Conceptualizing Decentralized Information Containers for Common Data Environments using Linked Data

Madhumitha Senthilvel, senthilvel@dc.arch.rwth-aachen.de
Design Computation Chair, RWTH Aachen University, Aachen, Germany

Jakob Beetz, j.beetz@dc.arch.rwth-aachen.de
Design Computation Chair, RWTH Aachen University, Aachen, Germany

Abstract

Common Data Environments (CDEs), which are crucial for ensuring data interoperability and reducing information loss during data exchange, use Information Containers to facilitate storing, retrieving, manipulating such information. While numerous efforts have proposed minimal conformance specifications for such containers, they don't address container-based federated information management. Since AEC industry relies on information represented in different formats, these specifications are not applicable. As Semantic Web (SW) technologies provide mechanisms for linking such disparate information, this paper conceptualizes Information Containers by combining these SW concepts with existing AEC-CDE specifications. Developed from our earlier experiments, we propose an architecture for complex functionalities like meta-containers, container interlinking, maintaining link integrity, version management etc. It considers information as the primary source (data blobs), which are encapsulated in containers (considered virtual). Demonstrated by a preliminary prototype, this work's main contribution is the conceptualization of Information Containers which uses specifications from existing approaches while accommodating interlinked data.

Keywords: Information Containers, CDE, ICDD, LDP, DIN SPE 91391, ISO 19650

1 Introduction

Construction projects are fundamentally "distributed" in nature owing to their numerous multi-disciplinary stakeholders participate across various phases of the lifecycle of the project, each bringing in their own data and tools. Due to the resulting complexity from the above system, where collaboration is crucial, Building Information Modelling (BIM) has indeed proved invaluable. In particular, Common Data Environments (CDEs), which belong to level 3 in the BIM Maturity framework, have facilitated a "Single Source of Truth", for managing such information. These CDEs rely on the concept of Information Containers for storing, retrieving, manipulating information (files or web resources). A CDE is a platform on which project actors can store their data, share them with other participants, potentially also modify the data and also use tools on them. Such a platform can be locally hosted, or also be available as a web-based CDE.

Presently there exists a range of specifications from various organizations for defining the functionalities, scope, of Information Containers in CDEs, which can be split into two groups:

national and international standards such as ISO 19650[1], DIN SPEC 91391-2[2], ISO 21597[3] and other approaches such Linked Data Platform (LDP) [4](W3C Recommendation), Open CDE-API[5] (buildingSMART). Each standard in the former group builds on the concepts introduced in PAS 1192:2003 for information containers for information exchange, by gradually dealing with increasingly complex functions of containers, such as the ISO 21597, which specifies linking documents inside containers. Despite the above specifications building and deriving concepts from each other, and the latest ISO 21597's detailed specification, presently there exists no comprehensive overview of how a functional information container, which follows the requirements set by all these standards should be implemented in a web-environment.

In our previous works, we implemented a preliminary version of the linked data container concepts of ISO 21597 in a decentralized web environment (Senthilvel et al. 2020). In this work we only looked at information linking as per ISO 21597 and complex container concepts such as meta-containers, container-to-container linking, maintaining link integrity, version management etc. were not addressed. In this paper, based on our earlier experiments, we propose a container architecture that deals with such complex specifications. The conceptualization proposed in this work is based on considering information as the primary source (data blobs), which are encapsulated in containers (which are considered virtual). The proposed approach's point of divergence from the existing proposals such as ISO 21597 and its associated standards/approaches, is that, in this approach: "containers are the primary sources for holding data, thereby necessitating replication of a file/information for use in different containers". This work's main contribution lies in the conceptualization of Information Containers for practical aspects which uses specifications from existing approaches and also accommodate linked data.

In section 2, the existing AEC standards for Information Containers and the Linked Data Platform are discussed, leading to a short discussion of the research gap identified, which served as the motivation for this paper. Based on the identified gaps, in Section 3, a conceptualization for Information Containers is presented. Finally, Section 4 presents concluding remarks, limitations of this work and planned future directions of research.

2 Existing Methods and Approaches

The data that resides inside Information Containers can be broadly classified as structured or unstructured. The former includes information that is machine-readable such as geometrical models, schedules, information in databases while the latter can include textual/imagery-based information such as photographs, documentations in the form of drawings, video clips, sound recordings etc. In a CDE, these containers are created, aggregated, processed and distributed through a controlled process (Hartmann *et al.*, 2019).

