

Ein Fallbeispiel zum Umgang mit Learning-Analytics-Forschungsdaten

Fachspezifische Konzeptentwicklung eines Data Librarians

Ian Wolff, Otto-von-Guericke-Universität Magdeburg

Sebastian Zug, Technische Universität Bergakademie Freiberg

Veit Köppen, Zentral- und Landesbibliothek Berlin

Zusammenfassung

Offene Wissenschaft in die Praxis umzusetzen, bringt für jede Fachwissenschaft und Wissenschaftler*innen Herausforderungen mit sich. Mit dem Aufkommen von Data Librarians hat sich ein neues Berufsbild etabliert, mit dem wissenschaftliche Bibliotheken die Prozesse im Forschungsdatenmanagement umsetzen können. In diesem Beitrag werden die grundsätzlichen Rollen und Aufgaben von Data Librarians erläutert und ihr spezifischer Einsatz im Rahmen des BMBF-Forschungsprojekts DiP-iT aufgezeigt. Ausgehend von einer gemeinsamen Verständigungsebene in Form eines Dateninterviews wurden fachspezifische Konzepte für die Learning Analytics in den Bereichen Datendokumentation, -kuration und -organisation entwickelt. Ergebnisse sind ein Datentransfermodell, das die Speicherung der Daten in verschiedenen Domänen unter Einbezug datenschutzrechtlicher Rahmenbedingungen ermöglicht, und die detaillierte und fachspezifische Dokumentation der Daten anhand eines Learning Analytics Metadatenmodells.

Summary

Putting open science into practice brings challenges for every discipline and every scientist. With the emergence of data librarians, a new occupational profile has been established by academic libraries to implement processes in research data management. This article explains basic roles and tasks of data librarians and shows their specific tasks within DiP-iT, a research project funded by the Federal Ministry of Education and Research. Starting from a common level of understanding by a data interview, subject-specific concepts for learning analytics in the areas of data documentation, curation, and data organization were developed. One of the results is a data transfer model that enables storing data in different domains while taking into account data protection requirements. Another result is the detailed and subject-specific documentation of the data by means of a learning analytics metadata model.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5896>

Autorenidentifikation:

Ian Wolff: ORCID: <https://orcid.org/0000-0002-0413-0035>; GND: 1238114954

Sebastian Zug: ORCID: <https://orcid.org/0000-0001-9949-6963>; GND: 102224731X

Veit Köppen: ORCID: <https://orcid.org/0000-0002-6068-3275>; GND: 1054320845

Schlagwörter: Forschungsdatenmanagement, Learning Analytics, Data Librarian, Data Librarianship, Metadaten

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0](https://creativecommons.org/licenses/by/4.0/).

1. Einleitung

Die Datenrevolution und Herausforderungen durch Open Science haben sich auf Forschende aller Fachrichtungen ausgewirkt.¹ Mit der Digitalisierung gehen für Forschende grundlegende Veränderungen in der Art und Weise einher, wie sie ihre Forschung gestalten. Dadurch bedingt sind auch Veränderungen in der Wissenschaftskommunikation, da unter anderem von Drittmittelförderern eine offene Art der Veröffentlichung von Ergebnissen gefordert wird, um Forschung und Methoden digital und transparent zu halten.² Somit werden Forschungsergebnisse im Sinne von Open Science vergleichbarer und reproduzierbarer, was positive Effekte auf die Wissenschaftskommunikation hat.

Dieser Wandel wird von wissenschaftlichen Bibliotheken unterstützt und begleitet, wobei der Umgang mit Forschungsdaten neue Anforderungsprofile mit sich bringt. Im Bereich der forschungsnahen Dienstleistungen hat dies zu neuen Rollen und Aufgaben geführt, die unter anderem von Data Librarians erfüllt werden.³

Im Folgenden soll anhand des Learning-Analytics-Projekts DiP-iT⁴ – Digitales Programmieren im Team – der Einsatz eines in den Projektkontext integrierten Data Librarians veranschaulicht werden. Learning Analytics nutzen dynamische Informationen über Lernende und Lernumgebungen, indem sie diese bewerten, erheben und analysieren, um Lernprozesse, Lernumgebungen und pädagogische Entscheidungen in Echtzeit zu modellieren, vorherzusagen und zu optimieren.⁵ Obwohl die Learning Analytics zu den datengetriebenen Wissenschaften gehören, besitzen sie noch keine Open-Data-Kultur oder Best-Practice-Ansätze im Forschungsdatenmanagement.⁶ Dies ist insbesondere problematisch, weil in der entsprechenden Community die Datensätze einzelner Untersuchungen statistisch gesehen eher klein ausfallen. Die avisierte übergreifende Nutzung des Datenmaterials ermöglicht in diesem Kontext neue methodische Untersuchungsansätze.

- 1 Kitchin, Roy: The data revolution. Big data, open data, data infrastructures & their consequences, Los Angeles 2014. Online: <<https://dx.doi.org/10.4135/9781473909472>>.
- 2 Koltay, Tibor: Are you ready? Tasks and roles for academic libraries in supporting Research 2.0, in: New Library World (1/2), 2016, S. 94-104. Online: <<https://doi.org/10.1108/NLW-09-2015-0062>>; Konkretisierung der Anforderungen zum Umgang mit Forschungsdaten in Förderanträgen, Information für die Wissenschaft Nr. 25, 14. März 2022, Online: <https://www.dfg.de/foerderung/info_wissenschaft/2022/info_wissenschaft_22_25/index.html>, Stand: 06.12.2022.
- 3 Eine Definition zu Data Librarianship gibt die CILIP: „Data librarians are engaged in managing research data, using that data as a resource and supporting researchers in these activities. As a data librarian it's likely that you'll be involved in developing or implementing an organisations data management plan, storing and managing data and determining retention and disposal periods. You'll be expected to engage and potentially train other staff in issues such as copyright, intellectual property, licensing of data, embargoes, ethics and reuse, data literacy and privacy.“ Data Librarianship, <<https://www.cilip.org.uk/page/DataLibrarians>>, Stand: 06.12.2022.
- 4 Siehe Projektwebsite zum Projekt DiP-iT, Digitales Programmieren im Team, <<https://www.dip-it.ovgu.de/>>, Stand: 06.12.2022.
- 5 Grandl, Maria; Taraghi, Begnam; Ebner, Martin et al.: Learning Analytics, in: Wilbers, Karl; Hohenstein, Andreas (Hg.): Handbuch E-Learning. Expertenwissen aus Wissenschaft und Praxis. Strategien, Instrumente, Fallstudien. Köln 2017, S. 1-16.
- 6 Biernacka, Katarzyna; Pinkwart, Niels: Opportunities for Adopting Open Research Data in Learning Analytics, in: Azevedo, Ana; Azevedo, José Manuel; Onohuome Uhoimobhi, James u.a. (Hg.): Advancing the power of learning analytics and big data in education, Hershey 2021, S. 29-60. Online: <<https://doi.org/10.4018/978-1-7998-7103-3>>.

Um diesen Datenaustausch sicherzustellen, werden im Fallbeispiel zur Zusammenarbeit mit der TU Freiberg durch den Data Librarian fachspezifische Lösungen im Bereich des Datenmanagements und der Dokumentation entworfen. Ausgangspunkt bilden die Daten aus der Anwendung von Learning Analytics Methoden auf kollaborative Programmierlernszenarien in einer Lernumgebung in Github. Vorrangiges Ziel ist es, den Datenaustausch zwischen den verschiedenen Projektteams und Standorten zu ermöglichen. In diesem Kontext erfolgt die Konzentration auf das Erstellen eines disziplinspezifischen Metadatenschemas, das gleichzeitig in eine Datenaustauschumgebung integriert wird. Insbesondere wird dabei auf den Austausch zwischen den Forschenden und dem Data Librarian eingegangen, die sich thematisch für ein gezieltes und disziplinspezifisches Forschungsdatenmanagement einander annähern mussten.

