

Year	Institution	Research area
2023	Institut für Computerbasiertes Entwerfen und Baufertigung (ICD), University of Stuttgart; Universitäts- und Landesbibliothek (ULB) Darmstadt, TU Darmstadt	45 Construction engineering and architecture
Project title	RDM-Workflows for construction engineering and architecture Development and integration of exemplary application profiles for timber construction data	
Project Outline		
<p>Contemporary research on the future of the built environment is becoming increasingly interdisciplinary. Ecology and sustainability are shaping and determining current debates. Against this background, emerging integrative methods such as "computational design" or "computational engineering" and their digitally entangled algorithmic design, planning and manufacturing processes promise innovative solutions. In these digital environments data is aggregated in a multitude of states. It is prepared, processed, and finally used to foster new approaches for contemporary planning and production.</p> <p>At the same time, the awareness in the community as well as the infrastructure for the publication of research data in architecture and civil engineering is still in its early stages. While certain amounts of research data have already been published, most of it remains inaccessible as they are stored on university servers, clouds and external harddrives. The available research data infrastructure is mainly organized in a decentralized way, i.e. there is no central access to a domain-specific infrastructure. In addition, datasets that are published can only be annotated as closed packages on a generic level and are therefore lacking domain-specific method- and entity-related metadata to describe them. In this context, it is currently not possible for researchers and machines to search, find, explore, process, and reuse domain-specific research data in respect to the FAIR principles¹.</p> <p>In order to create a domain-specific access point for the publication of and search for research data, a repository is created in the context of the Fachinformationsdienst BAUdigital, funded by the DFG since 2020. While the technological basis for the domain-specific research data infrastructure is under development, a solution to the development of domain-specific metadata is still needed. A monolithic approach with which all research data can be described universally is not suitable, since methods and data are particularly heterogeneous. Application profiles provide an alternative, modular approach to assemble flexible application- and domain-specific metadata based on existing terminologies. For such application profiles to meet the needs of researchers and be used successfully, they must be developed and discussed collaboratively by infrastructure providers and researchers. In combination with tools and</p>		

¹ <https://www.go-fair.org/fair-principles/>, accessed on 30.09.2022

services developed by NFDI4Ing, these can facilitate novel paths in dealing with domain-specific research data.

In order to enable the research community to explore these paths, the ICD Stuttgart and the ULB Darmstadt team up to develop exemplary application profiles based on representative research data sets in the NFDI4Ing Seed Funds. Therefore, it is planned to collect and evaluate required terminologies and to integrate them into the NFDI4Ing terminology service (S-3-2) if needed (WP1). Based on this, application profiles for the annotation of domain-specific datasets and data will be developed (WP2) and implemented into the infrastructure of the NFDI4Ing Metadata Profile Service (S-3-1) Infrastructure for reuse in the Fachinformationsdienst (FID) BAUdigital “Workbench” (WP3). The application profiles will then be presented as a showcase to selected partners from the research community and jointly discussed (WP4). In parallel, the described development process will be documented and analyzed in order to be published as a best practice example and to be reused for the development of further application profiles (WP5).

Brief description of the proposed project

The Institute of Computational Design and Construction (ICD)², led by Prof. Menges, has been researching integrative design processes, computer-aided design, and robotic fabrication, particularly in the area of timber construction and digital technologies, for over a decade. The ICD has recently significantly expanded its expertise and is researching multi-story wood construction as part of a larger research program of the Integrative Computational Design and Construction for Architecture (IntCDC) cluster of excellence³. In the course of this effort, a wide number of different types of data on multi-storey solid wood construction have been collected, generated, analyzed and processed over the past three years. This includes three datasets: (I) 350 recent buildings, one published article and one dataset^{4,5}, (II) one published dataset on stakeholders in wood construction⁶, and (III) another dataset to be published by the end of the year, with about 100 buildings and their associated drawings (plans and sections) available as CAD- and PDF-files in the same scale and drawing style.

Since there is no subject-specific infrastructure for publishing such research data to date, the datasets have been published in the local repository of the University of Stuttgart. In this context, two main problems can be highlighted:

The published datasets and data are not findable to others due to storage in a non-specific local repository, lack of domain-specific metadata as well as an unstandardized way of data collection.

The actual data of interest to researchers, providing valuable quantitative and qualitative information on a large number of entries, is not searchable, discoverable, interoperable, or reusable in the form of a closed research dataset with respect to FAIR principles.

