A data model for simulation models relying on spatio-temporal urban data
G Langlois, Vincent Tourre, Myriam Servières, G Gervais, Gilles Gesquière

To cite this version:
A data model for simulation models relying on spatio-temporal urban data

J. G. Langlois\textsuperscript{1,2,3}, V. Tourre\textsuperscript{1}, M. Servières\textsuperscript{1}, B. Gervais\textsuperscript{3} and G. Gesquière\textsuperscript{2}

\textsuperscript{1}AAU, Ecole Centrale de Nantes, Nantes, France
\textsuperscript{2}LIRIS UMR CNRS 5205, Lyon, France
\textsuperscript{3}ForCity Platform, Rennes, France

Abstract
To understand the complexity of modern cities and anticipate their expansion, experts from various fields conceive simulation models that can be very different. Those simulation models work with a variety of data with their own organization. Furthermore, because the urban objects are studied in the context of the evolution of a city or urban area, they carry temporal and spatial information. In this paper, we present the base classes of a common data model robust and flexible enough to serve the identified use cases. We then introduce a mechanism that allows to add thematic information to those classes. We also present some possibilities of the use of the data model. We end this paper by evoking the future improvements of the data model.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—I.2.10 [Computer Graphics]: Vision and Scene Understanding - Representation, data structure, and transforms—

1. Introduction

Nowadays, some urban areas have developed infrastructures to be better understood and organised. Those “smart cities” have become real laboratories filled with data that are used to understand and anticipate their possible changes. The project MUG (Modélisation Urbaine Gerland) in the district of Gerland in Lyon, France is a good example of the intent to model the different aspects of an urban area. Researchers and engineers develop a number of simulation models to optimize and improve existing utilities, to anticipate the impacts of potential projects, or to predict possible evolution based on collected and organised data. Those models, by their diversity, rely on different data organised in different data formats. Our work presented in this paper consists in the definition of a common data model to be used by the largest number of simulation models related to city planning or analysis as possible, as traffic simulation. This data model is to be part of a unique platform to execute simulation models and visualise their results. The conception of such a platform is the ambition of the company ForCity Platform who contributed in the founding of this work. The data visualisation aspect of the platform is not presented in this paper. Another ambition for the data model is not only to serve as a base for simulation and data visualisation, but also to be compatible with existing standards, some of which are introduced later in the paper. After introducing the needs and requirements of the data model in section 2, we present existing data models and their limitations in section 3. Section 4 details the proposed data model and its base classes. Section 5 then gives some examples of the possible uses of the data model, and presents the future works.

2. Goals and requirement

This section presents the concepts needed for the data model and the presents use cases it answers.

2.1. Concepts of the data model

The main objective is to provide a common data model robust enough to be used by simulation models of potentially any kind. Those data models can be very different in complexity, calculation time and can iterate over different time steps. The proposed data model must be flexible enough to provide any type of object any model needs. Those models can be very diverse and can be related to domains such as population evolution [Ben99], transportation [Weg04], energy demand [RTIC10], or pedestrian behaviour [KB14] among others. Such models can operate on a variety of objects and work over different periods of time. Identified needs include the ability to thematically describe objects, to define relations between objects such as hierarchies constituted of spatial objects, to model specific systems like transportation networks, and to include potentially complex data like population statistics. The data model objects should be reusable, for example, a simulation model that estimates the evolution of the energy need in a neighbourhood and a model that anticipates the evolution of the real estate market of a district would use the same “Building”
object. The data model should also be able to describe the evolution of a city and the different possible scenarios. As many simulation models rely on GIS data, the geometry is an important part of the data model as well. The geometry aspect of urban objects presented in this paper is limited to 2D geometry, but the integration of 3D geometry is an important element that will be later added to the data model. The geometry of objects is a crucial element because it is also used for visualisation purposes, but can also provide new information for urban consolidation as the density construction rules in France studied in [BPMW16]. The data model should finally be flexible enough to be used by simulation models related to various domains, and it should be able to model frequently used objects, like hierarchies or networks.

As the data model is still a work in progress, the current state of the data model presented in this paper is not final and is not sufficient for several simulation models, particularly those needing 3D data. The use cases the current data model handles are detailed in the next section.

2.2. Use cases

The main use cases the data model needs to handle are related to the study of the city at different scales, from the whole city to the premises of the buildings. A strong focus is made on the semantic information, that can be related to different thematic domains. On the other hand, the studied use cases do not require 3D data, so 2D geometries are needed.

