











Mit AIMS zu einem Metadatenmanagement 4.0: FAIRe Forschungsdaten benötigen interoperable Metadaten

Matthias Grönwald ¹, Patrick Mund ², Matthias Bodenbenner ², Marc Fuhrmans ¹, Benedikt Heinrichs ³, Matthias S. Müller ³, Peter F. Pelz ⁴, Marius Politze ³, Nils Preuß ⁴, Robert H. Schmitt ^{2,5}, und Thomas Stäcker ¹

¹Universitäts- und Landesbibliothek, TU Darmstadt

²WZL | RWTH Aachen University

³IT Center, RWTH Aachen University

⁴Institut für Fluidsystemtechnik, TU Darmstadt

⁵Fraunhofer-Institut für Produktionstechnologie IPT

Gute wissenschaftliche Praxis erfordert eine präzise und verständliche Dokumentation der Ergebnisse. Dies ist umso wichtiger, wenn Forschende ihre eigenen Forschungsdaten teilen und publizieren oder archivierte Daten Dritter nachnutzen möchten. Forschungsdatenmanagement (engl. research data management) auf Basis weitreichend standardisierter Metadaten ist daher von essenzieller Bedeutung. Standardisierte Metadaten sollen in strukturierter und maschinenlesbarer Form Informationen zu Entstehung, Inhalt und Kontext der beschriebenen Forschungsdaten liefern. Im Zuge der ingenieurwissenschaftlichen Forschung zu Industrie 4.0 werden enorme Mengen heterogener Daten generiert. Ein entsprechendes “Metadatenmanagement 4.0” soll es Forschenden ermöglichen, aus einer Vielzahl bereits vorhandener digitaler Datensätze den richtigen Datensatz für die eigene Forschung zu selektieren. Entsprechend der FAIR-Prinzipien müssen Metadaten für Mensch und Maschine interpretierbar sein.

Im Forschungsprojekt „Applying Interoperable Metadata Standards (AIMS)“, gefördert von der Deutschen Forschungsgemeinschaft (DFG), werden die Herausforderungen an ein Metadatenmanagement 4.0 für die ingenieurwissenschaftliche Forschung adressiert. Der Projektfokus liegt auf dem Maschinenbau und verwandten Disziplinen. Das interdisziplinäre Team, verteilt über mehrere Institutionen aus Infrastruktur und Wissenschaft, schafft hier eine Plattform, die es Forschenden ermöglicht, Metadaten schemata zu erstellen und zu teilen. Durch ein Modellierungskonzept, das auf Vererbung und Modularität setzt, kann eine hohe Spezifität bei maximaler Anwendbarkeit und Nachnutzbarkeit der Metadaten schemata erreicht werden. So wird die Akzeptanz der Forschenden erhöht strukturierte Metadaten in ihre Forschungsabläufe zu integrieren, um mit zunehmender Verbreitung den Weg zu gemeinsamen Metadatenstandards zu bereiten. Das Projekt ist von Anfang an als Kooperation zwischen Infrastruktur und Forschung ausgelegt. Dadurch soll das in AIMS realisierte Metadatenkonzept direkt in die Prozesse der beteiligten Forschenden

den integriert werden. Durch die Berücksichtigung von Interoperabilität der entstehenden Metadaten schemata und -schnittstellen wird explizit der generische Transfer der Lösungen auf andere Forschungsfelder verfolgt. In diesem Beitrag werden die Anforderungen aus der Forschung aufgezeigt und der Ansatz von AIMS für ein Metadatenmanagement 4.0 vorgestellt. Zudem wird ein Einblick in infrastrukturelle Herausforderungen und mögliche technische Lösungen der aus dem Projekt AIMS hervorgehenden Metadaten-Plattform gegeben.

1 Einleitung

In die Umsetzung von experimentellen Studien (physisch oder virtuell) und die Generierung von großen Mengen an Messdaten wird viel Geld, Zeit und Expert:innenwissen investiert. Neben Publikationen und Berichten sind diese erfassten Forschungsdaten ein wichtiges Produkt wissenschaftlicher Arbeit [1]. Die langfristig nachvollziehbare Dokumentation dieser Daten sowie des Prozesses ihrer Entstehung ist ein Qualitätsmerkmal guter wissenschaftlicher Praxis und Organisationskultur. Sie eröffnet wichtige Anschlussmöglichkeiten für die weitere Forschung und Produktentwicklung. Sie ist außerdem Voraussetzung für einen nachhaltigen Transfer der Forschungsergebnisse in die Industrie bzw. die Weiterverwendung von Daten [2]. Besondere Herausforderungen für die standardisierte Dokumentation von Forschungsdaten in vielen wissenschaftlichen Disziplinen, darunter auch dem Maschinenbau, sind Heterogenität und Komplexität der Daten, sowie ihrer Entstehungsprozesse.

Bei der Durchführung ingenieurwissenschaftlicher Experimente werden i.d.R. sehr individuell konfigurierte Versuchsstände eingesetzt, die zudem meist kontinuierlich weiterentwickelt werden. Infolgedessen stehen keine standardisierten Werkzeuge wie elektronische Laborbücher (ELBs) oder Laborinformationsmanagementsysteme (LIMS), wie sie aus der Chemie und der Biologie bekannt sind, für den Bereich des Maschinenbaus zur Verfügung. Gleiches gilt für kommerzielle Softwarelösungen. Diese setzen meist starre domänenspezifische Standards voraus und versagen bei der Integration von heterogenen und sich verändernden Versuchsaufbauten, Analysemethoden oder Metadaten [3].

