buildings

MDPI

Article

Building Digital Twin Data Model Based on Public Data

Dawoon Jeong, **Changyun Lee, Youngmin Choi and Taeyun Jeong** *

Korea Land and Geospatial Informatix Corporation (LX), Department of Digital Twin,
Jeonju 35244, Republic of Korea; dawoon@lx.or.kr (D.J.)
* Correspondence: tiger@lx.or.kr

Abstract: This study aims to propose a method for constructing basic digital twin data in South Korea by adhering to international standards and by utilizing publicly available data. Specifically, the study focuses on designing and proposing a digital twin data model for buildings, as building-related digital twin data are the most applicable among the basic digital twin data. To achieve this, the first section provides essential background information, introduces concepts and requirements related to basic digital twin data, and offers a brief overview of City Geography Markup Language (CityGML). The second section explains the methodology and the data used in this study. The third section presents the main findings: the selection of public data (building data) for constructing basic digital twin data, the mapping process using CityGML, and the creation of Unified Modeling Language (UML) diagrams. The fourth section discusses these findings. Finally, the conclusion and recommendations for future research are provided. This approach enhances the accuracy of building-related digital twin data and supports the use of digital twin services in both public and private sectors by enabling various spatial analyses.

Keywords: 3D data; ADE; building; CityGML; digital twin; public data; spatial data; UML

check for updates

Citation: Jeong, D.; Lee, C.; Choi, Y.; Jeong, T. Building Digital Twin Data Model Based on Public Data. *Buildings* **2024**, *14*, 2911. https://doi.org/10.3390/buildings14092911

Academic Editor: Sehyun Park

Received: 7 August 2024
Revised: 3 September 2024
Accepted: 13 September 2024
Published: 14 September 2024

1. Introduction

Research on digital twins has been ongoing both domestically and internationally, with various definitions proposed by international organizations and researchers. One study [1] defined a digital twin as "a virtual representation of a physical object used to optimize performance using real-time data". The European Commission's Joint Research Centre (JRC) Technical Report [2] categorizes these definitions into standard, industry, and scientific literature. After comparing and reviewing these definitions, common and key terms such as "virtual representation", "physical object", "real-time data", and "optimizing performance" were identified [3]. Consequently, digital twins can be categorized by "semantics", "means", and "purpose".

Digital twins are applied across a wide range of fields. Nativi et al. [2] explored their use in manufacturing, energy, smart cities, farming, building, and healthcare. Rasheed et al. [4] reviewed their application in health, manufacturing process technology, meteorology, education, urban areas, transportation, and energy sectors. Fuller et al. [5] focused on their application in smart cities, manufacturing, and healthcare, while Barricelli et al. [6] highlighted their use in manufacturing, aviation, and healthcare, as shown in Table 1. Based on this discussion, the manufacturing sector is a prominent field where digital twins are utilized. Additionally, digital twins are frequently employed in urban areas, healthcare, and medical sectors. For instance, some studies suggest that Michael Graves [7] first introduced the concept of the digital twin in 2002 in the context of product lifecycle management. From another perspective, transportation and energy may also be considered part of the urban sector.

Table 1. Key application areas of digital twin technology.

Author	Areas
Nativi et al. (2020: 12)	Manufacturing, energy, smart city, farming, building, and healthcare
Rasheed et al. (2019: 3–5)	Health, manufacturing, process technology, meteorology, education, city, transportation, and energy sectors
Fuller et al. (2020: 108954–108955)	Smart city, manufacturing, and healthcare
Barricelli et al. (2019: 167660)	Manufacturing, aviation, and healthcare

Source: Nativi, S. et al. [2]; Rasheed, A. et al. [4]; Fuller, A. et al. [5]; Barricelli, B.R. et al. [6].

Based on the definitions and application fields of digital twins, it can be concluded that digital twin data must be both efficiently used and trustworthy, as digital twins are applied across various domains. In other words, the success of digital twin projects depends significantly on the effective use of digital twin data.

A common obstacle in data utilization is a lack of understanding of the data and its applications, which hinders active data use in other fields or applications. Specifically, smart city and digital twin projects in South Korea aim to collect data for specific areas or fields to achieve particular objectives or services. However, the interoperability of data across various services remains weak. To address these limitations, it is essential to establish a method for actively sharing and utilizing digital twin data based on international standards. Additionally, a plan for developing digital twin data should be devised by utilizing publicly available data in conjunction with adherence to international standards.

In this study, basic digital twin data is defined as follows.

The basic data for digital twins should support interoperability between heterogeneous systems without constraints and be provided in a machine-readable format with a well-established foundation. In other words, it is essential to fundamentally support an environment where data exchange between different platforms and systems can occur bidirectionally. Additionally, accessibility should be ensured to allow easy manipulation of digital twin data by all users. This can be described in detail, as shown in Table 2.

Table 2. Requirements for basic data in digital twin technology.

Classification	Description
Interoperability	The basic data for digital twins should comply with international standards, enabling the data to be shared, provided, and supported in a format that allows exchange and processing between diverse systems.
Machine-readable	The basic data for digital twins should be available in a machine-readable format, facilitating users' ability to classify, process, and extract the data without requiring additional processing steps.
Accessibility	The basic data for digital twins should be created using publicly available materials, such as public data, to enable ordinary users to construct digital twin data.

