Placeholder Name of the conference

Placeholder Session title

https://doi.org/10.52825/xxxx........ DOI placeholder (WILL BE FILLED IN BY TIB Open Publishing)

Published: (WILL BE FILLED IN BY TIB Open Publishing)

Conceptualising an Integrated Data Model Using RO-Crates for Sports Facility Building Information

Linking Distributed Data, Persistent Identifiers, and Building Information for Sports Facilities

Dirk Betz[1,2,*] https://orcid.org/0000-0002-6411-4758, Claudia Biniossek[1,2] https://orcid.org/0000-0002-2202-7875, Sören Wallrodt[3] https://orcid.org/0000-0003-1105-1156, Eli Chadwick[4] https://orcid.org/0000-0002-0035-6475, Stian Soiland-Reyes[4] https://orcid.org/0000-0001-9842-9718, and Carole Goble[4] https://orcid.org/0000-0003-1219-2137

[1]TIB – Leibniz Information Centre for Science and Technology, Germany

[2]Center for Empirical Research in Economics and Behavioral Sciences, University of Erfurt, Germany

[3]Koblenz University of Applied Sciences, Germany

[4]University of Manchester, GB

*Correspondence: Dirk Betz, dirk.betz@tib.eu

Abstract. The digital transformation of the sports facility domain has been hindered by fragmented, non-interoperable, and non-reusable data across scientific, industrial, and public data spaces. This paper explores an integrated data model based on RO-Crates, a lightweight implementation of the FAIR Digital Object principles, to enable the structured representation, interconnection, and reuse of data related to sports facilities. We present a dual approach in which persistent identifiers (PIDs) are assigned to sports facilities and RO-Crates are created to capture relevant building information. PIDs for Sports Facilities utilize existing sports atlas metadata e.g., whereas RO-Crates encapsulate specific Building Information Modelling (BIM), Global Unique Identifier (GUID), Global Location Number (GS-1) and Geographic Information System (GIS) metadata. The proposed concept aims to establish semantic and technical interoperability across industrial, public, and scientific data spaces. This will facilitate the efficient planning, construction, renovation, management and sustainability assessment of around 200,000 sports facilities in Germany alone.

Keywords: FAIR Digital Objects, RO-Crates, Sports Facilities, Building Information Modelling (BIM), Asset Administration Shell, Sports Science, Data Interoperability

1 Introduction

Sports facilities, ranging from municipal sports halls to large stadiums, are an integral component of urban and rural built environments. An estimated 200,000 sports facilities in Germany alone require coordinated management for planning, construction, renovation and operation throughout their lifecycle. However, data describing these facilities are often siloed across heterogeneous systems: municipal databases, architectural models, energy performance certificates, and academic data repositories. These data are rarely interoperable, often lack persistent identifiers, and are not reused efficiently across institutional boundaries.

The **FAIR Digital Object (FDO)** framework [1] provides a conceptual and technical foundation for data that are Findable, Accessible, Interoperable, and Reusable (FAIR principles) [2].

Within this framework, **Research Object Crates (RO-Crates)** have emerged as a practical and community-driven implementation [3].

This paper explores how the FAIR principles and the FAIR Digital Objects framework can be applied to the building information of sports facilities through an integrated data model based on RO-Crate.

2 Problem Statement

FAIR Digital Objects (FDOs) and their flavours (e.g. RO-Crates-based or handle-based solutions) have been proven in research practice. The question is whether FDOs can efficiently link existing datasets and metadata within and across scientific, industrial, and public data spaces in the field of sports facilities, successfully addressing the opportunities and challenges of digitalisation. This question arises because the efficient, needs-based construction, renovation and maintenance of around alone 200,000 core sports facilities in Germany has remained far below its potential due to a lack e.g., of consolidated information on sports facilities, regional supply levels, energy renovation requirements, and building and property data. Our approach consists of two parts. One necessary step is to create a PID for each sports facility. For instance, a record for each German swimming pool has already been created in the recently developed *Bäderleben* repository (https://baederleben.de/), within the framework of the digital sports facilities atlas [4], so the next step is to assign PID. The other part comprises creating an RO-Crate to integrate building information, including data on energy efficiency, persistent identifiers (e.g. GUID, GS1) and data models (BIM, GIS), as well as metadata and unstructured data. The RO-Crate is interlinked with the facility PID, administrative data and the scientific PID ecosystem.

While PIDs for facilities can be assigned in atlases or registries by public institutions in Germany on local, state, and federal levels, **Building Information Models (BIM)** evolve continuously throughout the lifecycle of a construction structure, involving multiple creators and changing metadata. Static PID assignment cannot capture this dynamic complexity. A complementary RO-Crate-based representation can address this by encapsulating evolving datasets within a structured, versioned, and semantically rich digital envelope.