Dies führt zu einer hohen Heterogenität der generierten Daten, was sich in Verbindung mit den steigenden Raten der Datengenerierung zunehmend als problematisch hinsichtlich ihrer Nutzbarkeit darstellt [4].

Mitarbeiter:innen sind gezwungen einen beträchtlichen Teil ihrer Zeit aufzuwenden, um Daten aus nicht maschinenlesbaren Quellen (bspw. Grafiken oder unkommentierte Wertetabellen) zu identifizieren, zusammenzuführen und mit relevanten Datensätzen über jeweils spezielle Programmierschnittstellen zu interagieren [5].

Die Umsetzung der FAIR-Prinzipien [6] (engl. findable, accessible, interoperable, re-usable) ist unerlässlich, um Forschungsdaten nachhaltig zu generieren und vorzuhalten, bzw. zu interpretieren und zu nutzen [7, 8]. Dies gilt sowohl für die aktive Nutzung während des Forschungsprozesses als auch für die Nachnutzung nach Abschluss eines Forschungspro-

jektes. Insbesondere beim Austausch und der gemeinsamen Nutzung von Forschungsdaten über Projektgrenzen hinweg werden interoperable Standards für Metadaten und Software-Schnittstellen benötigt, um ihre Nutzbarkeit zu gewährleisten [9, 10]. Es besteht der dringende Bedarf nach der Schaffung übergreifender Plattformen, Etablierung verbindlicher Standards und einheitlicher Dienste, die es erlauben, Forschungsdaten zu sichern, optimal nach zu nutzen und zu vernetzen.

2 Ziel und Ansatz

Das Projekt „Applying Interoperable Metadata Standards“ (AIMS) hat die Zielstellung (1) eine interoperable Abbildung komplexer Experimente, Methoden und Datensätze zu erreichen und (2) gleichzeitig die nachträgliche oder manuelle Dokumentation entsprechender Metadaten zu vermeiden.

Zu diesem Zweck werden in AIMS zwei Kernergebnisse erarbeitet: Zum einen wird eine Umgebung geschaffen, die es Forschenden ermöglicht spezifische Metadaten schemata zur Beschreibung ihrer individuellen Forschungsdaten zu erstellen, zu teilen und nach zu nutzen. Sie werden so in die Lage versetzt für ihr direktes Umfeld oder ihre Fachdomäne nach und nach Quasi-Standards zu entwickeln, zu pflegen und zu etablieren. Zum anderen werden Werkzeuge und Abläufe erarbeitet, die sich in den wissenschaftlichen Forschungsalltag integrieren und so, durch eine verbesserte Effizienz der Auswertung, Bearbeitung und Dokumentation von Daten, einen direkten Mehrwert für die Forschung bieten.

Realisiert wird dies auf Basis eines modularen Modellierungsansatzes (siehe Abschnitt 3.1), nutzbar gemacht durch die zur Verfügung gestellte Infrastruktur (siehe Abschnitt 3.2) und umgesetzt durch das Einbinden von mehreren Anwendungsfällen aus der ingenieurwissenschaftlichen Forschung direkt in den Projektablauf (siehe Abschnitt 3.3). Durch diesen integrativen Dialog zwischen Forschenden und Infrastruktur werden Möglichkeiten für ein verbessertes Metadatenmanagement geschaffen, validiert und demonstriert: Spezifische Metadatenbeschreibungen erleichtern nicht nur die Dokumentation, Publikation und Archivierung von Forschungsdaten, sondern auch den Forschungsprozess als solchen. Insbesondere Forschungsprozesse mit großen Datenmengen oder einem hohen Grad an Automatisierung können hier profitieren.

Die Implementierung der Konzepte und Werkzeuge aus AIMS innerhalb der bestehenden Forschungsprozesse ist ein wesentlicher Aspekt auf dem Weg zu Metadatenstandards. In einer Art Konzeptbeweis für ausgewählte Teilbereiche der Ingenieurwissenschaften können hier die Vorteile und Potentiale aufgezeigt werden, die sich für die Forschenden durch ein erfolgreiches Metadatenmanagement bereits während der Forschung ergeben. Dies führt zu einer deutlich gesteigerten Akzeptanz in der Forschungsgemeinschaft als ganzer. Integraler Bestandteil der erfolgreichen Umsetzung des Ansatzes ist deshalb die Verbreitung der positiven Anwendungsfälle aus dem Projekt in der Forschungsgemeinschaft. Durch eine verstärkte Interaktion über bestehende Projekte und Initiativen kann so ein Weg zu gemeinsamen Standards bereitet werden.

