Geospatial Database for the Generation of Multidimensional Virtual City Models Dedicated to Urban Analysis and Decision-Making

Andreas Fricke[(✉)] and Hartmut Asche

Hasso Plattner Institute, Digital Engineering Faculty, University of Potsdam,
Prof.-Dr.-Helmert-Str. 2-3, 14482 Potsdam, Germany
{andreas.fricke,hartmut.asche}@hpi.de

Abstract. The provision of a data base for the digital representation of a virtual spatial model is used as an example to demonstrate how a geographical problem solution can be achieved using modern service-based techniques and methods. Moreover it is explained which problems exist in the spectrum of digital twins and spatial databases with regard to the handling of ubiquitous geodata, which necessitate the conception of a new service-oriented approach. The database as the central medium towards a virtual representation of the urban environment is a particular focus. Ubiquitous geospatial data demand meaningful tools for integration and harmonisation, which are also included as part of a modular process chain. In this approach, procedures and workflow are designed on the basis of generic data models and standard schemes, with the aim of ensuring transparency and transferability, both methodologically and spatially. A workflow describing data acquisition, processing and utilisation is presented and exemplified by the Middle Eastern agglomeration region of East Jerusalem.

Keywords: Multidimensional virtual city model · Web-based platform · Database · Data model · Stereo imagery · Service-oriented computing · East Jerusalem

1 Introduction

Current geoinformation systems have not been designed to amalgamate and process heterogeneous, multidimensional geospatial data from different sources. It is hence not without limitations that this data can be used effectively and universally. What is lacking, in particular, is a geospatial processing and analysis environment including a harmonised, dynamic geospatial database continuous updating procedures as well as easy user access to elementary and specialist functionalities. The immense amount of unstructured, unfiltered data and a multitude of extensive software systems simply make it difficult for the non-professional user, in particular, to manage even simple problems with spatial data.

2   Study Area

In order to make a comprehensive statement about the effectiveness of the approach, the special form of the Middle Eastern urban agglomeration area of East Jerusalem is chosen as the research area. In general, the form of connected human settlements is characterised as an urban agglomeration or extended urban space [1]. It consists of built-up areas of a central location and related suburbs linked by a corresponding contiguous urban area. According to INSEE [2], the concept of urban units is based on the continuity of buildings and the number of inhabitants. An urban unit is a municipality or a group of municipalities with a continuous building zone (not more than 200 m between two buildings) with at least 2,000 inhabitants. If the urban unit is located in a single municipality, it is an isolated city. If the urban unit extends over several communes and each of these communes concentrates more than half of its population in the contiguous construction zone, it is a multi-montane agglomeration. Rural municipalities are those that do not fit into the constitution of an urban unit. This means municipalities without a continuous construction zone of 2,000 inhabitants and those in which less than half of the urban population is located in a continuous construction zone [1]. Agglomeration regions with a strong urban character can also be described as metropolitan regions. The functional urban region and urban agglomerations are highly integrated functional areas - complementary functions of different levels at different locations to provide the population with all necessities - from housing function to jobs to education, shopping and use of various services; related to Christaller's theory of central places. Urban agglomerations allow for services of all hierarchies up to the highest to be offered efficiently, as the various hierarchical services are located in close distances and serve a large number of people living closely together or collaborating (Fig. 1).

Fig. 1.   Aerial view of a typical urban agglomeration in the Middle East. Source: wikimedia.org

The study area of a Middle Eastern urban agglomeration represents a peculiarity and adaptation of the classical concept of urban agglomeration in that it also takes up the fine granularity in the structure of vertical and horizontal building, distribution and diverse use. Both urban and rural units are interconnected and allow an almost complete geographical analysis and assessment of the contrasts. The area under study includes East Jerusalem – the Old Town, excluding the city of Jerusalem - which was a part of Jordan from the Palestinian War of 1948 until conquered by Israel in the Six-Day War of 1967. East Jerusalem is the urban type of a classical oriental city. This urban type is interesting as there have been only a few major urban transformations over time. What characterises the region in particular is a constant urban expansion following Western and European models, which is reflected in the heterogeneity of the urban region. Basically, this process, which took place in many oriental cities, is transferable. It is therefore of great interest to understand the structure and context of the urban agglomeration of East Jerusalem and also to generate the digital representation of this region. The digital representation can allow to obtain a clear insight into the difficult acquisition, modelling and processing of the geographical conditions, in also allowing a comparability of possible future changes.