The first use case identified is the storage of statistical data related to various aspects of a city. Those data are described at different levels. The role of some computational models developed by ForCity is to aggregate and disaggregate data from one level to another, e.g. from data given at a district level to the building level. The statistical data can be divided into themes, among which are the sociological information about the inhabitants, their energy consumption or waste production. This use case requires to model the different divisions of the territory, with their 2D geometries, as well as the relations between those divisions (e.g. "contains" and "is contained").

Another use case identified is the modelling of networks and facilities, especially roads and utility networks. For such network, a 2D geometry (line) is needed, as well as the geometry of other city objects, mainly buildings and natural obstacles (water bodies). The area of interest for the networks are usually the whole city or the district. Those networks are used by LUTI (Land Use Transport Interaction) models that also need the territory division defined in the previous use case. An overview of the LUTI models is presented in [Weg04].

Finally, a various number of use cases have been studied. Those use cases need a variety of city objects with their semantic information and a 2D geometry (points, line or surfaces). An example is the estimation of the available surface in public places, computed from the data on pavements, roads, parking spots and bicycle ways.

For all the use, the evolution of the data over time is of capital importance, as well as the modelling of the different scenarios possible. Also, the same object must be universally usable by all the simulations models.

For the moment, no use case needing 3D data have been studied, but the proposed data model should be able to store data computed by a simulation model that use external 3D data. An example is the visibility of a building from a point in space. The 3D building cannot be modelled but the sky-view factor as well as a list of the buildings visible from the point of view can be expressed.

The next section presents different existing data models related to urban data.

3. Related Works

This section presents several existing data models specifically created to model urban data, or data models that integrate interesting elements or concepts related to our use cases.

3.1. CityGML

[Kol09] describes CityGML as "an open data model and XML-based format for the representation and exchange of virtual 3D city models." It is a standard of the Open Geospatial Consortium (OGC). The latest specification of CityGML is the 2.0 version approved in 2012. CityGML defines a semantic 3D data model that provides 5 levels of detail (LoD) for the 3D city objects. CityGML defines thirteen modules: Appearance, Bridge, Building, CityFurniture, CityObjectGroup, Generics, LandUse, Relief, Transportation, Tunnel, Vegetation, WaterBody and TexturedSurface (deprecated). CityGML also provides a mechanism of Application Domain Extensions (ADE) that can be used to add objects or attributes specific to a particular domain that is not covered by the base modules. For instance [dLVB11] developed an ADE to integrate the Industry Foundation Classes (IFC) into CityGML.

CityGML provides several interesting elements to integrate in the data model, objects like building, road, city furniture, or water body. But it does not provide any mean to define an object temporarily for now. [MG14], [CSG*15], [CK15] propose different enhancements to integrate temporal information into CityGML, and the integration of time is expected in a future version of the CityGML specification.

Also, CityGML being a format for the description of 3D city models, the representation of the city objects is the core of CityGML. In the data model that we want to propose, the representation is just a part of a city object that may be optional, therefore CityGML seems too restrictive for the many different targeted use cases. That means that the data model should be able to integrate different representation of an object, including CityGML Levels of Detail. Furthermore, the semantic and geometric aspect of CityGML is very restrictive as any semantic information belonging to a given level of detail must correspond to a geometric information. For instance, semantic information about interior in CityGML can only be made in LoD4 that is the most detailed LoD and required a very precise geometric representation. Those limitation have already been identified by [LBGH13].

3.2. INSPIRE Directive

The INSPIRE directive issued by the European Commission in 2007 is presented in [BK11]. The directive defines 34 data specific
themes among which cadastral parcels, buildings, transport networks and population distribution and demography. The building specification defines BaseBuildings with only semantic information that can be extended with 2D or 3D geometry. This modelling concept is closer to our needs than CityGML because they view geometry as a particular attribute of an object whereas in CityGML geometry is the core of an object. It is worth noting that the INSPIRE 3D building specification relies a lot on CityGML. The INSPIRE directive alone is too restrictive to model all the urban objects we intend to use, but the proposed data model should be able to integrate data specified by the INSPIRE directive.

3.3. The OH_FET model

The OH_FET model described in [RS07] is a data model designed for the study of the evolution of a city over long periods of time. This data model was created for the particular use case of archaeological studies. In this data model, an object is decomposed into three entities: a thematic entity, a temporal entity, and a spatial entity. The thematic entity is a value limited to a list defined by a thesaurus, the temporal entity can be either a period or a point in time, and the spatial entity is a 2D geometry such as those found in Geographic Information Systems (GIS).