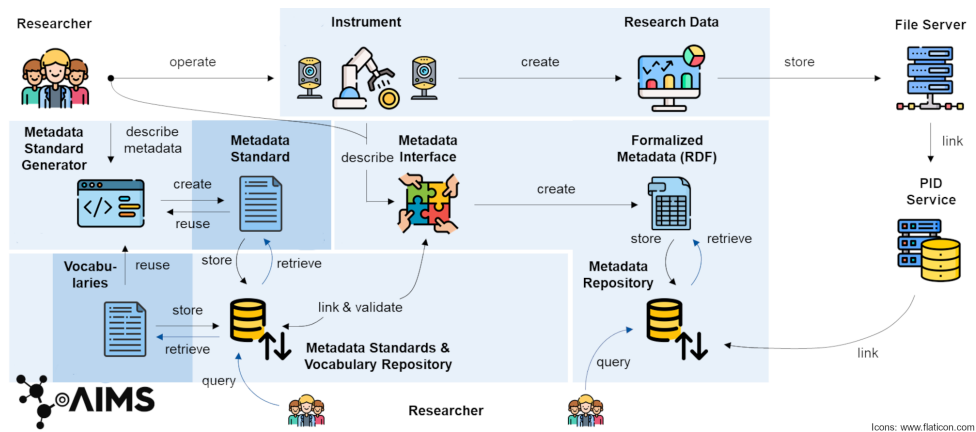


Abbildung 1: Projekt AIMS stellt vernetzte Infrastruktur zum Metadatenmanagement für die Forschenden bereit und bildet damit einen Baustein hin zu einer Linked Open Data Cloud.

3 Applying Interoperable Metadata Standards - AIMS

AIMS widmet sich den beschriebenen Herausforderungen und Lösungsansätzen seit Ende des Jahres 2020 im Rahmen einer Förderung durch die Deutsche Forschungsgemeinschaft (DFG). Von entscheidender Bedeutung ist, dass AIMS keine Initiative rein aus dem Infrastrukturbetrieb ist, sondern maßgeblich von Forschenden initiiert und getragen wird. Konkret arbeiten in einem interdisziplinären Team aus Ingenieur:innen, Informationswissenschaftler:innen und Informatiker:innen der Universitäts- und Landesbibliothek (ULB) und dem Institut für Fluidsystemtechnik (FST) der TU Darmstadt sowie dem IT Center (ITC) und dem Werkzeugmaschinenlabor WZL der RWTH Aachen University Forschende und Infrastrukturbetreibende aus vier unterschiedlichen Institutionen verteilt über zwei Universitäten gemeinsam an einer Plattform zur Erstellung und zum Teilen interoperabler Metadaten-Schemata und deren Integration in die wissenschaftlichen Abläufe.

AIMS beginnt mit der Integration in einzelnen Anwendungsfällen der beteiligten Institutionen, um von da aus kontinuierlich den Nutzerkreis durch Training und Verbreitung zu erweitern. Ausgehend von Fachgebieten, über Institutionen und Verbundforschung, dient das Projekt als Baustein in einer größeren Initiative zu einem verbesserten Forschungsdatenmanagement in der ingenieurwissenschaftlichen Forschungsgemeinschaft. Hervorzuheben ist dabei zum einen die Integration in die Nationale Forschungsdateninfrastruktur (NFDI) durch intensive Wechselwirkung speziell mit dem Konsortium Nationale Forschungsdateninfrastruktur für die Ingenieurwissenschaften (NFDI4Ing) und die Fokussierung auf offene Schnittstellen, beispielsweise das OAI Protocol for Metadata Harvesting (OAI-PMH, engl.), und der Einsatz standardisierter Abfragesprachen (siehe Abschnitt 3.2).

In AIMS sollen für die zu entwickelte Plattform auch Dienste aus anderen Projekten eingebunden werden, die die Verbreitung und Kuratierung bereits vorhandener Schemata adressieren, beispielsweise CoSciNe [11], oder die über offene Schnittstellen Ontologien



Abbildung 2: Das Konzept von Bausteinen, die von Forschenden als Module zu unterschiedlichen Einheiten zusammengefügt werden können, ist ein zentrales Element hinter AIMS.

verfügbar machen, wie der in NFDI4Ing entwickelte Terminology Service [12]. In seiner Gesamtheit wird so ein weiterer Beitrag zu den Bemühungen einer offenen und vernetzten Forschungslandschaft geleistet.

3.1 Modell

Im Umgang mit der Heterogenität und Komplexität der Untersuchungsgegenstände, Methoden und Werkzeuge in den Ingenieurwissenschaften setzt AIMS auf die flexible Erstellung und Nachnutzung fachspezifischer und anwendungsbezogener Metadatenschemata. Neben der in Abschnitt 4 beschriebenen Plattform (insbesondere dem Generator und dem Repository für Metadatenschemata) erfordert dies auch einen Modellierungsansatz für Metadaten, der Spezifität und Flexibilität auf der einen Seite mit einer hohen Nutzbarkeit, einer breiten Anwendbarkeit sowie einer maximalen Interoperabilität auf der anderen Seite vereinbart. Dazu wird in AIMS auf ein Baukastenprinzip gesetzt, das auf den folgenden Designkomponenten beruht:

Applikationsprofile

Metadatenschemata werden aus Termen, die aus kontrollierten Terminologien entnommen werden, in Form sogenannter Applikationsprofile [13] zusammengesetzt.

