Notizen zu den Spotlights

Zusammenstellung von Thomas Arndt (FZJ)

Review für Forschungsdaten: Wie sollte Qualitätssicherung von Datenpublikationen aussehen?

Das Thema wurde von den Organisatoren der Konferenz vorgeschlagen.
Moderation und Notiz: Oliver Bertuch (FZJ); Teilnehmer: ca. 25
Im Weiteren wird der Gesprächsfaden der Session dargestellt. Durch die offene Diskussion werden unmittelbar verbundene Themen gestreift. Diese Gebiete sind durch Überschriften kenntlich gemacht.

Eröffnung

- Wie sollte Qualitätssicherung von Datenpublikationen aussehen?
- Wer darf entscheiden, ob Daten veröffentlicht werden?
- Welche Entscheidungsverfahren für Textpublikationen sind geeignet?

Die TU Dresden arbeitet bei ihrem institutionellen Repositorium mit internem Reviewer. Ein Workflow ist etabliert mit Benachrichtigungssystem und Vergabe von Zugriffsrechten auf die Daten für den Reviewer.

Nach dem Review erfolgt ein Kurationsprozess, in den der Datenkurator involviert ist. Der Prozess ist "verpflichtend" (verankert durch die Datenpolicy) und wird im Repositorium technisch erzwungen. Derzeit sind 30 Datensätze so publiziert, bislang ohne negative Rückmeldungen.

Aufgrund der geringen Stichprobengröße ist eine Aussage über die Qualität der Daten und Metadaten (noch?) nicht möglich. Kommentare von Autor, Reviewer, Kurator etc. werden aufgezeichnet. Sie sind nur intern für die direkt Beteiligten sichtbar.

Probleme für Reviews

Idealerweise sollte ein Review analog zu Textpublikationen im Doppelblind-Verfahren ablaufen, um jegliche Beeinflussung auszuschließen.

Forschungsdaten können oft aufgrund ihrer Größe, ggf. Rechten und Ähnlichem nur schwer anonym zugänglich gemacht werden.

Anhand der URL des Repositoriums, in dem die Daten liegen, sind Autoren mit wenig Rechercheaufwand zu identifizieren. Es braucht bessere Werkzeuge für diese Use-Cases.

Kriterien zur Beurteilung
Daraus ergibt sich eine elementare Fragestellung:

- Was sind bessere Kriterien zur Beurteilung der Qualität eines Datensatzes während eines Reviews, statt subjektives Empfinden?

Die TU Dresden orientiert sich hier an Archivierungskriterien: "Ist der Datensatz nachnutzbar?"

Einige Orientierungshilfen wurden genannt:

- Identifizierung von archiv-würdigen Daten mit einigen Fragen: DCC (2014). Five steps to decide what data to keep: a checklist for appraising research data v.1 (http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep).
- Edinburgh: Digital Curation Centre. UK Data Archive hat "Quality Levels" als Kriterien veröffentlicht. (https://www.dataarchive.ac.uk/sharing-best-practice/digital-curation-and-data-publishing/quality-control/)

Quintessenz: Meist wird erst über Qualität auf inhaltlicher/fachlicher Ebene gesprochen, danach über technische Aspekte (wie z. B. Datenformate).

Stellenwert von Reviews

- Wie findet man Reviewer und wieso sollte das jemand tun?
- Gibt es Überlegungen zu Reputationssystemen oder Anerkennung von Gutachtertätigkeit als wissenschaftliche Leistung?

Einhellige Antwort: Es sind viele Absichtserklärungen im Raum. Die Details sind noch unklar.

Erste Lösungsansätze liefern:

- publons (https://publons.com/) dient als Plattform für das Finden von Reviewerns sowie der Durchführung und Dokumentation der Reviews. Motivation ergibt sich aus der Anerkennung für durchgeführte Reviews, die in Profilen öffentlich gelistet werden. Dieses kann für Bewerbungen z. B. in Lebensläufen referenziert werden.
- Die Open Science Bewegung kennt die "Open Peer Reviews", die von verschiedenen Zeitschriften praktiziert werden. Siehe dazu: AG OpenScience (https://ag-openscience.de/openpeer-review/) und Wikipedia (https://de.wikipedia.org/wiki/Offenes_Peer-Review)
- Einige Fachdisziplinen und ihre Publikationskanäle arbeiten mit der Vergabe von Badges für Reproduzierbarkeit.
    - Beispiel im Bereich Informatik: ACM (https://www.acm.org/publications/policies/artifact-review-badging)
    - Ein Kommentar erwähnte, dass dies auch in der Fachdisziplin "Psychologie" genutzt würde. Siehe dazu: Open Science Badges (https://cos.io/our-services/open-sciencebadges/)

Lessons learned: Peer Review bei Textpublikationen ist kein perfektes Kriterium für Qualität, keine 100% Sicherung für Qualität und ggf. manipulier- und ausnutzbar. Für Forschungsdaten sollte man es daher nicht überbewerten. Qualität kann auch selbst in eigenen Richtlinien und Communities definiert werden.

Kuration vs. Selbstregulierung
Zenodo, figshare und ähnliche Repositorien sind zweischneidige Schwerter. Einerseits sind Wissenschaftler froh, dass das Einstellen von Daten so leicht ist. Andererseits mehren sich die Stimmen, dass die Nachnutzung sehr schwierig ist: Schrott und Gold voneinander zu trennen fällt dort schwer.

Eine Frage wird in die Runde geworfen: Ist die Nachnutzung von Datensätzen ein Indikator für Qualität? Die These: Würde Nachnutzung eines Datensatzes honoriert und belohnt, würde der Anreiz zu besserer Qualität stärker. Mit "einfach Hochladen" wäre es nicht mehr getan, denn die Chance der Auffindbarkeit und Nachnutzung sinken.

