OGC Data Exchange Toolkit: Interoperable and Reusable 3D data at the end of the OGC Rainbow

Francesca Noardo[1], Rob Atkinson[1], Ingo Simonis[1], Alejandro Villar[1], and Piotr Zaborowski[1]

Open Geospatial Consortium Europe, Technologielaan 3, B - 3001 Heverlee, Belgium
fnoardo, ratkinson, isimonis, avillar, pzaborowski@ogc.org,
WWW home page: https://www.ogc.org

Abstract. Standard data models are powerful to enable data interoperability. However, due to their ambition to comprehensively represent a domain, they tend to be over-specified. On the other hand, they allow too much flexibility in implementation and modelling details, to adapt to several use cases needs. The OGC Data Exchange Toolkit intends to overcome the two issues by means of a methodology to define profiles through linked data technologies, operated via the OGC Rainbow. Moreover, attributes are defined to specify the desired object representation requirements, solving the under-specification issue. A template is provided, intended to support manual input of data requirements, generating in turn a human-readable description and a machine-readable input for the profiling tool itself. The profile allows automatic datasets validation against the defined requirements. The prototype is developed and tested on a sample CityJSON dataset.

Keywords: standard data models, data model profile, data requirements, data validation, ontologies, OGC Rainbow

1 Introduction

FAIR principles (Findable, Accessible, Interoperable, Reusable) (Wilkinson et al, 2016) are essential to support an optimised and efficient production and use of data in distributed, multi-user environments. Distributed ecosystems such as *Data Spaces*, S*patial Data Infrastructures*, *Digital Twins*, etc. are increasingly being conceptualised and realised to support reuse of data and system capabilities to deliver on policy and operational objectives. The rapid emergence of 3D, 4D and AI capabilities has resulted in more forms, and larger sets, of data with different structural characteristics. Hence the need to utilise very large quantities of data across many dimensions, with varying structures, has resulted in the renewed recognition attention on semantic interoperability across heterogeneous systems, rather than trying to enforce structural similarity.

Open standards are essential to facilitate this. Standard data models are intended to describe domains' information through agreed schemata in order to

produce consistent and unambiguous data. However, open standard data models are either intended to cover a huge number of use cases or are closely tied to a specific purpose. General models are complex to use and yet still unable to deal with the full spectrum of requirements for related use cases.

In practice, comprehensive data models with a fixed schema simultaneously both over-specify and under-specify the information needed to describe a use case's needs. Over-specification occurs when many aspects are represented even though they may not be relevant to a particular application, in an attempt to be a "one-size-fits-all" solution. Under-specification occurs when aspects of the model are oversimplified, or left to different implementation and presentation decisions. Under-specification leads to implementation-specific choices that hinder the reusability and safe interpretation of standards-compliant data. In fact, several datasets may represent a different selection of features with different modelling and filling choices but still remain data model-compliant in the same way. For example, different approaches may choose to use either geometric solids or component surfaces, and have a range of different conventions regarding how such geometric elements are identified and topologically related. Different selection of attributes to characterise objects may be used, with similar or ambiguous naming and rules for content standardisation. Consequently, this makes it difficult for users to understand these datasets, and the use of datasets that conform to a specific data schema does not *per se* guarantees consistent or valid results in every application (Wagner et al, 2012; Noardo, 2022).

For example, many spatial systems describe their own profile of the underlying ISO 19107 Geometry model[1] by using some words in the introduction of a document. For example the GeoJSON specification[2] declares *"GeoJSON supports the following geometry types: Point, LineString, Polygon, MultiPoint, MultiLineString, and MultiPolygon"*. When extrapolating to 3D, 4D geometries, and support for topology in 2D and 3D the variability between specifications becomes greater still, and there will obviously be great difficulty in finding, accessing and interpreting such informal wording to determine the level of compatibility and the transformations possible for data based on these specifications.

Defining a set of formal profiles can capture the intent in a simplified way and support data processing pipelines that can exploit these relationships.

Software developers can implement tools for standard data models (such as importers, exporters, analyzers, etc.) that consider the various options for different functions (Noardo et al, 2021). However, the problem always arises here that any leeway drives up implementation costs. If these cannot be borne, they lead to restrictions in the functional space of the respective tool. Even if these restrictions are clearly documented, the acceptance of such tools in the user groups is limited. Users and practitioners may have reservations about using automated tools, even if they are based on standardization of data sets that conform to the data model, because they find potential inaccuracies that make them not 100% reliable. Often the barrier is simply discovering the relevant documentation to

[1] https://www.iso.org/standard/66175.html
[2] https://geojson.org/

allow interpretation of how an application has used an underlying standard, even if the standard itself is well understood.

Ultimately, these problems affect trust in data and data processing. As a result, the power of open standards is not fully exploited by supporting automation and interoperability.

The solution to this problem lies in the combination of modularity of data models and dynamic, unambiguous definition of specific constraints and assignment of the degrees of freedom available to the model. This allows users to use only the elements necessary for their data and not be burdened with unnecessary complexity by the overall model. Specific applications in the domain can formally profile the overall model, which unambiguously defines the parts used as well as corresponding implementation decisions for a particular community of practice. By dynamically providing such application context-dependent definitions, data providers can provide users with any information necessary for unambiguous processing and interpretation of the data at runtime. Crucially, the provision of this information must be directly processable by both machines and human users.

