

## Business Economic and Related Data @ NFDI (BERD@NFDI)

*Sprecher: Prof. Dr. Florian Stahl, Universität Mannheim, [florian.stahl@uni-mannheim.de](mailto:florian.stahl@uni-mannheim.de)*

Die **Forschung in der Betriebswirtschaftslehre, Volkswirtschaftslehre und verwandten Fachgebieten** zielt auf die Interaktion zwischen Menschen und Organisationen innerhalb einer Gesellschaft. Um ein tieferes Verständnis über diese komplexen Systeme zu gewinnen, werden in diesen Fachgebieten traditionell strukturierte Daten und entsprechende empirische Methoden genutzt. Für ein umfassenderes Gesamtbild werden jedoch zunehmend unstrukturierte Daten unentbehrlich, d.h. Informationen, die kein vordefiniertes Datenschema haben oder nicht in einer vordefinierten Art und Weise organisiert sind (z.B. Bilder, Videos, Social Media). Die weltweite digitale Transformation von Gesellschaft und Wirtschaft verstärkt diesen Trend: Schätzungen gehen davon aus, dass bis 2025 etwa 80 % der Daten, die in der Wirtschaft entstehen, unstrukturierter Natur sein werden. Aufgrund des enormen Datenvolumens und vor allem wegen der Unstrukturiertheit und Heterogenität der Rohdaten benötigen Forschende innovative und nachnutzbare Methoden, insbesondere aus den Gebieten der Künstlichen Intelligenz und des Maschinellen Lernens, um diese Daten so organisieren und strukturieren zu können, dass sie für wissenschaftliche Analysen nutzbar sind. Folglich werden Algorithmen zunehmend ein integraler Bestandteil des Lebenszyklus von Forschungsdaten in den Wirtschaftswissenschaften und darüber hinaus und müssen daher genauso verwaltet werden wie die Daten selbst.

Vor diesem Hintergrund ist die Mission von BERD@NFDI, eine zukunftsorientierte und leistungsfähige Forschungsdateninfrastruktur zu entwickeln, dauerhaft bereitzustellen und zu betreiben, die das Potential hat, das Forschungsdatenmanagement in den Wirtschaftswissenschaften und verwandten Bereichen in Deutschland auf ein neues, herausragendes Niveau zu heben. Mit seinem **Fokus auf dem integrierten Management von (un-)strukturierten Daten und zugehörigen wissenschaftlichen Algorithmen** sowie seinem klaren Bekenntnis zur Offenheit (z.B. open software, open standards) und Nachvollziehbarkeit (insb. FAIR data principles) von Forschung kann BERD@NFDI als ein wichtiger, wenn nicht gar einzigartiger Mosaikstein für den Aufbau der NFDI betrachtet werden.

Die erklärten **Ziele bzw. Maßnahmen und Services** von BERD@NFDI sind in klar definierte „Task Areas“ gegliedert:

