
Das Data Science Center an der Universität Bremen: Interdisziplinärer Knotenpunkt und Service-Infrastruktur für die datenintensive Forschung

Lena Steinmann*, Heike Thöricht, Sandra Zänkert, Rolf Drechsler

Data Science Center, Universität Bremen;

*Korrespondierende Autorin: lena.steinmann@uni-bremen.de

Das Data Science Center (DSC) der Universität Bremen ist ein interdisziplinäres Institut und dient als Knotenpunkt für die datenintensive Forschung. Mit seinen Aktivitäten bricht das DSC disziplinäre Silos auf und stärkt die kooperative Forschung, steigert die Datenkompetenzen von Forschenden und ermöglicht eine wertschöpfende Nutzung von Daten in allen Forschungsbereichen. Es bietet umfassende Angebote und Services für datenintensiv Forschende wie Trainings, Beratungsangebote, eine IT-Infrastruktur und finanzielle Unterstützung. Das DSC kann als institutionelles Best-Practice für andere Standorte dienen, um eine strukturelle Zusammenführung und Zentralisierung von Forschungsdatenmanagement (FDM) und Data Science umzusetzen und so die Effizienz im Forschungsprozess zu steigern.

1 Die Bedeutung eines Data Science Centers

Der Wissenschaftsrat hat bereits in einem Positionspapier im Jahr 2020 die Errichtung von Data Science Centern als interdisziplinäre Forschungsstrukturen, die auch Möglichkeiten zur Weiterbildung und Kompetenzentwicklung bieten, empfohlen (Wissenschaftsrat 2020). Der Hintergrund dieser Empfehlung ist, dass die datenintensive Forschung mittlerweile Einzug in alle Disziplinen gehalten hat und das Wissenschaftssystem auf die einhergehenden Herausforderungen wie der Schnellebigkeit von Datenangeboten, Hard- und Software-Entwicklungen sowie dem immer schnelleren Aufkommen neuer Methoden stellen muss. Um auf diese Entwicklungen angemessen zu reagieren, sind neue Strukturen sowie Beratungs- und Lernangebote erforderlich. Data Science Center können dabei eine wichtige Rolle einnehmen, indem sie geeignete Rahmenbedingungen für datenintensiv Forschende schaffen und zur strategischen Positionierung der Hochschulen beitragen.

Publiziert in: Vincent Heuveline, Nina Bisheh und Philipp Kling (Hg.): E-Science-Tage 2023. Empower Your Research – Preserve Your Data. Heidelberg: heiBOOKS, 2023. DOI: <https://doi.org/10.11588/heibooks.1288.c18083> (CC BY-SA 4.0)

Die Universität hat bereits 2019 mit Unterstützung des Landes Bremen das Data Science Center (DSC@UB; im Folgenden durch DSC abgekürzt) als interdisziplinäres Institut und Knotenpunkt für die datenintensive Forschung gegründet. Die übergeordneten Ziele des DSC sind die wertschöpfende Nutzung von Daten in allen Forschungsbereichen nachhaltig zu ermöglichen und den kulturellen Wandel im Sinne einer „FAIR Data-Kultur“¹ voranzutreiben. Dazu bündelt das DSC die Kompetenzen von Forschenden unterschiedlicher Wissenschaftsbereiche in seinem Mitgliedernetzwerk. So werden die Vernetzung und der Transfer von Datenkompetenzen über fachliche Grenzen hinweg gefördert, disziplinäre Silos aufgebrochen und neue Wege für die kooperative Forschung geschaffen. Darüber hinaus stärkt das DSC durch gezielte Trainings die Datenkompetenzen von Forschenden und bietet umfassende Forschungsdatenmanagement (FDM)- und Data-Science-Services für datenintensiv Forschende, die in Abschnitt 2 näher erläutert werden. Durch die enge Verzahnung von FDM- und Data-Science-Kompetenzen ist es möglich, Forschende während des gesamten Datenlebenszyklus zu unterstützen und so eine bestmögliche Wertschöpfung aus Daten zu erzielen (siehe auch Steinmann und Drechsler 2021, für weitere Informationen). Das DSC repräsentiert somit einen wichtigen strukturellen Pfeiler für die datenintensive Forschung an der Universität Bremen und fungiert auch als Bindeglied zwischen Forschenden und zentralen Infrastruktureinrichtungen (wie Bibliotheken oder Datenzentren). Auf regionaler Ebene bildet es gemeinsam mit dem Promovierenden-Trainingsprogramm „Data Train – Training in Research Data Management and Data Science“ (Hörner u. a. 2021) und im Zusammenspiel mit den Bremer Konsortien der Nationalen Forschungsdateninfrastruktur (NFDI) einen entscheidenden Baustein im Leitprojekt „Forschungsdatenmanagement und Data Science“ der U Bremen Research Alliance – ein Kooperationsnetzwerk der Universität und zwölf außeruniversitärer Institute.

