



Rainer Niekamp. Foto: Vladimir Unkovic

*Ein verantwortungsbewusster und transparenter Umgang mit Forschungsdaten ist für die Qualität zeitgemäßer wissenschaftlicher Forschung von wesentlicher Bedeutung. Mit einer großen Dynamik hat der „digitale Wandel“ auch die Denkweise im Bereich der Materialwissenschaften beeinflusst und die Zusammenarbeit durch gemeinsame und kombinierte Forschungsdaten in den Fokus gerückt.*

# Von einzelnen Daten zu vernetztem Wissen

Forschungsdatenmanagement in einem interdisziplinären Verbundprojekt

Von Rainer Niekamp und Matthias Grönwald

Der Umgang mit Daten ist schon immer ein grundlegender Schritt hin zu neuem Wissen. Seit den Anfängen der Menschheit werden auf experimentelle Art und Weise natürliche Phänomene begriffen und zunehmend beherrschbar. Die systematische Durchführung von Versuchen als Frage an die Natur bildet zusammen mit der Auffindung immer neuer Fragestellungen das Paradigma der experimentellen Wissenschaft. Das Erkennen von Regelmäßigkeiten in den Versuchsergebnissen und deren Beschreibung durch Naturgesetze führte zum Paradigma der theoretischen Wissenschaft, die zu tiefgreifenden Erkenntnissen kam und das Weltbild mehrfach grundlegend veränderte. All dies ist verbunden mit dem Erheben, der Verarbeitung und auch dem Teilen von Daten, genauer Forschungsdaten. Unter anderem über solche gemeinsamen Forschungsdaten bleibt die theoretische Wissenschaft weiterhin an Experimente gebunden, um aufgestellte Vermutungen widerlegen oder bestätigen zu können. Die Anwendung der so gefundenen Gesetze auf komplexe Systeme führt zu komplizierten Berechnungen, die erst unter Verwendung digitaler Systeme handhabbar wurden. Die computergestützte Berechnung von Modellen als Annäherungen an die Wirklichkeit kam zum Paradigma der simulierenden Wissenschaft. Die entwickelten Modelle müssen hier zum einen die Naturgesetze abbilden und zum anderen mittels experimenteller Forschungsdaten kalibriert werden. Eine zunehmende Bedeutung bekommt die datengetriebene Wissenschaft [1,2]. Durch die Verwendung von Algorithmen aus dem Bereich des maschinellen Lernens und unter massivem Einsatz von digitalen Systemen können aus experimentellen oder simulierten Daten Vorhersagen gewonnen werden, die beispielsweise Eigenschaften (noch) nicht existierender Materialien betreffen. Die Umsetzung dieses Paradigmas benötigt umfangreiche, aber vor allem gleichartige, formalisierte Daten,

die als Trainingssatz das maschinelle Lernen erst ermöglichen. Die modernen wissenschaftlichen Methoden brauchen somit ein modernes Forschungsdatenmanagement. Als Bestandteil einer guten wissenschaftlichen Praxis dient das sogenannte FAIR-Prinzip [3] als Wegweiser. Das Konzept beschreibt, dass wissenschaftliche Daten über einfache Mechanismen auffindbar (findable) sind, einer (gegebenenfalls eingeschränkten Gruppe) jederzeit zur Verfügung gestellt werden können (accessible), dass sie mit anderen Daten verknüpfbar sind sowie flexibel einer Weiterverarbeitung mit verschiedenen Werkzeugen zugefügt werden können, also interoperabel sind und damit einer Nachnutzung verfügbar (reusable). Daneben bietet sich für ein optimiertes Forschungsdatenmanagement (FDM) der Anreiz, gewonnene Daten in systematischer Form als Daten-Veröffentlichungen beziehungsweise allgemein im Kontext der Open Science Bewegung [4] unter Erhaltung der Zitierfähigkeit nutzbar zu machen. [5]

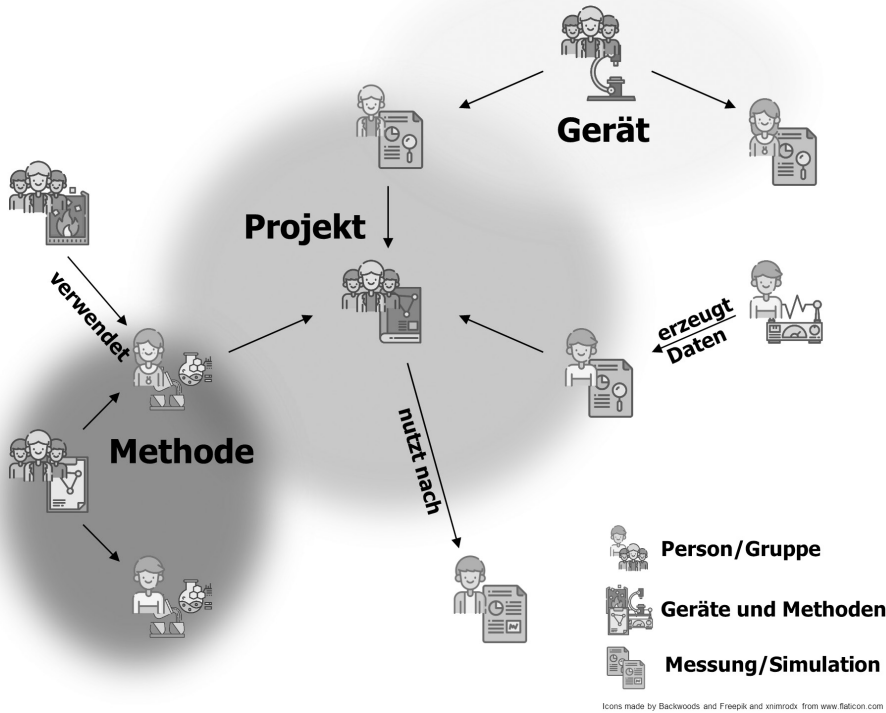
### **Unterstützung aus den eigenen Reihen**

