

# Implementing information container to link multi-models for Semantic Digital Twins

Yingying Zhang<sup>1</sup>

<sup>1</sup>RWTH Aachen University, Schinkelstraße1, 52062, Aachen, Germany

E-mail(s): yzhang@dc.rwth-aachen.de

**Abstract:** Infrastructure Digital Twins (DTs) are commonly understood as up-to-date virtual representations of the physical infrastructure assets. The Digital Twins data pattern contains not only geometric and spatial aspects of the physical entities but also incorporates heterogeneous information including specifications, sensor data, performance metrics, maintenance records, and operational reports. At the same time, the continued expansion of the Digital Twin system without proper modeling, decoupling, and optimization of control workflows will further reduce its manageability and reliability. This paper presents a data archiving method using an Information Container for linked Document Delivery (ICDD), which is primarily based on ISO 21597, to containerize heterogeneous Digital Twins payload in arbitrary formats and provide validation services. The proposed container uses a standardized structure to implement services such as data linking, version control, and access control, thus making the Digital Twin data archiving more efficient and reliable.

**Keywords:** Digital Twin (DT), Information Container for linked Document Delivery (ICDD), Linked Data, Semantic Web



Erschienen in Tagungsband 35. Forum Bauinformatik 2024, Hamburg, Deutschland, DOI: 10.15480/882.13551  
© 2024 Das Copyright für diesen Beitrag liegt bei den Autoren. Verwendung erlaubt unter Creative Commons Lizenz Namensnennung 4.0 International.

## 1 Introduction

The emergence of digital twins has led to the enrichment of static BIM systems in the temporal dimension, which has improved AEC productivity to some extent, but at the same time, more complex data management challenges need to be faced [1]. The initial challenge is how the digital twin infrastructure in the AEC domain can manage data from heterogeneous sources efficiently. Heterogeneous data formats and communication protocols require standardized processes. Linked data provides a way to facilitate the integration of heterogeneous datasets that might otherwise be incompatible, and can address interoperability issues to some extent to provide a more comprehensive DT model [2].

Secondly, the analysis of digital twin systems is highly reliant on large datasets, usually consisting of a collection of dynamic data of the temporal evolution of physical entities. This type of dynamic data is usually stored for access in time series databases, whereas contextual information about the DT system such as geometric data, and sensor contextual information is often stored separately in a static format. The underlying raw data does not need to be exposed to the top users, either for security or

administrative reasons, who only need access to the high-level knowledge or optimization conclusions generated by synthesizing the raw data. Therefore, the integration and unified access to dynamic data and static contextual information is far from easy and is still an active research area.

The emergence of the ICDD (Information Container for Linked Document Delivery) standard provides a container that can be used to standardize the management and delivery of data and documents in AEC engineering projects. It can provide a common platform to support multiple data formats and standards to ensure interoperability between different systems and tools, and at the same time can create links for heterogeneous resources. Currently, ICDD has been applied to solve some complex heterogeneous project unified information management, such as the infrastructure documentation management [3], or the automatic linkage of BIM model and external systems [4].

DT systems that collaborate across multiple domains rely on high-quality native data, and the use of information containers allows for effective management of DT projects to ensure that digital twins are always synchronized with physical entities. However, ICDD is an information container designed for static data, and the representation of DT, a system based on time-series iterations, becomes more complex. In this paper, we propose a research method that aims to link heterogeneous dynamic DT data in ICDD container and realize fully automatic updating of the linked information in the information container. This approach achieves effective linking of DT heterogeneous data while ensuring data storage efficiency, and saves the effort required to repeatedly create information in the container.

## 2 Background

### 2.1 DT data archival

Interoperability has always been one of the issues facing in the AECO as it is an industry that requires multi-disciplinary collaboration resulting in heterogeneous and highly fragmented data. While DT compared to static BIM systems the temporal evolution characteristic is added, which requires well-defined data structures to maintain the accurate operation of this system for decision support and prediction.

Currently the BIM domain is starting to evolve to address interoperability issues by using Linked Data approaches to ensure the exchange of heterogeneous data, such as the evolution of the IFC format to ifcOWL, and the emergence of open data paradigms such as the building topology ontology (BOT), where closed BIM systems are starting to move towards the stage of open BIM based on semantic modeling. The Semantic Web has proven its feasibility in addressing the interoperability of static BIM systems, e.g., using Linked Data to link BIM with IoT systems. Still, most of these studies have converted the data uniformly into RDF or OWL formats for storage and querying.