Presently there exists a range of specifications from various organizations for defining the minimum functionalities and scope of these containers in CDEs. In this section a concise summary of the above standards is presented which is used to assess and establish where we stand in terms of Information Containers, and what is required of them.

[1] Builds on the previous PAS1192:2003 standard with additional specifications for meta-data for container

[2] German standard defining OpenCDE through requirements for communication interface between CDEs and software

[3] Information Container for linked Document Delivery (ICDD), formerly known as "Information Container for Data Drop"

[4] Linked Building Data (LBD) is defined by W3C Linked Building Data Community Group https://www.w3.org/community/lbd/

[5] https://github.com/buildingSMART/OpenCDE-API

2.1 AEC Approaches

Most of the efforts in the Architecture, Construction and Engineering (AEC) domain for specification of Information Containers stem from the initial introduction of CDEs in PAS 1192: 2003[6]. The BIM Maturity levels conceptualize how at each stage, information is gradually transitioned from a file-based, i.e. the least level of digitalization to storing information in databases to full leveraging of web-based functionalities, which can be seen in Figure 1. While different organizations, countries and sub-domains in AEC are in different BIM maturity stages, the standards which address CDEs have also gradually evolved. These include the now-withdrawn PAS 1192:2003 which is superseded by ISO 19650[7], DIN SPEC 91391[8] and ISO 21597. Each of these builds on the concepts introduced in PAS 1192:2003 for information containers for

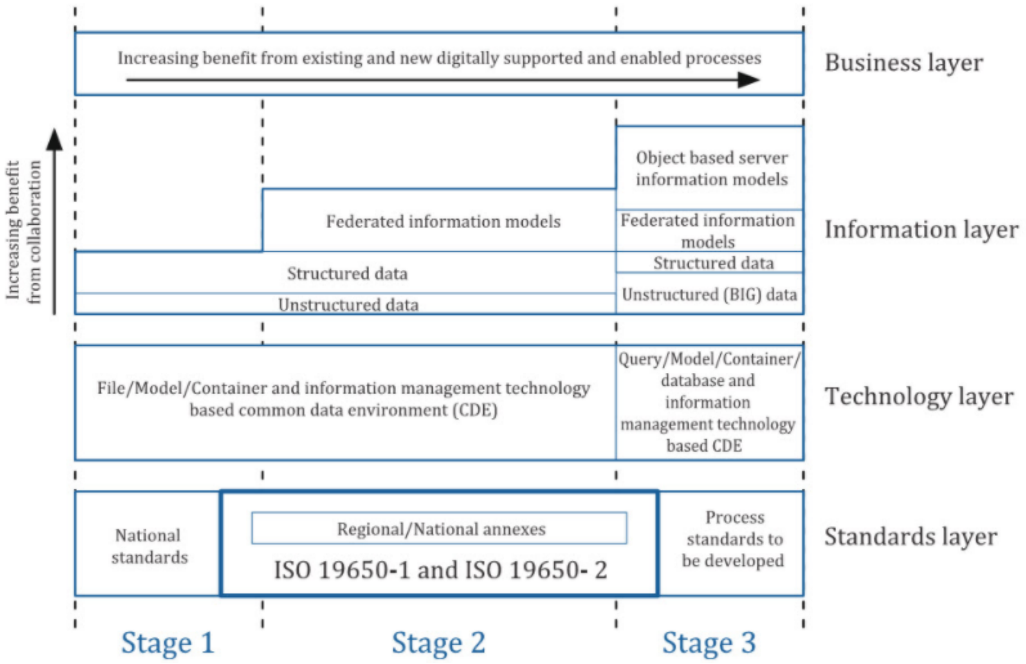


Figure 1 BIM Maturity Stages as per ISO 19650:1-2018

information exchange, by gradually dealing with increasingly complex functions of containers, such as the ISO 21597, which specifies linking documents inside containers. Most of the presently available standards contain specifications for catering to stage 2, while the work on implementation of these standards falls within the range of stage 1 & 2.

In this sections, three of these standards are discussed: the ISO 19650, which kicked-off the concept of Information Containers in level 2, the DIN SPE 91391 which continued by delving into the functional aspects of such containers, and the ISO 21597 which focuses on the breakdown structure of containers, from a Linked Data perspective.

