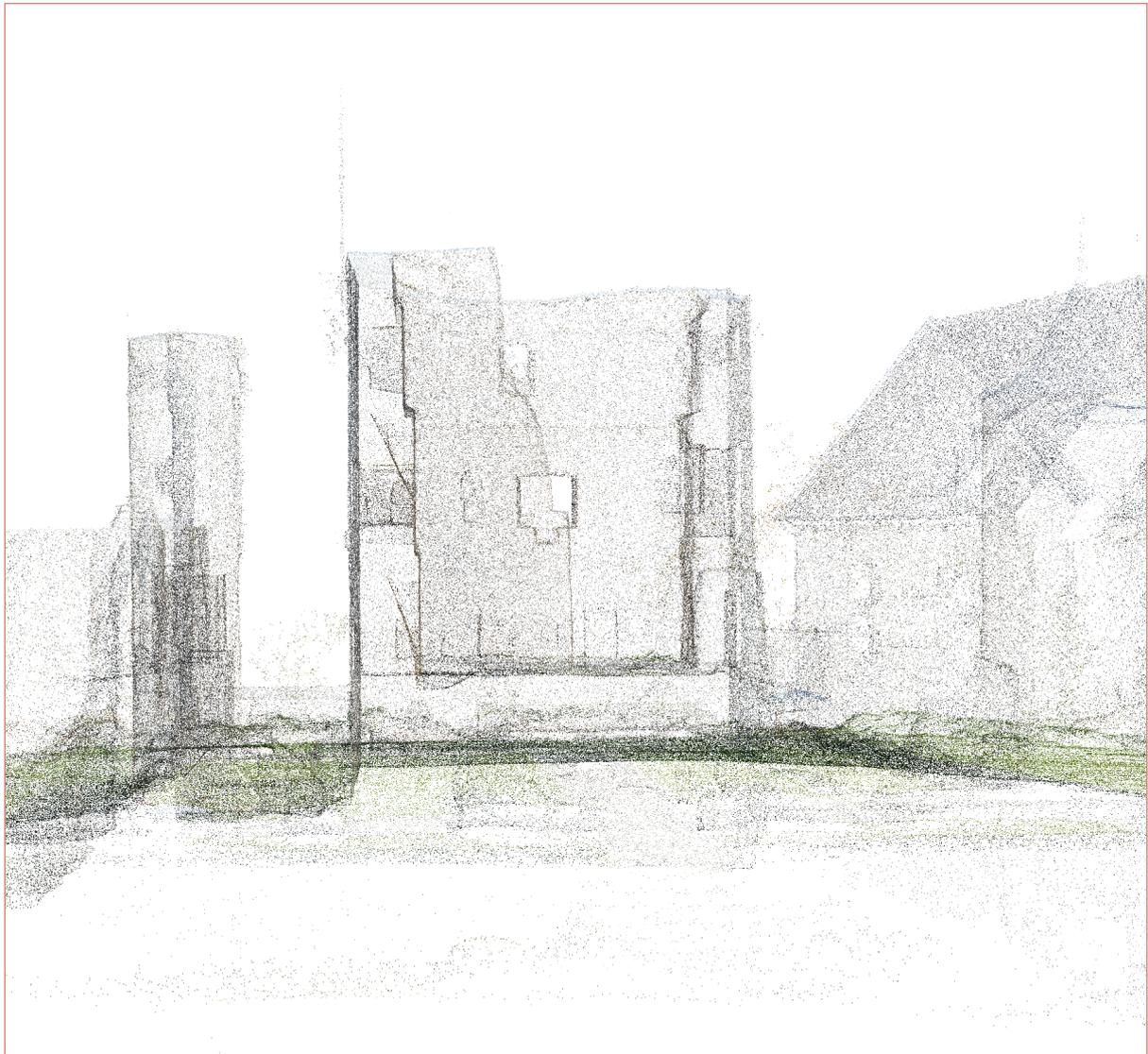




BUILD ON DATA

AUF DATEN BAUEN

Forschungsdaten in der Historischen Bauforschung und Denkmalpflege



Tagungsband zur Community-Tagung des DFG-Projekts baureka.online
am 4. und 5. Mai 2023 an der Technischen Universität Berlin

Abbildung Titelblatt

Sparse Point Cloud des Haus Heyden in Aachen-Richterich (Christian Klosterkötter / Felix Martin)

Impressum

Build on Data / Auf Daten bauen. Forschungsdaten in der Historischen Bauforschung und Denkmalpflege.

