Using ICDD Containers for documentation data archives

Semi-automated creation of linked documentation data archives using an *Information Container for linked Document Delivery*

Anne GÖBELS, Chair of Design Computation RWTH Aachen, Germany
Jakob BEETZ, Chair of Design Computation RWTH Aachen, Germany

Keywords: *ICDD Container — linked data — multimedia links — archiving*

CHNT Reference: Göbels, A. and Beetz, J., (2022). 'Using ICDD Containers for documentation data archives - Semi-automated creation of linked documentation data archives using an *Information Containers for linked Document Delivery*', in CHNT Editorial board. *Proceedings of the 27th International Conference on Cultural Heritage and New Technologies, November 2022.* Heidelberg: Propylaeum.
DOI: xxxxxxx.

Introduction

The recording and documentation of existing objects in the built environment include data from various media and provenances. Each medium represents different aspects of the real-world object and has specific functionalities: Textual descriptions are precise and easy to create; photographs can show the as-is state of a surface or represent the entire object with minimal effort. 2D plans and drawings allow uniform graphic documentation and communication about the object, which are not limited to a specific domain. Laser scans provide three-dimensional representations and can reflect the surface's condition and the object's volume. BIM models allow the combination of geometric and semantic data, as well as the integration of data from different domains.

Thus, comprehensive and accurate documentation of an object consists of a multitude of heterogeneous but complementary data sets. The spatial or semantic relationships between these data sources are usually drawn by the creator or user through intensive study. This knowledge is often limited to specific persons or project teams, so third parties users have to acquire this knowledge from scratch. For computer-based applications, the relationships between the data sets are also not apparent and must be implemented manually or assembled in a central model. Furthermore, the provenance of the data can get lost if the distributed, individual data sets are not managed in a structured way complicating the validation of the data quality.

This paper presents an approach using an *Information Container for linked Document Delivery* (ICDD) according to ISO 21597-1 (ISO, 2020) to make the processing and archiving of documentation data more efficient and reliable. An ICDD container allows interlinked storing and exchanging heterogeneous data sets, including their metadata. Moreover, it is shown how the links between the data sets can be partially automated to minimise manual effort. The approach is demonstrated by implementing an ICDD container for existing and newly recorded data of a road bridge. It is part of the TwinGen project funded by the German Ministry for Ministry and Transport, where digital twins for bridges and other infrastructure buildings are created.

Information Container for linked Document Delivery

An *Information Container for Linked Document Delivery* (ICDD) can combine all data belonging to an object of the built environment in one information delivery and semantically interlink the individual files from different data sources.

It enables efficient data management and provides a uniform and open information exchange format. All necessary data sources (models, drawings, texts, sheets, pictures, etc.)) that are needed for a specific purpose (e.g., archiving) are stored in this container. The individual, heterogeneous information units are linked with each other so that all information and data from different sources can be retrieved directly. This approach ensures reliable and comprehensive data transfer/archiving.

The ISO 21597 standard defines the structure and technical implementation of an ICDD. Each container represents a zip file with three main folders and one index file (see figure 2). The folder "Payload documents" contains all documents and datasets to be transferred. The links between those files are stored in the "Payload Triples" folder. They are implemented in the *Resource Description Framework* (RDF[1]), that allows a unique identification of data objects and the semantically defined linking between heterogeneous data resources. The datasets/files can not only be linked at the document level but also at the object level through the use of identifiers (string-, URI-, or query-based). Thus, detailed and sophisticated links can be created, e.g., between parts of point clouds and drawings or image areas and text sections.

The index file also uses RDF and lists all contained records, including their metadata, which enables to track provenances, versions and responsibilities. The ontology for the index file (container ontology) and the link file (linkset ontology) is stored in the third folder ("Ontology Resources"). The ontologies provide the classes and properties to describe the metadata (URI, creator, creation date, modification date, version, file type, file location, etc.), document types (internal, external, etc.,) and link types (e.g. binary or directed links). The created ICDD Container can be queried and read by human users and software using the *SPARQL Protocol And RDF Query Language* (SPARQL[2]).