Dies erlaubt die flexible Definition maßgeschneiderter Metadatenschemata, die aufgrund ihres Rückgriffs auf kontrollierte Terme mit anderen Schemata, die dieselben Terme verwenden, interoperabel sind.

Hierarchische Modellierung

Die Applikationsprofile werden darüber hinaus hierarchisch modelliert, indem spezifische Profile nicht losgelöst, sondern als Subklassen generischer Klassendefinitionen definiert werden, die, wie in der objektorientierten Programmierung, die Attribute ihrer Eltern erben und um neu hinzukommende auf die Kinder passende Attribute ergänzen. Durch dieses (mehrstufige) Verfahren sind verwandte Profile stets auf Ebene des letzten gemeinsamen Vorfahrs interoperabel. Ebenso kann die Definition neuer Profile stets an dem passenden Knoten in der Hierarchie erfolgen, die sich auf diesem Weg in die Baumstruktur einfügen.

Modularität

Um die Anwendbarkeit und Nachnutzbarkeit der Applikationsprofile zu erhöhen, werden trennbare Bereiche von Metadaten durch separate Schemata beschrieben. Solche Bereiche sind z. B. Untersuchungsgegenstand, Methode und Werkzeug. Dadurch wird eine unnötige Einschränkung der Anwendbarkeit, die bei Abdeckung mehrerer dieser Bereiche in einem einzelnen Schema aufträte, vermieden. Zur vollständigen Beschreibung von Forschungsdaten werden die einzelnen Profile kombiniert, etwa durch ein Profil für einen “Processing Step”, das Instanzen sämtlicher relevanter Profile bündelt und mit In- und Output des beschriebenen Processing Steps in Verbindung setzt. Darüber hinaus sind weitere direkte Verbindungen zwischen Profilen (sowohl innerhalb eines Bereichs als auch über Bereichsgrenzen hinweg) Teil des Konzepts. So kann etwa ein Werkzeug ein Attribut “enables” besitzen, welches eine durch das Werkzeug implementierte Methode angibt.

Die Realisierung der genannten Designprinzipien erfolgt auf Basis des Resource Description Framework (RDF, siehe Abschnitt 3.2), und fügt sich zu einer Beschreibung von Forschungsdaten in Form eines Knowledge Graph zusammen.

3.2 AIMS Infrastruktur

Zentraler Bestandteil von AIMS ist eine Web-Plattform, die intuitive Werkzeuge zur Erstellung, Veröffentlichung, Weiterentwicklung und Nachnutzung von Metadatenschemata als Applikationsprofile zur Verfügung stellt. Sie basiert auf dem in Abschnitt 3.1 beschriebenen Ansatz und ermöglicht die Dokumentation von Forschungsdaten durch Publikation von Metadatensätzen, die konform zu den mittels der Plattform verwalteten Schemata sind.

In Abbildung 3 ist der geplante Workflow gezeigt, über welchen Nutzende mit der AIMS-Plattform interagieren können. Zur Verdeutlichung des Fokus auf Nachnutzung starten Nutzende in der AIMS Plattform immer bei der Suche für Applikationsprofile (siehe Abb. 3 “search”). Auf Basis der eingegebenen Werte für die Metadaten eines Applikationsprofils (z. B. “Name”, “Beschreibung”, “Fachgebiet” usw.) werden passende Applikationsprofile gefunden. Nutzende haben nun die Möglichkeit eines der gefundenen Applikationsprofile zu verwenden, zu modifizieren oder zu erweitern.

Die Modifikation und Erweiterung finden in einem sogenannten Applikationsprofil-Generator statt. Der Applikationsprofil-Generator erhält Terme von einer “vocabulary database”, die zu einem Applikationsprofil hinzugefügt werden können. Diese Datenbank ist mit den für die AIMS Anwendungsfälle relevanten Terminologien befüllt und mit bestehenden Diensten verknüpft, die existierende Ontologien zur Verfügung stellen. Ein so erstelltes Applikationsprofil wird dann in einer “application profile database” gespeichert und steht für folgende Suchen zur Verfügung. Die Applikationsprofile können anschließend als Grundlage zur Dokumentation der eigenen Forschungsdaten verwendet werden (s. Abschnitt 3.3). Zur Dokumentation von Forschungsdaten anhand der Applikationsprofile steht außerdem ein “metadata store” zur Verfügung. Dieser ist ein Repositoryum in dem mittels Applikationsprofilen validierte Metadatenätze abgelegt werden können. Dieses Repositoryum für konforme Metadatenätze ist öffentlich zugänglich und durchsuchbar.

