architecture

MDPI

Article

Re-(De)fined Level of Detail for Urban Elements: Integrating Geometric and Attribute Data

Benjamin Boswick, Zachary Pankratz, Matthew Glowacki and Yuhao Lu *

Faculty of Architecture, University of Manitoba, Winnipeg, MB R3T 2M6, Canada;
boswick3@myumanitoba.ca (B.B.); pankra13@myumanitoba.ca (Z.P.); glowackm@myumanitoba.ca (M.G.)
* Correspondence: yuhao.lu@umanitoba.ca

Abstract: The level of detail (LOD) differentiates multi-scale representations of virtual 3D city models; however, the LOD tends to relay primarily the geometric details of buildings. When the LOD extends to other datasets, such as vegetation, transportation, terrain, water bodies, and city furniture, their LODs are not as clearly defined. Despite the general acceptance of this categorization, existing LOD formats also neglect non-geometric attributes. Integrating geometric and attribute data enables geometrically accurate and data-rich 3D models, ensuring that representations are as accurate as possible and that analyses contain as much information as possible. This paper proposes a family of LOD definitions considering both geometric and attribute data based on the geometric complexity and difficulty of obtaining, archiving, processing, and distributing the data. These definitions are intended to apply to all datasets by determining divisions in the LOD typically experienced across urban 3D model elements and their associated datasets, including buildings, vegetation, roads, relief, water bodies, and city furniture. Universally applicable definitions for datasets allow individuals to recreate studies or representations of 3D models to ensure the relevant data are present. These definitions also assist data providers in evaluating their data infrastructure and further strategizing and prioritizing updates or upgrades.

Keywords: level of detail (LOD); urban digital twins; 3D modelling; geometric data; attribute data

check for updates

Academic Editor: Zhonghua Gou

Received: 18 October 2024
Revised: 26 November 2024
Accepted: 17 December 2024
Published: 25 December 2024

Citation: Boswick, B.; Pankratz, Z.; Glowacki, M.; Lu, Y. Re-(De)fined Level of Detail for Urban Elements: Integrating Geometric and Attribute Data. *Architecture* **2025**, *5*, 1. https://doi.org/10.3390/architecture5010001

1. Introduction

Urban digital twins (UDTs) are one of the most fundamental methods of representing and studying a city in 2D and 3D models. More specifically, visualization (data takers) and analyses (data producers) are two primary functions for UDTs. Recent scholarly work has further cemented the benefits of creating centralized data-driven models capable of inventorying, assessing, and simulating both existing and future urban environments. Extending from conventional computer-aided design (CAD) urban models, data-driven models integrate geometric shapes (objects) with semantic attributes (data) (e.g., [1,2]). With modern remotely sensed data and processing techniques, urban models become more complex in terms of their geometric precision (level of details), extent, and analytical capabilities.

Sophisticated urban models like those in Great Britain [3], Zurich [4], and Singapore [5] are typically found in resourceful cities. This is largely because creating and maintaining such models are costly, requiring massive amounts of clean and structured datasets representing different subsystems of a city. These subsystems are then interconnected as layers to form a model of models, which further increases the level of complexity, and resources needed. These integrations necessitate a clearly defined framework that considers both geometric and semantic data for efficiency, cost saving, and better alignment between geometric objects and their associated attributes.

UDTs use geometries and datasets encompassing urban elements such as buildings, vegetation, transportation, water bodies, and infrastructure. The granularity of these datasets, known as the level of detail (LOD), varies significantly from one dataset to another. The LOD provides a framework for UDT modellers to construct, codify, and visualize elements commonly seen in a city. Evaluation and comparison among datasets in UDTs therefore rely on the clearly defined LOD of each dataset. In science and mathematical models, a clearly defined LOD of model attributes reduces the ambiguity of the modelling process, therefore improving transparency and reproducibility in decision-making [6]. However, finding the appropriate LOD can be one of the most challenging aspects of modelling [7].

CityGML, a standard for data representation, storage, and exchange for UDT, defines the LOD for building-related datasets but lacks a clearly defined LOD for other urban elements. Attempts and standards have been made in the existing literature to create definitions similar to the LOD of a building dataset (e.g., [8]). However, the current work falls short by neglecting the attributes associated with the geometry from each dataset, resulting in an incomplete and ambiguous definition. In addition, the current LOD is limited to only buildings and lacks the capacity for non-geometrically explicit attribute data [9]. With increased interest in adopting data-driven UDT visualization and analyses [10], it is imperative that the LOD must be able to accommodate a wide range of elements (e.g., vegetation, streets, terrain, water bodies, and city infrastructures) and their associated attributes beyond geometries. Our work re-(de)fines the LOD for both geometric and attribute data.

This paper will first review the existing framework, definitions, and contributions made by previous scholarly work (Section 2). Section 3 provides an overview of the methods and reasoning we used to develop our refined LOD. Section 4 consists of the results of our work, including graphic depictions and verbal definitions of the LOD definition refinements. Section 5 contains applications of our LOD definition refinements to three case study scenarios. Section 6 discusses the LOD, and points of further research associated with our work.

2. Literature Review

2.1. CityGML

CityGML is a standard that facilitates data representation, storage, and exchange for the 3D modelling of cities [11]. The current version contains modules for datasets including vegetation, transportation, relief, water bodies, and city furniture.

In the CityGML standard, the Vegetation module contains two types of vegetation representation. The first class is called *SolitaryVegetationObject* (SVO) and represents individual vegetation objects, such as boulevard trees. The second class is called *PlantCover* and represents areas with dense and continuous coverage, such as forests [12]. An SVO may be geometrically implicit and represented by LODs of 1 to 4 with increasing detail. A *PlantCover* is represented by either a *MultiSolid* or *MultiSurface* object [12].

The Transportation module in CityGML represents transportation corridors such as roads, railways, and public squares. The main *TransportationComplex* class differentiates *TrafficAreas* and *AuxiliaryTrafficAreas*, such as roads or sidewalks, in contrast with boulevards. The higher the LOD, the greater the distinction between these spaces.

Terrain is modelled in the CityGML format as a *ReliefFeature* consisting of one or more entities called *ReliefComponents*. These entities may include a TIN (Triangulated Irregular Network), which is a network of non-overlapped triangles formed by the interconnection of irregularly spaced points [13]. Additionally, other representation methods include mass points (*MasspointRelief*), break lines (*BreaklineRelief*), or grids (*RasterRelief*). There is potential for different LODs for *ReliefFeature* and *ReliefComponents*.

A series of surface boundaries define WaterBody geometry. The *WaterGroundSurface* represents the very bottom of the water body, the submerged surface such as riverbeds, whereas *WaterSurface* represents the upper surface, which represents the boundary between the water body and the atmosphere [14]. LOD 0 and LOD 1 represent the lowest level of illustration and have a high level of generalization [12]. A combination of geometries may be used to represent the WaterBody at each LOD.

Previously, CityFurniture objects could be represented by explicit or implicit geometry. Explicit geometry refers to using a specific, instanced object, whereas implicit geometry refers to using a prototypical object repeatedly within the model [12]. No additional information regarding the LOD is provided for the CityFurniture module.

Although CityGML is just one standard for modelling UDTs, its framework and modules provide a starting point for working with the organization of datasets. CityGML utilizes a LOD for each of its modules and corresponding datasets. The LOD of building datasets is the only one that is explicitly defined. The LOD is certainly referenced in every other module, but the explicit LOD definition of each remains vague. This vagueness and lack of refinement of CityGML has drawn attention and proposals for refinements. In the following section, we review peer-reviewed publications on this topic and summarize research gaps in LOD refinement for buildings, vegetation roads, terrain/relief, water bodies, and city furniture.

