


ORIGINAL ARTICLE

Open Access



A metadata ADE for CityGML

Anna Labetski* , Kavisha Kumar, Hugo Ledoux and Jantien Stoter

Abstract

While there exist international standards for geospatial metadata (ISO 19115), these are rarely used in practice for 3D datasets, and one of the OGC standards for 3D city models, CityGML, does not offer a mechanism to store metadata in a structured way. Having metadata in CityGML files, which are in practice often very large and complex, would provide us with the ability to quickly understand the nature of a dataset and to determine if it is relevant for a specific task. A lack of metadata introduces uncertainty into models that are already full of assumptions and estimations. In this paper, we first examine the metadata needs that are specific for 3D geographical datasets and propose ISO 19115 compliant categories. We then describe how these can be used within CityGML by defining an Application Domain Extension (ADE), which allows us to store metadata for existing city objects of CityGML, as well as objects in other domain-specific ADEs. Our ADE, its schema in both UML and XSD, and sample datasets is openly accessible, and it can be easily extended to support application specific metadata. In addition the metadata elements have been added to the core of CityJSON. We also offer software to generate automatically many of the metadata categories and we propose coupling it with the source 3D dataset.

Keywords: 3D city models, Metadata, CityGML, Application domain extension, UML, Automation

Introduction

On the basic level, ‘metadata’ is simply data describing other data, but more broadly it can be in reference to a range of information technology resources (e.g. data, services, knowledge stores) and it has four primary functions: locate, evaluate, access, and employ [7]. The boundary between data and metadata can be a fuzzy one but one distinction that we use in this paper is: metadata is structured to some degree and this structuring is what converts “raw information into actionable metadata” [23].

The use of metadata to describe and document data has several advantages. First, it is crucial for ensuring the interoperability of data, i.e. after data is exchanged it can be easily understood and is usable for different applications [7]. Second, metadata also ensures that data creators and data users from different domains can understand and communicate about data requirements and usability [18]. We can state that the presence of metadata is as crucial for achieving transparency as the presence of bibliography in an academic print publication [15]. Third, metadata facilitates *resource discovery*, which means that users can discover (e.g. on the web or in a portal) relevant datasets

for a specific application. This ensures that the data will be utilised correctly.

Metadata exists in most knowledge sectors and has unique specifications and standards depending on the domain. In geographical domains, the international standard ISO 19115 (we describe it in details in “ISO 19115” section) is relatively mature, several GIS software implement it, and is used by practitioners. Its use is however mostly restrained to 2D datasets (both raster and vector), 3D datasets very rarely have metadata information stored. One cause is probably because, as highlighted by Danko [7], the specifications of ISO 19115 do not cover several aspects specified to 3D datasets. If we consider specifically 3D city models then we can state that metadata are almost never stored. One probably cause is that CityGML, the international standard for storing them [22], offers no mechanisms to store metadata in a structured way. Practitioners often need to define their own definitions for CityGML metadata and create their own methodology for storing it [11, 25]. Metadata is also necessary in the development of CityGML extensions in order to track and manage the diversity of data sources and data qualities [20]. As further explained in “CityGML, ADEs, and metadata support” section, only a generic “container” for metadata can be used, and any

*Correspondence: a.labetski@tudelft.nl
3D geoinformation, Delft University of Technology, Julianalaan 134, 2628BL Delft, The Netherlands

XML elements can be used. This is ironic because, in our opinion, CityGML files would most benefit from having (structured) metadata explicitly stored: in practice the size and the complexity of a given CityGML file are way larger and higher than a 2D dataset of the same region. This means that parsing an unknown CityGML dataset to extract information is no easy task for practitioners. Currently, we have observed that many people write their own parsers and a great deal of time is lost on simply analysing a dataset to understand what it contains; examples are the bounding box, the year of creation, the levels of detail in the dataset, etc. Also, because they are too big, many real-world datasets can simply not be opened by text editors, let alone be visualised by a GIS viewer. While several datasets are tiled into subparts, e.g. the openly available dataset of Berlin¹, they are still very large in size, often 5GB+ for a single tile. Having metadata attached to the CityGML file would help in assessing the fitness-for-purpose of data for use within a specific application.

In this paper, we first analyse the metadata needs for 3D city models and for CityGML datasets in particular. In “[3D geospatial metadata needs](#)” section, we built upon the work of Dietze et al. [8] and propose a set of metadata categories, which are ISO 19115 compliant, that ought to be stored. Based on these, we propose in “[Result #1: Our 3D metadata ADE](#)” section an Application Domain Extension (ADE) to extend CityGML so that these can be easily included in CityGML datasets in a structured manner. As explained in “[Extendability](#)” section, our ADE allows us to store metadata for existing city objects of CityGML, as well as for other ADEs. That is, one could define specific metadata elements in her ADE (let say for energy or noise) and use our ADE to store these specific metadata; we believe to be the first to allow an ADE to use (be used by) another ADEs. In “[Result #2: automation and discovery](#)” section we explain our methodology for automating the metadata creation and storing it in a database, this is to ease the discovery of 3D city models. We have implemented our ADE and we offer the UML model, the XML schemas, sample datasets, and Python code to automatically create/extract the metadata from a given CityGML file. Finally, our results have also been added in the core of CityJSON², a JSON encoding of the CityGML data model.

Background

ISO 19115

While there exist multiple metadata standards, such as the American Content Standard for Digital Geospatial Metadata³ (CSGDM) and the European INSPIRE Metadata Directive⁴, their usage is being phased out in favour of ISO 19115 due to its international focus [7]. ISO 19115 is the metadata standard specifically for *geographic information* developed by the International Organization

for Standardization. ISO 19115 (latest revision is from 2014) defines mandatory, conditional and optional metadata attributes such as dataset title, responsible party and conditions of use [13]. The standard also provides guidance for the minimum set of metadata attributes required to serve most metadata applications, these are: data discovery, determining data fitness for use, data access, data transfer, and use of digital data and services [13]. It is composed of 13 packages, individual packages may be used alone to provide separate components of metadata to meet specific use case requirements [13].

The previous version of ISO 19115 (2003 and including later revisions) did not provide any encoding and an XML encoding was specified in ISO 19139 [12]. With the 2014 version of ISO 19115 there was a re-definition of the standard by splitting it into 3 parts: Part 1 - *Fundamentals*, Part 2 - *Extensions for imagery and gridded data*, and Part 3 - *XML schema implementation of metadata fundamentals* [13]. Part 3 was published in 2016 and defines XML schemas for encoding Parts 1 and 2, effectively superseding ISO 19139 [14].

The first function of metadata is data discovery, Table 1 outlines the ‘*Metadata for the discovery of geographic datasets and series*’ as defined by ISO [13]. The discovery list of attributes matches the attributes that were previously part of the metadata *core* concept present in previous versions of ISO 19115 and are therefore the attributes most commonly associated with the standard and featured in the metadata packages of many GISs. The list was also specifically designed to contain the metadata to be exposed to facilitate discovery, particularly in metadata catalogues ISO [13]. This is why our metadata proposal (see “[Result #1: Our 3D metadata ADE](#)” section) also focuses on meeting the requirements necessary for discovering 3D city model datasets.