This data model, although created for a purpose distant from the one we study, has an interesting view of the different characteristics of a city object and integrates space and time. But the thematic entity of the OH_FET model appears too limited.

3.4. OLAP/SOLAP

The On-Line Analysis Process (OLAP) presented in [CD97] is a unique way to model data. OLAP is used for example in business intelligence or data mining. OLAP presents a measure defined by different dimensions. For example, a measure "sales" can be defined the dimensions "type of product", "store" and "time period". This allows to know the sales of a given product, in a given store during a given time period. Dimensions can also be hierarchical, for instance stores can be aggregated into higher divisions (all the stores in a city), and a hierarchical time dimension can allow to get a measure by year, month, week, etc. All measures are then aggregated into a hypercube of n dimensions.

[RBM01] presents the Spatial OLAP concept or SOLAP. It introduces GIS elements to OLAP to handle spatial measures or dimensions. Each element of a hierarchical dimension can be associated with a 2D geometry and allow all traditional GIS operations. In our case, this approach is interesting as it allows us to decompose a city into several spatial levels, with each level aggregating data from its sub-levels. This notion of hierarchical geometric levels is crucial in the case of data aggregation and is not developed in CityGML.

But OLAP is too far from what we want to achieve, as those concepts are not designed to handle urban object models. The aggregation of data from one level to another (from a district to a city block for instance) is frequently addressed in our use-cases, so the hierarchies formalised by SOLAP is still interesting. [MZ04] identifies several hierarchies that the proposed data model should be able to implement.

3.5. Summary

The urban approach of CityGML is close to what we want to achieve, but the lack of a temporal aspect and monitoring of the objects prevent it to fully suit our needs. Furthermore, CityGML data model is primarily made for storing 3D city models. The semantic geometry is thus a core principle of CityGML, whereas in our case the geometry is an object attribute among others, depending if it is needed by a simulation model or not. If a 2D geometry is sufficient most of the time, the integration of 3D data must be possible in our proposed data model. Also, the ADE mechanism of CityGML is a strong concept and a similar concept will be part of our proposed data model.

The INSPIRE directive data specifications do not cover all aspects we want to study, but in the case of the buildings, contrary to CityGML, geometry is seen as an extension of a core base building, and this modular feature fits our needs.

The OH_FET model is also an interesting approach as it separates the spatial, temporal and semantic aspect of an urban object, so we retain its core principle. However, this data model was designed for a particular use case (archaeological research) and is thus too limited. Finally, the OLAP/SOLAP approach is interesting, as it allows the modelling of hierarchies of objects and answers the need for aggregation and disaggregation of data that is mandatory for many simulation models. However, its core concept is quite far from what we want to propose, as it was designed for a particular purpose (data mining and business intelligence).

The next sections present our proposed data model, its core classes, a few examples of what we are able to model, and the future improvements we want to add.

4. Data model objects presentation

The idea behind the design of the proposed data model is to provide base objects that could be used for many simulations related to urban data analysis. Furthermore, it should provide a way to specialize those base objects to answer specific use cases defined by the applications. For example, a base object "Building" would contain information not related to an application in particular, like its height. Furthermore, an application might need specific information about the building that are related to a particular domain, like its annual electricity consumption.

The UML diagram presented in figure 1 presents the base classes of the data model. The model is composed of a core class City Object that is described temporally and spatially. This class will serve as a parent class for all base classes. The thematic aspect of objects will be described using the City Object Extension class that also has temporal information. Those classes are then aggregated in a City Context that is versioned. Detailed descriptions of the classes are presented next.

4.1. City Objects

Base objects of the data model are identified as City Objects. The City Object class is generic enough to represent any urban object. Every base object in the data model inherits from this City Object class. Those base classes are able to specify spatial, temporal, and
Figure 1: UML diagram of the data model's base classes

semantic information. First, a City Object specify its geometry. It must be a 2D SIG geometry: a Point, Multipoint, Linestring, Multi-Linestring, Polygon, or MultiPolygon. Nonetheless, non-geometric City Objects can exist and be used to model abstract concepts like organisations or households, or to store statistical data. A City Object is also defined by a period of validity. Finally a City Object is composed of several semantic attributes. Those attributes can have various forms: primitive types like boolean, number or string, or more complex attributes. Those attributes must define the object intrinsically and not be related to any thematic domain. City Objects can be for example a building, a road, or any kind of administrative zone. This City Object resembles the _CityObject class of CityGML as it is the base class of a number of CityGML classes, but it does not incorporates the principle of expressing an object geometry, temporal existence and semantic attributes.