2.2. Related Work

This study reviewed the existing literature on LOD refinement across various urban datasets. Searches for academic papers concerning the LOD refinement of other datasets included the following keywords/phrases: LOD, Level of Detail, Definition, Refined, 3D Model, Urban Digital Twin, Digital Twin City, Trees, Vegetation, Solitary Vegetation Object, Plant Cover, Road, Intersection, Transportation, Terrain, Relief, Waterbody, Water, Surface, and CityFurniture. We reviewed around seventy-seven sources, ten of which were explicitly related to refining the LODs of datasets. These related works contributed to the refinement and formalization of the LODs of other 3D model datasets, such as buildings [15], trees and vegetation [16], roads and transportation [17–20], terrain [13], water [21,22], and city furniture [23].

Ambiguities in CityGML's building LOD definitions have prompted efforts for supplementary refinement. The authors of [15] refined building LOD definitions in the context of the CityGML format. The authors of [15] also identified the lack of precise LOD definitions in the CityGML format, which allowed for ambiguity. Ambiguity was observed in instances of models with varying differences in architectural features that were considered part of the same LOD. The intent behind refining the building LOD definition was not to extend CityGML but to provide a supplementary specification that reflects the practices and current concepts while also solving the ambiguities identified.

LOD definitions for Solitary Vegetation Objects (SVOs) are less common compared to those for buildings. Reference [16] is the only article that provided use cases and a definition framework for SVOs in the CityGML Vegetation module. Ortega-Cordova identified five categories of applications for urban vegetation models and data: managing, maintaining, sustaining urban vegetation, urban and landscape planning, and policy making. The use cases within these categories are used to inform their LOD definition refinement. Ortega-Cordova provided a refined framework for the geometric LODs of SVO modules, but did not accommodate the inclusion of potential attribute data.

Refinements in road and transportation LOD definitions have focused on geometric representation and use-case applicability. The author of [17] focused on roads and the associated vehicular transportation portion of the CityGML Transportation module. Boersma's LOD refinement was built on existing research by [18,19]. Additionally, use cases for road data were provided as a foundation for forming the refined definitions. Boersma identified three primary categories of use cases for road and transportation data, including transport and traffic models, navigation, and road maintenance. Based on street design, The authors of [20] emphasized the importance of the traffic area, driving lanes, and traffic logic in modelling roads. The author of [17] provided refinement in the geometric representation of traffic areas and driving lanes but lacked the attribute data associated with these components and traffic logic.

Efforts to refine terrain LOD definitions focus primarily on geometric representation. The authors of [13] provided a method for refining terrain LOD definitions. Based on the existing state of terrain representation in the CityGML standard, they emphasize the relationship between the Construction module, which includes Bridges and Tunnels, and the Relief module. They also explained in detail the method behind their proposal. Although the authors of [13] redefined the LOD associated with the geometric representation of terrain, other attribute data, such as geological composition or land use, are not included in their work. The geometric-focused approach of [13] neglects other important features/attributes that may be associated with terrain but are only sometimes geometrically represented, such as materials and geologic features.

No published peer-review articles were found on refining WaterBody LOD definitions. A WaterBody is defined as a significant and permanent or semi-permanent accumulation of surface water. Examples may include rivers, canals, lakes, and basins [14]. No published peer-review articles were found on refining WaterBody LOD definitions. Existing CityGML documentation was reviewed to understand the components needed for a WaterBody object. However, the literature on hydrology [21] identified difficulties in the modelling and simulation process that resulted from a lack of clearly defined LODs for water bodies. This is largely due to computing needs, as the LOD increases in complex 2D and 3D hydrological simulations in a UDT (or a city information model, CIM). In addition, while there is no direct literature on WaterBody LOD refinement, The authors of [22] highlighted that accurate flood simulation and assessment in a UDT requires clearly defined LODs of not only water but other interdependent urban structures such as buildings and road networks.

CityFurniture is one of the most open-ended datasets, as different cities will have different urban infrastructure objects. CityFurniture objects involve immovable objects that include decorations, explanations, or controls [14]. Examples may include street signs, traffic signals, streetlamps, benches, and fountains [14]. No published peer-review articles were found on refining CityFurniture LOD definitions. The literature on street design and mobility analysis [23], however, highlighted the importance of CityFurniture in mapping and managing public streets as the geometric condition and materials of CityFurniture can significantly impact the quality and experience of streetscapes. However, such data and a LOD framework to organize these datasets are lacking and not well aligned with other urban elements. With advancements in sensing technologies such as mobile laser scanning (MLS), increasingly rich and accurate datasets are expected to become more readily available. This study anticipates the influx of such high-dimensional datasets and proposes an LOD framework that integrates both geometric and semantic information for city furniture, addressing the growing need for comprehensive data representation in UDT models.

In summary, LOD refinements from the existing literature try to retrofit their corresponding dataset within the same definition framework as [15], and therefore similar shortcomings are experienced in each urban element reviewed above. Although significant attention is given to sorting geometric differences, The authors of [15] acknowledged the limited effort in associating embedded attributes with geometries in their refinement in favour of delineating the geometry divides between their LOD families. Embedded attributes refer to data that are typically associated with and may influence other geometric data but are not necessarily geometrically represented in 3D models. The existing literature addresses the ambiguity in graphic depictions mainly associated with geometry, but all lack attribute data. (See Table 1).

Table 1. Summary of reviewed literature on LODs.

Dataset	Published Work	Shortcomings
Buildings	[24]	Focus on refining geometric LODs to encompass a variety of geometric forms; no embedded attribute data included in present LOD definition
Vegetation/Trees	[16]	Focus on refining geometric LODs based on specific uses identified in thesis; no embedded attribute data included in present LOD definition
Roads	[17–19]	Focus on refining geometric LODs, lacks geometric and attribute data associated with transportation logic, no embedded attribute data included in present LOD definition
Terrain	[13]	Focus on refining representation and methods of representation of geometric LODs, no embedded attribute data included in present LOD definition
WaterBody	None found	Existing ambiguous LOD definitions beyond LOD 2, no embedded attribute data included in present LOD definition
CityFurniture	None found	No geometric or embedded attribute data included in present LOD definition, completely ambiguous

Excluding attributes in the refinement of LOD definitions hinders research necessitating comparable and consistent datasets. Most datasets from municipal and federal governments feature non-geometric data that would not fit these existing definition refinements. While some are geometrically accurate, the existing definition refinements do not allow for evaluations and comparisons of datasets among cities because of limited attributes defined using a standardized framework.

Our work proposes to formulate a family of definitions considering both geometric and attribute data. These definitions are intended to be applicable to all datasets by determining divisions in LODs typically experienced across all UDT elements and their associated datasets. Universally applicable definitions for datasets would also enable individuals to recreate studies or representations of UDTs to ensure all relevant data are present. These definitions would also assist data providers in evaluating their data infrastructure, and therefore further strategizing and prioritizing updates or upgrades. This level of evaluation is not possible with the current ambiguously defined LODs of datasets.

3. Methods

A good definition should be concise and applicable by researchers and practitioners. The role of data in the definition must be universal, accessible, and self-explanatory. The definition of LODs should consider the physical form (geometry + visual) and analytical function (attributes + data). The definition should be based on a careful selection of the involved parameters, and the parameters should be data-driven to avoid arbitrary choices. The definition should seek validation based on a procedure that is general and widely accepted by the community, including one or more quantitative criteria to assess the quality of the definition.

The studies identified earlier focused on the geometric data in different datasets used primarily in 3D modelling. Our proposal includes geometric and attribute data with increments of LOD based on the geometric complexity and difficulty of obtaining, archiving, processing, and distributing the data. We define complexity as the amount of geometric and attribute detail the dataset provides. Difficulty refers to the accessibility of the geometric and attribute data as annotated below in Table 2.

Table 2. Geometric and attribute LOD verbal descriptions.

	Geometric Definition	Attribute Definition
LOD 0	1-Dimensional point	Easily obtained, potentially crowd-sourced data
LOD 1	2-Dimensional line segments, polygons	Open data made available by different levels of local government (e.g., municipal, provincial, etc.)
LOD 2	2-Dimensional polygons extruded to 3-Dimensional height based on proxy	Data obtained by contractor that are not readily available through open data portals
LOD 3	3-Dimensional replica ranging in granularity from simple massing to complete, realistic	Specialty data that are obtained for academic or individual studies or other purposes that are not readily available through open data portals

Therefore, LOD (0,0) would be the least complex geometrically, with limited attributes embedded within each geometry, and have the lowest difficulty of data acquisition from the actual object. This contrasts with LOD (3,3), which would have the richest geometric and attribute data, and therefore would be time-consuming and expensive to obtain. Each LOD is an ordinal measurement rather than a ratio, meaning that LOD 3 is not necessarily 3 times more complicated or detailed than LOD 1.

