

KOLUMNE

Vom Gedeihen des Forschungsdatenmanagements – wie die Pflanzen – bis hin zu Künstlicher Intelligenz

Annette Strauch-Davey

► Nach einer langen Sommerpause, die zum Teil auch ein wenig in den Arbeitsgemeinschaften der Forschungsdatenmanagement (FDM)-Initiativen spürbar war, bringt der Herbst 2024 nun weiterhin wieder sehr viel Arbeit und neue Veranstaltungen zum Umgang mit Daten in der Forschung „an Land“. Was beschäftigt mich gerade neben dem FDM-Alltagsgeschäft?

Entwicklungen! – Ja, wie sich im Herbst die Blätter verfärben, so entfaltet sich FDM ständig weiter und kann an Universitäten und Forschungseinrichtungen, auch mit Hilfe von Bibliotheken, innovativen Services und mittels Künstlicher Intelligenz (KI) auf- und ausgebaut werden.



*Annette Strauch-Davey M.A. ist am DESY in Hamburg beschäftigt. Sie arbeitet dort als Section Liaison Officer für den Verbund Base4NFDI.
annette.strauch-davey@desy.de*

und um Anwendungen von KI im Forschungsdatenmanagement der Disziplinen. In Prozessen und innerhalb des Datenlebenszyklus (digitalen Räumen) helfen uns KI-Werkzeuge längst bei der Generierung von (neuen) Forschungsdaten, und Automatisierungen in der Medizin können für sehr gute Datensätze sorgen (Stichwort Datenqualität). Dateninfrastrukturen sollen auf sicherer KI basieren, denn das ist unter anderem essentiell für das Datenteilen, eine Datensouveränität oder den Umgang mit sensiblen Forschungsdaten und so weiter. Data Science, durch die Informatik vorangetrieben, und KI sind auch in der Nationalen Forschungsdateninfrastruktur (NFDI) allgegenwärtig. Das NFDI-Konsortium NFDI4DS¹



*Abb. 1: FDM-Aufbau, ein bisschen wie Feng Shui bei der Gestaltung von Lebensräumen, wie ein Garten
<https://pixabay.com/de/illustrations/feng-shui-steine-meditationen-1019929/> (Peggy_Marco)*

Die neuen Möglichkeiten des Datenmanagements sind durch gute Vernetzungsarbeit und den Ideenaustausch gegeben, denn wenn Fallbeispiele (Use Cases) besprochen werden, kann gleich Neues gedeihen.

Den Datenmanagenden geht es um Support der Forschenden, um Sichtbarkeit von Forschungsleistungen und vor allem darum, dass die Wissenschaftlerinnen und Wissenschaftler mit interoperablen Forschungsdaten FAIR arbeiten können. Es geht um Datenkompetenzen, aktuell noch mehr um KI-Kompetenzen wie bisher

¹ <https://www.nfdi4datascience.de/>; Arbeitsprogramm: <https://www.nfdi4datascience.de/about/> (Zugriff am 11.09.2024)



Abb. 2: KI (z.B. bei Generierung von Forschungsdaten oder bei der Datenorganisation, u.v.v.m.) [https://pixabay.com/de/photos/mockup-schreibmaschine-wort-5288034/\(viarami\)](https://pixabay.com/de/photos/mockup-schreibmaschine-wort-5288034/(viarami))

entwickelt offene Forschungsdateninfrastrukturen, die auf Wissensgraphen² basieren. Das Arbeiten mit großen Datenmengen kann unterstützt werden.

“Task Area 2: Research Knowledge Graphs aims at improving FAIRness of Data Science artifacts including research datasets, benchmarks, machine learning models and research software (code and executables). The last years have seen a paradigm shift in Data and Computer Science towards data-driven and deep learning-based methods, which often rely on a combination of code, models, and underlying datasets. However, lack of transparency about data, code, or models is a cause for significant reproducibility and reusability issues, which surfaced across various domains. Therefore, we will follow an integrated approach towards representing and linking Data Science artifacts into a joint Research Knowledge Graph (RKG), enabling a transparent understanding of code/data provenance, model configurations, and training/testing data.”