Creating an ICDD Container for the documentation of bridge data

The aim of creating the ICDD container for the TwinGen project is the *long-term* provision of a *complete* and *linked* dataset, providing a *reliable data basis and structure* that is *accessible* for further analyses and investigations. In the first step, all existing data sources were identified. There is mandatory documentation of the bridge construction and its inspections, which consists of database tables, 2D plans and (damage) pictures stored in a proprietary software system ("SIB-Bauwerke"). The basic structure and condition information was determined beforehand. Additionally, laser scans were created to obtain a 3D representation and information about the current condition. On its basis, an IFC model was created. As seen in figure 1, the documentation of the bridge finally consists of six heterogeneous data sources, each having a different data format.

[1] www.w3.org/RDF
[2] https://www.w3.org/TR/sparql11-query/

Since the data of the ICDD container should be freely accessible and easy to process for this approach, the database tables of the proprietary software system were converted into the open RDF format (Göbels, 2021). For this purpose, an OWL ontology (ASB-ING Ontology[3]) was developed that is compatible with the structure of the database tables (Göbels and Beetz, 2021). A converter[4] also transformed the IFC model into an RDF representation for easier processing. All existing and created data were then added to the ICDD container and described in the index file.

The heterogeneous data sources are then linked at the object level. As these describe aspects of the same real-world object, it is possible to achieve rule-based, automated linking for parts of the data sources based on semantical or spatial comparisons. Since both the inspection data and the IFC model data were available as RDF data based on defined ontologies, it is possible to automatically create the links between them by mapping the classes of the ASB-ING Ontology to the IFC classes (e.g. “Retaining Wall” to “IFC Wall”). Components in the bridge documentation that were only implicitly linked are now assigned to a 3D representation in the IFC model in an explicit, machine-readable way.

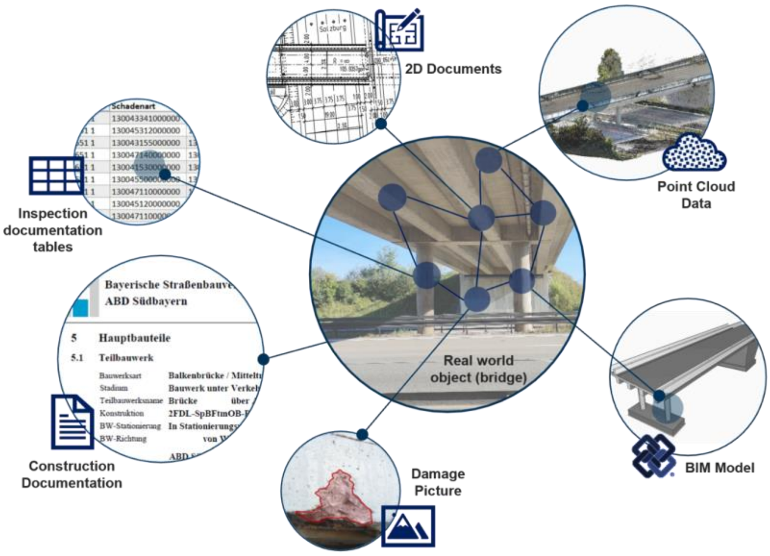

Fig. 1. Heterogeneous data sources of a bridge documentation

The recorded damages of the inspection data should also be assigned to model elements. The damage to components is only described and located textually in the existing inspection reports but with a defined vocabulary (right, left, rear, front, top, large, small, xx cm distance from.., etc.). This vocabulary enabled automated processing and linkage with areas in the IFC model using spatial filter processes and SPARQL queries. The links between damage descriptions and damage images are implemented based on an existing textual assignment in the inspection reports.

Finally, the automatic creation of object-specific links between the existing documentation/inspection reports and the new IFC model and between the damage images and damage descriptions is achieved (see figure 2). In the following processes, links will also be created between the 2D plan contents and the semantic and three-dimensional components, as well as between the point cloud segments and the corresponding other data fragments.

[3] ASB-ING Ontology: https://w3id.org/asbingowl/core
[4] IFCtoLBD Converter: https://github.com/jyrkioraskari/IFCtoLBD

The resulting ICDD Container represents a complete, multimedia interlinked information model of heterogeneous data sets with provenance and versioning metadata, independent of software solutions that can be used for archiving and further processing.