2.1.1 ISO 19650:1 & 2 - 2018

A four-part series, this standard describes the principles, concepts, terminology and implementation specifications for collaborative information management. Part 1 defines certain requirements for containers such as a unique identification of information container (GUID), an agreed-upon codification of the filename, display status, revision history and classification attributes, state transitions logs, authoring information and access control at the container level (ISO 19650-1, 2018). Containers, according to this standard has to be collaborative in order to

---

[6] Specification for information management for capital/delivery phase of construction projects using building information modelling

[7] Organization and digitalization of information about buildings and civil engineering works, including building information modelling (BIM) – Information management using building information modelling

[8] Common Data Environments (CDE) for BIM projects – Function sets and open data exchange between platforms of different vendors

produce federated models, which in turn require a federation strategy. Though it is mentions that this strategy should include explanation on how the federated model should be divided in containers, how such containers would be linked, and also the breakdown structures for such a container, it does not contain additional specifications or offer an example for the same.

While the ISO 19650 kick started the initiative on conceptualizing how an Information Container with minimum set of functionalities, they are generic and refer to other standards (either local or regional or national). Consequently, while this standard broadly defines the role of the parties involved in such information collaboration, their tasks, responsibilities, and also certain metadata such as file naming conventions, etc. it does not address a majority of the technical details and other metadata such as versioning, federated models, container aggregation. Furthermore, it acknowledges that its applicability is aimed towards maturity level 2, while being partially applicable to level 1 & 3. Some parts of these gaps are addressed by the DIN SPEC 91391 series, which is elaborated in the next sub-sub-section.

2.1.2 DIN SPEC 91391 – 2019

This series of standard, which also consists of 4 parts, is a German national standard which focuses on addressing the functional and detailed specifications for CDEs and also containers. Unlike the ISO 19650, this standard is intended for exploration of specifications for the development of a technology that is yet to be standardized (Hartmann *et al.*, 2019). Mainly led by interested stakeholders, it focuses on the principles of data management which were introduced in the previous sub-section by elaborating on the functionalities required for establishing and configuring containers and the data inside them, in a CDE. This includes specifications which extend the concepts of metadata for container states, revision history, classification, state logging for when containers transition to different states, history of authors/modifiers and access control at the container level. A broad visualization of the container is shown in Figure 2. For brevity, a concise summary of the concepts from this standard is given below.

- Data integrity to be maintained through having an unrestrictedly comprehensible history and verifiable processing
- Nested/aggregated containers through which multi-models can be formed based on information in each of the containers
- Metadata description for normal and nested/aggregated containers
- Filtering functionality: creation of subset of containers through metadata-based queries
- Structuring of containers through linking
- Accept external references to internal data/container structures (e.g: external model linked to individual model elements in a container)

Comparing with the ISO 19650, it is clear that the DIN SPEC addresses the practical aspects of information management by going into detail on the various forms containers can take, what kind of information and linking should be theoretically possible, and the metadata for both the documents and the containers. Additionally, it also contains specifications for the functioning of these containers in a web-based CDEs. From the above, it is also evident that there are some overlaps between the functionalities of LDP, and ISO 21597 (both of which would be explained in the upcoming subsection). For instance, the concept of aggregated information containers which can build hierarchically upon each other in order to form a much larger container introduced in this specification, is one of the functionalities put forth by LDP.

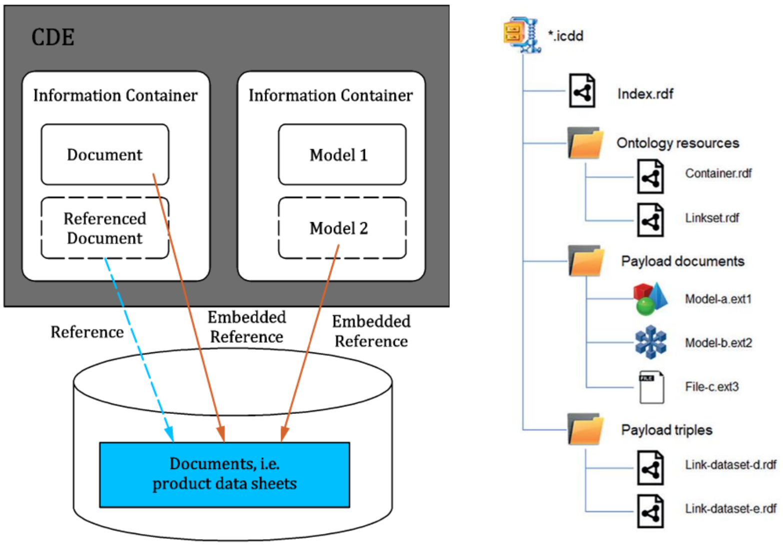


Figure 2 Information Containers as per DIN SPEC 913961 (left image), and as per ISO 21597 (right image)