The direct implication of this hard-won experience with standardisation data models is that we need a more flexible and dynamic ecosystem of standards components, with simplified general conceptual models, multiple schema options, application profiles and the support for the validation and vocabulary aspects - a wide spectrum of resources that are not easily contained in the existing approach of an ever-growing set of heterogeneous document formats distributed across the Web.

The same issue was addressed by the ISO standard 19131.2 on Geographic information - Data Product Specifications (ISO/TC 211, 2022). It provides a data model to support the conceptual formalisation of semantics and data structure related to specific requirements of use cases, for the same reasons of supporting data requirements definition and following datasets' description. Because of this last aim, the direct relation with the ISO 19115 standard on Metadata (ISO/TC 211, 2014) is foreseen in the model.

However, the use of standardised data models to formally describe the data semantics and structure is not foreseen within the ISO 19131 standard, with consequent loss of the potential interoperability advantage given by the use of standard data models.

To overcome this gap, in this study, a prototype of the *OGC Data Exchange Toolkit* is developed starting from a use case, in order to support the formal specification of a standard data model profile, including detailed definition of semantics and geometry of the single contained objects.

A template for data requirements definition was defined starting from a use case (i.e., the data requirements to check a building permit regulation in the municipality of Ascoli Piceno, in Italy) . The data requirements defined for the test refer to the CityJSON v.1.1 standard data model (Ledoux et al, 2019), which was mapped to OWL and identifiers were assigned. These were organised

in a template, useful to support similar operations, as well as facilitating the following development of a compliant standard data model profile (Section 3.1).

After cross-referencing the requirements in the template to the model, a profiling tool was configured to extract the same profile defined in the template into a machine readable (semantic) model and use it for validation purposes (Section 3.3).

A sample data is mapped to an equivalent semantic model, and the validation is performed using standard tools ensuring transparency, openness and trust of the process (Section 3.4).

2 Methodology

A bottom-up approach starting from a use case was adopted to iteratively build and test the necessary profile and supporting tools. Within the Horizon Europe project 'Change toolkit for Digital Building Permit' (CHEK)[3], on the digitalisation of building permits, an initial CityJSON profile was defined, useful to support automatic checks of a city regulation (Figure 1). Diverse information is involved in the digitalisation of building permits, including Building Information Models (BIM) and Geographical Information Systems (GIS). In this paper the focus is on the specification of requirements for the GIS data, as a 3D city model compliant to the CityJSON standard data model.

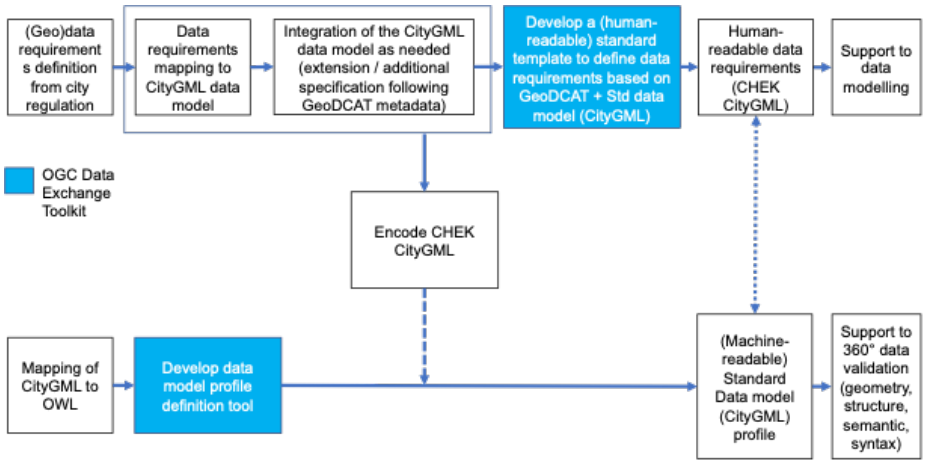


Fig. 1. Methodology workflow

[3] https://chekdbp.eu

2.1 The definition of a template for data requirements specification

Starting from the formalisation of a selected city regulation as a use case, the information requirements were defined and mapped on the chosen standard data model, in this case the OGC Community standard CityJSON[4].

To comprehensively represent the data requirements, a subset of the relevant attributes of metadata schemas, such as ISO 19115 or GeoDCAT (Raes, Lieven and Vandenbroucke, Danny and Reznik, Tomas, 2019) was selected, and integrated with further features relevant to support interoperability and possibly integration with different sources Noardo (2022). The obtained information was compared and properly mapped to the contents of ISO 19131. An extension to the ISO 19131 was defined to allow direct reference to standard data models in the Application Schema Information description.

The result of such mapping was gathered in tables that describe the reference template for the *OGC Data Exchange Toolkit*. In addition, they will be implemented as an interactive tool facilitating the filling of the necessary information in a more user-friendly interface. The tool will facilitate, for example, the mapping of the selected standard data model or related class to their respective URIs. The filled requirements will be later suitably exported in both a human readable data requirements specification, supporting data modelling, and in a more formal format able to support the linked data-based profiling tool.

