

# Extended Abstract

## Coscine.nrw Landesweite Basisversorgung zur Verwaltung von Forschungsdaten im Open Source Modell

Marius Politze<sup>1</sup>[\[https://orcid.org/0000-0003-3175-0659\]](https://orcid.org/0000-0003-3175-0659), Ilona Lang<sup>1</sup>[\[https://orcid.org/0000-0002-7202-5982\]](https://orcid.org/0000-0002-7202-5982), and  
Katja Jansen<sup>1</sup>[\[https://orcid.org/0009-0005-7076-9848\]](https://orcid.org/0009-0005-7076-9848)

<sup>1</sup>RWTH Aachen University, Germany

**Abstract:** n.a.

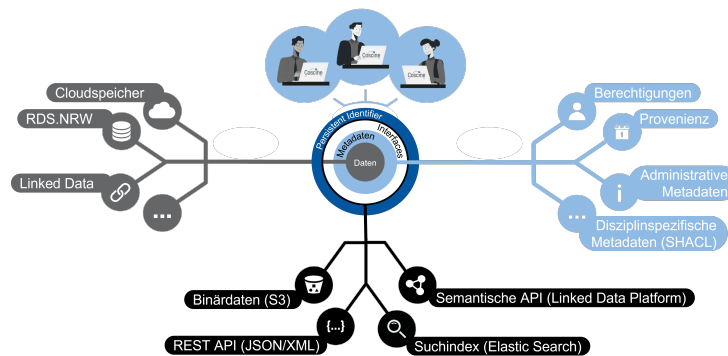
**Keywords:** FAIR, FAIR Digital Object, Data Storage, Metadaten, Scrum

### 1 Abstract

An der RWTH Aachen wird seit 2018 die Forschungsdatenplattform Coscine als Open-Source-Software entwickelt und für die Verwaltung von Forschungs(meta)daten, sowie zur Kontingentierung und Provisionierung von Speicherressourcen für Forschungsdaten eingesetzt. Coscine ist dazu gemäß der FAIR-Prinzipien entwickelt und implementiert Schnittstellen für sog. FAIR Digital Objects. Für Forschende bietet Coscine den Zugriff auf alle Forschungsdaten eines Forschungsprojekts, die Verknüpfung mit projekt- oder fachspezifischen Metadaten sowie die Verwaltung von Projektmitgliedern. Dank des niederschweligen Zugangsmanagements kann Coscine als Kollaborationsplattform über Hochschulgrenzen hinaus verwendet werden. In die Entwicklung von Coscine fließen zudem die Erkenntnisse und Anforderungen aus nationalen (NFDI, NHR) und internationalen Vorhaben (EOSC, gaia-x, RDA) ein und ermöglichen sowohl eine fachspezifische als auch eine fachübergreifende Verwendung der Plattform. Im Rahmen des Serviceangebots Coscine.nrw wird die Software Coscine für alle Hochschulen der DH.NRW zur Verfügung gestellt.

#### 1.1 Funktionsumfang der Datenmanagementplattform

Coscine bietet Forschenden eine Vielzahl von Funktionalitäten zur Verwaltung der Forschungsdaten unter Berücksichtigung der FAIR-Prinzipien. Coscine bietet dazu eine Abbildung einer Projektstruktur zur Verwaltung von Zugriffsberechtigungen und der Forschungs(meta)daten. Durch die Integration verschiedener Services können Forschende alle Projektdaten an einem Ort sehen und verwalten (vgl. Figure 1).



**Figure 1.** Auf Basis von definierten Schnittstellen kombiniert Coscine Speicherressourcen und Metadaten und stellt auf dieser Basis die Umsetzung der FAIR-Prinzipien nach dem Konzept des FAIR-Digital Object sicher.

### 1.1.1 Authentifizierung

Coscine nutzt die Authentifizierungs- und Autorisierungsinfrastruktur des DFN (DFN-AAI). Die DFN-AAI ist eine Kernkomponente der IAM4NFDI, sodass sich Coscine.nrw in die bestehende Landschaft integriert. Zur Authentifizierung von Nutzenden, die nicht direkt mit einer Hochschule oder Forschungseinrichtung in der DFN-AAI assoziiert sind, bietet Coscine die Möglichkeit der Authentifizierung über die Open Researcher and Contributor ID (ORCID).