Therefore, this study aims to propose a method for constructing basic digital twin data in South Korea by adhering to international standards and utilizing publicly available data. Specifically, the study focuses on designing and proposing a digital twin data model for buildings, as building-related digital twin data are the most applicable among the basic digital twin data. Unlike previous studies, this research extends international data model standards centered on publicly accessible data, making it easier for public institutions, private enterprises, and citizens to construct data based on the actual data model, thereby enhancing its effectiveness and usability. This study is expected to have significant academic implications in this regard. Additionally, it aims to optimize the understanding, access, management, and updating of basic digital twin data to support effective application and implementation in smart city and digital twin projects in South Korea.

The structure of this study is as follows: The first section explains the methodology and data used in this study. The second section provides essential background information, introduces concepts and requirements related to Application Domain Extension (ADE), and offers a brief overview of data models. The third section presents the main findings: the selection of public data (building data) for constructing basic digital twin data, the mapping process using CityGML, and the creation of UML diagrams. The fourth section discusses these findings. Finally, the conclusion and recommendations for future research are provided.

2. Materials and Methods

This study establishes the fundamental guidelines for the design of a building digital twin data model as follows. The key approach is to emphasize the importance of "data standardization" and to aim for a data-centric digital twin platform. The goal is to transition from the current data ecosystem—characterized by non-real-time, database-centric, and standardized data in specific areas—to one that enables the creation of new value and services through the convergence of real-time data (e.g., sensor data) and location data from various areas of the city. Developing smart city-related services in the country requires a primary understanding of data, with the challenge being to process the data structure (format) to suit its intended use. Currently, poor data interoperability between various domains (or services) limits the sharing, distribution, and utilization of application services and private data.

This study aims to create a digital twin data ecosystem that addresses these issues and meets the diverse demands of smart cities by enabling the activation of the smart city/digital twin data domain (such as buildings). The basic digital twin data serves as a conceptual tool for digitizing and constructing a database of indoor and outdoor, underground and aboveground, static and dynamic objects and phenomena in the real world.

To ensure consistent and systematic construction of basic digital twin data, this study establishes guidelines based on four key principles: continuous updating, multi-perspective support, scalability assurance, and standardization. The first principle is the continuous renewal of configuration. The identification and construction of objects required for the digital twin cannot be completed in a single operation. These objects need to be continuously updated and added based on changes in the urban environment and advancements in application technology. Therefore, it is necessary to create a feature catalog to facilitate the ongoing updating and addition of objects required for digital twin services, thus promoting continuous data management. The second principle is configuring multiple perspective support. In cities, the required attributes and representation methods for the same object can vary depending on the area of interest or activity. Consequently, it is essential to consider the diversity of urban environments and address the varying requirements of service users and stakeholders when managing basic digital twin data. The third principle is ensuring scalability in configuration (by defining scalability rules). The basic digital twin data are designed to be open-ended rather than complete, allowing for the continuous expansion of the data model through defined rules for its extension. In other words, the abstract urban data model cannot be finalized at a specific point in time, as it must continuously evolve in response to changes in the urban environment, administrative factors, and emerging technologies. Therefore, ensuring scalability is essential. The fourth principle is standardization-based configuration. In the field of smart cities, a digital twin should visualize diverse modeled data (or connected/converged data) through the creation of spaces and objects. Therefore, adherence to international standards, such as spatial data, is necessary during the pre-data construction process. However, if the content specified by international standards differs from the contextual environment of the digital twin, the standards should be adapted and applied accordingly.

In this study, a method consisting of four steps was used to establish the basic digital twin data. These steps included selecting a reference model, choosing public data, conducting the mapping process using the international standard CityGML, and creating UML diagrams, as shown in Figure 1.

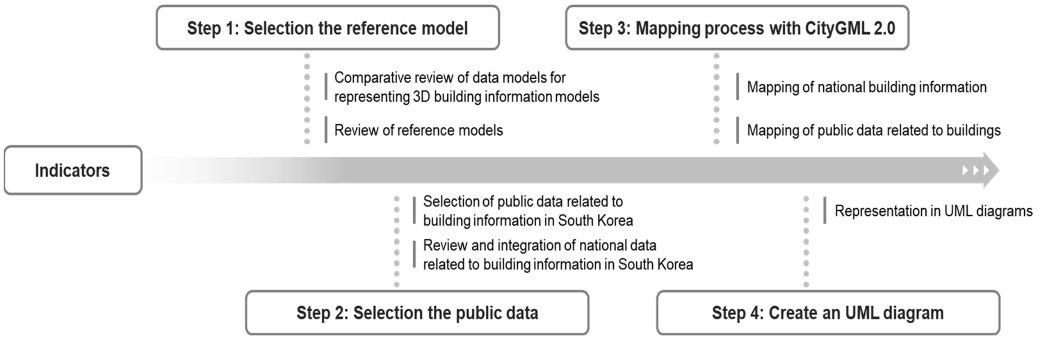

Figure 1. Methodology followed in this paper.

3. Theoretical Background

3.1. Review of Reference Models

In this study, various international standards related to existing data models were reviewed, including the following: Industry Foundation Classes (IFC), Land eXtensible Markup Language (LandXML), 3 Dimensional File Geographic Markup Language (3DF-GML), Keyhole Markup Language (KML), CityGML, and Graphics Library Transmission Format (glTF).