3 Data Modelling, Acquisition, and Management

In order to maintain a complete database of the relevant spatial entities as up-to-date as possible, it is necessary to consider a variety of available data sources. Nowadays, the problem is usually not the availability of geodata, but its heterogeneous nature. It is not easily possible to utilise all available sources with and among each other, since there is a lack of data models and schemes that could guarantee a successful integration and harmonisation [3]. In order to check the usability of existing data, it is necessary to select the relevant objects and combine them in an object type catalogue. The object type catalogue provides the basis on which existing schemes can be selected, adapted or newly developed. The schemes represent the link between the object type catalogue as data modelling, and the technical implementation in as structure and mapping in a database. Schemes should be designed as generic as possible and as detailed as necessary. The result is an abstract model that organises and standardises elements of data and how they relate to each other and to the properties of the real world.

3.1 Data Model, Objects, and Schemes

Data modelling is a term used in methodologies for the formal representation of objects relevant in a defined context by means of their attributes and relationships at the conceptual level. The main goal of data modelling is the unambiguous definition and specification of objects and the assignment and management of their attributes required for information purposes and their relations and dependencies between objects in order to obtain an overview of the data view for an information system [4].

The result of modelling are data models which, after potentially recursive modelling steps, finally lead to usable data schemes or data bases (conceptual, logical, physical schemata). Databases are thus the technically enhanced conceptual data model

and allow for its application with certain adaptations. As a rule, data models have a much longer lifespan and persistence than functions and processes (software systems), since they are detached from the technical level by the basic conceptual idea and thus are less susceptible to changes in a process [5]. It is true that data is more stable, but functions are not. Transferred to the use case, this means the identification of the relevant information requirement (attributes), whereby entity types and relationship types are identified and assigned. In addition, possible attribute values or proposals of attributes to be identified are determined. Finally, relationship cardinalities and the functional description of the entity and relationship types as well as the attributes are fixed. Under the premise of simplicity transferred to the application case of a Middle Eastern urban agglomeration, this means that an extremely generic basic data model can be derived: Each relevant object is unique and can be represented by multiple geometries and by multiple semantic attributes (rich information) as shown in Table 1.

Table 1. Comparison for the data model of relevant urban objects and rich information.

Relevant urban objects	Rich information
Built environments (buildings, infrastructure)	Geometries or appearances
Natural features (terrain, vegetation, land use)	Semantics or attributes
Utility networks	Decompositions and relations

Schemas are used to adapt the conceptual data models for application and transfer into a database and thus to coin them onto a database system. The INSPIRE schemes serve to ensure a long-term and as generic as possible technical convergence of the systematics, as the European Union has committed its member states to maintain all official geodata based on these schemes in the future. All functionalities and processes of integration and harmonisation, as already described in [8], take place via direct access in an INSPIRE compliant exchange scheme on the database. A feature catalogue provides an informative overview of the spatial object types and data types defined in the INSPIRE, providing implementers and decision makers with an easy introduction to the INSPIRE data models and data specifications. The INSPIRE Implementing Rules on Interoperability of Spatial Data Sets and Services and Technical Guidelines (Data Specifications) define common data models, code lists, map layers and additional metadata on interoperability to be used when exchanging spatial data sets [6].