Given the deliberate division between geometric and attribute data, those prioritizing data may fall within the data makers category. In contrast, those prioritizing geometric features may fall within the data takers category. Visualization would involve obtaining data to represent a space within the UDT. Such uses include rendering perspectives of architectural elements within the street space to demonstrate modifications or improvements proposed by a design and planning firm. The intention is to create an image (2D) or a scene (3D) as photorealistically as possible. Although a scene is produced, no additional data are directly created during the process. Simulation and data analytics, on the other hand, involve utilizing both non-geometric (attributes) and geometric data to typically study the interactions among UDT elements and with other external factors, such as climatic characteristics on specific buildings or socio-demographics of neighbourhoods, producing additional data as an outcome.

This paper proposes a framework (Figure 1) that contains two axes, reflecting both geometric and attribute LODs of urban elements. The geometric axis is divided into four increments that increase in complexity furthest from the axis.

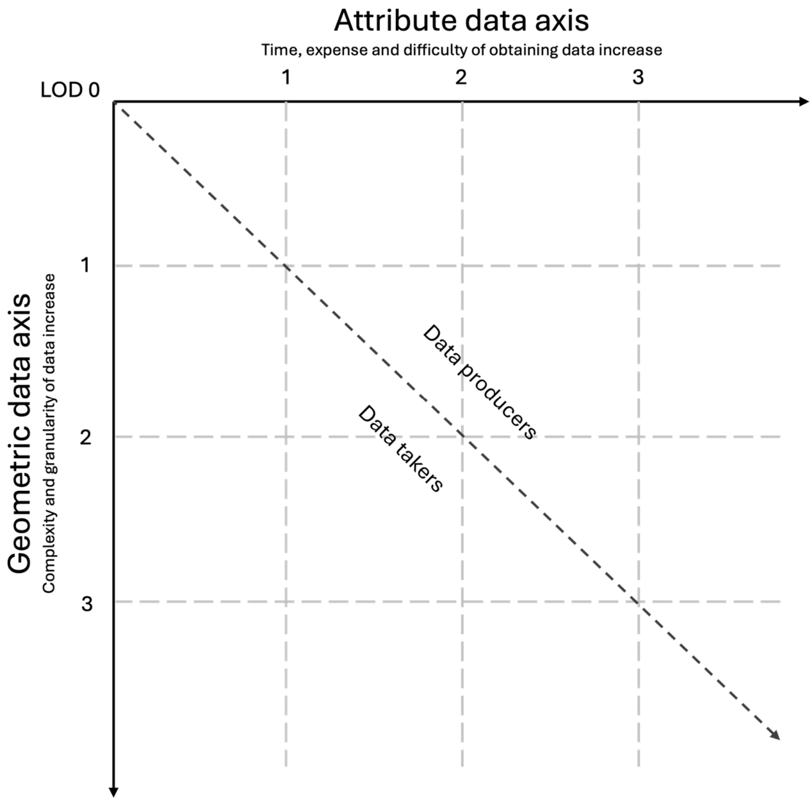

Figure 1. Geometric and attribute data LOD conceptual matrix.

Geometrically, the LOD is defined by divisions in dimensionality. Dimensionality is generally defined by the increase in geometric complexity of an object, involving point, area (length, width, space), and height. LOD 0 is 1D (restricted to points or centrelines), LOD 1 is 2D (restricted to lines and 2D polygons), LOD 2 is 2.5D (2D polygons extruded to height based on a proxy), and LOD 3 is 3D (graphic depictions may range from basic solid geometry to a 1:1 digital model of the object geometry).

There are a few exceptions in our LOD definitions. First, considering the geometric nature and common datasets of roads and water bodies, using points to geometrically define LOD 0 for these two elements is not practical. Therefore, we use basic centrelines to represent LOD 0 for roads and water bodies. Secondly, terrain is often perceived as a surface that is represented by contour lines or digital elevation models. Therefore, relief (terrain) LOD 0 (geometry and attribute) is absent in our definition.

In terms of data attributes, the LOD may be defined by divisions in accessibility and cost. Accessibility is defined by the ability to obtain data through different sources. These sources may include government or related open data portals, or private datasets made available by other institutions. Cost is defined by the amounts of financial cost and time given to obtaining data. LOD 0 is easily observed, widely available data, while LOD 1 is open data made available by a city or municipal government. LOD 2 refers to data obtained and made available by a third-party contractor, while LOD 3 represents data obtained directly by the group or individual, likely academic institutions or individual researchers. LODs 0 to 3 increase in cost as the scope and specificity grow and therefore would not be readily available within one of the lower LODs.

Visualizing the datasets was a straightforward process which was accomplished by translating geometric definitions into graphic depictions using Rhino. To illustrate the different LODs, a hypothetical square city block was created where each LOD and geometric definition was showcased over the same area. The definitions progressed from simple point/curve representations to more complex 2D polygons/areas, 2.5D shapes/volumes, and finally, detailed 3D likeness models.

4. Results

4.1. Buildings

Figure 2 below illustrates a hypothetical building dataset at four LODs with the attribute examples outlined in Table 3. Figure 2a uses building footprint centroids and coordinates, exemplifying the most straightforward geometry (point) and attributes (coordinates). The colours represent the variations in coordinates. Figure 2b adds a second geometric dimension (polygon) and a slightly more complex attribute, such as building classification. Colours represent civic (blue), residential (yellow), and commercial (red). Figure 2c features simple 2.5D forms representing previous building modifications, where the colour gradient indicates how recent a renovation is. Figure 2d showcases three-dimensional geometry with detailed architectural features, while the colour gradient represents annual energy use.

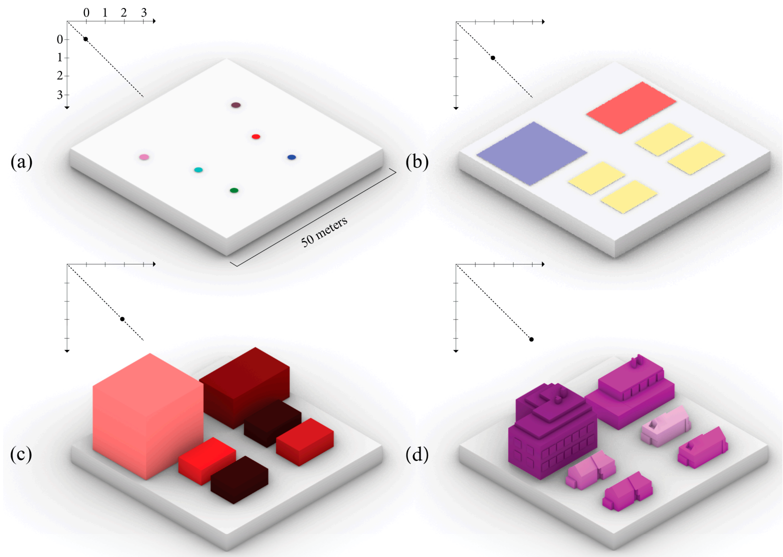

Figure 2. LOD graphical definitions of buildings. In addition to geometric form, different examples of attributes are represented by different colour palettes and gradients. The model spans a 30 by 30 m area to represent a variety of buildings of (**a**) LOD 0.0, (**b**) LOD 1.1, (**c**) LOD 2.2 and (**d**) LOD 3.3.

Table 3. Verbal descriptions of building LODs.