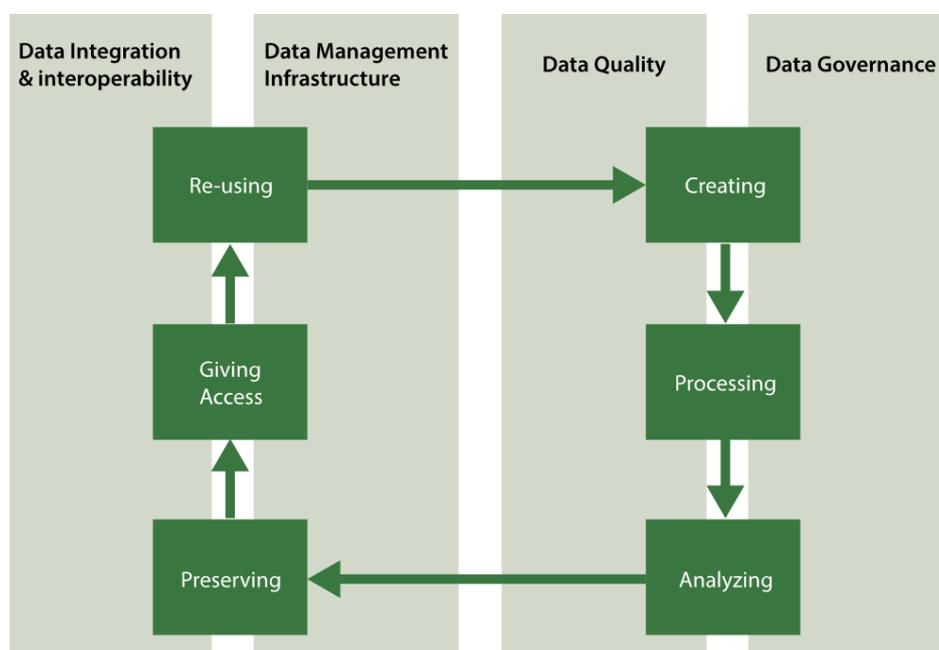
- **BERD@NFDI community involvement:** Die Einbindung der Nutzer spielt bei BERD@NFDI eine wichtige Rolle. Ein auf die Nutzer zugeschnittenes Design sowie agile Entwicklungsmethoden stellen sicher, dass die Infrastruktur auf die tatsächlichen Bedürfnisse der Nutzergemeinschaft ausgerichtet ist.
- **Creating BERD:** In den Sozialwissenschaften im Allgemeinen und in den Wirtschaftswissenschaften im Besonderen hat sich sowohl die Menge als auch die Art der Datenquellen weiterentwickelt. Durch erhöhte Transparenz beim Datenzugang und bei der Datenerhebung ist es möglich, frühere Datenerhebungsmethoden für deren Replizierbarkeit, Erweiterung und Anwendung auf neue Forschungsprojekte wiederzuverwenden.
- **Processing BERD:** Wissenschaftlerinnen und Wissenschaftler werden mit einer Vielzahl von Daten und unterschiedlichen Niveaus der Datenqualität konfrontiert. BERD@NFDI wird die Fachcommunity bei der Auswahl geeigneter Methoden zur Verarbeitung von wirtschaftswissenschaftlichen Forschungsdaten sowie bei deren Dokumentation und Bereitstellung unterstützen.
- **Analyzing BERD:** Zur Auseinandersetzung mit Forschungsfragen bedarf es nicht nur relevanter Daten, sondern auch entsprechender Algorithmen, um diese Daten zu analysieren. BERD@NFDI ermöglicht Nutzern, die Leistung von Algorithmen zu

bewerten, um potenzielle neue Anwendungen schnell zu verstehen und Leistungsunterschiede zwischen verschiedenen Anwendungen zu vergleichen.

- **Preserving BERD:** Eine große Herausforderung im Management von Forschungsdaten ist die langfristige Archivierung und Zugänglichmachung digitaler Inhalte. BERD@NFDI wird die Aufbewahrung und Verarbeitung von Daten übernehmen und darüber hinaus Strategien zur Datenmigration entwickeln.
- **Access to BERD:** Der Zugang zu einer leistungsstarken Forschungsdateninfrastruktur bietet die Möglichkeit, Dienste sowohl für menschliche als auch für maschinelle Szenarien (erneut) zu nutzen. Die zukünftige Plattform BERD@NFDI wird alle notwendigen Voraussetzungen bündeln und Zugriff auf weitere Funktionalitäten wie Suche, Fernzugriff und persistente Identifizierung bieten.
- **Re-using BERD:** Wissenschaftlerinnen und Wissenschaftler werden beim Austausch mit erzeugten Daten häufig mit ressourcenbedingten oder rechtlichen Einschränkungen konfrontiert. Selbst wenn sie ihre Daten zur Verfügung stellen, befinden sich diese oft in Archiven mit nur wenigen Nutzern und haben somit nur einen geringen Einfluss auf weitere wissenschaftliche Erkenntnisse. BERD@NFDI wird dabei unterstützen, diese bestehenden Barrieren zu überwinden.

