D3.3.1 Metadata schema extension for archival systems

DURAARK

FP7 – ICT – Digital Preservation
Grant agreement No.: 600908

Date: 2014-01-31
Version 1.0
Document id. : duraark/2013/D.3.3.1/v1.0

Grant agreement number	:	600908
Project acronym	:	DURAARK
Project full title	:	Durable Architectural Knowledge
Project's website	:	www.duraark.eu
Partners	:	LUH – Gottfried Wilhelm Leibniz Universitaet Hannover (Coordinator) [DE] UBO – Rheinische Friedrich-Wilhelms-Universitaet Bonn [DE] FhA – Fraunhofer Austria Research GmbH [AT] TUE – Technische Universiteit Eindhoven [NL] CITA – Kunstakademiets Arkitektskole [DK] LTU – Lulea Tekniska Universitet [SE] Catenda – Catenda AS [NO]
Project instrument	:	EU FP7 Collaborative Project
Project thematic priority	:	Information and Communication Technologies (ICT) Digital Preservation
Project start date	:	2013-02-01
Project duration	:	36 months
Document number	:	duraark/2014/D.3.3.1
Title of document	:	Metadata Schema Extension for Archival Systems
Deliverable type	:	Report
Contractual date of delivery	:	2014-01-31
Actual date of delivery	:	2014-01-31
Lead beneficiary	:	TUE
Author(s)	:	Jakob Beetz `<j.beetz@tue.nl>` (TUE) Michelle Lindlar `<michelle.lindlar@tib.uni-hannover.de>` (LUH) Stefan Dietze `<dietze@l3s.de>` (L3S) Ujwal Gadiraju `<gadiraju@l3s.de>` (L3S) Qinqin Long `<q.long@tue.nl>` (TUE)
Responsible editor(s)	:	Jakob Beetz `<j.beetz@tue.nl>` (TUE)
Quality assessor(s)	:	Raoul Wessel `<wesselr@cs.uni-bonn.de>` Martin Tamke `<Martin.Tamke@kadk.dk>` Östen Jonsson `<osten.jonsson@ldb-centrum.se>`
Approval of this deliverable	:	Stefan Dietze `<dietze@L3S.de>` (LUH) – Project Coordinator Marco Fisichella `<fisichella@L3S.de>` (LUH) – Project Manager
Distribution	:	Public
Keywords list	:	Metadata, Preservation

Executive Summary

In this report strategies are outlined how relevant metadata for the digital preservation of architectural data and related external vocabularies can be structured and stored. Related requirements are identified in WP7, further specified in WP2 and documented in the deliverables D 2.2.1, D.2.2.2, D.2.2.3 (Requirements documents) and D.7.7.1. Metadata for data identified relevant within the scope of the DURAARK project will be captured in OWL/RDF metadata schemas that re-use existing and recognized metadata standards and only extend them where deemed necessary. The proposed domain-specific metadata structures cover three main aspects:

Descriptive metadata that stores domain-specific summaries and classification information of **buildings as an intellectual entity** type in a LTP system.

Technical metadata that captures information about **Building Information Models (BIM's)** and **point cloud models** which are the two main representation forms of buildings the DURAARK project focuses on.

Linked Data metadata that captures information about the vocabularies and datasets that are used for the semantic enrichment of building models. These are stored separately alongside the individual building model Information Packages (IP) in the archive in a **Semantic Digital Archive (SDA)** for the storage and a **Semantic Digital Observatory (SDO)** for the processing of such volatile, evolving data.

The use of this metadata from IFC models and inference methods are described in generic vendor-neutral terms and exemplified as an extension implementation to the existing archival system. The suggested reuse of existing schemas and vocabularies is illustrated in a number of use cases in section 4.
In the appendices of this document complete listings of the respective vocabularies and schemas are provided.
The results of this document have been created in collaboration with the work and deliverables in other work packages and specifically with work packages 2 and 7.

Table of Contents

1 Introduction

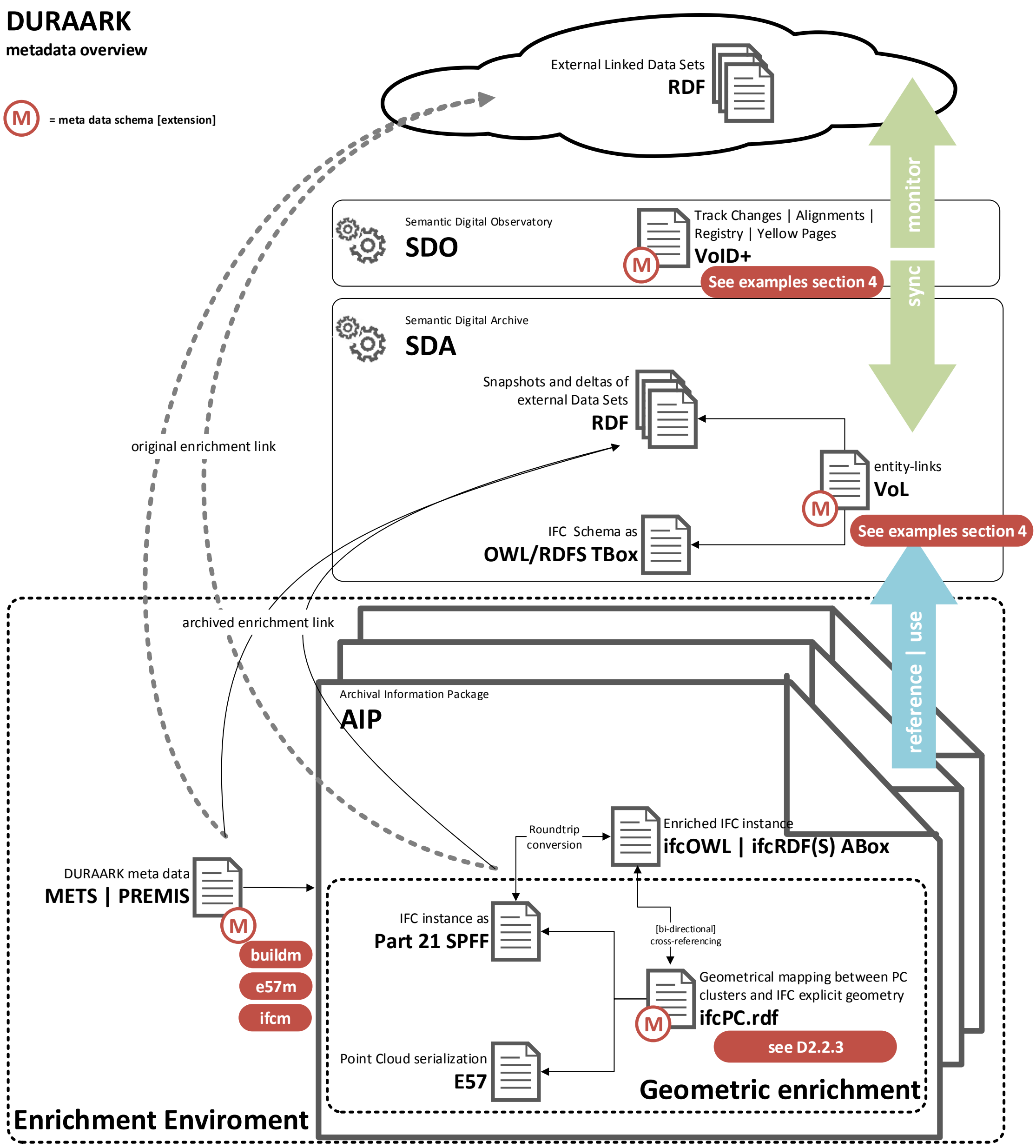


Figure 1: **Usage of Metadata in the DURAARK system**. The structure of this deliverable is matched with the main parts of this illustration. The buildm, ifcm and e57m vocabularies are embedded into the METS|PREMIS containers in the lower left hand corner. The Linked Data metadata vocabularies VoID and VoL are used in the SDA/SDO components. The preliminary metadata vocabulary linking BIM and point cloud representation can be found in Deliverable 2.2.3

The main aim of the DURAARK framework is the digital Long Term Preservation of data about buildings. In a preservation context, buildings are considered the main Intellectual Entity that the DURAARK efforts are focused upon. Like other types of intellectual entities, such as paintings, books or music, only representations of these artefacts can be archived in a digital form. For buildings, a wide spectrum of such digital representations can be used to capture and preserve information covering the various states of a building during its life cycle. This includes representations of the intended ('as-planned') and the observed ('as-built') characteristics of buildings. A major component of the execution and implementation of digital preservation is the collection and organization of data objects of the digital representations into Information Packages (IP) which are stored in an archival system. For many textual and audio-visual forms of digital representations this can be achieved with readily available systems. However, there are forms of digital representations that are specific to the domain of architecture which cannot be properly archived yet. The DURAARK project aims at facilitating the preservation three new main types of digital representations: **1) semantically rich, interoperable Building Information Models (BIM's)** that also contain explicitly modelled 3D geometry and **2) point cloud datasets** acquired from scanning devices. In addition to self-contained information in BIMs, relevant information is also contained in external datasets, such as product information by vendors or classification systems. Many of these datasets can be accessed and linked to on the Web. For complete preservation, these **3) Linked Data** sets have to be stored in addition to the building models and point clouds. These three main types of information forming the **original content** data are described in more detail in section 1.2 of this document.

In order to preserve, describe, index, organize and interrelate information captured in these original content files in accessible, interoperable ways, **metadata** is created that is stored alongside and inside the Information Packages. Such metadata captures information on different levels. Some metadata describes various aspects of the building - the intellectual entity - and its context itself, e.g. location, dimensions, use, ownership, style etc. Such metadata cannot be found and reused in existing metadata vocabularies describing e.g. books or music. Similarly, technical characteristics of the files representing this metadata has to be tailored to the specifics of the respective file formats. While e.g. current preservation systems can generate, maintain various image and document formats using specialized metadata and tools, no such metadata and tools exist yet for BIMs, point clouds and Linked Datasets. In this document the requirements for such metadata are

highlighted, existing metadata vocabularies and specification are examined and finally three new metadata vocabularies are proposed to address the requirements identified. These **three new metadata vocabularies 'buildm', 'ifcm' and 'e57m'** are the **main result** of the work documented in this report.
The work is partially based on and coordinated with the use cases and requirements identified in the work package 2 and further specified in work package 7. The concrete use of the metadata and its role in the whole context of the DURAARK system and the Semantic Digital Archive (SDA) and the Semantic Digital Observatory (SDO) are further detailed in the Deliverable 3.3.2 which describes the operational and ontological framework of the archival components. A brief overview illustration of the relevant parts in the DURAARK system with regard to the metadata proposed here is provided in Figure 1.

1.1 Document structure and reading guide

The document is structured as follows:

- In the remainder of the this section a brief overview of the original content data and metadata is provided.
- In **Section 2** a detailed review of existing standards and related work is provided. To accomplish this, for the three main areas **building information, preservation**and **Linked Data** individual sub-sections are dedicated to 1) the information relevant for preservation, 2) existing standards and their 3) shortcomings and gaps that are targeted by the DURAARK project in general and with regard to metadata specifically in this report.
- In **Section 3** metadata schemas addressing the gaps identified in **Section 2** are proposed .
- The usage as well as examples of the application of these proposed metadata formats are demonstrated in **Section 4**.

The document is concluded by a brief summary, recommendations concerning the proposed metadata and an outlook to related and subsequent work in the DURAARK. The appendix provides the listings of the vocabularies as well as the examples used as proofs of concept throughout the text.

1.2 Original content data - overview

The **original content** representations of buildings considered in the scope of the DURAARK systems include three main data models along with their respective serialization file formats. These are:

IFC STEP/ISO 10303 part 21 SPF as the main carrier of explicit building information. The Industry Foundation Classes (IFC) model is an standardized, extensive, vendor-neutral and widely-accepted data model for the exchange of building information. Concrete instantiations of this data model, e.g. a model of a particular building, are usually stored in the STEP Physical File (SPF) format. This is a commonly used clear-text encoded data format used in many other engineering domains and governed by the ISO 10303 family of standards, commonly referred to as STandard for the Exchange of Product data (STEP). Extended introductions of the model and the file format can be found in the Deliverable Reports D6.6.1 and D7.1.1 as well as in section 2.2.2 of this document. Depending on the use case, these representations might or might not be enriched with additional geometric representations (e.g. point clouds) and semantic meaning (e.g. using external vocabularies such as the buildingSMART Data Dictionary (bSDD)[1] or Linked Datasets provided in the Resource Description Format (RDF) found elsewhere "in the cloud" or in a network. Semantically equivalent versions of these file can also be captured as RDF(S)/OWL ("ifcOWL")[Beetz et al., 2009] documents. Generating and preserving this representation alongside the part 21 file is highly redundant, but has a number of useful applications in different scenarios. Part 21 SPF and ifcOWL can be converted bi-directionally without loss of information.

E57 point cloud data scans. These are either unstructured raw data dumps from measurements or data which has been grouped and clustered through post-processing, e.g., a number of scans grouped and spatially arranged to represent one building(part) from different scanning perspectives; a single scan segmented into different areas that corresponds to different building parts (door, window, "room Alice" etc.). In many use case scenarios, scans of the same objects and buildings would be taken at different moments in time. Such multiple states of the same object would be stored in separate archival packages. The necessity to keep past versions solely depends on the archival policies employed by the curator.

[1]see also section 2.2.2

Linked data sets used for the semantic enrichment of building models. These datasets can help to semantically enrich building models either at an intrinsic level (for instance, a designer references a particular material, product or color not found in the IFC schema), or at an extrinsic level (for instance, a curator captures social aspects by harvesting Twitter-feeds on a particular building or geo-references a building linking it to a DBPedia entity). Since many of these external references are used and referred to by multiple archival packages, and are often too big to be stored redundantly to guarantee archival autonomy and independence, these datasets will be managed in an individual system, the "Semantic Digital Archive" (SDA).

In real-world scenarios it is desirable to preserve further building-related information captured in a number of additional file types, such as text documents, images, native file formats of engineering applications or even full version histories of documents throughout a building project etc. However, numerous research projects have addressed this already and they are hence considered out of scope for the DURAARK project.

1.3 Metadata - overview of some key schemas

Metadata schemas are found in a number of areas of the DURAARK ontological framework and system architecture. These metadata schemas comprise those, that are stored in the Semantic Digital Archive (SDA) itself as well as the ones being used in the core architecture of the system. These include a number of key schemas that is provided in the following non-exclusive list. Their role and function is also illustrated in figure 1:

buildm, ifcm and e57m contained in METS|PREMIS : There are different metadata formats which describe the content of the individual Open Archival Information System (OAIS) information packages at their different OAIS process stages (SIP, AIP, DIP). Buildm, ifcm and e57m are domain specific metadata vocabularies which will be introduced in this deliverable. They will be embedded into the metadata wrapper format METS[2] (Metadata Encoding and Transmission Standard) which will furthermore include descriptive and structural metadata as well as preservation metadata in the PREMIS[3]. (PREservation Metadata: Implementation Strategies) schema. METS and PREMIS are described in more detail in chapter2.1.

[2]urlhttp://www.loc.gov/standards/mets/
[3]urlhttp://www.loc.gov/standards/premis/

ifcPC.rdf : format that links E57 and IFC representations of buildings (see section 1.2); the preliminary version of this metadata model is currently documented in D3.3.2 and will be refined and extended in later phases of the DURAARK project.

VoL : vocabulary of Links based metadata that captures versioned, relational information between individual entity classes of the IFC model and external vocabularies beyond the mere links between IFC schema entities and their meanings.

VoID+ : that is extended to meet the needs of evolving vocabularies which are monitored by the SDO.

2 Metadata: related work

In this section existing metadata efforts and approaches are described and discussed for their potential reuse, integration and extension into the DURAARK framework. Each section on the three main areas –Long Term Preservation, Building Information Model and Linked Data –is subdivided into three subsections: 1) A generic description of the kinds of data that need to be stored, 2) concrete models and vocabularies that allow to bind such information in vocabularies and data models and 3)a concluding section in which the open issues and unmet requirements of the related work are discussed per topic. In order to identify desirable metadata for buildings and building models representing the real-world artefacts, user requirement studies are being carried out in the context of the DURAARK project. These are documented in the respective deliverables and use cases of work packages 2 and 7.
In section 2.1 an introduction to the notion of metadata in the context of Long Term Preservation is introduced and discussed. In particular the concepts of "technical metadata" and "descriptive metadata" are introduced which determine the requirements and implementation of the vocabularies proposed in later sections of this report.

2.1 Metadata for Archival Systems

2.1.1 Types of data

The standard work of reference describing archival processes is the "Reference Model for an Open Archival Information System (OAIS)". The model describes an OAIS as an archive which has accepted the responsibility to preserve data for a specified target group called the designated community. The reference model defines these terms in the following way:

- An **archive** is not only software or hardware but a combination of an organization, people and systems.
- To **preserve** means to store and to maintain the accessibility to information.
- The **designated community** is a group of people identified by the archive as potential consumers.[CCSDS, 2012]

In order to fulfil the task of preservation, the archive needs to have a full understanding of the object it wants to preserve and the designated community it wants to preserve the object for. This includes knowledge about technical and contextual criteria - only if we understand how the object can be rendered technically and interpreted semantically can we guarantee accessibility and understandability over time. An “information object” in the sense of the OAIS reference model is therefore not only a singular file like a 3D scan, but the file accompanied by different types of information or metadata.

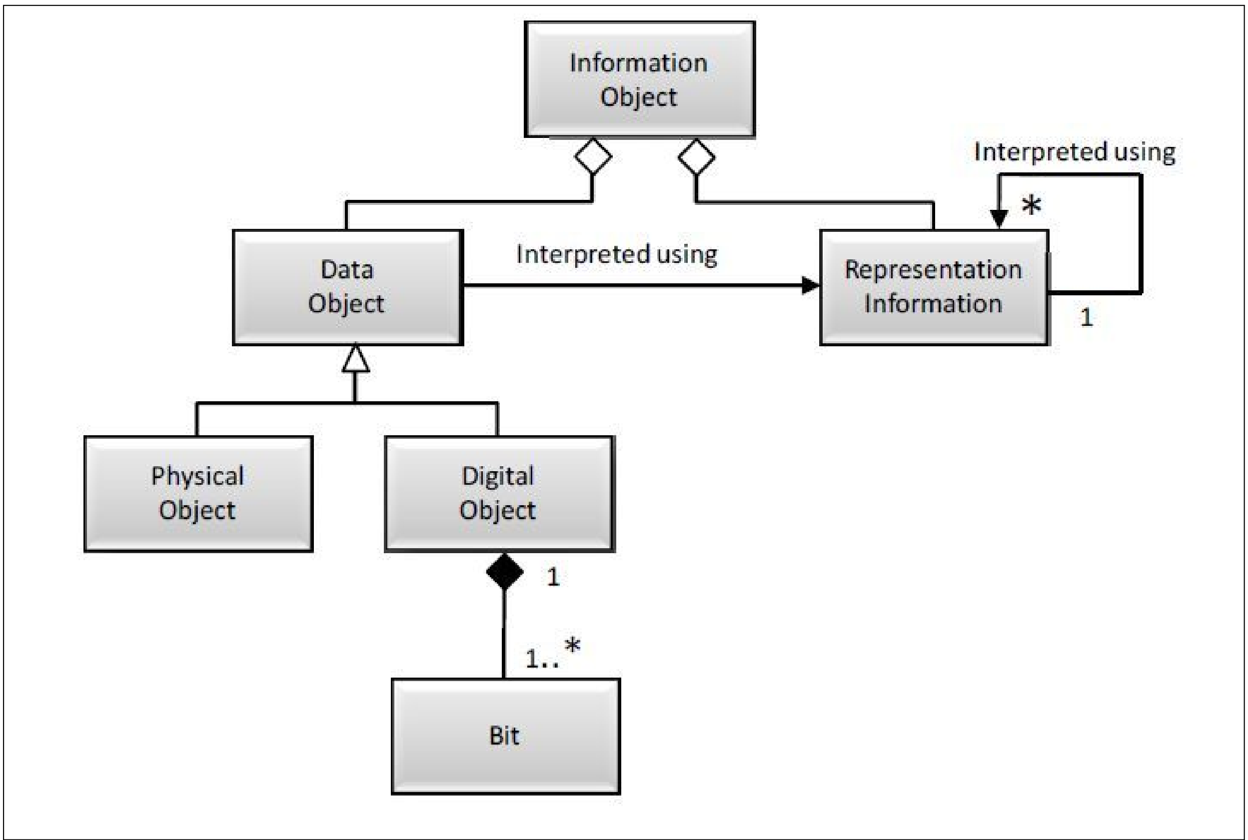


Figure 2: **Information Object in the context of the OAIS reference model [CCSDS, 2012]**

The first level of relevant information the reference model defines is that of representation information. Representation information includes structural as well as semantic information. Structural information explains how the bits at the basic layer of the objects are organized into a render-able object and how information within the file may be structured. To give a concrete example: in the case of E57 the archive should store the format standard as well, in the case of IFC-SPFF a schema description should be stored in the archive as well.