3D geospatial metadata needs

Despite the popularity and growing implementations of ISO 19115, Dietze et al. [8] examined its applicability for 3D city models and found that while there exist several attributes that are important there are further attributes that are missing, the most prominent being the level of detail and semantic object classes (e.g. buildings, roads, etc.). Additionally, Biljecki et al. [4] found that the modelling choices made by practitioners when acquiring, processing and utilising 3D city models are rarely documented in the metadata of a dataset, often because there is no way to store this information. This information is necessarily not only for dataset discovery but also to ensure interoperability between various 3D city models [26].

A lack of metadata being present in 3D city models means it is more difficult to integrate them in 3D spatial data infrastructures (SDIs) where metadata is an important base [19]. Furthermore ISO 19115 alone was found to

Table 1 ISO 19115-1:2014 - Table F.1: *Metadata for the discovery of geographic datasets and series* [13]

Metadata element	Obligation	Comment
Metadata reference information	Optional	Unique identifier for the metadata.
Resource title	Mandatory	Title by which the resource is known.
Resource reference data	Optional	A date which is used to help identify the resource.
Resource identifier	Optional	Unique identifier for the resource.
Resource point of contact	Optional	Name of the person, position, or organisation responsible for the resource.
Geographic location	Conditional ^a	Geographic description of coordinates (latitude/longitude) which describes the location of the resource.
Resource language	Conditional	The language and character set used in the resource.
Resource topic category	Conditional	A selection of the 20 elements in the MD_TopicCategory enumeration which describe the topic of the resource.
Spatial resolution	Optional	The nominal scale and/or/spatial resolution of the resource.
Resource type	Conditional	A resource code identifying the type of resource.
Resource abstract	Mandatory	A brief description of the content of the resource.
Extent information for the dataset (additional)	Optional	The temporal or vertical extent of the resource.
Resource lineage	Optional	A description of the source(s) and production process(es) used in producing the resource.
Resource on-line Link	Optional	Link (URL) in the metadata for the resource.
Keywords	Optional	Words or phrases describing the resource to be indexed and searched.
Constraints on resource access and use	Optional	Restrictions on the access and use of the resources.
Metadata date stamp	Mandatory	Reference date(s) for the metadata, especially creation.
Metadata point of contact	Mandatory	The party responsible for the metadata.

^a Conditional means that certain elements become mandatory based on the values of other elements

be lacking specifications to facilitate the development of a 3D SDI, especially for 3D city models [2].

Resource/data discovery

Resource discovery is understood as the problems associated with locating resources of interest, based on spatial location or attributes, in large-scale resource-intensive environments [1]. The size and complexity of 3D city models make their discovery (e.g. number of specific objects types or temporal and spatial extent) difficult due to the complicated encoding (very large XML, with a nested structure). There are further difficulties introduced when attempting to visualise datasets given that many are larger than the memory of a typical computer. One of the ways in which resource discovery can easily be facilitated is through the usage of the world wide web and there is a growing body of research into the querying, storage and usage of the web to facilitate metadata discovery [24]. The ease comes from reducing the need for data users to individually parse every file they encounter.

Zamyadi et al. [26] state that metadata for 3D geoinformation is very different from one online portal to another and even for 3D datasets within the same portal. Furthermore the integration of 3D datasets within traditional 2D portals is difficult because 2D portals do not contain specific and common metadata categories and attributes required by 3D datasets [26].

CityGML, ADEs, and metadata support

CityGML is an open 3D standard for the representation, storage, and exchange of 3D city models [22]. It models the geometry, semantics, and graphical appearance associated with 3D city models. The data model of CityGML comprises of a core module and several thematic modules such as Building, Relief, Bridge, Transportation, Vegetation, and WaterBody. Further objects which are not explicitly modelled in the thematic model of CityGML such as pipes and road noise barriers can be stored by extending the data model using either Generic city objects/attributes or ADEs (Application Domain Extensions). ADEs extend the data model of CityGML and come in the form of additional XML schemas. ADEs are used to create application specific extensions such as the Noise ADE for noise mapping [22], the Energy ADE for energy modelling [20], and the iTINs ADE for handling massive terrains [16, 17]; see Biljecki et al. [5] and CityGML[dot]org [6] for an overview of existing ADEs.

CityGML has very limited support for metadata and of the limited number of elements that are supported, i.e. name, description, bounding box, and coordinate system, they are not stored explicitly as metadata, thus making the integration of CityGML within current resource discovery databases a difficulty. In practice, such as in the example

below [22], the elements are normally present near the top of the file, almost directly after the <CityModel> tag, or sometimes at the very end before the closing of the <CityModel> tag.

```
<CityModel>
  <gml:name>Simple 3D city model LOD0
    without Appearance</gml:name>
  <gml:boundedBy>
    <gml:Envelope srsDimension="3"
      srsName=
        "urn:ogc:def:crs,crs:EPSG::25832,
        crs:EPSG::5783">
      <gml:lowerCorner>458868.0
        5438343.0
        112.0</gml:lowerCorner>
      <gml:upperCorner>458892.0
        5438362.0
        117.0</gml:upperCorner>
    </gml:Envelope>
  </gml:boundedBy>
  [...]
</CityModel>
```

CityGML inherits the metadata property (which can be information about the author/creator of the dataset, lineage of the dataset, reference system, etc.) from GML but this only hosts very basic attributes and is not implemented in practice. For example, the <gml:metaDataProperty> tag can be utilised to define the specifications necessary in the usage of a local coordinate system (Appendix G.9 in Open Geospatial Consortium [22]). As an alternative approach, some CityGML users tend to write ad hoc metadata elements as comments in XML, such as the following snippet from a 3D city model of Montreal⁵:

```
<!-- File Written With RhinoCity Software CopyRight
RhinoTerraain 2012 -->
```

Result #1: Our 3D metadata ADE
Metadata elements for 3D city models

Our methodology for including the metadata elements required for the discovery of 3D city models included:

- A an analysis of ISO 19115 for appropriate elements
- B a literature review of 3D city models and their applications to determine necessary elements not currently supported by ISO 19115
- C a study of the CityGML schema to understand which elements are missing
- D an analysis of which elements are required based on the hierarchy levels of CityGML, i.e. city model level, thematic module level and feature level
- E we interviewed software developers and asked them what they like to know to help them in dealing with large files

We utilised Table 1 to guide the ISO elements that we have included in our metadata ADE for the 3D city model

Table 2 Our inclusion of ISO 19115 metadata elements for data discovery

Metadata element	Inclusion	Comment
Metadata reference information	Included	-
Resource title	Included	-
Resource reference data	Included	-
Resource identifier	Included	-
Resource point of contact	Included	-
Geographic location	Modified	Coordinate representation is supported in the Extent element and therefore geographic location was restricted to a string representation.
Resource language	Included	-
Resource topic category	Included	-
Spatial resolution	Excluded	This category is supported for rasters in the Relief module but is not applicable at the city model level
Resource type	Included	-
Resource abstract	Included	-
Extent information for the dataset (additional)	Modified	This was renamed to Extent and follows the Extent package in ISO 19115. Modifications include removing <i>Vertical Extent</i> as a separate category and instead making <i>Geographic Extent</i> explicitly 3D. <i>Geographic Extent</i> is given a mandatory obligation and <i>Temporal Extent</i> is optional.
Resource lineage	Included	-
Resource on-line Link	Included	-
Keywords	Included	-
Constraints on resource access and use	Included	-
Metadata date stamp	Included	-
Metadata point of contact	Included	-

level, and we maintained the same obligation level (i.e. mandatory, conditional or optional). Elements that were not relevant to CityGML were excluded and certain elements were modified to be explicitly 3D. Our inclusions, modifications and exclusions are summarised in Table 2.