Nachnutzen von Daten ist leichter in fachspezifischen Repositorien. Dort kann mit fachspezifischen Metadaten gearbeitet werden und die hochgeladenen Daten unterliegen einer stärkeren Selbstregulierung durch die Community. Durch die fachliche Nähe fällt Datenmüll schlicht schneller auf. Ebenso hilft, dass diese Repositorien oftmals nur Schreibzugriff für hinlänglich bekannte WissenschaftlerInnen erlauben, sodass der eigene Ruf in Gefahr ist. Ein explizites Peer Review oder enge Kuration könnten sich hier erübrigen.

Nichtsdestotrotz hilft es insbesondere den Unerfahrenen, wenn Daten innerhalb der Arbeitsgruppe reviewt werden (z. B. Co-Autoren reviewen die Daten des Hauptautors, etc.). Eventuell könnte eine Trennung zwischen der Sicherung inhaltlicher Qualität durch die Arbeitsgruppe und technischer Qualität z. B. durch Kuratoren helfen, um "näher" am Forschungsprozess zu sein.

Kultur und Ängste
Die Runde mahnt zur Vorsicht vor Qualität als Selbstzweck, heiligem Gral oder gar Cargo-Kult. Es könnte WissenschaftlerInnen insbesondere in der Anfangszeit des FDM verschrecken, ihre Daten zu veröffentlichen. Dies wird größtenteils als kulturelles Problem wahrgenommen: einerseits schlechte Fehlerkultur, in der Angst vor Angreifbarkeit herrscht und andererseits vor Konkurrenz: Es könnte ja in den Daten noch der Nobelpreis versteckt sein.

Verfechter des Open-Science-Gedankens sind zutraulicher was die Veröffentlichung ihrer Daten betrifft. Die zunehmenden Diskussionen über und um Open Science färbt auch auf geschlossenere Gruppen ab. Erste Veränderungen, z. B. in der Juristen-Fachcommunity, wurden von einigen Teilnehmern der Session beobachtet.

Druck zur Veröffentlichung baut sich auch aus der Ecke der Forschungsinformationssysteme (FIS) und das daran angeschlossene Wissenschafts-Reporting auf. Anreize sind hier entsprechende positive Honorierung bis hin zu negativen Konsequenzen: "Was nicht im FIS ist, existiert nicht."

Allgemein ist die Veröffentlichung der Daten leichter, wenn Leidensdruck vorhanden ist. Ein Beispiel dazu ist der Datenaustausch in Kollaborationen. Um das gegenseitige Verständnis der Daten zu sichern, wird besser dokumentiert, sodass die Qualität ansteigt. Die Veröffentlichung fällt hier oft leicht, da bereits Konsens herrscht und die oben beschriebenen Ängste weitgehend abgebaut sind.

Insgesamt herrscht laut Beobachtungen große Angst vor Entblößung unter den Wissenschaftlern. Die reine Existenz der Daten zu dokumentieren, ist unproblematisch und argumentativ schwer auszuhebeln, jedoch leichte Zugreifbarkeit behagt den meisten nicht.

Reviews: Freiwillige?
Eventuell hilft es Ängste abzubauen, in Arbeitsgruppen und Institutionen "Critical Friends" zu etablieren (letztlich stellt dies ein internes Review dar).

Doch wer könnte dies tun, neben seiner eigentlichen Arbeit? Beispiel: Während ein Doktorand an seiner Dissertation schreibt, wird der Betreuer meist maximal die Text-Publikation Korrektur lesen. Ein Review der Daten fällt wohl meist aus, weil zu aufwändig und zeitintensiv.

Ein Review auf gleicher Ebene (Doktoranden helfen einander) oder auch Vier-Augen-Prinzip könnte Abhilfe schaffen, aber es braucht dafür Freiraum (Arbeitszeit). Lernen ließe sich das Reviewen in den Doktorandenprogrammen.

Kritisch zu sehen ist aber, dass unter dem großen Druck von Publikationen, etc. die Doktoranden nicht diejenigen sein können, die den großen Strukturwandel bringen sollen. Das schwächste Mitglied sollte nicht die institutionellen Probleme lösen müssen...

Besser wäre möglicherweise, bereits bei den Forschungsförderern anzusetzen und dies einzufordern. Daraus ergibt sich: Ist dazu eine stichprobenhafte Überprüfung der Qualität seitens der Förderer notwendig?

Eine weitere Idee wird in diesem Zusammenhang eingebracht: Ebenso könnten externe Peer Reviews der Dokumentation der Daten dienen, was wiederum gute Grundlage für eine Veröffentlichung in einem Datenjournal wäre.

Dies führt zurück auf einen Wert für sich: Ist offene Verfügbarkeit an sich bereits ein Qualitätskriterium? Immerhin kann es überhaupt genutzt werden, was ein explizites (formelles Verfahren) oder implizites (Versuch der Nachnutzung) Review ermöglicht. Und nur durch diese Offenheit ist exakt reproduzierbare Wissenschaft zu erreichen.

Technische Validierung und Reproduzierbarkeit

- Kann eine technische Validierung zwischen Dokumentation und Datensatz stattfinden?

Dazu wäre eine tiefe semantische Beschreibung der Daten notwendig, damit Maschinen dies verarbeiten können. Eventuell ist hier "Machine Learning" eine vielversprechende Hilfe, ohne genauer auszuführen, wie dies funktionieren könnte. Einige Teilnehmer berichten, dass bereits heute darüber Mindeststandards geprüft werden können. Beispiele dazu sind Lizenzen, aber auch Wertebereiche, Kodierungen etc. Der technische und personelle Aufwand für die Beschreibung und die nötigen Computing Ressourcen ist hoch, die Gefahr von Cargo Cult steigt.

