

Netzwerk Forschungsdaten Berlin-Brandenburg

Virtuelles Treffen am 27. Mai 2020 an der Humboldt-Universität zu Berlin

Thema: Open Data - Services, Weiterbildungen, Policy

Agenda

1. Begrüßung
2. Berichte zu Open Science Aktivitäten aus der Praxis
 - Evgeny Bobrov, Berlin Institute of Health (BIH)
 - Adrian Krolczyk, Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF)
3. Kleine Kaffeepause (entfallen)
4. Diskussion
5. Informeller Austausch (wurde aus Zeitgründen nicht mehr angeboten)

Nach einer kurzen Begrüßung von Kerstin Helbig (HU Berlin) berichtet zunächst **Evgeny Bobrov** von den Aktivitäten am Berlin Institute of Health (BIH) bzw. der Charité zum Thema Open Data. Dort werden u.a. eine Reihe von Schulungen für die Forschenden zu Open Data und Open Science angeboten. Darüber hinaus wurden Incentives entwickelt und umgesetzt, und es wird an einer Forschungsdaten-Policy gearbeitet. Der Referent verweist auf einen Artikel (<https://doi.org/10.1371/journal.pbio.3000576>), in dem alle Aktivitäten des BIH QUEST Centers näher beschrieben sind.

Evgeny Bobrov geht nun verstärkt auf die **Forschungsdaten-Policy** ein, die derzeit abgestimmt wird. Im Rahmen der Policy-Entwicklung gab es von den Forschenden sowohl den Wunsch nach stärkeren Vorgaben als auch den Hinweis, dass einzelne Formulierungen zu weit gehen würden. Ein Teilnehmer fragt, wie die Forschenden für die Gesprächsrunden zur Abstimmung der Policy ausgewählt wurden und wie die Beteiligungsrunden mit Forschenden aussahen. Evgeny Bobrov antwortet, dass die Forschungskommission beteiligt war und zur Beratung der Policy eine Untergruppe gebildet wurde. Ein weiterer Teilnehmer fragt nach begleitenden oder integrierten Handlungsempfehlungen zur Policy wie sie bspw. an der HU Berlin oder an der Uni Potsdam bestehen. Herr Bobrov antwortet, dass Handlungsempfehlungen im Nachgang zur Policy verfasst werden sollen. Jemand fragt, wer in die Überführung des DFG-Kodex in die Satzungen an den jeweiligen Institutionen beteiligt ist. Herr Bobrov antwortet, dass die derzeitige Entwurfsfassung der Policy nicht explizit auf den DFG-Kodex abgestimmt ist, mit diesem jedoch gut vereinbar ist. Eine Anpassung der Policy an den Kodex ist momentan nicht geplant, es soll jedoch eine Abstimmung zwischen Policy(-Prozess) und möglichen Änderungen an der Satzung Gute Wissenschaftliche Praxis der Charité erfolgen. Eine Teilnehmerin schreibt im Chat, dass die BTU Cottbus ihre Handlungsempfehlungen daraufhin angepasst hat. Die Policy der BTU soll erst nächstes Jahr entsprechend überarbeitet werden. Es wird nach der Dauer der Policy-Erstellung gefragt. Evgeny Bobrov antwortet, dass die Erarbeitung über ein Jahr gedauert hat und verschiedene Abteilungen involviert waren. Ein Teilnehmer fragt, ob man einschätzen könnte, wie viele Forschungsdaten in den Lebenswissenschaften auf Dauer nur mit eingeschränktem Zugang bereitgestellt werden können. Evgeny Bobrov antwortet, dass aus seiner Sicht mehr als die Hälfte der Forschungsdaten aus der klinischen Forschung nur mit eingeschränktem Zugang verfügbar gemacht werden können.