2. Data Librarianship

Das Aufkommen von Data Librarians kann als Reaktion auf die immer weiter fortschreitenden Innovationen im Zuge der Digitalisierung der wissenschaftlichen Forschung betrachtet werden. Denn mit der Digitalisierung und der Open-Science-Bewegung ging eine Öffnung der wissenschaftlichen Kommunikation einher, die Wissenschaftler*innen eine möglichst schnelle Informationsversorgung ermöglicht, aber auch Herausforderung bei ihrer Umsetzung mitbringt.⁷ Forschungsdatenmanagement als Open-Science-Teilkomponente wurde bereits vor ca. zehn Jahren von verschiedenen Bibliotheksverbänden als zukünftiges bibliothekarisches Feld erkannt, genauso wie die dafür neu zu entwickelnden Rollen, Konzepte und Berufsbilder.⁸ Doch bereits vor der Positionierung der Fachgesellschaften zur Notwendigkeit von Forschungsdatenmanagement und dem dafür benötigten spezialisierten Personal, entwickelte sich bereits das Berufsbild Data Librarian in Spezialbibliotheken. Ihren Ursprung besitzen sie in den social-science-Disziplinen (Politik, Wirtschaft, Psychologie, Soziologie, Anthropologie etc.) in Großbritannien, den USA und Kanada, wo insbesondere Serviceleistungen in Bezug auf Daten und deren Archivierung institutionell verankert wurden. Das GESIS – Leibniz-Institut für Sozialwissenschaften ist als Pendant im deutschen Raum zu betrachten.⁹ Im Rahmen des IASSIST auf dem World Congress of Sociology of Toronto 1974 erfolgte dann die erste Erwähnung des Data Librarianships für ein zukünftiges Berufsbild als Entwicklung von Datenunterstützungsdiensten, Festlegung von Standards für die Verwaltung und den Austausch von computerlesbaren sozialwissenschaftlichen Daten.¹⁰ Ausgehend von der sozialwissenschaftlichen Herkunft des Begriffs konstatieren Semeler, Pinto und Rozados, dass Rollen für Data Librarians „*focus on disseminating the important findings relating to their research in the form of relevant information by gathering data from various sources and*

7 Heise, Christian: Von Open Access zu Open Science. Zum Wandel digitaler Kulturen der wissenschaftlichen Kommunikation, Lüneburg 2018, 101f. Online: <<http://library.oapen.org/handle/20.500.12657/37554>>.

8 ERU Roadmap for Research Data, League of European Research Universities Research Data Working Group, Leuven 2013. Online: <<https://www.leru.org/files/LERU-Roadmap-for-Research-Data-Full-paper.pdf>>; Christensen-Dalgaard, Britte: Ten Recommendations for Libraries to Get Started with Research Data Management, LIBER, 2012. Online: <<https://libereurope.eu/wp-content/uploads/2020/11/Theresearch-data-group-2012-v7-final.pdf>>.

9 Data Librarianship. Responding to research innovation, in: Rice, Robin; Southall, John (Hg.): The Data Librarian's Handbook, 2018, S. 1-18. Online: <<https://doi.org/10.29085/9781783301836>>.

10 Adams, Margaret O.: The Origins and Early Years of IASSIST, in: IASSIST Quarterly 30 (3), 2007, S. 6. Online: <<https://doi.org/10.29173/iq108>>; Thompson, Kristi Anne; Kellam; Lynda: Introduction, in: Thompson, Kristi Anne; Kellam; Lynda (Hg.) Databrarianship. The Academic Data Librarian in Theory and Practice, 2016, S. 3-4. Online: <<https://scholar.uwindsor.ca/leddylibrarypub/47>>.

organizing and cleaning them. The role of data librarians is to act as facilitators in all stages of scientific research, contributing with potential services that might be useful for the data management and curation processes.”¹¹ Im Jahr 2007 publizierte Anna Gold zwei Artikel, in denen sie die Parallele zwischen Forschungszyklus und Datenlebenszyklus sowie eine Analogie zwischen Downstream und Upstream zur Beschreibung von Data-Librarian-Aufgabenfeldern zieht. „Downstream“ meint die Prozesse bei der Auswahl und Lizenzierung wissenschaftlicher Daten, deren Beschreibung, Entdeckung, Organisation und Dokumentation. Mit „Upstream“ ist die enge Verknüpfung zwischen Bibliotheken und Forschung gemeint, die bereits zum Beginn des Datenlebenszyklus ansetzt.¹² Aus diesem Rollenverständnis heraus bildete sich der Begriff „databrarian“¹³ (data + librarian), der im gleichnamigen Buchtitel von Kellam und Thompson insbesondere Rollen im Bereich des Zugangs zu Daten, deren Dokumentation und Erhalt umfasst und der fachspezifischen Unterstützung einzelner Disziplinen.¹⁴ Grundsätzlich lässt sich ableiten, dass es sich bei Data Librarianship um die Anwendung der Grundsätze, Praktiken und Ressourcen des traditionellen Bibliothekswesens auf digitale Daten handelt. Das umfasst alle traditionellen Phasen der bibliothekarischen Arbeit, wie Erwerbung (Entwicklung von Sammlungen), Organisation (Katalogisierung und Metadaten) und die Umsetzung von Serviceleistungen, wozu sich noch die Auswertung und Analyse von Daten hinzukommen.¹⁵ Daraus wird bereits ersichtlich, dass die neuen datenorientierten Aufgaben mit Herausforderungen für bereits bestehendes und zukünftiges Bibliothekspersonal einhergehen. Dementsprechend wird diesem Trend im deutschsprachigen Raum mit Fort- und Weiterbildungen sowie Anpassungen der LIS-Curricula begegnet.¹⁶

- 11 Semeler, Alexandre Ribas; Pinto, Adilson Luiz; Rozados, Helen Beatriz Frota: Data science in data librarianship: Core competencies of a data librarian, in: Journal of Librarianship and Information Science 51 (3), 2019, S. 772. Online: <<https://doi.org/10.1177%2F0961000617742465>>.
- 12 Gold, Anna: Cyberinfrastructure, Data, and Libraries. Part 1. A Cyberinfrastructure Primer for Librarians, in: D-Lib Magazine 13 (9/10), 2007. Online: <<http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>>; Gold, Anna: Cyberinfrastructure, Data, and Libraries. Part 2. Libraries and the Data Challenge. Roles and Actions for Libraries, in: D-Lib Magazine 13 (9/10), 2007. Online: <<http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>>.
- 13 Erstmals wurde der Begriff im Library Journal genannt, Maatta, Stephanie L., Placement & Salaries 2013. The Emerging Databrarian, in: Library Journal 138 (17), S. 26-43. Online: <<https://www.libraryjournal.com/story/the-emerging-databrarian>>.
- 14 Thompson, Kristi Anne; Kellam; Lynda (Hg.) Databrarianship. The Academic Data Librarian in Theory and Practice, 2016.
- 15 Koltay, Tibor: Data literacy for researchers and data librarians, in: Journal of Librarianship and Information Science 49 (1), 2017, S. 3–14. Online: <<https://doi.org/10.1177%2F0961000615616450>>; Xia, Jingfeng; Wang, Minglu: Competencies and Responsibilities of Social Science Data Librarians: An Analysis of Job Descriptions, in: College & Research Libraries 75 (3), 2014, S. 362–388. Online: <<https://doi.org/10.5860/crl13-435>>; Khan, Hammad Rauf; Du, Yunfei: What is a data librarian? A content analysis of job advertisements for data librarians in the United States academic libraries, in: IFLA World Library and Information Congress 2018, Kuala Lumpur 2018, S. 1-9. Online: <<https://digital.library.unt.edu/ark:/67531/metadc1225772/>>; Eclevia, Marian Ramos; La Torre Fredeluces, John Christopher; Eclevia, Carlos; Saguibo Maestro, Roselle: What Makes a Data Librarian? An Analysis of Job Descriptions and Specifications for Data Librarian, in: Qualitative and Quantitative Methods in Libraries 8 (3), 2019, S. 273-290. Online: <<http://qqml-journal.net/index.php/qqml/article/view/541>>. Siehe weiterhin Rice, Robin: Supporting Research Data Management and Open Science in Academic Libraries: a Data Librarian's View, in: Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare 72 (2), 2019, S. 263–273. <<https://doi.org/10.31263/voebm.v72i2.3303>>.
- 16 Rezniczek, Alina; Blumesberger, Susanne; Bargmann, Monika u. a.: Der Zertifikatskurs „Data Librarian“ und seine erstmalige Durchführung, in: Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare 72 (2), 2019, S. 274–283. Online: <<https://doi.org/10.31263/voebm.v72i2.3176>>; Fühles-Ubach, Simone; Schaer, Philipp; Lepsky, Klaus et al.: Data Librarian – ein neuer Studienschwerpunkt für wissenschaftliche Bibliotheken und Forschungseinrichtungen, in: Bibliothek Forschung und Praxis 43 (2), S. 255-261. Online: <<https://doi.org/10.1515/bfp-2019-2053>>; Fühles-Ubach, Simone: Bibliothekare und Data Librarians – neue Profile für das bibliothekarische Fachpersonal der Zukunft, in: o-bib. Das offene Bibliotheksjournal 5 (4), 2018, S. 7-17. Online: <<https://doi.org/10.5282/o-bib/2018H457-17>>. Tappenbeck, Inka; Meinhardt, Haike: MALIS Reloaded. Der berufsbegleitende