Forschungssoftware als Forschungsdaten
Nichtdestotrotz: "reproducible science" wäre ein starkes Hilfsmittel und ein im Prinzip erstrebenswerter Zustand. Dem steht entgegen: a) der hohe technische Aufwand und b) der dazu notwendige Aufbau von Kompetenzen bei den WissenschaftlerInnen. Hier bietet die "Research Software Engineering" (RSE) Bewegung großes Potential. Die Diskussionsrunde sieht jedoch auch hier neue Herausforderungen: Eigentlich müsste auch die Forschungssoftware reviewed werden, was ungleich komplexer ist. Und wo zieht man die Grenze zwischen dem Review der Daten und der Software - müsste nicht beides gleichzeitig geschehen, auch wenn dies die Aufgabe noch schwieriger macht?

Es ist im Falle von Publikationen der Ergebnisse (Text) zwingend notwendig für ein Review der Daten, nicht nur Dokumentation und Datensatz abzugeben, sondern auch die Software zur Analyse zu haben. Damit wird Software im Prinzip auch zu Forschungsdaten. Das ist natürlich eine noch weitergehende Forderung in Richtung der Wissenschaftler (Herausgabe; Reproduzierbarkeit von Anfang an) und auch an die Reviewer (nicht nur Daten verstehen, sondern auch die Software).

Forschungsdaten in der Blockchain: Hype oder die neue Art der Datenpublikation?

Das Thema wurde von den Organisatoren der Konferenz vorgeschlagen.
Moderation und Notiz: Claudia Frick (FZJ)

- Die Wichtigkeit des Themas ist bekannt, es fehlt jedoch das Grundwissen.
- Insgesamt gibt es einen Mangel an grundlegenden Kompetenzen dazu, um überhaupt abschätzen zu können, ob die Technologie (für die eigene Einrichtung) relevant ist und wofür.
- Es wird Bedarf gesehen zu Einführungen in das Thema für Bibliothekarinnen (Konzept und Technik/Praxis). Hands-On-Workshops könnten das leisten und begleitet von einer Informationsplattform die Wissenslücken schließen. Beispiel: Blockchain Experience Lab der RWTH Aachen
- Fragen, z.B. nach der Begrenzung der Datenmengen in der Blockchain und den Metadaten von Daten in der Blockchain, mussten offen bleiben.
- Ob die Blockchain ein Teil der FDM-Utopie ist oder nur ein Hype, darüber gab es ebenfalls keine Einigkeit.

Sinn und Unsinn des Verschiebens von Terabytes: Wo speichert man Rohdaten?

Das Thema wurde von den Organisatoren der Konferenz vorgeschlagen.
Moderation und Notiz: Torsten Bronger (FZJ)

Motivation
Große Datenmengen bewegen ist eine aufwendige und teure Angelegenheit. Während es im Falle von Terabytes über das Internet bloß lästig ist und Geduld erfordert, kann es im Falle von Petabytes bereits innerhalb einer Institution unpraktikabel sein, die Daten überhaupt zu bewegen. Kuriose Lösungen wie das Verschicken von Festplatten können dieses Problem nicht lösen.

Deshalb sollte das Ziel sein, den Datentransfer auf das Nötige zu beschränken. Leider sorgen gerade Repositorien dafür, dass Daten bei ihnen hochgeladen werden, obwohl die Wahrscheinlichkeit dafür, dass sie nachgenutzt werden, für den einzelnen Datensatz sehr klein ist. Dies gilt in besonderem Maße für Long-Tail-Daten. Gerade diese Daten können sehr groß sein.

Andererseits bieten Repositorien eine besonders zuverlässige Art der Datenarchivierung, die die Institute und Forschergruppen, die die Daten dort hochladen, in-house auf diesem Niveau nicht bieten können.

In diesem Spannungsfeld ist zu diskutieren, in welchen Szenarien ein Datentransfer ohne direkte Nutzung am Ziel sinnvoll ist, und wann er vermeidbar ist.

Umfang
Erster Diskussionspunkt war der Umfang archivierter und publizierter Daten. Letztlich muss der Wissenschaftler entscheiden, was aufbewahrt werden soll. Das muss ausdrücklich nicht alles sein. Publizierte Daten sind heute meistens die Daten, die unmittelbar für die Aussagen und Abbildungen der Textpublikation benutzt wurden. Das sind u.U. nur wenige Megabytes. Im Sinne der Reproduzierbarkeit ist das jedoch zu wenig, es sind stattdessen auch rohere Daten einzubeziehen. Außerdem kann und sollte die Datenpublikation über ihre verknüpfte Textpublikation im Scope hinausgehen, Stichwort „Supplemental Material“. Gleichzeitig bedeutet allerdings eine umfängliche Datenbereitstellung wieder einen u.U. sehr großen Datentransfer.

Status bei den Diskussionsteilnehmern
Keiner der Diskussionsteilnehmer kam von einer Institution, bei der bereits ein Datenrepositorium im Produktivbetrieb ist. Im Gegenteil gab es an mindestens zwei Institutionen die Maßgabe an die Forscher, externe Repositorien zu nutzen, vorzugsweise disziplinäre, und wenn keine solchen vorhanden, generische wie Zenodo.