In Anlehnung an den UK Data Archive Lifecycle fokussiert BERD@NFDI auf ein **ganzheitliches Datenmanagement** für die Erhebung, Aufbereitung, Analyse, Archivierung, das Teilen und die Nachnutzung insb. unstrukturierter Daten. Das Fundament dieser Maßnahmen bilden vier Säulen:

- Data Governance (einschließlich ethischer Gesichtspunkte),
- Datenintegration und -interoperabilität,
- Datenqualität (einschließlich der Bereinigung, Integrität, Anreicherung und Qualitätssicherung von Daten),
- Infrastruktur für das Datenmanagement (einschließlich Datenbank- und Speichermanagement, integriertes Management von Daten und Algorithmen).



Um den Spezifika der unstrukturierten Daten (Datenumfang, Heterogenität) gerecht zu werden, wird BERD@NFDI einen **besonderen Schwerpunkt** auf die Einbindung von Expertise und Ressourcen in den Bereichen **Künstliche Intelligenz** und **Machine Learning** sowie Rechen- und Speicherleistung setzen.

Im Hinblick auf **Querschnittsthemen** unterstützt BERD@NFDI die Berlin Declaration of handling NFDI-cross-cutting topics<sup>1</sup> und wird daher bei den Themen “Collaborative governance and general framework”, “Community (User) involvement“, „Technical infrastructures and concepts“ und „Legal and ethical aspects“ in konsortiumsübergreifenden Arbeitsgruppen mit bestehenden und zukünftigen NFDI-Konsortien kooperieren. Insbesondere kann BERD@NFDI aufgrund der Expertise und Schwerpunktsetzung der teilnehmenden Institutionen in folgenden Querschnittsbereichen einen Beitrag leisten:

- **Aus- und Fortbildung:** BERD@NFDI besitzt herausragendes Know-How in der Konzeption und Implementierung von E-Learning-Kursen und -Workshops im Bereich Data Science und wird dies in die NFDI einbringen.
- **Nutzereinbindung:** Als Initiative aus der wirtschaftswissenschaftlichen Community heraus und unter Einbindung entsprechender Spezialisten auf dem Gebiet des nutzerzentrierten Designs ist die Nutzereinbindung ein wesentliches Kernelement von BERD@NFDI.
- **Rechtliche Fragestellungen:** BERD@NFDI fokussiert v.a. auf rechtliche Fragen rund um das Teilen unstrukturierter Daten unter Berücksichtigung von Anonymisierungsmethoden.
- **Qualitätssicherung:** Im Mittelpunkt steht hier die Entwicklung von Standards für die Qualitätssicherung, die Normalisierung und das Pre-Processing von unstrukturierten Daten sowie deren standardisierte Dokumentation in Forschungsarbeiten.
- **Standardisierung und Homogenisierung von Terminologien:** Vertiefte Kenntnisse in den Bereichen der in den Wirtschaftswissenschaften genutzten Vokabularien und Ontologien sowie die geplante Entwicklung einer Firmenontologie im Projekt leisten einen Beitrag für ein Semantic Web wissenschaftlicher Daten.
- **Interoperabilität von Metadaten:** Die disziplinspezifischen Anforderungen an die Metadaten werden mit der Möglichkeit der inter-disziplinären Nachnutzung und allgemeinen Leitlinien wie bspw. den FAIR-Prinzipien in Einklang gebracht. Mit dem Go-Fair-Büro für Deutschland an der ZBW und Prof. Dr. Tochtermann als Sprecher der deutschen Go-Fair-Community ist FAIRes Forschungsdatenmanagement fest in BERD@NFDI verankert.