The IFC is an open Building Information Modeling (BIM) standard, recognized as the International Organization for Standardization (ISO) 16739-1:2024 [8] international standard for data models. It is designed to manage, share, and ensure interoperability of information exchanged between applications throughout the lifecycle of a construction project.

LandXML is a data model suitable for modeling information from a civil engineering perspective, such as linear, longitudinal, and cross-sectional data related to roads and bridges, rather than urban spatial dimensions.

3DF-GML is a standard for effectively storing and exchanging 3D national spatial data constructed in Korea. Currently, as the integration with indoor spatial information becomes increasingly important, 3DF-GML faces limitations in supporting different Levels of Detail (LoDs) for such applications.

KML is an XML-based markup language developed to store, represent, and share geospatial information on Google Earth and Google Maps. While KML implements zoom and pan features to represent continuous levels of detail, it is characterized by the technique of combining 2.5D planar objects with terrain to derive three-dimensional representations.

CityGML is an open data model by the Open Geospatial Consortium (OGC) for storing and exchanging 3D urban models based on the GML. It includes four key aspects—geometry, topology, semantics, and appearance—of virtual 3D city models, and is structured as an extensible application schema. Moreover, CityGML comprehensively addresses various urban objects, including buildings, transportation infrastructures, bridges, and water bodies, enabling efficient modeling of 3D objects at both macro and micro scales.

glTF is a format introduced by the Khronos Group, a leading consortium in the graphics industry. It provides standardized tools and functions to facilitate the import/export of assets across various graphics libraries, including OpenGL, OpenGL Embedded Systems (ES), and WebGL.

3.2. Prior Research

This section reviews application examples and previous research related to the OGC's ADE technique, which is the primary methodology of this study.

The Indoor ADE developed in South Korea is a model focused on enhancing the indoor characteristics of CityGML [9]. This ADE provides models in two aspects: space and facilities within the building module. The former includes spatial features such as meeting rooms, while the latter is designed from the perspective of facilities like fire hydrants. One of the key features of this ADE is its extension of CityGML. The LoD of this data is dependent on CityGML's standard features, which are only available at LoD4, indicating its LoD-dependent nature. The study proposed an extended model by defining additional objects and relationships necessary for representing indoor spaces within CityGML's CityObject, Building, and Transportation modules.

LandInfra is a relatively new OGC open standard that integrates the concepts of BIM and GIS for land and infrastructure features [10]. Consequently, it partially overlaps with CityGML, the primary standard in the 3D GIS domain, covering elements such as buildings, roads and railways (transportation in CityGML), and the ground surface (ReliefFeature in CityGML). However, due to the lack of supported software, practical implementation is minimal, which has hindered the effective utilization of LandInfra and InfraGML's advantages. To address this issue and promote the adoption of LandInfra features, an Infra ADE was developed to integrate the concepts of LandInfra into CityGML.

The waterlogging simulation process can be divided into three independent processes: rainfall, surface runoff yield and confluence, and drainage network. This study extended the model within the waterlogging domain by combining the CityGML ADE mechanism [11]. The three modules were designed to store relevant entities and attributes corresponding to the three simulation processes. Additionally, WaterBody and TimeSeries modules were designed to visualize the flooding process and support the storage of dynamic information.

W. Tegtmeier et al. [12] developed the 3D Geotechnical Extension Model (3D-GEM), an ADE aimed at handling information on subsurface geology and underground terrain. This is an example of UML integration that incorporates specifications for discrete LoD. The UML class diagram of CityGML as presented in this study, with the concept of ADE added, is shown below, where the prefix "newlod" was attached to the extensions. Mapping the classes is constrained when aligning the classes of CityGML objects with their elements. For instance, a building's garage and balcony both fall under BuildingInstallation, with no semantic distinction between them. Moreover, they cannot be further decomposed or nested, such as a garage roof being part of the garage itself. This issue was addressed in the study by extending the classes as subclasses of the CityGML class.

4. Results

4.1. Selection of the Reference Model

A digital twin data model should be able to accommodate information from multiple domains and be easily compatible with other 3D data and application systems. Therefore, selecting a reference framework that meets these criteria is crucial. For this purpose, this study selected the most widely used 3D data models across various fields. The models used for analysis include IFC, LandXML, 3DF-GML, KML, CityGML, and glTF, as specified in standard documents and research reports.

The items used for comparison and analysis between models include:

- Geometry, which contains 3D geometric information;
- Topology, which contains topological information;
- Texture, which supports the mapping of real-life captured images onto spatial objects;
- LoD, which indicates the level of detail of 3D objects;
- Semantic information, which reflects the conceptual data model;
- Support for attributes and georeferencing (geo ref.), which includes attributes and georeferencing of data [13], as shown in Table 3.

After comparing and analyzing the 3D data models based on the criteria mentioned, it was found that CityGML 2.0 is the most suitable reference model for basic building digital twin data. CityGML 2.0 supports all of the evaluated criteria and allows for flexible model extensions. Additionally, CityGML comprehensively covers various urban objects, including bridges, tunnels, and roads, beyond just buildings [14], which are the focus of this study. It facilitates efficient modeling of digital twin objects at both macro and micro scales. Furthermore, OGC's official standards, which address international spatial information standards, ensure a high level of reliability.