Außerdem wurde festgestellt, dass der Umfang der Forschungsdaten extrem stark vom Institut abhängt. Es ist nicht ungewöhnlich, dass wenige Prozent der Arbeitsgruppen einer Institution den Großteil des Datenaufkommens dieser Institution stellen. Umgekehrt kann man dann mit verhältnismäßig kleinem Ressourcenaufwand die Bedarfe des Großteils der Forscher einer Institution adressieren.

Registry statt Repositorium
Mindestens im Forschungszentrum Jülich und im DLR besteht der Plan, statt eines institutionellen Repositoriums eine Metadaten-Registry zu installieren, deren Datensätze auf die Rohdaten zeigen statt eine Kopie von ihnen zu enthalten. (Optional können Forscher ihre Daten dennoch hochladen, dies ist aber nicht die bevorzugte Methode.) Dadurch werden Datentransfers in den allermeisten Fällen vermieden. Mittels Datendurchleitung kann man das für Nutzer transparent machen, d.h. Rohdaten können von der Registry heruntergeladen werden als wären sie dort lokal gespeichert gewesen. Hinter den Kulissen werden die Daten von der Instituts-Lokation durchgeleitet.

Dieses Modell wurde positiv in der Runde aufgenommen, aber es ergeben sich auch Fragen. Die größte Herausforderung ist, dass eine zuverlässige langfristige Datenhaltung von den Instituten geleistet werden muss, die darin nicht ihre Kernkompetenz haben. Insbesondere sind sie nicht gewöhnt, Daten dauerhaft an derselben Lokation zu halten. Aber sonst würden die Rohdaten aus Sicht der Registry (und aller an die Registry angeschlossenen Systeme und Nutzer) verloren gehen.

Als Alternative wird vorgeschlagen, PIDs nicht nur für die Landing Page in der Registry zu benutzen, sondern auch für den Link von der Registry zu den Rohdaten im Institut. (Das müssen keine DOIs sein, sondern können z.B. ePICs sein.) Dies würde technisch einen Umzug der Daten möglich machen, und organisatorisch bei den Instituten das Bewusstsein dafür schaffen, dass dieser Punkt überhaupt wichtig ist, und das Institut dafür verantwortlich ist.

Ein weiterer offener Punkt bei diesem Modell ist das Rechtemanagement. Im Falle eines klassischen Repositoriums ist das vergleichsweise einfach: Die Rohdaten werden lokal vorgehalten und sind mit gewissen Rechten versehen, die im Repositorium gesetzt werden. Werden die Daten nicht lokal vorgehalten, sollen aber durchgeleitet werden, muß dauerhaft (also mindestens 10 Jahre) Lesezugriff der Registry auf alle Instituts-Storages bestehen, die angeschlossen sind.

Man kann die Zuverlässigkeit einer solchen Methode stärken und die Probleme beim Rechtemanagement abmildern, wenn man entweder sehr großen eigenen Storage an die Registry anschließt, oder den Instituten empfiehlt, die Daten in das lokale Rechenzentrum oder IT-Departement zu spiegeln und dann dort auf die Rohdaten zugreift.

Zentrale vs. dezentrale Datenhaltung
Das Problem mit großen Datentransfers wirft auch die Frage auf, inwieweit es überhaupt sinnvoll ist, dass Daten in den Instituten einer Institution lokal gespeichert werden. Es wäre nicht ungewöhnlich, dass eine große Storage-Farm, die institutionsweit angeboten wird, pro GB billiger ist als lokaler Storage in einem Institut. Allein schon Personenstellen, die dadurch eingespart werden können, können dafür sorgen. Andererseits können sehr große Datenmengen nicht sinnvoll selbst durch interne Netze geleitet werden, und die zentrale Datenhaltung eignet sich nicht in Fällen, in denen niedrige Latenzen gebraucht werden.

In Jülich gibt es ein laufendes – zur Zeit schlafendes – Projekt mit dem Ziel, einen Schlüssel zu liefern (Excel-Sheet), mit dem Institute bestimmen können, ob für ihr Szenario sich besser die zentrale oder die dezentrale Datenhaltung eignet.

Probleme der zentralen Datenhaltung sind aber auch auf anderen Ebenen:

- es gibt unter den Institutionsmitgliedern schlicht Unkenntnis der zentralen Angebote, insbesondere bei Neulingen in der Institution
- die Kosten können im Einzelfall höher sein als für die lokale Lösung
- die lokale Lösung wird von den Forschenden als bequemer empfunden
- zentrale Angebote sind manchmal technikverliebt, überkompliziert und am Kunden vorbeigedacht.

Neben allgemeinen Awareness-Maßnahmen wurden folgende Lösungsansätze diskutiert:

- Nutzung zentraler Dienste (z.B. für die Archivierung von Forschungsdaten) in die Promotionsordnung tragen
- Nutzung zentraler Dienste (z.B. das Datenbackup) in die jeweilige Dienstordnung tragen.

FD-Strategieentwicklung mit dem RISE-DE Framework

Das Thema wurde von Boris Jacob von der UB Potsdam vorgeschlagen und von ihm selbst moderiert.
Notizen von Irene Barbers (Forschungszentrum Jülich) und Boris Jacob (Universität Potsdam), weitere Bearbeitung: Niklas K. Hartmann (Universität Potsdam)

Darstellung des RISE-DE Frameworks
Das RISE-DE Framework ist im Projekt FDMentor entstanden und online verfügbar unter: http://doi.org/cz23

Nach der Präsentation des RISE-DE Frameworks am ersten Konferenztag wurde am zweiten Tag eine Spotlight-Session zu dem Thema veranstaltet, um einen Rahmen zum direkteren Erfahrungsaustausch und die Möglichkeit zur detaillierteren Betrachtung von FDM-Strategieprozessen zu geben.