3. Der Data Librarian im DiP-iT-Projekt

Im Projekt DiP-iT (Digitales Programmieren im Team) wird im Verbund mit der Technischen Universität Freiberg, Humboldt-Universität zu Berlin und Otto-von-Guericke-Universität Magdeburg das kollaborative Erlernen von Programmiersprachen untersucht. Dafür kommen Ansätze und Methoden der interdisziplinär arbeitenden Learning Analytics zum Einsatz. Im Fall des DiP-iT-Projekts werden für die kollaborativen Programmierlernszenarien verschiedene Softwareframeworks zum Messen der Lernaktivitäten durch Teams an den Fakultäten für Informatik an allen drei Universitäten entwickelt. Zur didaktischen Gestaltung der Lehr-Lernveranstaltungen und um die Lernprozesse sichtbar zu machen, begleitet ein Team an der humanwissenschaftlichen Fakultät der Magdeburger Universität die Lehrveranstaltungen. In den vergangenen zwei Projektjahren und noch fortlaufend für das Wintersemester 2022/2023, wurden Daten in Lehrveranstaltungen zum Erlernen der Programmiersprache SQL an der Universität in Magdeburg und C++ an der Universität Freiberg erhoben. Das Forschungsdatenmanagement wurde bereits in der Projektantragsphase mitbedacht, indem ein Data Librarian mit einem Forschungsanteil im Projekt für die Universitätsbibliothek Magdeburg beantragt wurde. Dieser adressiert die Bedürfnisse der Wissenschaftler*innen im Projekt und der Learning Analytics auf übergeordneter Ebene. Insgesamt sind den Forschenden im Bereich der Learning Analytics Offenheit und Transparenz der erhobenen Daten und Privacy-Aspekte von besonderer Wichtigkeit.¹⁷ Dies sind auch gleichzeitig die größten Herausforderungen der Learning-Analytics-Community bei der Speicherung von Datensätzen. Weitere Herausforderungen sind die dementsprechenden Lösungen zur Anonymisierung der Daten, die standardisierte Dokumentation von Datensätzen sowie die Vergabe geeigneter Lizenzen, die eine Nachnutzung erlauben.¹⁸ Den genannten Herausforderungen nimmt sich der Data Librarian als Teil des Projektteams in Zusammenarbeit mit den Forschenden an.

Zuerst wurde nach der ersten Erhebungsphase am Standort in Freiberg zur internen Verständigung ein Dateninterview durchgeführt, um sich einander anzunähern. Im Projektverlauf folgten dann noch weitere Interviews. Danach analysierte der Data Librarian die Vorgänge der kollaborativen Datenentstehung von Programmierlernszenarien und kontextualisierte sie mit den Forschungsabläufen der Learning Analytics. Entlang der Spezifika der Learning-Analytics-Forschung wurde dann ein Datentransfermodell für die im Projekt anfallenden Daten entwickelt. Das Transfermodell bezieht dabei insbesondere datenschutzrechtliche Grundsätze mit ein, die relevant sind für einen Austausch der Daten im Projektkontext und für eine spätere Publikation. Diese Prozesse müssen von Beginn an von den Forschenden mitgedacht werden. Nachdem die ersten Datensätze generiert wurden, erfolgte ein weiteres Gespräch mit den Forschenden, um Beschreibung und Dokumentation der Daten in Form eines Metadatenmodells zu entwickeln. Das Metadatenmodell kommt dann in den Datenaustauschumgebungen (Repositorien) zum Einsatz und wird erprobt und angepasst.

Masterstudiengang „Bibliotheks- und Informationswissenschaft“ der TH Köln präsentiert sich mit einem neuen Curriculum, in: o-bib. Das Offene Bibliotheksjournal 8 (2), 2021, S. 1–9. Online: <<https://doi.org/10.5282/o-bib/5708>>.

17 Scheffel, Maren; Drachsler, Hendrik; Stoyanov, Slavi u.a.: Quality Indicators for Learning Analytics. Educational Technology & Society, 17 (4), 2014. Online: <<https://www.jstor.org/stable/jeductechsoci.17.4.117>>.

18 Greller, Wolfgang; Drachsler, Hendrik: Translating Learning into Numbers. A Generic Framework for Learning Analytics. Educational Technology & Society 15 (3), 2012, S. 48f. Online: <<https://www.jstor.org/stable/jeductechsoci.15.3.42>>.

3.1 Dateninterviews

Im Vortrag auf dem Bibliothekskongress 2022 „Ein Fallbeispiel zum Umgang mit Learning Analytics Forschungsdaten – fachspezifische Konzeptentwicklung eines Data Librarians“ in Leipzig, stand auf einer abstrakten Ebene das Verdeutlichen der Wichtigkeit von Dateninterviews zwischen den Learning Analytics Forschenden der TU Freiberg und dem Data Librarian im Vordergrund.¹⁹ Die Ausgangssituation bildeten dabei der Data Librarian, für den Learning-Analytics-Forschung vollkommen neu war, sowie der Learning Analytics Forschenden, der bisher nur wenige Berührungspunkte mit Aspekten des Forschungsdatenmanagement hatte. Nach der ersten Erhebungsphase an der TU Freiberg fand das erste Dateninterview statt. Hier wurde von der Freiburger Forschendenseite eine Vorstellung des Forschungsvorhabens vorgenommen. Relevante Informationen wurden durch den Data Librarian dokumentiert und erste Erläuterungen zu Maßnahmen bezüglich Datenformaten und Aufbau der Datensätze vorgenommen. Darauf folgten im Projektverlauf noch weitere Dateninterviews mit dem Ziel, dass beide Parteien sich thematisch einander annähern und Teile eines Datenmanagementplans erstellen, um spezifische Aspekte der Learning-Analytics-Daten in Erfahrung zu bringen. Dabei ist zu erwähnen, dass diese internen Datenmanagementpläne vorrangig den Umgang mit den einzelnen Datensätzen zum Ausgangspunkt haben. Die weiteren gemeinsamen Gespräche bildeten eine Basis, um alle Aspekte der Datenkuratation und -dokumentation der anfallenden Learning-Analytics-Daten zu erfassen. Eine Grundlage für die sonst freien Gespräche bildete die Checkliste aus Ludwig und Enkes Leitfaden zum Forschungsdatenmanagement.²⁰ Ein weiteres Ziel der Gespräche war auch, die Daten gemeinsam in die für das Projekt zum Test vorgesehenen Forschungsdatenrepositorien einzufügen und Metadaten zu erfassen, beziehungsweise das durch den Data Librarian angefertigte Metadatenschema gemeinsam zu testen und bedarfsgerecht weiterzuentwickeln. Darüber hinaus ist zu erwähnen, dass die Interviews neben den inhaltlichen Punkten auch Einfluss auf die Beziehungsbildung besaßen. Somit sensibilisierten sich beide Seiten gegenseitig, für die in der jeweiligen Disziplin gegebenen Anforderungen. Von Seiten der Forschenden wurde dem Data Librarian ein detaillierter Einblick in die immer weiter fortschreitende Forschung gegeben, wobei im Gegenzug Aspekte des Datenmanagements gleichzeitig in die Forschung integriert wurden.

3.2 Datenentstehung kollaborativer Programmierlernszenarien

An der TU Bergakademie Freiberg wurden in den vergangenen zwei Projektjahren und fortlaufend Übungen und Vorlesungen im Bereich der Softwareentwicklung gegeben. Die Datenerhebungen für das DiP-iT-Projekt werden im Rahmen dieser Veranstaltungen durchgeführt und basieren auf den Aktivitäten von Studierenden verschiedener technischer und weniger technischer Studiengänge. Die Studierende bearbeiten in Teams mit je zwei Mitgliedern über das gesamte Semester verteilt sieben Aufgaben, die sowohl den Softwareentwurf, die Verwendung der entsprechenden Werkzeuge, als auch die konkrete

19 Wolff, Ian; Köppen, Veit; Zug, Sebastian: Ein Fallbeispiel zum Umgang mit Learning Analytics Forschungsdaten - fachspezifische Konzeptentwicklung eines Data Librarians, 2022. Online: <<https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/docId/17908>>.

20 Ludwig, Jens; Enke, Harry (Hg.): Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt, Glückstadt 2013, S. 87-100. Online: <https://publications.goettingen-research-online.de/bitstream/2/14366/1/leitfaden_DGRID.pdf>., siehe weiterhin Witt, Michael; Carlson, Jake R.: Conducting a Data Interview. Libraries Research Publications 81, 2007. Online: <http://docs.lib.purdue.edu/lib_research/81>., weiterhin siehe Jake Carlsons Artikel, aus dem ein Toolkit für Data Interviews hervorging: Carlson, Jake: Demystifying the data interview, in: Reference Services Review 40 (1), 2012, S. 7-23. Online: <<https://doi.org/10.1108/00907321211203603>>., Data Curation Profile Toolkit, Online: <<https://docs.lib.purdue.edu/dcptoolkit/>>.