Apart from this, some research has started to focus on the effectiveness of the Semantic Web in representing dynamic data [5] [6]. Hybrid storage approaches have been proposed for the data characteristics of DTs, either in the form of data lakes or decentralized data modeling, which are able to combine data from multiple sources to provide a unified view and enable more efficient native data storage for data-intensive analytics (e.g., machine learning). Nevertheless this hybrid storage

approach has increased difficulty in unified access to data, and a number of data federated querying approaches have been proposed [7]. Overall, hybrid data storage approaches can effectively utilize the potential of linked data in solving interoperability problems as well as enabling flexible and efficient storage of massive amounts of data.

## 2.2 ICDD as the information container for AECO sectors

Document-based collaboration has been used in the AECO field until now, but traditional document management methods can lead to information fragmentation and confusion. Originally an information container for multi-domain data exchange was developed in the Dutch COINS project, and this approach to information containers was eventually standardized as ICDD. The ICDD, i.e. Information Container for Linked Document Delivery, defines a standardized method for managing and delivering linked documents in projects in the AEC field. In ISO 21597 [8], the structure of ICDD container is defined in detail and two ontology resources, Container ontology and Linkset ontology, are proposed to be used for interlinking documents inside the container.

The effectiveness of ICDD in managing static AEC documents has been demonstrated many times and continues to evolve, whether it is for the effective management of BIM models and drawings [4], to facilitating the integrated delivery of BIM models and other heterogeneous information. Ye [9] in this study explores the use of information containers and blockchain to ensure automated payment for construction, through the use of ICDD, where the BIM model, BoQs (Bill of Quantities) and QTO (Quantity Take-Off) are effectively interconnected in the information container. A micro-service based approach for BIM project document delivery has also been proposed, applying ICDD in open CDE can provide web-based online services [10]. However, projects in the AEC domain are based on the process of continuous time synchronization, so the dynamic functionality extension for information containers is concerned, and a dynamic information container feature is proposed to enable automatic update of version control [11]. The question of how to handle high volume time series data, which is the essential part of Digital Twin, can be efficiently represented in information containers will be explored in this paper.

## 3 Prototype design: ISO 21597-ICDD

Linked Data has been applied to address the interoperability of DT heterogeneous data sources, meanwhile it has become common sense in recent years Linked Data is not optimally efficient in representing high throughput data like sensor data, so this study envisions an ICDD-based Digital Twin data modeling approach that manages dynamic data in chunks in information containers and abstracts them into knowledge represented on Linked Data, thus facilitating interlinking of contextualized DT data and faster query paths.

Research Questions (RQs) outline the scope of how the data modeling method can effectively archival heterogeneous DT data, below lists the research questions:

- **RQ1:** How can heterogeneous DT data be more efficiently archived?
- **RQ2:** How to facilitate real-time collaboration and communication of DT's dynamic data?

- **Hypothesis 1:** Hybrid storage scheme: the time series raw data is decomposed into different json files according to the chunking principle, and descriptive labels are generated and stored in the knowledge graph for querying.
- **Hypothesis 2:** Use ICDD to facilitate standardized DT data exchange: ICDD is appropriate for federating heterogeneous project data while adhering to industry standards and regulatory document management and delivery processes, which helps to ensure that projects efficiently facilitate standardized DT data exchange in a compliant manner.

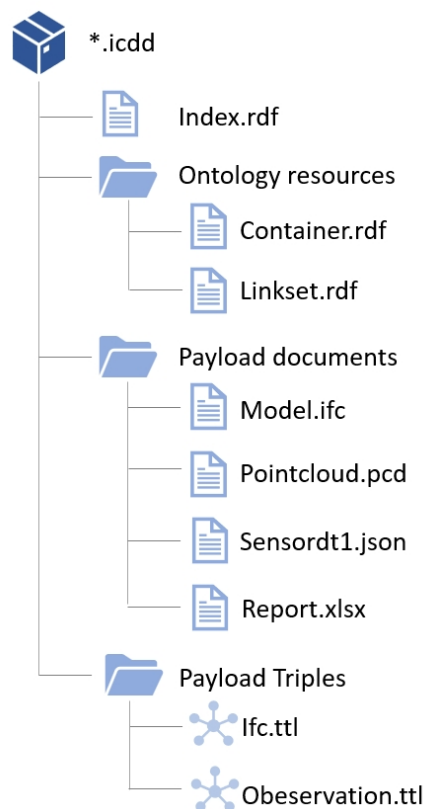


Figure 1: The file structure of the proposed scheme

### 3.1 Automated Dynamic data linking in ICDD

The growing volume and complexity of data generated by DT compared to static BIM systems require more effective data archiving strategies for efficient data storage, retrieval, and analysis. The DT model in the AEC domain, mainly consists of several data models, which also can be seen in Figure 2:

- **Time series model:** it mainly consists of real-time monitoring information generated from actuators and sensors, and is mainly stored in databases that are optimized for time-stamped or time series data, such as InfluxDB.
- **Geometric model:** the virtual representation of physical entities, used to represent relationships and interactions within a DT system and to facilitate pattern recognition.

- **Relational model:** mainly includes structured data, including system log records, texts, spreadsheets, etc.
- **NoSQL model:** includes both semi-structured and structured data, such as document-based storage, key-value storage, and wide-column storage, etc.

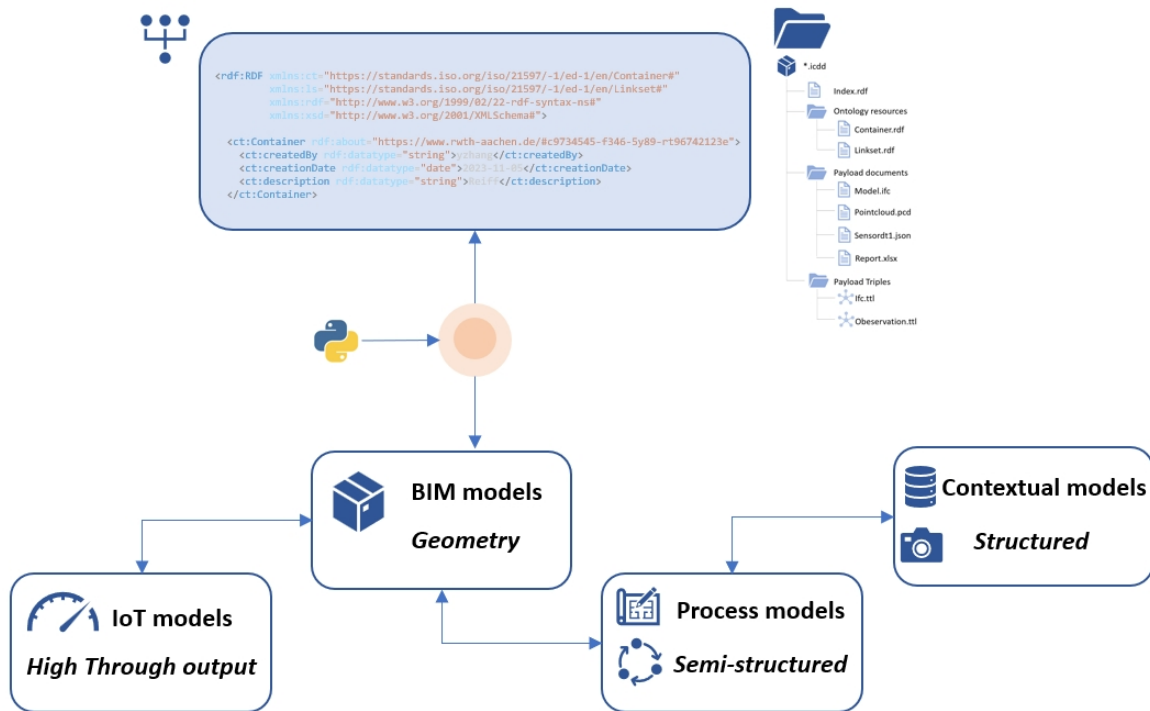


Figure 2: Information container work as the media for Multi-model digital twin data

So in this paper we designed a multi-data model for DT data from physical entity to virtual migration, and manage the DT data dynamically in the information container, using the function of ICDD to create links between different data elements in the container and also between multiple containers. The specific steps are as follows:

- 1 Physical entity information is transmitted via MQTT to the time series database for storage
- 2 Automatically create links to ICDD containers from InfluxDB datastore buckets chunked according to the chunking principle
- 3 The time series data is divided into different payload documents to store and automatically generate Identifier and RDF document reference, which is stored in index.rdf