Begonnen wurde mit einer Vorstellungsrunde der Teilnehmenden, welche aus Hochschulen, außeruniversitärer Forschung und aus dem wissenschaftspolitischen Bereich kamen, u.a. von der Universität Graz, Freie Universität Berlin (FDMentor-Projektpartner), FH Aachen, TIB Hannover, Universität Vechta, RWTH Aachen, Deutsches Institut für Entwicklungspolitik Bonn und European University Alliance. Die Situation in den jeweiligen Institutionen beim FDM ist unterschiedlich, einige stehen noch am Anfang und wollen Dienste strategisch aufbauen, andere wollen bereits existierende Dienste evaluieren und das Servicespektrum erweitern.

Strategieentwicklung am Beispiel UB Potsdam
Anschließend wurde vom Strategieprozess an der Universität Potsdam berichtet, der vor einem Jahr mit dem Ziel FDM-Dienste der Zentraleinrichtungen von Grund auf aufzubauen gestartet wurde. Grundlage war das RISE-Framework (Research Infrastructure Self-Evaluation) des Digital Curation Centres. Das FDM-Team von Universitätsbibliothek (UB) und Zentrum für Informationstechnologie und Medienmanagement (ZIM) hat es übersetzt, an den deutschen Wissenschaftskontext angepasst und in einem Multi-Stakeholder-Prozess mit einer von der Forschungskommission des Senats eingesetzten Arbeitsgruppe verwendet. Die AG Forschungsdaten besteht aus Vertreter*innen aus Forschung, Universitätsleitung und Verwaltung und traf sich im Verlauf eines Jahres zu neun Treffen in der die Themen von RISE-DE diskutiert, teilweise angepasst und schließlich Bewertungen zum Ist- und Soll-Zustand vorgenommen wurden. Außerdem wurden immer wieder Gäste aus den Bereichen zu einzelnen Themen hinzu geladen. Insgesamt waren etwa 30 Personen beteiligt. Aus den dokumentierten Ergebnissen wurde ein Strategiepapier zur Weiterentwicklung der Dienste der Zentralen Einrichtungen mit Roadmap und Governance entwickelt, eine Forschungsdaten-Policy sowie diese ergänzende Handlungsempfehlungen. Alle drei Dokumente befinden sich nun auf dem Weg durch die Gremien.

Fragen und Antworten zum Framework
Die folgende Diskussion wurde über verschiedene Aspekte der Strategieentwicklung im FDM geführt und wird hier in Form eines Frage-Antwort-Schemas dargestellt:

Frage: Also gab es bei euch eine „grüne Wiese“ die ihr bepflanzen konntet, das hört sich alles so einfach an?
Antwort: Für den Strategieprozess hatten wir sowohl Unterstützung von der Universitätsleitung, als auch eine durch das BMBF im Projekt FDMentor finanzierte Stelle, von daher waren die Voraussetzungen gut. Legitimation nach innen ist für die institutionelle Strategieentwicklung entscheidend. Stakeholder (auch bspw. SFBs) von Anfang an einbeziehen, dadurch Lernen aus den Bedürfnissen anderer und Bildung von Multiplikatoren. Man muss aber nicht gleich den Multi-Stakeholder-Prozess starten, sondern kann RISE-DE auch alleine ausfüllen bzw. gezielt einzelne Stakeholder mit einbinden und das als Diskussionsgrundlage und Argument für einen Strategieprozess nehmen. Auch macht eine „grüne Wiese“ nicht alles einfacher: Wenn nicht so viel vorhanden ist, gibt es natürlich weniger „historisch gewachsene“ idiosynkratrische Strukturen, aber eben auch weniger praktische Erfahrung und Dinge, auf die aufgebaut werden kann.

F: Wie sollten solche FDM-AGs idealerweise besetzt werden?
A: Am besten Schöpfung aus möglichst schon vorher bestandener Vernetzung, Besetzung einerseits politisch, andererseits kompetenzorientiert. Der FDM-Prozess sollte wissenschaftsgetrieben sein bzw. nach den Bedürfnissen der Wissenschaft vorangetrieben werden. Am besten sitzt am Tisch Expertise zu allen wichtigen Arten von Daten bzw. Arten von datengetriebenen Methoden, die an der Einrichtung regelmäßig eingesetzt werden. Es ist auch wichtig, dass die Beteiligten zwar eine positive Grundhaltung mitbringen (keine „Verhinderer“), aber nicht nur „Datenchampions“ teilnehmen. Das Vorwissen der beteiligten Wissenschaftler*innen ist daher unterschiedlich. Alle sollten auf ein Level gebracht werden.

F: Müssen bei einer Weiterentwicklung alle Stufen eines Themas durchlaufen werden?
A: Prinzipiell sind die Stufenbeschreibungen so formuliert dass sie aufeinander aufbauen. Es kann aber auch eine bewusste Entscheidung sein, davon abzuweichen.

F: Bei einer Bewertungsskala von 0 bis 3, wird dann nicht immer der höchste Wert gewählt?
A: Bei der Bewertung in den Stufen geht es nicht darum generell möglichst hohe Werte zu wählen, sondern jeweils die, welche für die jeweilige Einrichtung am sinnvollsten zu erreichen scheint. Die im Framework vorgegebenen Kategorisierungen zur Einordnung der eigenen Leistungen sind nicht als Bewertung oder Kennzahlen zu verstehen, sondern eher als Einordnung der Möglichkeiten. Unter Umständen ist eine niedrige Einordnung in einzelnen Bereichen sogar gewünscht. Aus der Kategorisierung sollte eine eigene Roadmap entwickelt werden

F: Sollte man das Tool nutzen, wenn es mit der FDM-Planung eigentlich erst in einem Jahr losgeht?
A: Unbedingt, je früher man einen Überblick über den Ist-Zustand hat, desto fundierter kann man Strategieprozess beginnen.