To make the published datasets searchable and findable by third parties, they must be published in a central subject-specific repository and annotated with subject-specific metadata. To make the contained data searchable and findable, it must be structured in a

² [The Institute of Computational Design and Construction, ICD](#), accessed on 30.09.2022

³ [the Integrative Computational Design and Construction for Architecture \(IntCDC\) cluster of excellence](#), accessed on 30.09.2022

⁴ <https://doi.org/10.18419/darus-2733>, accessed on 30.09.2022

⁵ <https://www.mdpi.com/2075-5309/12/4/404>, accessed on 30.09.2022

⁶ <https://doi.org/10.18419/darus-2740>, accessed on 30.09.2022

standardized way based on subject-specific vocabularies and converted from tabular data into separate entity-based repository entries (data-to-metadata conversion). Only then it will be possible to semantically link the currently separated data about individual buildings, the actors involved in the construction process as well as with the associated planning material such as building drawings.

At the same time, the ULB Darmstadt has teamed up with TIB Hannover, UB Braunschweig and Fraunhofer IRB to establish the DFG-funded Fachinformationsdienst BAUdigital⁷ for the disciplines of civil engineering, architecture and urban studies. A core task of the FID BAUdigital is to develop the needed infrastructures and processes for annotation of heterogeneous research data and thus to offer a domain-specific access point for searching and publishing corresponding research data. For this purpose, the so-called "Workbench"⁸ – a modular collection of tools in a Kubernetes cluster, that can be accessed beyond the FID BAUdigital – is being developed at the ULB Darmstadt and offers researchers specific tools for processing research data during upload as well as annotation with domain-specific metadata. Functionalities for processing and preparing research data currently include the automatic generation of web-enabled preview objects of 2D and 3D data (.svg, .gltf) as well as the automatic extracting of floor plans from digital building models such as .ifc files. In the area of metadata, the Workbench currently enables automated classification of plan data using machine learning (e.g., floor plan, elevation, section), access to vocabularies (FINDEX Bau⁹) from the Terminology Service to describe methods and materials, localization of a dataset on an interactive map, and assignment of permanent identifiers (DOI) and licenses. In addition, it currently provides an initial prototype interface for searching and previewing research datasets with corresponding metadata, which will be continuously developed. While the development of the core infrastructure as well as the preparation of subject-specific data has been a core task in the development of the Workbench so far, the integration of the NFDI4Ing Metadata Profile Service (S-3-1)¹⁰ shall enable the development of subject-specific metadata schema (application profiles) in close cooperation between infrastructure and research community. To date, the development of the Metadata Profile Service is focused on the Community Cluster 41, "Mechanical and industrial engineering". An extension and evaluation of its generalizability is promising and of great interest to both the researching community of the NFDI4Ing Community Cluster 45¹¹ and the FID BAUdigital as well as the National Research Data Infrastructure. This connection promises to be the basis for both researchers and machines to search, find, explore, process, and thus reuse different types of domain-specific research data as defined by FAIR principles.

While the development of an initial representative application profile will provide a common ground for this connection as a solid infrastructure foundation, the goal of this work is intended to be an accelerator for a research proposal that promises to enable far-reaching innovations in the architecture and engineering research landscape: Collecting, curating, analyzing, and comparing data from discipline-specific entities such as buildings, structures, sites, stakeholders, plans, and models has long been a typical task for researching the built

⁷ <https://doi.org/10.3897/rio.8.e82563>, accessed on 30.09.2022

⁸ <https://www.fid-bau.de/forschungsdaten/>, accessed on 30.09.2022

⁹ <https://service.tib.eu/ts4tib/ontologies/bau>, accessed on 30.09.2022

¹⁰ <https://nfdi4ing.de/base-services/s-3/>, accessed on 30.09.2022

¹¹ <https://nfdi4ing.de/about-us/community-clusters/>, accessed on 30.09.2022

environment. Datasets often have overlaps and similarities in description, and are laboriously re-aggregated to feed specialized digital process chains and generate new knowledge. This project is intended to be the starting point for developing scientific and open digital editions of collections of buildings, bridges, floor plans, 3D models and numerous other entities, which can be expanded flexibly and easily and maintained over the long term. In this context, tools for user-friendly and simple data visualization, access options, and cross-dataset comparison options can add essential value to contemporary research and sustainably increase the attractiveness for reuse and publication of research data in the community. The necessary tools and infrastructures as well as an active and networked research community to fill them are available.