Das Verbundforschungsprojekt SFB/TRR 270 „Hysteresis design of magnetic materials for efficient energy conversion“ – kurz HoMMage – widmet sich der Erforschung neuer magnetischer Materialien zur Energieumwandlung. Dies umfasst sowohl harte Permanentmagneten mit maximierter Hysteresese, als auch weiche Magnete mit minimierter Hysteresese. Beide Arten haben unterschiedliche Anwendungsfelder, erstere unter anderem im Bereich der Windgeneratoren oder der Elektromobilität, letztere im Bereich magnetische Kühlung unter Anwendung des magnetokalorischen Effekts. Allen Anwendungen ist gemein, dass sie von, für den jeweiligen Einsatz gezielt verbesserten, neuen Magnetmaterialien profitieren. Deshalb suchen innerhalb des SFB/TRR 270 Forschende aus unter-

schiedlichen Disziplinen wie der Materialwissenschaft, der Physik, der Chemie oder der Fertigungstechnik, verteilt über mehrere Arbeitsgruppen an fünf Standorten, nach neuen innovativen Magnetmaterialien und Wegen zu deren Verarbeitung.

Neben den klassischen Paradigmen wird in diesem multidisziplinären Gemeinschaftsprojekt der datengetriebene Ansatz verfolgt, neue magnetische Materialien mit verbesserten Eigenschaften zu finden. Das INF-Teilprojekt (Informationsmanagement und -infrastruktur) unterstützt hierbei vor allem in Fragen des FDM, unter anderem damit die generierten Forschungsergebnisse zentral und strukturiert gespeichert und ausgetauscht werden können.

Als zentrales Dienstleistungsprojekt innerhalb des SFB/TRR 270 zielt es darauf ab, ein nachhaltiges Management von Forschungsdaten gemäß dem FAIR-Prinzip für alle Beteiligten standortübergreifend zu ermöglichen. In einer abstrakten Betrachtung sind, neben rein infrastrukturellen Herausforderungen, die wichtigsten: Komplexität, Heterogenität und Größe der Daten. Eine der ersten Aufgaben war die Definition von Plänen für die Datensammlung und -verwaltung (sog. Datenmanagementpläne, kurz DMP). Ein elektronisches Laborbuch (ELB) auf Basis von freier Software (FLOSS) steht zur Verfügung und wird laufend erweitert. Weite Teile der entstehenden Forschungsdaten können darüber geladen und geteilt werden und sind über eine Benutzeroberfläche nach grundlegenden Kriterien auffindbar. Über eine Schnittstelle stehen die Inhalte für automatisierte Datenkonvertierungen und -reduktionen zur Verfügung. In Verbindung mit einer Speicherinfrastruktur und Schnittstellen für besonders große Forschungsdaten ist es Teil des notwendigen Werkzeugesatzes, um die gesammelten Forschungsdaten und Metadaten mithilfe von Algorithmen des maschinellen Lernens zu analysieren und neue Materialien zu entdecken. Eine



(1) Schematische Struktur der Forschungsdaten im ELB. Bei den abstrakten Objekten handelt es sich um gemeinsam oder individuell genutzte Forschungsdaten bzw. -metadaten, die durch Beschreibungen, Kategorien und Bereiche einer grundlegenden Semantik unterliegen. Personen teilen sich Bereiche, wie eine Methode oder ein Gerät. Einzelne Vorgänge, wie Messexperimente, werden einem Teilprojekt zugeordnet. Bezüge zwischen den Objekten erhalten eine implizite Bedeutung und erlauben es, Forschungsdaten besser in Kontext zu setzen, um Synergien zu erzeugen. Aus individuellen Abläufen können so gemeinsame Wissensnetzwerke gebildet werden.

Quelle: eigene Darstellung

weitere wesentliche Aufgabe sind Schulungsangebote für alle Mitglieder des SFB/TRR 270 zu Themen im Bereich FDM durch regelmäßige Workshops und Beratung sowie die Sensibilisierung für mögliche Verbesserungen im Umgang mit Daten. Um einem Silodenken vorzubeugen, ist auch die Integration von Daten und Werkzeugen in die breitere FDM-Landschaft ein wichtiges Thema.