Eine Community-Tagung des DFG-Projekts baureka.online

© 2024 by Anke Naujokat and Sophie Helas

Herausgegeben von

Anke Naujokat, Sophie Helas

RWTH Aachen University

Lehrstuhl für Architekturgeschichte | Chair of Architectural History

Schinkelstraße 1

D-52062 Aachen

Germany

Layout und Grafik

Lara Draschoff, Felix Martin

Redaktion und Satz

Annika Dues, Sophie Helas, Svenja Kruijer

Founding

Diese Publikation wird von der Deutschen Forschungsgemeinschaft, Projektnummer 454194613, finanziert.



This work is licensed under Creative Commons Attribution 4.0 International.

Ausgenommen von der Lizenz sind Fotos, auf denen Personen erkennbar abgebildet sind, sowie alle Logos.

Verfügbar über das institutionelle Repositorium der RWTH Aachen University. DOI: 10.18154/RWTH-2024-04215

BUILD ON DATA

AUF DATEN BAUEN

Forschungsdaten in der Historischen Bauforschung und Denkmalpflege

Tagungsband zur Community-Tagung des DFG-Projekts baureka.online
am 4. und 5. Mai 2023 an der Technischen Universität Berlin

Herausgegeben von Anke Naujokat und Sophie Helas

In Gedenken an

Hermann Schlimme (1969-2023)



Inhalt

1.	Auf Daten bauen – Build on Data	6
2.	Auf Daten bauen – nicht ohne Forschungsdatenmanagement! <i>Matthias Razum</i>	10
3.	Das Forschungsdatenportal baureka.online. Die Vision <i>Anke Naujokat</i>	12
4.	Das Forschungsdatenportal baureka.online. Ein Fachrepositorium für die Historische Bauforschung <i>Nadine Marcinczik</i>	16
5.	Das Forschungsdatenportal baureka.online. baureka.index als zentraler Nachweiskatalog für Bauforschungsdaten <i>Tobias Glitsch</i>	22
6.	Madīnat al-Zahrā' <i>Heike Lehmann / Simon Trischberger</i>	28
7.	Forschungsdaten im Lehrkontext <i>Lukas Stampfer / Eva Kodžoman / Marina Döring-Williams</i>	38
8.	Die Datenbank Bauforschung/Restaurierung Baden-Württemberg <i>Claudia Mohn</i>	46
9.	Austausch und Standards für Daten und Metadaten photogrammetrischer Baufnahmen <i>Claudia Mächler / Andreas Noback</i>	56
10.	Forschungsdaten im Projekt „Synagogen-Gedenkbuch Hessen“ <i>Fani Gargova / Tilmann Gempp-Friedrich</i>	62
11.	Historischen Baubestand digital erfassen, modellieren, publizieren <i>Jörg Richter</i>	66

12.	Fehlende Bausteine in der GND: Normdaten zu Bauwerken	70
	<i>Julia Rössel / Hanna-Lena Meiners</i>	
13.	Das DFG-Projekt IDOVIR	78
	<i>Markus Wacker</i>	
14.	Coscine – Make Your Research Data FAIR	82
	<i>Lukas C. Bossert</i>	
15.	Bauforschung in der Lehre	88
	<i>Melanie Kim-Lan Nguyen / Liang Song</i>	
16.	3D Infrastructure for Digital Reconstructions	94
	<i>Igor Bajena / Clemens Beck</i>	
17.	Rechtsfragen in der Historischen Bauforschung	100
	<i>Grishka Petri / Oliver Vettermann</i>	
18.	Tagungsprogramm	104