LOD	Geometric Definition	Attribute Definition
0	Building centroid point	Street address, lat/long location
1	Polygon of building footprint	Area/square footage, classification, year built, number of floors, construction material, assessment value
2	Polygon of building footprint extruded to approximate height based on number of floors	Years of renovations/modifications
3	3D extrusion with accurate height ranges from simple massing to 3D replica—may include interior and underground structure	Foundation type, HVAC system, energy use, water use

4.2. Vegetation (Trees)

Figure 3 below illustrates a hypothetical tree dataset at *four* LODs with the attribute examples outlined in Table 4. Figure 3a uses tree locations, exemplifying the most straightforward geometry (point) and attributes (coordinates). Colours represent the variations in coordinates. Figure 3b adds a second geometric dimension (polygon) and a slightly more complex attribute, such as canopy height. Colours represent an approximation of general height in three categories: tallest, medium, and *shortest*. Figure 3c illustrates species with colour variations, highlighting their *2.5D* forms and general relationships. Figure 3d shows the specific tree form characteristics necessary for environmental simulations *such as canopy shading*. The colour gradient represents the physiological conditions of trees, an attribute that is typically difficult to obtain.

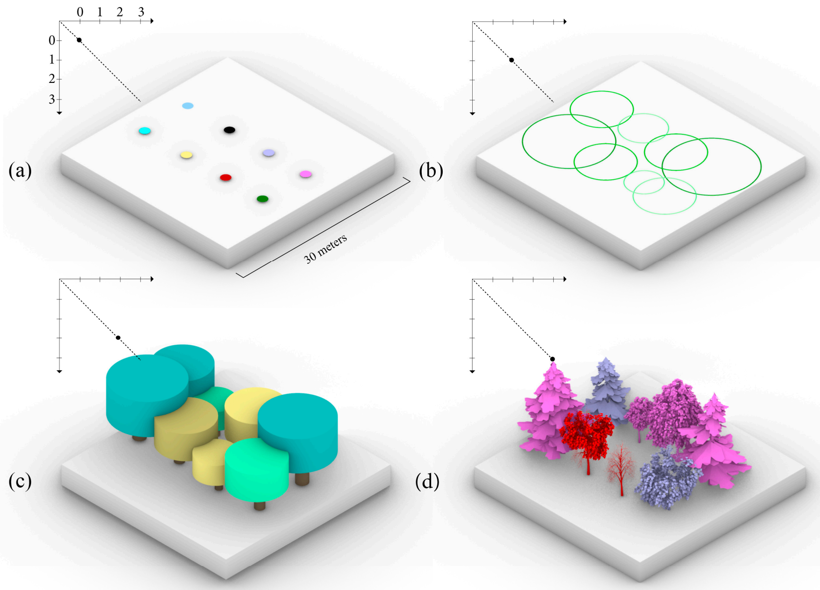

Figure 3. LOD graphical definitions of vegetation. In addition to geometric form, different examples of attributes are represented by different colour palettes and gradients. The model spans a 30 by 30 m area to represent the spatial distribution of street vegetation of (**a**) LOD 0.0, (**b**) LOD 1.1, (**c**) LOD 2.2 and (**d**) LOD 3.3.

Table 4. Verbal descriptions of vegetation LODs.

LOD	Geometric Definition	Attribute Definition
0	Point location for treetops	Coordinates
1	Polygon of estimated drip line of tree or forest	Coordinates, address, canopy width, canopy height, pesticide treatments
2	Polygon extruded to an estimated height based on canopy width and known species	Canopy density, species, DBH
3	3D extrusion with accurate height ranges from simple massing to 3D replica—may include underground root massing	Physiological character, pathological condition (health)

4.3. Transportation (Roads)

Figure 4 below illustrates a hypothetical road dataset at four LODs with the attribute examples outlined in Table 5. The graphic represents a typical boulevard and residential street intersection. Figure 4a uses road centrelines, exemplifying the most straightforward geometry for roads and attribute (street name). Figure 4b adds a second geometric dimension (polygon) and slightly more complex attributes, such as street class. Colours differentiate the two street class attributes (boulevard and residential street). Figure 4c differentiates the pedestrian space from the road with varying curb heights. The colour gradient demonstrates the age of the road surface. Figure 4d showcases painted road signage, lane width, and the number of lanes necessary for traffic-related simulations. The colour gradient represents the road quality attribute.

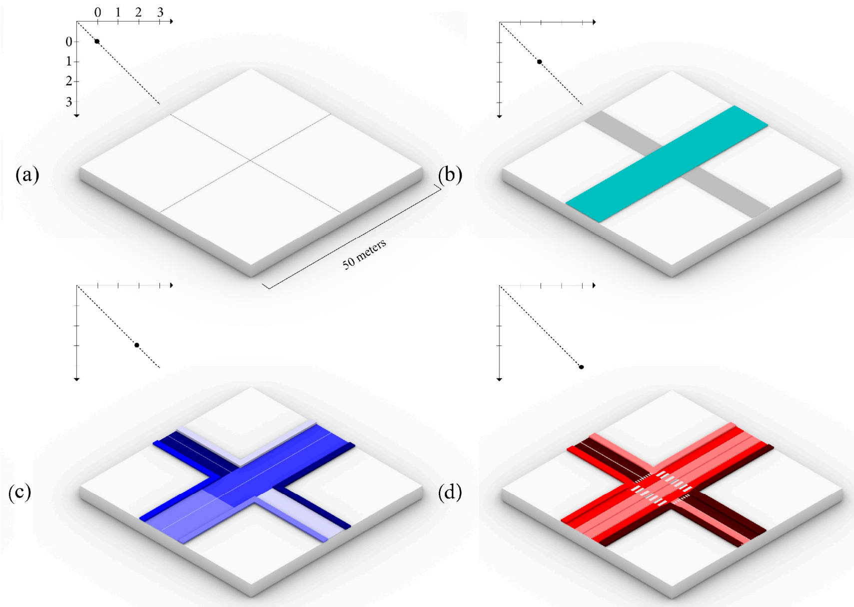

Figure 4. LOD graphical definitions of transportation (roads) of (**a**) LOD 0.0, (**b**) LOD 1.1, (**c**) LOD 2.2 and (**d**) LOD 3.3. In addition to geometric form, different examples of attributes are represented by different colour palettes and gradients. The model spans a 30 by 30 m area to represent the spatial layout of a typical street.

Table 5. Verbal descriptions of road LODs.

LOD	Geometric Definition	Attribute Definition
0	Centreline	Street names
1	Road area and its components are represented by polygons; centrelines may also be included	Street class, address range, length, upcoming construction, posted speeds
2	Road area is further divided into separate components and heights are assigned to curbs, roads, sidewalks, etc., with standardized height estimates	Maintenance scheduling, last replaced, painted road signs, surface material, traffic direction, number of lanes, lane size
3	Road area and its components are extruded to standard heights as experienced in reality	Road surface quality

4.4. Relief (Terrain)

Figure 5 below illustrates a hypothetical relief dataset at three LODs with the attribute examples outlined in Table 6. The graphic depicts an example of a valley relief. As described in the Methods, Figure 5a is absent. Figure 5b represents a 2D relief (polygon) and an attribute, such as elevation above sea level. Colours differentiate the various contour elevations. Figure 5c features extruded contour lines, while different land uses are projected on top in colour. Figure 5d showcases the specific form of the relief, exemplifying geological compositions.