2.2 The profiling tool development

For the developemnt of the profiling tool, to facilitate the mapping, first an Ontology for CityJSON was developed to provide a unique and unambigous identifier (in the form of a Universal Resource Identifier - URI) for each Class and Property in the model, and to capture relationships. The OWL, RDFS and SHACL languages can then be used to describe each of these within a knowledge graph so that profiles can access underlying concepts when required. These URIs then allow us to define mappings in a way that can be published in a FAIR fashion, using the OGC Registry for Accessible Identifiers of Names and Basic Ontologies for the Web (Rainbow)[5].

The OGC Rainbow is a "full-spectrum" interoperability environment that provides for publication and cross-referencing any type of resource used in creating or using a spatio-temporal dataset. OGC Rainbow enables model-driven standards with continuous integration and testing, data model and building block registries, and many other tools that ultimately enable semantic interoperability, the long-awaited next interoperability level.

The OGC Rainbow supports communities to publish profiles of existing baseline standards.

At this stage the focus is on the machine-readability of such profiles, with an emphasis on provision of validators, and forms of the profile that are directly

[4] https://www.cityjson.org
[5] https://www.ogc.org/resources/rainbow/

validatable. Exploiting semantic technologies to model the standards baseline provides a set of identifier about which statements can be made.

Technologies such as SHACL[6] allow "shapes" to be defined that control how elements of such models may link to related data with its own data models, and rules defined for validation of data against such linked models. The "shape" refers to the pathways that are present in an interconnected graph of related data, and it can be seen that such an approach is naturally suited to the integration of related systems with their own preferred data models.

The profiling toolkit supports cataloguing of such profiles using the Profiles Vocabulary[7] which allows profiles to be incrementally developed on top of more general profiles.

The OGC Rainbow uses such a profile catalogue to manage the heterogeneous forms of resources from many different sources - each is validated against a range of requirements for the specific resource type, but also further augmented and even transformed using specific profiles describing the input data source and the final form that can be integrated into the linked knowledge graph.

With this powerful decomposition of complex specification environments into standardised pieces, specific application challenges can then be supported by a common toolkit. As we shall describe below, this toolkit allows for a best practice "Continuous Integration/Continuous Testing" (CI/CT) approach to defining the specific profiles of standards.

The required subset of CityJSON can then be defined as a formal profile of CityJSON, or a set of related profiles for different levels of detail.

2.3 Validation

We test the template and the standard data model profile in conjunction with each other - to ensure the tool deliver useful forms of profiles, and that the profiles meet the business (use case) objectives. The profile itself is tested using sample data[8], and the testing process is validated through feedback from data producers as well as software developers who benefit from the clarity and flexibility of the tested examples.

3 Results

The prototyped toolkit includes a template to specify data requirements (i.e., complying to the same data model profile) in human-readable format, supporting the production and modelling of data.

Moreover, a standard data model profiling tool allows the definition of the specific part of a standard data model and the specific implementation requirements in a machine-readable format, leveraging the OGC Rainbow semantic

[6] https://www.w3.org/TR/shacl/
[7] https://www.w3.org/TR/dx-prof/
[8] https://www.cityjson.org/tutorials/getting-started/

capabilities. Profiles rely on underlying specifications having unambiguous identifiers for the component elements, which is a natural consequence of using semantic technologies as the underlying expression language, since these are based on links (predicates) between identifiable "subjects" and other objects or literal data values. A generated machine-readable profile can support the automatic validation of datasets as well as software development and configuration to exploit the data. The extensible profile model used can even include specific implementation resources such as code functions if required.

A part of the urban regulation related to the minimum distances of the new building, submitted to the municipality for approval, in the municipality of Ascoli Piceno (art.61), in Italy, was chosen as a starting point. The data requirements for the representation of the context as a 3D city model were considered and extracted following the RASE methodology (Hjelseth and Nisbet, 2011).

The considered text of the regulation is (translated from Italian): *"2 - [...] between windowed walls of facing buildings, a minimum distance is prescribed, equal to the height of the tallest building and no lower than 10 m; if the facing facades overlap for more than 12 meters, the rule applies also when only one of them has windows.[...] In case of setback storeys, the rule applies per each floor. [...] 4 - Minimum distance between buildings with roads in-between, excluding cul-de-sac roads serving single buildings or settlements, must be equal to the road width plus: (1) 5.00 m per side, in case of streets width lower than 7.00 m, (2) 7.50 m per side, in case of streets width between 7.00 and 15.00 m, (3) 10.00 m per side, in case of streets width higher than 15.00 m. "*

From the analysis of the text, the relevant information was defined, and later mapped to the CityJSON v.1.1 data model. Additional requirements, for the overall dataset or the single objects were then defined and mapped, whether possible to the existing standards, in particular to metadata standards such as GeoDCAT or ISO 19115, or to the data product specification standard ISO 19131.

