaftlichen Karriere voran. (Sofern diese qualitativ hochwertig sind.) Zudem könnten über Ihre Daten auch Kollegen auch aus anderen Fachgebieten auf Sie aufmerksam werden und womöglich ergeben sich so neue Kollaborationen.

Folie 13

Und ganz abgesehen von Ihren persönlichen Vorteilen, fördert die Publikation von Daten auch die Allgemeinheit.

Die allgemeine Öffentlichkeit profitiert von einer transparenten Wissenschaft, indem das Vertrauen in diese gefördert wird.

Und die wissenschaftliche Gemeinschaft kann die Daten nachnutzen, Erkenntnisse verifizieren und neue Forschung darauf aufbauen.

Sie sehen also: Der geringfügige Mehraufwand für die Publikation von Daten kann einen ganzen Schwanz an Vorteilen für Sie, die Allgemeinheit und Ihre Kollegen nach sich ziehen.