This project originates from the sports facility databases developed by the partners (www.baederleben.de and www.sportstaettenatlas-rlp.de), which do not yet fully comply with the FAIR principles. Based on existing metadata schemas and infrastructures, we demonstrated how a static, relational database model can be linked to other information using PIDs and existing services, for example the TIB Open Research Knowledge Graph (ORKG). Pilot projects funded by the Federal Ministry of the Interior (BMI) have demonstrated the practical applicability of PID prototypes for sports facilities. Switching to PIDs and linking them to flexible RO-Crates for building information over the lifecycle represents a paradigm shift for the heterogenous stakeholder groups. To ensure that PIDs are assigned uniquely and consistently and that duplicates are avoided, appropriate governance and coordination with stakeholders are needed.

3 Proposed Integrated Data Model

3.1 Conceptual Architecture

The default for the persistent digital identification of physical objects is a PID due to its technology readiness level. However, PIDs are not suitable for building information which often fail to meet basic requirements such as ownership rights, data collection and entry regulation, data stability, and central control via a registration agency. Furthermore, building information follows the dynamic life cycle model and may change throughout the building's lifespan. Therefore, identifiers must be flexible enough to accommodate these changes. RO-Crates fulfil these requirements.

The proposed architecture links two levels of granularity:

1. **Facility Level (PID).** Each sports facility (e.g., a football field or a swimming pool) receives a persistent identifier (e.g., Handle, DOI [e.g., from DataCite]) in a sports facility registry. This ensures that the reference remains persistent across approximately 200,000 German sports facilities and more than 10 million international sports facilities, enabling PID-graph integration and connection with related data from the (sports-)scientific, (construction-)industrial, and public/ administrative domain including statistical data.
2. **Building and construction level (RO-Crate).** Each building hosting a facility is represented as an RO-Crate. The crate aggregates: (a) BIM models (IFC, gbXML, CityGML); (b) metadata on construction phases, renovations, and maintenance cycles; (c) links to material passports and energy certificates; (d) contextual datasets (geospatial data, soil data, population coverage).

The facility PID is referenced within the building’s RO-Crate metadata, establishing a bi-directional link between the two digital entities.

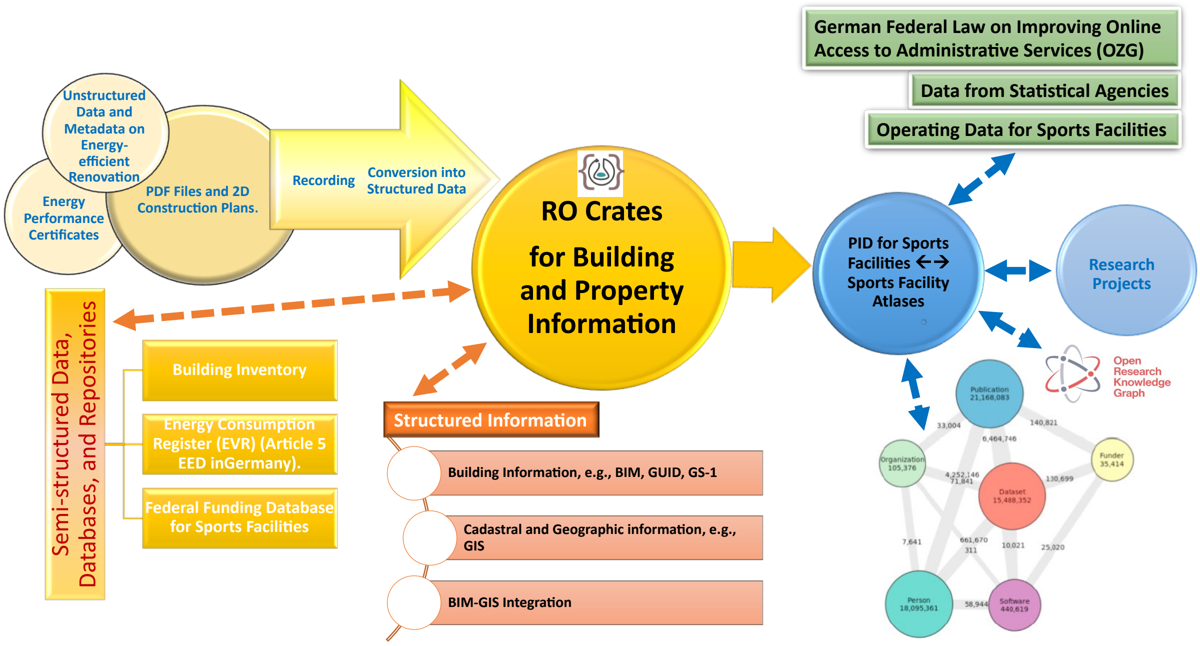

Figure 1. *Integrated data model for sports building information (PID-Graph see [5]).*