Bei der Fachtagung „Gesundheit Gemeinsam“,³ die vom 8. bis 13. September in Dresden stattfand, gab es zahlreiche Beiträge zu Dateninfrastrukturen in der Medizin, unter anderem vom NFDI-Konsortium NFDI4Health.⁴ Ein Tutorial hatte bei der Tagung beispielsweise das Thema Datenschutz in der medizinischen Forschung zum Inhalt, etwas, das unbedingt zum FDM dazugehört, denken wir an die Informierte Einwilligung, Datenschutzkonzepte, das Gesundheitsdatennutzungsgesetz sowie die Datenschutz-Folgenabschätzung. Weitere Themen dieser Ta-

gung war Cloudspeicher in deutschen Krankenhäusern, Datenmanagement in Studien, synthetische Daten,⁵ FHIR-kodierte Daten des Kerndatensatzes der Medizininformatik-Initiative⁶ unter Einsatz von Open-Source-Software neben ganz „viel anderem Stoff“. Die Diversität der Fächerkulturen spiegelt sich im Forschungsdatenmanagement und in den dazugehörigen Veranstaltungen in dem Bereich wider. Forschungsdaten müssen wie Pflanzen unterschiedlich behandelt werden, aber manchmal ist „die Pflege“ bei beiden sehr ähnlich, das heißt, gemeinsame Dienste können genutzt werden. Der richtige Dünger spielt beim Gärtnern, und übersetzt ins FDM als guter Dienst, eine wichtige Rolle. Forschende haben aktuell viele Möglichkeiten, ein ordentliches FDM mitzugestalten, unter anderem, indem sie Dienste und Werkzeuge ausprobieren, sie weiterentwickeln und schauen, was nützlich ist oder nicht. Das, was sich nicht bewährt, kann abgeworfen werden, so wie die Blätter im Herbst.

Die Landesinitiative für Forschungsdatenmanagement-fdm.nrw veranstaltet am 12. November 2024 ein NRW-weites Programm, online und in Präsenz, auf das ich hinweisen möchte. Aus der Forschungspraxis heraus soll es den Austausch am 4. Tag der Forschungsdaten unter dem Motto „Auf in den Daten-Garten: Bring deine Forschung zum Blühen“ geben.⁷ Die Metapher des Gedeihens für diese Kolumne ist im Zusammenhang mit dieser Veranstaltung entstanden:

Save the Date: 12. November 2024 | 4. Tag der Forschungsdaten in NRW | Weitere Informationen folgen auf dieser Webseite: <https://fdm-nrw.coscine.de/#/Mitmachen?id=veranstaltungen>

Kontakt: info@fdm.nrw (Team fdm.nrw)



Abb. 3: Ein Daten-Garten wie ein Garten durch Netzwerken und Austausch [https://pixabay.com/de/vectors/g%C3%A4rtner-gartenarbeit-garten-7096952/\(dandelion_tea\)](https://pixabay.com/de/vectors/g%C3%A4rtner-gartenarbeit-garten-7096952/(dandelion_tea))

2 Siehe: Natural Scientific Language Processing and Research Knowledge Graphs First International Workshop, NSLP 2024, Hersonissos, Crete, Greece, May 27, 2024, Proceedings <https://link.springer.com/book/10.1007/978-3-031-65794-8> (Zugriff am 11.09.2024)

3 <https://gesundheit-gemeinsam.de/> (Zugriff am 10.09.2024)

4 <https://www.nfdi4health.de/> (Zugriff am 10.09.2024)

5 Kühnel L, Schneider J, Perrar I, et al. Synthetic data generation for a longitudinal cohort study – evaluation, method extension and reproduction of published data analysis results. *Sci Rep.* 2024;14:14412. DOI: [10.1038/s41598-024-62102-2](https://doi.org/10.1038/s41598-024-62102-2)

6 <https://www.medizininformatik-initiative.de/de/der-kerndatensatz-der-medizininformatik-initiative> (Zugriff am 10.09.2024)

7 <https://fdm-nrw.coscine.de/#/Mitmachen?id=veranstaltungen> (Zugriff am 11.09.2024)