### 1.1.2 Speicherplatzverwaltung

Der Zugang zu Speicherplatz auf dem Landesweiten Forschungsdatenspeicher RDS.nrw wird über ein wissenschaftsgeleitetes Antragsverfahren vergeben. Zur Sicherstellung der Einhaltung der guten wissenschaftlichen Praxis (GWP) und der Umsetzung der FAIR-Prinzipien wird das jeweilige FDM-Konzept im Rahmen eines Datenmanagementplans dargelegt.

Neben RDS.nrw können auch GitLab (für textbasierte Forschungsdaten bspw. Softwarequellcodes) oder Linked Data (für eine Metadatenverwaltung und Referenzierung externer Daten) Ressourcen erzeugt werden. Weitere (Community-) Cloud-Anwendungen, wie Sciebo und Nextcloud, sind geplant (vgl. Figure 1).

### 1.1.3 Metadatenverwaltung

Das Metadatenmanagement auf Projekt-, Ressourcen- und Dateiebene basiert auf den W3C Standards RDF und der Shapes Constraint Language (SHACL) und sichert so die Interpretierbarkeit und Interoperabilität der Metadaten. Alle von Coscine verwalteten Elemente erhalten einen Persistenten Identifier (PID). Nutzende können aus fach- und projektspezifischen Metadatenprofile ausgewählt. Die individuelle Erstellung von Metadatenprofilen ist dank der Einbindung des AIMS Projekts auch ohne fundierte Kenntnisse im Bereich RDF und SHACL möglich. Coscine bietet so die Möglichkeit, die in verschiedenen Fachbereichen und -Konsortien erstellten Ontologien wiederzuverwenden oder zu kombinieren.

Als Anknüpfungspunkt für internationale Infrastrukturen implementiert Coscine FAIR Data Point (FDP). Diese ist zu den FAIR Digital Objects kompatibel und basiert technisch auf dem Resource Description Framework (RDF), dem Data Catalog Vocabulary (DCAT) und der Linked Data Platform (LDP).

### 1.1.4 Automatisierung

Coscine bietet Schnittstellen zur Automatisierung von Arbeitsprozessen. Über die REST-API können Dateien beispielsweise automatisch mit Metadaten versehen und hochgeladen werden. Durch die Möglichkeit, über S3-Clients direkt mit verwalteten Objektdatenspeichern zu interagieren, wird eine weitere Automatisierungsmöglichkeit angeboten.

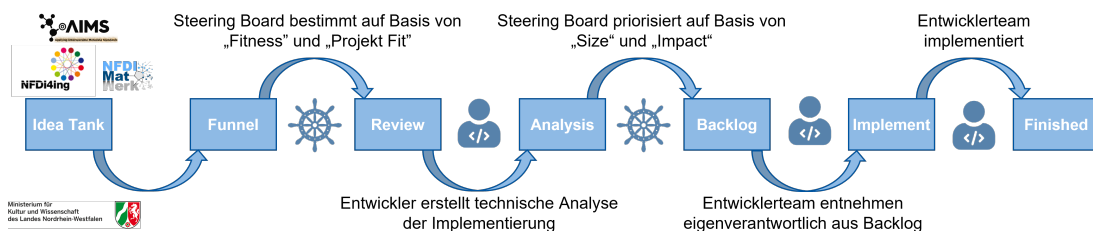
### 1.1.5 Archivierung

Neben der Speicherung von sogenannten aktiv genutzten heißen Forschungsdaten ermöglicht Coscine in Kombination mit RDS.nrw die Archivierung von kalten Forschungsdaten und Metadaten und wird für 10 Jahre nach Projektende gemäß der GWP gewährleistet. Danach sind der Übergang in ein geeignetes Langzeitarchivierungssystem und Maßnahmen für die Langzeitverfügbarkeit notwendig.

## 1.2 Communitygetriebene Weiterentwicklung im Open Source Modell

Die Entwicklung von Coscine unter einer freien Lizenz im Rahmen eines öffentlichen GitLab Projekts ermöglicht allen interessierten Nutzenden eine Teilhabe und Einbringung in die Coscine-Entwicklung.