An das FDM innerhalb eines interdisziplinären Verbundprojektes werden viele Anforderungen gestellt. Neben technischen Anforderungen und Inkompatibilitäten bei der Integration bestehender Infrastruktur erzeugt die Kollaboration zwischen den fachlich heterogenen Gruppen

auch neuen, zum Teil spezialisierten, Bedarf. Das INF-Projekt innerhalb des SFB/TRR 270 HoMMage hat sowohl das Ziel, diesen Herausforderungen mit einer Kombination selbstverwalteter und zentraler Infrastruktur zu begegnen sowie der Aufgabe, den Forschenden innerhalb des Verbunds Wissen und Unterstützung zu einem besseren FDM zu bieten, als auch den Anspruch einer wissenschaftlichen Nachnutzung im Kontext der künstlichen Intelligenz. So kann mithilfe strukturierter Forschungsdaten und moderner Computerverfahren ein erfolgreicher Beitrag zur Entwicklung neuer vielversprechender magnetischer Materialien geliefert werden.

**Werkzeuge sind notwendig, aber nicht hinreichend**

Als eine der in diesem Rahmen zentralen Aufgaben wird meist das Angebot von verschiedenen Diensten und Werkzeugen zum Forschungsdatenmanagement wahrgenommen. Dabei stehen oft Fragen des infrastrukturellen Betriebs, des Ressourcenaufwands sowie der Zugriffsverwaltung beziehungsweise Verfügbarkeit im Vordergrund. Das ist naheliegend, da an diesen Stellen die Kosten und der Aufwand für ein verbessertes FDM am direktesten sichtbar sind. Letztlich gibt es natürlich für jeden Abschnitt des klassischen Datenlebenszyklus [6] Werkzeuge, die Forschende unterstützen. Ein Fokus

liegt aber oft auf jenen, die entweder die Speicherung, Verwaltung oder Archivierung von Forschungsdaten begleiten und ermöglichen.

Für das Verbundprojekt SFB/TRR270 werden ebenso einige Dienste zur Verfügung gestellt, unter anderem um die teilweise speziellen Anforderungen abdecken zu können, die sich nicht sinnvoll durch zentrale Dienstleistungen befriedigen lassen. Diese vom INF-Projekt verwaltete Infrastruktur besteht dabei im Wesentlichen aus drei Säulen: einem zentralen elektronischem Laborbuch, Speicher- und Analyse-Hardware zur Verarbeitung von „lebenden Forschungsdaten“ [7] sowie dem Anschluss an ein institutionelles Repositorien zur richtlinienkonformen Archivierung. Aus der Perspektive des Datenlebenszyklus unterstützt diese Infrastruktur somit die Datenerhebung, die Datenprozessierung sowie die Datenarchivierung. Diese stehen im Verbund mit Diensten der verschiedenen Institutionen für die Planung sowie die Veröffentlichung von Forschungsdaten. Das sind unter anderem verschiedene Research Data Management Organizer (RDMO)-Instanzen an den einzelnen Standorten, mit tudmo [8] als übergreifende genutzte Instanz zur Erstellung und Verwaltung von DMP sowie die Möglichkeit Forschungsdaten aus dem Archiv, beispielsweise mittels tudatalib [9] oder einem fachspezifischen Forschungsdatenrepositorium zu veröffentlichen.

Als zentrales ELB kommt im Projekt elabFTW [10] zum Einsatz. Dabei handelt es sich um eine web-basierte freie Software, die kontinuierlich weiterentwickelt wird und an den beiden Kernstandorten des Verbundprojekts (Universität Duisburg-Essen und TU Darmstadt) auch als zentraler Dienst betrieben wird [11] beziehungsweise in der Planung ist. Die Stärke von elabFTW liegt unter anderem in einem recht freien Ansatz, der es trotzdem erlaubt, Forschungsdaten aus unterschiedlichen wissenschaftlichen

Abläufen zu verbinden (Abb. 1). Wie erwähnt, ist eine der Herausforderungen die Heterogenität der Daten, weswegen ein stark formalisiertes ELB mit hoher fachlicher Spezifität nahezu zwingend einzelne Gruppen behindert oder sogar ausschließt. Durch die Festlegung grundlegender Leitlinien zur Nutzung und einer Konfiguration bis auf tiefe Systemebenen konnte das Werkzeug für den im Projekt breiten Bedarf nutzbar gemacht werden. Aus infrastruktureller Perspektive ist dabei die für ein ELB ungewöhnliche hohe Speicherkapazität sowie der zentrale Zugriff mittels Single-Sign-On-Dienst hervorhebenswert, die beide eine Kooperation zwischen den Projekten erleichtern. Die Hauptzielgruppe für ein ELB ist hier wie allgemein die experimentelle Forschung, aber über die integrierte Datenbank können alle Projektinhalte untereinander verknüpft werden. Für größere Datenmengen, zum direkten Austausch mit der speziellen Hardware zur Datenprozessierung oder dem HHLR [12] schließt sich an das ELB eine zusätzliche Speicherinfrastruktur an.