Table 3. Comparison between 3D data models (adapted from Kim et al., 2018).

Classification	IFC	LandXML	3DF-GML	KML	CityGML 2.0	glTF
Geometry	++	+	+	+	+	+
Topology	+	+	0	-	+	-
Texture	-	+	+	0	+	+
LoD	-	-	0	-	+	-
Semantic	++	0	+	0	++	-
Attributes	+	+	+	0	+	-
Geo ref.	-	0	0	+	+	0

-: not supported; 0: basic; +: supported; ++: extended support.

As previously stated, CityGML is a data exchange format for representing 3D cities and landscapes, independent of any specific application. Sometimes, it may be necessary to extend the CityGML model by adding additional information. There are two main methods to achieve this. The first method involves adding CityGML objects to the application framework. The second method involves incorporating specific information into the CityGML instance document. If the information to be added follows the format defined by the CityGML schema, the second method is more effective. This method is divided into two approaches: one utilizing generic attributes and objects, and another utilizing CityGML ADE. All objects and attributes in CityGML are defined through standards, but the specific names and data types of generic attributes and objects are not explicitly defined. Consequently, there are limitations in ensuring semantic interoperability of the model, as comprehensive attributes and objects cannot be validated through XML. In contrast, ADE offers the advantage of adding information to the application in a more systematic manner.

The Building module of CityGML 2.0 represents various characteristics of a building, including its class, function (residential, public, industrial), purpose, construction year, demolition year, roof type, measured height, and the number of aboveground and underground floors. The building model allows for detailed structuring through different LoDs. In the Building module, the _AbstractBuilding class is central to this representation. The range of LoDs for each geometric and semantic theme within the _AbstractBuilding class is described, as shown in Table 4.

Table 4. Geometric/semantic themes of the _AbstractBuilding class (adapted from OGC, 2012).

Theme	Property Type	LoD0	LoD1	LoD2	LoD3	LoD4
Building footprint and roof edge	gml:MultiSurfaceType	■				
Volume part of the building shell	gml:SolidType		■	■	■	■
Surface part of the building shell	gml:MultiSurfaceType		■	■	■	■
Terrain intersection curve	gml:MultiCurveType		■	■	■	■
Curve part of the building shell	gml:MultiCurveType			■	■	■
Building parts	BuildingPartType		■	■	■	■
Boundary surfaces	AbstractBoundarySurfaceType			■	■	■
Outer building Installations	BuildingInstallationType			■	■	■
Openings	AbstractOpeningType				■	■
Rooms	RoomType					■
Interior building Installations	IntBuildingInstallationType					■

4.2. Selection of the Public Data

To design a building digital twin data model, it is crucial to integrate and extend South Korea-specific fundamental spatial information (2D) attributes into the CityGML 2.0 schema. A key aspect of this process involves applying the ADE method to enable seamless integration between 2D and 3D data. For this purpose, we will utilize the 2D building data model attributes defined in the "National Base Map" established by the National Geographic Information Institute of Korea. These attributes include "Class, Usage, Year of construction, Year of demolition, Measured height, Roof type, Building definition ID, Building name, Building main number, Building sub number, Poi ID, Parcel number, Road name, Road name code, Unique Feature Identifier (UFID), Zip code", among others. Among these, "Class, Usage, Year of construction, Year of demolition, Measured height, Roof type" directly correspond to existing attribute items specified in the CityGML 2.0 Building model, allowing their values to be used without modification. The other attributes, such as "Building definition ID, Building name, Building main number, Building sub number, Poi ID, Parcel number, Road name, Road name code, UFID, Zip code" will be mapped as separate classes within the Building model using the ADE method. Integrating these additional attributes through ADE allows the building digital twin data model to comprehensively represent the unique characteristics and features of Korean buildings, ensuring a seamless connection between 2D and 3D spatial data. This approach enhances the usability and accuracy of the building data model and promotes the standardization and interoperability of geospatial information in urban planning and development contexts [15].

The core of the building digital twin data model lies in its height attributes. Therefore, we plan to utilize the "Height Information Database" established by the National Geographic Information Institute of Korea for these attributes. The attributes defined in the "Height Information Database" include "Building furniture maximum height, Building ground height, Building height, Building maximum height". Since these attributes do not directly match any existing attributes in the CityGML 2.0 Building model, we will apply the ADE method to map them as separate classes within the building model. By integrating these height-related attributes through ADE, the building digital twin data model can accurately represent the vertical dimensions of buildings, which is essential for various urban planning and development applications. This approach ensures that the model is both comprehensive and interoperable, aligning with international standards while addressing specific regional requirements [16].