The resulting model, as represented in Figure 2 was on the one hand described in tables defining a data requirement definition template intended to receive users input (Section 3.1), and then converted into RDF format to support the profiling tool developments (Section 3.3)

3.1 Data requirement definition template

A template for the definition of the required information was developed. The goal of such a template is, on the one hand, to facilitate the municipalities officers, or the professionals commissioned by them, to define and publish the data requirements related to the geoinformation, as part of the digitalisation process of regulations and building permit overall. On the other hand, the result of the data requirement definition through the template will be made available to the data modellers and producers as a clear specification of the data to be produced. The format of such export should therefore be easily understandable by the modellers. In case the production of the 3D city model would use auto-

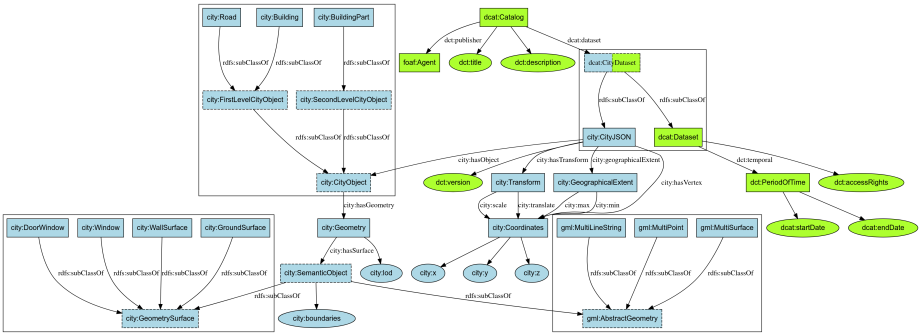


Fig. 2. Profile of CityJSON v.1.1 needed for the prototype test (in blue), integrated with the additional attributes intended to further specify the data, as represented in RDF. Rectanglesare classes are rectangles; the labels for the arcs define the relationship type (object property); abstract classes have dashed outlines; data properties are ellipses.

matic modelling algorithms, a different machine-readable format able to input the useful parameters to such algorithm could be developed in future.

In addition, the same information will be exported in a suitable format to be automatic input for the profiling tool explained in Section 3.3.

Table 1 summarises the information to be defined for the dataset to be modelled, including parameters useful to describe the necessary information and referring to metadata description standards, mostly GeoDCAT[9] and ISO19131[10]. This alignment will facilitate the storage of the same values as initial metadata set for the dataset publication, increasing its findability and re-usability later.

The field *'Access rights / rights holder'* is intended to store the kind of licence of the data. At the moment it is structured according to four criteria, compliant to the Creative Commons licenses model[11], based on which the user should indicate if it is acceptable that data license requires: *"Attribution"* (BY), *"Share alike"* (SA), *"Noncommercial use"* (NC) or *"No Derivative works"* (ND).

The field *'Reference (standard) data model'* is intended to store the reference to the used (and profiled or extended) standard data model and specific version. It is supposed to provide the semantics and the structure for the application schema[12] being defined as a profile of it. The reference to the schema should be an official unique identifier. The defined profile will later specify the required application schema in detail (Table 2).

[9] https://semiceu.github.io/GeoDCAT-AP/releases/2.0.0/
[10] https://www.iso.org/standard/85092.html
[11] https://creativecommons.org/about/cclicenses/
[12] The *'application schema'*, according to the definition from ISO 19101-1:2014, 4.1.2, is the conceptual schema, i.e. the formal description of a conceptual model, for data required by one or more applications or use cases.

Table 2 explains the needed features of each object which should be included in the dataset, properly mapped to the suitable part of the standard data model (yellow part) and further specified as useful for the use case.

'Reference data model subset' attribute is used to indicate standard data models' modules or any other kind of segmentation, as foreseen by the single standards. Application Domain Extensions, as foreseen in CityGML or other kinds of extensions, such as the CityJSON ones can be referred here as well. *'Class'*, in case of CityJSON, refers to the kind of semantics associated to the object (e.g. 'semantic surface').

Whenever possible, the filling of each field should refer to uniquely identified resources, which can be automatically integrated with related semantic resources in the OGC Rainbow or other shared repository. For example, the fields to be filled through URI links can preferably be: the kind of licence and access rights, the coordinate reference system, the standard data model which is the main reference for the semantic and structure, data format, as well as all the aspects of the profile defined (i.e., yellow fields in Fig. 2). Note that URI links may be in a convenient short form by identification of the "namespace" - URI base - for terms that form part of the canonical identifier - thus a simple term "Building" may be interpreted as "namespaceForCityGMLBuilding" provided the lexical form of the term is identical in each case - hence the value of the template derived from the underlying ontology to ensure this is robust.

Ref	GeoDCAT										-	ISO19131
Field	**Title**	**Spatial / geographic coverage**	**Temporal coverage**	**Description**	**Access rights / rights holder (Licence)**				**Geometry**		**Reference (standard) data model**	**Delivery Format**
									Reference system	**Unit of measure**		
Prescribed content	*Title of the dataset, or topic, intended to identify the scope*	*Spatial extent of the dataset, defined in the chosen CRS*	*Temporal framework of the data representation*	*Short description of the dataset*	*BY*	*SA*	*NC*	*ND*	*Coordinate Reference System EPSG - OGC ID in Rainbow*	*Unit of measure*	*Reference data model and version URI in Rainbow (to be selected via a codelist name – URI)*	*Encoding format of the dataset*
Example	Example Building Permit 3D city model	surrounding of the new building e.g. parcel in which the new building is designed + 1km buffer	Updated with latest and approved new works	3D city model LoD 3.3 intended to support automatic checks of urban regulation for new buildings within a GIS environment.	Yes	Yes	No	No	Local projected CRS	m	CityJSON v.1.1 URI	JSON

Table 1. Description of the dataset in the OGC Data Exchange tool template.