To summarize, it gives a moderately detailed picture of how containers in BIM Level 3 (refer to Figure 1) would work through conceptualizing that information would be stored in database structures, and they can be referred to at the individual model elements/attributes level (which would make it the smallest unit of information). However, it also acknowledges that the conceptualization for realizing the same was not addressed.

2.1.3 ISO 21597

Part 1 of this standard defines the container structure and the general linking concepts through the definition of a container ontology, corresponding data types and object properties, along with a linkset ontology with corresponding data types and properties (refer to Figure 2). Part 2 defines additional types of links which form the extended linkset. ICDD follows a folder structure which three major components: an Ontology folder containing the schema of the files in the ICDD Zip file format, a linkset folder containing the links, and a payloads folder containing the documents themselves (Technical Committee : ISO/TC 59/SC 13 Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM), 2020). A meta-file called index contains a summary of the ICDD container documents and how they are linked. In brief, the ICDD ontologies together with data types and properties define what kind of meta-information can be used for a file and the links between different files. ICDD includes both the Multi-model approach and the Linked Data approach (introduced in the following sub-section). The primary advantage offered by it is the object-level linking mechanism for connecting heterogeneous data. However, it does not yet have definitions or specifications for CDE-specific functionalities.

2.2 Linked Data Platform

LDP is a specification containing definitions for reading-writing Linked Data architecture (Spiecher *et al.*, 2015). An LDP can be a client or server or a combination of both while conforming to the specification set forth by the W3C Working Group of the same name. It focuses on the use of HTTP to Create, Read, Update, and Delete Linked Data Resources (files/documents) that are part of a collection (folder). While it defines containers of three types, it specifications are considered too broad. Consequently, the structure for linking, and managing resources/files in containers is dependent on the CDE designer.

2.3 Research Gap and Motivation

From the previous sub-sections, it is evident that while ISO 19650 and DIN SPEC 91391 present more specifications for Information Containers, they do not address a comprehensive breakdown structure for a container. In addition, since information in the AEC domain are usually heavily interlinked, these approaches also do not provide mechanisms for enabling such interlinking. Linked Data is seen as a potential solution for enabling such interlinked heterogeneous data (Pauwels *et al.*, 2015). While the container concepts from LDP and ICDD do address this gap, they by themselves do not contain specifications for addressing more complex container use-cases in a detailed manner such as nested containers, and how they would function in a CDE. Thus, this paper aims to combine both the detailed specifications on the overall requirements of CDEs and ICs from the AEC perspective, with the functional concepts and structure of Linked Data, in order to put forth a conceptual framework of Information Containers which caters to decentralized information storage and references.

3 Conceptualization of Information Containers for Linked Data in AEC

In this section, a conceptualization of an Information Container which combines the principles of Linked Data along with the ISO 21597 ICDD structure is presented. Collaborative environments such as CDEs aim to store, re-use and share information efficiently in order to reduce the risk of information loss, contradiction, misinterpretation, redundancies, non-timeliness of information. Metadata are a critical component of ICs since they facilitate capture, storing, organization, querying, and ultimately the use of the data itself. Conventionally, metadata can be encoded in the names of the resources (either the files or the container) through a previously agreed upon convention. While such an encoding a user can understand the meaning/purpose of a document without opening it, since it is not captured in a machine-readable and scalable way, the information it contains cannot be used for further improvement of the knowledge about the IC. Linked Data serializations, by representing such metadata using RDF graphs can help leverage such information and also make it extendable.