A second form of information included in representation information is semantic information. The semantic information is necessary knowledge which cannot be assumed for the designated community and should therefore accompany the digital object. [CCSDS, 2012] If our digital object contains information in for example Czech, the archive can only successfully guarantee understandability if the designated community speaks this language. In the architectural domain measurement parameters without an accompanying information on the system (metric or imperial) and unit size (e.g., cm) used may lead to different assumptions depending on the designated community.
The OAIS model describes further information types, such as preservation description information and packaging information. However, in the archival community a differentiation between "descriptive metadata", "structural metadata" and "administrative metadata" has established itself. "Administrative metadata" may be further divided into "rights management metadata" and "preservation metadata". [NISO, 2004] "Technical metadata", in return, is a relevant sub-part of preservation metadata. It describes file intrinsic metadata needed to maintain access to the file on a technical level, i.e. information needed in order to render the file.
The following section gives examples for each section of relevant metadata.

At a descriptive metadata level: Information about the author (person and/or organization), the creation date and a basic set of contextual information about the object described by the BIM / captured in the point cloud (i.e. location, status [planned, built]) should accompany the information as a basic set of semantic representation information. Depending on the intended designated community this set needs to be extended. Is the designated community for example a municipal building department, this information may need to be extended to include information like the property owner.

At a structural metadata level: In case the information object consists of multiple files, the hierarchical structure of files and/or the linking between different files and their associated metadata is described. Multiple files within one information object may consist in form of multiple representations for specific reasons - for example the archival file and a preview file or access file in a different file format. Furthermore, behaviour description in the case of executable code accompanying the file is described at the structural metadata level.

Administrative metadata: Information about the data producer, i.e. the person and

-if applicable- matching organization initiating the deposit to the archive should be captured (name, address, contact information). If the information object is part of a larger collection to be deposited this should be stated as well, including the name or other way to reference the collection.

On the rights management level of administrative metadata information about the intellectual property owner, the license of the object -if applicable- and access restrictions -if applicable- should be captured.

Preservation metadata includes technical metadata like the file format information (name, version), technical metadata / intrinsic file information such as byte-sequence, encoding or resolution as well as the information about the tools used to identify the file format and extract the technical metadata. This highly content specific information is relevant for maintaining the render-ability of the digital object over time.

A last part of preservation metadata are significant properties. They define characteristics which the archive considers important factors of an object's or environment's quality, structure or behaviour and which should be preserved over time, i.e. over the course of digital preservation actions such as migration to a new target format. Significant properties are often a subset of technical metadata but may additionally also draw on other information sources such as organizational parameters. The concept of significant properties is further discussed in D6.6.1.

Metadata type	Content description	Content examples	Metadata schemas
Wrapper Metadata and Structure Information	Structure and order of files in compound objects	File name, strucutral information like chapter	METS, EAD
Descriptive Metadata	Information needed for identification and discovery	Author, title, subject, URI	DC, MODS, MARC, EAD
Rights Management Metadata	Information needed for legal administration of object	Copyright, access restrictions	ccREL, ODRL
General Preservation Metadata	Content and format agnostic information needed to archive the object	Software and hardware used to create object, actions performed on object	PREMIS, LMER
Technical Metadata (Content Level)	Content and format specific information needed to archive the object	Image width, image height, Bit depth, encoding	MIX, TextMD. XMP, EXIF

Figure 3: Metadata types and examples of standards in archival practise

2.1.2 Data binding

The following section examines two de-facto metadata standards in archival practise: the wrapper schema METS and the preservation metadata schema PREMIS. While other preservation metadata frameworks, such as the National Library of New Zealand's Preservation Metadata [Searle and Thompson, 2003] or the German National Library's LMER[4] (LangzeitarchivierungsMetadaten für Elektronische Resourcen) schemas, were developed, PREMIS has established itself as the de-facto standard for preservation metadata. The OAIS reference model specifically recommends PREMIS as a standard for the submission of digital metadata about the object to an archive [CCSDS, 2012]. In a digital library and digital preservation context, PREMIS is frequently used embedded within a METS wrapper[5]. As opposed to other wrapper formats, such as the AV specific MPEG-7/MPEG-21[6] and MXF[7] (Material eXchange Format), METS is extendible and largely content and workflow agnostic.

2.1.2.1 PREMIS (PREservation Metadata: Implementation Strategies) is an international standard for preservation metadata and is maintained by the PREMIS Editorial Committee[8]. The standard consists of a data dictionary which is regularly being revised based on community input and is currently available in version 2.2. An XML schema as well as a draft OWL ontology of the data dictionary version 2.2 are available via the PREMIS website. The OAIS points out PREMIS as a suitable digital metadata standard [CCSDS, 2012]. Within the realm of digital metadata submitted to an OAIS compliant archive, PREMIS defines preservation metadata as a combination of information on a context descriptive, technical, business logic and rights level.

The five main entities of the PREMIS data dictionary and their relation to each other are shown in 5. The data dictionary defines semantic units as properties of those five entities. The PREMIS object entity, for example, can capture information such as the objects fixity information, significant properties, extracted technical metadata, and file format information to name a few. To support further granularity, PREMIS itself allows extensibility for the following semantic units: `significantProperties`, `objectCharacteristics`, `creatingApplication`, `environment`, `signatureInformation`, `eventOutcomeDetail`,

[4] http://nbn-resolving.de/urn:nbn:de:1111-2005051906
[5] seehttp://www.loc.gov/standards/premis/premis-mets.html
[6] http://mpeg.chiariglione.org/standards/mpeg-21/mpeg-21.htm
[7] http://tech.ebu.ch/docs/techreview/trev_2010-Q3_MXF-2.pdf
[8] http://www.loc.gov/standards/premis/

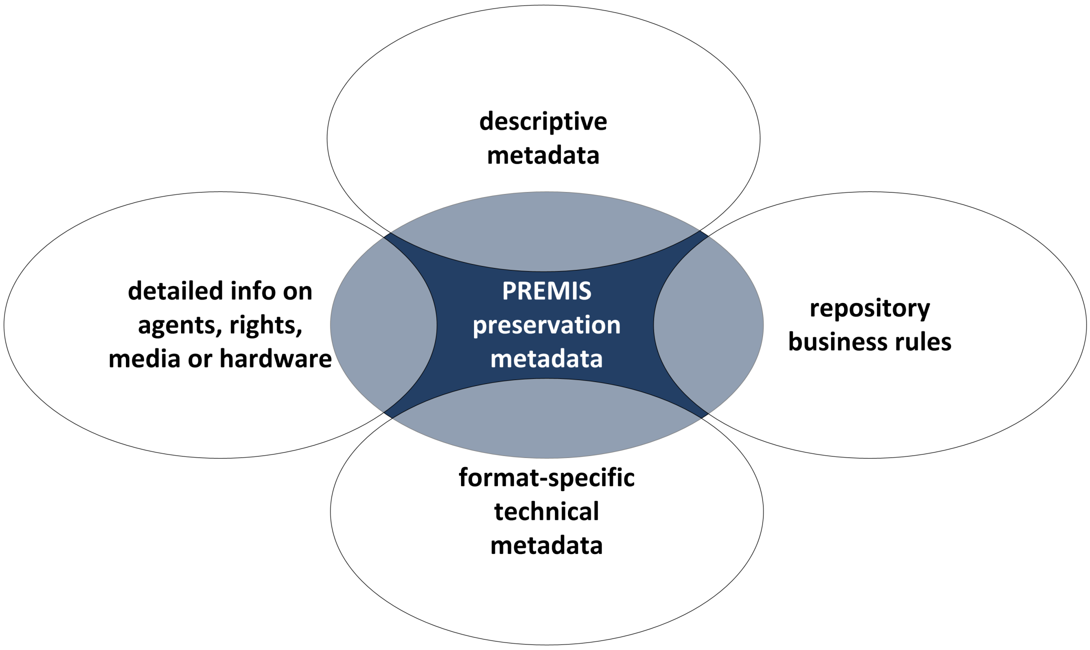


Figure 4: PREMIS as a subset of different information [Caplan, 2009]

`rights` and `agent` [PREMIS Editorial Committee, 2012]. The extendible semantic unit `objectCharacterisitcs` captures technical metadata about a file, which is very specific to the content type. The PREMIS data dictionary states that "Technical metadata describes the physical rather than intellectual characteristics of digital objects. [...] Further development of technical metadata is left to format experts. An extensibility mechanism is provided by including the semantic unit `objectCharacteristicsExtension`, which may be used with an external technical metadata scheme." [PREMIS Editorial Committee, 2012] Examples for content-specific technical metadata standards are MIX - the NISO metadata standard for still images[9] or MPEG-7 [10]. A full PREMIS example including MIX technical metadata has been supplied by the PREMIS Editorial Committee and is available at: `http://www.loc.gov/standards/premis/louis-2-1.xml`

2.1.2.2 METS (Metadata Encoding & Transmission Standard) is a wrapper format, tying together descriptive, administrative and structural metadata. The standard is expressed in an XML schema and maintained by the Library of Congress hosted Network

[9] `http://www.loc.gov/standards/mix/`
[10] url{http://mpeg.chiariglione.org/standards/mpeg-7

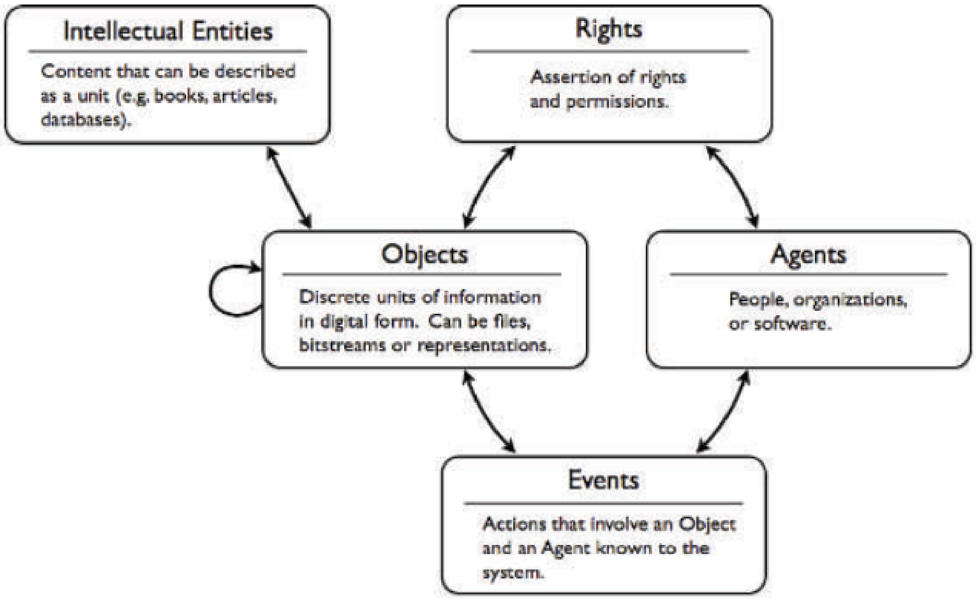


Figure 5: the PREMIS data motel [Caplan, 2009]

Development and MARC Standards Office. [11]. A METS file can include internal or external descriptive metadata (`<dmdSec>`), internal or external administrative metadata (`<amdSec>`), internal or external behavior metadata (`<behaviorSec>`) and a listing of files (`<fileSec>`). A structural map detailing the hierarchy of the objects (`<structMap>`) is the only required section for all METS files. The `<structLink>` section allows hyper-links between different `structMap` nodes and is typically used for webarchiving. An extensive list of METS example documents can be found on the METS website[12]. Furthermore, the METS Editorial Board maintains a METS Profile Registry, where digital library or archiving systems using the standard can register their profile[13]. Since METS allows for the integration of external metadata standards and sources, the PREMIS in METS combination is a common one. The PREMIS entities are mapped to different METS metadata sections. A simple view of the mapping is shown in figure 6. An example for a METS file containing PREMIS can be found on the PREMIS website[14]. Furthermore, the Florida Center for Library Automation has created an open source tool which validates

[11] http://www.loc.gov/standards/mets/
[12] http://www.loc.gov/standards/mets/mets-examples.html
[13] http://www.loc.gov/standards/mets/mets-profiles.html
[14] http://www.loc.gov/standards/premis/louis-2-1-mets.xml

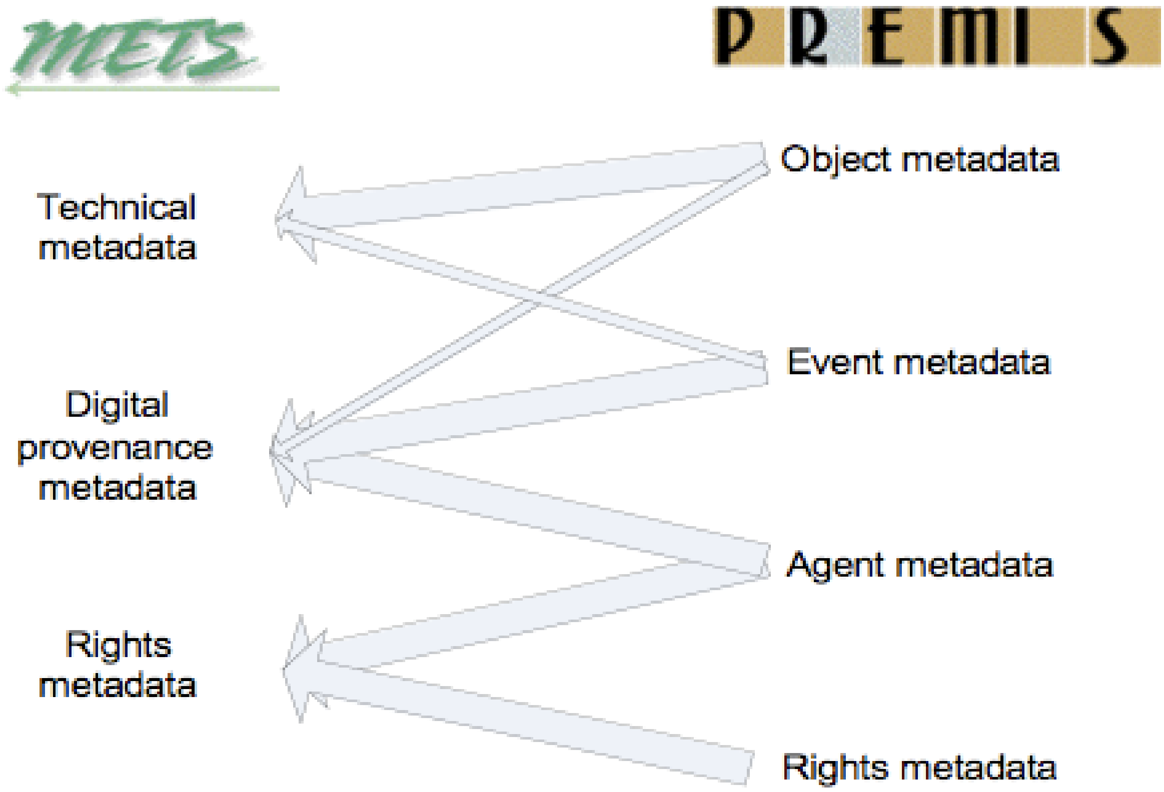


Figure 6: PREMIS entities mapped to METS sections [Guenther, 2008]

PREMIS in METS and converts PREMIS to PREMIS in METS[15].

2.1.3 Unmet Requirements of the DURAARK system

When preserving 3D architectural objects, the relevant information as defined in the Reference Model for an Open Archival Information System needs to be captured and stored alongside the digital objects. Figure 3 shows the different metadata types needed for this - relevant examples have been briefly discussed in this chapter.

The content agnostic wrapper format METS and the preservation metadata de-facto standard PREMIS have been chosen as suitable candidates within the DURAARK project. As no reference implementation for 3D objects is currently available in the METS[16] and PREMIS[17] implementation registries, the suitability will have to be proven within the

[15] http://pim.fcla.edu/
[16] http://www.loc.gov/standards/mets/mets-registry.html
[17] http://www.loc.gov/standards/premis/registry/premis-fulllist.php

DURAARK project. The biggest challenge in this regard is a current lack of a de-facto standard for technical metadata of 3D data. In this context, the suitability of a common group of elements for both scanned (e57) and planned (IFC) data should be evaluated. A necessary prerequisite for this is an identification of potential technical criteria based object and format capabilities and requirements alike.
Similarly to the identification of technical metadata criteria, descriptive metadata criteria for the specific content type architectural 3D data need to be identified and described in a schema.

2.2 Metadata for Building Models

2.2.1 Types of data

At present, there are a number of data models for Building Information Models capturing data about built environment artefacts. In general most of these standards contain both *data* and *metadata* (data about data) at the same time, where the line separating both categories is often blurred. In the archival use cases of the DURAARK project this distinction is defined as follows:

content data is about the artefact itself. This includes for instance, parameters of objects (“the width of the window is 1.01 m"), geometric representations of the objects (e.g. an extruded rectangular profile representing the glass pane of the window) and other information that is used as a description of how to for example, construct, produce or maintain buildings and their components.

metadata is data about *content data* that is not necessary and essential. For example, it is essential to describe that a window is 1.01 m wide, has a wooden frame and a burglary resistant class RC 3 according to EN 1627:2011 *while stating the name of the architect who designed it is not essential.*

For the purposes of Long Term Preservation, some of the required content data of building models can be extracted and transformed (e.g. through aggregation) into both descriptive and technical metadata. A typical example would be information about stakeholders such as the architect of a building [18], the owner and tenants. In other cases, information that is only *implicit* in the content data is made *explicit* in the metadata. As an example, a

[18] One can argue, that any information about the architect could be considered metadata *per se*

building model may provide area measures of individual rooms that reflect raw geometrical properties. For archival purposes however, it is desirable to store the aggregate gross floor area for the whole building. However, the inference and derivation of such metadata from content data often involves complex calculations. As an example, the “gross floor area" of a building storey that has slanted roofs is calculated by the projected areas with a head clearance of higher than 1.5 m in many countries such as Germany.

2.2.2 Data binding

2.2.2.1 IFC In the context of the DURAARK project, the IFC [16739:2013, 2013] model has been identified as the most relevant one based on its semantic richness, openness and acceptance in industry.
For a historic overview of the STEP data modelling series of standards in general and the developments of the IFC model in particular, also see D 7.7.1, where a historic introduction is given in chapter 2. An overview of the technical specifications regarding the clear-text encoding serialization format SPF relevant to preservation aspects is provided in D6.6.1. The model is specified using two general mechanisms:

- The model schema, defined in the EXPRESS modelling language. This schema currently contains more than 650 class definitions (called `ENTITY` in EXPRESS). For most `ENTITY`s several attributes are defined on a schematic level. Conceptually, all definitions on the schematic level are grouped into sub-schemas that are organized into four main architectural layers (Resource, Kernel, Shared and Domain layers) with increasing level of specialization. At the lowest level - the resource layer - a number of re-usable basic data structures have been defined. In some cases, such as the data structures capturing geometric and topological information, these are re-used and adapted from other ISO 10303 data standards.

- In addition to schema-level attribute definitions, object instances can be further specified using *properties*. These properties can be assigned to instantiated objects using a specialized objectified relationship type `IfcRelDefinesByProperties`. The properties themselves can be defined ad hoc by software tool, thus providing a flexible extension mechanism. Some property definitions however are standardized by the buildingSMART organization. They are defined external to the schema and are specialized to individual objects, e.g. “PSet_DoorCommon", “Pset_DoorWindowGlazingType" etc.