To facilitate the filling of URIs as well as the consistent selection of standard data model features, the tables can be filled through an online interface allowing the selection of the possible contents of each cell from a drop-down list, in the cases in which the content should be restricted to specific values, including all the information related to the standard data model prescribed aspects (classes, properties, model version, etc.) and related URIs.

Ref	-	Profile of standard data model					-	GeoDCAT	-
Field	**Object**	**Reference data model subset**	**Class**	**SuperClass with property**	**Property**	**Geometry object**	**Accuracy**	**Spatial Resolution / LoD**	**Additional notes on the spatial representation**
Prescribed content	*Mapped city object*	*URI in Rainbow (to be selected via a codelist name – URI)*	*Selected from drop-down list based on the standard data model (with URIs in Rainbow)*	*Selected from drop-down list based on the standard data model (with URIs in Rainbow)*	*Selected from drop-down list based on the standard data model (with URIs in Rainbow)*	*Selected from drop-down list based on the standard data model (with URIs in Rainbow)*	*Based from the use case requirements defined with stakeholders*	*According to Biljecki et al., 2016*	*Further specifying the kind of representation needed, based on the use case requirements*
Example	Existing buildings	Building	Building	Building	Geometry.type	MultiSurface	0.2 m	LoD 3.3	To be included: external walls with windows, all protruding parts of the building, height = max height of the building at the top of the roof
		Building	Window	Building Building.children*	Geometry.type	MultiSurface	0.2 m		
		Building	WindowSurface	AbstractThematicSurface	boundary	GM_MultiSurface	0.2 m		
		Building	WallSurface	AbstractThematicSurface	boundary	GM_MultiSurface	0.2 m		
		Building	WallSurface	AbstractThematicSurface	lod3MultiSurface	GM_MultiSurface	0.2 m		
	Road, including sidewaks, flowerbed parkings	Road	Road	Road	Geometry.type	MultiSurface	0.2 m	LoD0	
		Road	TrafficArea	AbstractThematicSurface	lod0MultiSurface	GM_MultiSurface	0.2 m	LoD0	
		Road	AuxiliaryTrafficArea	AbstractThematicSurface	lod0MultiSurface	GM_MultiSurface	0.2 m	LoD0	

Table 2. Detailed description of the dataset's objects in the OGC Data Exchange tool template. The yellow fields are intended to define the needed profile, while the others help further specifying the needed features. In the examples, the purple fields are refering to the CityJSON v.1.1 schema, while the green ones refer to CityGML v.3, showing the applicability of the tool to different standards.

3.2 The OGC Data Exchange Toolkit as an extension of ISO 19131

In the *DataContentAndStructureSection* of the *DataProductSpecification* standard ISO 19131.2, it is foreseen that the *DataContentAndStructure* might be represented by either Application schema information (MD_ApplicationSchemaInformation) or content information description (MD_ContentInformation), both borrowed from the metadata standard ISO 19115.

Application schema information as described in ISO 19115 supports the storage of information related to the application schema used, but it does not foresee explicitly the use of standardised domain data models or vocabularies, nor the use and definition of a profile of them.

Figure 3 depicts how the OGC Data Exchange Toolkit extends and further specify the ISO 19131 on 'Data Product Specification' (in grey) to allow a consistent representation of the application schema, referring to standardised data models and respective profiles or extensions (in blue).

3.3 The Profiling Toolkit

The OGC Profiling Toolkit is a suite of open source tools and resources for CI/CT/CD (Continuous Integration, Testing, Deployment) using GIT repositories.

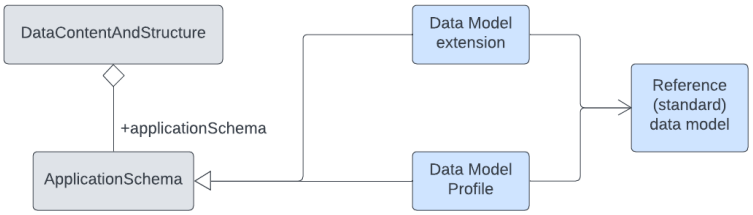


Fig. 3. Extension provided by the OGC Data Exchange Toolkit presented in this paper (in blue) to the ISO19131 DataContentAndStructure section (in grey).

Each tool is managed in separate repository and published as python packages to support reuse. Github actions are used to automatically apply these tools to a repository containing the profile being developed.

The tools leverage, and may publish outputs to **registers** of reusable assets, which are published as Linked Data using the Profiles vocabulary.

Thus the tools themselves embody the FAIR principles, the profile specifications developed using these can be published as FAIR Linked Data, and finally any service or dataset implementing the profiles generated can, by referencing this Linked Data describing the data, itself be published using FAIR principles. The ubiquitous nature of the URI identifiers and publication of the meta-models for profiles, specifications and profiles using the same semantic technologies ensures interoperability of all these different aspects.