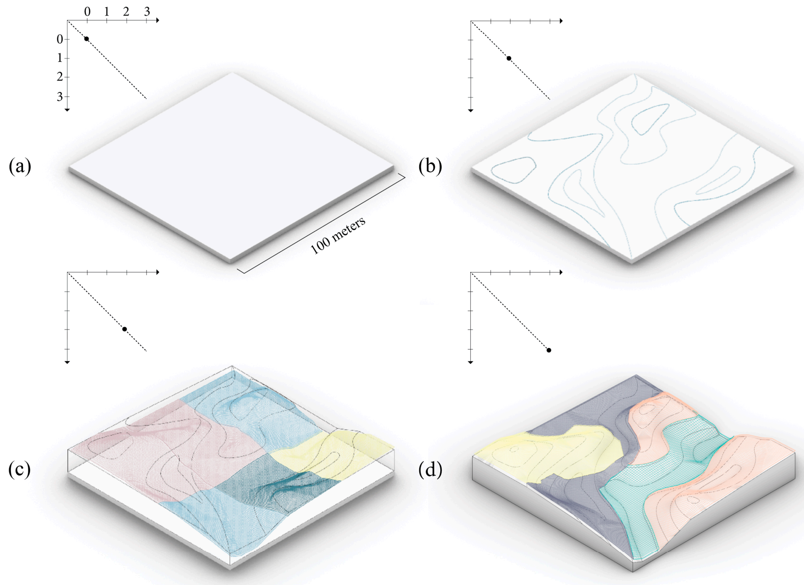

Figure 5. LOD graphical definitions of relief (terrain) of (**a**) LOD 0.0 (not applicable in this case), (**b**) LOD 1.1, (**c**) LOD 2.2 and (**d**) LOD 3.3. In addition to geometric form, different examples of attributes are represented by different colour palettes and gradients. The model represents a 100 m by 100 m area to accurately depict topographical variations and avoid a featureless appearance.

Table 6. Verbal descriptions of relief LODs.

LOD	Geometric Definition	Attribute Definition
0	None	None
1	2D contour curves at a given interval global 100 m (or higher)	Elevation (distance above or below sea level)
2	3D contour curves at a given resolution (national level 10–100 m)	Materials, land use
3	3D contour curves at a maximum resolution, perhaps from LiDAR (local < 1–10 m)	Geologic structure and composition, and the types of processes; land subsidence

4.5. Water Bodies

Figure 6 below illustrates a hypothetical water body dataset at four LODs with the attribute examples outlined in Table 7. The extent (e.g., area and depth) and characteristics (e.g., water quality) of a water body will affect how it appears in a visual application. Figure 6a uses water body centroids and centrelines, exemplifying the most straightforward geometry and attributes (classification). Figure 6b adds a second geometric dimension (polygon), while a slightly more complex attribute (volume) is represented in colour. Figure 6c illustrates a bathymetric estimation, highlighting the three-dimensional form of the water body. Colours demonstrate the water quality attribute in three categories: excellent, good, and poor. Figure 6d showcases specific geometric water body dimensions, while the flood risk attribute is represented in colour in three categories: high, moderate, and low flood risk.

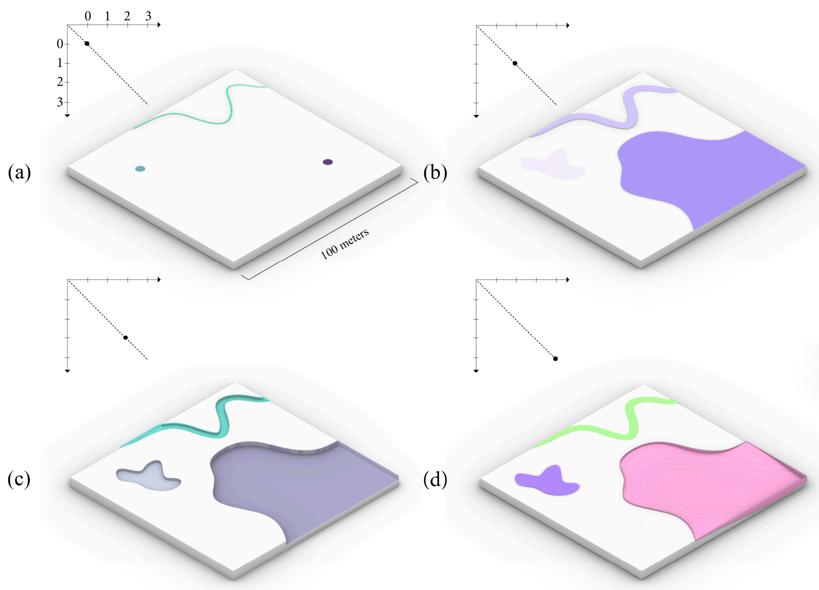

Figure 6. LOD graphic definition of water bodies of (**a**) LOD 0.0, (**b**) LOD 1.1, (**c**) LOD 2.2 and (**d**) LOD 3.3. In addition to geometric form, different examples of attributes are represented by different colour palettes and gradients. The model spans a 100 by 100 m area to accurately depict a permanent water body.

Table 7. Verbal descriptions of water LODs.

LOD	Geometric Definition	Attribute Definition
0	Point location (centroid) or centreline	Name of water body, classification, coordinates
1	Polygons showing the area or outline of a water body or the centrelines of a river	Tributaries, greater watershed, volume, and associated dimensions
2	Water body bathymetric estimation	Water quality data (and associated fields), pollution point sources, flow rate
3	3D volume of entire water body	Ground-water table, flood risk, sinusoidal measurements, alluvial distribution

4.6. CityFurniture

Figure 7 below illustrates a hypothetical city furniture dataset at four LODs with the attribute examples outlined in Table 8. The graphic depicts a hypothetical street space with streetlights, benches, and litter receptacles. Visually speaking, city furniture elements augment the realism to representations of the street space and provide a sense of scale and accessibility at the street level. In terms of analysis, city furniture can inform maintenance cost estimates, upgrade cost estimates, and connectivity with other systems such as water lines and electrical lines. Figure 7a uses city furniture object coordinates, exemplifying the simplest geometry (point) and attributes (coordinates). Colours represent the variations in coordinates. Figure 7b adds a second geometric dimension (polygons), while a slightly more complex attribute (e.g., maintenance) is visualized in colour. Figure 7c illustrates approximate object types, highlighting general shapes and 2.5D forms. Colour indicates the primary material attribute of the city furniture object. Figure 7d showcases specific object shapes, offering more realistic geometric representations necessary for more photorealistic renderings. The colour gradient represents the installation year attribute.

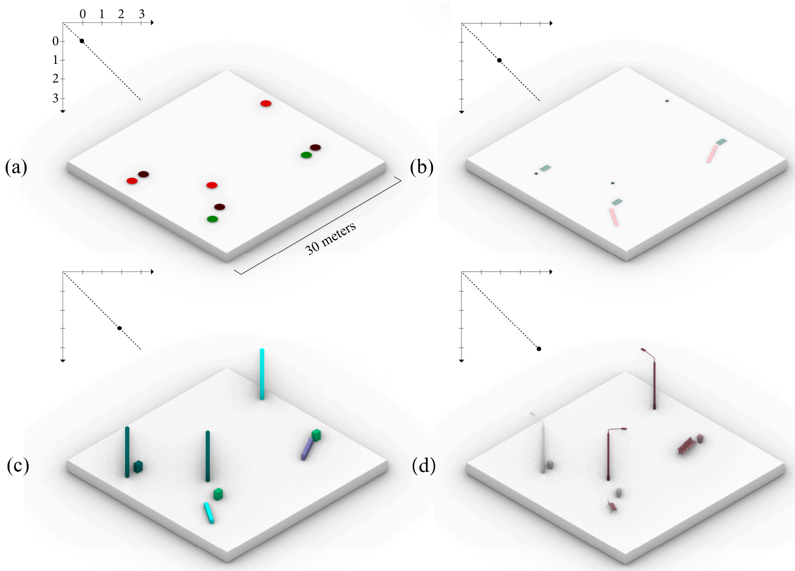

Figure 7. LOD graphical definitions of city furniture of (**a**) LOD 0.0, (**b**) LOD 1.1, (**c**) LOD 2.2 and (**d**) LOD 3.3. In addition to geometric form, different examples of attributes are represented by different colour palettes and gradients. The model spans a 30 by 30 m area to represent the spatial distribution and arrangement of city furniture along a typical street.

Table 8. Verbal descriptions of CityFurniture LODs.

LOD	Geometric Definition	Attribute Definition
0	Point location for city furniture	Location/coordinates, object type
1	Polygons showing length and width of city furniture object	Address of object, neighbourhood, maintenance requirements
2	Estimated 3D boundary of city furniture object	Material, accessibility
3	3D massing replica of city furniture objects—may include underground infrastructure	Installation date, potential replacement date, connectivity with greater systems (water lines, electrical lines)

Challenges were encountered while creating the graphic depictions, particularly with topography, streets, and water body definitions. Specifically, the LOD 0s for topography and streets differed from other LOD 0s because they represent networks that extend beyond the city block. Moreover, the subtle differences between LODs were difficult to depict in a hypothetical city block, as their intricacies exist at varying scales and spatial extents, making it challenging to convey their complexity graphically.