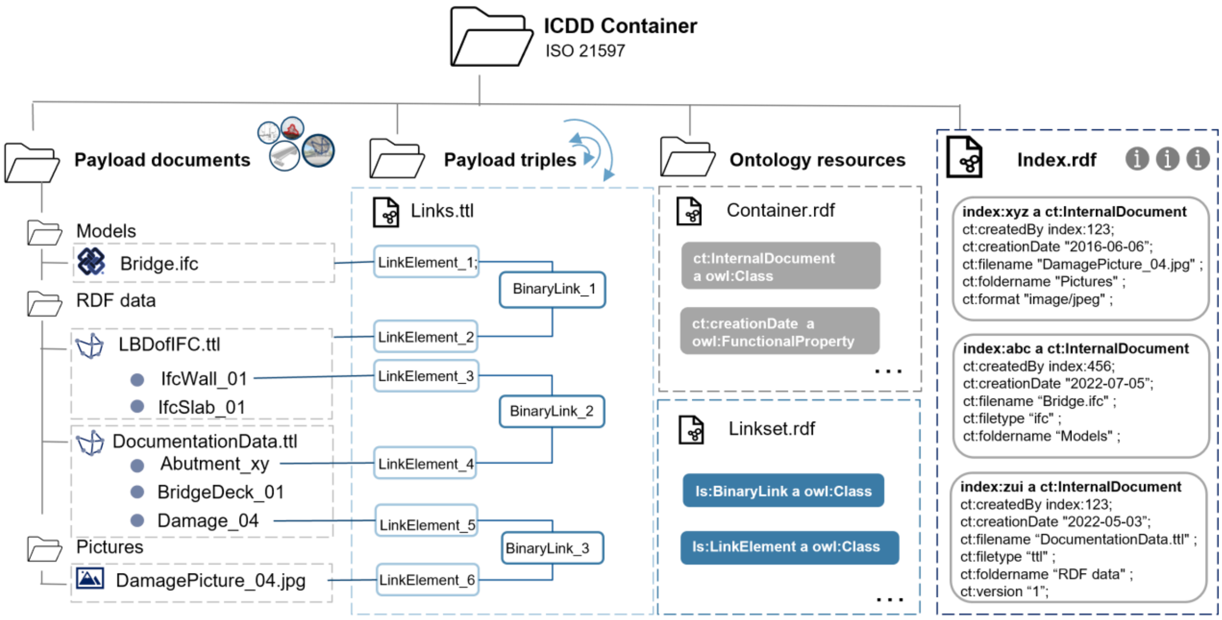

Fig. 2. Schematic illustration of the ICDD Container for bridge documentation data

Funding

This research is part of the "TwinGen" project funded by the German Federal Ministry for Digital and Transport (BMDV). We thank the "Autobahn GmbH des Bundes - Niederlassung Südbayern" for providing the existing bridge documentation for the case study.

Author Contributions

Conceptualization: Jakob Beetz, Anne Göbels
Data curation: Anne Göbels
Resources: ZPP Ingenieure, Autobahn GmbH des Bundes - Niederlassung Südbayern, Simon Vilgertshofer (TU München)
Supervision: Jakob Beetz
Writing – original draft: Anne Göbels
Writing – review & editing: Jakob Beetz

References

International Organization for Standardization (2020). *ISO 21597-1:2020 Information container for linked document delivery - Exchange specification - Part 1: Container.* Geneva. International Organization for Standardization.

Göbels, A. and Beetz, J. (2021).' Conversion of legacy domain models into ontologies for infrastructure maintenance', *Proceedings of the 9th Linked Data in Architecture and Construction Workshop*, Luxembourg. 11th – 13th October 2021. CEUR-WS, pp. 20 – 31. Available at: http://ceur-ws.org/Vol-3081/02paper.pdf (Accessed: 13 July 2022).

Göbels, A.(2021). 'Conversion of infrastructure inspection data into Linked Data models', 32. *Forum Bauinformatik 2021*, Darmstadt. 09th – 10th September 2021. TU Prints. Available at: https://tuprints.ulb.tu-darmstadt.de/21521/7/32.%20Forum%20Bauinformatik%202021_V1.1.pdf (Accessed: 13 July 2022).