Lukas C. Bossert

14. Coscine – Make Your Research Data FAIR

DOI: 10.18154/RWTH-2024-04226

Daten im Forschungsalltag

Forschungsdaten sind zentraler Bestandteil eines Erkenntnisgewinns. Unzertrennlich von diesen Daten sind Metadaten, also Informationen zu Forschungsdaten, die im Verlauf des Experiments, der Rekonstruktion oder Beobachtung gewonnen wurden. Erst durch die Kombination von Forschungsdaten mit Metadaten entsteht das Potenzial, dass Daten richtig eingeordnet, verstanden und schlussendlich auch nachgenutzt werden können. Metadaten werden in diesem Sinne nicht als Daten per se aufgefasst, sondern als Aussagen („Statements“), die etwas über Objekte mit potenziellem Informationsgehalt (gemeint sind Forschungsdaten) aussagen (Pomerantz, 2015).

Aus diesem Grund muss besonderes Augenmerk auf die Gewinnung, Strukturierung und Organisation von Metadaten und Daten gelegt werden, was insbesondere im laufenden Betrieb eines Forschungsalltags nicht immer leicht zu bewerkstelligen ist. Seit 2016 wird daher am IT Center der RWTH Aachen University an der Forschungsdatenplattform Coscine gearbeitet (<https://coscine.de>; Politze u.a., 2020). Coscine ist für Forschende die zentrale Dateninstanz für die Verwaltung und Aufbewahrung von Daten: während und nach einem Forschungsprojekt.

Während einer Projektlaufzeit fallen bereits wichtige Metadaten an, die jedoch nicht notwendigerweise mit der gleichen Sorgfalt wie die erhobenen Daten behandelt und dokumentiert werden. Eine nachträgliche Dokumentation und Aufbereitung der Daten am Ende eines Projekts ist kaum mehr zu leisten und wird daher oft vernachlässigt, sodass die Daten zwar archiviert werden, jedoch nicht den FAIR-Prinzipien* entsprechen und eine Nachnutzung nur schwer möglich ist.

Wenn jedoch mit Datenerhebung auch eine zeitgleiche Metadatendokumentation ohne erheblichen Mehraufwand stattfindet, wird nicht nur den FAIR-Prinzipien entsprochen, sondern den Forschenden ein signifikantes Arbeitspensum abgenommen.

Coscine und die FAIR-Prinzipien

Coscine gewährleistet eine FAIRe Datenorganisation, indem es auf technischer, administrativer und organisatorischer Ebene die Daten ganzheitlich betrachtet (Abb. 1).

Technische Komponenten

Die Daten in Coscine werden in einem georedundantem Speichersystem mit mehreren Standorten in NRW vorgehalten, um damit gängigen Anforderungen bei technischen oder

naturbedingten Beeinträchtigungen und Ausfällen zu genügen. Es handelt sich dabei um einen Objektspeicher, welcher Daten – im Gegensatz zu einem konventionellen Dateispeicher – nicht hierarchisch, sondern auf einer Ebene ablegt. Dabei wird für jede eingelieferte Datei ein zusätzlicher Metadateneintrag angelegt, sodass auf diesem Wege eine direkte Zuordnung und die Verknüpfung zwischen Datei und Metainformationen hergestellt werden kann. Speicherplatz wird Forschenden in einem wissenschaftsgeleiteten Bewirtschaftungsverfahren unentgeltlich zur Verfügung gestellt (Lang u.a., 2023), wofür verschiedene Ressourcentypen für die Organisation und Ablage der Daten bereitstehen.

Ressourcentyp Web

Alle Nutzenden bekommen für diesen Ressourcentyp 100 GB Speicherplatz zugewiesen. Dieser wird entweder über die grafische Oberfläche des Browsers oder über eine Programmierschnittstelle erreicht. Damit ist gewährleistet, dass mit den Daten ebenso die entsprechenden Metadaten abgelegt werden.