Reusability is further enhanced by the provision of validated examples, allowing developers and users of specifications, implementing software or data using this structure to be confident they can easily understand derived profiles.

Table 3 identifies the elements of the Profiling Toolkit and the standards or common infrastructure that underpin these components.

The result of this methodology is a spectrum of related resources, each adhering to a simple suite of standards, with a set of validatable profiles for each standard. The overall architecture and the full set of interoperability requirements for each component are beyond the scope of this description, however using Linked Data principles each resource can be linked to standards and validation processes, and the data management history (provenance) or any other relevant metadata.

3.4 Test and validation

To validate this approach a simple experiment was undertaken in advance of city-scale testing planned in the next phase.

The validation focuses on several questions:

Component	**Purpose**	**Interoperability Foundations**
Profiles Register	Source and target for re-use of profiles	PROF vocabulary
Profile repository	Development of profiles	GIT
Validators	Rules defining a profile	e.g. SHACL
Examples	Illustration and testing of validators	e.g. JSON-LD
Description	Additional documentation	Source + Rendered HTML
Source models	Authoritative standards	UML or encoding schema (e.g. XSD)
Source model OWL version	integration of model	OWL, SHACL
Controlled Vocabularies	Profile data rules	SKOS
Validation Pipeline	CI/CT/CD	GIThub actions, docker etc
Semantic Uplift	link data to profile	JSON-LD
Semantic transformers	data commonality (isomorphism) validation	SHACL-AF

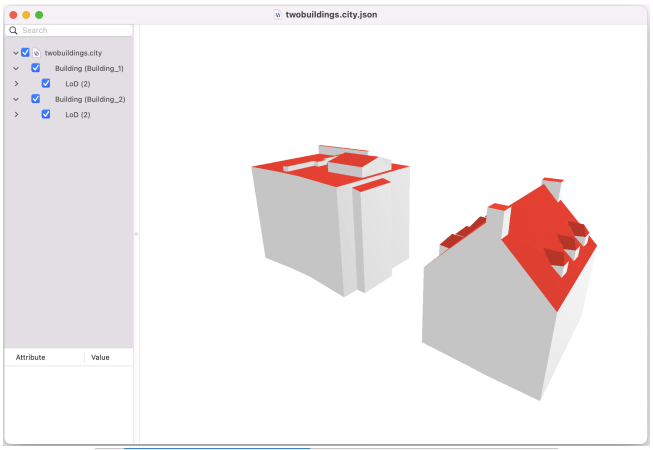
Table 3. Components of the OGC Profiling Toolkit

- *Which parts (profile) of CityJSON need to be present?*
- *Which specific additional requirements the dataset and data objects should comply to?*
- *Can such a profile be described and used to validate data meets these requirements?*
-

Figure 4 shows a snippet of the CityJSON test case (with verbose details such as coordinates and vertex references collapsed for readability). The "semantics" element (e.g. "WallSurface" provides a sub-typing approach that is equivalent to the Class specialisation in the CityGML UML model.

Using the OGC Profiling Toolkit a JSON-LD context document (Figure 5) can be used to transform the JSON into an explicit implementation of the CityGML model. The complexity of this process is beyond the scope of this experiment, however the tools to perform linked uplift and transformation pipelines are an important component of the OGC Profile Toolkit necessary to deal with the myriad of alternative ways of structuring information. The general approach is to map the data structure to the semantic concepts early, which then allows alternative representations to co-exist in the same Linked Data environment if necessary.

Figure 6 shows a snippet of the resulting form of this data matching the target ontology (CityGML). This RDF syntax (Turtle) represents the data in a




```
{
  "type": "CityJSON",
  "version": "1.1",
  "metadata": {...},
  "CityObjects": {
    "Building_1": {
      "geometry": [
        {
          "boundaries": [...],
          "lod": "2",
          "semantics": {
            "surfaces": [
              {
                "type": "GroundSurface"
              },
              {"type": "WallSurface"...},
              {"type": "RoofSurface"...}
            ],
            "values": [...]
          },
          "type": "MultiSurface"
        }
      ],
      "type": "Building"
    },
    "Building_2": {"type": "Building"...}
  },
  "vertices": [...],
```

Fig. 4. Sample CityJSON data for a simple test case. Visualization in azul (Arroyo Ohori, 2020) and JSON encoding (with some details collapsed)

```
"@context": {
  "@version": 1.1,
  "@base": "#",
  "city": "http://example.com/vocab/city/",
  "dct": "http://purl.org/dc/terms/",
  "xsd": "http://www.w3.org/2001/XMLSchema#",
  "CityJSON": "city:CityJson",
  "type": "@type",
  "version": "city:version",
  "CityObjects": {
    "@id": "city:hasObject",
    "@context": {
      "Bridge": "city:Bridge",
      "BridgeConstructiveElement": "city:BridgeConstructiveElement",
      "BridgeFurniture": "city:BridgeFurniture",
      "BridgeInstallation": "city:BridgeInstallation",
```

Fig. 5. Example section of JSON-LD context that maps JSON structural elements to the CityGML semantic model

canonical form that can be directly validated against the model and profile using standardised validation rules and toolkits. Figure 7 shows an example validation rule in the SHACL (Shapes Constraint Language) standard.