The decision to use the discovery table in particular and not the full suite of categories was determined based on several reasons:

- A The size of 3D city models means that their discovery tends to be one of the largest deterrents preventing their usage, the categories defined in Table 1 are already designed to combat this.
- B Many of the attributes listed in Table 1 can already be easily derived from existing models as will be described in “[Result #2: automation and discovery](#)” section, where we automate the process for CityGML files
- C Duval et al. [9] argues that metadata architects should utilise base schemas to facilitate the interoperability of various metadata systems thus ensuring higher usability with various applications, while still allowing for extensions based on domain-specific needs. Table 1 has most of the elements that are likely to be found in many metadata schemas, particularly due to the elements originating in the previous notion of the metadata core model in earlier versions of ISO 19115. Table 1 is therefore suited to act as the base schema. Base schemas can also promote the integration of 3D city models into larger 3D geospatial catalogues, where there are already difficulties in standardising base terminology without introducing the full suite of values [26].
- D The exclusion of a given category or element does not prevent their usage within our ADE as it is modularly designed for future possible extensions, as will be described in “[Extendability](#)” section.
- E Populating databases with metadata can be costly in both time and computation, and therefore there are strong incentives to create metadata with sufficient detail to meet the functional requirements of a domain to encourage their usage [9].

Building on the elements identified in the literature review, particularly in Dietze et al. [8], as well as assessing the CityGML schemas, we identified several elements that needed to be added to the ISO 19115 elements. These are summarised in Table 3. Note that the need to store the datasets used in acquisition and reconstruction as well as the model generation method [8, 26] is supported by the ISO element `Lineage` as described in the next section.

Not all of Dietze et al. [8]’s categories were incorporated as unique elements because some of the proposals were too highly focused on buildings exclusively, such as the categories in relation to ‘process of roof modelling method’ and ‘process of building height’. Such categories are better suited under the `Lineage` tag which can be further extended to individual features with the utilisation

Table 3 Additional metadata elements for 3D city models we include in our ADE

Metadata element	Description	Source
Levels of Detail (LoDs)	This includes the LoDs present in the city model and each thematic module with unique and aggregate counts for each. We support the improved specifications of buildings as developed by Biljecki et al. [3].	Biljecki et al. [3]; Dietze et al. [8]
Semantic Surfaces	The presence or absence of semantic surfaces in objects, e.g. roofs, walls, etc.	Dietze et al. [8]
Textures/Materials	The presence or absence of textures and/or materials in a city model which are representation of object surface characteristics	
XLinks	The presence or absence of <i>XLinks</i> in a city model, these are used to share geometry elements between features.	
External References	The presence or absence of external references, these are a reference of a 3D object to its corresponding object in an external data set	
Thematic Modules	A list of all thematic modules present in a city model, e.g. <code>Building</code> , <code>Transportation</code> , etc.	
ADEs	A list of the ADEs utilised in the city model and their corresponding metadata as described in “ Implementation ” section	

of *XLinks* between the two. Further categories that can take advantage of the `Lineage` element include geometric references (vertical or horizontal), which Biljecki et al. [4] argues have an important role in influencing a spatial analysis.

Most 3D city models are generated with a mixture of different methodologies and data sources and therefore top level metadata is not sufficient but rather there is a need for feature level metadata [8]. For CityGML this means metadata at the city model level, the thematic module level and the feature level. Note that there is often confusion between feature level metadata and feature attributes, attributes contain information about the *feature* while metadata contains information about the *feature data*. Attributes are already supported by CityGML: each feature has the attributes *class*, *function*, and *usage*. While feature-level metadata is currently not supported and is necessary for instances of `lineage`. The following section describes all of the above metadata values in greater detail and explains the hierarchy level at which it is implemented.

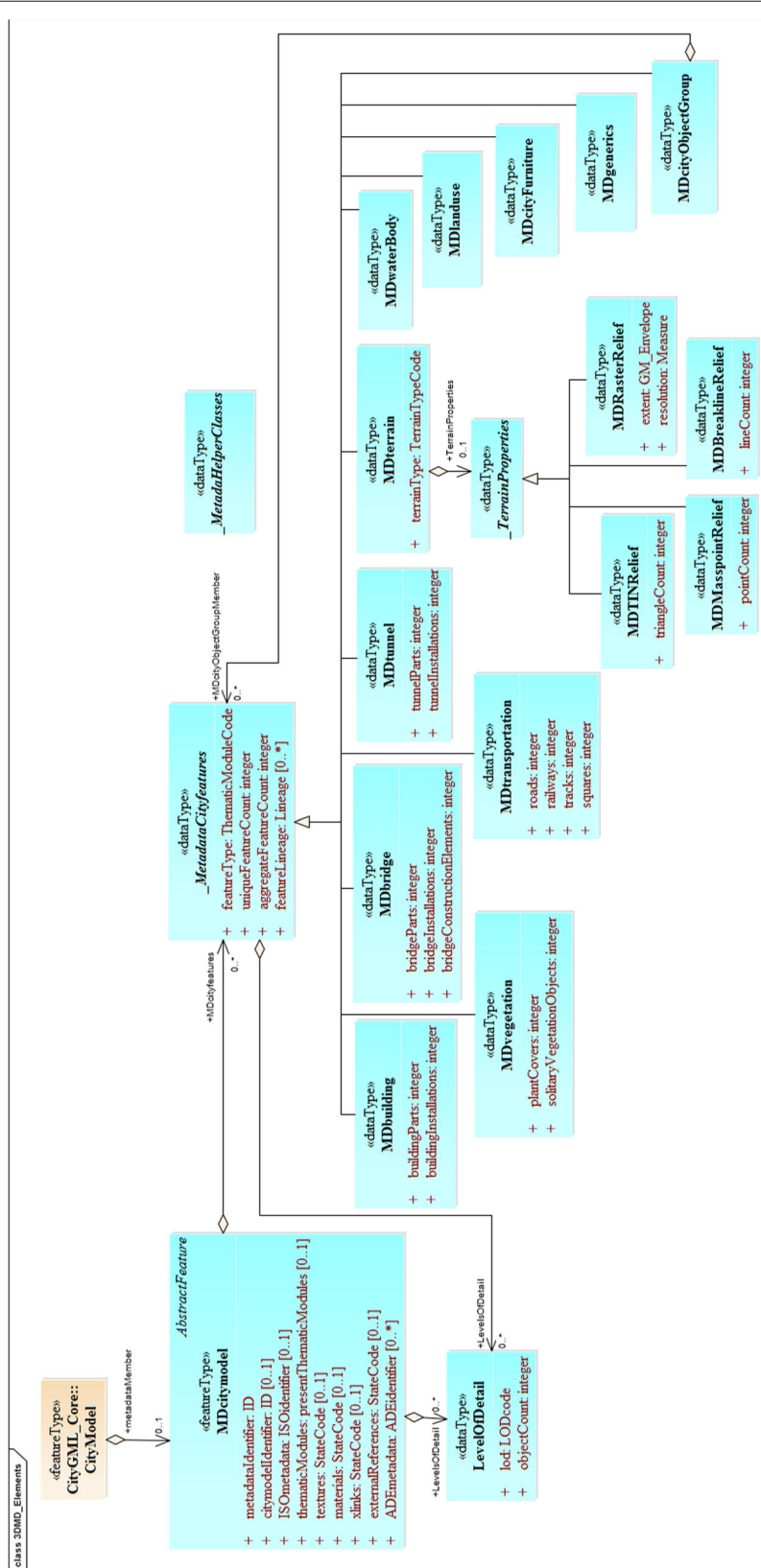


Fig. 1 UML model of the CityGML 3D Metadata ADE depicting the proposed metadata objects for storing metadata related to 3D city models (`_MetadataObjects`) (shown coloured in blue)