Bei allen solchen dezentral betriebenen Diensten zum FDM ist eine gute Integration und Vernetzung mit den zentralen Hochschuleinrichtungen von großer Bedeutung. Ohne sie verbleibt die Infrastruktur nach Projektende schnell ungenutzt und wird zur Belastung oder gar zum Risiko. Im Verbundprojekt wurde dem dadurch begegnet, dass man entweder bestehende Dienste nutzt und ausbaut oder im konstanten Dialog mit den zentralen Einrichtungen Kompatibilität wahrt. Dabei ist es für das INF-Projekt von entscheidender Bedeutung, die Koordination sowie ein bedarfsgerechtes Angebot zu gewährleisten. Nur durch diese Integrationsleistung in die bestehende Infrastruktur ergeben sich weitere Vorteile für den Betrieb der eigenen FDM-Dienste durch hohe Verfügbarkeit, zentrale Account-Authentifizierung und schnellen Support. So kann auch

die Wartung nach höheren professionellen Standards gewährleistet werden als es für rein wissenschaftlich genutzte Hardware oft der Fall ist. Ähnliches gilt für Speicher- und Backup-Systeme. Auch die Erfahrungen im Umgang mit Auflagen zu Datenschutz und Arbeitsrecht sind hilfreich für einen gesetzeskonformen Betrieb. Diese Aspekte bieten im Forschungsalltag wesentliche Vorteile.

Ein dezentraler Betrieb von Infrastruktur und das Angebot von einzelnen Diensten zum FDM ist ein sehr aufwendiger Schritt. So ist eine umfassende Bedarfserhebung im Vorfeld und eine angeschlossene Planung unerlässlich. Auch lohnt sich eine umfassende Evaluation von Fördermöglichkeiten beziehungsweise umgekehrt Auflagen von Förderinstitutionen vor der Auslegung von FDM-Strukturen. Ein eigenständiger und von zentralen Diensten unabhängiger Betrieb birgt immer das Risiko von Insellösungen und fördert eine Silodenken bei den Forschenden, weswegen er wohlüberlegt sein sollte. Allgemein darf ohnehin die Bedeutung von Werkzeugen nicht überschätzt werden, da sie zwar oft zu den sichtbarsten und greifbarsten Elementen eines FDM-Angebots zählen, in den meisten Fällen aber nicht die entscheidenden Bausteine sind. Ihren vollen Nutzen können sie letztlich nur in Verbindung mit angepassten Nutzungsabläufen und hoher Akzeptanz bei den Forschenden erzielen. Das gilt es, gerade mit Blick auf den oft hohen Investitionsbedarf von Infrastruktur zu berücksichtigen. Die Gestaltung von derartigen Abläufen zu unterstützen und Forschende durch Information, Motivation und Beratung mit den Vorzügen eines modernen FDM vertraut zu machen, ist deswegen vielleicht sogar von noch größerer Bedeutung.

## Die Integration in Forschungsabläufe muss gelingen

Wie nützlich Werkzeuge wie ELB oder Forschungsrepositorien im Alltag von Forschenden sind, und welchen Beitrag sie zu einem verbesserten FDM leisten können, hängt stark davon ab, wie gut sie in die bestehenden Abläufe integriert werden konnten oder wie erfolgreich bestehende Prozesse mit ihrer Hilfe durch optimierte Forschungsabläufe ersetzt werden können. Der Begriff „Workflow“ taucht hier oft in vielschichtigem Gebrauch auf und meint meist nur die direkte Aneinanderreihung von Prozessschritten oder das reine Nutzungserlebnis eine Bedienoberfläche (UI). Aber bei einer Gestaltung von FDM-Prozessen sind oft gerade wenig technische Aspekte für eine breite Akzeptanz entscheidend. Welche Zugriffsmöglichkeiten bestehen, wer Inhalte unkompliziert einsehen kann, ob bestehende Richtlinien zu Dokumentation mit neuen Werkzeugen kompatibel sind, sind nur einzelne Beispiele für Fragen, die oft übersehen werden. Teilweise wird auch der hohe Aufwand, der mit der Optimierung von bestehenden Prozessen verbunden wird, gescheut und deswegen vollständig auf ein verbessertes FDM verzichtet. Im Verbundprojekt besteht mit dem INF-Projekt deshalb eine besondere Chance, diese Herausforderungen zu überwinden, da explizit Ressourcen und Expertise für die Forschenden verfügbar sind, um für sie den Aufwand einer Umgestaltung der eigenen Abläufe zu reduzieren. Damit kann sich auch eine Strahlkraft entwickeln, da die mit Unterstützung des INF-Projekts entwickelten Abläufe und die Integration von Werkzeugen natürlich auch außerhalb des Forschungsverbundes als Erfahrungen und Ergebnisse verschiedenen Institutionen zur Verfügung stehen. Das INF-Projekt leistet deshalb besondere Unterstützung beim Entwurf von gemeinsamen Prozessen und einer gemeinsamen Beschrei-