The distinction between the spatial, temporal and semantic aspect of a City Object is similar to the OH_FET model proposed by [RS07]. However, the thematic entities of the OH_FET model are restricted to an archaeological thesaurus because archaeology is the only one application for this model. The proposed City Object class is much more flexible and can have all kinds of attributes, simple and complex, like stated before. A City Object is identified by an Object ID and a unique UID. Two City Objects can have the same Object ID only if their respective period of validity do not overlap. It means that those two City Objects are the same version of the same object at different moments. Each object also has an UID that cannot be duplicated. In [CSG*15], the mechanism of minor ID and major ID resembles the Object ID mechanism. In our case, the Object ID is similar to the major ID, and the UID can be the concatenation of the minor ID and the major ID.

4.2. City Object Extensions

After defining what a base object is, a mechanism to specialize those base object must be added. An application is indeed rarely thematic-free and therefore cannot rely solely on base objects. Hence we introduce the City Object Extensions. A City Object Extension is an object that has attributes related to one domain in particular. A City Object Extension must be linked to a City Object to exist. Several City Object Extensions can be linked to the same City object. A City Object Extension has its own period of validity that must be contained into the period of validity of the City Object it extends. It can be identical to the period of validity of the City Object, or shorter. This allows to express the evolution of thematic attributes that do not intrinsically qualify the object. For example, a building can have several extensions "energy consumption" with different period of validity that do not overlap, to indicate the evolution in time of different attributes related to the energy consumption. City Object Extensions can be assimilated to the ADE mechanism of CityGML. Unlike the City Objects, the City Object Extensions does not possess an Object ID because, as a City Object Extension must be linked to a City Object, the referred City Object served as version identification.

4.3. City Context

A City Context is an object that represents an urban study area. City Objects and City Object Extensions must belong to a City Context. City Object Extensions can only extend City Objects that belong to the same City Context. All City Contexts are versioned. The versioning of the City Context is similar to those of Version Control Systems. City Contexts can be forked and merged. A City Context can also be frozen. When it is frozen, no modifications can be added and it must be forked to be modified again. A frozen City Context can be viewed as a snapshot of the City Contexts evolution on its time-line. A City Context has a child or children if it is forked, and a parent or parents if it comes from a merge.

When a City Context is forked, all of its City Objects and Extensions are duplicated and kept the same Object ID. Therefore, two City Objects or Extensions that belong to two different City Contexts can easily be identified and compared. An application is having two City Contexts describing two possible scenarios for the same city. The evolution of the same object can be easily observed because the two versions of the object keep the same Object ID. An example of the evolution of a City Context is presented in figure 2.

Figure 2: Example of the evolution of a City Context

In addition to their period of validity, every City Object and City
Object Extension is monitored and has temporal attributes to indicate its creation and deletion times. Like in [CSG15], this allows bi-temporal modelling of City Objects and City Object Extensions. The next section presents some examples of concrete use of City Object and Extensions.

5. Results

The purpose of the data model is to serve as a support for different simulation models that manipulate urban objects. In this section some examples of the kind of classes and extensions created to handle certain use cases. The presented classes and extensions all inherit from the previously introduced classes _CityObject and _CityObjectExtension.

5.1. Building

The UML diagram presented in figure 3 gives the example of a Building object specialized with 3 extensions giving information about the energy consumption, the inhabitants and the occupancy of the building. Since the classes Building, Population, Occupancy and EnergyConsumption inherit from _CityObject or _CityObjectExtension, their objects must specify a period of validity. Only the class Building specifies a geometry, since it inherits from _CityObject. Those classes could be used for example by a simulation model to evaluate the need in energy depending on the number of inhabitants and their presence.

5.2. Attributes evolution

Since City Objects and Extensions have different validity ranges, Extensions can have a shorter longevity that the City Object they extend, and several extensions of the same type can extend the same City Object. Knowing that, we can express the evolution of an Extension without modifying the City Object. The figure 4 shows a Building object with a longevity of 10 years, and two Population objects that extend that Building. The Population objects both have a validity range shorter than the Building object. This expresses that the attributes of the Population extension change over time. In the implementation of the data model, there are strong constrains on validity ranges to assure that extensions no not overlap and are coherent with the object they extend.