2 Angebote und Services

Das DSC bietet vielfältige Angebote und Services, die angepasst an die Bedarfe der Forschenden kontinuierlich weiterentwickelt werden. Im Folgenden stellen wir diese eingeordnet in die drei Bereiche Datenmanagement, Data Science und Kulturwandel vor (für eine Zusammenfassung siehe Abbildung 1).

2.1 Angebote und Services im Bereich Datenmanagement

FDM-Beratung: Data Stewards unterstützen Forschende der Universität Bremen sowie der außeruniversitären Institute der U Bremen Research Alliance beim effektiven und nachhaltigen Management ihrer Forschungsdaten. Dabei arbeiten sie eng mit dem FDM-Beratungspersonal der Staats- und Universitätsbibliothek (SuUB) und des Referats für Forschung und wissenschaftlicher Nachwuchs der Universität Bremen sowie mit den in Bremen ansässigen NFDI-Konsortien und den Datenzentren PANGAEA und Qualiservice zusammen. Die FDM-Beratung erfolgt sowohl in der Antragsphase als auch forschungsbe-

¹ Gemäß der FAIR-Prinzipien sollen Forschungsdaten auffindbar (findable), zugänglich (accessible), interoperabel (interoperable) und nachnutzbar (reusable) sein (Wilkinson u. a. 2016).

gleitend. Typische Themen sind dabei Datenmanagementpläne, Datenorganisation, Datenspeicherung, FAIR-Prinzipien und die Nachnutzung von Forschungsdaten.

FDM-Trainings: Neben den Beratungstätigkeiten entwickeln die Data Stewards auch gezielte FDM-Trainings. Hierbei liegt der Fokus auf "hands-on-Formaten, die sich insbesondere an Promovierende und Postdocs richten. Bisher wurden Themen wie die Erstellung von Datenmanagementplänen (DMP) für Anträge, die Nutzung von Git und die Reproduzierbarkeit von Daten in Form eines „Reproducibility Hackathon“ adressiert. Die Trainings werden regelmäßig angeboten und angepasst an die Bedarfe um weitere Formate ergänzt.

Brücken bauen: Das DSC nimmt eine wichtige Rolle bei der Verknüpfung von Forschenden mit den zahlreichen FDM-Strukturen in der Bremer Forschungslandschaft ein. Ziel ist es, die bestehenden Angebote der unterschiedlichen Einrichtungen für die Zielgruppe sichtbar zu machen und Hürden bei der Nutzung zu verringern. Die Data Stewards agieren dabei im „Front Office“ als direkte Ansprechpartner:innen für Forschende, während die NFDI-Infrastruktur und Datenzentren im „Back Office“ die Daten entgegennehmen und zugänglich machen.

Peer2Peer Austausch: Der Austausch innerhalb der FDM-Community wird auf mehreren Ebenen gefördert. Zum einen werden lokale Community-Events in ungezwungener Atmosphäre für den kollegialen Austausch organisiert, wie z.B. „In Love With Data - A Coffee and Cake Get-Together of the Data-Steward-Network Bremen“ während der Love Data Week 2022. Zum anderen arbeitet das DSC aktiv in verschiedenen FDM-Netzwerken mit, wie dem lokalen Data-Steward-Netzwerk der U Bremen Research Alliance oder der DINI/nestor UAG-Schulungen/Fortbildung auf nationalem Level. Darüber hinaus beteiligt es sich am fachlichen Austausch auf den Workshops der NFDI oder FDM-Landesinitiativen und bei FDM-Konferenzen und -Vernetzungsevents wie dem Open Science Festival oder Data Stewardship goes Germany.

2.2 Angebote und Services im Bereich Data Science

Data-Science-Beratung: Das DSC unterstützt anwendungsorientierte Forschende bei der Einbindung von Data-Science-Komponenten in Anträgen und deren Umsetzung im Forschungsprozess. Dabei bietet es Beratung zu der Wahl der richtigen Methoden und Hardware, dem Aufbau einer passenden System-/IT-Architektur, der (Weiter-)Entwicklung von Algorithmen und dem Training neuronaler Netze. Als Partner für Anträge bietet es außerdem die Möglichkeit zur zentralen Ansiedlung von Data Scientists und Data Stewards zur bestmöglichen Bündelung und Nutzung von Kompetenzen.

Data-Science-Trainings: Data-Science-Trainings werden mit Fokus auf die am Standort Bremen häufig verwendeten Programmiersprachen Python und R angeboten. Thematisch werden Einführungskurse aber auch fortgeschrittene Workshops zu Maschinellem Lernen, Datenvisualisierung oder Zeitreihenanalyse angeboten. Ergänzend zur bereitgestellten IT-Infrastruktur gibt es regelmäßige Einführungskurse für die Nutzer:innen.