F: Ist es Sinnvoll RISE-DE zu nutzen auch wenn man weiß, dass momentan keine Mittel zum FDM-Ausbau zur Verfügung stehen??
A: Ich halte es für sinnvoll um das mögliche Spektrum aller Dienstleistungen an der Einrichtung zu bewerten. In der Selbstbewertung werden dann etliche Themen auf niedriger Stufe bewertet werden, dass Ergebnis kann dann aber zur internen Diskussion über die Zielvorstellung genutzt werden. Vielleicht wird dann am Ende entschieden, die vorhandenen Ressourcen nur in ein Kommunikationskonzept zu stecken und für andere Bedarfe Dienste von Dritten zu benutzen.

F: Wie geht es weiter, wenn man alle RISE-DE Themen bearbeitet hat?
A: Aus den dokumentierten Ergebnisse des kann man Roadmap, Policy und Strategie entwickeln. Für die Erstellung der Policy haben wir das Policy-Kit der Technischen Universität Berlin verwendet, welches sie im Rahmen des Projekts FDMentor entwickelt haben (Online verfügbar unter: http://dx.doi.org/10.14279/depositonce-8412).

F: Soll besser das englische RISE oder das deutsche RISE-DE benutzt werden?
A: Je nach Bedarf und Ausrichtung des Prozesses. RISE ist weniger konkret und hat dadurch weniger Text (pro Themenfeld eine Seite). RISE-DE ist umfangreicher und stärker an den deutschen Wissenschaftskontext angepasst.

F: Wird es ein Online-Tool für RISE-DE geben, so wie es für die englische Version von RISE durch SPARC Europe erstellt wurde, unter https://sparceurope.org/evaluate-your-rdm-offering/?
A: Das ist momentan nicht geplant, für den Herbst ist erstmal ein Update von RISE-DE geplant.

F: Wie hängen das RISE-DE Framework und das DIAMANT-Modell aus dem PODMAN-Projekt miteinander zusammen (Online verfügbar unter: https://fdm.uni-trier.de/)?
A: Sie sind komplementär, RISE-DE orientiert sich an FD-Diensten/Portfolio, ihren Voraussetzungen und „Qualitätsstufen“, beinhaltet aber kein detailliertes Modell aller inhaltlichen Elemente der Dienste oder benötigten Kompetenzen. DIAMANT dagegen orientiert sich an einem forschungszentriertem Modell des FD-Lebenszyklus und bildet detailliert die relevanten Elemente eines institutionellen FDM ab, inkl. Kompetenzmatrix für Forschende und Service-Personal.
Wir wollen uns das aber von beiden Seiten aus noch einmal genauer anschauen.

F: Sind RISE-DE Anwendertreffen geplant?
A: Dezidierte Anwendertreffen sind bisher nicht geplant, wir geben aber gerne Training-Sessions und Workshops, wenn sich die Gelegenheit ergibt. Wir versuchen auch die Nutzung von RISE-DE zu erfassen und stehen darüber im Erfahrungsaustausch mit deutschsprachigen Anwender*innen und auch mit internationalen über die Nutzung der englischsprachigen Version RISE v0.9. Sie können uns bei Fragen zur Anwendung beider Frameworks gerne kontaktieren unter: forschungsdaten@uni-potsdam.de

DMP/RDMO meets FDM Servicekatalog

Zusammenfassung der Spotlight – Diskussion im Rahmen der WissKom2019,
Das Thema wurde von Teilnehmern vorgeschlagen.
Ca. 15 Teilnehmer, etwa die Hälfte davon kennt das RDMO-Tool bereit aus eigener Erfahrung.
Moderation und Notiz: Eva-Maria Gerstner, 6. Juni 2019

- Nutzerfeedback: Der generische Fragenkatalog des RDMO-Tools ist viel zu umfangreich, daher wurde der kurze Fragenkatalog für RDMO entwickelt. Generell hat sich ein generischer Fragenkatalog als nicht so hilfreich erwiesen.
- Der Fragenkatalog ist für Studierende oft nicht eindeutig genug (fehlendes Hintergrundwissen).
- Vision: Weiterentwicklung des RDMO-Tools. Den Forschenden soll eine Hilfestellung in Form fachspezifischer FDM-Dienste geboten werden, die in den DMP-Fragenkatalog integriert werden und zwar in Form von Dropdown-Menüs (Tendenz: weg von Freitextfeldern); es können hier auch externe Services anderer Einrichtungen angeboten werden.
- Transparenter machen, was z.B. andere Fachcommunitys (Beispiel: GFBio, GESIS) anbieten und im DMP-Tool anzeigen.
  Beispielfrage: „Wo wollen Sie Ihre Daten publizieren?“ Wenn bereits „Sozialwissenschaften“ für das Auswahlfeld „Fachbereich“ ausgewählt wurde, könnte dann als Vorschlag „GESIS“ angeboten werden.
- In der „idealen Welt“ hat eine Institution einen großen Servicekatalog, der sich über entsprechend aufzubauende Filterkategorien sukzessiv einschränken lässt.
- Hat eine Einrichtung das Repositorium RADAR nicht vertraglich lizensiert, taucht es in der Auswahlliste gar nicht erst auf, usw.
- Dazu bedarf es einer guten Kategorisierung, um diese fachspezifischen Filter einbauen zu können. Das PODMAN – Projekt kann mit seiner fachspezifischen Service-Matrix dazu eine Hilfestellung geben.
- RDMO soll im Bausteinprinzip genutzt werden können: Jede Institution baut sich nicht nur ihren eigenen Fragenkatalog auf, sondern auch eine eigene Service-Liste. RDMO wird installiert und dann an den institutseigenen Servicekatalog „angedockt“.
- Dies dient auch der Awareness, da oft von den Forschenden gar nicht wahrgenommen wird, welche Angebote bereits bestehen.
- RDMO soll trotzdem kompatibel bleiben.
- Frage: Ist RDMO bereits mandantenfähig? Wunsch geht dahin, dass nach dem Einloggen für eine zuvor auswählbare Fachdisziplin (z.B. Maschinenbau) nur noch eine bestimmte Auswahl an Fragen / Servicevorschlägen dargeboten wird.
- Vor ein paar Tagen erging seitens des RDMO-Projekts der Aufruf an die Nutzer, ihre Fragenkataloge zu teilen (auf GitHub), diese werden zurzeit geprüft und in Kürze dann der Allgemeinheit zur Verfügung gestellt.
- Szenario: Ein Nutzer trägt im DMP-Tool / RDMO ein, dass er zu einer bestimmten Fachdisziplin gehört (z.B. Sozialwissenschaften) und mit einer bestimmten Software arbeitet: Welche Services werden ihm dann angeboten?
- Es ist auch denkbar, dass Links / Kontaktdaten zum jeweiligen Fachreferenten gezeigt werden.