3.1 Framework of a Linked Data supported Information Container

The previous section illustrated the gaps in the current standards for container conceptualization which can be used to frame low-level conceptual requirements for containers. Since containers would have to deal with files/resources that would be linked, they should have a structure which

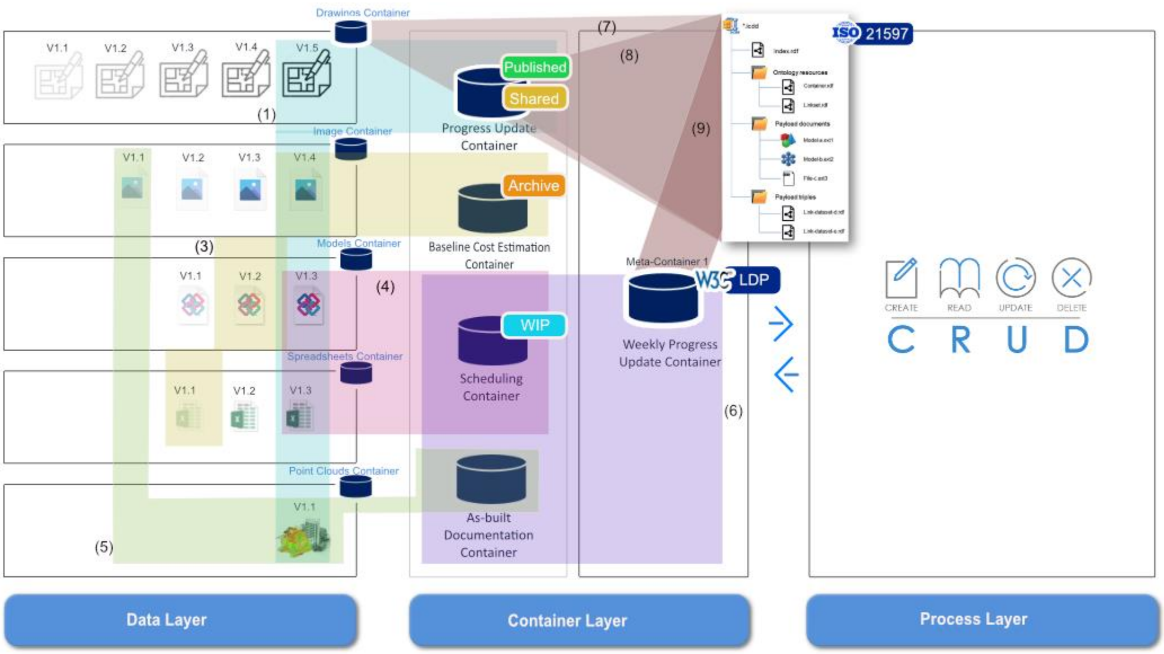


Figure 3 Conceptual framework of a linked-data supported Information Container using ISO 21597

facilitates two levels of linking: file/resource level linking (meta level) and also on the object and attribute level linking. Additionally, since ISO 21597's container structure satisfies these requirements, in this section a conceptualization of how this container structure can be applied in the context of CDEs as envisioned by ISO 19650 and DIN SPEC is presented.

This conceptualization, as seen in Figure 3 has a three layered architecture: the Data Layer, the Container Layer and the Process Layer. The Data Layer forms the fundamental backend of the proposed architecture, which stores the information from any stakeholder of the project. Each file uploaded by user is stored in its own container, which also stores all the succeeding versions of the document. From the ICDD schema, the terminology *ct:priorVersion*/*ct:nextVersion* along with *ct:VersionID* can be used for the above. The Container Layer that builds on top of this Data Layer lets stakeholders create containers based on project usecases. Such containers in this layer make use of the container states specification from ISO 19650 for labeling such as '*Published*', '*Shared*', '*WIP*', or '*Archived*'. This layer also provides a meta-container functionality which aggregated container or nested container wherein two containers can be combined together to form an even larger one. These types of containers utilize the ICDD schema on "link types" for specifying how information within each container is linked to another. All of the containers in the three layers follow the ICDD container structure (annotated as (7)(8)(9) in Figure 4) with the Index Graph storing the metadata and link information (for documents within the container and also between two or more containers). The figure also illustrates 3 major types of ways a