Abbildung 1: Zusammenfassende Darstellung der Angebote und Services des DSC bezogen auf die drei Bereiche Datenmanagement, Data Science und Förderung des Kulturwandel.

DSC Seed Grant: Über den DSC Seed Grant können Forschende finanzielle Unterstützung zur Umsetzung ihrer Data-Science-Vorhaben beantragen. Förderfähig sind beispielsweise datenwissenschaftliche Pilotstudien, die Durchführung von Konferenzen und Workshops in Bremen sowie Forschungsaufenthalte zur Vernetzung und Weiterentwicklung von Forschungsvorhaben. Die geförderten Vorhaben decken eine breite Spanne an datenwissenschaftlichen Themen ab mit Beteiligung von Forschenden aus unterschiedlichsten Disziplinen, wie Wirtschafts-, Sprach-, Geo-, Sozial-, und Kulturwissenschaften sowie der Informatik und Mathematik. Der DSC Seed Grants stärkt damit die datenwissenschaftliche, kooperative Forschung.

IT-Infrastruktur: Die spezialisierte IT-Infrastruktur ist auf die Durchführung von Data-Science-Anwendungen wie maschinelles Lernen ausgelegt. Sie bietet Kapazitäten für Virtualisierung, Speicherung und High-Performance-Computing, um das Arbeiten mit aufwendigen Algorithmen zu ermöglichen. Für Anwender:innen dient sie dabei als leicht zugängliche Struktur mit deutlich höherer Leistung als ein herkömmlicher PC. Somit schließt

sie die Lücke zwischen Rechenressourcen auf Arbeitsgruppenlevel und nationalen Supercomputern wie das System des Norddeutschen Verbunds für Hoch- und Höchstleistungsrechnen (HLRN). Im Sinne der Ressourcenschonung und -bündelung, ist die Architektur auf Skalierbarkeit ausgelegt und kann flexibel erweitert werden. Dies bietet die Möglichkeit flexibel auf die Bedürfnisse und Anforderungen unserer Nutzer:innen zu reagieren.

2.3 Angebote und Services zur Förderung des Kulturwandels

Strategieentwicklung: Das DSC unterstützt die Universität Bremen bei der Entwicklung von Strategien in Bezug auf Datenkompetenzen, Beratungsstrukturen, FDM-Leitlinien und technische Infrastruktur. Dabei wurde aus der Zusammenarbeit in der U Bremen Research Alliance im Jahr 2021 ein Whitepaper erstellt, das den aktuellen Stand sowie zukünftige Handlungsfelder für die disziplinübergreifende Etablierung eines kooperativen FDM skizziert (Pigeot u. a. 2021).

Wissenstransfer: Um einen wechselseitigen Transfer von Wissen zwischen Wissenschaft, Gesellschaft, Wirtschaft und Politik zu ermöglichen, engagiert sich das DSC bei der Umsetzung unterschiedlicher Wissenschaftskommunikationsformate. Hierzu gehören informationsvermittelnde Formate wie öffentlichen Vorlesungen oder digitale Beiträge (z.B. auf Social-Media-Kanälen oder Blogs) sowie dialogorientierte Formate wie Podiumsdiskussionen, Roundtables oder World Cafés.

Vernetzung: Die wissenschaftliche Vernetzung und der datenwissenschaftliche Austausch werden auf verschiedenen Ebenen gefördert. Lokal haben Forschende die Möglichkeit, Teil des Mitgliedernetzwerks zu werden und so ihre datenwissenschaftlichen Kompetenzen sichtbar zu machen. Seminarreihen wie das „Data Science Forum“ bieten eine Plattform für den wissenschaftlichen Dialog. Auf nationaler Ebene engagiert sich das DSC durch die Durchführung von oder den Beitrag zu Community-Events und Workshops. Im internationalen Kontext wird die Vernetzung von Forschenden und der Aufbau neuer Forschungsoperationen durch länderübergreifende Austauschformate wie dem „Bremen-Cardiff Data Science Workshop“ gefördert. Ziel dieser Aktivitäten ist es, die interinstitutionelle, disziplinübergreifende Zusammenarbeit zu fördern und den wechselseitigen Kompetenztransfer zu ermöglichen.

Bewusstsein schaffen: Um Forschende für einen FAIRen Umgang mit Daten zu sensibilisieren und ihnen die damit verbundenen Vorteile für sich und ihre Forschung aufzuzeigen, bietet das DSC unterschiedliche niederschwellige Informationsangebote. Dazu gehört beispielsweise das hybride Coffee-Lecture-Format „Data Snacks“, das in kurzweiligen Sessions zu aktuellen Themen, wichtigen datenwissenschaftlichen Aspekten und hilfreichen Diensten informiert. Darüber hinaus steht das DSC für Expert:innen-Inputs bei Early Career Events, an Instituten oder Fachbereichen sowie für Lehrveranstaltungen zur Verfügung. Durch die aktive Ansprache von Forschenden wird eine individuelle Sensibilisierung für einen verantwortungsvollen Umgang mit Daten erreicht.