Implementierung vermitteln. Die Arbeit selbst wird in GitHub-Classrooms²¹ organisiert; damit steht für jedes Team bei jeder Aufgabe eine eigene Entwicklungs- und Projektmanagementumgebung bereit, wie es auch in einem professionellen Entwicklungsvorhaben der Fall wäre. Zentrale Aufgabe von GitHub, das auf dem eigentlichen Versionsmanagementsystem git aufsetzt, ist die transparente Verwaltung der Entwicklungsschritte eines Softwareprojektes. Ein Team spezifiziert Ziele (Features/Issues), diskutiert dazu in einer Chat-ähnlichen Umgebung und legt Verantwortlichkeiten fest. Während der Umsetzung entstehen neue Versionen, die automatisch getestet und publiziert werden können. Dabei sind die Beiträge der einzelnen Entwickler*innen sowohl bei der Programmierung als auch der Organisation und Diskussion sichtbar. Sich im Nachhinein als ungeeignet herausstellende Änderungen können zudem automatisiert im Projekt entnommen oder korrigiert werden.

Für die Analyse dieser Abläufe wurden im Projekt DiP-iT die Werkzeuge github2pandas²² und github2pandas_manager²³ geschaffen, die eine skriptbasierte Aggregation anhand der Github-API ermöglichen. Pandas ist eine Datenrepräsentationsform der Programmiersprache Python. Die Lehrenden geben lediglich die Namen der Übungen an und wählen die zu extrahierenden Inhalte aus. In den Projektjahren 2020 bis 2022 wurden insgesamt 413 Repositories von 59 studentischen Teams erfasst. Der Gesamtumfang der erfassten Daten in Bezug auf das Tracking von Versionen, Feature-diskussionen oder Tests findet sich in der Dokumentation des Paketes.²⁴

3.3 Forschungsabläufe der Learning Analytics verstehen

Auf der Grundlage der Interviews zwischen den Learning Analytics Forschenden und dem Data Librarian erfolgte eine Analyse des Learning Analytics Forschungsablaufs. Dies dient dem Data Librarian dazu, die ihm fachfremden Spezifika der Learning Analytics zu verstehen und einen ganzheitlichen Blick auf die Forschung zu erhalten, um sie mit den Rahmenbedingungen des Forschungsdatenmanagements in Einklang zu bringen. Zum einen werden die Abläufe der im Projekt anfallenden Forschung zu kollaborativen Programmierlernszenarien betrachtet, zum anderen mit denen aus der Literatur verglichen.

21 Webseite von Github-Classrooms einem leichtgewichtigen Learning-Management-System für git Repositorien auf der Basis von Github, Online: <<https://classroom.github.com/>>, Stand: 06.12.2022.

22 Projektwebseite, Online: <<https://github.com/TUBAF-IFI-DiPiT/github2pandas>>, Stand: 06.12.2022.

23 Projektwebseite, Online: <https://github.com/TUBAF-IFI-DiPiT/github2pandas_manager>, Stand: 06.12.2022.

24 Übersicht der extraktionsfähigen Daten und der realisierten Datenstruktur, Online: <<https://github.com/TUBAF-IFI-DiPiT/github2pandas/wiki>>, Stand: 06.12.2022.

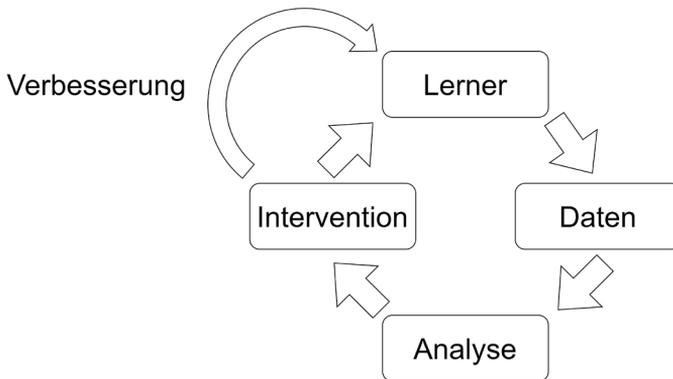


Abb. 1: Learning Analytics Zyklus nach Clow (2012) erweitert um die Phase der Verbesserung von Lernszenarien

Nach Clow sowie Khalil und Ebner teilt sich der Learning-Analytics-Forschungsablauf in vier Phasen.²⁵ Der in Abbildung 1 aufgezeigte Learning-Analytics-Zyklus startet mit den Lernenden. Dabei kann es sich um Studierende handeln, die einen Kurs an einer Universität belegen, wie in unserem Fall das Erlernen von Programmiersprachen in einer GitHub-Umgebung. In der zweiten Phase werden Daten aus den Learning-Management-Systemen generiert, oder wie in unserem Fall die Logdaten zu den verschiedenen Schritten des Programmierprozesses, die dann mit demographischen Informationen der Lernenden kombiniert werden. Die dritte Phase besteht in der Verarbeitung dieser Daten unter der Verwendung von Metriken oder Analysen, die einen Einblick in den Lernprozess geben. Dazu gehören Visualisierungen und Dashboards mit denen die Leistungen der Studierenden dargestellt werden können und die somit Vergleiche von Ergebnismessungen beispielsweise früherer Kohorten ermöglichen. Diese Phase ist der Kern der meisten Learning-Analytics-Projekte. Der Zyklus ist jedoch erst dann abgeschlossen, wenn die Daten und Messungen dazu verwendet werden, eine oder mehrere Maßnahmen zu ergreifen, die eine gewisse Wirkung auf die Lernenden haben. Dies könnte ein Dashboard für Lernende sein, das ihnen ermöglicht, ihre Aktivitäten mit denen ihrer Mitstudierenden oder früherer Kohorten zu vergleichen. Weiterhin kann eine Lehrkraft wichtige Informationen über ein Dashboard erhalten und Interventionen für beispielsweise abbruchgefährdete Studierende einleiten. Nach Leitner sowie Romero folgt darauf eine Phase der Verbesserung bzw. Anpassung der Lernmaterialien und -szenarien, was insbesondere für das Forschungsdatenmanagement wichtig ist. Denn um die Nachvollziehbarkeit der letztendlichen Datensätze zu gewährleisten, müssen die am Lehr-/Lernedesign vorgenommenen Veränderungen detailliert dokumentiert werden. Dies umfasst beispielsweise Veränderungen an Aufgabenstellungen innerhalb eines Lernszenarios.²⁶

²⁵ Clow, Dough: The Learning Analytics Cycle. Closing the loop effectively, in: Dawson, Shawn; Haythornthwaite, Caroline (Hg.): Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, New York 2012, S. 135. Online: <<https://doi.org/10.1145/2330601.2330636>>., Khalil, Mohammad; Ebner, Martin: Learning Analytics: Principles and Constraints, in: Carliner, S.; Fulford, C.; Ostashewski, N. (Hg.): Proceedings of EdMedia 2015. World Conference on Educational Media and Technology, S. 1790.

²⁶ Leitner, Philipp; Khalil, Mohammad; Ebner, Martin: Learning Analytics in Higher Education. A Literature Review, in: Peña-Ayala, Alejandro (Hg.): Learning Analytics. Fundaments, Applications, and Trends, Cham 2017, S. 1–23. Online:

3.4 Datentransfermodell

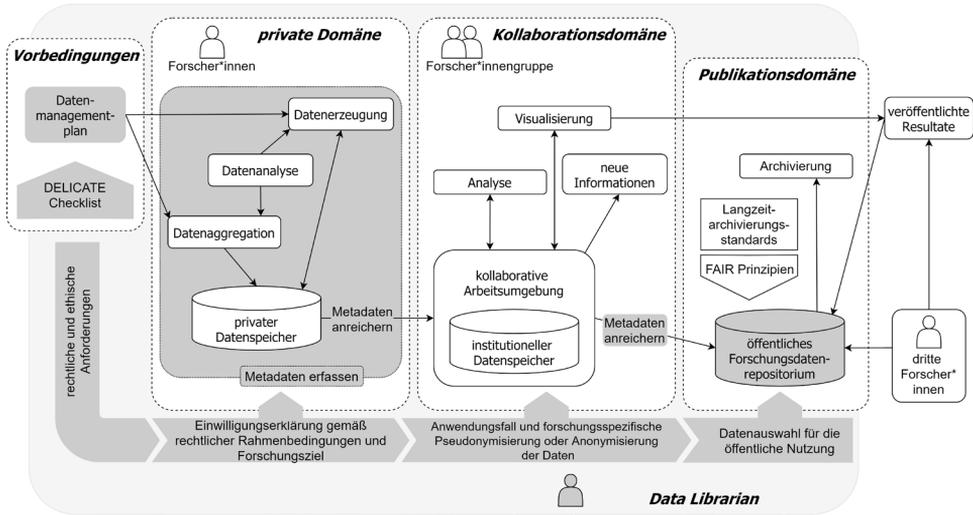


Abb. 2: Datentransfermodell adaptiert nach Treloar (2019)