After examining public data related to the building sector in South Korea for the purpose of designing a building digital twin data model using the CityGML 2.0 Building module, it was determined that the "GIS-Integrated Building Information" provided by the Ministry of Land, Infrastructure, and Transport in South Korea is the most suitable option. "GIS-Integrated Building Information" is a public dataset constructed in spatial data (shp) format, representing both the shape and attribute information of buildings [17]. Furthermore, as designated by South Korea's national core data, it is considered highly suitable for this study, given its accessibility and usability. "GIS-Integrated Building Information" integrates building spatial data based on continuous cadastral polygon information. This dataset includes the following information: "Original drawing ID, GIS building integration identifier, Unique number, Legal district code, Legal district name, Lot number, Special land code, Special land category, Building use code, Building use name, Building structure code, Building structure name, Building area, Date of use approval, Floor area, Land area, Height, Building coverage ratio, Floor area ratio, Building ID, Violation building status, Reference system link key, Data reference date". The design was refined by excluding redundant information and adding only newly expanded items for mapping between CityGML 2.0 building models and GIS-Integrated Building Information: "Usage (ADE), Class (ADE), Use approve date, Area, Land area, Building coverage ratio, Floor area ratio, Building ID, Illegal building, Data date".

4.3. Mapping Process

Before mapping, it is essential to understand the building model within CityGML 2.0. The _AbstractBuilding class plays a pivotal role among all classes, encapsulating characteristics such as building type, function (residential/public/industrial), usage, construction year, demolition year, roof type, measured height, and the number of above- and below-ground floors. Therefore, the extension method proposed in this study using the ADE technique is designed in relation to the _AbstractBuilding class. The Building class represents a single building composed of the same structural segments (roof, floors, entrances, etc.). The BuildingPart class models structural parts of a building, representing building parts that must be associated with either a single building or a building part object. The BuildingInstallation class represents external components that influence the building's external characteristics, such as chimneys, staircases, antennas, balconies, stairs, and roofs. The _BoundarySurface class represents the surfaces constituting the building's interior and exterior, including roofs, grounds, external ceilings, external floors, windows, entrances, floors, internal and external walls, and ceilings. The _Opening class represents entrances on internal and external boundary surfaces, such as chimneys, doors, and windows. The Room class represents the interior rooms of buildings, consisting of entrances, floors, ceilings, and internal walls, and can be represented as LoD4 Solid or MultiSurface. Lastly, the IntBuildingInstallation class is used for permanent, non-movable internal objects attached to the building structure, such as internal staircases, railings, and pipes [18].

The result of the mapping process of the digital twin data model related to buildings is shown in Table 5.

Table 5. Mapping process with geometric/semantic themes of the _AbstractBuilding class.

Theme	Property	Data Type
The fundamental spatial information from South Korea's National Geographic Information Institute (NGII)	- ufid	CharacterString
	- buildingName	CharacterString
	- buildingMainNumber	Integer
	- buildingSubNumber	Integer
	- roadName	CharacterString
	- roadNameCode	CharacterString
	- zipCode	CharacterString
	- parcelNumber	CharacterString
	- buildingGroundHeight	Integer
	- buildingHeight	Integer
	- buildingMaximumHeight	Integer
	- buildingFurnitureMaximumHeight	Integer
GIS-integrated building information	- usageADE	UsageADECode
	- classADE	ClassADECode
	- useApproveDate	Datetime
	- area	CharacterString
	- landArea	Integer
	- buildingCoverageRatio	CharacterString
	- floorAreaRatio	CharacterString
	- buildingId	CharacterString
	- illegalBuidling	Boolean
	- dataDate	Datetime

4.4. Creation of UML

According to ISO 19101-1:2014 [19], spatial information modeling refers to a series of processes that recognize a part of the real world, limited within a discussion domain, as a spatial information application context and define it as a conceptual model. The conceptual model is expressed in the form of a conceptual schema using UML diagrams. The procedures and methods for defining the rules and UML profiles necessary to convert the conceptual model of spatial information into an implementation schema follow the standards of ISO 19103:2015 [20] and ISO 19109:2015 [21].

In this study, as demonstrated in the previous process, we reference the 2D data model attributes included in the NGII's "National Base Map" and "Height Information Database". Additionally, we utilize South Korea's public GIS-Integrated Building Information as an extension element of the building digital twin data model. As previously described, attributes not present in CityGML 2.0 are added as new subclasses to the CityGML 2.0 Building model (_AbstractBuilding) using the ADE method. For attributes with different English labels but identical properties between the extension elements and CityGML 2.0, we standardize them using international standard terms. The data model defined through the mapping between the CityGML 2.0 Building model and the extension elements is illustrated using UML in Figure 2. In Figure 2, the classes highlighted in gray represent the newly extended elements based on the ADE method in the data model.