3.2 RO-Crate Metadata Schema

The RO-Crate metadata follows a JSON-LD structure extended with domain-specific vocabularies e.g., IFC [6] schema elements (building elements, spaces, systems); BOT (Building Topology Ontology) [7]; QUDT (Units and quantities for energy and material data) [8]; GeoSPARQL (Geospatial coordinates and bounding geometries representing and querying geospatial data on the Semantic Web) [9]; Generic descriptive metadata e.g., Dublin Core, Schema.org, ddi-cdi metadata standard.

3.3 Integration with Asset Administration Shells (AAS)

The Asset Administration Shell (AAS), a core concept from Industrie 4.0 [10], represents digital twins of assets with standardized submodels. Integrating RO-Crates with AAS allows the encapsulation of digital building components. Each submodel (e.g., energy data) can be represented as a nested crate.

3.4 Data Ingest

Data ingest follows a multi-step workflow: (1) extract metadata from BIM files using IFC parsers; (2) enrich metadata with contextual datasets (geospatial, demographic, administrative); (3) generate RO-Crates with appropriate ontology mappings (comparable: [11]); (4) register crates in a federated FDO registry; (5) enable discovery and querying via FAIR Data Points and Linked Data interfaces.

4 Conclusion and Outlook

The integration of RO-Crates into the sports facility domain represents a paradigm shift in managing and reusing building information. The dual approach - PIDs for facilities and RO-Crates for buildings - provides a scalable foundation for linking, discovering, and analyzing sports infrastructure data across institutional and sectoral boundaries. After the current pilot phase, a stakeholder engagement forum and a sustainable governance structure that relates to the challenges of community-driven physical object initiatives [12] are needed to implement this approach and prevent failed PID and failed Crates.

Data availability statement

Not applicable.

Underlying and related material

Not applicable.

Author contributions

Conceptualization (DB, CB, SW, EC, SS, CG); Data curation (N/A); Formal Analysis (N/A); Funding acquisition (N/A); Investigation (N/A); Methodology (N/A); Project administration (N/A); Resources (N/A); Software (N/A); Supervision (N/A); Validation (N/A); Visualization (N/A); Writing – original draft (DB, CB, SW, EC, SS, CG); Writing – review & editing (DB, CB, SW, EC, SS, CG).

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

References

[1] https://fairdo.org/

[2] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1), 1-9.

[3] Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L. J., Coppens, F., Fernández, J. M., ... & Goble, C. (2022). Packaging research artefacts with RO-Crate. *Data Science*, *5*(2), 97-138.

[4] Wallrodt, S., & Thieme, L. (2021). Grundlagen für einen digitalen Sportstättenatlas. *Entwicklung einer Systematik anhand von Parametern zur digitalen bundesweiten Erfassung von Sportstätten. Bonn: BISp. Zugriff unter: https://www. dosb. de/sonderseiten/news/news-detail/news/grundlagen-fuer-digitalen-sportstaettenatlas-veroeffentlicht*.

[5] Suominen, T., Himanen, L., Hellsten, L., Bennett, M., & Buys, M. (2024). Piloting the use of PIDGRAPH data in research. fi. *Procedia Computer Science*, *249*, 232-242.

[6] https://technical.buildingsmart.org/standards/ifc/ifc-schema-specifications/

[7] https://w3c-lbd-cg.github.io/bot/

[8] https://www.qudt.org/pages/HomePage.html

[9] https://www.ogc.org/standards/geosparql/

[10] Rohrmus, D., Simon, A., Biniossek, C., Betz, D., & Wittenburg, P. (2024). Manufacturing X–Solutions for the Manufacturing Industry: Industrial Data Standards to Support Industrial Data Ecosystems. In *Open Conference Proceedings* (Vol. 5).

[11] Grieb, J., Weiland, C., Wolodkin, A., Gordillo, J., & Islam, S. (2025). Implementing FAIR Semantic Mappings Leveraging on RO-Crate. Open Conference Proceedings. 5. 10.52825/ocp.v5i.1179.

[12] De Castro, P., Herb, U., Rothfritz, L., & Schöpfel, J. (2023). Building the plane as we fly it: the promise of Persistent Identifiers.