Unter Einbezug der Erkenntnisse aus der Analyse des Learning-Analytics-Forschungsablaufs ergibt sich eine spezifische Handhabung der Daten entlang eines Transfermodells, begleitet durch den Data Librarian.²⁷ Da besonders rechtliche Aspekte bezüglich des Datenschutzes beachtet werden müssen, hat dies von Anfang an im Forschungsdesign Beachtung gefunden. Zu Beginn des Projekts wurden Einwilligungserklärungen entlang der für die Learning Analytics entwickelten DELICATE Checklist erstellt, um möglichst hohe Transparenz zur Datenerhebung für die Studienteilnehmer und die spätere Speicherung der Daten zu gewährleisten.²⁸ Gleichzeitig wurde abgewogen, welche personenbezogenen Daten für die Erhebung notwendig sind. In der privaten Domäne (Forscher*in) werden die Daten dann erzeugt und aggregiert und in Absprache zwischen Forschenden und Data Librarian ein Metadatenmodell für die Daten entwickelt – hier findet das in 3.5 vorgestellte Metadaten- und Datenmodell Anwendung. Daraufhin erfolgten Test zur Speicherung der Daten und deren Beschreibung anhand des Metadatenmodells in Repositorien zur projektinternen Datenweitergabe in die Kollaborationsdomäne (Gruppe der Forschenden). Bereits hier erfolgt eine Pseudonymisierung der Daten durch die Forschenden, indem identifizierende Merkmale von weiteren getrennt werden. Der vollständige Datensatz wird in der privaten Domäne vorgehalten und beide Datensätze durch einen eindeutigen

<<https://doi.org/10.1007/978-3-319-52977-6>>, Romero, Cristobal; Ventura, Sebastian: Data mining in education. In: WIREs Data Mining Knowledge Discovery 3 (1), 2012, S. 12–27. Online: <<https://doi.org/10.1002/widm.1075>>.

27 Treloar, Andrew; Klump, Jens: Updating the Data Curation Continuum, in: International Journal of Digital Curation 14 (1), 2019, S. 87–101. Online: <<https://doi.org/10.2218/ijdc.v14i1.643>>.

28 Drachsler, Hendrik; Greller, Wolfgang: Privacy and analytics, in: Gašević, Dragan; Lynch, Grace; Dawson, Shane u. a. (Hg.): Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16, New York, New York, USA 2016, S. 89–98. Online: <<https://doi.org/10.1145/2883851.2883893>>.

Zuordnungsschlüssel miteinander verknüpfbar gemacht. Somit ist die Auswertung aller erhobenen Merkmale im Projektkontext möglich, ohne die Identitäten von Studienteilnehmern offen zu legen. Das Veröffentlichen der Daten wird dann in der Publikationsdomäne erfolgen unter Beachtung von Langzeitarchivierungsstandards und der FAIR-Prinzipien.²⁹ Für den Veröffentlichungsschritt werden dann externe Forschungsdatenrepositorien genutzt. Dafür werden die Daten nochmals selektiert und anonymisiert und somit alle identifizierenden Merkmale und personenbezogene Angaben entfernt.

3.5 Metadatenmodell und Datenmodell

Im Austausch zwischen Data Librarian und Learning-Analytics-Wissenschaftler*innen ergaben sich fachspezifische Merkmale der Learning Analytics, die in einem Metadatenchema repräsentiert werden, das zum Datenaustausch innerhalb der Repositorien zum Einsatz kommt. Die Ergebnisse der Interviews wurden zudem mit einem Literaturreview kontrastiert, was in einem Datenmodell (Abbildung 3) resultiert, das alle erforderlichen Entitäten des Learning-Analytics-Forschungsablaufs miteinander in Verbindung setzt.³⁰ Wenn möglich, werden die Metadaten entlang der FAIR Prinzipien entworfen.³¹ Eine erste Arbeitsversion des Metadatenchemas wurde zusätzlich auf ihre Anwendbarkeit in Repositorien getestet.³²

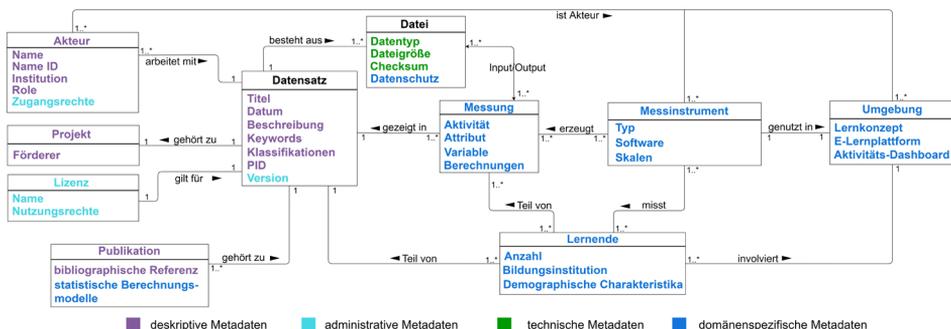


Abb. 3: Objektmodell mit domänenspezifischen Merkmalen der Learning Analytics

- 29 Weber, Andreas; Piesche, Claudia: 4.2 Datenspeicherung, -kuration und Langzeitverfügbarkeit, in: Putnings, Markus; Neuroth, Heike; Neumann, Janna (Hg.): Praxishandbuch Forschungsdatenmanagement, 2021, S. 327–356. Online: <<https://doi.org/10.1515/9783110657807>>., Wolff, Ian; Broneske, David; Köppen, Veit: FAIR Research Data Management for Learning Analytics, in: Lingnau, Andreas (ed.): Proceedings of DELFI Workshop 2021, Bottrop 2021, pp. 158–163. Online: <<https://repositorium.hs-ruhrwest.de/frontdoor/index/index/docId/733>>.
- 30 Wolff, Ian; Broneske, David; Köppen, Veit: First metadata schema for learning analytics research data management, in: o-bib. Das offene Bibliotheksjournal 8 (4), 2021, Online: <<https://doi.org/10.5282/o-bib/5735>>. Für das Datenmodell siehe weiterhin Wolff, Ian; Broneske, David; Köppen, Veit: Towards a Learning Analytics Metadata Model, in: Friend Wise, Alyssa; Martinez-Maldonado, Roberto; Hilliger, Isabel: Companion Proceedings of the 12th International Learning Analytics and Knowledge Conference (LAK'22), 2022, S. 51–53. Online: <https://www.solaresearch.org/wp-content/uploads/2022/03/LAK22_CompanionProceedings.pdf>.
- 31 Wilkinson, Mark; Dumontier, Michel; Aalbersberg, IJsbrand Jan u.a.: The FAIR Guiding Principles for scientific data management and stewardship, in: Scientific data 3, 2016. Online: <<https://doi.org/10.1038/sdata.2016.18>>.
- 32 Wolff, Ian; Broneske, David; Köppen, Veit: LAMM - Learning Analytics Metadata Model, 2022, Online: <<https://doi.org/10.24352/UB.OVGU-2022-013>>.

Im Folgenden wird der Aufbau des Metadatenmodells und seinen verschiedenen Entitäten in Abbildung 4 veranschaulicht. Der erste verwendete Metadatentyp sind *deskriptive Metadaten*³³, die den Bedürfnissen der Forschenden gerecht werden, indem sie die Auffindbarkeit der Daten erhöhen und den Datensatz durch die Angabe detaillierter Informationen beschreiben und identifizierbar machen. Dabei handelt es sich um elementare Angaben wie *Titel*, *Keywords*, *Beschreibung* und beteiligte *Personen* sowie den *Themenbereich* des Datensatzes. Darüber hinaus ist die Zuordnung der Daten zu einer etablierten Klassifikation wichtig. Für Datensätze der Learning Analytics bietet sich die Verwendung der Dewey-Dezimalklassifikation (DDC) an.³⁴

Forschende benötigen zudem einen einfachen Zugang zu ihren Daten, was mit *administrativen Metadaten* sichergestellt wird. Dies ist für Learning-Analytics-Daten aus Gründen des Datenschutzes und der Ethik von herausgehobener Bedeutung.³⁵ Die Zugänglichkeit wird vorrangig durch die Vergabe geeigneter *Lizenzen* sichergestellt, durch die der Umfang, in dem die Forschungsdaten wiederverwendet und genutzt werden können, festgelegt wird.