The most basic types of metadata about objects described in IFC models are realized on the level of the `IfcRoot ENTITY` definition. In Listing 1 these two definitions are provided for convenience.

```
ENTITY IfcRoot
ABSTRACT SUPERTYPE OF (ONEOF(IfcPropertyDefinition, IfcRelationship,
    IfcObjectDefinition));
    GlobalId     :   IfcGloballyUniqueId;
    OwnerHistory:   IfcOwnerHistory;
    Name         :   OPTIONAL IfcLabel;
    Description :   OPTIONAL IfcText;
UNIQUE
    UR1          :   GlobalId;
END_ENTITY;

ENTITY IfcOwnerHistory;
    OwningUser           :   IfcPersonAndOrganization;
    OwningApplication    :   IfcApplication;
    State                :   OPTIONAL IfcStateEnum;
    ChangeAction         :   IfcChangeActionEnum;
    LastModifiedDate     :   OPTIONAL IfcTimeStamp;
    LastModifyingUser    :   OPTIONAL IfcPersonAndOrganization;
    LastModifyingApplication: OPTIONAL IfcApplication;
    CreationDate         :   IfcTimeStamp;
END_ENTITY;
```

Listing 1: Excerpts of the IFC EXPRESS schema definition of the `IfcRoot` and `IfcOwnerHistory` entities.

Many of the descriptive and technical metadata items can be derived directly or indirectly from definitions on either on the level of the schema or the standardized property set definitions. However a number of issues and insufficiencies are described in later sections.

2.2.2.2 IFD and bsDD Since even the IFC model with its vast number of entity and standardized property definitions does not cover all of the information capturing and exchange needs that occur in building a construction, meta models have been developed and implemented to allow the controlled growth of semantic expressiveness in information models such as the IFC .

Currently, the most widely accepted one in the building industry is the ISO 12006 series of standards [ISO12006-3, 2006], [12006-2:2001, 2001] which has led to a number of suggested frameworks and reference implementations [Grant et al., 2008], [Bell et al., 2004], [Woestenenk, 2002] of controlled vocabularies. The conceptual approaches of these vocabularies are based on a long research and development history which goes back to the 1950s

with the Swedish SfB classification model. Other classification systems, like the Omniclass and the Uniclass systems, the Dutch SfB-NL, the German DIN 276 and many other similar standards have led to a rich landscape of extensive vocabularies with often considerable overlap in their facets. In order to allow the simultaneous storage of classifications and other vocabularies the development of the ISO 12006-3, referred to as the "International Framework for Dictionaries (IFD) was triggered. The aim of this framework has been formulated to enable *"the specification of a taxonomy model, which provides the ability to define concepts by means of properties, to group concepts, and to define relationships between concepts"*[ISO12006-3, 2006].
Where the IFD provides the metadata schema vocabulary, the reference collection of instances of such interrelated concepts is referred to as the buildingSMART Data Dictionary (bsDD). In its current state of implementation (at the end of 2013) the dataset consists of some more than 100.000 thousand concepts. It represents the merged data from several initiatives, including the Norwegian BARBi, the Dutch Lexicon, the Omniclass and Uniclass tables as well as all standardized IFC property sets[19] including their multilingual term associations, hierarchies (specialization, partonomy, decomposition, grouping, etc.) and associated properties including SI and other units. Figure 7 shows a subset of the bsDD dataset.

2.2.2.3 COBie Construction Operations Building information exchange (COBie)[20] is a standardized method for the documentation of various aspects of buildings throughout all of its lifecycle stages with a focus on operation and maintenance. The development of COBie has been initiated by the US Army corps of engineers where it is used for the Facility Management of their vast assets. COBie information is composed of tabular data (that can be captured e.g. in Excel® sheets) that allows easy specification of mandatory information exchange units (referred to as 'drops'). The COBie guide e.g. lists specific requirements for naming spaces, decomposing systems and provides extensive tables of properties for individual components such as doors, HVAC units or power outlets. Due to its intuitive tabular form, low technical threshold and liberal licensing it has quickly gained acceptance as an alternative interoperability format to more complex approaches like IFC and MVD. However the integration with the latter technologies has been a design

[19] recognizeable by their naming scheme, where the `IfcProppertySet` name starts with the string 'PSET'

[20] http://www.wbdg.org/resources/cobie.php

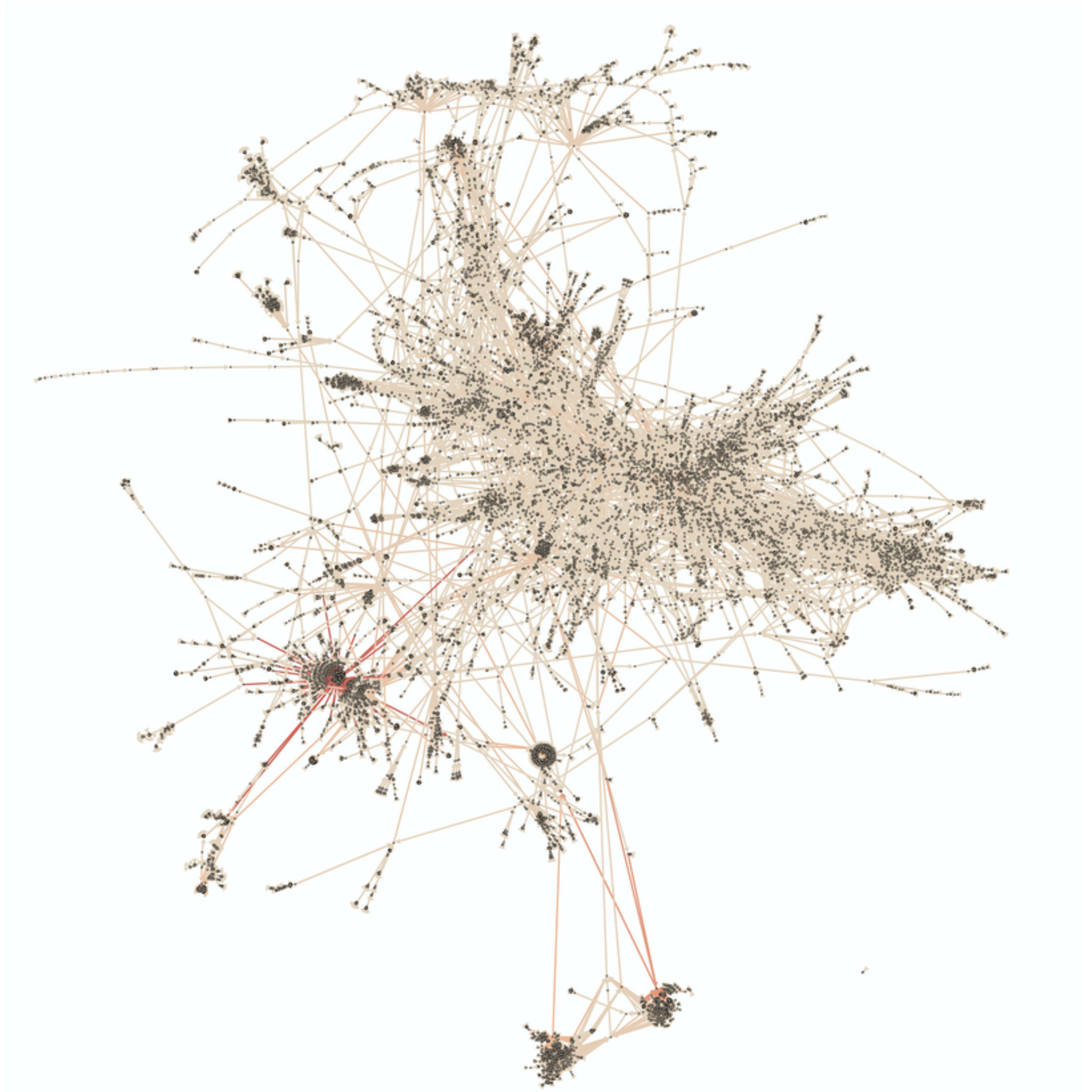


Figure 7: Visualization of a subset of the bsDD graph showing its specialization ('is-a') relationships

goal of COBie from early on and COBie is an endorsed standard of the buidlingSMART organization.

2.2.2.4 gbXML The Green Building XML (gbXML)[21] is another special-purpose BIM interoperability format. It addresses the specific needs of energy performance analysis tools that often need other and much simpler representations of building geometries, topologies and the (material) properties of individual components than can quickly be

[21]urlhttp://www.gbxml.org/

extracted from IFC files. Due to its compact XML form it has received wide support by software vendors.

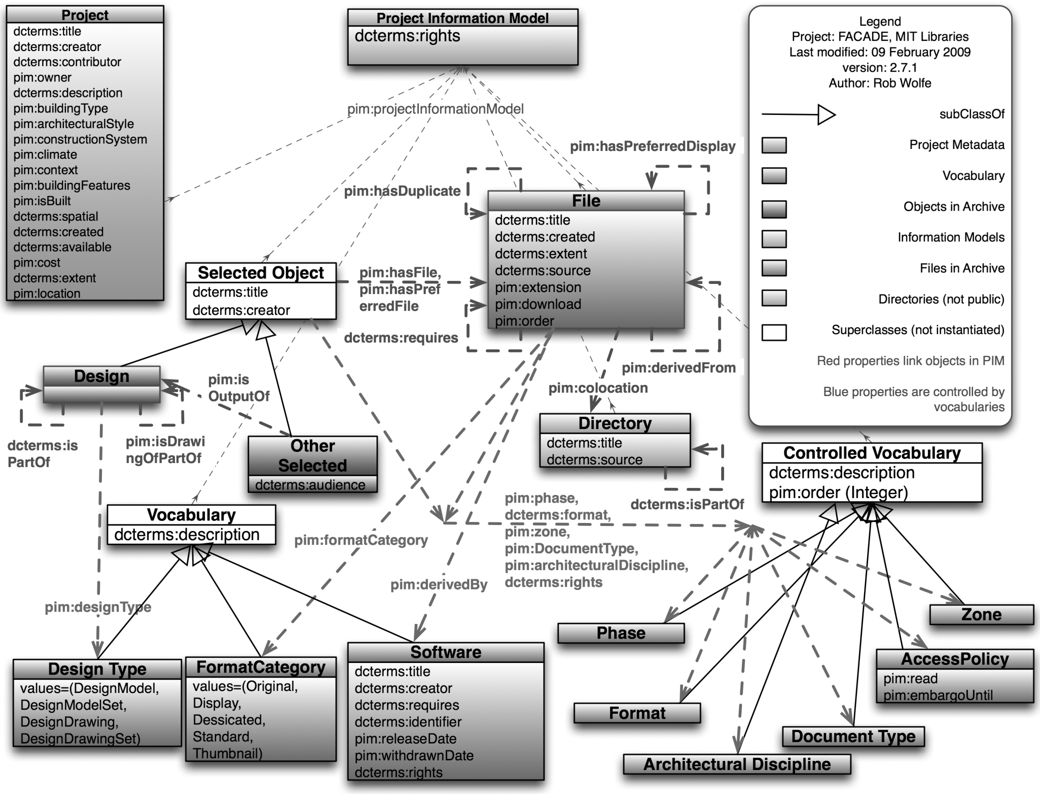
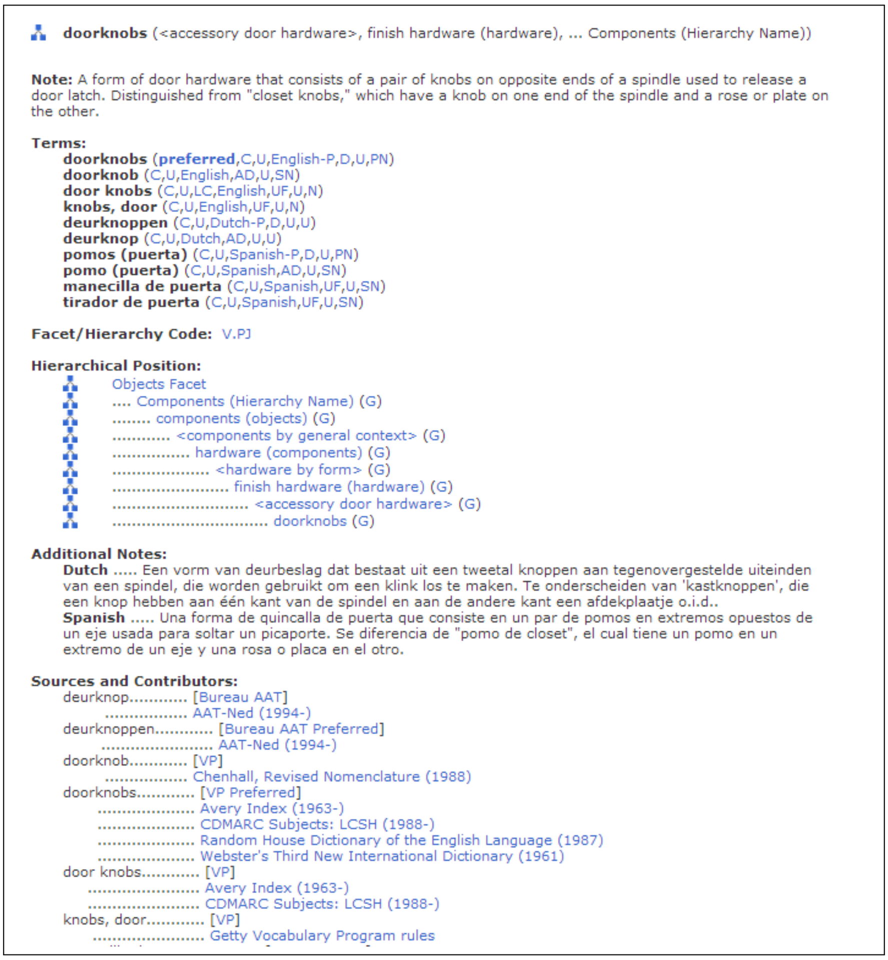


Figure 8: Overview of the PIM vocabulary. source: project website (see text)

2.2.2.5 PIM In the MIT FACADE project, the “Project Information Model" PIM has been suggested to store accumulated descriptive metadata about buildings and the files in archives that are representing them. The suggested vocabulary has been published as OWL RDF. It re-uses, integrates and extends the established Dublin Core metadata schema for generic information (`creator`, `contributor`, `isPpartOf`) and adds a number of additional OWL classes and properties where they are required specifically for architecture and archival (e.g. `pim:isBuilt`, `pim:ArchitecturalDiscipline` etc.). Even though this vocabulary is a valuable and inspiring contribution for the domain-specific metadata for archival of building information, a small number of issues render it inappropriate for the unmodified use in DURAARK: a) it does not cover all descriptive metadata identified in the user requirements (see D.7.7.1, e.g. gross floor area) b) some information is redundant with more sophisticated, e.t. technical and structural metadata in dedicated vocabularies such as METS/PREMIS and c) many `xsd:string` ranges of properties leave too much room for arbitrary wording, where controlled vocabularies would enable better classification and searching. A schematic overview of the vocabulary can be found in fig. 8[22].



doorknobs (<accessory door hardware>, finish hardware (hardware), ... Components (Hierarchy Name))

Note: A form of door hardware that consists of a pair of knobs on opposite ends of a spindle used to release a door latch. Distinguished from "closet knobs," which have a knob on one end of the spindle and a rose or plate on the other.

Terms:
doorknobs (**preferred**,C,U,English-P,D,U,PN)
doorknob (C,U,English,AD,U,SN)
door knobs (C,U,LC,English,UF,U,N)
knobs, door (C,U,English,UF,U,N)
deurknoppen (C,U,Dutch-P,D,U,U)
deurknop (C,U,Dutch,AD,U,U)
pomos (puerta) (C,U,Spanish-P,D,U,PN)
pomo (puerta) (C,U,Spanish,AD,U,SN)
manecilla de puerta (C,U,Spanish,UF,U,SN)
tirador de puerta (C,U,Spanish,UF,U,SN)

Facet/Hierarchy Code: V.PJ

Hierarchical Position:
Objects Facet
.... Components (Hierarchy Name) (G)
.... components (objects) (G)
.... <components by general context> (G)
.... hardware (components) (G)
.... <hardware by form> (G)
.... finish hardware (hardware) (G)
.... <accessory door hardware> (G)
.... doorknobs (G)

Additional Notes:
Dutch ..... Een vorm van deurbeslag dat bestaat uit een tweetal knoppen aan tegenovergestelde uiteinden van een spindel, die worden gebruikt om een klink los te maken. Te onderscheiden van 'kastknoppen', die een knop hebben aan één kant van de spindel en aan de andere kant een afdekplaatje o.i.d..
Spanish ..... Una forma de quincalla de puerta que consiste en un par de pomos en extremos opuestos de un eje usada para soltar un picaporte. Se diferencia de "pomo de closet", el cual tiene un pomo en un extremo de un eje y una rosa o placa en el otro.

Sources and Contributors:
deurknop............ [Bureau AAT]
.... AAT-Ned (1994-)
deurknoppen............ [Bureau AAT Preferred]
.... AAT-Ned (1994-)
doorknob............ [VP]
.... Chenhall, Revised Nomenclature (1988)
doorknobs............ [VP Preferred]
.... Avery Index (1963-)
.... CDMARC Subjects: LCSH (1988-)
.... Random House Dictionary of the English Language (1987)
.... Webster's Third New International Dictionary (1961)
door knobs............ [VP]
.... Avery Index (1963-)
.... CDMARC Subjects: LCSH (1988-)
knobs, door............ [VP]
.... Getty Vocabulary Program rules

Figure 9: Example listing the entries for the AAT vocabulary on the 'doorknob' concept

2.2.2.6 AAT The Art & Architecture Thesaurus (AAT) developed and maintained by the Getty Research Institute and other contributors is currently the largest structured vocabulary for artefacts used in the context of architecture. Its more than 250.000 concepts, terms and descriptions are part of a larger set of vocabularies (including CONA, TGN and

[22] http://facade.mit.edu/topics/03_Project_Information_Model.html

ULAN) which allow a multi-faceted classification and attribution of works of art. Figure 9 on page 17 shows an example of the vocabulary listing information about doorknobs. The recognized values, the wide use in library context, and professional curation make this vocabulary an ideal candidate to enrich archival information in the DURAARK system with concepts from this vocabulary. In addition, the imminent publication of the vocabulary as a LOD dataset fits well within the general technological approaches suggested in the project. It has to be noted however, that even though this vocabulary is extensive in terms of the number of different objects and has deep hierarchical specialization and classification structures, it lacks some significant semantic expressiveness. In particular, even though there are conceptual facets that allow the classification of a building as being 'low-rise' or 'high-rise' the vocabulary does not have predicates that would allow for statements, e.g. expressing the fact that "the building is 157 m high".

2.2.3 Unmet requirements of the DURAARK system

Even though the metadata facilities built into the IFC model itself provides a number of useful information items that could be used for archival purposes (e.g. provenance information regarding applications, users, modification dates etc.), there are a number of practical issues related to the immediate real-world usability of this data regarding the extraction of metadata. These issues can be found in many contemporary implementations in software tools:

- even though the provision of an `IfcOwnerHistory` instance per item is non-optional a number of attributes of the metadata record itself are[23]. For example, it is interesting in many maintenance scenarios, to query when components have been changed in a building model. Having such time stamps per component instead of global times on a file level would make it convenient to create complete timelines of building lifecycles where every change to the building itself is reflected in the model. Even though the IFC model provides a dedicated `LastModifiedDate`, this is attribute is optional. In most implementations this is value is never captured.
- In practice, many tools are re-using a single `IfcOwnerHistory` instance throughout a model. While this saves space by re-using constant information that does not change per context (e.g. the application creating or modifying an object) it renders volatile data such as time stamps useless.

[23] see the `OPTIONAL` keywords in the provided listing

- even though a great number of semantically meaningful object definitions are inherited from `IfcRoot`, the majority of object instances - mainly geometry definitions taken over from ISO 10303 part 42[24] - are not covered by this metadata.

This lax implementation of the standard is not only a missed chance to capture valuable metadata directly in the content carrier itself, it also makes versioning and collaboration more difficult[Beetz et al., 2010].
As a counter measure to overcome this arbitrariness and ensure the required quality of information a Model View Definition (MVD) is going to be developed in later stages of the DURAARK project. The Model View Definition is a standardized mechanism to address two main of the above mentioned issues of the IFC model and its implementation:

- enable the creation of constraints on model instantiations. Such constraint include simple cardinality restrictions (an attribute must have some value) and qualified cardinality restrictions (an attribute must have a value in the range between “1" to “10" or must be either “red" or “blue").