Implementation

Our Metadata ADE is realised as a UML model (Fig. 1). The objective of developing this Metadata ADE is to store and manage the metadata associated with a 3D city model in the CityGML format. All resources (UML, XSD, and documentation) related to the Metadata ADE are available on our public GitHub repository: https://github.com/tudelft3d/3D_Metadata_ADE

To avoid any conflict with the existing CityGML classes and attributes, the new 3D metadata classes and attributes are defined in a different namespace with identifier 'md'. Our metadata ADE has three main classes: (i) `MDcitymodel` (ii) `_MetadataCityfeatures` and (iii) `_MetadataHelperClasses`.

`MDcitymodel`. `MDcitymodel` is the core class of this ADE which stores the metadata about a CityGML dataset. It includes the following attributes (see snippet below for detailed attributes and their values):

- ID of the metadata dataset (< `md:metadataIdentifier`>)
- ID of the city model (< `md:citymodelIdentifier`>)
- ISO 19115 metadata elements as mentioned in Table 1 and 2 (< `md:ISOMetadata`>)
- metadata about the ADEs present in the dataset (< `md:ADEmetadata`>)
- metadata indicating which thematic modules are present in the dataset (< `md:thematicModules`>)
- metadata indicating if textures are present in the dataset (< `md:textures`>)
- metadata indicating if materials are present in the dataset (< `md:materials`>)
- metadata indicating if XLinks are present in the dataset (< `md:xLinks`>)
- metadata indicating if external references are present in the dataset (< `md:externalReferences`>)
- metadata about the city features present in the dataset e.g. total number of buildings, building parts, levels of details, etc. (< `md:MDcityfeatures`>)
- levels of detail present in the dataset e.g. (< `md:LevelsOfDetail`>)

`_MDcityfeature` is an abstract class which defines the metadata classes for different CityGML thematic modules. It defines the attributes information about different city features present in the dataset such as:

- the type of city feature (Building, Vegetation, etc.) (< `md:featureType`>)
- total number of a specific type of city feature (< `md:uniqueFeatureCount`>)

- total number of a specific type of city feature if it exists in more than 1 level of detail (< `md:aggregateFeatureCount`>)
- levels of detail of that specific city feature (< `md:LevelsOfDetail`>)
- lineage of that specific city feature (< `md:featureLineage`>)

Apart from the aforementioned metadata elements, `_MDcityfeature` has specialised subclasses (e.g. `md:MDbuilding`, `md:MDbridge`, etc.).

- `md:MDbuilding`. `md:MDbuilding` defines attributes to store metadata about the buildings present in the dataset such as:
 - number of building parts (< `md:buildingParts`>)
 - number of building installations (< `md:buildingInstallations`>)
- `md:MDbridge`. `md:MDbridge` defines attributes to store metadata about the bridges present in the dataset such as:
 - number of bridge parts (< `md:bridgeParts`>)
 - number of bridge installations (< `md:bridgeInstallations`>)
 - number of bridge construction elements (< `md:bridgeConstructionElements`>)
- `md:MDtunnel`. `md:MDtunnel` defines attributes to store metadata about tunnels present in the dataset such as:
 - number of tunnel parts (< `md:tunnelParts`>)
 - number of tunnel installations (< `md:tunnelInstallations`>)
- `md:MDtransportation`. `md:MDtransportation` defines attributes to store metadata about transportation features such as:
 - number of roads (< `md:roads`>)
 - number of railways (< `md:railways`>)
 - number of tracks (< `md:tracks`>)
 - number of squares (< `md:squares`>)
- `md:MDvegetation`. `md:MDvegetation` defines attributes to store metadata about vegetation present in the dataset such as:
 - number of plant covers (< `md:plantCovers`>)
 - number of solitary vegetation objects (< `md:solitaryVegetationObjects`>)

```
<md:MDcitymodel>
  <md:metadataIdentifier>MD_38b566d8-7ea3-11e8-9cde-a0999b076edf</md:metadataIdentifier>
  <md:citymodelIdentifier>GML_38bb4326-7ea3-11e8-8dfd-a0999b076edf</md:citymodelIdentifier>
  <md:ISOmetadata>
    <md:ISOidentifier>
      <md:datasetTitle>3D Building Model</md:datasetTitle>
      <md:datasetReferenceDate>2018-08-23</md:datasetReferenceDate>
      .....
    </md:ISOidentifier>
  </md:ISOmetadata>
  <md:thematicModules>
    <md:presentThematicModules> Building </md:presentThematicModules>
  </md:thematicModules>
  <md:textures>absent</md:textures>
  <md:materials>absent</md:materials>
  <md:xLinks>present</md:xLinks>
  <md:externalReferences>absent</md:externalReferences>
  <md:ADEmetadata>
    <md:ADEidentifier>
      <md:adeName>CityGML iTINs ADE</md:adeName>
      <md:adeVersion>0.1</md:adeVersion>
      <md:namespace>http://godzilla.bk.tudelft.nl/schemas/CityGML_iTINs_ADE</md:namespace>
      <md:status>Implemented</md:status>
      <md:authority>
        <md:OrganizationalContact>
          <md:contactName>3D Geoinformation Group</md:contactName>
          <md:phone>+31 666666666</md:phone>
          <md:address>Julianalaan 134, 2628 BL, TU Delft</md:address>
          <md:emailAddress>3dtud@tudelft.nl</md:emailAddress>
          <md:website>https://3d.bk.tudelft.nl</md:website>
        </md:OrganizationalContact>
      </md:authority>
      <md:summary>CityGML ADE for storing massive TIN terrains</md:summary>
      <md:xmlSchema>https://github.com/tudelft3d/CityGML_iTINs_ADE/
        blob/master/XSD/CityGML_iTINs_ADE.xsd</md:xmlSchema>
      <md:umlModel>https://github.com/tudelft3d/CityGML_iTINs_ADE/
        blob/master/UML/CityGML_iTINs_ADE_0_1.eap</md:umlModel>
      <md:documentation>https://github.com/tudelft3d/CityGML_iTINs_ADE/
        tree/master/Documentation</md:documentation>
    </md:ADEidentifier>
  </md:ADEmetadata>
  <md:MDcityfeatures>
    <md:MDbuilding>
      <md:featureType>Building</md:featureType>
      <md:uniqueFeatureCount>2</md:uniqueFeatureCount>
      <md:aggregateFeatureCount>2</md:aggregateFeatureCount>
      <md:LevelsOfDetail>
        <md:LevelOfDetail>
          <md:lod>2</md:lod>
          <md:objectCount>2</md:objectCount>
        </md:LevelOfDetail>
      </md:LevelsOfDetail>
      <md:buildingParts>2</md:buildingParts>
      <md:buildingInstallations>0</md:buildingInstallations>
    </md:MDbuilding>
  </md:MDcityfeatures>
  <md:LevelsOfDetail>
    <md:LevelOfDetail>
      <md:lod>2</md:lod>
      <md:objectCount>2</md:objectCount>
    </md:LevelOfDetail>
  </md:LevelsOfDetail>
</md:MDcitymodel>
```

- `md:MDterrain`. `MDterrain` defines attributes to store metadata about the terrain model present in the dataset such as:
 - the type of terrain representation (TIN, raster, etc.) (`<md:terrainType>`)
 - levels of detail of the terrain (`<md:LevelsOfDetail>`)

Depending on the type of terrain, it is possible to store additional properties (`<md:TerrainProperties>`) such as the number of triangles (`<md:triangleCount>`) in the case of a TIN or the spatial resolution (`<md:resolution>`) in the case of a raster.