Ein Datensatz kann auch aus mehreren Dateien mit Dateiattributen wie *Name*, *Typ* und *Dateigröße* bestehen und mit einer *Checksum* oder einem *persistenten Identifikator* angereichert werden, was ebenfalls die Auffindbarkeit gewährleistet. Wir kategorisieren solche Entitäten als *technische Metadaten*.

Der letzte Typ sind die *domänenspezifischen Metadaten*, die nicht zu den klassischen Metadaten gehören, aber für die Wiederverwendbarkeit und Interoperabilität eines Datensatzes wichtig sind. Sie definieren die Merkmale der Learning Analytics entlang des Lebenszyklus der Forschungsdaten. Hier werden der Lernende und die für die Analyse der Aktivitätsmessung verwendete *Umgebung* sowie das *Messinstrument* definiert, ebenso wie die für die Analyse des Lernprozesses verwendeten *Berechnungen*. Diese Entitäten ermöglichen es weiteren Learning Analytics Forschenden, die Datenherkunft im Detail zu verstehen, um das Studiendesign zu adaptieren oder die Daten für Sekundäranalysen zu verwenden.

33 Siehe zu den Metadatenkategorien Baca, Murtha: Introduction to metadata, Los Angeles 2008. Online: <<https://d2aohyo3d3idm.cloudfront.net/publications/virtuallibrary/0892368969.pdf>>.

34 Dewey-Dezimalklassifikation und Register: DDC 22 (deutsche Ausgabe), begründet von Melvil Dewey, Joan S. Mitchell u.a. (Hg.), 2 Bände, München 2005.

35 Pardo, Abelardo; Siemens, George: Ethical and privacy principles for learning analytics, in: British Journal of Educational Technology 45 (3), 2014, S. 438–450. Online: <<https://doi.org/10.1111/bjet.12152>>.

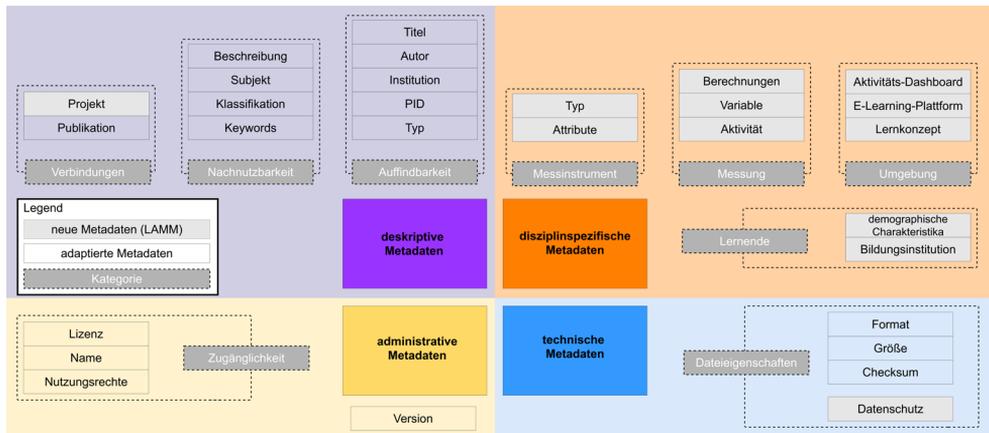


Abb.4: Komponenten des Learning Analytics Metadata Models (LAMM)

Kein bestehendes Metadatenchema führt alle beschriebenen Informationen auf, weshalb für das Learning Analytics Metadata Model (LAMM) auf verschiedene etablierte Schemata zurückgegriffen wird. Zu diesen gehören der DublinCore³⁶ und DataCite³⁷, welches Metadatenfelder für Zitate und allgemeine beschreibende Metadaten bietet, das RADAR Metadatenchema³⁸, was insbesondere für die Beschreibung von Software und Data Processing Schritten Verwendung findet, sowie PREMIS³⁹, das vorrangig für technische Aspekte zur Beschreibung der Datenpakete zum Einsatz kommt. Abbildung 4 zeigt alle Metadatenfelder, die im Metadatenmodell Anwendung finden und welche aus bereits bestehenden Modellen übernommen wurden. Unser Hauptbeitrag besteht in der Einführung neuer Felder, die vor allem für die Learning Analytics spezifisch sind. Diese Felder und ihre Semantik sind in Tabelle 1 zusammengefasst.

Metadatenelement	Semantik
Lernkonzept	Beschreibung der Abläufe und Spezifika des Unterrichtsszenario in das die Lernenden eingebunden ist.
Lernplattform	Beschreibung der Lernplattformen, mit denen die Aktivitäten Lernenden zur Analyse des Lernprozesses gemessen werden.
Lernender	Charakterisierung der Gruppe an Lernenden.

36 Dublin Core, Online: <<https://www.dublincore.org/schemas/>>, Stand: 06.12.2022.

37 DataCite Metadata Schema, Online: <<https://schema.datacite.org/>>, Stand: 06.12.2022.

38 RADAR Metadatenchema, Online: <<https://radar.products.fiz-karlsruhe.de/de/radarfeatures/radar-metadatenchema>>, Stand: 06.12.2022.

39 PREMIS, Online: <<https://www.loc.gov/standards/premis/v3/index.html>>, Stand: 06.12.2022.

Demografische Charakteristika der Lernenden	Information zu den durch Fragebögen erschlossenen Charakteristika der Lernenden für Lernanalysen.
Bildungseinrichtung	Angaben in welcher Ausbildungsform (Schule, Universität) die Daten erhoben wurden.
Aktivität	Beschreibt die vom Lernenden ausgeführte Aktivität.
Attribute	Definition der einzelnen gemessenen Attribute.
Variable	Ausgewählte Attribute die für späterer Berechnungen genutzt werden.
Berechnungen	Angaben zu verwendeten statistischen Verfahren für die Lernprozessmessungen.
Datenschutz	Informationen zu datenschutzrechtlichen Maßnahmen, die auf die Daten angewendet wurden.

Tab. 1: neu eingeführte Metadatenelemente und deren Semantik

Im direkten Einsatz bezugnehmend auf die in 3.2. vorgestellte Datenentstehung von kollaborativen Programmierlernszenarien zeigt sich für die Erhebungen folgendes Bild einer Dateninstanz in der kollaborativen Domäne zum projektinternen Datenaustausch: Das Beispiel in Abbildung 5 stellt die Metadatenfelder und Werte dar. Es werden zwei Instanzen aufgesetzt, da zwei verschiedene Messinstrumente für die Datenerhebungen verwendet werden, von denen jedes spezifische Merkmale aufweist. Im unteren Teil wird die Fragebogenerhebung der teilnehmenden Studierenden aufgezeigt. Die Art der Bildungseinrichtung, in unserem Falle der Hochschulbereich, und die Lernenden werden hinlänglich der erhobenen Charakteristika durch entsprechende Metadatenpunkte beschrieben. Die Instanz enthält zwei Datenpakete: zum ersten die Fragebögen als PDF-Dateien, um Fragen und Skalen nachvollziehen zu können, und zum zweiten die Ergebnisse aus den Fragebögen. Zweitere werden dann noch durch Metadaten beschrieben, indem die gemessenen Attribute, die Anzahl der gemessenen Items pro Attribut und der verwendete Skalen-Typ für die Messung angegeben werden. Die Instanz zu den Fragebögen ist als komplementär zu den Logdaten aus github2pandas zu betrachten. In der Instanz finden sich Angaben zur Umgebung, zur verwendeten Lernplattform und zum Lernkonzept, in das die Lernenden eingebettet sind. Die Lernenden werden neben den Angaben zur Anzahl und Bildungseinrichtung durch die aus den Fragebögen gemessenen demographischen Charakteristika näher beschrieben und das Datenset verlinkt. Als Datenpakete befinden sich in der Instanz die verwendeten Arbeitsblätter, das Softwarepaket und die Ergebnisse der Erhebung, die als gemessene Attribute näher durch ihren Namen und die Anzahl der gemessenen Einheiten charakterisiert werden. Kern der Learning Analytics sind Lernprozessmessungen. Die gemessenen Attribute werden dann als Variablen für Berechnungen in Publikationen genutzt.