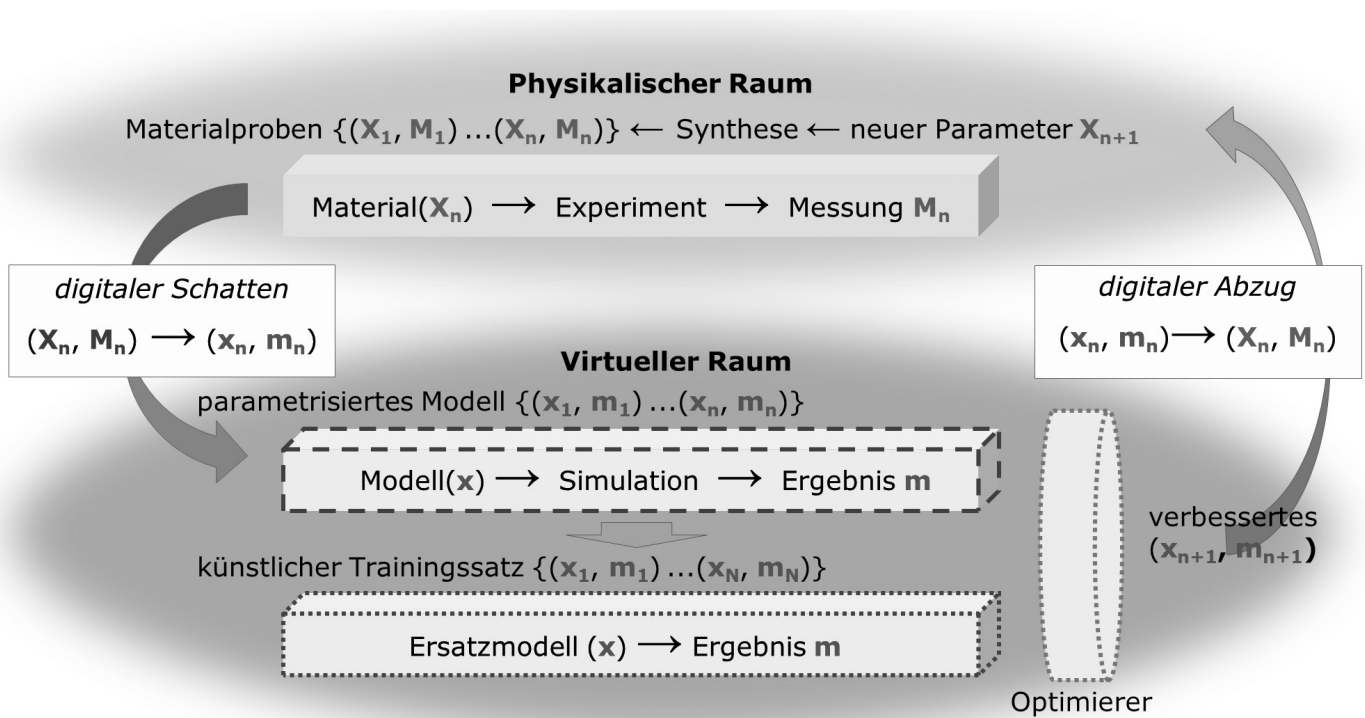
bung von Forschungsdaten. Diese geht über die Gestaltung von rein technischen Schnittstellen und Werkzeugen hinaus. Ein wichtiger Schwerpunkt ist das Koordinieren und Normieren von Forschungsabläufen. Dazu wurde unter anderem eine interne Workshop-Reihe gestaltet, in der die beteiligten Projekte auf Basis der Nutzung des ELB und der bis dahin erhobenen Forschungsdaten gemeinsame Formate, Beschreibungen und minimal notwendige Metadatensätze finden und abstimmen konnten. Diese Inhalte können dann durch das INF-Projekt formalisiert werden, beispielsweise in Form eines kontrollierten Vokabulars für Stichworte oder der Serialisierung von Metadatenschemata in allgemeinen Formaten wie JSON-LD.

Dem Angleichen von Formaten und Datentypen für Daten und Metadaten sowie der Interoperabilität von Forschungsabläufen kommt eine hohe Bedeutung zu. Um wissenschaftliche Daten für einen synergetischen Mehrwert zu vernetzen, ist dies unerlässlich. Im SFB/TRR270 ist dies aufgrund der Heterogenität der Forschung durchaus eine Herausforderung, zu deren Lösung aber mit einem Glossar sowie der Anpassung von gemeinsamen Abläufen im ELB bereits erste Schritte getan sind. Die Konzeption und Gestaltung von Forschungsabläufen benötigt dabei immer fachspezifisches Expert\*innenwissen. Eine reine Vorgabe allein durch infrastrukturell geprägte Teams, wie es auch das INF-Projekt ist, sind meist nicht erfolgreich. Derartige Strukturen sind meist nicht praxistauglich, selbst wenn sie allgemein anerkannte „Best Practices“ umsetzen. Entwicklungen unter Beteiligung von Forschenden, die ihren tatsächlichen Bedarf und ihre praktischen Erfahrungen im Umgang mit Forschungsdaten einbringen können, haben weit bessere Erfolgsaussichten. Im Verbundprojekt ist die Rolle des INF deswegen auch in unterstützender Funktion ausgearbeitet.