Ressourcentyp S3

Für diesen Ressourcentyp gibt es standardmäßig kein Speicherplatzkontingent, sondern es wird erst über ein Beantragungsverfahren zur Verfügung gestellt. Die Besonderheit dieses Ressourcentyps ist, dass neben den Zugriffsmöglichkeiten über die Weboberfläche oder die Programmierschnittstelle auch per externer Software (Client) Daten abgelegt werden können. Damit besteht die Möglichkeit besonders große oder sehr zahlreiche Datensätze abzulegen. Die Metadaten werden anschließend über die Weboberfläche nachgetragen oder in einem anderen Verfahren dokumentiert.

Ressourcentyp WORM

Für besonders schützenswerte Daten mit der Notwendigkeit zur garantierten Integrität steht der Ressourcentyp WORM bereit. Mit dem Konzept „Write Once, Read Many“ werden die einmal eingelieferten Daten für 10 Jahre unveränderbar vorgehalten, um beispielsweise bei einem Vorwurf von wissenschaftlichem Fehlverhalten die Datenintegrität belegen zu können. Speicherplatz für diesen Ressourcentyp wird erst nach einem Antrag und einem Beratungsgespräch freigeschaltet. Für die sichere Aufbewahrung der Daten im Sinne einer Archivierung sind die anderen beiden Ressourcentypen ausreichend.

Neben den erwähnten Ressourcentypen, bei denen Daten direkt in Coscine abgelegt werden können, wird mit dem Ressourcentyp Linked Data auch die Möglichkeit gegeben, Dateien, die an einem anderen Ort vorgehalten werden (zum Beispiel weil diese für einen Transfer zu groß sind oder weil rechtliche Anforderungen keine Migration in ein anderes System erlauben), ebenso mit Metadaten zu annotieren und somit strukturiert zu organisieren.

Administrative Komponenten

Coscine ist Projekt-orientiert aufgebaut und ermöglicht auf diese Weise den technischen Transfer einer Forschungsstruktur. Projekte werden neben einem Titel samt Abstract mit den geläufigen Angaben zu Principal Investigators, Laufzeit, Einordnung ins Wissenschaftsgebiet und Schlagworten versehen. Es können zu jedem Projekt weitere Unterprojekte angelegt werden. Die Dateien selbst werden in Ressourcen gespeichert, die sich nach den oben

genannten Typen (bspw. Web, S3, Worm) unterscheiden. Innerhalb eines Projekts können beliebig viele Ressourcen erstellt werden.

Die Nutzerverwaltung findet auf Ebene eines Projekts statt. Nach dem Login über einen DFN-AAI-Account oder über die ORCID-ID können Nutzende auf alle Ressourcen in ihrem Projekt zugreifen. Dabei kann einem Nutzenden eine der drei Rollen zugeteilt werden:

Gast

Kann auf Ressourcen lesend zugreifen, d.h. Daten und Metadaten ansehen und herunterladen.

Mitglied

Im Gegensatz zum Gast kann ein Mitglied auch Daten bearbeiten und neue Daten zu einer Ressource hinzufügen.

Besitzer

Zusätzlich zu den Berechtigungen eines Mitglieds kann ein Besitzer auch Projekteinstellungen ändern und neue Mitglieder hinzufügen oder entfernen oder deren Rolle ändern.