- Similarly, it has `md:MDwaterBody`, `md:MDlanduse`, `md:MDcityFurniture`, `md:MDcityObjectGroup`, and `md:MDgenerics` to store metadata about water bodies, landuse, city furniture, city object groups, and generic city objects/attributes.

We have created a sample dataset to implement the proposed 3D Metadata ADE. The following XML snippet is taken from that sample data⁶.

```

<md:MDcityfeatures>
  <md:MDbuilding>
    .... (same as the aforementioned
         snippet)
  </md:MDbuilding>
</md:MDcityfeatures>
<md:MDcityfeatures>
  <md:MDterrain>
    <md:featureType>Relief</md:featureType>
    <md:featureCount>1</md:featureCount>
    <md:LevelsOfDetail>
      <md:LevelOfDetail>
        <md:lod>3</md:lod>
        <md:objectCount>1</md:objectCount>
      </md:LevelOfDetail>
    </md:LevelsOfDetail>
    <md:terrainType>TINRelief</md:terrainType>
    <md:TerrainProperties>
      <md:MDTINRelief>
        <md:triangleCount>
          22022
        </md:triangleCount>
      </md:MDTINRelief>
    </md:TerrainProperties>
  </md:MDterrain>
</md:MDcityfeatures>

```

`_MetadataHelperClasses`. It defines the supporting classes and attributes required by `MDcitymodel` (see Fig. 2). It includes:

- `ISOidentifier`. It defines metadata elements for a city model according to ISO 19115 Table F.1: *Metadata for the discovery of geographic datasets and series* explained in Tables 1 and 2.

- `ADEidentifier`. It defines attributes to store metadata about the ADEs present in the dataset such as: the name of the ADE and its version, URI of the UML and XML schema and any other available documentation.
 - name of the ADE (`<md:adeName>`)
 - version of the ADE (`<md:adeVersion>`)
 - namespace of the ADE (`<md:namespace>`)
 - status of the ADE (`<md:status>`)
 - authority responsible for the ADE (`<md:authority>`)
 - short summary about the ADE (`<md:summary>`)
 - link to its XML schema (`<md:xmlSchema>`)
 - link to its UML model (`<md:umlModel>`)
 - link to any additional documentation (`<md:documentation>`)

- `LevelOfDetail`. It stores which LoDs are present in a city model (`<md:lod>`) and their count (`<md:objectCount>`). The LoDs are defined in a separate enumeration list (`LODCode`).
- `_Contact`. It stores information (such as name, address, role, etc.) about the person (`<md:IndividualContact>`) or organization (`<md:OrganizationalContact>`) responsible for the dataset.
- `Lineage`. It is possible to store two things with `<md:Lineage>`: (1) metadata about the data sources (`<md:source>`) and production steps (`processStep`) used in the generation of the whole dataset, (2) metadata about individual city features e.g. if a dataset has two buildings A & B created by different organizations/authorities using different methods and data (see snippet below).

Codelists & Enumerations. We defined 5 codelists (taken from ISO 19115) and 4 enumerations (Table 4, Fig. 3). These codelists are implemented as simple dictionaries according to CityGML specifications and can be further extended.

Extendability

Our Metadata ADE is modularly designed for future possible extensions to store metadata related to other domains and applications. It can be extended by other ADEs to incorporate domain-specific data needs. Figure 4 presents such an example for the existing CityGML

```
<md:Building>
  ....
  <md:featureLineage>
    <md:Lineage>
      <md:featureID> A_11 </md:featureID>
      <md:source>
        <md:LiSource>
          <md:description> AHN3 Point cloud </md:description>
          <md:sourceSpatialResolution> 10 points/m2
            </md:sourceSpatialResolution>
          <md:sourceCitation> www.pdok.nl/ahn3-downloads
            </md:sourceCitation>
        </md:LiSource>
      </md:source>
      <md:processStep>
        <md:LiProcessStep>
          <md:description> Point cloud processing using Lastools
            </md:description>
          <md:stepDateTime>2017-02-01</md:stepDateTime>
          <md:processor>
            <md:OrganizationalContact>
              <md:contactName>3D Geoinformation Group</md:contactName>
              <md:phone>+31 666666666</md:phone>
              <md:address>Julianalaan 134, 2628 BL, TU Delft</md:address>
              <md:emailAddress>3dtud@tudelft.nl</md:emailAddress>
              <md:website>https://3d.bk.tudelft.nl</md:website>
            </md:OrganizationalContact>
          </md:processor>
        </md:LiProcessStep>
      </md:processStep>
    </md:Lineage>
  </md:featureLineage>
  <md:featureLineage>
    <md:Lineage>
      <md:featureID> B_11 </md:featureID>
      <md:source>
        <md:LiSource>
          <md:description> BGT 2D </md:description>
          <md:sourceSpatialResolution> 30m2 </md:sourceSpatialResolution>
          <md:sourceCitation> www.pdok.nl/bgt-download </md:sourceCitation>
        </md:LiSource>
      </md:source>
      <md:processStep>
        <md:LiProcessStep>
          <md:description> Extrusion using 3Dfier </md:description>
          <md:stepDateTime>2015-04-01</md:stepDateTime>
          <md:processor>
            <md:OrganizationalContact>
              <md:contactName>Geonovum</md:contactName>
              <md:phone>+31 455678999</md:phone>
              <md:address>Geonovum, Amersfoort</md:address>
              <md:emailAddress>xyz@geonovum.nl</md:emailAddress>
              <md:website>https://geonovum.nl</md:website>
            </md:OrganizationalContact>
          </md:processor>
        </md:LiProcessStep>
      </md:processStep>
    </md:Lineage>
  </md:featureLineage>
  ....
</md:Building>
```