3.2   Multi-dimensional Geodata

In many disciplines, two-dimensional data sets are the ones to choose. While two- and higher-dimensional data sets are multi-dimensional, the term multidimensional tends to be applied only to data sets with three or more dimensions. Normally, when thinking about multi-dimensional data, they are characterised by time or multiple phenomena obtained. Typically, the geographical dimension is limited to 3+1D (spatially and temporally). In fact, combining time series and cross-sectional data can be thought of as daily issues in the geospatial domain. But often these analyses or standard tasks are

limited to one dimension only. As an example, weather forecast data sets provide forecasts for multiple target periods, conducted by multiple forecasters or automated models, and made at multiple horizons. Both semantic and numerical attributes help to generate an increased information space, which can be made addressable and usable. The multiple dimensions provide more information than can be gathered from two-dimensional data sets. Therefore, multi-dimensionality has two meanings which both are important but require different strategies [6]. The tricky task is to combine both in a flexible process chain to take advantage of different perspectives of one and the same object. The solution are partial reverse engineering techniques as a generalisation, where the most important facts are a multidimensional database based on different resources, while reverse engineering is used as a search for identical characteristics in different scales, harmonisation and homogenisation of geodata to enable geoprocessing at runtime.

3.3   Generalisation of Multi-dimensional Geodata as Data Integration

Generalisation is known and necessary for many different tasks, areas and applications. Nevertheless, their meaning could be described as simplification or abstraction, in order to ultimately obtain a more universal insight and evidence for a tested behaviour.

The process generally includes reducing complexity, emphasising the essential and suppressing the unimportant, maintaining the relationship between objects while maintaining a defined quality [7]. Last but not least, it is a process of semantic and numerical solution reduction. Quality is the selection and simplified representation of a detail, appropriate scope and/or purpose of an application, of various processes arising from the data source through the application of transformations, as derivation objectives are to reduce data in terms of quantity, type and cartographic representation, whilst maintaining consistency and clarity of representation. Thus, the term generalisation is divided into two main parts [7]. Model generalisation as new conceptualisation of phenomena (mostly qualitative extension) and visual generalisation as new structural representation of Phenomena (mostly quantitative). In either case, both types of generalisation aim at isolating the features that are independent of both feature and domain resolution. In addition, it also defines the capture in a software architecture. The proposed software architecture makes it possible to instantiate this method as a web-based service system. This means that users can adapt the system to their specific (micro-) task or workflow [8], regardless of their level of knowledge or professional affiliation. It is designed in such a way that it enables a complete division of the process chain as micro-tasks. The architecture is designed to meet two non-functional requirements: Scalability (vertical and horizontal) and adaptability (deployment and runtime). Included are experiments to evaluate these two requirements in terms of quality assurance and performance issues.

3.4   Data Acquisition

A problem with the use of existing data sources consists in the fact that these data sources are usually not available across the entire study area, be it in their extent, their characteristics or their quality [7]. In this approach, a basic data set is therefore derived from a uniform remote sensing data set, which is subsequently enriched and expanded with further existing data sources. This ensures that a basic level of comparability can be guaranteed as a reference in the study area.

Several methods exist which are appropriate to collect a comprehensive basic data set. Since the study area is characterised by a high heterogeneity and classical in-situ methods do not appear to be effective in terms of cost/benefit, a remote sensing campaign is used instead. For this purpose, high-resolution aerial imagery of the area is evaluated by methods of stereo photogrammetry. In the course of each photogrammetric analysis, pixels in at least two different images of the same spatial area are compared and three-dimensional coordinates are derived using compensation functions. This method is called matching in modern software systems. Influencing factors are primarily the visibility of two identical image-object points (image overlap) and the probability that the same image-object point is encountered. An increased resolution of the input data, however, only plays a subordinate role.