The defined rules and profile were succesfully tested against the '2buildings' sample dataset (Fig. 4). The resulting validation report (Figure 8) shows three different types of validation that can be made. The first relates to general requirements - such as data licences, date (currency) of information etc. The second violation relates to whether the particular dataset has the right level of detail - this is optional in the underlying CityJSON (and CityGML) models, but a requirement to meet the CHEK business rules.

Thus the OGC Profiling Toolkit can support the full process of bringing one or more different data forms into a common semantic framework where explicit rules can be executed to validate the data. The profiles allow sophisticated and fine-grained control over exact data requirements.

4 Discussion

This initial development allowed testing the feasibility of the approach, which will enable the development of a more complete profiling tool for more complex cases.

```
<#city-objects-Building_1> a city:Building ;
    city:hasGeometry [
            city:hasSurface [
                a city:GroundSurface, gml:MultiSurface ;
                city:boundaries ( ( ( <#vertices-0> <#vertices-1> <#vertices-2> ) )
            ] , [
                a city:WallSurface, gml:MultiSurface ;
                city:boundaries ( ( ( <#vertices-11> <#vertices-9> <#vertices-12> )
            ] , [
                a city:RoofSurface, gml:MultiSurface ;
                city:boundaries ( ( ( <#vertices-20> <#vertices-19> <#vertices-21> )
            ];
            city:lod "3.3" ;
    ] .
```

Fig. 6. Example canonical form of data ready for validation

```
<#Lod3.3>
    a sh:NodeShape ;
    sh:targetClass city:Building, city:BuildingPart, city:BuildingRoom, city:BuildingStorey,
     city:BuildingUnit, city:BuildingInstallation;
    sh:not [
      sh:sparql [
        sh:prefixes ex: ;
        sh:select """
          SELECT $this (city:lod as ?path) (?lod as ?value) WHERE {
            $this city:hasGeometry/city:lod ?lod .
            FILTER regex(str(?lod), '^(3.3|4.*)$')
          } LIMIT 1
      """ ;
      ] ;
    ];
    sh:message "Building has LoD < 3.3" ;
    sh:severity sh:Violation
.
```

Fig. 7. Example SHACL rule (LOD present)

```
Validation Report
Conforms: False
Results (6):
Constraint Violation in NotConstraintComponent (http://www.w3.org/ns/shacl#NotConstraintComponent):
        Severity: sh:Violation
        Source Shape: <file:///github/workspace/ogc/chek-profiles/cityjson/paper/shapes.ttl#License>
        Focus Node: <http://example.com/twobuildings.ttl#city>
        Value Node: <http://example.com/twobuildings.ttl#city>
        Message: License does not support commercial usage
Constraint Violation in NotConstraintComponent (http://www.w3.org/ns/shacl#NotConstraintComponent):
        Severity: sh:Violation
        Source Shape: <file:///github/workspace/ogc/chek-profiles/cityjson/paper/shapes.ttl#Lod3.3>
        Focus Node: <http://example.com/twobuildings.ttl#city-objects-Building_2>
        Value Node: <http://example.com/twobuildings.ttl#city-objects-Building_2>
        Message: Building has LoD < 3.3
Constraint Violation in NotConstraintComponent (http://www.w3.org/ns/shacl#NotConstraintComponent):
        Severity: sh:Violation
        Source Shape: <file:///github/workspace/ogc/chek-profiles/cityjson/paper/shapes.ttl#RoadsRequired>
        Focus Node: [ ]
        Value Node: [ ]
        Message: Dataset contains no Road objects
```

Fig. 8. Validation report from the testing of the developed profile over the '2buildings' sample dataset.

A significant challenge, was to choose a data encoding for sample data. In this case CityJSON was chosen to represent the CityGML model. The exact degree of conformance of CityJSON to the full CityGML model is an open question, however the working hypothesis is that for the subset required (the CHEK profile) CityJSON and CityGML terminology and semantics are equivalent, even though for the case of CityJSON the mapping of concepts to data structures for elements such as geometry and topology are not formally defined.

Some limitations and obstacles to be overcome were pointed out, to support the future efforts. A key obstacle, which is directly addressed by this approach, is the lack of formal descriptions of how existing data formats and datasets relate to common underlying concepts. The CHEK project will demonstrate that every aspect does not need to be resolved to demonstrate that for a specific profile of each standard sufficient interoperability can be established to automate data transformation and business rule evaluation.

A key outcome of this is the potential to characterise the commonality, differences, inter-relationships and transformation options between different 3D geometry and topology regimes, which in turn becomes a building block for describing integrations between different application domain models in the 3D environment. This outcome arises in part from the unique approach CityJSON uses to define a set of vertices and then reference them in nested array structures carrying implicit semantic information (e.g. the first level of nesting of a solid is the set of faces of that solid).

The inclusion of more detailed specification of the single objects in a structured format allows removing the ambiguity still remaining in many datasets, even though they may conform to various structural data models. Extending specifications to the specific semantic information required to support trans-

formation to other well-known standards enabling real interoperability and reusability of datasets. Datasets that declare conformance to specific profiles can be reliably validated and transformed using automated and robust pathways in a transparent and reproducible fashion, using tested reusable tools to reduce effort, cost and risk of invalid outcomes.