Die Umsetzung neuer strategischer Anforderungen wird dabei über Prozesse aus dem Scaled Agile Framework (SAFe) und der agilen Scrum Methode an der RWTH Aachen gemanagt (vgl. Figure 2). Neue Anforderungen werden als sogenannte Epics formalisiert, in denen die benötigten Weiterentwicklungen im Rahmen einer vorgefertigten Vorlage ausgearbeitet werden. Die Epics sind öffentlich einsehbar und werden in enger Kommunikation mit allen beteiligten Stakeholdern auf Richtigkeit, Vollständigkeit, Mehrwert für Nutzenden sowie zeitlicher Kritikalität geprüft. Eine interne Steuerungsgruppe prüft in einem zweiten Schritt die Umsetzbarkeit und strategische Einordnung der Epics. Die nachfolgende technische Analyse des Epics erfolgt durch das Scrum-Team. Anschließend werden alle neuen Epics basierend auf der WSJF-Formel priorisiert, wobei ein Strategiefaktor, der Profit für alle Nutzenden, das Risikomanagement, die zeitliche Kritikalität und der Arbeitsaufwand berücksichtigt wird. Die drei Epics mit der höchsten Priorisierung werden an den/das Scrum-Team zur Entwicklung übergeben. Nach erfolgter Entwicklung wird die Umsetzung durch das Scrum-Team den Stakeholdern präsentiert und an alle Nutzenden kommuniziert.



**Figure 2.** Input von den verschiedenen Stakeholdern wird im SAFe Prozess überprüft, priorisiert und umgesetzt.

Ein zusätzlich eingerichtetes Contributions-Repository bietet der Community eine Plattform für das Teilen eigener Entwicklungen und Anwendungen, die beispielsweise im Rahmen der Nutzung der offenen Schnittstellen von Data Stewards auf Projektbene beigetragen werden. Neben dem Teilen von Code ermöglicht die Plattform auch

einen Austausch für Forschende und Data Stewards an den verschiedenen Standorten zur gegenseitigen Unterstützung für die Automatisierung von individuellen Prozessen.

### **1.3 Zusammenfassung und Ausblick**

Dank offener Schnittstellen (REST-APIs), der fortwährenden Anbindung an anerkannte Konzepte wie FAIR Digital Objects ist Coscine ein wichtiger Baustein, um verteilte Speicherinfrastrukturen entsprechend der FAIR Prinzipien weiterzuentwickeln. Durch die Metadaten-Verwaltung von Coscine können Speichersysteme, die Industriestandard entsprechen, gemäß der FAIR-Prinzipien verwendet werden. Dies verbessert insbesondere die Partizipationsmöglichkeiten kleinerer Hochschulen an diesen wissenschaftlichen Infrastrukturen und erhöht so langfristig die Wirtschaftlichkeit der investierten Ressourcen. Die communitygetriebene Weiterentwicklung ermöglicht die Beteiligung einer Vielzahl von Stakeholdern an der Entwicklungsroadmap.

Die Software Coscine wird in NRW als Serviceangebot Coscine.nrw etabliert und steht mit dem beschriebenen Funktionsumfang den Hochschulen in NRW zur Verfügung. Darüber hinaus können auch Use Cases der NFDI-Konsortien außerhalb von NRW Coscine als Basisinfrastruktur nutzen, um Forschungsdaten zu verwalten oder Forschungsdatenspeicher zu kontingentieren und provisionieren.

### **Beiträge der Autoren**

Marius Politze — Project administration, Supervision, Writing – original draft  
Ilona Lang — Conceptualization, Investigation, Writing – original draft  
Katja Jansen — Writing – review & editing

### **Förderung**

Die Arbeiten wurden mit Ressourcen von NFDI4Ing, gefördert von der Deutschen Forschungsgemeinschaft (DFG) unter Projektnummer 442146713, NFDI-MatWerk, gefördert von der Deutschen Forschungsgemeinschaft (DFG) unter Projektnummer 460247524 unterstützt. Coscine.nrw ist ein Service des Ministeriums für Kultur und Wissenschaft des Landes Nordrhein-Westfalen.