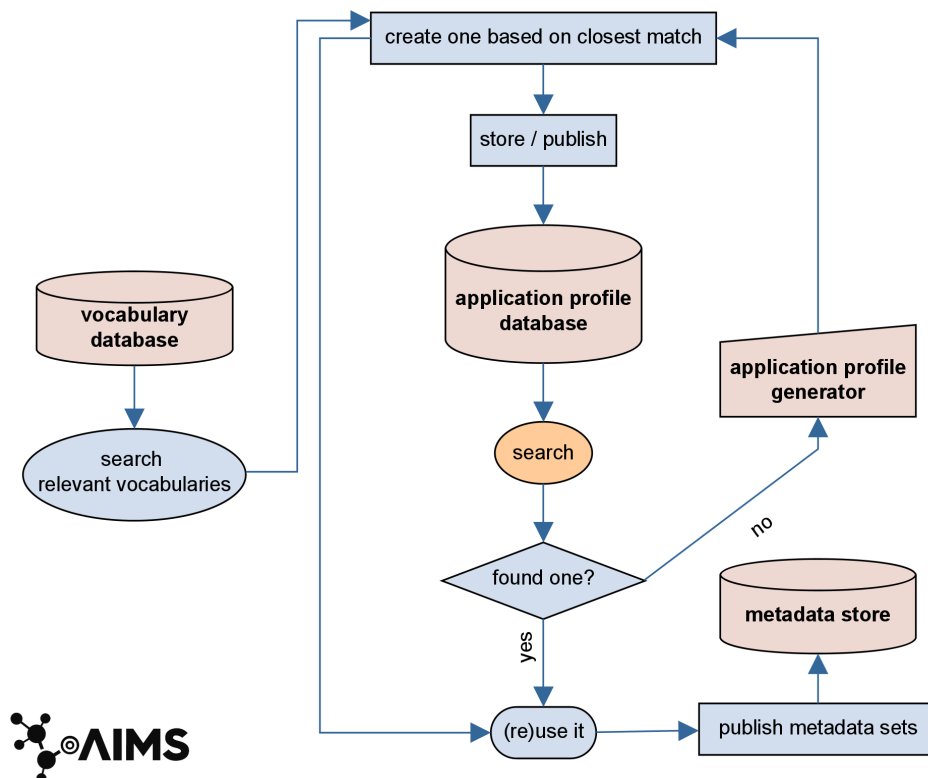


Abbildung 3: Abstrahierter Nutzer-Workflow der AIMS-Plattform.

Repräsentation von Metadaten und Applikationsprofilen

Zur Erstellung von Applikationsprofilen und Metadaten muss ein einheitliches Format gewählt werden, welches interoperabel agieren kann und welches bereits breit unterstützt wird. Auf Basis dieser Kriterien und da es ein bereits genutzter Standard für Ontologien ist, wird das Resource Description Framework (RDF) [14] zur Formulierung von Applikationsprofilen und Metadaten verwendet.

Zur Repräsentation von Applikationsprofilen ist es notwendig, eine einheitliche Beschreibung für diese zu finden. Diese sollte die Möglichkeit bieten, die gewünschte Struktur eines Metadatensatzes unter Rückgriff auf Terme aus kontrollierten Vokabularen nach dem in Abschnitt 3.1 beschriebenen Baukastenprinzip festzulegen. Zudem sollte die Beschreibungssprache bereits verbreitet sein und eine Notation in RDF möglich sein. Auf dieser Basis wird in AIMS die Shapes Constraint Language (SHACL) [15] genutzt, welche einen Standard zur Beschreibung von Applikationsprofilen in RDF darstellt. Zu deren Validierung stehen Implementierungen in den populärsten Programmiersprachen wie C#, JavaScript oder Python zur Verfügung. Mit SHACL ist es möglich für einzelne Attribute spezifische Einschränkungen zu beschreiben, sodass Applikationsprofile auch die Konformität der Metadaten mit bestehenden Definitionen validierbar machen können bspw. einen Wertebereich, ein Zahlenformat oder auch eine Klasse von Applikationsprofilen.

Applikationsprofil-Generator

Das Erstellen von Applikationsprofilen durch Auswahl geeigneter Terme aus kontrollierten Vokabularen wird für den Nutzer intuitiv durch eine grafische Oberfläche, die die Suche nach Termen und deren Zusammenfügen per Drag&Drop ermöglicht, ohne eine Auseinandersetzung mit der exakten Syntax von RDF und SHACL zu erfordern. Dieser Generator stellt eine der Hauptkomponenten von AIMS dar. Je nachdem, ob Forschende ein bereits bestehendes Applikationsprofil erweitern bzw. modifizieren oder ein neues Applikationsprofil erstellen möchten, starten sie mit einem jeweiligen Stand. Bei einer Erweiterung wird sich auf Vererbung gestützt, welche durch das RDF-Schema (RDFS) [16] mit dem Attribut “rdfs:subClassOf” gewährleistet wird.

Im Applikationsprofil-Generator besteht die Möglichkeit für Nutzende nach Terminologien aus Vokabularen zu suchen. Diese Vokabulare stammen aus einer “vocabulary database” und können auf Basis der AIMS Infrastruktur je nach Implementierung und Anwendungsfall verschieden sein. Die Ergebnisse der Suche werden je nach Relevanz, Benutzung und anderer Kategorien sortiert und Nutzenden aufbereitet dargestellt. Sie können nun mithilfe von Drag&Drop die einzelnen gewünschten Terme in ein Applikationsprofil hineinziehen und so zu diesem hinzufügen. Weiterhin können Nutzende ihrem Applikationsprofil und seinen Termen verschiedene Attribute geben, die die Struktur weiter verfeinern und etwa die Kardinalität, mit “sh:minCount” und “sh:maxCount” darstellt, oder den zulässigen Wertebereich festlegen. Über den Wertebereich kann dabei auf weitere Applikationsprofile verwiesen werden, sodass das in Abschnitt 3.1 beschriebene Prinzip der Modularität umgesetzt werden kann.