Zentraler Kern von Coscine ist die intuitive Datenablage und Annotation der Daten über sog. Applikationsprofile. In diesen Applikationsprofilen werden für den Datensatz auf Dateiebene alle relevanten Aussagen (Statements) notiert. Der entscheidende Unterschied zu dem mitunter geläufigeren Begriff Metadatenschema ist, dass Applikationsprofile von Forschenden maßgeschneidert erstellt werden können, um dezidiert die wichtigen Informationen zum Datensatz aufzunehmen. Während Metadatenschemata mitunter in längeren Prozessen definiert, diskutiert und schließlich publiziert wurden (Vergleiche beispielsweise das Metadatenschema Dublin-Core: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>), werden Applikationsprofile individuell und über einen speziellen Generator ohne größere Hürden oder Aushandlungsprozesse erstellt.**

Organisatorische Komponenten

Mit den organisatorischen Komponenten sind die Bereiche gemeint, mit denen man auf effiziente Art mit den Daten interagieren kann. Dafür wird nicht nur eine technische Validierung der Metadaten vorgenommen, die im Resource-Description-Framework-Graphen (RDF) abgelegt sind, sondern es wird über eine Programmierschnittstelle (REST-API) eine Möglichkeit gegeben, Daten und Metadaten hochzuladen und auch Auswertungen der Metadaten vorzunehmen. Schlussendlich sind damit alle Anforderungen erfüllt, sodass die Ressourcen anhand ihres persistenten Identifikators als ein FAIR Data Point (<https://www.fairdatapoint.org/>) gelten. Damit schließt sich der Kreis zu den FAIR-Prinzipien auch auf technisch-organisatorischer Ebene wieder (Politze u.a., 2023).

Fazit

Die an der RWTH Aachen entwickelte digitale Plattform, die darauf spezialisiert ist, nicht nur große Datenmengen, sondern auch die entsprechenden und gewinnbringenden Metadaten zu verwalten, ist mit dem Fokus auf eine intuitive Bedienung konzipiert. Darüber hinaus

wird mit umfangreichen Dokumentationsseiten (<https://docs.coscine.de/de/>) und einer stetig wachsenden FAQ-Sparte (<https://about.coscine.de/faq>) den Nutzenden ein stets aktueller Überblick über die Funktionen und Anwendungsmöglichkeiten angeboten. Für individuelle Fragen und Beratungswünsche kann ebenso über den ServiceDesk der RWTH per Email (servicedesk@itc.rwth-aachen.de) Kontakt aufgenommen werden.

Gleichwohl Coscine an der RWTH entwickelt wird, steht die Forschungsdatenplattform Forschenden in NRW zur Verfügung (Politze / Lang / Jansen, 2023).

*Das Akronym wird aufgelöst mit F: Findable, A: Accessible, I: Interoperable, R: Reusable. Der Kern dieser Prinzipien ist, dass Daten so abgelegt sind, dass sie für alle zugänglich, nachnutzbar und verständlich sind. Die einschlägige Publikation ist: Wilkinson u.a., 2016.

** Der Applikationsprofilgenerator wird in Zusammenarbeit mit dem Projekt AIMS (Applying Interoperable Metadata Standards; <https://gepris.dfg.de/gepris/projekt/432233186>) entwickelt und ist zu erreichen unter <https://coscine.rwth-aachen.de/coscine/apps/aimsfrontend/#/editor>.

Literatur

Ilona Lang / Marcel Nellesen / Lukas C. Bossert / Marius Politze: “Carrots and sticks: Motivating with storage for good rdm: Science led allocation of research data storage resources within an integrated rdm system.” In: *E-science-tage 2023: Empower your research – preserve your data*. heiBOOKS 2023.

Marius Politze / Florian Claus / Bela Darius Brenger / Mohammad Amin Yazdi / Benedikt Paul Anton Heinrichs / Annett Schwarz: “How to Manage IT Resources in Research Projects? Towards a Collaborative Scientific Integration Environment.” In: *European journal of higher education it 1* (2020/1) 5, <https://doi.org/10.18154/RWTH-2020-11948>

Marius Politze / Ilona Lang / Katja Jansen: “Coscine.nrw Landesweite Basisversorgung zur Verwaltung von Forschungsdaten im Open Source Modell.” In: *Proceedings of the conference on research data infrastructure*. Vol. 1. TIB Open Publishing (2023), <https://doi.org/10.52825/cordi.v1i.235>

Marius Politze / Yusra Shakeel / Siream Hunke / Philipp Ost / Rossella Aversa / Benedikt Heinrichs / Ilona Lang: “Long term interoperability of distributed research data infrastructures.” In: *Proceedings of the conference on research data infrastructure*. Vol. 1. TIB Open Publishing (2023), <https://doi.org/10.52825/cordi.v1i.348>

Jeffrey Pomerantz: *Metadata*. Cambridge – Massachusetts – London (2015).