Die aus diesem Konstrukt resultierenden Prozesse sind den Forschenden besser zugänglich, da sie mit ihrem individuellen Praxisbezug gestaltet wurden. Dadurch erfahren alle Werkzeuge und das FDM selbst eine höhere Akzeptanz, da der Hintergrund und der Nutzen zugänglicher sind, was die Motivation der Forschenden erhöht. Schon allein die interdisziplinäre Arbeit an FDM-Abläufen strahlt so über Veranstaltungen, Veröffentlichungen und die allgemeine Verbreitung über das Verbundprojekt hinaus. Erworbenes Wissen und Sensibilität für Prozesse werden in die Institutionen und in die breitere wissenschaftliche Gemeinschaft mitgenommen und schaffen so eine breite Zunahme an Kompetenz. Der Aufwand, Prozesse und Abläufe zu koordinieren und gemeinsam zu verbessern, zahlt sich also an mehreren Stellen aus und führt fast zwangsläufig zu einer besseren Datenqualität. Gut dokumentierte Daten in Verbindung mit reichhaltigen Metadaten sind dann letztlich ein entscheidender Baustein für die Nachnutzung von Forschungsdaten. Diese „FAIRen“ Forschungsdaten bieten nicht nur Potential erhöhter Aufmerksamkeit und Verbreitung als Selbstzweck, sondern öffnen auch den Pfad zu modernen Verfahren wie der Algorithmen-gestützten Analyse oder Techniken wie dem „Deep Learning“. Erst dadurch werden schließlich Konzepte wie der digitale Zwilling in der Praxis greifbar.

## Wissenschaftlicher Mehrwert durch vernetzte Daten

Unser Wissen von der realen Welt basiert letztendlich aus der Vernetzung von Information von Teilen derselben. Dieser Gedanke bietet einen Ansatz zur Virtualisierung von realen Objekten und Prozessen, indem komplexe Strukturen auf ihre einfacheren Komponenten und deren Wechselwirkung zurückgeführt werden. Im SFB/TRR 270



(2) Das Konzept des virtuellen Zwillings (digital twin) bildet, im Kontext dieses SFB, physische Proben mit der Nummer  $n$ , den Synthese bedingten Parametern  $X_n$  und den resultierenden messbaren Eigenschaften  $M_n$  auf eine digitale Repräsentation, dem sogenannten digitalen Schatten (digital shadow), ab. Es wird vorausgesetzt, dass zwischen den Größen  $X_n$  und  $M_n$  ein gewisser funktionaler Zusammenhang besteht, d.h. dass die Eigenschaften über die Parameter gesteuert werden können. Dabei sind  $x_n$  und  $m_n$  die digitalen Abbilder der Parameter bzw. Eigenschaften in der virtuellen Welt als numerische Darstellung. Des Weiteren muss der funktionale Zusammenhang zwischen virtuellen Parametern  $x_n$  und Eigenschaften  $m_n$  mit einer Simulation erfassbar sein, wodurch die digitale Simulationsberechnung die physikalische Messung ersetzt. Werden nun Parameter mit resultierenden optimalen Eigenschaften gesucht, kann dies im virtuellen Raum geschehen (möglich ist auch die Verwendung eines mit Simulationsdaten trainierten Ersatzmodells) und das Ergebnis der Optimierung als digitaler Abzug (digital trigger) in die Realität zurück übertragen werden, was über die Synthese zu einer neuen verbesserten Probe führen kann.

Quelle: eigene Darstellung

sind die betrachteten Objekte Materialien, die auf kleineren Ebenen aus vergleichsweise einfachen Komponenten zusammengesetzt sind. Die Prozesse sind hier die Synthese, Umformungen und Messungen von Eigenschaften dieser Materialien. Das Konzept des virtuellen Zwillinges angewandt auf das Verbundprojekt versucht, mittels digitaler Repräsentation der über deren magnetischen und mechanischen Eigenschaften definierten Materialien, die Prozesse Synthese und Umformung durch computergestützte Simulationen zu ersetzen. Das Ergebnis der Simulation sind die Eigenschaften des prozessierten Materials, womit der realen Messung auf der virtuellen Seite ein Auslesen der Simulationsergebnisse entspricht (Abb. 2).

Dadurch, dass im ELB die Daten für die Experimente mit den behandelten Materialien und diese mit deren Komponenten vernetzt zur Verfügung stehen, wird es im Prinzip möglich, die Virtualisierung gemeinsam auf Objekt- und Prozessebene durchzuführen, da die einem Prozess entsprechende Simulation aus dem im ELB vorliegenden „vernetzten Wissen“ (Daten) heraus gestartet werden kann. Darauf aufbauend soll es damit ermöglicht werden, statt einzelner unabhängiger Experimente und Simulationen mit den gewonnenen vernetzten Daten einen geschlossenen Trainingssatz zu erstellen. Dadurch wird es möglich, die hochdimensionale nichtlineare Abbildung zwischen Materialzusammensetzung (inklusive der Mikro-/Nanostruktur) auf deren magnetischen Eigenschaften mittels eines Deep Neural Networks zu approximieren. Eine virtuelle Extrapolation der Versuchsergebnisse würde es erlauben, interessante Kandidaten von Hochleistungsmagneten mit dem vernetzten Wissen, das aus einzelnen experimentellen Daten generiert wurde, vorherzusagen. Die Entwicklung von digitalen Zwillingen und datenbasierten Vorhersagen stößt bereits in vielen Bereichen der Materialwissenschaft