5.3. Hierarchy

The next example shown in figure 5 is a case where different classes with a hierarchy relation all have information in common. In that case, it is a hierarchy of administrative zones. All elements of the hierarchy are derived from the abstract class _AdministrativeZone. The extension DailyWasteProduction is applied to this abstract class meaning all its children classes can be linked to a DailyWasteProduction extension. The aggregation and disaggregation of data from one element of the hierarchy to another is an important use case studied in our work.

The representation of hierarchy is at the core of the OLAP concept. Hierarchy are traditionally modelled in OLAP using star or snowflake schema as seen in [CD97] where a fact is linked to the smallest element of a hierarchy that is then aggregated into higher levels of the hierarchy. The problem is that in our case, data is often obtained at a higher level and will have to be disaggregated to lower one, so the data model should allow to store data at any level of the hierarchy, as in the figure 5. On the other hand, it forces to add constraints between hierarchy levels to assure data integrity. For example if the level N and N-1 of a hierarchy have the same information, it must be assured that the sum of all N-1 data are equal to the N data.

5.4. Using extensions to be compatible with CityGML and INSPIRE

At this stage of development, the data model lacks the implementation of the 3D semantic geometry provided by CityGML. An object with a 2D geometry can however be considered as a LoD 0 in CityGML. Nevertheless, all attributes related to a CityGML Building for example can be modelled as a City Object Extension for the base Building class. Attributes that are not available in CityGML can be added using the ADE mechanism. The figure 6 shows the UML diagram of a City Object Extension that adds the CityGML Building attributes to the base Building class.
Other CityGML classes like Transportation (Road, Track, Railway, Square), Vegetation, WaterBody or CityFurniture can also be integrated as base classes or extensions of the data model.

The same way, City Object Extensions can be created to integrate into the data model the BuildingsBase and BuildingsExtendedBase schemas of the INSPIRE Data Specification on Buildings.

5.5. Implementation

The classes of the data model previously described are implemented in a PostGIS database, accessible via a REST API. On top of this database, tools written in Python have been developed to provide a development environment for simulation models developers. Those tools allow to fetch a City Context and get, update or create its City Objects and Extensions. A process for the integration of classes has been set up to ensure the coherence and consistency of classes and extensions. A schema describing this organisation is presented in figure 7.

6. Conclusion and future works

So far the current data model has managed to answer all the needs expressed by the simulation model developers of the ForCity company. 34 base classes and 35 extensions have been defined.

Those classes are used to handle use cases such as anticipate the evolution of needs of a population, simulate the traffic, estimate the evolution of the job market, or optimise the extension of an urban utility network. The data model is constantly changing and classes are expected to be created, modified or merged. However, some improvements are needed and some important cases of modelling appear to need stronger mechanics.

The most important features to develop is the capability to describe relations between objects of the data model, and most importantly to define constraint on those relations. Those constraints can be temporal (e.g. two objects existing at the same time), spatial (e.g. two objects must be separated by a given distance), or semantic (e.g. one object can be linked to another only if one of its attributes equals a given value). What’s more, information specific to those relation need to be integrated, for example traffic information between urban zones. To model those relations, a specific object Relation distinct from City Objects and Extensions could be integrated in the data model. It would be able to link several objects together and add information and constraint about the relation between those two objects.

Also, we saw previously that the modelling of hierarchies of objects is an important use case of the data model. Today, hierarchies are modelled using attributes parent and children but it does not allow to describe complex hierarchies. [MZ04] describe several types of possible hierarchies that the data model must be able to implement. This could be done by using the Relation object previously mentioned, or it might require a new object Hierarchy that would describe levels and constraints of a hierarchy.

Furthermore, the present temporal aspect of City Object could be improved by, for example, introducing an Event object that could be able to temporally link different City Objects and Extensions. It could be useful for example to indicate a point in time corresponding to the end of existence of an object and the beginning of a new one.

The geometric aspect of City Objects must also be improved. For now a City Object can only have one 2D geometry. One idea is, at first, to allow an object to have different 2D geometry semantically identified, thus allowing to have for example a City Object representing a zone with two geometries: a polygon for the edge of the zone and a point for its centroid. City Objects must also integrate semantic 3D geometry of different kind, and be able to handle Levels of Detail, and be flexible enough to be fully compatible with the CityGML data model.
Acknowledgments

The authors thank the ForCity Platform company and ANRT (CIFRE) for funding this work.

References


[RBM01] Rivest S., Bédard Y., Marchand P.: Toward better support for spatial decision making: defining the characteristics of spatial on-line analytical processing (solap). GEOMATICA-OTTAWA 55, 4 (2001), 539–555.