5. Implementation and Case Studies

This study introduces a comprehensive and data-inclusive LOD framework, which emerged from challenges encountered during our ongoing efforts to develop UDT models using open and multi-source spatial datasets at a city scale. While the application of this framework to real-world datasets is an important aspect of our research, it is not included in this manuscript. This is because the proposed LOD framework is designed to be generalized and broadly applicable to diverse UDT projects and focusing on a specific dataset at this stage might limit its versatility. These practical validations and case studies will be addressed in detail in a forthcoming publication, ensuring the framework remains robust and adaptable across different contexts.

To demonstrate the potential application and use cases, the following examples provide two possibilities for modelling with our proposed LOD definitions. We wanted to demonstrate possible use cases by applying our graphic depictions and verbal descriptions of our LOD definitions. Depending on the user's needs, there are many possibilities for 3D modelling scenarios. We wanted to create two different scenarios where different datasets are prioritized to show the contrast between the different combinations of LODs for different cases.

5.1. Scenario 1

This scenario (See Figure 8) demonstrates a possible representation of an urban tree canopy density model. In this scenario, the tree dataset is represented by LOD 2, as it is necessary to visualize the canopy location, size (height and coverage), and species of each tree. Building and road datasets are represented in LOD 1 as it is necessary to understand their position concerning the trees. However, more detailed models of these datasets are not required as their volume is not required for an urban tree canopy density model. The presence of terrain, water bodies, and city furniture datasets are not integral to this 3D model; therefore, they have been left out in this example.

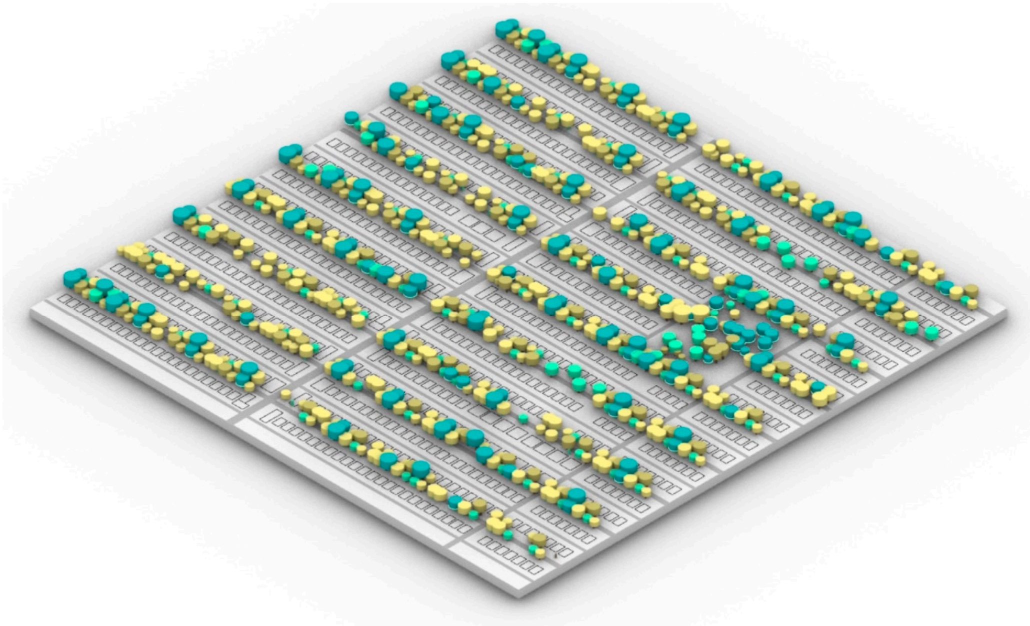

Figure 8. An example of an urban canopy density model featuring multiple tree types.

This urban tree canopy density model could be used to determine areas where more significant heat may be experienced in the street space. The presence of trees in a 3D model could also indicate the quality of spatial experience or environmental quality. Tree quantity could inform the amount of maintenance required in specific neighbourhoods.

5.2. Scenario 2

This scenario (See Figure 9) demonstrates a possible representation of building use and its relationship to neighbourhood density. In this scenario, the LOD 1 of the road dataset is used to connect neighbourhoods and represent density. The LOD 2 of building datasets illustrates building volume, density, and classification, with different colours indicating building uses. The presence of trees, terrain, water bodies, and city furniture datasets are not integral to this 3D model; therefore, they have been left out.

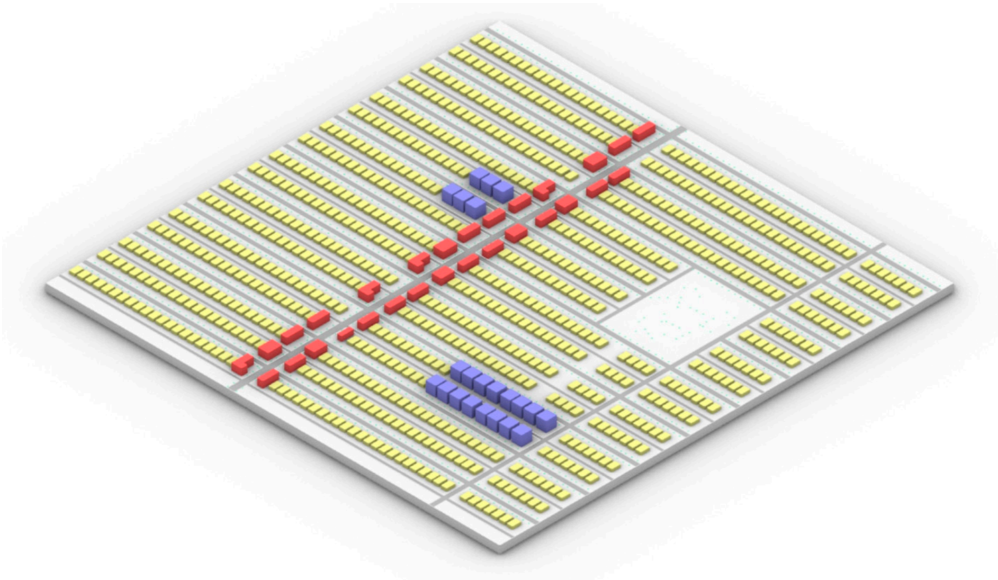

Figure 9. An example of a building density distribution model featuring multiple building types.

This building density distribution model could inform urban planners which areas of the city require specific services or would benefit from more infill. Based on the types of buildings or the frequency of their use, the model could also indicate areas where additional traffic controls or road improvements may be needed. Known as a shoebox model [25], buildings at this LOD are more computationally efficient and can generate comparable energy simulation results at scale [26]. More recently, [27] also developed a methodological framework that enables the connection between building geometric data and non-geometric attributes in a geographic information system (GIS), enabling more comprehensive urban-scale building energy modelling as well as affordability concerns.

Appendix A provides access to the 3D files (.stl format) used in creating individual LOD illustrations (Section 4) as well as combined scenarios (Section 5). We made these files available for readers so that they can visualize their own tiled urban 3D models for their specific needs and LODs.

6. Discussion

We reviewed the existing literature associated with LODs and identified shortcomings (Table 1) associated with LOD definitions in the context of data-driven 3D UDT models. We established a universal geometric and attribute LOD framework that extends beyond only building data. While we cannot provide exact quantitative measurements in this study to demonstrate how the proposed LODs can improve UDT modelling accuracy and/or efficiency, this work achieved the following:

- Outlining key considerations when defining LODs for common urban elements;
- Providing a framework integrating both geometric and attribute data;
- Allowing users to compare and evaluate the quality and feasibility of a dataset;
- Aligning data LODs with UDT modelling needs in efficient and effective large-scale and urban-scale UDTs.