- limit the scope of data structures defined in the schema. Supporting the more than 650 `ENTITY` schema definitions in software implementations is often unnecessary and demands considerable efforts. In order to limit this implementation effort, an application can instead implement compliance to a certain MVD that was created for a specific exchange scenario (e.g. quantity take-off, energy simulation etc.). Every `ENTITY` and attribute that is *not* mentioned in the MVD does not have to be supported by the software tool.

Such MVD definitions can be specified in a standardized schema (mvdXML). Such a schema is being created to enable the validation of model instances with regard to required information defined in the metadata vocabularies introduced in this documents (see section 3.2.2 for technical metadata ‘ifcm’ and section 3.2.1 for the descriptive metadata ‘buildm’).

2.3 Metadata for Contextual Knowledge and Linked Data

The Web of data is a rich source of contextual knowledge pertaining to numerous domains. Due to the vast contextual coverage of the Web, this section is restricted to the closer

[24]this also is the main reason while future versions of the model are very unlikely to include such (verbose?) metadata for geometries in order not to break the little compatibility with other STEP models there is left

context of DURAARK. This section provides descriptions of the related types of data that need to be stored, as well as datasets and vocabularies that allow to bind such data.

2.3.1 Types of data

Architectural/infrastructural data may be contextually enriched by considering publicly available web data as a potential, rich knowledge base. Within the DURAARK context, we have identified different kinds of knowledge that can be captured with respect to an individual built structure (depicted in the Figure 10), as well as corresponding datasets that can help us extract and encapsulate relevant information.
A few potential categories are presented here, bearing in mind the existing Linked Data and vocabularies.

- Location (location of the building, weather/climatic conditions, accessibility, etc.)
- Function (purpose of the building)
- Usage (who uses the building, how often, mobility patterns, etc.)
- Context (built environment, surrounding structures, transport, environment, crime, architectural style, etc.)
- History (construction period, evolution of the building over time, culture & heritage, structural changes over time, etc.)
- Social context (sentiments, movements, events, etc.)
- Sustainability (energy consumption, impact on external environment, indoor environment quality, cost, efficiency of retrofit, fire safety measures, etc.)
- Aesthetics/Materials (building materials, design, etc.)
- Structural Information (blueprints, 2D/3D drawings/models, IFC, Point Clouds, BIM Models, etc.)

A number of such vocabularies identified as potentially relevant for the DURAARK project are further introduced in section 2.3.2.1.

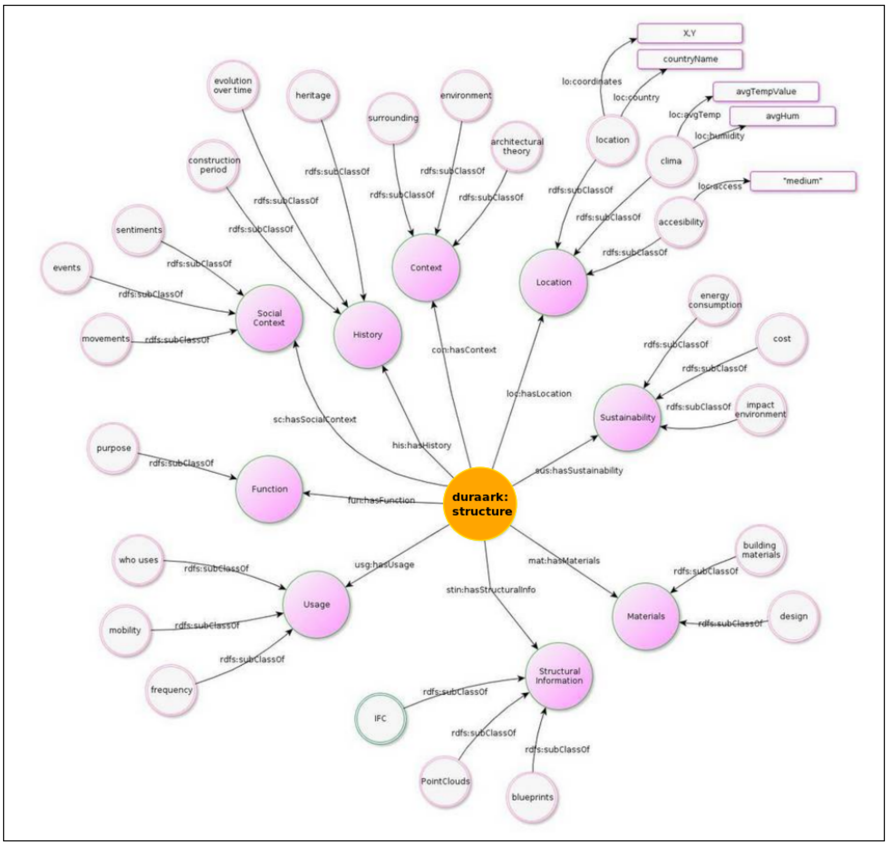


Figure 10: Potential knowledge types for semantic enrichment.

2.3.2 Data binding

The extent of stability or the level of maturity of currently available metadata schemas varies greatly. Many such metadata schemas are a result of ongoing projects, or the outcome of individual initiatives, and are essentially prototypes rather than mature standards. Having observed that, several established institutions are working towards committing resources to Linked Data projects. For instance, this includes the national libraries of Germany, France, Sweden, Hungary, the Library of Congress and the British Library, the Food and Agriculture Organization of the United Nations, apart from the Online Computer Library Center (OCLC). Some relevant state-of-the-art metadata standards for Linked Data are presented below. Before providing an overview of these existing metadata schemas however, we are providing some data bindings of potentially interesting Linked Datasets identified in section 2.3.1.

2.3.2.1 Examples of Linked Data In this section we identify a number of possible datasets/vocabularies relevant to some potential categories of the knowledge types that have been identified in 2.3.1 These include the following :

- The Dublin Core[25] vocabulary terms are typically used to describe physical resources. These include text files, books, digital materials (like videos, images and audio clips), apart from media like web pages. Metadata records based on Dublin Core are intended to be used for cross-domain information resource description and have become standard in the fields computer science and libraries. The Simple Dublin Core Metadata Element Set (DCMES)[26] consists of 15 metadata elements: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage and Rights. The implementations of Dublin Core typically make use of RDF/XML.

- Geographical data from GeoNames[27] provides us with location coordinates and names, elevation but also population and weather information. For instance, the following RDF listing 2 obtained from the feature document of Cologne, describes the location and population of the German city.

```
<rdf:RDF>
    <gn:Feature rdf:about="http://sws.geonames.org/2886242/">
        <rdfs:isDefinedBy
            rdf:resource="http://sws.geonames.org/2886242/about.rdf"/>
        <gn:name>Cologne</gn:name>
        <gn:countryCode>DE</gn:countryCode>
        <gn:population>963395</gn:population>
        <wgs84_pos:lat>50.93333</wgs84_pos:lat>
        <wgs84_pos:long>6.95</wgs84_pos:long>
    </gn:Feature>
</rdf:RDF>
```

Listing 2: Excerpt from the RDF feature document of Cologne describing the location and population.

- GeoLife GPS Trajectories [28] from Microsoft Research contains the information of latitude, longitude and altitude of users' outdoor activities.

[25] http://dublincore.org/
[26] http://dublincore.org/documents/dces/
[27] http://www.geonames.org/
[28] http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/

- Data from the Social Web can be harnessed for analysis in order to gauge sentiments, evolutionary changes in perceptions, apart from other applications like triggering retrofit scenarios. For example, tweets from Twitter, Flickr metadata, Facebook places and status updates, Foursquare check-ins, News articles and blog posts, etc. can be exploited.
- Eurostats Linked Data [29] contains statistical data regarding population and social conditions, agriculture, transport environment, and energy apart from other statistics. The bulk download facility [30] provides means to download individual datasets.
- Wikipedia and/or DBPedia can be exploited to extract historical facts. The Europeana LOD [31] contains cultural information, historical photos and sounds. The YAGO LOD [32] contains more than 120 million facts about entities such us places, people and cities.


```
<a:RDF>
    <ore:Aggregation
        a:about="http://data.europeana.eu/aggregation/provider/03486/2BE8A1953AA9F7BE5D5E17950640E
        <edm:aggregatedCHO
            a:resource="http://data.europeana.eu/item/03486/2BE8A1953AA9F7BE5D5E17950640ECFCA645D
        <edm:dataProvider>Bayerische Staatsbibliothek</edm:dataProvider>
    </ore:Aggregation>
    <edm:EuropeanaAggregation
        a:about="http://data.europeana.eu/aggregation/europeana/03486/2BE8A1953AA9F7BE5D5E1795064
        <dc:creator>Europeana</dc:creator>
    </edm:EuropeanaAggregation>
    <ore:Proxy
        a:about="http://data.europeana.eu/proxy/provider/03486/2BE8A1953AA9F7BE5D5E17950640ECFCA6
        <dc:title>The history of the Boroughs and municipal Corporations of
            the United Kingdom / T. 2 (1835)</dc:title>
        <dc:creator>Merewether, Henry Alworth</dc:creator>
        <dc:creator>Stephens, Archibald John</dc:creator>
        <ore:proxyFor
            a:resource="http://data.europeana.eu/item/03486/2BE8A1953AA9F7BE5D5E17950640ECFCA645D
        <dc:relation>Signatur: Brit. 386 m-2</dc:relation>
        <edm:type>TEXT</edm:type>
        <dc:type>Druck</dc:type>
        <dc:date>1835</dc:date>
        <dc:publisher>London u.a.</dc:publisher>
```


[29] http://eurostat.linked-statistics.org/
[30] http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
[31] http://pro.europeana.eu/linked-open-data
[32] http://www.mpi-inf.mpg.de/yago-naga/yago/

```
            <ore:proxyIn
                a:resource="http://data.europeana.eu/aggregation/provider/03486/2BE8A1953AA9F7BE5D5E1
            <dc:identifier>bvb-id :
                BV004616483</dc:identifier><dc:identifier>oclc :
                165640651</dc:identifier>
            <dc:identifier>urn : urn:nbn:de:bvb:12-bsb10280955-4</dc:identifier>
            <dct:created>1835</dct:created>
        </ore:Proxy>
    </a:RDF>
```

Listing 3: Excerpt from the RDF description of the history of the Boroughs and municipal Corporations of the United Kingdom.

- Public data about energy efficiency guidelines can be exploited (for example, Linked Data from GBPN's Energy Efficiency Guidelines[33]), including dedicated vocabularies and ontologies. GBPN provides access to Linked Data about energy efficiency in European buildings[34] apart from energy rating policies[35]. LOD Clean Energy Datasets can be exploited similarly[36]. Statistics from the World's Energy Information and Data can be used to complement this information [37].
- The European Building and Construction Materials Database for the Semantic Web, BauDataWeb[38] exposes a major dataset of the European building and construction materials market for the Semantic Web on the basis of the GoodRelations Web Vocabulary for E-Commerce. In addition, IFD-Library/bsDD[39], is the official buidlingSMART framework. It provides an implementation of buidling concepts spanning various classifications from multiple contexts.
- Geo[40] is a basic RDF vocabulary that provides the Semantic Web community with a namespace for representing latitude, longitude and other information about spatially-located objects, using the World Geodetic System (WGS84) as reference data.

2.3.2.2 Schemas for bibliographic assertions

[33] http://www.gbpn.org/
[34] http://www.gbpn.org/databases-tools/data-hub-energy-performance-buildings
[35] http://www.gbpn.org/databases-tools/building-energy-rating-policies
[36] http://data.reegle.info/
[37] http://en.openei.org/datasets/
[38] http://semantic.eurobau.com/
[39] http://www.ifd-library.org/
[40] http://www.w3.org/2003/01/geo/

BIBO [D'Arcus and Giasson, 2009], the Bibliographic Ontology[41] is largely based on the Dublin Core model. It extends Dublin Core for the description of bibliographic things like books or magazines. This specification provides concepts and properties which in turn facilitate the description of bibliographic references and citations.

FRBR [O'Neill, 2002, FBF,], Functional Requirements for Bibliographic Records[42] is a conceptual entity-relationship model. It was developed by the International Federation of Library Associations and Institutions (IFLA). The key aspect that this model introduces is that it relates user tasks of retrieval and access, in online library catalogues and bibliographic databases from a user's perspective. Thus it represents a more holistic approach to retrieval and access as the relationships between the entities provide links to navigate through the hierarchy of relationships. FRBR standardises a set of terms and relationships that are essential to any cataloguer.

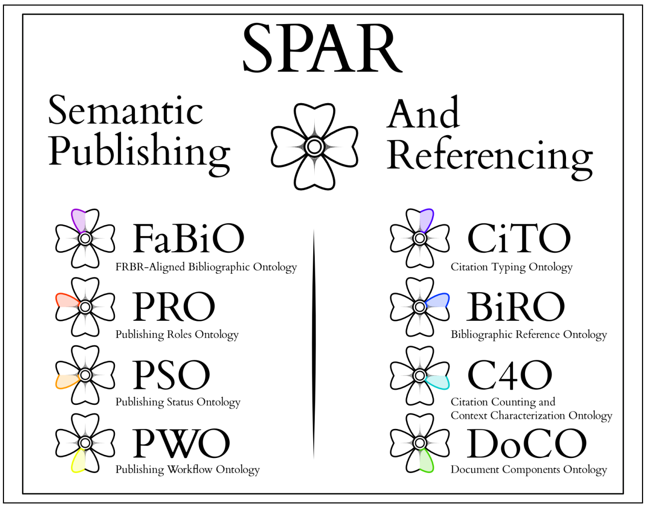


Figure 11: **Component ontologies within SPAR**

SPAR (the Semantic Publishing and Referencing Ontologies)[43] is a collection of orthogonal and complementary ontology modules that can be used in order to create comprehensive machine-readable RDF metadata for numerous aspects pertaining to semantic publishing and referencing. Figure 11 presents the component ontologies of SPAR. The ontologies can be used either in isolation or in conjunction, as per the requirement. Each is encoded in the Web ontology language OWL 2.0 DL. Together,

[41] http://bibliontology.com/
[42] http://vocab.org/frbr/core.html
[43] http://sempublishing.sourceforge.net/

they provide the ability to describe far more than simply bibliographic entities such as books and journal articles, by enabling RDF metadata to be created to relate these entities to reference citations, to bibliographic records, to the component parts of documents, and to various aspects of the scholarly publication process.

RDA [Coyle and Hillmann, 2007](Resource Description and Access)[44] is a recent set of library cataloguing rules, and is supported by an element set that is defined in RDF. RDA is an implementation of the FRBR model. It has about 1400 properties and over 60 term lists, that cover several media including text, sound, film, cartographic materials, and objects, as well as archival materials.

2.3.2.3 Schemas for dataset metadata

OMV [Hartmann et al., 2005] (Ontology Metadata Vocabulary)[45] specifies reusability-enhancing ontology features for human and machine processing purposes. OMV organize the metadata elements according to the type and purpose of the contained information as follows; *General*, *Availability*, *Applicability*, *Format*, *Provenance*, *Relationship*, and *Statistics*. OMV organize the metadata elements according to the impact on the prospected re-usability of the described ontological content as follows; *Required*, *Optional*, and *Extensional*.

VoID [Alexander and Hausenblas, 2009] (The Vocabulary of Interlinked Datasets)[46] is an RDF Schema vocabulary which provides a means to express the interlinking between RDF datasets in a powerful and flexible manner. VoID covers different areas of metadata including general metadata that follows the Dublin Core model described earlier. It also describes how RDF data can be accessed by exploiting various protocols. In addition, the structure and schema of datasets is also described in terms of structural metadata. The metadata expressed, thus serves as a bridge between the publishers and users of RDF data, opening a floodgate of possible applications ranging from data discovery to cataloging and archiving of datasets.

VoL (Vocabulary of Links)[47] is an RDF Schema vocabulary for expressing metadata about links between resources and entities within and across RDF datasets. While there

[44] http://rdvocab.info/
[45] http://omv2.sourceforge.net/index.html
[46] http://vocab.deri.ie/void
[47] http://data.linkededucation.org/vol/

do exist limited ways to express links, for instance with RDFS/OWL expressivity such as owl:sameAs or rdfs:seeAlso or SKOS, up to now, there is no vocabulary at hand to further describe the semantics of a link. This is of particular importance for links obtained with automated entity linking techniques, where links are usually associated with some form of probability (such as a connectivity score) instead of simple links where the decision of a connection is binary. In addition, often some of ranking has to be expressed to, for instance, rank the relevance of a particular link for a given entity or dataset (e.g. the association of a category with a particular RDF resource or dataset). VoL provides a vocabulary to describe a link with respect to obtained scores and utilised methods for obtaining the links. It has been specifically designed to be used together with VoID dataset descriptions, as part of VoID linksets. However, it is intended for general-purpose description of links between resources within and across distinct RDF datasets.

2.3.2.4 Schemas for geographical data

GeoNames Ontology [48] makes it possible to add geo-spatial semantic information to the Word Wide Web. Over 8 million GeoNames toponyms now have a unique URL with a corresponding RDF web service. The listing 4 presents an example RDF description of some geographical features that can be ascertained using the GeoNames feature document of Berlin available through the RDF Webservice at the corresponding URI[49].

```
<rdf:RDF
xmlns:cc="http://creativecommons.org/ns#"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:gn="http://www.geonames.org/ontology#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:wgs84_pos="http://www.w3.org/2003/01/geo/wgs84_pos#">
<gn:Feature rdf:about="http://sws.geonames.org/2950159/">
<rdfs:isDefinedBy
    rdf:resource="http://sws.geonames.org/2950159/about.rdf"/>
<gn:name>Berlin</gn:name
<gn:alternateName xml:lang="en">Berlin</gn:alternateName>
```

[48] http://www.geonames.org/
[49] http://sws.geonames.org/2950159/about.rdf

```
<gn:alternateName xml:lang="eu">Berlin</gn:alternateName>
<gn:alternateName xml:lang="fr">Berlin</gn:alternateName>
<gn:featureClass rdf:resource="http://www.geonames.org/ontology#P"/>
<gn:featureCoderdf:resource="http://www.geonames.org/ontology#P.PPLC"/>
<gn:countryCode>DE</gn:countryCode>
<gn:population>3426354</gn:population>
<wgs84_pos:lat>52.52437</wgs84_pos:lat>
<wgs84_pos:long>13.41053</wgs84_pos:long>
<wgs84_pos:alt>74</wgs84_pos:alt>
<gn:parentFeature rdf:resource="http://sws.geonames.org/6547539/"/>
<gn:parentCountry rdf:resource="http://sws.geonames.org/2921044/"/>
<gn:parentADM1 rdf:resource="http://sws.geonames.org/2950157/"/>
<gn:parentADM3 rdf:resource="http://sws.geonames.org/6547383/"/>
<gn:parentADM4 rdf:resource="http://sws.geonames.org/6547539/"/>
<gn:nearbyFeatures
    rdf:resource="http://sws.geonames.org/2950159/nearby.rdf"/>
<gn:locationMap
    rdf:resource="http://www.geonames.org/2950159/berlin.html"/>
<gn:wikipediaArticle rdf:resource="http://en.wikipedia.org/wiki/Berlin"/>
<rdfs:seeAlso rdf:resource="http://dbpedia.org/resource/Berlin"/>
<gn:wikipediaArticle
    rdf:resource="http://uk.wikipedia.org/wiki/%D0%91%D0%B5%D1%80%D0%BB%D1%96%D0%BD"/>
</gn:Feature>
<foaf:Document rdf:about="http://sws.geonames.org/2950159/about.rdf">
<foaf:primaryTopic rdf:resource="http://sws.geonames.org/2950159/"/>
<cc:license rdf:resource="http://creativecommons.org/licenses/by/3.0/"/>
<cc:attributionURL rdf:resource="http://sws.geonames.org/2950159/"/>
<cc:attributionName
    rdf:datatype="http://www.w3.org/2001/XMLSchema#string">GeoNames</cc:attributionName>
<dcterms:created
    rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2006-01-15</dcterms:created>
<dcterms:modified
    rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2012-09-19</dcterms:modified>
</foaf:Document>
</rdf:RDF>
```

Listing 4: Excerpts from a sample RDF description of the GeoNames feature document of Berlin.

2.3.2.5 Schemas for concepts and relationships metadata

DBPedia Ontology[Bizer et al., 2009] [50] is a shallow, cross-domain ontology, which has been manually created based on the most commonly used info-boxes within Wikipedia. The ontology currently covers 529 classes which form a subsumption hierarchy and are described by 2,333 different properties.