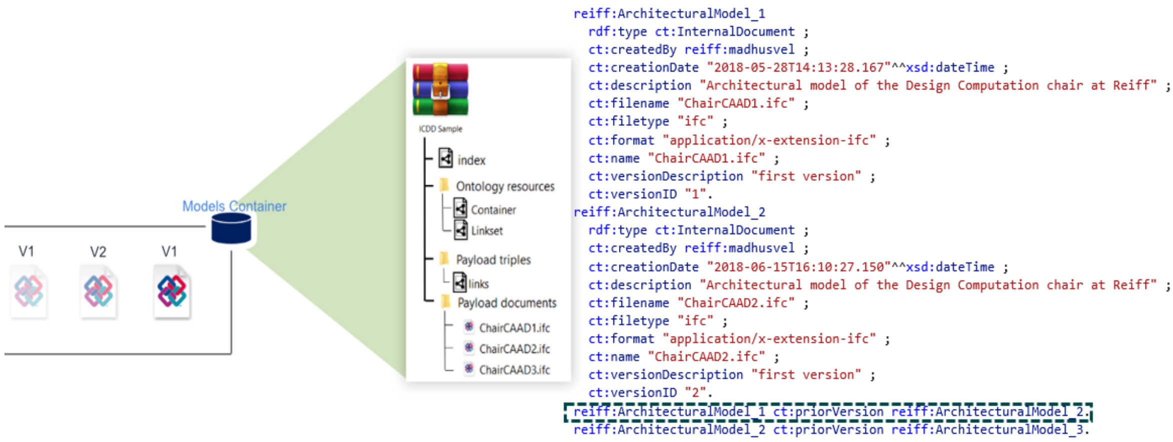
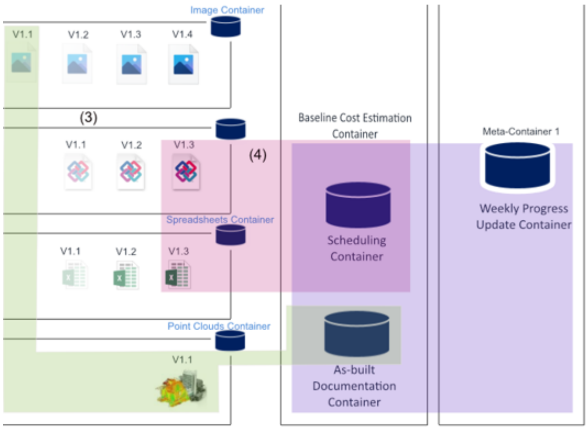


Figure 4 Leveraging ICDD container structure for the Data Layer

container can be realized in this conceptualization. In the section annotated as (1) in the Figure 3, a container has internal reference to all up-to-date files, while the section annotated as (2) is a container which points files with different versions virtually contained. The section annotated as (4) and (5) are containers which are virtually contained by a meta-container (annotated as (6)), thereby utilizing the concept of container aggregation and container nesting.

The proposed architecture puts forth the data blobs as the smallest referable unit. This concept is a departure from the ISO 19650 and the DIN SPEC 91391, where ICs are regarded as the smallest referable unit. The rational for this approach stems from the complications that arise in considering the IC as the smallest transferable unit, because in that case files/resources would be stored in multiple containers, resulting in duplication of data, which finally results in data redundancy, thus conflicting with one of the core requirements of all the CDE standards. However, it should also be noted that, DIN SPEC mentions that for the current file-based OS, a file by itself can be considered as an IC.

Figure 3 shows a short snippet of the how the ICDD vocabulary is leveraged to capture meta data about the different models in the Models Container. The highlighted portion of the code shows the use of *ct:priorVersion* for linking versions which when used in tandem with *ct:versionID* and *ct:versionDescription* can be used for version management.




```
reiff:MetaContainer
  rdf:type ct:Container ;
  ct:conformanceIndicator "IFC Architectural Models" ;
  ct:containsDocument reiff:ChairCAAD1 ;
  ct:containsDocument reiff:ChairCAAD2 ;
  ct:containsDocument reiff:ChairCAAD3 ;
  ct:createdBy reiff:madhusvel ;
  ct:creationDate "2020-07-29T14:24:28.167"^^xsd:dateTime ;
  ct:description “Weekly Progress Update Container" ;
  ct:publishedBy reiff:teamreiff ;
  ct:versionDescription "first version" ;
  ct:versionID "1“.
reiff:Container1
  rdf:type ct:Container ;
  ct:containsDocument reiff:ChairCAAD1 ;
  ct:containsDocument reiff:excel1 ;
  ct:createdBy reiff:madhusvel ;
  ct:creationDate "2020-07-29T14:24:28.167"^^xsd:dateTime ;
  ct:description “Scheduling Container" ;
  ct:publishedBy reiff:teamreiff ;
  ct:versionDescription "first version" ;
  ct:versionID "1".
reiff:Container2
  rdf:type ct:Container ;
  ct:containsDocument reiff:Image1 ;
  ct:containsDocument reiff:Pointcloud1 ;
  ct:createdBy reiff:madhusvel ;
  ct:creationDate "2020-07-29T14:24:28.167"^^xsd:dateTime ;
  ct:description “As-built Documentation Container" ;
  ct:publishedBy reiff:teamreiff ;
  ct:versionDescription "first version" ;
  ct:versionID "1".
reiff:Container2 els:PartOf reiff:Container1.
```