Evgeny Bobrov berichtet nun über die Entwicklung von **Incentives** für Forschende zur Veröffentlichung ihrer Daten. Im Rahmen der Leistungsorientierten Mittelvergabe (LOM) wird auch ein kleiner Anteil für die Veröffentlichung von Forschungsdaten ausgezahlt. Die FAIR-Kriterien werden dabei nicht zu Grunde gelegt. Für die Suche nach Datenpublikationen der Forschenden der Einrichtung wurde ein Algorithmus von einem Data Scientist des BIH entwickelt (ODDPub,

<https://doi.org/10.1101/2020.05.11.088021>). Ein Teilnehmer fragt, welche Incentives durch eine entsprechende Publikation konkret erlangt werden können. Evgeny Bobrov antwortet, dass ca. 800 € pro Erst- und/oder Letzt-Autorenposition über die LOM (max. ca. 1600 € pro Publikation) ausgezahlt werden. Darüber hinaus gibt es einen QUEST-Preis von 1000€ für die Nachnutzung offener Daten. Jemand fragt, ob mehrere Datensätze pro Publikation für die LOM zugelassen werden. Herr Bobrov antwortet, dass es nicht mehr Geld für mehr Datensätze gibt. Andersherum kann aber ein Datensatz mehrere Publikationen erzeugen. Dies wurde als Ausnahme zugelassen.

Zuletzt geht Evgeny Bobrov kurz auf die entwickelten **Fortbildungsmaßnahmen** und **hilfreiche Ressourcen** ein. Es gibt Open Data Workshops, die monatlich angeboten werden. Diese dauern drei Stunden, haben max. 20 Teilnehmende und sind in der Regel sehr gut besucht. Die bisherigen Fortbildungsteilnehmer sollen demnächst befragt werden, ob sie ihre neuen Kenntnisse nutzen konnten, um den Schulungserfolg zu evaluieren. Herr Bobrov verweist darüber hinaus auf ein Seminar zum Thema „Rechtssicheres Teilen von Daten“, das am 3. Juni 2020 als Webinar angeboten wird. Eine Anmeldung von Externen ist möglich

(https://www.bihealth.org/de/aktuell/?L=0&tx_news_pi1%5Bnews%5D=2876&tx_news_pi1%5Bcontroller%5D=News&tx_news_pi1%5Baction%5D=detail&cHash=0d5d6e8e9f38a811559a314c5482825c). Das Webinarmaterial soll soweit möglich im Nachgang verfügbar gemacht werden. Des Weiteren wurde ein Open Data FAQ (<https://www.bihealth.org/de/forschung/quest-center/mission-ansatze/open-science/faq-open-data/>) und ein Guide zur Veröffentlichung von Datensätzen sowie Null-Resultaten (<https://s-quest.bihealth.org/fiddle/>) entwickelt, welche auch für andere nützlich sein könnten.

Nun berichtet **Adrian Krolczyk** von den Aktivitäten am Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF). Nach einem kurzen Überblick zur Forschung am ZALF, die von einer hohen Datenheterogenität geprägt ist, berichtet er zur AG RDA. Ein Teilnehmer fragt, von wem die Initiative zur Gründung der AG ausging. Herr Krolczyk antwortet, dass die Initiative von der Leitung des ZALF und der Leibniz-Gemeinschaft ausging. Die Arbeitsgruppe berät schwerpunktmäßig Forschungsprojekte, von denen aktuell ca. 30 am ZALF angesiedelt sind. Die Services des ZALF sind dabei abgestimmt auf die Projektlaufzeit und den Forschungsprozess (ZALF FDM Prozess) und richten sich nicht am Forschungsdatenlebenszyklus aus. Ziel ist das möglichst gute Ineinandergreifen von Services. Jemand fragt, wie die Unterstützung der Forschenden am Forschungsprozess in der Realität unterstützt werden. Gibt es eine direkte Anbindung in Projekten oder einen individuellen Zuschnitt der Dienste? Herr Krolczyk antwortet, dass die Projekte idealerweise ab der Projektskizze unterstützt werden, aber recht unterschiedliche Kontexte und Einstellungen der Forschenden bestehen, sodass eine enge Kommunikation notwendig ist. Die AG ist dabei so eng wie möglich und gewünscht in das Projekt eingebunden.