- Auch die Einbindung weiterer Servicekataloge, die in der Einrichtung bereits vorhanden sind (z.B. bei der IT) können integriert werden.
- Überfordert ein solch komplexes Tool die Nutzer? Ist es eine Zukunftsversion für in 10 Jahren? Gradwanderung, die Forschenden nicht zu erschlagen
- Zu lange Hilfstexte / Auswahlfilter könnten die Forschenden wieder abschrecken. Auch eine verringerte Repositoriumsauswahl ist nichtzwingend eindeutig genug und macht eine persönliche Beratung nicht redundant.
- Fassen Forschende ein DMP-Tool derzeit eigentlich überhaupt an ohne externen Zwang?
- Kategorien müssen sich am Lebenszyklus orientieren.
- DINI-AG (gelbe Seiten) beschäftigt sich derzeit mit dem Aufbau (als Liste) von fachspezifischen Services / Kategorien. Dieses Projekt könnte der Grundstein sein zur weiteren Umsetzung der verfolgten Vision.
  https://www.forschungsdaten.org/index.php/UAG_Gelbe_Seiten_der_AG_Forschungsdaten
- Metadaten sollten in jedem Fall nur einmal eingegeben werden müssen und dann für ein anderes Projekt (innerhalb von RDMO) weitergenutzt werden können; auch sollten diese Metadaten nicht in anderen Systemen gedoppelt werden (GEPRISs, FIS, Datenbanken für Förderanträge, Repositorien, …).
  → ambitioniertes Projekt, für all diese Systeme API-Schnittstellen zu integrieren in beide Richtungen: push und pull

Konkrete Zukunftspläne von RDMO

- Anbindung an DataCite
  (erfordert ein Mapping zwischen RDMO und DataCite-Metadatenschema)
- Anbindung an re3data war ursprünglich geplant, aber da re3data derzeit keine Ressourcen für die Weiterentwicklung des Portals hat, ist dies schwierig. re3data müsste auch einige Updates einbauen. Z.B. wird angemerkt, dass man keine Übersicht darüber ausgeben lassen kann, welche Repositorien kürzlich dazugekommen sind (Update-Filter) oder welche für eine bestimmte Fachdisziplin Open Access unterstützen. ZB MED hat dies gerade für die Lebenswissenschaften manuell herausgesucht; für die Geowissenschaften gibt es dazu das Portal
  https://repositoryfinder.datacite.org/
- Eine REST-API-Schnittstelle an Repositorien erfordert generell, dass das Rechtemanagement mit abgebildet wird
- Frage: Kann man in RDMO abbilden, wo die Daten einer Institution physikalisch liegen und diese durchsuchbar machen?
- Es sollte auch eine Anbindung an die institutionelle IT geben, damit diese beispielsweise benachrichtigt wird, wenn angegebene Hardwareanforderungen im DMP-Tool eine bestimmte Größe überschreiten, so dass hierfür schon im Antrag Gelder mitgedacht werden müssten.

Generelles Fazit:
Die Umsetzung der hier angedachten Vision wäre vielleicht ein geeignetes Thema für ein NFDI-Konsortium oder jedes Konsortium baut sich seinen eigenen RDMO-Service (RDMO4Life, 4 Earth, usw.).

RADAR Metadata Schema meets discipline-relevant Schema

Das Thema wurde von Teilnehmern der WissKom2019 vorgeschlagen.
Moderation und Notiz: Barbara Scheidt (FZJ)

Zum Start in das Thema

- Ein Diskussionsteilnehmer der ULB Halle schildert das Problem der unterschiedlichen Datensätze aus den Fachrichtungen der Universität, die sich mit einem Metadatenschema nicht abbilden lassen. U.a. hätten sich Historiker an die Bibliothek gewandt und um eine Erweiterung des Metadatenschemas gebeten. Die Historiker arbeiten mit DDI = Data Documentation Inititative.
- Die Unibibliothek bietet ein Repositorium auf Basis der Software D-Space an.
- Eine Teilnehmerin aus dem RADAR-Projektteam geht auf die daraus resultierende Frage nach den Möglichkeiten und Grenzen einer Schemaerweiterung ein:
    - Entstehung RADAR-Schema: Im Rahmen des Projekts wurden verschiedene Metadaten-Schemata gesichtet und ein Basis-Dienst, der den Grundbedarf aller Fächer abdeckt, entwickelt.
    - Darin sind 10 Pflichtelemente enthalten.
    - Relevant für die Datenerfassung ist eine DOI-Registrierung der Daten.
    - Nach bisherigen Erfahrungen reichen die Basisfelder i.d.R. aus, aber für die Beschreibung der Daten stehen darüber hinaus 13 optionale Felder zur Verfügung.
    - Das Metadatenschema wurde durch Naturwissenschaftler auf seine Anwendbarkeit mit positivem Ergebnis geprüft.
    - Aber: Es fehlt standardisiertes Vokabular.
    - Angedacht: Erweiterungen in Pakten.