Sobald ein Applikationsprofil zur Zufriedenheit erstellt wurde, müssen zur Auffindbarkeit und Nachnutzung des Applikationsprofils relevante Metadaten, wie “Name”, “Beschreibung” oder “Fachgebiet”, eingetragen werden. Erst danach kann das Applikationsprofil in der Datenbank veröffentlicht werden.

Applikationsprofil Datenbank

Neben dem Generator stellt die Applikationsprofil-Datenbank ein weiteres zentrales Element der AIMS-Plattform dar. Die dort gespeicherten Applikationsprofile sind über einen kombinierten Ansatz durchsuchbar. Dieser kombinierte Ansatz ergänzt einen aus Titel und Beschreibung gespeisten Suchmaschinenindex zur Textsuche mit Filteroptionen. Die Filterung erfolgt über die zu jedem Applikationsprofil erfassten Metadaten und erlaubt es außerdem Informationen aus den Beziehungen zwischen Applikationsprofilen zur Navigation und weiteren Eingrenzung zu verwenden. Dadurch wird verstärkt die Nachnutzung von Applikationsprofilen angeregt, was Parallelentwicklungen vorbeugt. Gleichzeitig können gewonnene Erkenntnisse aus der Nutzungs- und Suchstatistik für eine kuratierte Terminologieentwicklung im Rahmen von NFDI4Ing genutzt werden.

Um die notwendige Freiheit zur Gestaltung von individuellen, spezifischen Applikationsprofilen zu bieten bleibt die Verantwortung für eine Qualitätssicherung und Pflege der Inhalte der AIMS Infrastruktur letztlich bei den Nutzenden. Nicht zuletzt um dies zu erleichtern stellt die Plattform ein Rechte- und Rollenmodell zur Verfügung.

3.3 Standardisierte Metadaten in der Praxis

Disziplinspezifische Metadaten-Schemata sind eine Schlüsselkomponente zur Standardisierung von Metadaten und eröffnen neue Möglichkeiten zur Optimierung der Datengenerierung und -analyse während der aktiven Forschung. Allerdings ist hinsichtlich der Akzeptanz bei den Forschenden darauf zu achten, dass die Nutzung von disziplinspezifischen Schemata nicht als einschränkend aufgefasst wird und mit möglichst geringem Mehraufwand verbunden ist.

In AIMS wird daher ein integrativer Forschungs- und Entwicklungsansatz verfolgt, bei dem der in Abschnitt 3.1 beschriebene Modellierungsansatz und die zu entwickelnde Plattform an verschiedenen Prüfständen am FST der TU Darmstadt sowie am WZL der RWTH Aachen praktisch angewendet und iterativ verbessert wird. Die Entwicklung der Plattform wird also aktiv von den Forschenden mitgestaltet und getragen. Dabei ist die aufwandsarme Integration einer strukturierten Dokumentation von Forschungsdaten (in Form von durch Applikationsprofile beschriebener Metadaten) in den Forschungsalltag zentraler Bestandteil des Projekts. Im Vordergrund stehen bei der Entwicklung außerdem, wie durch die entwickelte Plattform und die dort verwalteten Applikationsprofile die Generierung und -analyse von Metadaten vereinfacht und automatisiert werden kann.

Etablierung von Quasistandards für die Generierung von Metadaten

Die Plattform von AIMS unterstützt Forschende bei der Standardisierung der Dokumentation von Prüfständen und Messdaten. In der Praxis müssen die über die Plattform generierten Schemata dafür auf eine Vielzahl unterschiedlicher, disziplinspezifischer Prüfstände anwendbar sein und gleichzeitig eine hohe Benutzerfreundlichkeit aufweisen.

Hier besteht das Risiko, dass sich diese beiden Anforderungen diametral gegenüberstehen. Durch den partizipativen Ansatz, der die aktive Beteiligung der Forschenden fordert und fördert, kann den Herausforderungen, die sich insbesondere aus der Heterogenität des Anwendungsfeldes ergeben, Rechnung getragen werden: Anstatt eine große, aber zwangsweise unvollständige Menge definierter Schemata vorzugeben, ermöglicht AIMS den Forschenden disziplinspezifische, individualisierte Metadatenschemata auf Basis kontrollierter Bausteine und unter Rückgriff auf bereits durch andere Forschende erstellte Schemata eigenständig zu definieren. Durch diese Wiederverwendung und Weiterentwicklung wird in einem evolutionären Prozess die Etablierung von Quasi-Standards unterstützt. D.h. De-Facto-Standards, die sich in der Praxis durchgesetzt haben, aber keinen formalen Prozess zur Erlangung eines Konsenses durchlaufen haben [17].

Ein solcher De-facto-Standard kann durch ein formelles Normungsverfahren in eine weitreichend anerkannte Norm überführt werden (beispielsweise PDF, HTML). Als anerkannte Regeln der Technik können diese so regulierende Bedeutung in Forschung und Produktion erlangen. Durch die in Abschnitt 3.1 beschriebene Modularität und Hierarchie kann dabei ein hohes Maß an Wiederverwendbarkeit bei maximaler Interoperabilität gewährleistet werden.