Integrating the proposed LOD framework into existing tools and workflows, such as GIS software and urban modelling platforms, offers the potential to enhance their utility and adaptability in urban design and planning processes. Features such as dynamic LOD filtering could allow users to visualize and analyze data at appropriate levels of abstraction, optimizing the balance between computational efficiency and data precision. This would support workflows that involve large-scale urban projects or complex multi-source datasets, where inconsistent LODs often pose challenges. For example, the recently released 3D BAG dataset [28] allows the user to choose a specific geometric LOD depending on their project's needs or computational limits. Platforms such as Autodesk InfraWorks, Rhino with Grasshopper, or building information model (BIM) tools (e.g., Revit) could adopt this framework to better align LOD classifications across urban elements beyond just buildings. This would ensure a standardized approach to defining levels of detail for buildings, terrain, vegetation, and infrastructure within simulation and design processes. For instance, modellers could define or adjust the LOD requirements for specific project phases—such as preliminary conceptualization versus detailed construction planning—using the framework as a reference.

Typically, municipal governments manage and host open data portals, including the acquisition, quality, and frequency of updates of data. Data made available or data withheld, in their quality and temporality, could be politically motivated [29]. Regarding quality, a municipal government may restrict access to a higher LOD of a dataset, such as utility infrastructure, where the cost of obtaining and maintaining the data is greater than that of a lower LOD. A high LOD, both in its geometric details and/or attribute datasets, may be limited to internal department use only and not be publicly available. Regarding temporality, a government that leans more conservatively may not see the value in frequently refreshing specific datasets, such as tree or road inventories, thus limiting updates to, for example, every four years instead of every four months. Academic organizations may require specific data for different studies that may not be available through a government portal or OSM. Data made available by local governments may also be of low granularity. If an academic study requires data with higher fidelity, it may be collected independently. These data could be available to others within the academic organization, but others outside would not be able to benefit. If these data become publicly available, they may not be as easily accessible or available in as many formats as open data from government portals.

Crowd-sourced data and citizen scientists are means of obtaining less unbiased data or data with little to no political affiliation. Institutions, such as the Open Data Inventory [30], evaluate the data openness of different governments worldwide. Another example is the Global Open Data Index [31]. Both organizations consider datasets that involve social and economic statistics and environmental data. Datasets within these categories are evaluated based on coverage and openness and given an overall score. Openness is defined by Open Knowledge in terms of data and content as "anyone can freely access, use, modify, and share for any purpose". The evaluation of data openness by Open Data Watch occurs at the national scale, focusing on data supplied by national governments and organizations. Currently, no organizations or services rank and compare the quality of open data made available at the urban scale. To understand the openness of urban data within a country, we can extrapolate based on the country's score.

Bijecki et al. [32] discuss the quality of crowd-sourced open data in the context of building information from OpenStreetMap (OSM). They acknowledge OSM's benefits of accessibility and inclusivity but also its drawbacks in maintaining uniform quality. The primary disadvantage of the service is scattered and voluntary data inputs. One city may benefit from a diligent and enthusiastic contributor, while another city may not [32].

One potential future work topic is to define a keystone dataset containing data that can potentially benefit multiple disciplines or departments and needs of research and practices. Merriam–Webster defines a keystone as "something on which associated things depend for support." Coined by ecologist Robert Paine in 1969 [33], the idea of a keystone in a scientific field tends to be associated with keystone species in ecology and conservation.

Architecturally and structurally speaking, a keystone often refers to a stone block located at the pinnacle of a Roman-style archway [34]. A keystone species is a species that is recognized as having a broadly reaching, cascading effect on associated species within an ecosystem, intertwined at all levels of interaction [35]. Although the keystone in the archway tends to appear decorative, it is structurally integral to supporting the physics of the arch and any other adjacent structural forces. This concept of an integral and dependent unit can also fit into the field of data science as a keystone dataset. Identifying keystone datasets would offer insights into data infrastructure prioritization, particularly for cities with limited resources. This would tell governments which datasets to prioritize. Keystone datasets are important because they provide recommendations for data investments for cities and governments to improve data infrastructure—collecting, analyzing, storing, distributing, and maintaining data is costly.

As real-time open data become more readily available for use in UDTs [35], it will be necessary to incorporate this temporal quality within the LOD definition framework. Another research topic furthering this work therefore concerns data temporality: a variable of time based on the attributes of a dataset to be updated or rebuilt. Temporality refers to the frequency of how often a dataset is updated. The frequency of dataset updates may vary depending on the magnitude and rate of change of the given event for which the dataset is built. Buildings may experience renovations and retrofitting. Trees will grow each season, changing their trunk diameter, canopy width and height, in addition to their ever-changing physiological attributes between seasons and years. New trees will be planted, and some trees may be cut down. For certain attributes, trees may remain relatively the same. For instance, the species of an existing tree should not change, although maintenance schedules may. In terms of temporality, real-time data are as close to realistic as possible. The use of real-time data in UDT is one of their main benefits. The author of [36] utilizes real-time data to create UDT transportation and traffic-related simulations. The authors of [37] have suggested the inclusion of sensors linked to the Internet of Things (IoT) to provide real-time data about the public use and occupation of a public space, for example. Although this definition framework may work well in theory, it is a different matter of how it may be incorporated in practice. The authors of [38] specified that the definition's results should seek validation from procedures widely accepted by the community. The temporal quality may require adaptation within the current framework. This adaptation could include a refresh rate for some datasets or a ratio or scale for growth or modification. This adaptation could apply to either geometric or attribute data.

7. Conclusions

In this paper, we propose a new family of level of detail (LOD) definitions that integrate geometric complexity with the challenges of data acquisition, archiving, processing, and distribution. Our LOD definitions extend beyond buildings to encompass trees, roads, terrain, water bodies, and city furniture, addressing critical gaps in existing standards. By distinguishing between geometric and attribute LODs, we offer a flexible framework that can be universally applied across datasets, enabling more consistent, precise, and information-rich urban models. This work highlights the importance of incorporating both static and real-time datasets in urban digital twin (UDT) applications. The proposed LOD framework not only facilitates better data infrastructure evaluation but also supports strategic decision-making for urban planning, sustainability, and resource management.

Looking forward, we see significant potential in applying this framework to diverse disciplines and real-world projects. The concept of keystone datasets, introduced here, could guide cities and governments in prioritizing data investments to maximize impact across multiple domains. Further research into real-time data integration and the temporality of datasets will enhance the framework's adaptability to rapidly changing urban environments. By bridging the gaps in existing LOD standards, this framework lays the groundwork for future innovations in urban modelling, offering a path toward more.

Author Contributions: Conceptualization, Y.L.; methodology, B.B. and Y.L.; formal analysis, B.B. and Z.P.; investigation, B.B.; resources, Y.L.; data curation, M.G.; writing—original draft preparation, B.B.; writing—review and editing, B.B., M.G., Z.P. and Y.L.; supervision, Y.L.; project administration, Y.L.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the University of Manitoba's URGP (Early Career Stream) and the Faculty of Architecture. The APC was funded by the Faculty of Architecture, University of Manitoba.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: We thank the University of Manitoba's URGP (Early Career Stream) and the Faculty of Architecture for funding this research. We also appreciate the valuable comments and suggestions provided by all the reviewers. Additionally, we would like to acknowledge Filip Bijecki (National University of Singapore) for generously offering his time and insights during the early stages of this project.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

LOD 3D Models (Rhino, ~75 MB) Available at: https://object-arbutus.cloud.computecanada.ca/fes_dataShare/LOD_Models/LOD_CaseStudy_Model.zip (accessed on 10 October 2024).