POWDER [Archer et al., 2008] (the Protocol for Web Description Resources)[51] is the

[50] http://dbpedia.org/Ontology
[51] http://www.w3.org/TR/powder-primer/

W3C recommended method for describing web resources. It specifies a protocol for publishing metadata about Web resources using RDF, OWL, and HTTP. The primary 'unit of information' within POWDER is the Description Resource (DR), which comprises:

1. attribution (assertions about both the circumstances of its own creation and the entity that created it);
2. scope (to which resources does the description apply);
3. the description itself.

CIDOC/CRM (CIDOC Conceptual Reference Model)[52] is a formal ontology that provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modelling. It provides an integrated framework for different kind of resources like archives, images, places and objects.

PRISM (Publishing Requirements for Industry Standard Metadata) [53] is an industry-standard metadata that can be used to build efficient, multi-channel publishing solutions. With PRISM a publisher can create, manage, aggregate, produce, distribute and reuse content. PRISM specifications serve as the foundation for two XML publishing solutions:

1. PRISM Aggregator Message: An XML tag set developed in 2004, designed to package articles for delivery to content aggregators.
2. PRISM Source Vocabulary: An XML tag set developed in 2011, designed to encode a wide variety of content, ranging from articles, to advertisements, to the chapters of books.

MADS (Metadata Authority Description Schema)[54] is a standard for describing subjects, names, and other 'authorities'. There is an RDF vocabulary available for this purpose. The US Library of Congress currently uses it to export authority information.

[52] http://www.cidoc-crm.org/
[53] http://www.idealliance.org/specifications/prism-metadata-initiative
[54] http://www.loc.gov/standards/mads/

It is important to note that for some categories, we will need to deal with unstructured and heterogeneous data. So, an important issue that we aim to handle is generating and associating such unstructured data to our model. Knowledge acquisition from unstructured data involves particular challenges. This heterogeneous extracted data usually is ambiguous and provides limited information. Thus numerous techniques, including NLP, NER, WD and EL, should be applied aiming at (i). extracting entities (ii). event/relation extraction (iii). disambiguation and (iv). interlinking with other existing entities.

2.3.3 Unmet requirements of the DURAARK system

As observed in the previous sections, there is an abundance of contextual knowledge that can be captured using various Linked Data resources on the Web. However, the real value of such independent knowledge bases and vocabularies is only realized when used in unison in order to generate correlations with an enriched semantic value. Especially with respect to the DURAARK context, there is thereby a need to design and extend existing vocabularies and schemas, which when utilised solely, fall short in expressing useful high and low level semantic details, important from a preservation perspective.

Apart from a schema that is capable of such extended expressibility, there are several other challenges that surface with respect to preservation of Linked Datasets. These challenges have been studied in more detail and documented in Chapters 2 and 4 of the Deliverable D6.6.1 [55]. Some of the gaps that have thus been identified in the state-of-the-art are presented below.

- Strategies to track and maintain changes within chosen datasets have not been made available in tools exposed to the domain thus far.
- There are no standard methods to preserve and link changing semantic concepts in metadata available.
- Currently there are no standard methods which have been tested in a domain specific setting, to calculate the impact of changes pertaining to one entity within a chosen knowledge graph on other entities.

[55] Current state of 3D object digital preservation and gap-analysis report

3 Proposed metadata schemas

In this section, the metadata schemata to be used in the DURAARK project are described. Firstly, in the section 3.1 the general approach to the design and use of the proposed metadata schemas are provided. Categorized into different domains and purposes the individual schemas are briefly introduced and discussed for their applicability in the project context.
The definition of the proposed schemas will be provided in section 3.2

3.1 Design rationale

There are a number of general design approaches taken in the DURAARK metadata schema design which can be summarized in the following rules:

- re-use existing schemas where possible
- give precedence to widely accepted, frequently used schemata where alternatives exist.
- limit extensions to bare necessary amount
- limit extensions to domain specific aspects
- favour RDF formats over XML schemata even where RDF representations of respective XML models are not mature, complete or stable yet.

In addition to the overview of the schemas that has already been provided in section 1 and fig. 1, the data captured using the metadata schemas described here can fall in to either one of the three following categories:

Descriptive metadata extension For other artefacts such as textual and audio-visual media a wide range of available metadata schemas such as DC, MARC can readily be applied to capture information about the respective **Intellectual Entities**. However, no such vocabularies dedicated to the domain of architecture and buildings are available. One of the approaches to address this problem in the DURAARK project could have been to just reuse readily available ifcOWL and ifcRDF vocabularies that have been developed and published by a number of authors [Beetz et al., 2009]. However, as already discussed in section 2.2.2, potentially interesting information often is unreliable or 'hidden' in implicit structures and needs to be derived and inferred from the original data.

Technical metadata extension The domain specific schemas are reflecting the main file types of the DURAARK archival framework. In particular they cover the file formats IFC, E57 and RDF in different serialization formats. For these three formats `objectCharacteristicExtension`s to the PREMIS schema are specified that will cover the technical particularities of the respective formats. Based on the requirements that have been specified in section 3 of the deliverable data dictionaries are created that describe respective technical and non-technical metadata aspects. These dictionaries are modelled as OWL vocabularies which can be instantiated in an archival description and linked to a PREMIS `Object` (usually of the `File` variant)

Metadata for Linked Data used in the SDA and the SDO. The two vocabularies used are the VoL and VoID schemas. Unlike the descriptive and technical metadata schemas mentioned above, the VoL and VoID vocabularies are considered to meet all requirements of the DURAARK project and thus do not need to be extended. The recommended use with regard to multiple versions of evolving Linked Datasets stored in the SDA are further outlined and specified in the report D3.3.2.

VoL is used to capture mappings between classes / entity definitions of the IFC model schema and the vocabularies stored in the SDA. In the example provided in section 4.1 and illustrated in fig. 14 such a mapping has been created between a wall entity and the 'color' concept of the semantic concepts vocabulary. This can be used by engineers to semantically enrich a building model during the creation (e.g., by referencing the concepts 'from inside' the IFC model) or from a librarian to curate and classify the ingested model (e.g., by adding further descriptive metadata)

The **VOiD** vocabulary on the other hand is used in the main registry of the SDO, where it summarizes and classifies datasets and vocabularies that are stored in the SDA and whose evolution and changes are monitored and updated by the SDO. Figure 12 illustrates an example structure for the type of linkset description that would fit the DURAARK context.

```
:DURAARK_to_External_Datasets a void:Linkset;
```

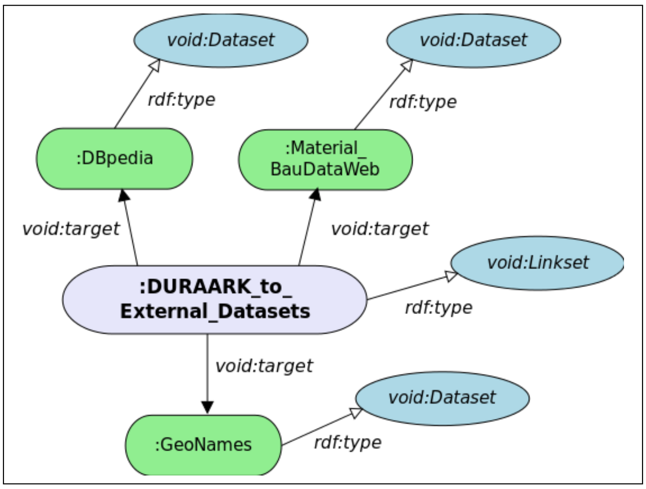


Figure 12: **Example structure for a VoID linkset description.**

```
void:target :DBpedia;
void:target :Material_BauDataWeb;
void:target :Geonames;
.
```

Listing 5: An example excerpt of a VoID linkset description in RDF/Turtle.

3.2 DURAARK Schema Extensions

In this section the implemented metadata extensions are introduced. At present these schema extensions consist of three individual vocabularies that can be integrated and used with the PREMIS OWL[56] representation published by the Library of Congress. These three parts are

buildm , the DURAARK **build**ing descriptive **m**etadata schema intended to be dropped into PREMIS records describing `premisowl:IntelectualEntity` items

ifcm , the DURAARK **IFC** technical **m**etadata schema capturing information describing IFC object representations.

e57m , the DURAARK **e57** technical **m**etadata schema describing E57 point clouds.

[56]Currently the RDF/OWL representation of the PREMIS vocabulary is strictly experimental and not supported by any commercial or R&D software tool yet. However, the RDF/OWL metadata vocabularies here can easily be converted into the required XML schema documents that are still required by contemporary tools. It is to be expected however, that the gradual adoption and uptake of RDF over XML in the preservation community will allow the direct use of the vocabularies here in the future

3.2.1 buildm

The descriptive metadata schema **buildm** captures descriptive information about buildings as it has been required by users. These user requirements have been evaluated through interviews and questionnaires. Their results have been captured in D.7.7.1.
Table 3.2.1 shows the tabular summary of the information items that can be captured using **buildm**. In appendix A.1 the complete listing of the meatdata vocabulary is provided which also contains additional information in the respective `rdfs:comment` section of each property.

Table 1: overview of the properties of the **descriptive metadata** to describe the intellectual entities that are documented by the two main forms of representation in (IFC, E57)

Metadata	Property Type	IFC Mapping[57]
Name	`xsd:string`	IfcProject.Name, Pset_BuildingCommon. BuildingID
Location	Location object[58]	IfcBuiling.Adress
Type of building/project	`xsd:string`	Pset_BuidlingUse, Pset_BuildingCommon
Year of construction	`xsd:date`	Pset_BuildingCommon. YearOfConstruction
Year of modification	`xsd:date`	Pset_BuildingCommon. YearOfLastRefurbishment
Owner	`foaf:person`	IfcPersonAndOrganization
Indoor building area	`qudt:AreaMeasure`	Pset_BuildingCommon. GrossAreaPlanned
Number of floors	`xsd:nonNegativeInteger`	Pset_BuildingCommon
Number of rooms	`xsd:nonNegativeInteger`	*Derived*
Model type	`xsd:string`	IfcActorRole, *derived*

3.2.2 ifcm

The metadata schema **ifcm** captures technical information that is crucial for LDP (long-term digital preservation) systems. It includes items such as the schema in use, MVDs

[57] using the dot-syntax name that delimits an ENTITY definition from a schema level attribute
[58] using one of the geographical datasets|

to which this model complies, the kind of geometry that is being used in the dataset representing a building[59]. It also comprises statistical metrics like the number of optional parameters that have been instantiated with values, the ratio of the number of entities to the geometrical density and other parameters. The table 3.2.2 provides an overview of the vocabulary to capture this information and descriptions of the respective means to derive this information automatically from ingested IFC files. The complete listing of the proposed vocabulary is provided in appendix A.2. This listing also contains additional information for each property that can be read in the respective `rdfs:comment` sections.

3.2.3 e57m

The metadata schema **e57m** captures technical information about the point cloud files that are stored in the open E57 file format. The metadata stored here consists of a number of technical parameters that help to determine the content and quality of the point clouds and the required software tools to render them in their original form. Table 3.2.3 shows an overview of the information that is captured along with their intended meaning. For a full listing of the schema refer to appendix A.3. This listing also includes additional description of the individual properties.

[59] which is crucial to the applications that should render the information: While some kinds of geometries can be implemented in lightweight visualization engines (2D line-based information, simple forms of explicitly pre-cut polygonal representations of e.g. walls that already have cut gaps into the geometry for openings (windows, doors etc.) etc., more complex geometries like NURBS, BRep and CSG operation might need other applications to be rendered correctly)

Table 2: overview of the properties of the **technical metadata** to describe **IFC files** and how the values can be extracted or derived from the models themselves

Metadata	Property Type	Method of derivation
Original file type	`xsd:string`	file format of the originating package that is documented in `IfcApplication` of the `IfcOwnerHistory`
File size	`xsd:nonNegativeInteger`	from file
IFC Version	`xsd:string`	from SPFF header
Applicable MVD	`xsd:string`	from SPFF header
Implementation Level	`xsd:string`	from SPFF header
Number of entities	`xsd:nonNegativeInteger`	the number of instantiated `ENTITY`s in the file.
Component count	`xsd:nonNegativeInteger`	number of `IfcProduct` instances including its subtypes
Detail level / Information level	*TBD*	metric indicating the ratio of the number objects per spatial quantity
Parameters in use	`xsd:nonNegativeInteger`	metric indicating how many schema-level attributes that are marked `OPTIONAL` and applicable Pset_[...] properites are filled in. Helps determine the level of detail
Components library	`xsd:string`, URI	useage of component libraries for products
Geometric representation	`xsd:string`	metrics on the types of geometry in use, e.g. “BRep", “CSG", “NURBS"

Table 3: overview of the properties of the **technical metadata** to describe **E57 files** and how the values can be extracted or derived from the models themselves

Metadata	Property Type	Method of derivation
# of scans	`xsd:nonNegativeInteger`	total number of scans constituting the scan project
# of points	`xsd:nonNegativeInteger`	total number of points
# of refernces	`xsd:nonNegativeInteger`	total number of references
intensity	`xsd:boolean`	whether or not intensity values are included in the scan
color	`xsd:boolean`	whether or not color values are included in the scan
Exterior Façade Amount	`xsd:double`	The amount of the building architecture within the scan project, which is covered by exterior scans in percentage
Room Distribution	`xsd:nonNegativeInteger`	The amount of singled out rooms with no connectivity to other rooms or neighbouring rooms or façades within an appropriate distance threshold set by architectural standards.
Element distribution (architectural)	`xsd:double`	Distribution of detected elements is an indicator in percentage, of how many of the points within a scan project are labelled as architectural elements and how many can be labelled as random elements.
Element distribution (non-architectural)	`xsd:double`	Distribution of detected elements is an indicator in percentage, of how many of the points within a scan project are labelled as architectural elements and how many can be labelled as random elements.
Point distribtution	*TBD*	see D.7.1.1
Noise	*TBD*	see D.7.1.1
Quality of Registration	*TBD*	see D.7.1.1
Quality chain	*TBD*	see D.7.1.1

4 Examples and use cases

In this section the intended use of the metadata is illustrated in two examples from the DURAARK use cases and scenarios. Their location within the context of the DURAARK system is illustrated in figure 13. The first examples in section 4.1 demonstrates how metadata for the semantic enrichment of IFC models using external Linked Datasets is being carried out using the VoL vocabulary. In section 4.2 an example is provided which illustrates the use of the VoID dataset in using a number of vocabularies to describe a building through social media datasets.

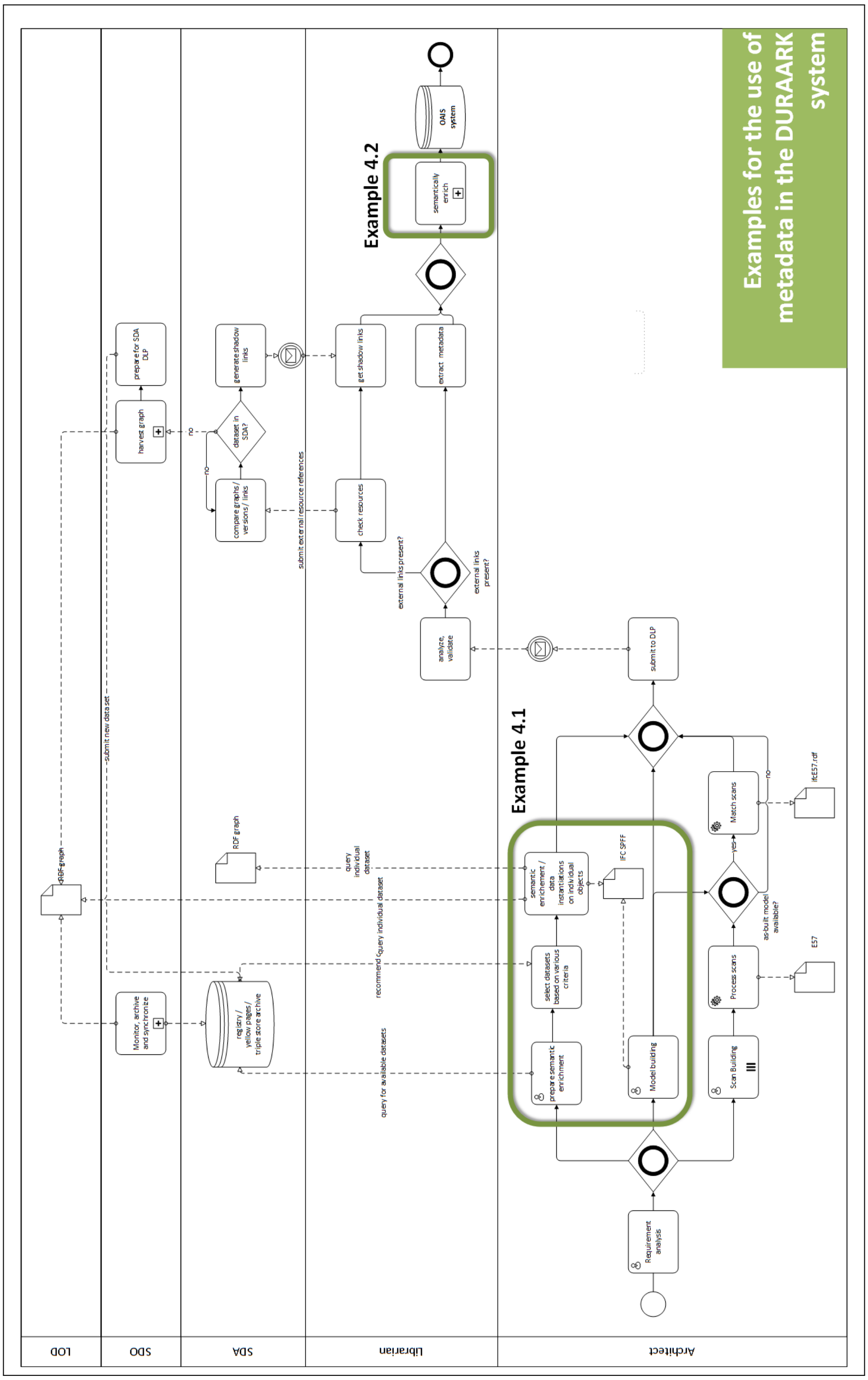


Figure 13: **Overview of the two examples provided in this section that illustrate the use of metadata in the overall DURAARK system)**

4.1 Linking IFC and external RDF datasets on a schematic level using VoL

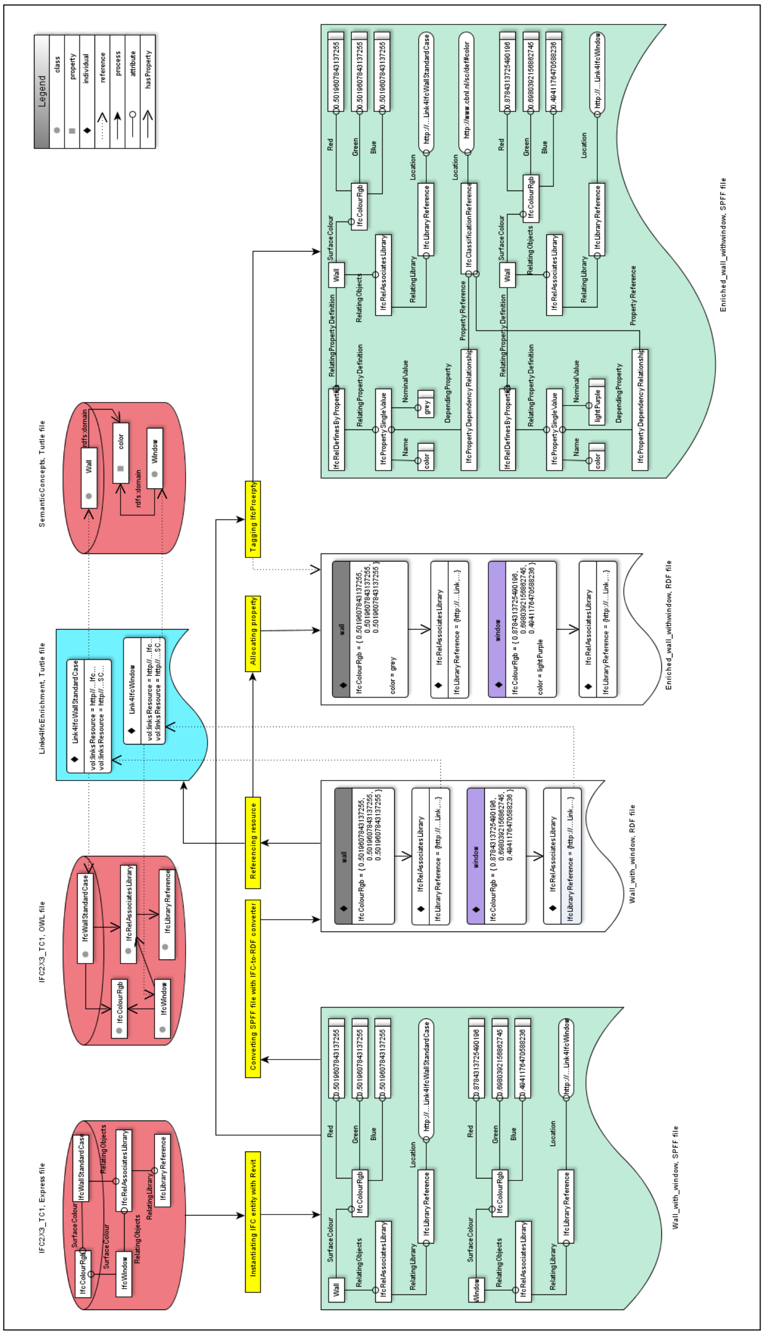


Figure 14: **Overview of using VoL to entities from the IFC schema to an external vocabulary (SemanticConcepts in this case)**