Zudem ermöglicht ein verfügbares Metadatenschema, Forschungsabläufe so zu gestalten, dass sie die größtenteils automatisierte Generierung von Metadaten unterstützen. Dies gewährleistet die Qualität der Metadaten hinsichtlich des abgebildeten Prüfstandes und der tatsächlichen Messdaten. Benutzereingaben wie Informationen über den Benutzer, das Projekt oder Steuerungsparameter können ebenso adressiert werden. Software, die Forschungsdaten generiert, muss gegebenenfalls beide Wege (manuell und automatisiert) zur Generierung von Metadaten in die Funktionalität jedes ihrer Module integrieren. So ist jedes Modul verantwortlich für die Protokollierung seiner entsprechenden Metadaten. Dadurch wird die Notwendigkeit einer nachträglichen Zuordnung vermieden und Informationsverlust minimiert. Der Prozess wird unterstützt durch grundlegende Kommunikations- (I-O-Link-Kommunikation), Steuerungs- und Online-Verarbeitungsmechanismen.

Automatisierung bei der Datenanalyse zur Ergebnisinterpretation

Über die Dokumentation der Forschungsdaten hinaus, liegt der Nutzen einer standardisierten Dokumentation in der Vereinfachung von Prozessen für Abruf, Kuration und Auswertung von Forschungsdaten. Durch die Verwendung definierter Terminologien sind Forschungsdaten, die durch einheitliche Schemata gemäß der FAIR-Prinzipien dokumentiert sind, leichter auffindbar, verständlich und wiederverwendbar. Die in AIMS verwen-

deten Methoden und Technologien garantieren die Maschinenlesbarkeit der Metadaten, sodass Prozesse zur Verarbeitung und Analyse von Forschungsdaten leicht automatisiert werden können. Durch den Export dieser Metadaten in das in AIMS entwickelte Repository können die Indizierungs- und Suchfunktionen des Repositoriums genutzt werden, um selektiv alle Datensätze abzurufen, die gewünschte Kriterien erfüllen. Dies erleichtert die Wiederverwendung vorhandener Daten, auch außerhalb des Projekts in dem sie entstanden sind. Darüber hinaus entsteht durch die einheitliche, mittels Applikationsprofilen standardisierte Dokumentation von Arbeitsabläufen und resultierenden Datenprodukten ein vernetzter Wissensgraph, der das generierte Wissen in einer nachvollziehbaren und verständlicheren Weise bewahrt, wovon andere Projekte in hohem Maße profitieren.

Außerdem können durch die standardisierte Beschreibung der Metadaten nicht nur die (Meta-) Daten selbst, sondern auch Programme, die auf diesen arbeiten, wiederverwendet werden. AIMS ermöglicht es somit, selbst entwickelte Programme nachhaltiger zu gestalten und die Routineschritte in der Datenhandhabung und -analyse wie die Verarbeitung von Rohdaten oder die Quantifizierung der Unsicherheit wiederverwendbar zu automatisieren. Dadurch kann redundanter Code vermieden werden, der üblicherweise bei der Verarbeitung nicht standardisierter Daten entsteht. Die Analysewerkzeuge werden durch breitere Anwendung umfassender und insgesamt weniger fehleranfällig.

4 Zusammenfassung

Es wurde dargelegt, dass die ingenieurwissenschaftliche Forschung mit ihren oft sehr heterogenen Themengebieten, Forschungsfragen und Untersuchungsgegenständen besondere Herausforderungen an die Gestaltung von Metadatenbeschreibungen stellt. Maschinenlesbare Metadatenansätze stehen zunehmend im Fokus als Lösungsansatz unter anderem für den Umgang mit zunehmenden Datenmengen, auch im Kontext der Industrie 4.0. Beides erzeugt Anforderungen, für die bisher keine anwendbaren, etablierten Lösungen gefunden werden konnten.

In AIMS wird ein aus der Forschung getriebener Ansatz verfolgt, der diese Anforderungen zielgenau adressiert. Zentrale Elemente sind dabei ein Modellierungsansatz, der flexible und maximal interoperable Metadaten schemata in Form hierarchischer und sich modular referenzierender Applikationsprofile erlaubt, die auf Basis kontrollierter Terminologien im Baukastenprinzip trotzdem hochspezifische Beschreibungen ermöglichen. Dieser Ansatz wird in eine Web-Plattform überführt, die unter einer benutzerorientierten Oberfläche ein Repository für fachspezifische Applikationsprofile zur Nachnutzung und einen Generator zu deren Erzeugung und Weiterentwicklung bereitstellt. Diese wird ergänzt um ein Repository, in dem den Applikationsprofilen entsprechende validierte Metadatenansätze veröffentlicht werden können. Forschenden wird so ein niederschwelliger Zugang zu einem verbesserten Metadatenmanagement ermöglicht.

Ebenso wichtig wie die Erzeugung fachspezifischer Applikationsprofile und Terminologien sind Wege, die es erlauben, sie in die praktische Forschung zu integrieren. Zu diesem Zweck


werden in den beteiligten Forschungsgruppen Verfahren entwickelt, um die Erzeugung und Nutzung standardisierter Metadaten zu optimieren.