To validate the proposed method of multi-modeling application, a case study was executed using sensor monitoring data in a university building, which is more complex compared to a static BIM model, where firstly, spatial information is modeled as an IFC model, while three air quality sensors are placed in different rooms to monitor the air quality in real-time. The specific data schema is shown in Figure

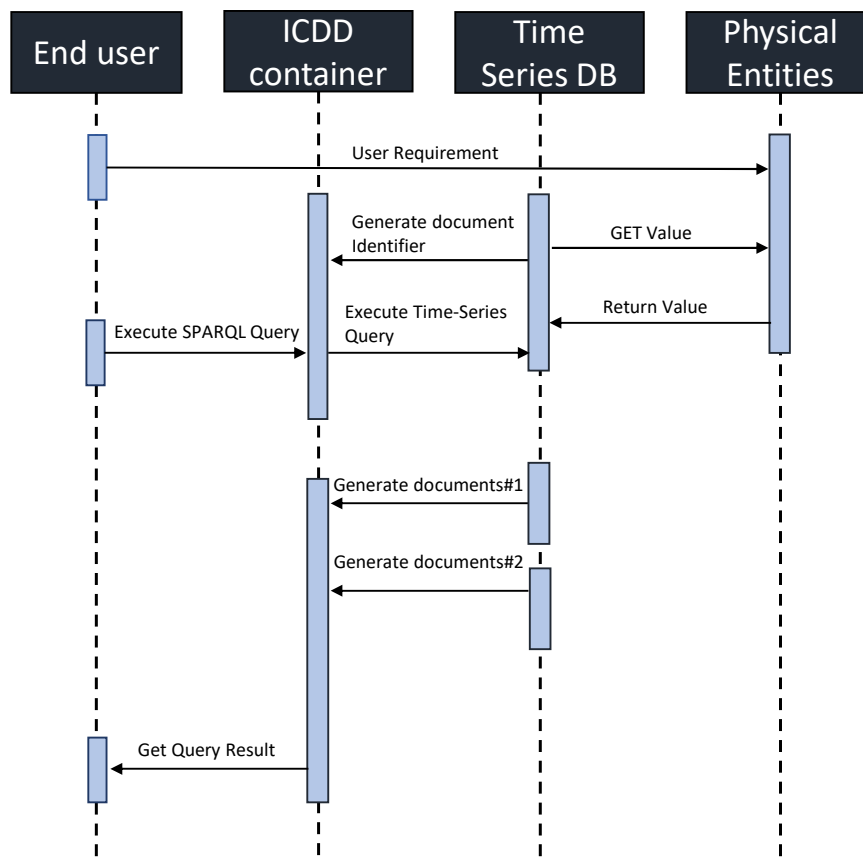


Figure 3: Overview sequence diagram of the proposed solution

4 Information from the environment is collected and converted into digital signals by micro-sensors and Raspberry Pi, subscribed and stored in buckets in the time series database through the MQTT protocol, and converted by the Python middleware, which allows linking the time series database directly to information containers and dividing them into different json files according to the chunking principle.

In this case study it is possible to provide the client with: links between the representation of the physical entity through the IFC space and the values of the aggregation operations on the observation values, links between physical entity model elements and inspection reports and photos.

## 4 Discussion and Conclusion

This paper highlighted how dynamic data and heterogeneous data models can be efficiently stored and interlinked with each other in the information container. It explored how dynamic and contextualized data interoperability can be achieved while ensuring storage and querying efficiency. Hence, an approach based on linking data in information containers is used, which by dynamically linking time-series databases and automatically generating summarized information in information containers, is able to create formal links between dynamic data and contextualized information is necessary for the