Relevance to the community: Researchers are currently faced with the problem that their produced datasets and data that are of interest to others are not accessible with the currently available infrastructures, because the data itself is not published in a reusable way as datasets, is not provided with subject-specific metadata, and the data collection could therefore not be carried out in a standardized and interoperable way. The envisioned solution for this lies in intertwining the developments of the FID BAUdigital (subject-specific research data infrastructure) and NFDI4Ing through integration with the Metadata Profile Service (S-3-1) and Terminology Service (S-3-2). This approach will be solidified and published through documentation as best practices for research data management workflows for "Defining and Delivering RDM Training Content for Engineers" (S-6-1). Beyond the Base Services we will pursue an exchange primarily with the task area "Frank" due to the heterogeneous nature of sources in architecture and civil engineering (text, numerical data, images) as well as "Ellen", since many aspects are based on analyses of complex systems with interdependencies. In summary, this project therefore aims to bring together different actors from the basic services, task areas and the community cluster 45 "Civil Engineering and Architecture" of the NFDI4Ing, the "Fachinformationsdienst BAUdigital" for civil engineering, architecture and urban planning as well as the addressed research community, represented by the ICD Stuttgart, to promote an exchange for the exemplary development of application profiles. This is particularly important because, although it is possible to develop individual solutions on a small scale, it is only through exchange that transfer on a broad scale can be made possible.

Networking with the community: The disciplines involved in the built environment are almost impossible to enumerate. For this very reason, it is particularly important for us to seek in this project a basis for exchange. The researchers community to be served through the FID and the NFDI4Ing first within the community of CC-45, i.e. starting with architecture and civil engineering within our networks, but also going beyond to intersections into historical building research, geodesy or urban planning, in order to develop a sustainable network between the NFDI4Ing, the FID BAUdigital and the researching community. In addition to the development of standards for documenting and describing entities such as buildings, this also involves the development of discipline-specific data management work processes and concepts for conducting complex research efforts. Both the processes and the tools developed in the Workbench of the FID BAUdigital will be beneficial for the NFDI4Ing infrastructure and strengthen the cooperation and network.

Networking with other consortia: The problems to be addressed in this project, as described above, cannot be solved in the long term without interdisciplinary exchange and coordination. Entities such as buildings or floor plans are objects that overlap in the context of the NFDI, especially with the NFDI4Culture¹² community "Architecture". The work in the task areas "Data capture and enrichment of digital cultural assets" and "Standards, data quality and curation" deals with similar problems in order to capture and describe buildings as cultural artifacts, especially in the historical context. The ULB Darmstadt is in close contact to these developments, which aim to develop similar infrastructures to those in the FID BAUDigital, f.e. with a triple store based on Wikibase¹³. Based on this, the application profiles to be developed promise to provide a long-term solution for the exchange of data between the infrastructures. We see further potential in the exchange with NFDI Text+¹⁴, which deals with language- and text-based research data infrastructures, since buildings have always been also described by texts with descriptions can be very heterogeneous in style, length, format and content.

Degree of innovation: Existing solutions in the area of research data management can be divided into three categories. University repositories (e.g. DaRUS¹⁵) offer local services, but these do not allow domain-specific search scenarios, especially not at the level of data; the same applies to generic repositories (e.g. Zenodo). Data-based repositories (e.g. DFG-3D Viewer¹⁶, Baureka¹⁷) offer very specific connection points, but are not relevant for the addressed community. Beyond these commercial solutions can be named such as conventional architecture and engineering magazines, that offer access to building databases.

The main therefore innovation thus lies in the modularity and flexibility that comes to the fore through the use of application profiles in the context of a domain-specific repository, and the scalability that connection to the NFDI4Ing infrastructure makes possible in the first place.

The proposed project goes beyond the current developments in the field as it is currently difficult to locate data, and even more difficult to parse through, filter, and visualize data in architecture and construction. An outlook could hint towards the development in this project would be to combine several different datasets, link the individual data and filter the data, as well as to optimize the processes for researchers publishing data or datasets through automatic metadata generation.