In this example a link is created between an IFC Entity definition and an owl:Class from an external dataset. The dataset in this example is an OWL version of the Dutch "Semantic Concepts" data base. The developer of the dataset, Kees Woestenenk, has been involved in the standardization effort for classification models in the building industry which led to the ISO 12006 series of standards. Initially, the dataset was developed under the name "LexiCON" for the Dutch STABU foundation[60]. STABU made a joint effort with the Norwegian BARBi initiative by the Norwegian Statsbygg organization to develop a common reference model for the building sector that could be tailored to individual needs. The result of these developments, which have been continued by a small community for over ten years eventually led to the reference IFD instance database bsDD.
In the example illustrated here, which is depicted in Fig 14, the `IfcDoor` and IfcWallStandardCase entities are linked to the respective nodes of the SemanticConcepts vocabulary. Snippets from the two respective models are shown in listings 7 and 8.

```
@prefix : <http://www.duraark.eu/Links4IfcEnrichment/def#>.
@prefix vol: <http://purl.org/vol/ns/>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix sd: <http://www.w3.org/ns/sparql-service-description#> .
@base <http://www.duraark.eu/Links4IfcEnrichment/def> .


:Link4WallStandardCase a vol:Link;
    rdfs:label "Link4WallStbdarCase";
    rdfs:comment "A link instance describing a link between IfcWallStandardCase
        and a Wall Concept in Semantic Concept set";
    vol:linksResource <http://www.duraark.eu/schema/IFC2X3#IfcWallStandardCase>;
    vol:linksResource <http://www.cbnl.nl/sc/def#SC4411>;
    vol:hasScore "1";
    vol:hasType "Semantic connectivity";
    .

:Link4Window a vol:Link;
    rdfs:label "Link4Window";
    rdfs:comment "A link instance describing a link between IfcWindow and a Window
        Concept in Semantic Concept set";
    vol:linksResource <http://www.duraark.eu/schema/IFC2X3#IfcWindow>;
    vol:linksResource <http://www.cbnl.nl/sc/def#SC4411>;
    vol:hasScore "1";
    vol:hasType "Semantic connectivity";
```

[60] http://stabu.nl

.

Listing 6: example of a a mapping between an IFC Entity definition and an external dataset

```
ENTITY IfcDoor;
    ENTITY IfcRoot;
        GlobalId         :   IfcGloballyUniqueId;
        OwnerHistory     :   IfcOwnerHistory;
        Name             :   OPTIONAL IfcLabel;
        Description      :   OPTIONAL IfcText;
    ENTITY IfcObjectDefinition;
    INVERSE
        HasAssignments   :   SET OF IfcRelAssigns FOR RelatedObjects;
        IsDecomposedBy   :   SET OF IfcRelDecomposes FOR RelatingObject;
        Decomposes       :   SET [0:1] OF IfcRelDecomposes FOR RelatedObjects;
        HasAssociations :   SET OF IfcRelAssociates FOR RelatedObjects;
    ENTITY IfcObject;
        ObjectType       :   OPTIONAL IfcLabel;
    INVERSE
        IsDefinedBy      :   SET OF IfcRelDefines FOR RelatedObjects;
    ENTITY IfcProduct;
        ObjectPlacement :   OPTIONAL IfcObjectPlacement;
        Representation   :   OPTIONAL IfcProductRepresentation;
    INVERSE
        ReferencedBy     :   SET OF IfcRelAssignsToProduct FOR RelatingProduct;
    ENTITY IfcElement;
            Tag          :   OPTIONAL IfcIdentifier;
    INVERSE
        FillsVoids       :   SET [0:1] OF IfcRelFillsElement FOR
            RelatedBuildingElement;
        ConnectedTo      :   SET OF IfcRelConnectsElements FOR RelatingElement;
        HasCoverings     :   SET OF IfcRelCoversBldgElements FOR
            RelatingBuildingElement;
        HasProjections   :   SET OF IfcRelProjectsElement FOR RelatingElement;
        ReferencedInStructures    :      SET OF IfcRelReferencedInSpatialStructure
            FOR RelatedElements;
        HasPorts         :   SET OF IfcRelConnectsPortToElement FOR RelatedElement;
        HasOpenings      :   SET OF IfcRelVoidsElement FOR RelatingBuildingElement;
        IsConnectionRealization      :      SET OF
            IfcRelConnectsWithRealizingElements FOR RealizingElements;
        ProvidesBoundaries   :   SET OF IfcRelSpaceBoundary FOR
            RelatedBuildingElement;
        ConnectedFrom        :   SET OF IfcRelConnectsElements FOR RelatedElement;
        ContainedInStructure     :   SET [0:1] OF IfcRelContainedInSpatialStructure
            FOR RelatedElements;
    ENTITY IfcBuildingElement;
    ENTITY IfcDoor;
        OverallHeight     :      OPTIONAL IfcPositiveLengthMeasure;
        OverallWidth      :      OPTIONAL IfcPositiveLengthMeasure;
END_ENTITY;
```

Listing 7: Excerpt from the IFC schema showing the definition of the IfcDoor class with its complete inheritance hierarchy

```
:SC4411 rdf:type owl:Class ;

        rdfs:label "Wall"@en ,
                   "Wand"@nl ,
                   "Wand (NEN 2767-4)"@nl ;

        rdfs:subClassOf :SC273405 ,
                        :SC332046 ,
                        [ rdf:type owl:Restriction ;
                          owl:onProperty :actsUpon ;
                          owl:allValuesFrom [ rdf:type owl:Class ;
                                              owl:unionOf ( :SC325652
                                                          )
                                            ]
                        ] ;

        dct:date "2013-04-25"^^xsd:dateTime ;

        cbnl:hasStatus "DRAFT"^^xsd:string ;

        dct:conformsTo "[BS Glossary]: Wall: Vertical construction, usually in
            masonry or in concrete, that bounds or subdivides a space an usually
            fulfils a loadbearing function."^^xsd:string ,
                       "[NEN 2767-4 1619]: Wand: "^^xsd:string ;

        dct:hasPart :SC136074 ,
                    :SC142018 .
```

Listing 8: Excerpt from the SemanticConcepts vocabulary (in RDF/Turtle) showing the definition of a door.

Equivalent to this example, concepts from the bsDD vocabulary can be used to semantically enrich IFC models. An OWL representation of the official reference database has been developed in the context of the DURAARK project. This dataset, containing some 80.000 concept nodes and a rich set of relations between these nodes will be used as the main test datasets for the SDA in later stages of the project.

4.2 Enriching a building dataset with social media

We investigated the problem of ranking architectural structures based on their sentiments extracted from social media[61]. We identify the need for type-specific tailoring of mecha-

[61] How does a Building make you feel?-Towards Tailored Mechanisms for Entity Ranking. Under review at the 23rd International World Wide Web Conference, Seoul, Korea. April 7-11, 2014.

nisms to arrive at accurate structure rankings. While showing that specific mechanisms are required in order to cater to the specifics of the type of architectural structure, we propose a crowd-sourced approach to identify *influential factors* for specific entity types. Building around a use case centered on architectural structures, we investigate the performance gains from designing type-specific methods. For instance, the public perception of buildings and architectural structures is influenced by a wide range of factors, such as functional or aesthetic aspects, dependent on the building type.

There have been research works that have tried to gauge comparison metrics for buildings. However, such works have been restricted to the functionality level of buildings, for example, sustainability, indoor-environment quality, energy consumption, construction waste management and so on. Although these factors are very important in accessing the performance of a building, they do not help in understanding the effect of the buildings on the built environment and the perception of their aesthetic appeal among people.

While detecting and measuring the impact of architectural changes, traditionally is an important issue for urban planners, architects as well as policy makers, it is a cumbersome and costly task. On the other hand, the Social Web yields invaluable information which can be exploited to address this problem. In our work, we develop a processing pipeline aimed at retrieving, filtering and analyzing social media towards creating precise sentiment scores, and eventually rankings of buildings. We explore the possibilities of exploiting data from the Social Web including tweets from Twitter, News articles and blog posts accumulated through crawling the search results from Google and metadata from Flickr images. We identify Flickr metadata to be well suited to this end. Extensive experimental results depict that methods tailored to specific building types allow an accurate measurement of public perception of architectural structures.

We can easily integrate such sentiment scores pertaining to various buildings alongside the data that is to be archived, on a temporal basis. By enriching the archived data with corresponding scores reflecting the perception of the structure, a stakeholder can easily gauge the evolution in perception of the particular architectural structure. This can be implemented at the schema level and incorporated for instance, as an RDF property as reflected in the listing below.

```
<rdf:RDF

    xml:base="http://purl.org/DURAARK_structure/"
```

```
         xmlns:foaf="http://xmlns.com/foaf/0.1/"
         xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
         xmlns:owl="http://www.w3.org/2002/07/owl#"
         xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
         xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:gn="http://www.geonames.org/ontology#"
         xmlns:pos="http://www.w3.org/2003/01/geo/wgs84_pos#">

    <!-- http://purl.org/DURAARK_structure/hasSentimentScore -->

        <rdf:Property rdf:about="hasSentimentScore">
            <rdfs:label xml:lang="en">Sentiment Score</rdfs:label>
            <rdfs:comment xml:lang="en"> This score depicts the perception of this
                architectural structure in the designated time period. </rdfs:comment>
        </rdf:Property>

    <!-- http://purl.org/DURAARK_structure/Oberbaum_Bridge -->

        <owl:NamedIndividual rdf:about="Oberbaum_Bridge">
            <rdf:type rdf:resource="ArchitecturalStructures/Bridge"/>
            <rdfs:label xml:lang="en">Oberbaum Bridge</rdfs:label>
        </isLocatedIn>
            <foaf:isPrimaryTopicOf>http://en.wikipedia.org/wiki/Oberbaum_Bridge
                    </foaf:isPrimaryTopicOf>
            <duraark:hasSentimentScore></hasSentimentScore>
            <pos:lat_long>52.501900,13.445600</pos:lat_long>
            <gn:countryCode>DE</gn:countryCode>
            <owl:sameAs
                xml:lang="en">http://dbpedia.org/resource/Oberbaum_Bridge</owl:sameAs>
        </owl:NamedIndividual>

    </rdf:RDF>
```

Listing 9: Excerpt from a possible RDF schema depicting the perception property corresponding to an archival object (in RDF/XML).

5 Summary and conclusions

In this report the existing state of the art of metadata vocabularies for the three main topics Preservation Systems, Building Information Models and Linked Data has been examined. We have shown that none of the current standards or proposals is able to fully meet the requirements for the DURAARK system that have been specified in the work packages 2 and 7. We have identified a number of data models and vocabularies that can be used to a) derive the metadata required for the DURAARK archival framework and to b) capture this metadata by extending and augmenting existing, commonly used metadata standards for archival systems. In particular, we have suggested three concrete RDF vocabularies that can be used as domain-specific extensions of the PREMIS metadata standard for preservation system information. These vocabularies are

- **buildm** as a **descriptive metadata** vocabulary to describe a building as an intellectual entity
- **ifcm** and **e57m** as **technical metadata** vocabularies to describe the two open file formats for the representations of buildings IFC and the geometric description of building using point clouds (E57)

Furthermore, we have demonstrated how the **VoL** and **VoID** vocabularies can be used to capture metadata about Linked Data sets that are referenced by Building Models. While in this report the focus has strictly been put on capturing metadata required for the DURAARK archive and the accompanying SDA and SDO, the parallel report D.3.3.2 is focused on the operation, processes and integration of the SDA and SDO. To get a full overview this document is necessary as is are the reports from WP 2 and 7 which deliver the greater context of the whole system. Since the whole project is still in its early stages and subject to constant evolution and change and depends heavily on the outcomes and results of other activities, the vocabularies suggested here are very likely to change and should be considered preliminary.

References

[FBF,] Functional requirements for bibliographic records @ONLINE.

[12006-2:2001, 2001] 12006-2:2001, I. (2001). ISO 12006-2:2001 building construction – organization of information about construction works – part 2: Framework for classification of information. Text.

[16739:2013, 2013] 16739:2013, I. (2013). ISO 16739:2013 industry foundation classes, release 2x, platform specification (IFC2x platform).

[Alexander and Hausenblas, 2009] Alexander, K. and Hausenblas, M. (2009). Describing linked datasets - on the design and usage of void, the 'vocabulary of interlinked datasets. In *In Linked Data on the Web Workshop (LDOW 09), in conjunction with 18th International World Wide Web Conference (WWW 09*.

[Archer et al., 2008] Archer, P., Smith, K., and Perego, A. (2008). Protocol for web description resources (powder): Description resources. *W3C Working Draft*, 14.

[Beetz et al., 2010] Beetz, J., van Berlo, L., de Laat, R., and van den Helm, P. (2010). Bimserver.org – an open source IFC model server. In *Proceedings of the CIB W78 2010: 27th International Conference*, Cairo, Egypt.

[Beetz et al., 2009] Beetz, J., Van Leeuwen, J., and De Vries, B. (2009). IfcOWL: a case of transforming EXPRESS schemas into ontologies. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*, 23(1):89–101.

[Bell et al., 2004] Bell, H., Bjørkhaug, L., and Rønning, J. (2004). ICT-Platform for object oriented knowledge in the building and construction industry. In *B4E conference "Building for a European Future*.

[Bizer et al., 2009] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., and Hellmann, S. (2009). Dbpedia-a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165.

[Caplan, 2009] Caplan, P. (2009). Understanding premis. Technical report, Library of Congress Network Development and MARC Standards Office.

[CCSDS, 2012] CCSDS (2012). Reference model for an open archival information system (oais) - magenta book. Technical report.

[Coyle and Hillmann, 2007] Coyle, K. and Hillmann, D. (2007). Resource description and access (rda): Cataloging rules for the 20th century. *D-Lib magazine*, 13(1):3.

[D'Arcus and Giasson, 2009] D'Arcus, B. and Giasson, F. (2009). Bibliographic ontology specification. specification document, 4 november 2009. *Retrieved August*, 10:2011.

[Grant et al., 2008] Grant, R., Bell, H., Bjorkhaug, L., and Bjaaland, A. (2008). IFD library white paper.

[Guenther, 2008] Guenther, R. (2008). Battle of buzzwords - flexibility vs. interoperability when implementing premis in mets. *D-Lib Magazine*, 14(7/8).

[Hartmann et al., 2005] Hartmann, J., Palma, R., Sure, Y., Suárez-Figueroa, M. C., Haase, P., Gómez-Pérez, A., and Studer, R. (2005). Ontology metadata vocabulary and applications. In Meersman, R., Tari, Z., and Herrero, P., editors, *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops*, number 3762 in Lecture Notes in Computer Science, pages 906–915. Springer Berlin Heidelberg.

[ISO12006-3, 2006] ISO12006-3 (2006). Building construction — organization of information about construction works — part 3: Framework for object-oriented information. certification state at time of this writing: FDIS.

[NISO, 2004] NISO (2004). Understanding metadata. Technical report, NISO - National Information Standards Organisation, Bethesada, MD.

[O'Neill, 2002] O'Neill, E. T. (2002). Frbr: Functional requirements for bibliographic records. *Library resources & technical services*, 46(4):150–159.

[PREMIS Editorial Committee, 2012] PREMIS Editorial Committee (2012). Premis data dictionary for preservation metadata, version 2.2. Technical report.

[Searle and Thompson, 2003] Searle, S. and Thompson, D. (2003). Preservation metadata. *D-Lib Magazine*, 9(4).

[Woestenenk, 2002] Woestenenk, K. (2002). The LexiCon: structuring semantics. In *Distributing Knowledge In Building*.

A Listings of schemas and examples

A.1 DURAARK building descriptive metadata schema (buildm)

```
# baseURI: http://duraark.eu/vocabularies/buildm
# imports: http://purl.org/dc/elements/1.1/
# imports: http://purl.org/dc/terms/
# imports: http://qudt.org/schema/qudt
# imports: http://qudt.org/vocab/unit
# imports: http://www.geonames.org/ontology
# imports: http://www.qudt.org/qudt/owl/1.0.0/quantity.owl
# imports: http://www.w3.org/TR/owl-time
# imports: http://www.w3.org/ns/regorg

@prefix :       <http://duraark.eu/vocabularies/buildm#> .
@prefix owl:    <http://www.w3.org/2002/07/owl#> .
@prefix qudt:   <http://qudt.org/schema/qudt#> .
@prefix rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:   <http://www.w3.org/2000/01/rdf-schema#> .
@prefix regorg: <http://www.w3.org/ns/regorg#> .
@prefix skos:   <http://www.w3.org/2004/02/skos/core#> .
@prefix time:   <http://www.w3.org/2006/time#> .
@prefix xsd:    <http://www.w3.org/2001/XMLSchema#> .

<http://duraark.eu/vocabularies/buildm>
      rdf:type owl:Ontology ;
      owl:imports <http://purl.org/dc/terms/> ,
          <http://www.qudt.org/qudt/owl/1.0.0/quantity.owl> ,
          <http://qudt.org/schema/qudt> , <http://www.geonames.org/ontology> ,
          <http://www.w3.org/TR/owl-time> , <http://www.w3.org/ns/regorg> ,
          <http://purl.org/dc/elements/1.1/> , <http://qudt.org/vocab/unit> ;
      owl:versionInfo "Created with TopBraid Composer"^^xsd:string .

:AreaUnit_1
      rdf:type qudt:AreaUnit ;
      rdfs:comment ""^^xsd:string ;
      rdfs:label "Area unit 1"^^xsd:string ;
      skos:prefLabel "Area unit 1"^^xsd:string .

:Building
      rdf:type owl:Class ;
      rdfs:label "DURAARK Building"^^xsd:string ;
      rdfs:subClassOf owl:Thing ;
      skos:definition "The DURAARK Building class is a domain-specific concept to
          capture descripte metadata about artefacts of the built enironement. It
          is desgigned to be used in the context of archival systems, especially
          such systems that are capable of storing Building Information Models
          (BIM) in archives for the purpose of Digital Longterm Preservation
          (DLP). In particular, this vocabulary has been designed to capture
          information that can be explicitly extracted or inferred from
          information residing in Industry Foundation Classes (IFC)
          models."^^xsd:string ;
      skos:prefLabel "Buidling"^^xsd:string .
```

```
:Building_1
      rdf:type :Building ;
      rdfs:comment ""^^xsd:string ;
      rdfs:label "Building 1"^^xsd:string ;
      :has_building_area
              [ rdf:value "150.0"^^xsd:double ;
                qudt:unit qudt:SquareMeter
              ] ;
      skos:prefLabel "Building 1"^^xsd:string .

:Resource_1
      rdf:type rdfs:Class .

:constructionDateEnd
      rdf:type owl:DatatypeProperty ;
      rdfs:comment "The year of completion of the initial construction.
          "^^xsd:string ;
      rdfs:label "construction date"^^xsd:string ;
      skos:prefLabel "construction date"^^xsd:string .

:constructionTime
      rdf:type owl:ObjectProperty ;
      rdfs:comment ""^^xsd:string ;
      rdfs:label "construction time"^^xsd:string ;
      rdfs:range time:ProperInterval ;
      skos:example "The construction of the Cologne Cathedral was started in 1248
          and finished in 1880"^^xsd:string ;
      skos:prefLabel "construction time"^^xsd:string .

:hasOwner
      rdf:type owl:ObjectProperty ;
      rdfs:comment ""^^xsd:string ;
      rdfs:label "has owner"^^xsd:string ;
      rdfs:range <http://www.w3.org/ns/RegisteredOrganization> ;
      skos:prefLabel "has owner"^^xsd:string .

:has_Location
      rdf:type owl:ObjectProperty ;
      rdfs:comment ""^^xsd:string ;
      rdfs:label "has Location"^^xsd:string ;
      rdfs:range <http://purl.org/dc/terms/Location> ,
          <http://www.geonames.org/ontology#Country> ;
      skos:prefLabel "has Location"^^xsd:string .

:has_building_area
      rdf:type owl:DatatypeProperty ;
      rdfs:comment "the gross floor according to local quantification method
          (Depends on legal etc. practises)"^^xsd:string ;
      rdfs:domain :Building ;
      rdfs:label "has building area"^^xsd:string ;
      rdfs:range xsd:double ;
      skos:prefLabel "has gross floor area"^^xsd:string .
```

```
:has_function
      rdf:type owl:DatatypeProperty ;
      rdfs:comment ""^^xsd:string ;
      rdfs:domain :Building ;
      rdfs:label "has function"^^xsd:string ;
      rdfs:range xsd:string ;
      skos:prefLabel "has function"^^xsd:string .

:has_name
      rdf:type owl:DatatypeProperty ;
      rdfs:comment ""^^xsd:string ;
      rdfs:domain :Building ;
      rdfs:label "buildingname"^^xsd:string ;
      rdfs:range xsd:string ;
      skos:example "\"Vertigo Building TU Eindhoven\"; \"Cologne
          cathedral\""^^xsd:string ;
      skos:prefLabel "buildingname"^^xsd:string .

:lengthUnit
      rdf:type owl:ObjectProperty ;
      rdfs:comment "Determines in which unit other properties are interpreted,
          e.g. qudt:SquareMeter"^^xsd:string ;
      rdfs:domain :Building ;
      rdfs:label "length unit"^^xsd:string ;
      rdfs:range qudt:AreaUnit ;
      skos:prefLabel "length unit"^^xsd:string .

:modificationDate
      rdf:type owl:DatatypeProperty ;
      rdfs:comment ""^^xsd:string ;
      rdfs:domain :Building ;
      rdfs:label "modification date"^^xsd:string ;
      rdfs:range xsd:dateTime ;
      skos:prefLabel "modification date"^^xsd:string .

:numberFloor
      rdf:type owl:DatatypeProperty ;
      rdfs:comment ""^^xsd:string ;
      rdfs:label "number of floors"^^xsd:string ;
      rdfs:range xsd:integer ;
      skos:prefLabel "number of floors"^^xsd:string .

:numberRoom
      rdf:type owl:DatatypeProperty ;
      rdfs:comment "The number of all rooms in the builing "^^xsd:string ;
      rdfs:label "number room"^^xsd:string ;
      rdfs:range xsd:integer ;
      skos:prefLabel "number room"^^xsd:string .
```