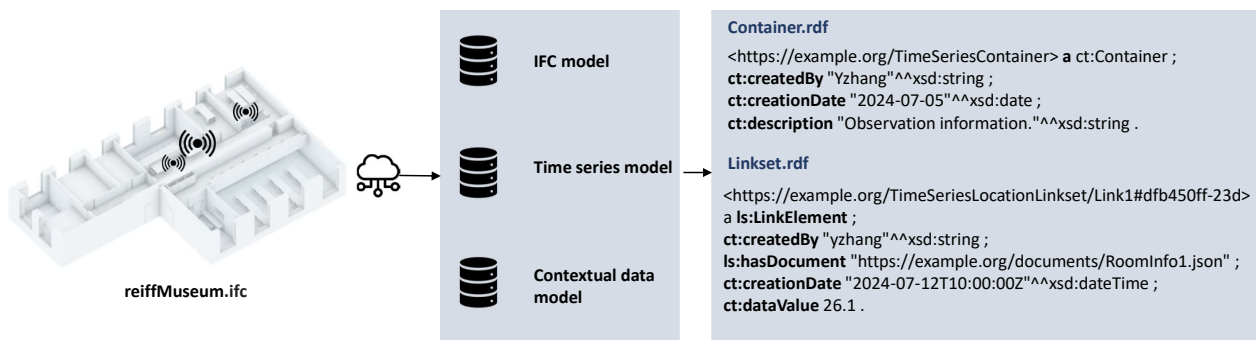


Figure 4: Using multiple containers to link multiple models in the use case Digital Twin scenario

Digital Twin system for data archiving, pattern validation and support for retrieval of context-supported observation information.

The AEC domain has brought great progress to the industry with the collaboration of multiple stakeholders, from drawing design to virtual modeling of physical entities, the intercommunication between multiple specialized models and systems plays an increasingly important role. From BIM to dynamic DT systems, the constantly proposed models put high demands on the matching of legacy data, so the significance of this research is to explore the usage of information containers as a medium to link BIM and complex DT systems by trying to use ICDD as the top-level data model architecture, heterogeneous geometric models, time series models, process models, and structured context models can be effectively integrated and linked in the information container.

Meanwhile, using ICDD containers can facilitate standardized data exchange workflows, thus resulting in increased reliability and consistency throughout the project lifecycle. However, ICDD currently stays at the standard level, and the compatibility with existing tools and legacy data is yet to be developed. In addition, whether the use of information containers to federate digital twin data is an optimal solution needs to be explored and verified, and the ability of Linked Data to effectively federate heterogeneous data has been demonstrated, but its current ability and efficiency to handle dynamic data still needs to be explored.

## References

- [1] A. Borrmann, J. Schlenger, N. Bus, and R. Sacks, "Aec digital twin data-why structure matters", in *International Conference on Computing in Civil and Building Engineering*, Springer, 2022, pp. 651–669.
- [2] C. Boje, A. Guerriero, S. Kubicki, and Y. Rezgui, "Towards a semantic construction digital twin: Directions for future research", *Automation in construction*, vol. 114, p. 103 179, 2020.
- [3] A. Göbels and J. Beetz, "Using icdd containers for documentation data archives: Semi-automated creation of linked documentation data archives using an information container for linked document delivery", in *Proceedings of the 27th International Conference on Cultural Heritage and New Technologies*, 2022.

- [4] A. Borrmann, J. Abualdenien, and T. Krijnen, “Information containers providing deep linkage of drawings and bim models”, in *Proc. of the CIB W78 Conference 2021*, 2021.
- [5] S. Hu, E. Corry, E. Curry, W. J. Turner, and J. O'Donnell, “Building performance optimisation: A hybrid architecture for the integration of contextual information and time-series data”, *Automation in Construction*, vol. 70, pp. 51–61, 2016.
- [6] Y. Zhang and J. Beetz, “Semantic sensor data federation in dynamic knowledge graphs using rdf-star”, in *Proceedings of 30th EG-ICE: International Conference on Intelligent Computing in Engineering*, European Group for Intelligent Computing in Engineering, vol. 30, 2023.
- [7] M. Bakken and A. Soylu, “Chrontext: Portable sparql queries over contextualised time series data in industrial settings”, *Expert Systems With Applications*, vol. 226, p. 120 149, 2023.
- [8] ISO, “Information container for linked document delivery—exchange specification—part 1: Container”, *ISO 21597-1*, 2020.
- [9] X. Ye, K. Sigalov, and M. König, “Integrating bim-and cost-included information container with blockchain for construction automated payment using billing model and smart contracts”, in *IS-ARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, IAARC Publications, vol. 37, 2020, pp. 1388–1395.
- [10] M. Senthilvel, J. Oraskari, and J. Beetz, “Implementing information container for linked document delivery (icdd) as a micro-service”, in *EG-ICE 2021 Workshop on Intelligent Computing in Engineering*, Universitätsverlag der TU Berlin, 2021, pp. 66–73.
- [11] N. N. Al-Sadoon, R. Scherer, and K. Menzal, “From static to dynamic information containers”, in *EC3 Conference 2023*, European Council on Computing in Construction, vol. 4, 2023, pp. 0–0.