Mark D. Wilkinson u.a.: “The FAIR guiding principles for scientific data management and stewardship.” In: *Scientific data 3*, 160018 (2016), <https://doi.org/10.1038/sdata.2016.18>

Abbildung

- Abb. 1 Funktionsumfang von Coscine hinsichtlich technischer, administrativer und organisatorischer Bereiche. In Kombination dieser Bereiche stellt Coscine die Daten als ein FAIR Digital Object dar.
(Ilona Lang, Marius Politze, Lukas C. Bossert)

Lukas C. Bossert
RWTH Aachen University
IT Center
- Research Process and Data Management -
Kackertstr. 10
52072 Aachen
bossert@itc.rwth-aachen.de
ORCID: 0000-0003-3076-3968

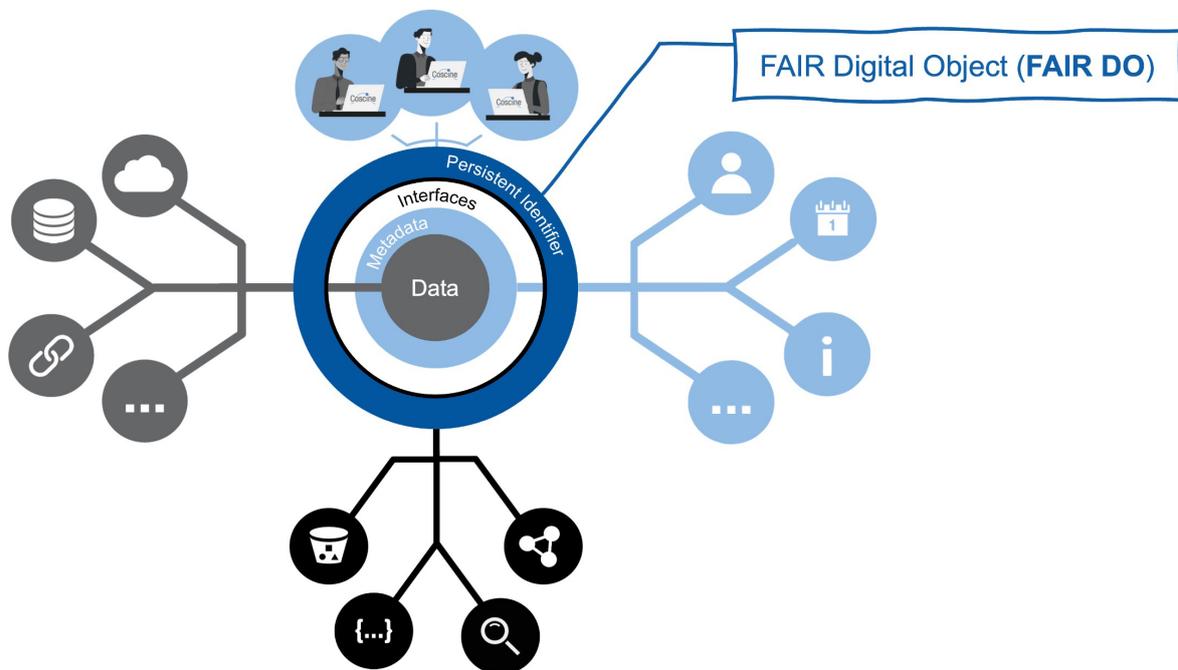


Abb. 1 Funktionsumfang von Coscine hinsichtlich technischer (grau), administrativer (hellblau) und organisatorischer (schwarz) Bereiche. In Kombination dieser Bereiche stellt Coscine die Daten als ein FAIR Digital Object dar.