References

1. Manandhar, P.; Rafiq, H.; Rodriguez-Ubinas, E. Current status, challenges, and prospects of data-driven urban energy modeling: A review of machine learning methods. *Energy Rep.* **2023**, *9*, 2757–2776. [CrossRef]
2. Acosta, M.P.; Vahdatikhaki, F.; Santos, J.; Hammad, A.; Dorée, A.G. How to bring UHI to the urban planning table? A data-driven modeling approach. *Sustain. Cities Soc.* **2021**, *71*, 102948. [CrossRef]
3. Batty, M.; Milton, R. A new framework for very large-scale urban modelling. *Urban Stud.* **2021**, *58*, 3071–3094. [CrossRef]
4. Schrotter, G.; Hürzeler, C. The Digital Twin of the City of Zurich for Urban Planning. *PFG J. Photogramm. Remote Sens. Geoinf. Sci.* **2020**, *88*, 99–112. [CrossRef]
5. *Virtual Singapore*; Singapore Land Authority (SLA): Singapore.
6. Brooks, R.J.; Tobias, A.M. Choosing the best model: Level of detail, complexity, and model performance. *Math. Comput. Model.* **1996**, *24*, 1–14. [CrossRef]
7. Law, A.M. Simulation-models level of detail determines effectiveness. *Ind. Eng.* **1991**, *23*, 16.
8. Abualdenien, J.; Borrmann, A. Levels of detail, development; definition, and information need: A critical literature review. *J. Inf. Technol. Constr.* **2022**, *27*, 363–392. [CrossRef]
9. Floros, G.; Pispidikis, I.; Dimopoulou, E. Investigating integration capabilities between IFC and CityGML LOD 3 for 3D city modelling. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *XLII-4/W7*, 1–6. [CrossRef]
10. Ferré-Bigorra, J.; Casals, M.; Gangolells, M. The adoption of urban digital twins. *Cities* **2022**, *131*, 103905. [CrossRef]
11. Ohori, K.A.; Biljecki, F.; Kumar, K.; Ledoux, H.; Stoter, J. Modeling cities and landscapes in 3D with CityGML. In *Building Information Modeling: Technology Foundations and Industry Practice*; Borrmann, A., König, M., Koch, C., Beetz, K., Eds.; Springer: Cham, Switzerland, 2018; pp. 199–215. [CrossRef]
12. Kolbe, T.H.; Groger, G.; Plumer, L. City Geography Markup Language. 2006. Available online: https://www.ogc.org/standard/gml/ (accessed on 10 October 2024).
13. Kumar, K.; Labetski, A.; Ledoux, H.; Stoter, J. An improved LOD framework for the terrains in 3D city models. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *4*, 75–82. [CrossRef]

14. Kolbe, T.H.; Kutzner, T.; Smyth, C.S.; Nagel, C.; Roensdorf, C.; Heazel, C. OGC City Geography Markup Language (Citygml) Part 1: Conceptual Model Standard. 2021. Available online: https://doi.org/10.62973/20-010 (accessed on 10 October 2024).
15. Biljecki, F.; Ledoux, H.; Stoter, J. An improved LOD specification for 3D building models. *Comput. Environ. Urban Syst.* **2016**, *59*, 25–37. [CrossRef]
16. Ortega-Córdova, L. Urban Vegetation Modeling 3D Levels of Detail. Master's Thesis, Delft University of Technology, Delft, The Netherlands, 2018.
17. Boersma, F. Modelling Different Levels of Detail of Roads and Intersections in 3D City Models. Master's Thesis, Delft University of Technology, Delft, The Netherlands, 2019.
18. Beil, C.; Kolbe, T.H. CityGML and the streets of New York-A proposal for detailed street space modelling. In Proceedings of the 12th International 3D Geoinfo Conference, Melbourne, Australia, 26–27 October 2017; Volume 2017, pp. 9–16.
19. Labetski, A.; van Gerwen, S.; Tamminga, G.; Ledoux, H.; Stoter, J. A proposal for an improved transportation model in CityGML. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-4/W10*, 89–96. [CrossRef]
20. Beil, C.; Ruhdorfer, R.; Coduro, T.; Kolbe, T.H. Detailed streetspace modelling for multiple applications: Discussions on the proposed CityGML 3.0 transportation model. *ISPRS Int. J. Geoinf.* **2020**, *9*, 603. [CrossRef]
21. Li, Q.; Unger, A.J.A.; Sudicky, E.A.; Kassenaar, D.; Wexler, E.J.; Shikaze, S. Simulating the multi-seasonal response of a large-scale watershed with a 3D physically-based hydrologic model. *J. Hydrol.* **2008**, *357*, 317–336. [CrossRef]
22. Ghaith, M.; Yosri, A.; El-Dakhakhni, W. Digital twin: A city-scale flood imitation framework. In Proceedings of the Canadian Society of Civil Engineering Annual Conference, Montreal, QC, Canada, 26–29 May 2021; pp. 577–588.
23. Fernández-Arango, D.; Varela-García, F.A.; González-Aguilera, D.; Lagüela-López, S. Automatic generation of urban road 3D models for pedestrian studies from LiDAR data. *Remote Sens.* **2022**, *14*, 1102. [CrossRef]
24. Stoter, J.; Vosselman, G.; Goos, J.; Zlatanova, S.; Verbree, E.; Klooster, R.; Reuvers, M. Towards a national 3D spatial data infrastructure: Case of the Netherlands. *Photogramm. Fernerkund. Geoinf.* **2011**, *2011*, 405–420. [CrossRef] [PubMed]
25. Reinhart, C.; Dogan, T.; Jakubiec, A.; Rakha, T.; Sang, A. UMI–an urban simulation environment for building energy use, daylighting and walkability. *Build. Simul.* **2013**, *2013*, 476–483.
26. Dogan, T.; Reinhart, C. Shoeboxer: An algorithm for abstracted rapid multi-zone urban building energy model generation and simulation. *Energy Build.* **2017**, *140*, 140–153. [CrossRef]
27. Keena, N.; Friedman, A.; Parsaee, M.; Klein, A. Data Visualization for a Circular Economy: Designing a Web Application for Sustainable Housing. *Technol.|Archit. + Des.* **2023**, *7*, 262–281. [CrossRef]
28. 3D BAG. *Documentation: 3D Basisregistratie Adressen en Gebouwen (BAG)*; 3D BAG: Delft, The Netherlands, 2024.
29. Kitchin, R.; McArdle, G. Urban data and city dashboards: Six key issues. In *Data and the City*; Routledge: London, UK, 2017; pp. 111–126.
30. Open Data Inventory. Open Data Watch: Global Rankings; Open Data Inventory. Available online: https://odin.opendatawatch.com/report/rankings (accessed on 10 October 2024).
31. Open Knowledge. Global Open Data Index: Places; Open Knowledge. Available online: http://index.okfn.org/place.html (accessed on 10 October 2024).
32. Biljecki, F.; Chow, Y.S.; Lee, K. Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes. *Build. Environ.* **2023**, *237*, 110295. [CrossRef]
33. Paine, R.T. A Conversation on Refining the Concept of Keystone Species. *Conserv. Biol.* **1995**, *9*, 962–964. Available online: http://www.jstor.org/stable/2387008 (accessed on 10 October 2024). [CrossRef]
34. Ching, F.D.K. *A Visual Dictionary of Architecture*; John Wiley: Hoboken, NJ, USA, 1995.
35. Farsi, M.; Daneshkhah, A.; Hosseinian-Far, A.; Jahankhani, H. *Digital Twin Technologies and Smart Cities*; Springer Nature: London, UK, 2020.
36. Argota Sánchez-Vaquerizo, J. Getting real: The challenge of building and validating a large-scale digital twin of Barcelona's traffic with empirical data. *ISPRS Int. J. Geoinf.* **2022**, *11*, 24. [CrossRef]
37. Mendula, M.; Bujari, A.; Foschini, L.; Bellavista, P. A data-driven digital twin for urban activity monitoring. In Proceedings of the 2022 IEEE Symposium on Computers and Communications (ISCC), Rhodes, Greece, 30 June–3 July 2022; pp. 1–6.
38. Dong, L.; Duarte, F.; Duranton, G.; Santi, P.; Barthelemy, M.; Batty, M.; Bettencourt, L.; Goodchild, M.; Hack, G.; Liu, Y.; et al. Defining a city—Delineating urban areas using cell-phone data. *Nat. Cities* **2024**, *1*, 117–125. [CrossRef]