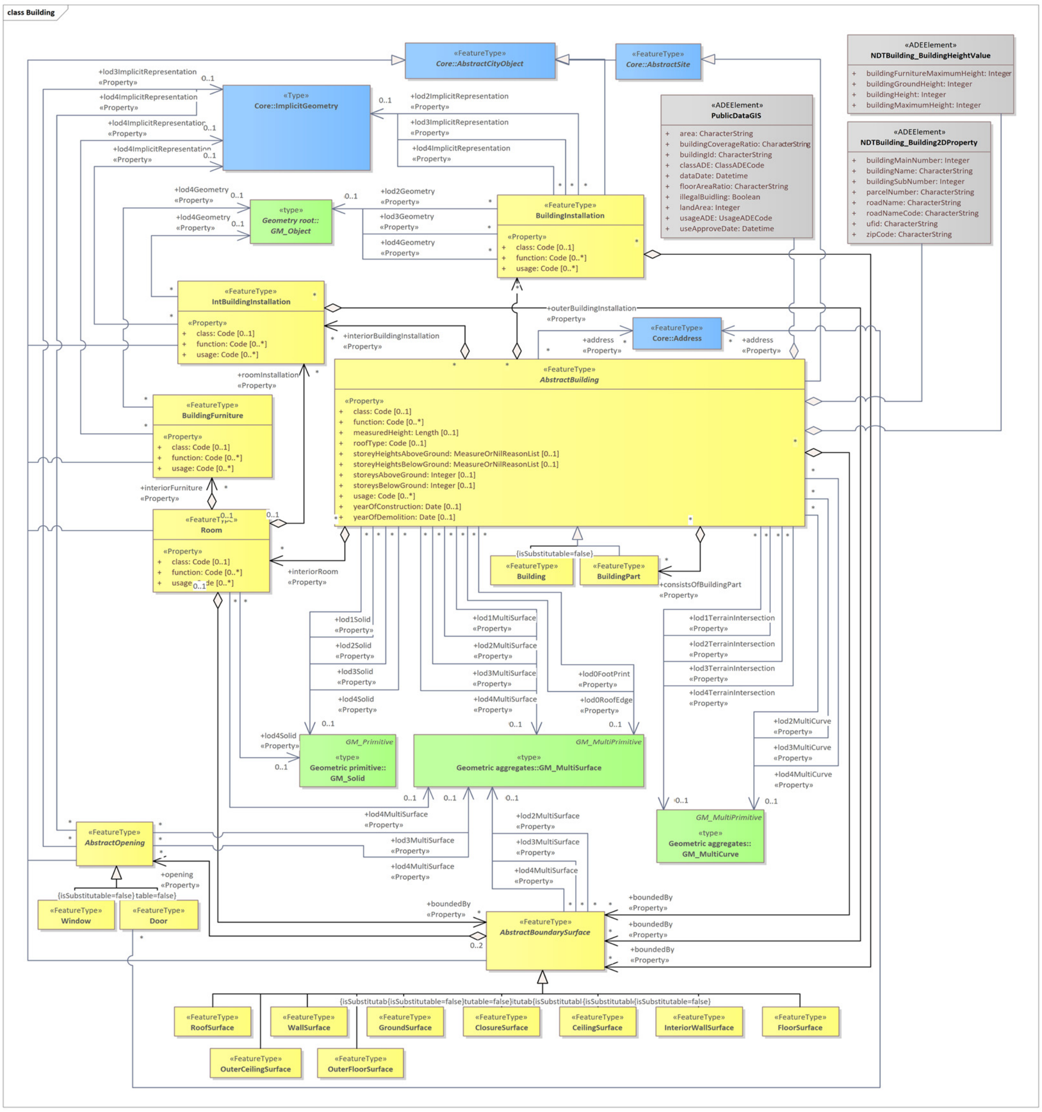

Figure 2. UML diagram of the building digital twin data model.

The detailed examination of the building digital twin data model is as follows. As previously described, attributes included in NGII's "National Base Map" have been extended using the ADE method, resulting in the "NDTBuilding_Building2Dproperty". Additionally, to further extend building height information, the NGII's "Height Information Database" was referenced to expand the CityGML 2.0 Building model, resulting in the "NDTBuilding_BuildingHeightValue", as shown in Figure 3.

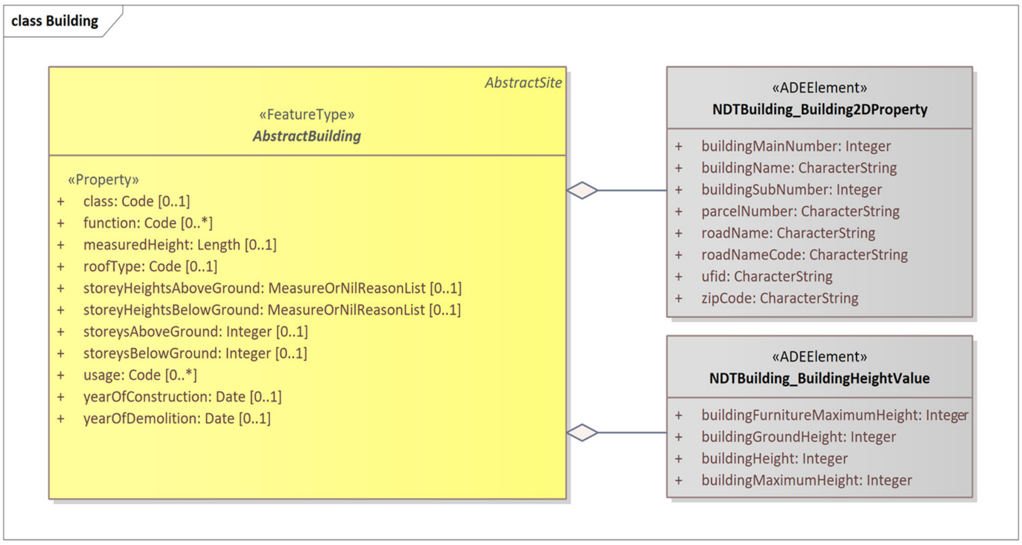

Figure 3. UML diagram of NDTBuilding_Building2Dproperty and NDTBuilding_BuildingHeightValue.