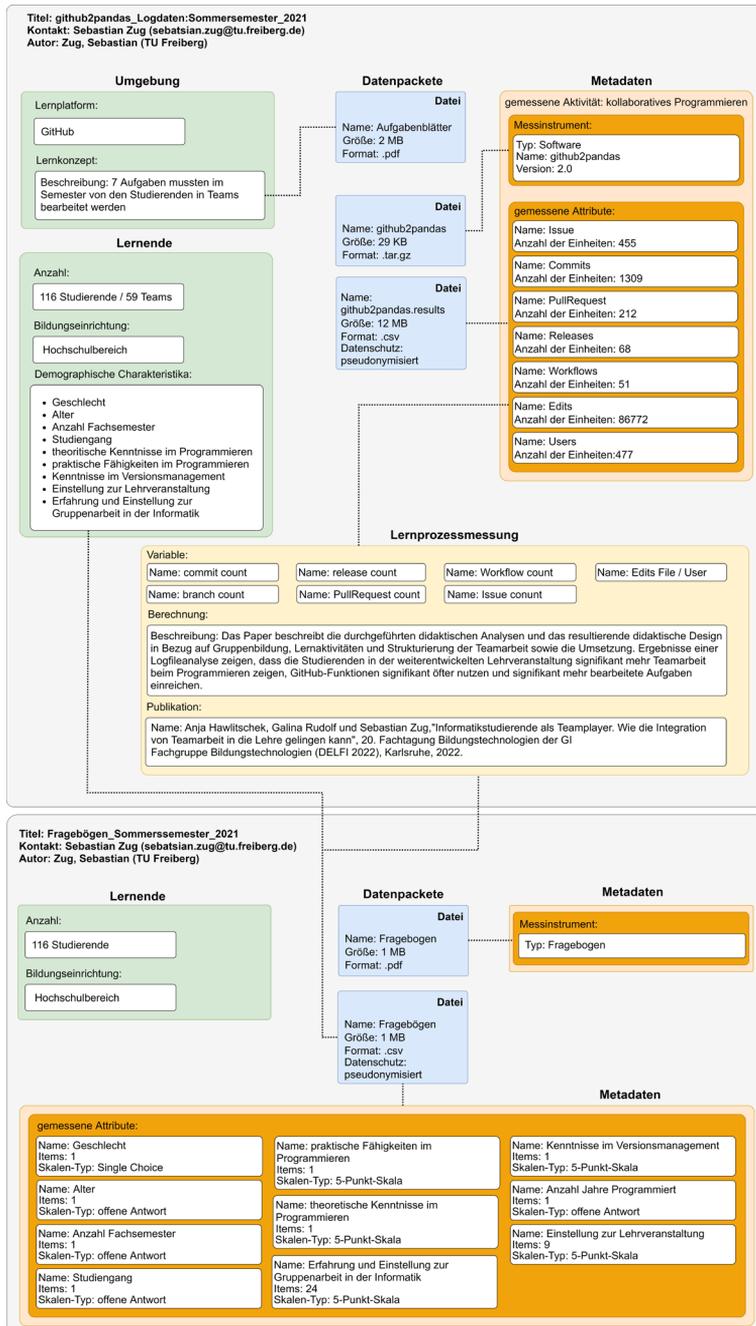


Abb. 5: Beispielinstanz eines Datensets in den Learning Analytics

4. Zusammenfassung

Prozesse im Forschungsdatenmanagement bringen für Wissenschaftler*innen neue Herausforderungen mit sich. Im Projekt DiP-iT werden diese durch den eingesetzten Data Librarian bearbeitet und in Zusammenarbeit mit den Wissenschaftler*innen fachspezifische Lösungen entwickelt. Dabei wurden durch immer wiederkehrende Treffen in Form eines Datainterviews die Forschungsprozesse der Learning Analytics untersucht, um daran anknüpfend insbesondere in den Bereichen der Auffindbarkeit und der Beschreibung der Daten, Organisation und Dokumentation sowie der Datenkuration und Bewahrung der Daten zu unterstützen. Konkrete Maßnahmen umfassten das Erstellen eines Transfermodells zum Umgang mit sensiblen Daten für den projektinternen Datenaustausch, sowie der späteren Publikation, unter Beachtung datenschutzrechtlicher Rahmenbedingungen. Zur Dokumentation der Daten wurde ein disziplinspezifisches Metadatenschema entwickelt, das den Forschungsabläufen der Learning Analytics Rechnung trägt und die spätere Publikation der Daten auch mittels bereits etablierter Metadatenschemata ermöglicht.

Die entwickelten Ansätze zum Datentransfer unter Berücksichtigung datenschutzrechtlicher Aspekte eignen sich auch zur Anwendung in anderen Disziplinen, die ebenso mit personenbezogenen Daten umgehen müssen. Der Aufbau der Datensets besitzt ebenfalls einen generischen Ansatz, indem alle notwendigen Instrumente und Methoden zur Datenerhebung den letztendlichen Ergebnisdaten zugeordnet werden und untereinander verlinkt sind. Diese können auch entlang bereits bestehender Metadatenschemata beschrieben werden, die einen generischen Ansatz besitzen, Erhebungsprozesse standardisiert zu erfassen.

Acknowledgements

Finanzierung: Diese Arbeit wurde im Rahmen des Projektes DiP-iT in der Förderlinie zur Digitalen Hochschulbildung vom deutschen Bundesministerium für Bildung und Forschung unterstützt [Förderkennzeichen 16DHB 3007 und 16DHB 3008].

Literaturverzeichnis

- Adams, Margaret O.: The Origins and Early Years of IASSIST, in: IASSIST Quarterly 30 (3), 2007, S. 6. Online: <<https://doi.org/10.29173/iq108>>.
- Baca, Murtha: Introduction to metadata, Los Angeles 2008. Online: <<https://d2aohiyo3d3idm.cloudfront.net/publications/virtuallibrary/0892368969.pdf>>.
- Biernacka, Katarzyna; Pinkwart, Niels: Opportunities for Adopting Open Research Data in Learning Analytics, in: Azevedo, Ana; Azevedo, José Manuel; Onohuome Uhomobhi, James u.a. (Hg.): Advancing the power of learning analytics and big data in education, Hershey 2021, S. 29–60. Online: <<https://doi.org/10.4018/978-1-7998-7103-3>>.

- Carlson, Jake: Demystifying the data interview, in: Reference Services Review 40 (1), 2012, S. 7-23. Online: <<https://doi.org/10.1108/00907321211203603>>.
- Christensen-Dalsgaard, Brite, Ten Recommendations for Libraries to Get Started with Research Data Management, LIBER, 2012. Online: <<https://libereurope.eu/wp-content/uploads/2020/11/Theresearch-data-group-2012-v7-final.pdf>>.
- Clow, Dough: The Learning Analytics Cycle. Closing the loop effectively, in: Dawson, Shawn; Haythornthwaite, Caroline (Hg.): Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, New York 2012, S. 135. Online: <<https://doi.org/10.1145/2330601.2330636>>.
- Data librarianship. responding to research innovation, in: Rice, Robin; Southall, John (Hg.): The Data Librarian's Handbook, 2018, S. 1-18. Online: <<https://doi.org/10.29085/9781783301836>>.
- Dewey-Dezimalklassifikation und Register: DDC 22 (deutsche Ausgabe), begründet von Melvil Dewey, Joan S. Mitchell u.a. (Hg.), 2 Bände, München 2005.
- Drachsler, Hendrik; Greller, Wolfgang: Privacy and analytics, in: Gašević, Dragan; Lynch, Grace; Dawson, Shane u. a. (Hg.): Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16, New York, New York, USA 2016, S. 89-98. Online: <<https://doi.org/10.1145/2883851.2883893>>.
- Dublin Core, Online: <<https://www.dublincore.org/schemas/>>, Stand: 06.12.2022.
- Eclevia, Marian Ramos; La Torre Fredeluces, John Christopher; Eclevia, Carlos; Saguibo Maestro, Roselle: What Makes a Data Librarian? An Analysis of Job Descriptions and Specifications for Data Librarian, in: Qualitative and Quantitative Methods in Libraries 8 (3), 2019, S. 273-290. Online: <<http://qqml-journal.net/index.php/qqml/article/view/541>>.
- Fühles-Ubach, Simone; Schaer, Philipp; Lepsky, Klaus et al.: Data Librarian – ein neuer Studienschwerpunkt für wissenschaftliche Bibliotheken und Forschungseinrichtungen. Unter Mitarbeit von Humboldt-Universität zu Berlin, in: Bibliothek Forschung und Praxis 43 (2), S. 255-261. Online: <<https://doi.org/10.1515/bfp-2019-2053>>.
- Fühles-Ubach, Simone: Bibliothekare und Data Librarians – neue Profile für das bibliothekarische Fachpersonal der Zukunft, in: o-bib. Das offene Bibliotheksjournal 5 (4), 2018, S. 7-17. Online: <<https://doi.org/10.5282/o-bib/2018H457-17>>.
- Gold, Anna: Cyberinfrastructure, Data, and Libraries. Part 1. A Cyberinfrastructure Primer for Librarians, in: D-Lib Magazine 13 (9/10), 2007. Online: <<http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>>.