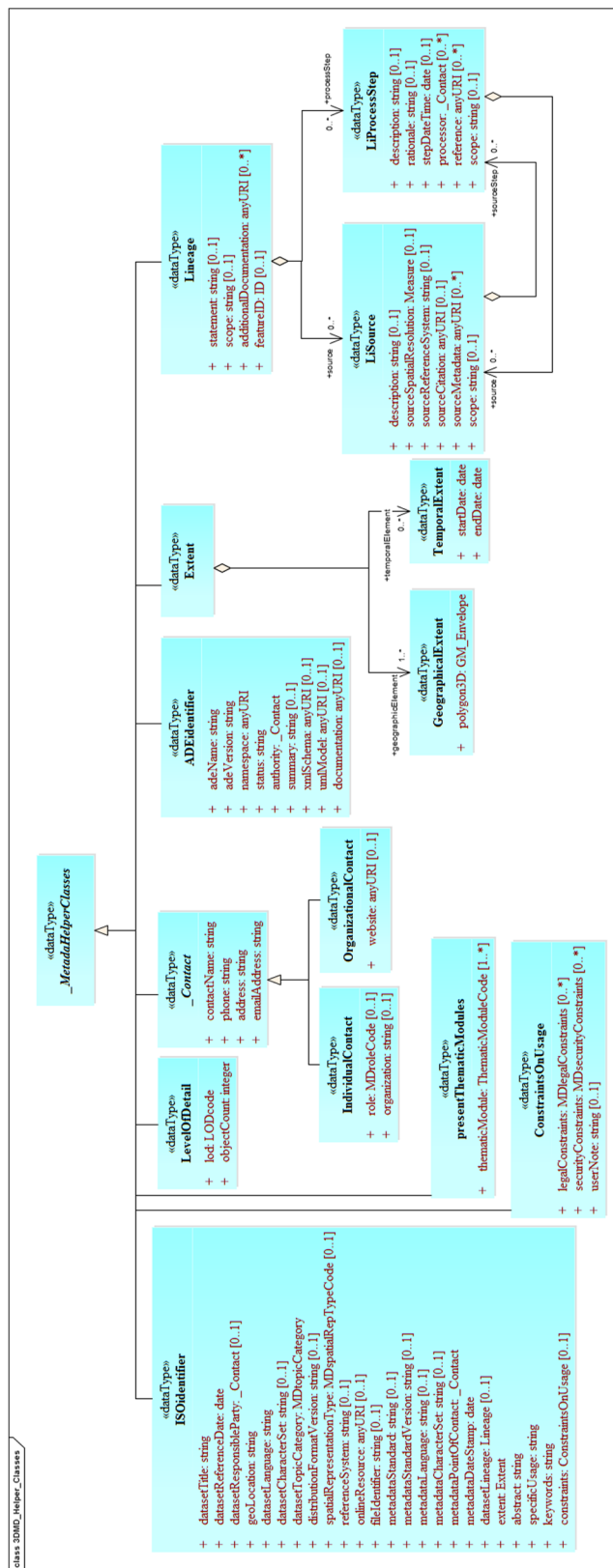


Fig. 2 UML model of the CityGML 3D Metadata ADE depicting the proposed supporting classes (*_MetadataHelperClasses*) (shown coloured in blue)

Table 4 Description of codelists proposed for the 3D Metadata ADE

Codelist	Description
MDtopicCategory	ISO 19115 codelist of themes (such as environment, atmosphere, climatology) for classification of datasets.
MDroleCode	ISO 19115 codelist of the functions performed by the person responsible for the dataset.
MDlegalConstraints	ISO 19115 codelist of restrictions and legal prerequisites for accessing and using the dataset or metadata.
MDsecurityConstraints	ISO 19115 codelist of the restrictions imposed on the data or metadata for national security or similar security concerns.
MDspatialRepType	ISO 19115 codelist of the methods (such as raster or vector) used to represent the geoinformation present in the dataset.
Enumeration	Description
ThematicModelCode	Enumeration of different thematic models present in CityGML such as Building, Vegetation, etc.
TerrainTypeCode	Enumeration of different terrain types present in CityGML such as TINRelief, RasterRelief, etc.
LODcode	Enumeration of the CityGML LoDs (0-4). We also included the LoDs proposed by Biljecki et al. [3]
stateCode	Enumeration with values to identify if a feature is present or absent.

Noise ADE. A new class `MDnoiseBuilding` is created by extending the Metadata ADE `MDbuilding` class. `MDnoiseBuilding` has a new attribute to store information such as the number of buildings enriched with Noise ADE attributes (`numberOfNoiseBuildings`). Similarly, `MDnoiseRoads` is a subclass of `MDtransportation` with a new attribute to store the number of noise road segments present in the data (`numberOfNoiseSegments`).

Result #2: automation and discovery

Automatic metadata generation

Metadata generation, to populate the metadata extension with data, does not need to be a painful task and while the seemingly time-consuming and uninteresting nature of the topic is often seen as a deterrent [10], it is more often hindered by a lack of definition in metadata for 3D city model data in particular. Many of the values for the categories discussed in this paper can be easily accessed during the city model generation process and therefore having a solidly defined metadata ADE is advantageous to guide data creators.

To further ensure the usability of our work we offer a Python software⁷ that derives categories such as the levels of detail present, thematic models, extent, etc. It parses

a dataset looking to see if other metadata information is present and has default values to indicate which values are missing. Due to the large file size of most CityGML files, we chose to generate the metadata as a separate file which ensures faster parsing but users can write to the original file if they wish to.

Database population and discovery

In order to ease the process of resource discovery we have developed an online database⁸ that contains all of the aforementioned metadata for 3D CityGML datasets. The database is open access for both viewing and contributing and can be queried with SQL. The database is composed of seven tables that are summarised in Table 5. In cases where the tables have a one-to-one relationship, i.e. only one unique value exists for the 3D city model (e.g. dataset abstract), then the city model identifier is the *primary* key for the table, in one-to-many relationships it is the *foreign* key (e.g. lineage).

Populating the database is an easy step after the automation script has been executed because the automation script already ensures that the output is conformant to the definition of our Metadata ADE. The output file can be uploaded and the database is updated. A database also helps in issues where metadata is decoupled from the 3D city model dataset by establishing a link through the unique ids of both. It also addresses a major barrier to dataset discovery which is a lack of user-friendly interfaces which encourage usage [21].

Conclusions

In this paper we have addressed the lack of metadata in CityGML and specifically we have focused on the discovery of datasets which are often too large to be parsed easily. We have proposed an ADE that is both ISO 19115 compliant and incorporates further elements as required by users of 3D city models. Given the open nature of CityGML and its collaborative evolution process, this ADE could serve as the model for the next version of CityGML, version 3.0. We have modelled our ADE to reuse CityGML elements as much as possible to realise an easy transition from ADE to a core module of the CityGML standard.

Furthermore, as is argued by many metadata practitioners, including Ellul et al. [10] and Olfat et al. [21], creating metadata after dataset generation requires a considerable amount of effort and the availability of information may be reduced which leads to incomplete metadata. They argue that metadata generation should be a process that is run parallel to data generation. Having a well-defined metadata ADE can aid data-creators by providing neat guidelines to follow.