Next, to define the core attribute information of the building digital twin data model based on public data, we have further expanded the attributes related to "GIS-Integrated Building Information" (PublicDataGIS). The detailed contents of this expansion are illustrated in Figure 4.

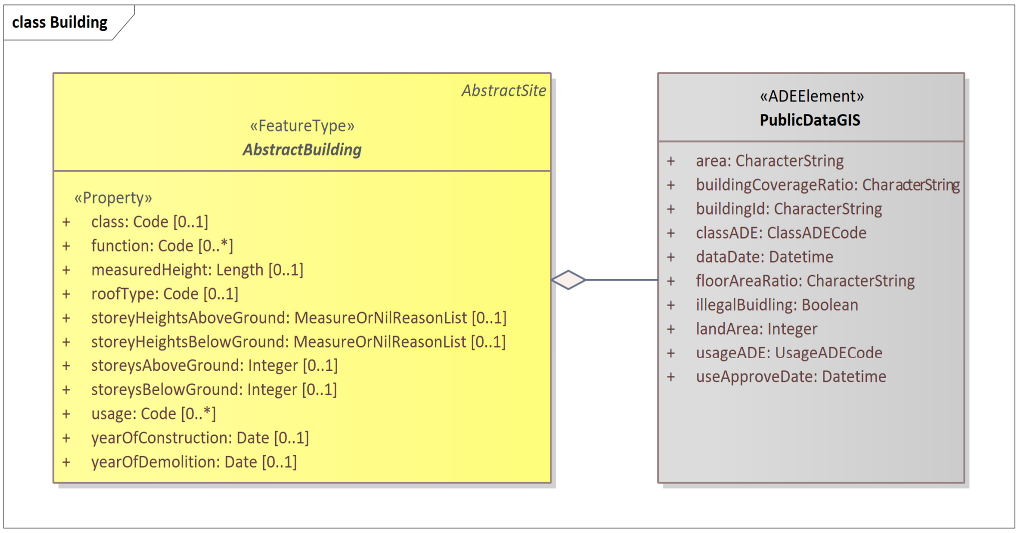

Figure 4. UML diagram of PublicDataGIS.

5. Discussion

The method of defining a digital twin data model allows real-world objects to be represented in various forms in the virtual world. For example, the same real-world object can appear in different forms and have different attributes in the virtual world, depending on the digital twin data model used. These data models influence spatial analysis methods and systems that utilize spatial analysis.

Due to the unique characteristics and broad applicability of digital twin data models, various spatial information-related data models currently exist. In particular, 3D spatial information is expected to be highly versatile compared to 2D spatial information, serving as an important tool for developing noise maps, analyzing 3D urban landscapes, and planning smart cities that connect real and virtual spaces. Recently, there has been a demand for high-precision data and rich semantic information that allows devices to independently recognize spatial situations, particularly in building-related data within digital twin frameworks.

Therefore, establishing principles and proactive design processes for digital twin data models is essential to expand their use and ensure interoperability in building-related applications. In this context, the building digital twin data model proposed in this study is designed by integrating unique aspects such as building attribute information from Korea's fundamental spatial information while adhering to the international standard of OGC's CityGML 2.0. This approach enhances the accuracy of building-related digital twin data and supports the use of digital twin services in both public and private sectors by enabling various spatial analyses.

Following the research procedures, the study presents the following implications:

1. Selection of the reference model: This study utilizes the reference model (CityGML) based on international standards like those from OGC, rather than a vendor-specific model. This allows the creation and use of open data applicable in any country, including South Korea, ensuring interoperability between heterogeneous systems and data.
2. Selection of public data: While the reference model CityGML, established as an international standard, provides guidelines for shape, additional and expanded attribute information tailored to the building conditions of each country (e.g., South Korea) is necessary. This study proposes utilizing publicly available data for attribute information, facilitating easy access, acquisition, processing, and use by both public and private sectors.
3. Mapping process with CityGML 2.0: Using the ADE technique officially guided by CityGML, this study demonstrates how building attribute information can be extended into the CityGML 2.0 Building model. This offers an effective mapping method for practical physical model construction or derivative services in various fields.
4. Creation of UML diagram: By mapping basic spatial information and public data related to buildings in South Korea to the CityGML 2.0 Building model, this study presents a UML diagram for utilizing the building digital twin data model. This simplifies the framework process for building data construction and enhances efficiency in data construction and distribution at the national level.

Alongside these implications, the study acknowledges several research limitations that need further consideration. First, with the advent of more advanced versions like CityGML 3.0, additional measures should be devised to flexibly incorporate these advancements. Second, the added and expanded attribute information in this study primarily pertains to South Korea's basic spatial information and public data. To further enhance usability, expanding the range of ADE to include diverse attribute and shape information related to buildings is necessary. Third, since the version presented in this study corresponds to a conceptual/logical model, there must be a capability to transform it into a physical data model for practical application in building-related systems.

6. Conclusions

Spatial information data models determine how and what real-world spatial information is represented. Therefore, spatial information data models play a crucial role in the visualization and spatial analysis capabilities of systems, with the performance of these systems varying based on the data model used. In this context, while data models for 2D spatial information have been developed in South Korea, there is a notable absence of data models for 3D spatial information, which is anticipated to have higher utility in the future.