Die von der AG innerhalb von vier Monaten verfasste **Data Policy** (<http://doi.org/10.5281/zenodo.2567723>) wurde in Abstimmung mit acht Stakeholdern des ZALF entwickelt. Dabei waren die Ombudsperson, der Datenschutzbeauftragte, eine Forschende, der Betriebsrat, die UB-Leitung, die IT-Leitung, die juristische Beauftragte und der FDM. Die Policy wurde im Rahmen von Workshops im Think-Tank-Format erstellt. Darüber hinaus wurden per Text Mining die bisherigen veröffentlichten Data Policies auf [forschungsdaten.org](https://www.researchdata.org) analysiert und mit der ZALF-Policy verglichen. Jemand fragt, ob bei der Analyse spezifische Unterschiede zwischen den Data Policies von Hochschulen und Forschungseinrichtungen aufgefallen sind. Herr Krolczyk antwortet, dass die Analyse signifikante Unterschiede zwischen den Policies von Hochschulen und Forschungseinrichtungen zeigte.

Die veröffentlichte Data Policy nimmt unter anderem Bezug auf die Leitlinie der Leibniz-Gemeinschaft (<https://www.leibniz->

[gemeinschaft.de/fileadmin/user_upload/Bilder_und_Downloads/Forschung/Open_Science/Leitlinie Forschungsdaten 2018.pdf](https://www.gemeinschaft.de/fileadmin/user_upload/Bilder_und_Downloads/Forschung/Open_Science/Leitlinie_Forschungsdaten_2018.pdf)), enthält eine Verpflichtung zur DMP-Erstellung und hat eine Laufzeit von zwei Jahren. Danach ist eine Aktualisierung oder Verlängerung erforderlich. Handlungsempfehlungen zur Policy werden aktuell erarbeitet. Ein Teilnehmer fragt, wie die Forderung umgesetzt wird einen DMP zu erstellen. Wie viele DMP wurden schon erstellt? Werden sie den anderen am ZALF zur Verfügung gestellt? Adrian Krolczyk antwortet, dass dies nicht direkt kontrolliert wird, hierzu aber viel Überzeugungsarbeit geleistet wird und die Forschungsdatenmanager der Projekte aktiv, Stück für Stück begleitet werden. Unterstützt wird u.a. mit Templates für verschiedene Forschungsförderer, Beratung und Services. Ein institutionsspezifisches ZALF Template wird derzeit getestet.

Zur leichten Umsetzung der Data Policy wird derzeit die **Kollaborationsplattform RDA Service base** entwickelt, die als ein zentraler Dienst im internen ZALF-Netz zur Verfügung gestellt werden soll. Als One-Point-of-Entry sollen dort verschiedene Funktionalitäten für die Forschenden abgedeckt und über Schnittstellen Informationen und Daten ausgetauscht werden. Ein Teilnehmender fragt, ob der Code der Kollaborationsplattform ausschließlich für das ZALF oder als Open Source entwickelt wird. Adrian Krolczyk antwortet, dass der Code Open Source sein wird, er jedoch derzeit allein daran programmiert und es dementsprechend noch bis Mitte/Ende 2021 dauern kann bis eine erste Version des Codes zur Verfügung gestellt wird. Jemand fragt, wie Personendaten im System eingebunden sind. Herr Krolczyk antwortet, dass derzeit hierfür nur eine Excel-Liste zur Verfügung steht - entsprechende interne Webservices sind in der Entwicklung. Die ORCID soll implementiert werden.

Ein Teilnehmender fragt, ob das ZALF an einem NFDI-Konsortium beteiligt ist. Adrian Krolczyk bejaht dies und nennt [NFDI4Agri](#) und [NFDI4Earth](#).

Das virtuelle Treffen endet mit einem Dank an die Referenten und die über 30 interessierten Teilnehmerinnen und Teilnehmer.