Listing 10: Listing of the buildm vocabulary for the description of buildings in PREMIS records

A.2 DURAARK IFC techncial metadata schema (ifcm)

```
# baseURI: http://duraark.eu/vocabularies/ifcm
# imports: http://bloody-byte.net/rdf/dc_owl2dl/dcmitype.ttl
# imports: http://purl.org/dc/terms/
# imports: http://www.loc.gov/premis/rdf/v1
# imports: http://www.w3.org/2004/02/skos/core
# imports: http://www.w3.org/TR/prov-o/

@prefix :        <http://duraark.eu/vocabularies/ifcm#> .
@prefix owl:     <http://www.w3.org/2002/07/owl#> .
@prefix premisowl:  <http://www.loc.gov/premis/rdf/v1#> .
@prefix provo:   <http://www.w3.org/TR/prov-o#> .
@prefix rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:    <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos:    <http://www.w3.org/2004/02/skos/core#> .
@prefix xsd:     <http://www.w3.org/2001/XMLSchema#> .

<http://duraark.eu/vocabularies/ifcm>
      rdf:type owl:Ontology ;
      owl:imports <http://bloody-byte.net/rdf/dc_owl2dl/dcmitype.ttl> ,
          <http://www.w3.org/TR/prov-o/> , <http://www.w3.org/2004/02/skos/core> ,
          <http://purl.org/dc/terms/> , <http://www.loc.gov/premis/rdf/v1> ;
      owl:versionInfo "Initital draft M12"^^xsd:string .

:IFCCharacteristics
      rdf:type owl:Class ;
      rdfs:comment "A domain specific PREMIS ObjectCharacteristics extension that
          describes File entities of the Industry Foundation Classes in their part
          21 STEP Physical File Format (SPFF)"^^xsd:string ;
      rdfs:label "IFC Characteristics"^^xsd:string ;
      rdfs:subClassOf premisowl:ObjectCharacteristicsExtension ;
      skos:prefLabel "IFC Characteristics"^^xsd:string .

:LODmetric
      rdf:type owl:DatatypeProperty ;
      rdfs:comment "a percentage ratio that determines the number of object per
          quibical meter"^^xsd:string ;
      rdfs:label "LODmetric"^^xsd:string ;
      rdfs:range xsd:float .

:attributeMetric
      rdf:type owl:DatatypeProperty ;
      rdfs:comment "the percentage of the OPTIONAL schema-level attributes that
          have are provied with values in this file"^^xsd:string ;
      rdfs:domain :IFCCharacteristics ;
      rdfs:label "attribute metric"^^xsd:string ;
      rdfs:range xsd:float .

:componentLibaries
      rdf:type owl:DatatypeProperty ;
```

```
        rdfs:comment "name of external component library that has been used. Can be
            instantiated multiple times"^^xsd:string ;
        rdfs:domain :IFCCharacteristics ;
        rdfs:label "component libaries"^^xsd:string ;
        rdfs:range xsd:string .

:creationTime
        rdf:type owl:DatatypeProperty ;
        rdfs:comment "timestamp of the generation of this file from the header
            "^^xsd:string ;
        rdfs:label "creation time"^^xsd:string ;
        rdfs:range xsd:dateTime .

:ifcVersion
        rdf:type owl:DatatypeProperty ;
        rdfs:comment "The version of the IFC file. Can be extracted from the header
            e.g. FILE_SCHEMA ((’IFC2X3’)). Will be IFC2x3 most of the times by the
            time of this  initial version"^^xsd:string ;
        rdfs:domain :IFCCharacteristics ;
        rdfs:label "has IFC version"^^xsd:string ;
        rdfs:range xsd:string .

:implemenationLevel
        rdf:type owl:DatatypeProperty ;
        rdfs:comment "The implementation level according to the ISO 10303:21, most
            often this will be ’2;1’"^^xsd:string ;
        rdfs:domain :IFCCharacteristics ;
        rdfs:label "implemenation level"^^xsd:string ;
        rdfs:range xsd:string .

:lastProducingApplication
        rdf:type owl:ObjectProperty ;
        rdfs:comment "the LAST Software application that has been involved in
            creating this model. In contrast to teh producingApplication property,
            this information should be taken from the header"^^xsd:string ;
        rdfs:domain :IFCCharacteristics ;
        rdfs:label "originating application"^^xsd:string ;
        rdfs:range <http://purl.org/dc/dcmitype/Software> .

:mvd    rdf:type owl:DatatypeProperty ;
        rdfs:comment "the Model View Definition this file complies to. Can be
            extracted from the FILE_DESCRIPTION section in the SPFF
            header"^^xsd:string ;
        rdfs:label "has MVD"^^xsd:string ;
        rdfs:range xsd:string .

:numberComponents
        rdf:type owl:DatatypeProperty ;
        rdfs:comment "The number of IfcProduct subtypes that have been instantiated
            (doors, windows, roofs, walls etc)"^^xsd:string ;
        rdfs:domain :IFCCharacteristics ;
        rdfs:label "number or products"^^xsd:string ;
        rdfs:range xsd:integer .
```

```
:numberEntities
      rdf:type owl:DatatypeProperty ;
      rdfs:comment ""^^xsd:string ;
      rdfs:domain :IFCCharacteristics ;
      rdfs:label "number entities"^^xsd:string ;
      rdfs:range xsd:integer .

:producingApplication
      rdf:type owl:ObjectProperty ;
      rdfs:comment "one of the  software application that has been involved in
          creating this model. This information can be extracted on a per-object
          record from the asscociated IfcOwnerHistory record. In theory. In
          practice this provenance metadata record is not used to its intended
          purpose in most implementing applications."^^xsd:string ;
      rdfs:domain :IFCCharacteristics ;
      rdfs:label "producing Application"^^xsd:string ;
      rdfs:range <http://purl.org/dc/dcmitype/Software> .

premisowl:ObjectCharacteristicsExtension
      rdf:type owl:Class ;
      rdfs:comment "A Container Class to implement domain-specific
          extensions"^^xsd:string ;
      rdfs:label "Object characteristics extension"^^xsd:string ;
      rdfs:subClassOf premisowl:ObjectCharacteristics ;
      skos:prefLabel "Object characteristics extension"^^xsd:string .
```

Listing 11: Listing of the ifcm vocabulary for the technical metadata of IFC models in PREMIS ObjectCharacteristics sections

A.3 DURAARK E57 techncial metadata schema (e57m)

```
# baseURI: http://duraark.eu/vocabularies/e57m
# imports: http://purl.org/dc/elements/1.1/
# imports: http://purl.org/dc/terms/
# imports: http://www.loc.gov/premis/rdf/v1
# imports: http://www.w3.org/2004/02/skos/core

@prefix :          <http://duraark.eu/vocabularies/e57m#> .
@prefix owl:       <http://www.w3.org/2002/07/owl#> .
@prefix premisowl:  <http://www.loc.gov/premis/rdf/v1#> .
@prefix rdf:       <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:      <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd:       <http://www.w3.org/2001/XMLSchema#> .

<http://duraark.eu/vocabularies/e57m>
      rdf:type owl:Ontology ;
      owl:imports <http://www.loc.gov/premis/rdf/v1> ,
          <http://purl.org/dc/elements/1.1/> , <http://purl.org/dc/terms/> ,
          <http://www.w3.org/2004/02/skos/core> ;
      owl:versionInfo "M12 initial draft "^^xsd:string .

:E57Characteristics
      rdf:type owl:Class ;
      rdfs:comment "A domain specific PREMIS ObjectCharacteristics extension that
          describes File entities in the E57 format to capture point cloud
          data"^^xsd:string ;
      rdfs:label "E57Characteristics"^^xsd:string ;
      rdfs:subClassOf premisowl:ObjectCharacteristicsExtension .

:distribtutionDetectedElements
      rdf:type owl:DatatypeProperty ;
      rdfs:comment """Distribution of detected elements is an indicator in
          percentage, of how many of the points within a
scan project are labelled as architectural elements and how many can be labelled
    as random elements
like e.g. planes not part of the architectural features. This can indicate the
    presence of other elements
like furniture etc. within the scan project."""^^xsd:string ;
      rdfs:label "distribtution elements"^^xsd:string ;
      rdfs:range xsd:float .

:distributionDetectedArchitecturalElements
      rdf:type rdf:Property ;
      rdfs:comment ""^^xsd:string ;
      rdfs:domain :E57Characteristics ;
      rdfs:label "distribution detected architectural elements"^^xsd:string ;
      rdfs:range xsd:float .

:exteriorFacadeAmount
      rdf:type owl:DatatypeProperty ;
```

```
      rdfs:comment """The amount of the building architecture within the scan
          project, which is covered by exterior scans  as a
percentage"""^^xsd:string ;
      rdfs:label "exterior facade amount"^^xsd:string ;
      rdfs:range xsd:float .

:hasColor
      rdf:type owl:DatatypeProperty ;
      rdfs:comment "shows if the project file contains RGB color values for the
          point cloud in the scan project."^^xsd:string ;
      rdfs:label "has color"^^xsd:string ;
      rdfs:range xsd:boolean .

:hasIntensity
      rdf:type owl:DatatypeProperty ;
      rdfs:comment """shows if the project file contains intensity values for the
          point cloud in the scan
project."""^^xsd:string ;
      rdfs:domain :E57Characteristics ;
      rdfs:label "has intensity"^^xsd:string ;
      rdfs:range xsd:boolean .

:libE57version
      rdf:type owl:DatatypeProperty ;
      rdfs:comment "The version of the .E57 lib of the specific .E57
          file"^^xsd:string ;
      rdfs:domain :E57Characteristics ;
      rdfs:label "lib Ee57 version"^^xsd:string ;
      rdfs:range xsd:string .

:numberPoints
      rdf:type owl:DatatypeProperty ;
      rdfs:comment "the total number of scan points stored in this e57 point cloud
          data file"^^xsd:string ;
      rdfs:label "number points"^^xsd:string ;
      rdfs:range xsd:integer .

:numberReferences
      rdf:type owl:DatatypeProperty ;
      rdfs:comment """In a scan if several references have angles that are close
          to each other they will be combined to one
distinct reference. Too few distinct references and the scan will get lower
    quality"""^^xsd:string ;
      rdfs:label "number of distinct references"^^xsd:string ;
      rdfs:range xsd:integer .

:numberScans
      rdf:type owl:DatatypeProperty ;
      rdfs:comment "number of scans that have been included into this e57 file
              "^^xsd:string ;
      rdfs:domain :E57Characteristics ;
      rdfs:label "number of scans"^^xsd:string ;
      rdfs:range xsd:integer .
```

```
:roomDistribution
      rdf:type owl:DatatypeProperty ;
      rdfs:comment """The amount of singled out rooms with no connectivity to
          other rooms or neighbouring rooms or
facades within an appropriate distance threshold set by architectural
    standards."""^^xsd:string ;
      rdfs:label "room distribution"^^xsd:string ;
      rdfs:range xsd:integer .

:spatialDisturbance
      rdf:type owl:DatatypeProperty ;
      rdfs:comment " indicating the amount of non-architectural elements, like
          furniture, noise etc. Should be \"high\",\"medium\", or \"low\"
          "^^xsd:string ;
      rdfs:domain :E57Characteristics ;
      rdfs:label "spatial disturbance"^^xsd:string ;
      rdfs:range xsd:string .

premisowl:ObjectCharacteristicsExtension
      rdf:type owl:Class ;
      rdfs:comment ""^^xsd:string ;
      rdfs:label "Object characteristics extension"^^xsd:string ;
      rdfs:subClassOf premisowl:ObjectCharacteristics .
```

Listing 12: Listing of the e57m vocabulary for the technical metadata of e57 point cloud files in PREMIS ObjectCharacteristics sections

A.4 VoL example

```
@prefix : <http://www.duraark.eu/Links4IfcEnrichment/def#>.
@prefix vol: <http://purl.org/vol/ns/>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix sd: <http://www.w3.org/ns/sparql-service-description#> .
@base <http://www.duraark.eu/Links4IfcEnrichment/def> .


:Link4WallStandardCase a vol:Link;
    rdfs:label "Link4WallStbdarCase";
    rdfs:comment "A link instance describing a link between IfcWallStandardCase
        and a Wall Concept in Semantic Concept set";
    vol:linksResource <http://www.duraark.eu/schema/IFC2X3#IfcWallStandardCase>;
    vol:linksResource <http://www.cbnl.nl/sc/def#SC4411>;
    vol:hasScore "1";
    vol:hasType "Semantic connectivity";
    .

:Link4Window a vol:Link;
    rdfs:label "Link4Window";
    rdfs:comment "A link instance describing a link between IfcWindow and a Window
        Concept in Semantic Concept set";
    vol:linksResource <http://www.duraark.eu/schema/IFC2X3#IfcWindow>;
    vol:linksResource <http://www.cbnl.nl/sc/def#SC4411>;
    vol:hasScore "1";
    vol:hasType "Semantic connectivity";
    .
```

Listing 13: Example of a mapping between an IFC Entity definition and an external dataset.

A.5 VoID example

```
@prefix void: <http://rdfs.org/ns/void#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix gn: <http://www.geonames.org/ontology#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix duraark : <http://purl.org/duraark/>
@prefix : <#> .

:DURAARK a void:Dataset;
    dcterms:title "DURAARK";
        dcterms:description "DURAARK Contextual Knowledge for archival";
        dcterms:contributor :L3S Research Center;
        dcterms:contributor :LUH;
        dcterms:contributor :TU_Eindhoven;
        dcterms:contributor :DURAARK_Consortium;
        dcterms:created "2013-10-07"^^xsd:date;
        dcterms:modified "2014-01-17"^^xsd:date;
        dcterms:publisher :Ujwal;
        foaf:page <http://duraark.eu/>;
        void:exampleResource <http://purl.org/duraark/Berliner_Dom>;
        void:exampleResource <http://purl.org/duraark/Empire_State_Building>;
        void:triples 768994500;
        void:entities 457800;
    void:classPartition [
        void:class duraark:wall;
        void:entities 398000;
    ];
    void:classPartition [
        void:class duraark:roof;
        void:entities 75000;
    ];
        void:propertyPartition [
            void:property duraark:hasDoor;
            void:triples 124590;
        ];
        void:propertyPartition [
            void:property duraark:hasWindow;
            void:triples 156510;
        ];
        void:propertyPartition [
            void:property rdf:type;
        ];
        .
    :L3S Research Center a foaf:Organization;
        rdfs:label "Forschungszentrum L3S";
        foaf:homepage <http://www.l3s.de/>;
```

```
        .
    :Ujwal a foaf:person;
        rdfs:label "Ujwal Gadiraju";
        foaf:mbox <gadiraju@l3s.de>;
        .


:DURAARK_to_DBpedia a void:Linkset;
    void:target :DURAARK;
    void:target :DBpedia;
    void:linkPredicate owl:sameAs;
        void:triples 500000;
        .

:DURAARK2DBpedia a void:Linkset;
    void:target :DURAARK;
    void:target :DBpedia;
    void:linkPredicate foaf:primaryTopicOf;
        void:triples 304500;
        .

:DURAARK_to_GeoNames a void:Linkset;
    void:target :DURAARK;
    void:target :GeoNames;
    void:linkPredicate foaf:primaryTopicOf;
    void:triples 917500;
    .

:DURAARK_to_BauDataWeb a void:Linkset;
    void:target :DURAARK;
    void:target :BauDataWeb;
    void:linkPredicate duraark:hasMaterial;
    void:triples 63400;
    .
```

Listing 14: Example of a VoID description of datasets and links alongside provenance data.