Figure 5 Structuring and linking of Meta Container for nesting/aggregating using ICDD concepts

As already mentioned in section 2.3.1, the concept of nested/aggregated containers has been consistently mentioned in almost all the standards. However, in this work, we present how such nested containers can work by leveraging the container structure of ICDD. Figure 5 sketches out this through the use of an example of the annotated section (6) of Figure 3 through the use of link types as defined by part 2 of the ICDD standard which defines Extended Link Types for linking RDF resources.

4 Discussion and Conclusion

This paper presented a conceptualization of Information Containers using Linked Data principles (and by extension LDP) by combining the container structure as proposed by the ISO 21597 standard for Information Container for linked Document Delivery with the existing specifications for CDEs from ISO 19650, DIN SPEC 91391. The current proposed conceptualization using ICDD is would fall in-between level 2 and 3 because it supports both file based (stored locally) and web-based resources (stored in database) and supports federated model building, it is yet to fully implement concepts such as object based server information model, querying, etc. However, it is planned in the future to explore and integrate these.

4.1 Limitations

On the topic of versioning of data and containers, there are two aspects for which such versioning can be applicable: information change (such as models being iterated) or changes to the linking between the information. With regards to the former, the CDE and the Information Container themselves do not have the capability to either change the data or the containers. Such changes, regardless of whether they are additions, modifications or deletions of information snippet in the data blob are authored by the project participants through overwriting. Hence, these are dealt with through external tools such as model authoring software. However, for the latter, wherein metadata is being changed, these can be captured and serialized as RDF graphs. However, further investigation is necessary to determine the vocabulary, concepts and level of change to be captured.

The framework proposed in paper makes use of the containers themselves for access control. However, if graph based control mechanisms are employed, it would also be possible to define access control to certain data blobs within a container.

4.2 Future Work

At the moment of writing, the API interface for the proposed conceptualization is in development and a preliminary version is available on Github[9]. The proposed conceptualization does not introduce any new vocabulary for capturing meta-data. However, the existing ICDD schema does have its own limitations such lack of terminology for defining the history of a link between files/resources which was created by stakeholder, additional meta-data for defining delta changes between versions of files/resources, how to link external resources in such containers. These topics are under active research.

Acknowledgements

This research was funded through a doctoral research grant from the Deutscher Akademischer Austauschdienst (DAAD) and partly through H2020 EU project on BIM-based Tools for Fast and Efficient Renovation (BIM4Ren - Grant No. 820773).

References

Hartmann, U., Weiß, F., Blume, R., Mueller, T., Aengenvoort, K., Krieger, V., Wang, M., *et al.* (2019), "Common Data Environments (CDE) for BIM projects - Function sets and open data exchange between platforms of different vendors - Part 1 :Components and function sets of a CDE", DIN.

ISO 19650-1. (2018), "Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) — Information management using building information modelling — Part 1: Concepts and principles", ISO, December.

Pauwels, P., Törmä, S., Beetz, J., Weise, M. and Liebich, T. (2015), "Linked Data in Architecture and Construction", *Automation in Construction*, Vol. 57, pp. 175–177.

Spiecher, S., Arwe, J. and Malhotra, A. (2015), "Linked Data Platform 1.0", 26 February, available at: https://www.w3.org/TR/ldp/ (accessed 6 May 2021).

Technical Committee : ISO/TC 59/SC 13 Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM). (2020), "ISO 21597-1:2020 Information container for linked document delivery Exchange Specification - Part 1: Container", ISO, April, available at: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/43/74389.html (accessed 14 March 2021).

Senthilvel, M.,Oraskari, J. and Beetz, J. (2020) Common Data Environments for Information Container for linked Document Delivery. 8th Linked Data in Architecture and Construction Workshop, LDAC 2020, Dublin, Ireland, 17 Jul 2020 - 19 Jul 2020. pp. 132-145.

[9] https://github.com/sbalot/ICDDOnline