Nicht zuletzt ist es ein wichtiges Anliegen von AIMS die Projektergebnisse in der Ingenieur- und FDM-Community zu vermitteln und einen stetigen Austausch zu den Projektergebnissen und dem entstandenen Wissen zu initiieren, um über vernetzte Initiativen und Projekte die verbundenen Institutionen und letztlich eine größtmögliche Forschungs- und Anwendergemeinschaft im Umfeld der Ingenieurwissenschaften zu erreichen. Zusammen sollen diese Schritte und Maßnahmen ein verbessertes Metadatenmanagement 4.0 gewährleisten und durch AIMS gemeinsam eine Umgebung geschaffen werden, in der Metadaten wachsen können.

Danksagungen

Mit besonderem Dank für die Unterstützung durch Gerald Jagusch (ULB), Ina Heine (WZL), Wolfgang Stille (ULB) sowie durch das eScience Team der RWTH Aachen University vertreten durch Annett Schwarz und Florian Claus. Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Projekt-ID 432233186.

ORCID IDs

- Matthias Grönewald  <https://orcid.org/0000-0002-3480-9102>
- Patrick Mund  <https://orcid.org/0000-0003-0890-5472>
- Matthias Bodenbenner  <https://orcid.org/0000-0002-9413-1874>
- Marc Fuhrmans  <https://orcid.org/0000-0002-9826-018X>
- Benedikt Heinrichs  <https://orcid.org/0000-0003-3309-5985>
- Matthias S. Müller  <https://orcid.org/0000-0003-2545-5258>
- Peter F. Pelz  <https://orcid.org/0000-0002-0195-627X>
- Marius Politze  <https://orcid.org/0000-0003-3175-0659>
- Nils Preuß  <https://orcid.org/0000-0002-6793-8533>
- Robert H. Schmitt  <https://orcid.org/0000-0002-0011-5962>
- Thomas Stäcker  <https://orcid.org/0000-0002-1509-6960>

Literaturverzeichnis

- [1] J. Ludwig. "Leitfaden zum Forschungsdaten-Management.", 2013.
- [2] Deutsche Forschungsgemeinschaft e. v, "Leitlinien zum Umgang mit Forschungsdaten.", 2015.
- [3] Laboratory Informatics Institute Inc. "The Complete Guide to LIMS & Laboratory Informatics.", ed. J. Johnes, Atlanta, 2017.
- [4] Z. Chen, D. Wu, J. Lu, and Y. Chen. "Metadata-based Information Resource Integration for Research Management." *Procedia Computer Science*, vol. 17, pp. 54-61, January 01, 2013.
- [5] M. Franklin, A. Halevy, and D. Maier. "From databases to dataspace: A new abstraction for information management.", *Sigmod Record*, vol. 34, no. 4, pp. 27-33, December, 2005.
- [6] M. D. Wilkinson, et al., "The FAIR Guiding Principles for scientific data management and stewardship." *Sci. Data* 3. 2016.
- [7] R. Van Noorden. "Data-sharing: everything on display." *Nature*, vol. 500, 2013.
- [8] J. Greenberg, S. Swauger, and E. Feinstein. "Metadata Capital in a Data Repository." *International Conference on Dublin Core and Metadata Applications; DC-2013—The Lisbon Proceedings*, September 2013.
- [9] M. Greenwald, T. Fredian, D. Schissel, and J. Stillerman. "A metadata catalog for organization and systemization of fusion simulation data." *in English, Fusion Engineering and Design*, vol. 87, no. 12, pp. 2205-2208, December 2012.
- [10] D. P. Schissel et al. "Automated metadata, provenance cataloging and navigable interfaces: Ensuring the usefulness of extreme-scale data." *(in English), Fusion Engineering and Design*, vol. 89, no. 5, pp. 745-749, May, 2014.
- [11] M. Politze, et al. "How to Manage IT Resources in Research Projects? Towards a Collaborative Scientific Integration Environment." *European journal of higher education IT*, 1(2020/1), 5. 2020.
- [12] R. H. Schmitt, V. Anthofer, et. al., "NFDI4Ing - the National Research Data Infrastructure for Engineering Sciences", 2020, doi:<https://doi.org/10.5281/zenodo.4015201>.
- [13] K. Coyle, T. Baker. "Guidelines for Dublin Core™ Application Profiles" <http://dublincore.org/documents/profile-guidelines/>, 18.05.2009. [Stand 14.04.2021]
- [14] O. Lassila. "Resource Description Framework (RDF) Model and Syntax Specification." <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>, 22.03.1999. [Stand 14.04.2021].

- [15] H. Knublauch, Dimitris Kontokostas. “Shapes Constraint Language (SHACL).”<https://www.w3.org/TR/shacl/> (20.07.2017) [Stand 14.04.2021].
- [16] D. Brickley, R. Guha. “RDF Schema 1.1.” <https://www.w3.org/TR/rdf-schema/>. (25.02.2014) [Stand 14.04.2021].
- [17] T. Carpenter.”9 - Electronic publishing standards.”, Academic and Professional Publishing, pp. 215-241, Chandos Publishing, 2012. <https://doi.org/10.1016/B978-1-84334-669-2.50009-3>.