Im BERD@NFDI-Konsortium finden sich hochgradig relevante Institutionen aus Infrastruktur und Forschung unter der Leitung der Universität Mannheim zusammen, die auf den vier Säulen der BERD-Infrastruktur **ausgewiesene Expertise** aufweisen: (empirische) Wirtschaftswissenschaften und verwandte Gebiete, Künstliche Intelligenz & Maschinelles Lernen, leistungsfähige Recheninfrastruktur und FAIRes Datenmanagement. Viele dieser Institutionen kooperieren schon lange erfolgreich, sodass in der Zusammenarbeit ein hohes Maß an Effizienz und Effektivität gegeben ist, die das Konsortium dafür nutzen wird, das ehrgeizige Arbeitsprogramm von BERD@NFDI erfolgreich durchzuführen. Darüber hinaus besteht auf internationaler Ebene eine enge Verknüpfung mit der inhaltlich ähnlich gelagerten US-amerikanischen Coleridge Initiative. Schließlich wird die transparente Governance-Struktur mit klaren Verantwortlichkeiten in Verbindung mit einem nutzerorientierten Multi-Stakeholder-Ansatz den Erfolg von BERD@NFDI sicherstellen und einen strukturellen Beitrag zum Erfolg der NFDI insgesamt leisten.

Mit der **Teilnahme an der 2. NFDI-Konferenz** soll ein Austausch mit NFDI-Konsortien, die bereits zur Förderung vorgesehen sind oder sich in Vorbereitung finden, angestoßen werden, übergreifenden Themen identifizieren und Kooperationen anbahnen zu können und ggf. spezifische Schwerpunkte bilden zu können.

---

<sup>1</sup> <https://doi.org/10.5281/zenodo.3457213>.

**Vorgesehene Mitglieder des Konsortiums (Co-Sprecherinnen/Co-Sprecher und die weiteren, beteiligten Institutionen):**

<b>Co-Sprecher/in</b>	<b>Zugehörige Institution</b>
Prof. Dr. Bernd Bischl, Lehrstuhl für Statistical Learning & Data Science, bernd.bischl@stat.uni-muenchen.de	Ludwig-Maximilians-Universität München Ludwigstr. 33, 80539 München
Prof. Dr. Marc Fischer, Chair of Marketing and Market Research, marc.fischer@wiso.uni-koeln.de	Universität zu Köln, Albert-Magnus-Platz, 50923 Köln
Dr. Sabine Gehrlein, Leitende Bibliotheksdirektorin, sabine.gehrlein@bib.uni-mannheim.de	Universität Mannheim, Universitätsbibliothek, Schloss Schneckenhof West, 68161 Mannheim
Prof. Dr. Mark Heitmann, Professur für Marketing & Customer Insight, mark.heitmann@uni-hamburg.de	Universität Hamburg, Mittelweg 177, 20148 Hamburg
Prof. Dr. Hartmut Höhle, Professor für Enterprise Systems, hartmut.hoehle@uni-mannheim.de	Universität Mannheim, L15,1-6, 68161 Mannheim
Prof. Dr. Göran Kauermann, Professor für Statistik und ihre Anwendungen in Wirtschafts- und Sozialwissenschaften, goeran.kauermann@stat.uni-muenchen.de	Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, 80539 München
Prof. Dr. Frauke Kreuter, Professor of Statistics and Social Research Methods, frauke.kreuter@uni-mannheim.de	Universität Mannheim, Schloss, 68161 Mannheim / Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit (IAB), Regensburger Str. 100, 90478 Nürnberg
Dr. Georg Licht, Leiter Forschungsbereichs "Innovationsökonomik und Unternehmensdynamik", georg.licht@zew.de	ZEW – Leibniz Center for European Economic Research GmbH, L7,1, 68161 Mannheim
Alexander Pfister, Kerstin Bein, Leitung Universitäts-IT alexander.pfister@uni-mannheim.de, kerstin.bein@uni-mannheim.de	Universität Mannheim, Universitäts-IT, L15,1-6, 68161 Mannheim
Prof. Dr. Klaus Tochtermann, Direktor der ZBW - Leibniz- Informationszentrum Wirtschaft und Professor für Digitale Informationsinfrastrukturen an der Christian-Albrechts-Universität zu Kiel, k.tochtermann@zbw.eu	Leibniz Information Center for Economics (ZBW), Düsternbrooker Weg 120, 24105 Kiel