Given this background, the objective of this study is to design a building digital twin data model utilizing UML and focusing on buildings, as building-related digital twin data have the highest applicability among digital twin data. This study has designed a building digital twin data model that adheres to international standards, ensures interoperability with South Korea's existing fundamental spatial information and public data, and can be utilized across various fields. Based on the results of this study, it is expected that digital twin data models can be designed and implemented for various domains, such as roads and bridges, in addition to buildings. Furthermore, there is a need to develop and establish these models as official standards, such as national standards (Korean Industrial Standards, KS) or standards by organizations like South Korea's TTA (Telecommunications Technology Association).

However, since the primary objective of this study was to present a building data model at the conceptual and logical levels, there remains a need to further develop this model into a physical model, including the encoding aspect, to facilitate its application and use in practical settings. Additionally, future research should focus on developing specific methodologies (guidelines) for constructing CityGML-based building digital twin data, building upon the data model proposed in this study.

Author Contributions: Conceptualization, D.J., C.L., Y.C. and T.J.; Methodology, D.J. and T.J.; Software, D.J.; Formal analysis, D.J.; Resources, Y.C.; Data curation, D.J.; Writing—original draft, D.J. and T.J.; Writing—review & editing, D.J.; Supervision, C.L.; Project administration, T.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Institute for Information & Communications Technology Promotion (IITP) grant, funded by the Korean government (MSIP) (No. 2022-0-00622, Digital Twin Testbed Establishment).

Data Availability Statement: https://www.data.go.kr/.

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

References

1. Park, J.; Choi, W.; Jeong, T.; Seo, J. Digital twins and land management in South Korea. *Land Use Policy* **2023**, *124*, 106442. [CrossRef]
2. Nativi, S.; Delipetrev, B.; Craglia, M. *Destination Earth*; European Commission JRC Technical Report; Publications Office of the European Union: Luxembourg, 2020.
3. Jeong, D.; Park, J.; Choi, W.; Cho, K.; Hong, S. Comparative analysis of digital twin trends in urban area. *J. Cadastre Land InformatiX* **2023**, *53*, 123–137.
4. Rasheed, A.; San, O.; Kvamsdal, T. Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access* **2020**, *8*, 21980–22012. [CrossRef]
5. Fuller, A.; Fan, Z.; Day, C.; Barlow, C. Digital twin: Ena-bling technologies, challenges and open research. *IEEE Access* **2020**, *8*, 108952–108971. [CrossRef]
6. Barricelli, B.R.; Casiraghi, E.; Fogli, D. A survey on digital twin: Definitions, characteristics, applications, and design implications. *IEEE Access* **2019**, *4*, 167653–167671. [CrossRef]
7. Grieves, M.; Vickers, J. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. *Transdiscipl. Perspect. Complex Syst.* **2016**, *17*, 85–113.
8. *ISO 16739-1:2024*; Industry Foundation Classes (IFC) for data sharing in the construction and facility managemnet industries Part 1: Data schema. International Organization for Standardization (ISO): Geneva, Switzerland, 2024.
9. Kang, H.; Hwang, J.; Lee, J. A Study on the development of indoor spatial data model using CityGML ADE. *J. Korea Spat. Inf. Soc.* **2013**, *21*, 11–21.
10. Kumar, K.; Labetski, A.; Ohori, K.A.; Ledoux, H.; Stoter, J. Harmonising the OGC standards for the built environment: A CityGML extension for LandInfra. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 246. [CrossRef]
11. Shen, J.; Zhou, J.; Zhou, J.; Herman, L.; Reznik, T. Constructing the CityGML ADE for the multi-source data integration of urban flooding. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 359. [CrossRef]
12. Tegtmeier, W.; Zlatanova, S.; Oosterom, P.V.; Hack, H.R. 3D-GEM: Geo-technical extension towards an integrated 3D information model for infrastructural development. *Comput. Geosci.* **2014**, *64*, 126–135. [CrossRef]
13. Kim, B.; Jeong, D.; Hong, S. Constructing 3D geo-spatial model for building data. *J. Korean Soc. Geospat. Inf. Sci.* **2018**, *26*, 57–67.
14. Jeong, D.; Shin, D. A study on development of 3D data model for underground facilities using CityGML ADE. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* **2021**, *39*, 245–252.
15. NGII. *Strengthening the Use of Institutional Standards and Expanding National Standardization*; National Geographic Information Institute Research Report: Suwon, Republic of Korea, 2020.
16. NGII. *Creation of 3D Spatial Information to Establish the Foundation for Digital Twins*; National Geographic Information Institute Research Report: Suwon, Republic of Korea, 2020.
17. Public Data Portal in South Korea. Available online: https://www.data.go.kr/ (accessed on 29 December 2023).
18. *OGC 12-019*; OGC City Geography Markup Language (CityGML) Encoding Standard. Open Geospatial Consortium (OGC): Arlington, TX, USA, 2012.
19. *ISO 19101-1:2014*; Geographic Information—Reference Model. International Organization for Standardization (ISO): Geneva, Switzerland, 2014.

20. *ISO 19103:2015*; Geographic Information—Conceptual Schema Language. International Organization for Standardization (ISO): Geneva, Switzerland, 2015.
21. *ISO 19109:2015*; Geographic Information—Rules for Application Schema. International Organization for Standardization (ISO): Geneva, Switzerland, 2015.