auf großes Interesse und wird in Zukunft für noch mehr Forschungsfelder relevant werden. Aktuell sind die Modelle und Anwendungen dabei in der Regel noch entkoppelt, das heißt jede Institution arbeitet mit eigenen Datensätzen oder liest Datensätze aus dritten Quellen nach oft umfassender manueller Anpassung in die eigenen Systeme ein. Weiter sind Inhalte meist nach Disziplinen, wie Maschinenbau, Chemie, Physik, oder Methoden, wie Dichtefunktionaltheorie (DFT), Röntgenpulverdiffraktometrie (XRD) oder Verfahren der mechanischen Charakterisierungen sortiert. Ein erweitertes modernes FDM kann hier mittels semantischer Technologien wie Resource Description Framework (RDF) und der Entwicklung von gemeinsamen Standards für Formate und Datentypen innerhalb von Fachgemeinschaften oder dem NFDI e.V. in Zukunft einen wertvollen Beitrag leisten. Durch fachspezifische maschinenlesbare Metadaten können Repositorien ausgebaut und qualitativ umfassend beschriebene Datensätze direkter in Prozesse wie dem Deep Learning integriert werden, was den Weg zu einer echten semantischen Datenökologie in der Materialwissenschaft bereitet.

### **Vernetzte Forschungsdaten sind ein gemeinsames Ziel**

Digitale Forschungsdaten und die damit verbundenen Werkzeuge und Infrastruktur spielen schon heute eine wichtige Rolle in der modernen Forschung. Speziell interdisziplinäre Verbundforschung stellt dabei zum Teil besondere Anforderungen. Diese liegen dabei nicht nur im Bereich des Aufbaus und Betriebs von Soft- und Hardwarelösungen als Werkzeuge für ein verbessertes FDM, sondern vor allem im Bereich der Entwicklung von gemeinsamen Forschungsabläufen und Leitlinien. Durch Synergien und Nachnutzung können hier Ressourcen zu Optimierung des institutionellen FDM

besonders effizient genutzt werden, wovon vor allem junge Forschende profitieren. Das INF-Projekt kann dabei als zentrale unterstützende Dienstleistung für das gesamte Verbundprojekt einen klaren Mehrwert bieten.

Ein modernes FDM bildet schlussendlich ein wichtiges Element in einer erfolgreichen Digitalisierungsstrategie der Hochschulen und sichert langfristig eine hohe Datenqualität. Als Grundstein für eine zunehmende Vernetzung von Forschungsdaten, die wiederum verschiedene Methoden der computergestützten Forschung erst ermöglicht, ist es aus der zukünftigen Forschung nicht mehr wegzudenken. Aktuell stellen zwar Teile der Umsetzung noch eine Herausforderung dar, allerdings zeigen sich schon jetzt Vorzüge und Erfolge einer erfolgreichen Integration mit den Forschungsabläufen. In Verbindung mit einem Klima der forschungspolitischen Initiative öffnet es Raum für eine Vision einer wissenschaftlichen Datenökologie, in der „FAIRer“ Umgang mit vernetzten Forschungsdaten, von der Erhebung, dem Verarbeiten, dem Speichern von Daten über das Archivieren, das Veröffentlichen sowie das Publizieren und Nachnutzen von Daten, einen umfassenden gesellschaftlichen Mehrwert bringt.



## Summary

Digital research data and the associated tools and infrastructure already play an important role in modern research. In particular, interdisciplinary collaborative research has special requirements. These are not limited to the area of building and operating software and hardware solutions as tools for improved research data management. It is even more important that joint research processes and guidelines be developed. Thanks to synergies and reuse, resources for optimizing institutional research data management can be used particularly efficiently, and especially early career researchers benefit from this. In this context, the INF project offers clear added value as a central supporting service for the entire collaborative project.

Ultimately, modern research data management constitutes an important element of a successful digitalization strategy for universities, ensures high data quality in the long term and makes various methods of computer-aided research possible in the first place. As a cornerstone for increased networking of research data, it is impossible to imagine future research without it. Although parts of the implementation are still a challenge, the benefits and gains of a successful integration with the research processes are already becoming apparent. In combination with a climate of progress in research policy, it opens up space for a vision of a scientific data environment in which the 'FAIR' handling of networked research data – from the collection, processing and storage of data to its archiving, publishing and reuse – brings comprehensive added value to society.