Creating profiles of complex standards such as OGC CityGML and buildingSMART Industry Foundation Classes (IFC)[13] for Building Information Models (BIM) allow for data sharing, including supporting exchanges between Geographical Information Systems and engineering environments by focusing on the "Goldilocks zone". Furthermore, the flexible nature of semantic expressions of this simplified views means these can be extended for specific purposes later, leaving a simplified core neither over-specific nor under-specified for the practical purposes of integrating separate systems for specific purposes.

Vocabularies, syntactical rules for data, inferencing rules for related data and consistency checking and transformations between related data models can all then be defined against the profile, without requiring the full standard data model to be treated in the same way. Example data objects can be used to test these models (and create useful, validated examples in documentation), using the same profiling toolkit used for each part of the overarching spectrum of resources.

5 Conclusions

A prototype of a data exchange toolkit for 3D geoinformation was developed in form of a data specification template intended to be filled (for the moment) manually, according to the use case needs, and an OWL-based profiling tool to support automatic extraction and validation of datasets against the defined profile.

In future work the tool will be further tested with more extended data requirements.

In addition, it will be completed to include the whole CityGML v.3 standard model, as well as extending the functionalities, by following the same methodology, to other standard data models from OGC and possibly external ones (e.g. the INSPIRE thematic model[14]). Finally, the management of Application Domain Extensions or similar mechanisms to extend the standard data models will be added and tested in future releases of the tool.

An additional attribute, relevant for the dataset description and possible data requirement could be 'Language', as also proposed within the GeoDCAT model. However, for the moment we only consider datasets in English as a prescription. Future works could be intended to extend the tool to foresee the use of different languages. Other inegrations and extensions may be considered within the future developments.

[13] https://technical.buildingsmart.org/standards/ifc/ifc-schema-specifications/
[14] https://inspire.ec.europa.eu/data-model/approved/r4618-ir/html/

Moreover, the toolkit will be refined within the CHEK project as well as in other research projects in which the OGC is involved (e.g. USAGE[15], AD4GD[16], ACCORD[17], ILIAD[18]) as well as becoming part of the OGC developments and discussions with OGC members, to improve the performances and test within multiple use cases.

Using the same approach, further checking can be performed on the objects themselves against the business rules (permit checking in this case). Such rules are not limited to SHACL, however by expressing these against the canonical (semantic) model such business rules can be applied to many forms of original source data, and can be more transparent because specialised code to apply to specialised data structures is avoided.

The results of the initial experimentation has confirmed positive answer to the feasibility questions and show a mechanism to define a profile can be based on a spectrum of standardised components, linked by open and transparent data transformation and validation pipelines.

The complexity of the structural transformations (such as geometry and subtyping classifications) highlights how important it is to isolate these aspects from the business rules. The business rules can be defined against the single semantic model, and any complex structural transformations between semantically equivalent encodings can be done using reusable tools.

Acknowledgements

This work was funded by the 'Change toolkit for Digital Building Permit' (CHEK) project, funded by the European Union's Horizon Europe programme under Grant Agreement No.101058559

[15] http://usage-project.eu
[16] https://ad4gd.eu
[17] https://accordproject.eu
[18] https://www.ocean-twin.eu

Bibliography

Arroyo Ohori K (2020) azul: A fast and efficient 3d city model viewer for macos. Transactions in GIS 24(5):1165–1184

Hjelseth E, Nisbet N (2011) Capturing normative constraints by use of the semantic mark-up rase methodology. In: Proceedings of CIB W78-W102 Conference, pp 1–10

ISO/TC 211 (2014) Iso 19115-1:2014, geographic information — metadata — part 1: Fundamentals. Standard ISO 19115-1:2014, International Organization for Standardization, URL https://www.iso.org/standard/53798.html

ISO/TC 211 (2022) Iso 19131:2022, geographic information — data product specifications. Standard ISO 19131:2022, International Organization for Standardization, URL https://www.iso.org/standard/85092.html

Ledoux H, Arroyo Ohori K, Kumar K, Dukai B, Labetski A, Vitalis S (2019) Cityjson: A compact and easy-to-use encoding of the citygml data model. Open Geospatial Data, Software and Standards 4(1):1–12

Noardo F (2022) Multisource spatial data integration for use cases applications. Transactions in GIS 26(7):2874–2913

Noardo F, Arroyo Ohori K, Biljecki F, Ellul C, Harrie L, Krijnen T, Eriksson H, van Liempt J, Pla M, Ruiz A, et al (2021) Reference study of citygml software support: The geobim benchmark 2019—part ii. Transactions in GIS 25(2):842–868

Raes, Lieven and Vandenbroucke, Danny and Reznik, Tomas (2019) Geodcat-ap. OGC Discussion paper 18-001r1, Open Geospatial Consortium, URL https://portal.ogc.org/files/?artifact_id=82475

Wagner D, Wewetzer M, Bogdahn J, Alam N, Pries M, Coors V (2012) Geometric-semantical consistency validation of citygml models. In: Progress and new trends in 3D geoinformation sciences, Springer, pp 171–192

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, et al (2016) The fair guiding principles for scientific data management and stewardship. Scientific data 3(1):1–9