A.6 IFC reference example

```
ISO-10303-21;
HEADER;
FILE_DESCRIPTION(('ViewDefinition [CoordinationView]'),'2;1');
FILE_NAME('Project Number','2013-10-30T19:29:09',(''),(''),'Autodesk Revit
    Architecture 2011 - 1.0','20100903_2115(x64)','');
FILE_SCHEMA(('IFC2X3'));
ENDSEC;
DATA;
#1=IFCORGANIZATION($,'Autodesk Revit Architecture 2011',$,$,$);
#2=IFCAPPLICATION(#1,'2011','Autodesk Revit Architecture 2011','Revit');
#4=IFCCARTESIANPOINT((0.,0.));
#5=IFCDIRECTION((1.,0.,0.));
#6=IFCDIRECTION((-1.,0.,0.));
#10=IFCDIRECTION((0.,0.,-1.));
#11=IFCDIRECTION((1.,0.));
#12=IFCDIRECTION((-1.,0.));
#13=IFCDIRECTION((0.,1.));
#14=IFCDIRECTION((0.,-1.));
#15=IFCSIUNIT(*,.LENGTHUNIT.,.MILLI.,.METRE.);
#16=IFCSIUNIT(*,.AREAUNIT.,.MILLI.,.SQUARE_METRE.);
#17=IFCSIUNIT(*,.VOLUMEUNIT.,.MILLI.,.CUBIC_METRE.);
#18=IFCSIUNIT(*,.PLANEANGLEUNIT.,$,.RADIAN.);
#19=IFCDIMENSIONALEXPONENTS(0,0,0,0,0,0,0);
#20=IFCMEASUREWITHUNIT(IFCRATIOMEASURE(0.0174532925199432 8),#18);
#21=IFCCONVERSIONBASEDUNIT(#19,.PLANEANGLEUNIT.,'DEGREE',#20);
#22=IFCSIUNIT(*,.TIMEUNIT.,$,.SECOND.);
#23=IFCUNITASSIGNMENT((#15,#16,#17,#21,#22));
#26=IFCAXIS2PLACEMENT3D(#3,$,$);
#27=IFCGEOMETRICREPRESENTATIONCONTEXT($,'Model',3,1.E-006,#26,$);
#28=IFCGEOMETRICREPRESENTATIONCONTEXT($,'Plan',3,1.E-006,#26,$);
#29=IFCGEOMETRICREPRESENTATIONSUBCONTEXT($,'Plan',*,*,*,*,#28,0.01,.PLAN_VIEW.,$);
#30=IFCPERSON($,'LONG','Qinqin',$,$,$,$,$);
#31=IFCORGANIZATION($,'','',$,$);
#32=IFCPERSONANDORGANIZATION(#30,#31,$);
#33=IFCOWNERHISTORY(#32,#2,$,.NOCHANGE.,$,$,$,0);
#35=IFCPOSTALADDRESS($,$,$,$,('Enter address here'),$,'<Default>','','','');
#39=IFCBUILDINGSTOREY('2Hh0Ap8hrFGOA15cviZI2C',#33,'Level
    1',$,$,#38,$,$,.ELEMENT.,0.);
#40=IFCCARTESIANPOINT((0.,0.,4000.));
#41=IFCAXIS2PLACEMENT3D(#40,$,$);
#47=IFCCARTESIANPOINT((3400.,0.));
#48=IFCPOLYLINE((#4,#47));
#49=IFCSHAPEREPRESENTATION(#27,'Axis','Curve2D',(#48));
#50=IFCCARTESIANPOINT((1700.,0.));
#51=IFCAXIS2PLACEMENT2D(#50,#12);
#52=IFCRECTANGLEPROFILEDEF(.AREA.,$,#51,3400.000000000001,200.);
#53=IFCAXIS2PLACEMENT3D(#3,$,$);
#54=IFCEXTRUDEDAREASOLID(#52,#53,#9,4000.000000000195);
#55=IFCCOLOURRGB($,0.5019607843137255,0.5019607843137255,0.5019607843137255);
```

```
#56=IFCSURFACESTYLERENDERING(#55,0.,$,$,$,$,IFCNORMALISEDRATIOMEASURE(0.00390625),
    IFCSPECULAREXPONENT(64.),.NOTDEFINED.);
#57=IFCSURFACESTYLE('Default Wall',.BOTH.,(#56));
#58=IFCPRESENTATIONSTYLEASSIGNMENT((#57));
#59=IFCSTYLEDITEM(#54,(#58),$);
#60=IFCSHAPEREPRESENTATION(#27,'Body','SweptSolid',(#54));
#61=IFCPRODUCTDEFINITIONSHAPE($,$,(#49,#60));
#103=IFCMATERIAL('Default Wall');
#104=IFCPRESENTATIONSTYLEASSIGNMENT((#57));
#105=IFCSTYLEDITEM($,(#104),$);
#106=IFCSTYLEDREPRESENTATION(#29,'Style','Material',(#105));
#107=IFCMATERIALDEFINITIONREPRESENTATION($,$,(#106),#103);
#108=IFCMATERIALLAYER(#103,200.,$);
#109=IFCMATERIALLAYERSET((#108),'Basic Wall:Generic - 200mm');
#110=IFCMATERIALLAYERSETUSAGE(#109,.AXIS2.,.NEGATIVE.,100.);
#111=IFCCARTESIANPOINT((-724.0288406218866,-2789.352874261761,0.));
#112=IFCAXIS2PLACEMENT3D(#111,#9,#8);
#62=IFCWALLSTANDARDCASE('3hE4njHGr57Puzv5XHW9UO',#33,'Basic Wall:Generic -
    200mm:127694', $, 'Basic Wall:Generic - 200mm:398',#46,#61,'127694');
#114=IFCCARTESIANPOINT((750.,457.4999999999997));
#115=IFCAXIS2PLACEMENT2D(#114,#11);
#116=IFCRECTANGLEPROFILEDEF(.AREA.,$,#115,1499.999999999998,914.9999999999993);
#117=IFCAXIS2PLACEMENT3D(#3,#7,#9);
#118=IFCEXTRUDEDAREASOLID(#116,#117,#9,200.);
#119=IFCSHAPEREPRESENTATION(#27,'Body','SweptSolid',(#118));
#120=IFCPRODUCTDEFINITIONSHAPE($,$,(#119));
#124=IFCOPENINGELEMENT('2Hh0Ap8hrFGOA15cviZDBz',#33,'M_Fixed:0915 x 1500mm:0915 x
    1500mm:127814:1',$,'Opening',#123,#120,$);
#125=IFCRELVOIDSELEMENT('3EMqedU0XFxhlFYnnHrGLE',#33,$,$,#62,#124);
#126=IFCAXIS2PLACEMENT2D(#4,#11);
#127=IFCRECTANGLEPROFILEDEF(.AREA.,$,#126,12.,788.9999999999972);
#128=IFCCARTESIANPOINT((457.5000000000056,159.,63.));
#129=IFCAXIS2PLACEMENT3D(#128,#9,#8);
#130=IFCEXTRUDEDAREASOLID(#127,#129,#9,1374.000000000007);
#131=IFCCOLOURRGB($,0.,0.5019607843137255,0.7529411764705882);
#132=IFCSURFACESTYLERENDERING(#131,0.8999999761581421,$,$,$,$,IFCNORMALISEDRATIOMEASURE(0.00390625),
    IFCSPECULAREXPONENT(12.),.NOTDEFINED.);
#133=IFCSURFACESTYLE('Glass',.BOTH.,(#132));
#134=IFCPRESENTATIONSTYLEASSIGNMENT((#133));
#135=IFCSTYLEDITEM(#130,(#134),$);
#136=IFCCARTESIANPOINT((-394.4999999999997,-687.));
#137=IFCCARTESIANPOINT((-394.4999999999997,687.));
#138=IFCCARTESIANPOINT((394.4999999999975,687.));
#139=IFCCARTESIANPOINT((394.4999999999975,-687.));
#140=IFCCARTESIANPOINT((-394.4999999999997,-687.));
#141=IFCPOLYLINE((#136,#137,#138,#139,#140));
#142=IFCCARTESIANPOINT((-438.4999999999999,-731.));
#143=IFCCARTESIANPOINT((438.4999999999999,-731.));
#144=IFCCARTESIANPOINT((438.4999999999999,731.));
#145=IFCCARTESIANPOINT((-438.4999999999999,731.));
#146=IFCCARTESIANPOINT((-438.4999999999999,-731.));
#147=IFCPOLYLINE((#142,#143,#144,#145,#146));
#148=IFCARBITRARYPROFILEDEFWITHVOIDS(.AREA.,$,#147,(#141));
```

```
#149=IFCCARTESIANPOINT((457.5000000000067,137.,750.));
#150=IFCAXIS2PLACEMENT3D(#149,#7,#5);
#151=IFCEXTRUDEDAREASOLID(#148,#150,#9,44.00000000002744);
#152=IFCCOLOURRGB($,0.8784313725490196,0.6980392156862745,0.4941176470588236);
#153=IFCSURFACESTYLERENDERING(#152,0.,$,$,$,$,IFCNORMALISEDRATIOMEASURE(0.00390625),
    IFCSPECULAREXPONENT(128.), .NOTDEFINED.);
#154=IFCSURFACESTYLE('Sash',.BOTH.,(#153));
#155=IFCPRESENTATIONSTYLEASSIGNMENT((#154));
#156=IFCSTYLEDITEM(#151,(#155),$);
#157=IFCCARTESIANPOINT((-438.5000000000011,-731.));
#158=IFCCARTESIANPOINT((-438.5000000000013,731.));
#159=IFCCARTESIANPOINT((438.4999999999986,731.));
#160=IFCCARTESIANPOINT((438.4999999999987,-731.));
#161=IFCCARTESIANPOINT((-438.5000000000011,-731.));
#162=IFCPOLYLINE((#157,#158,#159,#160,#161));
#163=IFCCARTESIANPOINT((-457.4999999999998,-750.));
#164=IFCCARTESIANPOINT((457.4999999999998,-750.));
#165=IFCCARTESIANPOINT((457.4999999999998,750.));
#166=IFCCARTESIANPOINT((-457.4999999999998,750.));
#167=IFCCARTESIANPOINT((-457.4999999999998,-750.));
#168=IFCPOLYLINE((#163,#164,#165,#166,#167));
#169=IFCARBITRARYPROFILEDEFWITHVOIDS(.AREA.,$,#168,(#162));
#170=IFCCARTESIANPOINT((457.500000000008,137.,750.));
#171=IFCAXIS2PLACEMENT3D(#170,#7,#5);
#172=IFCEXTRUDEDAREASOLID(#169,#171,#9,63.00000000002743);
#173=IFCPRESENTATIONSTYLEASSIGNMENT((#154));
#174=IFCSTYLEDITEM(#172,(#173),$);
#175=IFCCARTESIANPOINT((-425.4999999999999,-718.));
#176=IFCCARTESIANPOINT((-425.4999999999999,718.));
#177=IFCCARTESIANPOINT((425.5000000000002,718.));
#178=IFCCARTESIANPOINT((425.5000000000002,-718.));
#179=IFCCARTESIANPOINT((-425.4999999999999,-718.));
#180=IFCPOLYLINE((#175,#176,#177,#178,#179));
#181=IFCCARTESIANPOINT((-457.5000000000002,-750.));
#182=IFCCARTESIANPOINT((457.5000000000002,-750.));
#183=IFCCARTESIANPOINT((457.5000000000002,750.));
#184=IFCCARTESIANPOINT((-457.5000000000002,750.));
#185=IFCCARTESIANPOINT((-457.5000000000002,-750.));
#186=IFCPOLYLINE((#181,#182,#183,#184,#185));
#187=IFCARBITRARYPROFILEDEFWITHVOIDS(.AREA.,$,#186,(#180));
#188=IFCCARTESIANPOINT((457.5000000000084,0.,750.));
#189=IFCAXIS2PLACEMENT3D(#188,#7,#5);
#190=IFCEXTRUDEDAREASOLID(#187,#189,#9,136.9999999999726);
#191=IFCPRESENTATIONSTYLEASSIGNMENT((#154));
#192=IFCSTYLEDITEM(#190,(#191),$);
#193=IFCSHAPEREPRESENTATION(#27,'Body','SweptSolid',(#130,#151,#172,#190));
#194=IFCCARTESIANPOINT((63.,159.));
#195=IFCCARTESIANPOINT((852.,159.));
#196=IFCPOLYLINE((#194,#195));
#197=IFCCARTESIANPOINT((883.,137.));
#198=IFCCARTESIANPOINT((883.,0.));
#199=IFCPOLYLINE((#197,#198));
#200=IFCCARTESIANPOINT((915.,0.));
```

```
#201=IFCCARTESIANPOINT((883.,0.));
#202=IFCPOLYLINE((#200,#201));
#203=IFCCARTESIANPOINT((852.,137.));
#204=IFCCARTESIANPOINT((896.,137.));
#205=IFCPOLYLINE((#203,#204));
#206=IFCCARTESIANPOINT((852.,181.));
#207=IFCCARTESIANPOINT((852.,137.));
#208=IFCPOLYLINE((#206,#207));
#209=IFCCARTESIANPOINT((896.,181.));
#210=IFCCARTESIANPOINT((852.,181.));
#211=IFCPOLYLINE((#209,#210));
#212=IFCCARTESIANPOINT((896.,200.));
#213=IFCCARTESIANPOINT((896.,137.));
#214=IFCPOLYLINE((#212,#213));
#215=IFCCARTESIANPOINT((913.8782114087891,200.));
#216=IFCCARTESIANPOINT((896.,200.));
#217=IFCPOLYLINE((#215,#216));
#218=IFCCARTESIANPOINT((32.,137.));
#219=IFCCARTESIANPOINT((32.,0.));
#220=IFCPOLYLINE((#218,#219));
#221=IFCCARTESIANPOINT((63.,181.));
#222=IFCCARTESIANPOINT((63.,137.));
#223=IFCPOLYLINE((#221,#222));
#224=IFCCARTESIANPOINT((19.,200.));
#225=IFCCARTESIANPOINT((19.,137.));
#226=IFCPOLYLINE((#224,#225));
#227=IFCCARTESIANPOINT((19.,137.));
#228=IFCCARTESIANPOINT((63.,137.));
#229=IFCPOLYLINE((#227,#228));
#230=IFCCARTESIANPOINT((19.,181.));
#231=IFCCARTESIANPOINT((63.,181.));
#232=IFCPOLYLINE((#230,#231));
#233=IFCCARTESIANPOINT((32.,0.));
#234=IFCPOLYLINE((#4,#233));
#235=IFCCARTESIANPOINT((0.,200.));
#236=IFCCARTESIANPOINT((19.,200.));
#237=IFCPOLYLINE((#235,#236));
#238=IFCGEOMETRICSET((#196,#199,#202,#205,#208,#211,#214,#217,#220,#223,#226,#229,#232,#234,#237));
#239=IFCSHAPEREPRESENTATION(#29,'Plan','GeometricSet',(#238));
#240=IFCAXIS2PLACEMENT3D(#3,$,$);
#241=IFCREPRESENTATIONMAP(#240,#193);
#242=IFCREPRESENTATIONMAP(#240,#239);
#243=IFCWINDOWLININGPROPERTIES('1buJOWNff03QYYiPJ_yE5z',#33,'M_Fixed:0915 x
    1500mm:0915 x 1500mm:127814',$,$,$,$,$,$,$,$,$,$);
#244=IFCWINDOWSTYLE('3FJuE3T117zATlIp3trRDs',#33,'0915 x
    1500mm',$,$,(#243),(#241,#242),'127814',.NOTDEFINED.,.NOTDEFINED.,.F.,.F.);
#245=IFCCARTESIANTRANSFORMATIONOPERATOR3D($,$,#3,1.,$);
#246=IFCMAPPEDITEM(#241,#245);
#247=IFCSHAPEREPRESENTATION(#27,'Body','MappedRepresentation',(#246));
#248=IFCCARTESIANTRANSFORMATIONOPERATOR3D($,$,#3,1.,$);
#249=IFCMAPPEDITEM(#242,#248);
#250=IFCSHAPEREPRESENTATION(#29,'Plan','MappedRepresentation',(#249));
#251=IFCPRODUCTDEFINITIONSHAPE($,$,(#247,#250));
```

```
#8=IFCDIRECTION((0.,-1.,0.));
#253=IFCAXIS2PLACEMENT3D(#252,#9,#8);
#252=IFCCARTESIANPOINT((-811.028840621876,-2331.852874261744,915.));
#123=IFCLOCALPLACEMENT(#46,#122);
#46=IFCLOCALPLACEMENT(#38,#45);
#38=IFCLOCALPLACEMENT(#25,#37);
#24=IFCAXIS2PLACEMENT3D(#3,$,$);
#37=IFCAXIS2PLACEMENT3D(#3,$,$);
#3=IFCCARTESIANPOINT((0.,0.,0.));
#9=IFCDIRECTION((0.,0.,1.));
#7=IFCDIRECTION((0.,1.,0.));
#45=IFCAXIS2PLACEMENT3D(#44,#9,#7);
#44=IFCCARTESIANPOINT((-711.028840621878,-4386.352874261744,0.));
#122=IFCAXIS2PLACEMENT3D(#121,$,$);
#121=IFCCARTESIANPOINT((1139.5,-100.,915.));
#255=IFCCARTESIANPOINT((915.,200.,0.));
#256=IFCAXIS2PLACEMENT3D(#255,#9,#6);
#257=IFCLOCALPLACEMENT(#123,#256);
#258=IFCWINDOW('3hE4njHGr57Puzv5XHW9OG',#33,'M_Fixed:0915 x 1500mm:0915 x
    1500mm:127814',$,'0915 x
    1500mm',#257,#251,'127814',1499.999999999998,914.9999999999997);
#293=IFCRELFILLSELEMENT('3DR03AYWz6BQxNbUh1fYC2',#33,$,$,#124,#258);
#294=IFCAXIS2PLACEMENT3D(#3,$,$);
#295=IFCLOCALPLACEMENT($,#294);
#297=IFCRELAGGREGATES('0SpLwQTm11JxOHjN76Iv03',#33,$,$,#34,(#296));
#298=IFCRELAGGREGATES('3lcY8x5mD19QQwb6bAT$Np',#33,$,$,#296,(#36));
#36=IFCBUILDING('2Hh0Ap8hrFGOA15cwJSjvA',#33,$,$,$,#25,$,$,.ELEMENT.,$,$,#35);
#296=IFCSITE('2Hh0Ap8hrFGOA15cwJSjv9',#33,'Default',$,'',#295,$,$,.ELEMENT.,(42,21,30,344238),
    (-71,-3,-35,-194702), -0.,$,$);
#25=IFCLOCALPLACEMENT(#295,#24);
#317=IFCRELCONTAINEDINSPATIALSTRUCTURE('3nNeWI6YP5SuEpAIjWMzgx',#33,$,$,(#62,#258),#39);
#318=IFCRELAGGREGATES('3XlW0aawn4BPkVCcCFgRZq',#33,$,$,#36,(#39));
#330=IFCRELASSOCIATESMATERIAL('1VAxNo9HX4Thv4XOgWxcgs',#33,$,$,(#62),#110);
#331=IFCRELDEFINESBYTYPE('2fVfL$sfnDD9HH4vhZ88_W',#33,$,$,(#258),#244);
#342=IFCPRESENTATIONLAYERASSIGNMENT('A-GLAZ',$,(#193,#239,#247,#250),$);
#343=IFCPRESENTATIONLAYERASSIGNMENT('A-WALL-MBNI',$,(#49,#60,#119),$);
#34=IFCPROJECT('2Hh0Ap8hrFGOA15cwJSjvB',#33,'Project Number',$,$,'','Project
    Status',(#27,#28),#23);
#344=IFCRELASSOCIATESLIBRARY('1k7LvqpHjCpQXRvTBnrOc8',#33,$,$,(#62),#345);
#345=IFCLIBRARYREFERENCE('http://www.duraark.eu/Links4IfcEnrichment/def#Link4WallStandardCase',$,$);
#346=IFCRELASSOCIATESLIBRARY('3HE0BD9pbBIxZAAa7cCRLP',#33,$,$,(#258),#347);
#347=IFCLIBRARYREFERENCE('http://www.duraark.eu/Links4IfcEnrichment/def#Link4Window',$,$);
ENDSEC;
END-ISO-10303-21;
```

Listing 15: Showing the uses of external links in an minimal IFC SPFF file.

List of Terms

BIM Building Information Model (BIM). 3, 8, 86

LTP Long Term Preservation. 3, 86

OAIS Open Archival Information System. 86

SDA Semantic Digital Archive. 86

SDO Semantic Digital Observatory. 86

SPF STEP Physical File (SPF). 10, 30, 86

STEP STandard for the Exchange of Product data (STEP). 10, 86

Glossary

Building Information Model Object-oriented, parametric and process-oriented data structures to organize information relevant to buildings. 13, 86

IFC The Industry Foundation Classes (IFC) is an open interoperability model for the exchange of information reltated to building and construction. 10, 30, 31, 64, 86

Long Term Preservation The internationally widely accepted Open Archival Information System (OAIS) framework, defines preservation as

> "The act of maintaining information, Independently Understandable by a Designated Community, and with evidence supporting its Authenticity, over the Long Term."

(source: OAIS "Pink Book"). 8, 13, 29, 86

Open Archival Information System Framework defining general concepts and best practises for Digital Long Term Preservation. 11, 13, 86

Semantic Digital Archive a part of the DURAARK framework that stores snapshots of linked data sets referenced from archives and their descriptions. 86

Semantic Digital Observatory a part of the DURAARK system that crawls, fetches, monitores and updates external data sets stored for preservation in the SDA. 86

STandard for the Exchange of Product data (STEP) A group of information standards covering a wide spectrum of engineering domains grouped under the ISO 10303 series of standards. 86

STEP Physical File A clear-text file format defined in the STandard for the Exchange of Product Data (STEP). Specified in the ISO 10303, part 21 and often referred to as "Part 21 file" or "SPFF" (STEP Physical File Format). 86