Erfahrungsberichte aus der Gesprächsrunde und Diskussionen

- Universitätsbibliothek Rostock: Datenpublikationen werden seit 1½ Jahren als neuer Publikationstyp in das institutionelle Repositorium RosDok (MyCoRe) eingepflegt. Damit sei eine gute Basis vorhanden, aber die Möglichkeiten würden für viele Fächer nicht ausreichen.
- Das RADAR-Team hofft auf die Ergebnisse der RDA-Working Group zu Metadaten.
- Die Universitätsbibliothek Kaiserslautern wird immer wieder mit sehr speziellen Wünschen seitens der Fachwissenschaftlern hinsichtlich der Datenbeschreibungen konfrontiert. So wünschten sich die Biologen eine spezielle Ontologie.
- Alle Teilnehmer der Gesprächsrunde sehen die fachspezifischen Anforderungen im Rahmen einer institutionellen Lösung als problematisch (Aufwand, Machbarkeit) an und wünschten sich eine universellere Lösung.
- In Halle empfiehlt man den Wissenschaftlern als Übergangslösung das Anhängen spezieller Datenbeschreibungen in Form einer ZIP-Datei zu den Forschungsdaten.

- Zur Nutzung von Forschungsdatenrepositorien überhaupt wird einvernehmlich festgestellt, dass dies u.a. in direktem Zusammenhang mit Forderungen von Verlagen steht. Sind im Rahmen von Veröffentlichungen Datenpublikationen gefordert (Beispiel: PLOS ONE), benötigen die WissenschaftlerInnen DOIs und ein Repositorium zum Veröffentlichen der Daten.
- In Rostock gehören Forschungsdaten zum Bibliotheksbestand und werden daher im Katalog nachgewiesen. Offensichtlich laufen Planungen, diese Daten auch in der Verbundkatalogisierung (GBV) zu ergänzen (Handout für Ende des Jahres geplant?)
- In diesem Zusammenhang wird erwähnt, dass RADAR auch eine Schnittstelle für das Harvesting der Daten von teilnehmenden Einrichtungen anbietet
- Ein Teilnehmer aus der Unternehmensentwicklung des Forschungszentrums Jülich fragt grundsätzlich, ob Forschungsdaten nach Bibliotheksstandard erschlossen werden sollten oder ob der Aufwand zu groß ist. In einer Diskussion wird aber deutlich, dass zum Wiederauffinden der Daten eine sehr gute Erschließung notwendig ist.
- Für eine automatisierte Sacherschließung wäre Text notwendig, was nicht bei allen Forschungsdaten der Fall ist. Es werden weitere automatisch erfassbare Datenbestandteile diskutiert und dabei erneut die Vielfalt von FD offensichtlich.
- In der Diskussionsrunde besteht Einigkeit darüber, dass die Kompetenz für die Daten eher bei den Forschern liegt als den Bibliothekaren.
- RADAR wird als ein sehr generisches Tool charakterisiert, es sind viele Freitexte möglich. Das soll sich in Zukunft mit Erweiterungen für disziplinspezifischen Metadatenformaten ändern.
- Auf die Nachfrage über die Bedeutung der NFDI-Förderung der Communities im RADAR-Kontext erläutert die Teilnehmerin aus dem RADAR-Projektteam, dass der Input aus den Communities für die Entwicklung von RADAR äußerst wichtig ist.
  → Die Einbindung weiterer Metadatenschemata ist wünschenswert.
    - RADAR sei zudem an drei NFDI-Konsortien beteiligt.

Thema Datensuche

- Frage nach disziplinspezifischen Aggregatoren: Was ist schon da?
    - Google Datensätze sind sehr generisch, es werden wenig bis tief erschlossene Datensätze gefunden.

Thema Aufwand für die WissenschaftlerInnen:

- Problematisch
- Hinweis, dass neben einem passenden Schema für jede Disziplin auch eine Beratung zur Verbesserung der Datenqualität hilfreich wäre. Das scheint an vielen Einrichtungen noch ausbaufähig zu sein.
- Weitere Diskussionspunkte: Einheit von Systemen, Schnittstellen-Notwendigkeit und die Bereitstellung von Workflows mit Arbeitsprozessen, um den Aufwand zu reduzieren

- Kontrovers diskutierte Frage, ob Bewertung ein Anreiz für WissenschaftlerInnen sein kann, ihre Daten bereitzustellen

Vorteile institutioneller Repositorien:

- Gute Kenntnisse über die Vor-Ort-Gegebenheiten → wichtig für Beratung der Forscher
- Vorteile disziplinspezifischer Repositorien: besser für die Disziplin

Fazit: Die weitere Entwicklung wird mit Spannung erwartet. Alle wären froh, rund um das Thema Forschungsdaten, aber vor allem im Hinblick auf die Metadaten, Unterstützung aus einer größeren Community zu finden.