We see big potential in extending the proposed project into a larger follow-up research project to establish a novel (user-oriented) interactive data repository in architecture, engineering and construction (AEC) capable of interconnecting multiple repositories through the tools provided by the infrastructure. The developed methods, in particular establishing information on individual data points, would serve as groundwork for different applications such as (a) automatic data visualization, f.e. generating maps for the buildings or stakeholders, as well as generating individual graphs for interactive data comparison, (b) an AI-based metadata-extraction and automatic classification, f.e. for drawings such as floorplans (prototype

¹² <https://nfdi4culture.de>, accessed on 30.09.2022

¹³ <https://doi.org/10.5281/zenodo.6359284>, accessed on 30.09.2022

¹⁴ <https://www.text-plus.org>, accessed on 30.09.2022

¹⁵ <https://darus.uni-stuttgart.de>, accessed on 30.09.2022

¹⁶ <https://dfg-viewer.de/dfg-3d-viewer>, accessed on 30.09.2022

¹⁷ <https://baureka.online/de>, accessed on 30.09.2022

developed at FID BAUdigital¹⁸), as well as (c) a prototypical ingest of building surveys into the repository and the ability to match tabular data with metadata fields for item based building objects. At the same time the developments in the projects can serve as groundwork for exchange with other researchers that work on similar topics in other disciplines, f.e. DFG-3D Viewer, or Baureka, or to exchange to other repositories via a connection to the NFDI4ing Metadata Hub (S-3-3).

Timeline, milestones, metrics, KPI

WP1: Indexing of existing terminologies to be added to the NFDI4ing Terminology Service

Application profiles must use the terminology of a discipline - ideally in a consistent way to permanently reference research data. For this purpose, existing terminologies and ontologies are collected and evaluated for their applicability. If they do not already exist in the NFDI4ing Terminology Service (S-3-2), they should be integrated to be used in the development of the application profiles in the Metadata Profile Service (Milestone 1, terminology integration).

WP2: Develop application profiles as extension of the NFDI4ing Metadata Profile Service

The development of application profiles is a complex process that the proposing partners from research and infrastructure are approaching together in this project. Therefore, the processes and data to be captured are mapped in detail as well as content dependencies in the metadata are identified. Existing metadata schemas - especially from research (e.g. DFG 3D-Viewer, Baureka) and existing databases (DETAIL Inspiration, Building Types Online, Wikidata) - will be used for comparison, discussion and decision making. On this basis, concepts for the application profiles will be developed, coordinated and successively expanded.

On the level of research datasets three application profiles serve to describe the datasets on „Multi-storey Timber Buildings Data“, on „Stakeholders in Multi-storey Timber Data“ and on the technical drawings (to be published in November 2022) including information on descriptive, technical, and administrative metadata as well as on metadata to discrete the domain-specific research methods (f.e. quantitative building analysis, morphology, etc.). On the level of data it is planned to develop application profiles to describe three main entities. The building data application profiles will describe the buildings in general (project title, location, construction, etc.), complemented with application profiles for the domain-specific data (program, materials, structural system, etc.). A similar approach will be pursued for the stakeholder and plan data.

The application profiles are finally described in an RDF schema as SHACL shapes and are available in this form for further use from a server run by the FID BAUdigital (Milestone 2, application profiles on FID BAUdigital).

WP3: Integrate application profiles infrastructure into the workbench for data ingest

¹⁸ <https://api.fid-bau.de/?urls.primaryName=Floorplan%20classifier>, accessed on 30.09.2022

As a first exemplary application, it is planned to make the application profiles in the ingest process of research data available via the Workbench, the central research data infrastructure of the FID BAUdigital. For this purpose, the Workbench accesses the application profiles of the server and provides them to the user in the interface for entering metadata. After selecting one or more application profiles from a drop-down menu, the individual form fields for entering metadata are automatically generated, validated after the metadata is entered, and provide visual feedback according to the constraints specified in the application profiles. During the development phase, the developed application profiles will initially be retrieved from an internal server; the long-term goal is to integrate them into the NFDI4Ing Metadata Profile Service infrastructure and access them from the Workbench (Milestone 3, Workbench Access).

WP4: Discuss developed application profiles with community

The key aspect for the development of standardized application profiles is a broad reuse by the research community. For this reason, it is essential that both the development of the application profiles themselves and the (interim) results are discussed and further developed together with the research community (Milestone 4, community meeting). This is the only guarantee that the application profiles and the research data annotated with them will be usable in the long term.

WP5: Document the workflow as best practice example for the community

In order to encourage the use of standardized application profiles in the community, not only the application profiles themselves must be discussed and shared with the community, but also the processes that have been evolved for their development. Therefore, the entire process described here will be documented, optimized, and made available to researchers as a best-practice example (Milestone 5, documentation). This will provide a flexible basis both for new workflows for preparing to carry out complex and data-intensive research processes and for promoting research data publications among researchers.

Timeline

1	2	3	4	5	6	7	8	9	10	11	12
WP 1	WP 1	M1									
	WP 2	M2									
				WP 3	WP 3	WP 3	M3				
			WP 4	WP 4			WP 4	WP 4	M4		
									WP 5	WP 5	M5