- Gold, Anna: Cyberinfrastructure, Data, and Libraries. Part 2. Libraries and the Data Challenge. Roles and Actions for Libraries, in: D-Lib Magazine 13 (9/10), 2007. Online: <<http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>>.
- Grandl, Maria; Taraghi, Begnam; Ebner, Martin et al.: Learning Analytics, in: Wilbers, Karl; Hohenstein, Andreas (Hg.): Handbuch E-Learning. Expertenwissen aus Wissenschaft und Praxis. Strategien, Instrumente, Fallstudien. Köln 2017, S. 1-16.
- Greller, Wolfgang; Drachsler, Hendrik: Translating Learning into Numbers. A Generic Framework for Learning Analytics. Educational Technology & Society 15 (3), 2012, S. 48f. Online: <<https://www.jstor.org/stable/jeductechsoci.15.3.42>>.
- Heise, Christian: Von Open Access zu Open Science. Zum Wandel digitaler Kulturen der wissenschaftlichen Kommunikation, Lüneburg 2018. Online: <<http://library.oapen.org/handle/20.500.12657/37554>>.
- Khalil, Mohammad; Ebner, Martin: Learning Analytics: Principles and Constraints, in: Carliner, S.; Fulford, C.; Ostashewski, N. (Hg.): Proceedings of EdMedia 2015. World Conference on Educational Media and Technology, S. 1790.
- Khan, Hammad Rauf; Du, Yunfei: What is a data librarian?. A content analysis of job advertisements for data librarians in the United States academic libraries, in: IFLA World Library and Information Congress 2018, Kuala Lumpur 2018, S. 1-9. Online: <<https://digital.library.unt.edu/ark:/67531/metadc1225772/>>.
- Kitchin, Roy: The data revolution : big data, open data, data infrastructures & their consequences, Los Angeles 2014. Online: <<https://dx.doi.org/10.4135/9781473909472>>.
- Koltay, Tibor: Are you ready? Tasks and roles for academic libraries in supporting Research 2.0, in: New Library World (1/2), 2016, S. 94-104. Online: <<https://doi.org/10.1108/NLW-09-2015-0062>>.
- Koltay, Tibor: Data literacy for researchers and data librarians, in: Journal of Librarianship and Information Science 49 (1), 2017, S. 3-14. Online: <<https://doi.org/10.1177/%2F0961000615616450>>.
- Leitner, Philipp; Khalil, Mohammad; Ebner, Martin: Learning Analytics in Higher Education. A Literature Review, in: Peña-Ayala, Alejandro (Hg.): Learning Analytics. Fundamentals, Applications, and Trends, Cham 2017, S. 1-23. Online: <<https://doi.org/10.1007/978-3-319-52977-6>>.

- LERU Roadmap for Research Data, League of European Research Universities Research Data Working Group, Leuven 2013. Online: <<https://www.leru.org/files/LERU-Roadmap-for-Research-Data-Full-paper.pdf>>.
- Ludwig, Jens; Enke, Harry (Hg.): Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt, Glückstadt 2013, S. 87-100. Online: <https://publications.goettingen-research-online.de/bitstream/2/14366/1/leitfaden_DGRID.pdf>.
- Maatta, Stephanie L., Placement & Salaries 2013. The Emerging Databrarian, in: Library Journal 138 (17), S. 26-43. Online: <<https://www.libraryjournal.com/story/the-emerging-databrarian>>.
- Pardo, Abelardo; Siemens, George: Ethical and privacy principles for learning analytics, in: British Journal of Educational Technology 45 (3), 2014, S. 438-450. Online: <<https://doi.org/10.1111/bjet.12152>>.
- Rezniczek, Alina; Blumesberger, Susanne; Bargmann, Monika u. a.: Der Zertifikatskurs „Data Librarian“ und seine erstmalige Durchführung, in: Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare 72 (2), 2019, S. 274-283. Online: <<https://doi.org/10.31263/voebm.v72i2.3176>>.
- Rice, Robin: Supporting Research Data Management and Open Science in Academic Libraries: a Data Librarian's View, in: Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare 72 (2), 2019, S. 263-273. <<https://doi.org/10.31263/voebm.v72i2.3303>>.
- Romero, Cristobal; Ventura, Sebastian: Data mining in education. In: WIREs Data Mining Knowledge Discovery 3 (1), 2012, S. 12-27. Online: <<https://doi.org/10.1002/widm.1075>>.
- Scheffel, Maren; Drachsler, Hendrik; Stoyanov, Slavi u.a.: Quality Indicators for Learning Analytics. Educational Technology & Society, 17 (4), 2014. Online: <<https://www.jstor.org/stable/jeductechsoci.17.4.117>>.
- Semeler, Alexandre Ribas; Pinto, Adilson Luiz; Rozados, Helen Beatriz Frota: Data science in data librarianship: Core competencies of a data librarian, in: Journal of Librarianship and Information Science 51 (3), 2019. Online: <<https://doi.org/10.1177/0961000617742465>>.
- Tappenbeck, Inka; Meinhardt, Haike: MALIS Reloaded. Der berufsbegleitende Masterstudiengang „Bibliotheks- und Informationswissenschaft“ der TH Köln präsentiert sich mit einem neuen Curriculum, in: o-bib. Das Offene Bibliotheksjournal 8 (2), 2021, S. 1-9. Online: <<https://doi.org/10.5282/o-bib/5708>>.

- Thompson, Kristi Anne; Kellam; Lynda: Introduction, in: Thompson, Kristi Anne; Kellam; Lynda (Hg.) Databrarianship. The Academic Data Librarian in Theory and Practice, 2016, S. 3-4. Online: <<https://scholar.uwindsor.ca/leddylibrarypub/47>>.
- Treloar, Andrew; Klump, Jens: Updating the Data Curation Continuum, in: International Journal of Digital Curation 14 (1), 2019, S. 87-101. Online: <<https://doi.org/10.2218/ijdc.v14i1.643>>.
- Übersicht der extraktionsfähigen Daten und der realisierten Datenstruktur, Online: <<https://github.com/TUBAF-IFI-DiPiT/github2pandas/wiki>>, Stand: 06.12.2022.
- Weber, Andreas; Piesche, Claudia: 4.2 Datenspeicherung, -kuration und Langzeitverfügbarkeit, in: Putnings, Markus; Neuroth, Heike; Neumann, Janna (Hg.): Praxishandbuch Forschungsdatenmanagement, 2021, S. 327-356. Online: <<https://doi.org/10.1515/9783110657807>>.
- Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, Ijsbrand Jan u.a.: The FAIR Guiding Principles for scientific data management and stewardship, in: Scientific data 3, 2016. Online: <<https://doi.org/10.1038/sdata.2016.18>>.
- Witt, Michael; Carlson, Jake R.: Conducting a Data Interview. Libraries Research Publications 81, 2007. Online: <http://docs.lib.purdue.edu/lib_research/81>.
- Wolff, Ian; Köppen, Veit; Zug, Sebastian: Ein Fallbeispiel zum Umgang mit Learning Analytics Forschungsdaten - fachspezifische Konzeptentwicklung eines Data Librarians, 2022. Online: <<https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/docId/17908>>.
- Wolff, Ian; Broneske, David; Köppen, Veit: FAIR Research Data Management for Learning Analytics, in: Lingnau, Andreas (ed.): Proceedings of DELFI Workshop 2021, Bottrop 2021, pp. 158-163. Online: <<https://repositorium.hs-ruhrwest.de/frontdoor/index/index/docId/733>>.
- Wolff, Ian; Broneske, David; Köppen, Veit: first metadata schema for learning analytics research data management, in: o-bib. Das offene Bibliotheksjournal 8 (4), 2021, Online: <<https://doi.org/10.5282/o-bib/5735>>.
- Wolff, Ian; Broneske, David; Köppen, Veit: LAMM – Learning Analytics Metadata Model, 2022, Online: <<https://doi.org/10.24352/UB.OVGU-2022-013>>.
- Wolff, Ian; Broneske, David; Köppen, Veit: Towards a Learning Analytics Metadata Model, in: Friend Wise, Alyssa; Martinez-Maldonado, Roberto; Hilliger, Isabel: Companion Proceedings of the 12th International Learning Analytics and Knowledge Conference (LAK'22),

2022, S. 51-53. Online: <https://www.solaresearch.org/wp-content/uploads/2022/03/LAK22_CompanionProceedings.pdf>.

- Xia, Jingfeng; Wang, Minglu: Competencies and Responsibilities of Social Science Data Librarians: An Analysis of Job Descriptions, in: *College & Research Libraries* 75 (3), 2014, S. 362–388. Online: <<https://doi.org/10.5860/crl13-435>>.