Future work will deal with examining if there are further specifications necessary for individual thematic

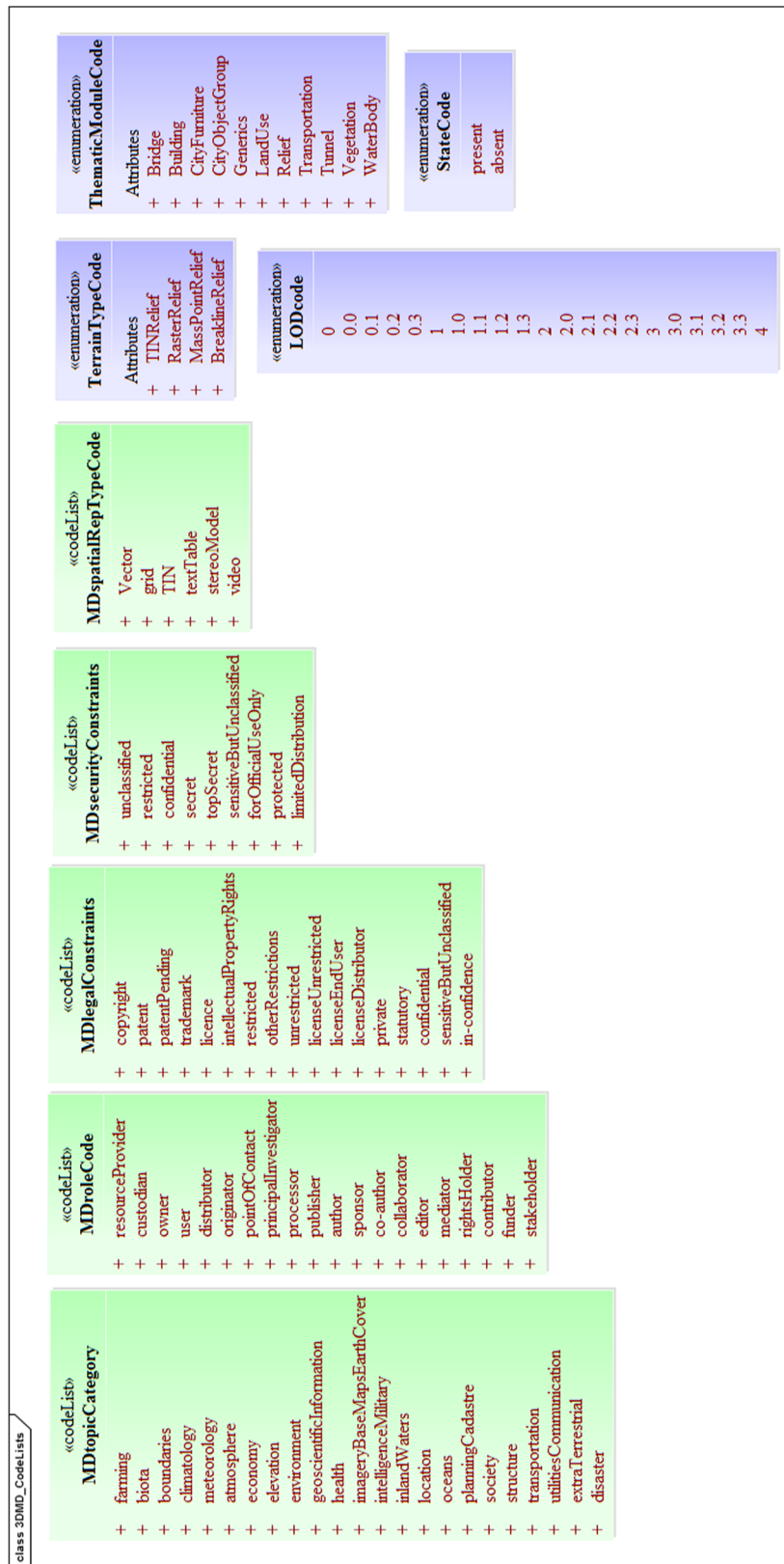


Fig. 3 Implemented ISO codelists (shown coloured in green) and proposed new codelists (shown coloured in purple) in the CityGML 3D Metadata ADE

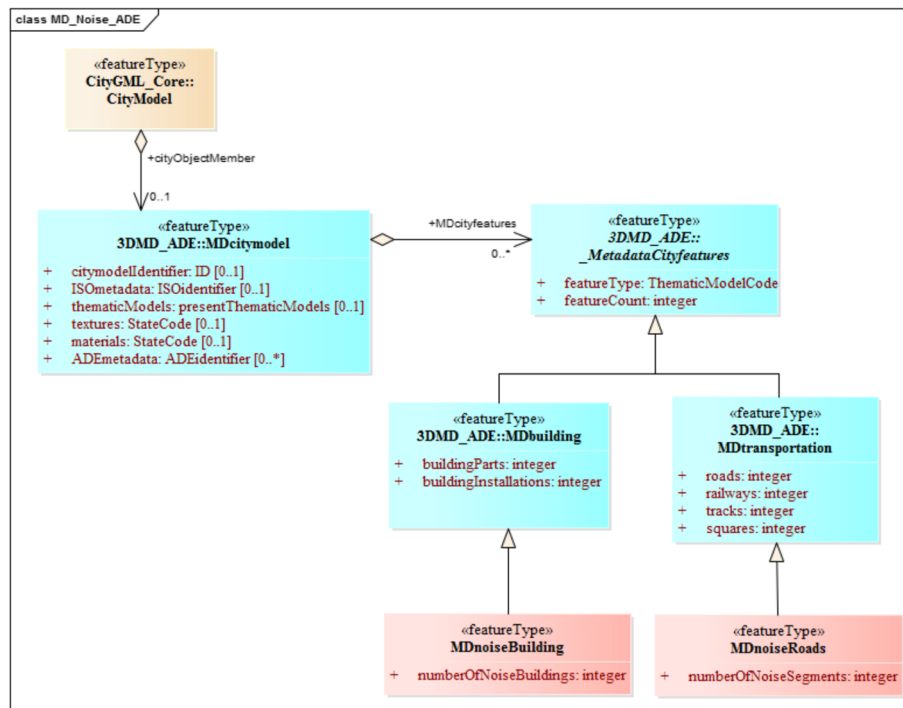


Fig. 4 Excerpt of the UML diagram of the 3D Metadata ADE, depicting how to extend it to incorporate the metadata related to other ADEs (such as Noise ADE)

Table 5 Description of tables present in the database for the discovery of 3D city models as related to the 3D Metadata ADE

Table	Description	Unique city model identifier
Master Table	Containing the unique city model identifier, the unique metadata identifier, the file names of both the city model and metadata, and the presence or absence of textures and materials	Primary key
ISO Categories	All of the categories as listed in ISO 19115-1:2014 - Table F.1	Primary key
Thematic Models	Presence or absence of each thematic model and the count of those features	Primary key
Thematic Model LoDs	The count by subClass and LoD for city features within a thematic model	Foreign key
Lineage	The source and process steps that track the lineage of a 3D city model	Foreign key
ADEs	Containing all of the categories associated with the ADEIdentifier _MetadataHelperClass for every ADE present	Foreign key
Terrain	Containing counts by terrain representation types	Foreign key

modules that can be investigated and added as extensions, this is also true for various ADEs. Given the easy nature of extensibility we can easily implement various extensions, and we have already begun working with the Energy ADE working group to determine how they can extend our ADE for their domain-specific metadata.

The ISO 19115 compliant categories and elements that were defined for the ADE can also easily be used to extend the metadata capabilities of CityJSON⁹, which is a format that encodes the CityGML data model using JavaScript Object Notation (JSON). CityJSON offers an alternative to the GML encoding of CityGML, in which objects can be defined in a very large number of possible ways, and which can therefore be verbose and complex (and thus rather frustrating to work with). Since it is based on JSON, anyone can add any attributes to a given city objects, and the categories and elements can simply be added to the "metadata" entry of a CityJSON file. Our metadata categories and elements are available in the core of the recent version of CityJSON, version 0.8.