## Anmerkungen/Literatur

- [1] Tolle, K. M.; Tansley, D. S. W.; Hey, A, J. G. (2011): The Fourth Paradigm: Data-Intensive Scientific Discovery [Point of View], Proceedings of the IEEE, 99 (8): 1334–1337, <https://doi.org/10.1109/JPROC.2011.2155130>.
- [2] Draxl, C.; Scheffler, M (2018): NOMAD: The FAIR concept for big data-driven materials science, MRS Bulletin, 43: 676–682, <https://doi.org/10.48550/arXiv.1805.05039>.
- [3] Wilkinson, M. D.; Dumontier, M.; Mons, B. (2016): The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data 3, <https://doi.org/10.1038/sdata.2016.18>.
- [4] OECD (2015): Making Open Science a Reality, OECD Science, Technology and Industry Policy Papers, 25, <https://doi.org/10.1787/5jrs2f963zs1-en>.
- [5] SPARC Europe (2017): The Open Data Citation Advantage, <https://sparceurope.org/open-data-citation-advantage/>, 27.07.2022.
- [6] research data lifecycle, <https://www.reading.ac.uk/research-services/research-data-management/about-research-data-management/the-research-data-lifecycle>, 28.07.2022.
- [7] Grönewald, M.; Niekamp R.; Schröder J.; Gutfleisch O. (2022) Forschungsdatenmanagement für ein interdisziplinäres Verbundprojekt, in: Heuveline, V. und Bisheh, N. (Hrsg.): E-Science-Tage 2021: Share Your Research Data, 249–255, <https://doi.org/10.11588/heibooks.979.c13734>.
- [8] tuDMO, <https://tudmo.ulb.tu-darmstadt.de/>, 28.07.2022.
- [9] TUDatalib-Repository, <https://tudatalib.ulb.tu-darmstadt.de/>, 28-07.2022.
- [10] elabFTW – a free and open source electronic lab notebook, Nicola Carpi, <https://www.elabftw.net>, 28.07.2022.
- [11] eScience Infrastruktur des ZIM, <https://www.uni-due.de/zim/forschung/escience.php>, 28.07.2022.
- [12] Lichtenberg-Hochleistungsrechner, <https://www.hhhr.tu-darmstadt.de/>, 28.07.2022.

## Die Autoren

**Rainer Niekamp** ist wissenschaftlicher Mitarbeiter des Instituts für Mechanik der Universität Duisburg-Essen. Er studierte 1986 bis 1993 an der Leibniz Universität in Hannover Mathematik und Informatik mit den Schwerpunkten Topologie und Stochastik. Von 1994 bis 2000 war er wissenschaftlicher Mitarbeiter am Institut für Numerische Mechanik und promovierte dort 2001 unter Professor Erwin Stein mit dem Titel „Verteilte Algorithmen für h-, p- und d-adaptive Berechnungen in der nichtlinearen Strukturmechanik“. Nach einer zweijährigen Zeit als Chef-Entwickler von Internet-Anwendungen in einem Startup-Unternehmen der Universität Hannover kehrte Rainer Niekamp 2002 bis 2014 in den wissenschaftlichen Betrieb als akademischer Rat am Institut für wissenschaftliches Rechnen der Technischen Universität Braunschweig zurück. In verschiedenen DFG- und EU-geförderten Projekten fanden seine Forschungsschwerpunkte multivariate stochastische Modelle, parallele Algorithmen und komponentenbasierte Softwareentwicklung Anwendungen. Von 2014 bis 2016 entwickelte Rainer Niekamp für eine aktuariellen Gesellschaft Algorithmen zur Gefahrenabschätzung von Naturgewalten. Seit 2016 ist er am Institut für Mechanik in Forschung und Lehre tätig. Seine Forschungsschwerpunkte liegen hier in der Entwicklung und Anwendung von Simulationen im Bereich der angewandten Mechanik. Seit 2020 betreut Rainer Niekamp zusammen mit seinem Darmstädter Kollegen Matthias Grönewald das Z-INF Projekt des SFB/TRR270 HoMMage.

**Matthias Grönewald** ist wissenschaftlicher Mitarbeiter der TU Darmstadt, dabei ist er zu gleichen Teilen Angehöriger des Instituts für Materialwissenschaft sowie der Universitäts- und Landesbibliothek. Er studierte 2003 bis 2009 an der Philipps-Universität Marburg Chemie mit den Schwerpunkten physikalische Chemie und Polymerphysik. Von 2010 bis 2013 war er wissenschaftlicher Mitarbeiter am Deutschen Kunststoff-Institut (seit Juli 2012 Fraunhofer LBF) und promovierte 2013 mit dem Thema „Ionenkomposite und Ionensiloxanopolymere und ihre Eignung als DEA“. 2013 bis 2014 betreute er verschiedene Forschungsprojekte im Bereich der polymeren Materialwissenschaften als wissenschaftlicher Mitarbeiter des Eduard-Zintl-Instituts für Chemie der TU Darmstadt, um sich 2016 bis 2019 weiter in Richtung wissenschaftlicher EDV-Dienstleistungen zu profilieren und als selbstständiger IT-Dienstleister tätig zu sein. Darauf aufbauend kehrte er 2020 zur TU Darmstadt in den Bereich Forschungsdatenmanagement zurück und beschäftigt sich vor allem mit den Themenfeldern elektronische Laborbücher und Metadaten sowie Infrastrukturlösungen für ein modernes, bedarfsgerechtes Forschungsdatenmanagement. Unter diesem Aspekt betreut er auch zusammen mit seinem Kollegen Rainer Niekamp von der Universität Duisburg-Essen seit 2020 das Z-INF Projekt des SFB/TRR270 HoMMage.



Matthias Grönewald. Foto: Vladimir Unkovic

# DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT  
DUISBURG  
ESSEN

*Offen im Denken*

ub | universitäts  
bibliothek

Dieser Text wird via DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt. Die hier veröffentlichte Version der E-Publikation kann von einer eventuell ebenfalls veröffentlichten Verlagsversion abweichen.

**DOI:** 10.17185/duepublico/81683

**URN:** urn:nbn:de:hbz:465-20240410-100722-7

Erschienen in: UNIKATE 60 (2024), S. 88-97

Alle Rechte vorbehalten.