3 Ausblick

Das DSC hat sich bereits jetzt als bedeutende Institution an der Universität Bremen und in der Bremer Forschungslandschaft positioniert und kann für andere Regionen als Modell für die institutionelle Zusammenführung von FDM- und Data-Science-Kompetenzen dienen. Im nächsten Schritt soll die Zusammenarbeit und der Kompetenzaustausch mit anderen „Data Science Centern“ und Data-Science-Initiativen vertieft werden. Dafür organisiert das DSC im Rahmen der INFORMATIK 2023 einen Community-Workshop zum Thema „Aktuelle Entwicklungen und Perspektiven (an Hochschulen) im Bereich Data Science“². Die Angebote und Aktivitäten des DSC werden kontinuierlich weiterentwickelt, basierend auf den Bedürfnissen der Nutzenden. Die Bedarfserfassung erfolgt beispielsweise durch quantitative Umfragen und im Dialog mit Forschenden z.B. in Beratungen. Die Wirksamkeit der Maßnahmen wird über verschiedene Kennzahlen bewertet, darunter die Anzahl und fachliche Diversität der Nutzer:innen in Bezug auf Beratungen, Trainings, Seed Grants, IT-Infrastruktur, Netzwerkevents und Wissenstransferformate. Auch die Anzahl der unterstützten Drittmittelanträge und Forschungsprojekte sowie die eingeworbenen Drittmittel werden berücksichtigt. Die Qualität der Trainings wird außerdem über Evaluationen sichergestellt. Zusätzlich betrachtet das DSC die Teilnahme an (inter)nationalen Community-Events sowie die Veröffentlichung von Materialien und Konzepten (wie Trainingsinhalte und Coffee Lectures) zur Nachnutzung als essenziell, um einen standortübergreifenden Kompetenztransfer zu ermöglichen und den kulturellen Wandel voranzutreiben.

Um der steigenden Nachfrage nach Unterstützungsangeboten gerecht zu werden, ist perspektivisch eine Erweiterung der personellen Unterstützungsstruktur geplant. Hierbei sollen zusätzliche zentral koordinierte Data Scientists und Data Stewards für unterschiedliche Disziplinen eingebunden werden. Eine Finanzierung kann beispielsweise durch Verbundprojekte wie Exzellenzcluster, Sonderforschungsbereiche oder Graduiertenkollegs sichergestellt werden.

Danksagung

Die Autor:innen danken dem Land Bremen für die Förderung des Data Science Centers.

Literaturverzeichnis

Hörner, Tanja, Iris Pigeot, Frank Oliver Glöckner und Rolf Drechsler. 2021. „Disziplinübergreifendes Modell zur Ausbildung von Forschungsdatenmanagement und Data Science Kompetenzen: 'Data Train – Training in Research Data Management and Data Science'“. *Bausteine Forschungsdatenmanagement*, Nr. 3: 56–69. DOI: <https://doi.org/10.17192/bfdm.2021.3.8343>.

² <https://informatik2023.gi.de> und <https://www.dsc-ub.de/GI-data-science-workshop.php>;
Zuletzt aufgerufen am 11.Mai 2023.

- Pigeot, Iris, Frank Oliver Glöckner, Rolf Drechsler, Tanja Hörner, Derk Hergen Schönfeld, Lena Steinmann und Björn Oliver Schmidt. 2021. *Etablierung eines kooperativen Forschungsdatenmanagements in der U Bremen Research Alliance*. Technischer Bericht. U Bremen Research Alliance. DOI: <https://doi.org/10.5281/zenodo.4775371>.
- Steinmann, Lena, und Rolf Drechsler. 2021. „Verzahnung von Data Stewardship und Data Science – Wege und Perspektiven“. *Bausteine Forschungsdatenmanagement*, Nr. 3: 82–91. DOI: <https://doi.org/10.17192/bfdm.2021.3.8342>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersbergand, Gabrielle Appleton, Myles Axtonand, Arie Baakand, Niklas Blombergand u. a. 2016. „The FAIR Guiding Principles for scientific data management and stewardship“. *Scientific data* 3 (1): 1–9. DOI: <https://doi.org/10.1038/sdata.2016.18>.
- Wissenschaftsrat. 2020. „Zum Wandel in den Wissenschaften durch datenintensive Forschung“. Besucht am 11. Mai 2023. <https://www.wissenschaftsrat.de/download/2020/8667-20.html>.