Availability and requirements

Project name: 3D Metadata ADE

Project home page: https://github.com/tudelft3d/3D_Metadata_ADE

Documentation:

https://github.com/tudelft3d/3D_Metadata_ADE/tree/master/Documentation/BrowsableSchema

Operating system(s): Platform independent

Programming language: Python

Other requirements: None

License: MIT License

Endnotes

¹ <http://www.businesslocationcenter.de/en/downloadportal>

² <http://www.cityjson.org>

³ <http://www.fgdc.gov/metadata/csdgm>

⁴ <http://www.inspire.ec.europa.eu/metadata>

⁵ <http://donnees.ville.montreal.qc.ca/dataset/maquette-numerique-batiments-citygml-lod2-avec-textures>

⁶ https://github.com/tudelft3d/3D_Metadata_ADE/tree/master/Code/citygmldatasets

⁷ https://github.com/tudelft3d/3D_Metadata_ADE/tree/master/Code

⁸ https://github.com/tudelft3d/3D_Metadata_ADE

⁹ <http://www.cityjson.org>

Acknowledgements

We gratefully acknowledge the helpful feedback of the reviewers.

Funding

The research leading to this paper has received funding from the European Research Council under the European Unions Horizon2020 ERC Agreement no. 677312 UMnD: Urban modelling in higher dimensions. This work is also a part of the research project 3D4EM (3D for Environmental Modelling) in the Maps4Society programme (grant No. 13740) which is funded by the NWO (Netherlands Organisation for Scientific Research), and partly funded by the Ministry of Economic Affairs.

Authors' contributions

AL conceived the paper, and carried out the literature review together with KK. AL and KK wrote the paper in consultation with HL and JS. KK designed the schema and UML for the ADE and the automation script. AL designed the database for discovery. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 11 October 2018 Accepted: 2 November 2018

Published online: 30 November 2018

References

1. Aktas MS, Pierce M, Fox GC, Leake D. A web based conversational case-based recommender system for ontology aided metadata discovery. In: Proceedings of the 5th IEEE/ACM international workshop on grid computing. Pittsburgh: IEEE Computer Society; 2004. p. 69–75.
2. Basanow J, Neis P, Neubauer S, Schilling A, Zipf A. Towards 3D spatial data infrastructures (3D-SDI) based on open standards—experiences, results and future issues. In: Advances in 3D geoinformation systems. Berlin: Springer; 2008. p. 65–86.
3. Biljecki F, Ledoux H, Stoter J. An improved LOD specification for 3D building models. *Computers, Environment and Urban Systems*. 2016a;59: 25–37. <http://doi.org/10.1016/j.compenvurbsys.2016.04.005>.
4. Biljecki F, Ledoux H, Stoter J, Vosselman G. The variants of an LOD of a 3D building model and their influence on spatial analyses. *ISPRS J Photogramm Remote Sens*. 2016b;116:42–54. <http://doi.org/10.1016/j.isprsjprs.2016.03.003>.
5. Biljecki F, Kumar K, Nagel C. CityGML Application Domain Extension (ADE): overview of developments. *Open Geospatial Data, Software and Standards*. 2018. <https://doi.org/10.1186/s40965-018-0055-6>.
6. CityGML[dot]org. Application Domain Extensions (ADEs). 2018. <https://www.citygml.org/ade/>. Accessed 7 May 2018.
7. Danko DM. Geospatial metadata. In: Kresse W, Danko DM, editors. *Springer Handbook of Geographic Information*. Berlin: Springer Science & Business Media; 2010. p. 359–91.
8. Dietze L, Nonn U, Zipf A. Metadata for 3D city models analysis of the applicability of the ISO 19115 standard and possibilities for further amendments. In: Proceedings of the 10th AGILE International Conference on Geographic Information Science. Aalborg: AGILE; 2007. p. 1–9.
9. Duval E, Hodgins W, Sutton S, Weibel SL. Metadata principles and practicalities. *D-lib Magazine*. 2002;8(4):1082–9873.
10. Ellul C, Tamash N, Xian F, Stuver J, Rickles P. Using Free and Open Source GIS to Automatically Create Standards-Based Spatial Metadata in Academia. *OSGeo J*. 2014;13(1):51–9.
11. Huang F. CityGML-based 3D GIS visualization in the cloud environment. *Comput Model New Technol*. 2014;18(12):91–8.
12. ISO. Geographic information - Metadata - XML schema implementation (ISO/TS 19139:2007,IDT). 2009.
13. ISO. Geographic information - Metadata - Part 1: Fundamentals (ISO 19115-1:2014). 2014.
14. ISO. Geographic information — Metadata — Part 3: XML schema implementation for fundamental concepts (ISO 19115-3:2016). 2016.
15. Koller D, Frischer B, Humphreys G. Research challenges for digital archives of 3D cultural heritage models. *J Comput Cult Herit (JOCCH)*. 2009;2(3):7.
16. Kumar K, Ledoux H, Stoter J. A CityGML extension for handling very large TINs. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci*. 2016;IV-2/W1: 137–43. <https://doi.org/10.5194/isprs-annals-IV-2-W1-137-2016>, <https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/IV-2-W1/137/2016/>.
17. Kumar K, Ledoux H, Stoter J. Compactly representing massive terrain models as TINs in CityGML. *Trans GIS*. 2018;1–28. <https://doi.org/10.1111/tgis.12456>.
18. Maravelakis E, Konstantaras A, Kritsotaki A, Angelakis D, Xinogalos M. Analysing user needs for a unified 3D metadata recording and exploitation of cultural heritage monuments system. In: *International Symposium on Visual Computing*. Berlin: Springer; 2013. p. 138–47.
19. Neubauer S, Zipf A. Suggestions for extending the OGC styled layer descriptor (SLD) specification into 3D—towards visualization rules for 3D city models. *Appl Sci*. 2007;7:133.
20. Nouvel R, Bahu JM, Kaden R, Kaempf J, Cipriano P, Lauster M, Haeefe KH, Munoz E, Tournaire O, Casper E. Development of the CityGML application domain extension energy for urban energy simulation. In: 14th International Conference of the International Building Performance Simulation Association (IBPSA). Hyderabad: IBPSA; 2015b.
21. Olfat H, Kalantari Soltanieh S, Rajabifard A, Senot H, Williamson IP. Spatial metadata automation: A key to spatially enabling platform. *Int J Spat Data Infrastruct Res*. 2012;7:173–95.
22. Open Geospatial Consortium. OGC City Geography Markup Language (CityGML) Encoding Standard 2.0.0. 2012.
23. Riley J. Understanding metadata - what is metadata, and what is it for? National Information Standards Organization (NISO) Primer Series. 2017.
24. Shi R, Maly F, Zubair M. Automatic metadata discovery from non-cooperative digital libraries. In: Proceedings of the IADIS International Conference on e-Society. Lisbon: IADIS; 2003.
25. Wan Y, Bian F. A Extended Web Feature Service Based Web 3D GIS Architecture. In: *Wireless Communications, Networking and Mobile*

Computing, 2007. WiCom 2007. International Conference on. Shanghai: IEEE; 2007. p. 5947–50.

26. Zamyadi A, Pouliot J, Bédard Y. Towards Precise Metadata-set for Discovering 3D Geospatial Models in Geo-portals. In: 8th 3D Geoinfo Conference. Istanbul: Springer; 2013.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