In sum, digital surface models can be derived for overlapping image areas. Nowadays, digital surface models are usually made available on a raster basis. A characteristic of raster data, however, is the basic regular generalisation through its cell-based data structure (Fig. 2).
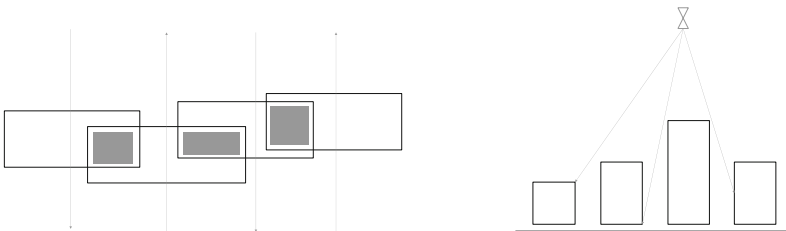

Fig. 2. Schematic overview of different flight strips (rectangles) and overlapping areas in dark grey (left) and potential image-object rays with respect to visibility of buildings (right).

In order to process a surface model as objectively as possible, the result of the image matching is used for the further workflow instead. This result is characterised by irregular and inhomogeneously distributed points of the overlapping area of the photogrammetric analysis. This intermediate product is also called a point cloud. A distinction is made between the sparse cloud and the dense cloud. Depending on the software system and implementation, the sparse cloud is the result of a pre-allocation of all images and overlapping areas in a bundle block (Fig. 3).

In a first iteration a mostly regular assignment of image-object points is made for each overlapping area, which is condensed in a second iteration, thus becoming a dense cloud until the squared resolution of the input data is reached at most; unless the

subpixel area is used. This means that at an input resolution of 10 cm Ground Sampling Distance a theoretical dense cloud with a density of 10 * 10 points per square meter (100 points) can be derived.

Fig. 3. Schematic overview of the visible image-object points of buildings and ground (left). The representation including building and ground points to be reconstructed (right).

3.5   Data Management

With the advent of digitalisation and the ever-growing volume of data, new disciplines (data science) have been opened up in information technology and related disciplines to address the scaling problem facing today's software systems and users. File sizes and file occurrences are multiplying in many domains, although this fact does not constitute the whole problem. Rather it is the awareness and the way in which data is stored, analysed and made usable [9]. Especially in the domain of geoinformatics, this functionality can be implemented with a database and database management system even with growing data and data volumes. In software development, working with databases is a must. But for people in other subdomains of computer science (e.g. with GIS) the advantages of a database may not always be so obvious. People tend to use the most familiar tools, such as classical monolithic GIS, although it is not the most efficient way to achieve goals. But abandoning the quasi-standards can provide great benefits. The data management is implemented as part of a modular process chain [8], which aims at realising the digital representation of the urban agglomeration of the East Jerusalem region in a web-based software environment on the basis of the derived area-wide basic data set [10].

4   Digital Twin vs Model

A digital twin is a digital replica of a physical object, device, system, place or person. Originally, digital twins were used in manufacturing to represent real devices, machines, factories and plants. For example, sensors attached to machines over an entire production line route data through an IoT network that can be accessed interactively.

The myth of the digital twin is based on ensuring a complete digital representation of a real object. This does not match the concept of model, since the fundamental principle is to abstract and generalise. Thus, it is the data of a digital twin as a visual

model that assumes the form of representation and functions as an information medium. In contrast to geoinformatics, the digital twin is regarded in computer graphics as the most accurate image of a real object. In geoinformation science, however, the focus is always on the digital model, whose degree of abstraction serves as the basis for a variety of applications [9]. Models serve as abstraction of reality for better analysis in a wide range of applications. They differ in the degree of similarity defined and required by a particular application. Digital twins, on the other hand, are usually the most accurate image of reality, although they represent only one model due to possible fine-granular differences to the reference, reality. Therefore, digital twins are also called "Look-a-Like", "Feel-a-Like", "Work-a-Like", because the focus is often not on the solution of an application problem, but rather on the representation of reality in certain forms. The functions of a digital twin are usually limited to a few individuals. The spatial reference and the application context usually play a subordinate role. In contrast, digital twins as a model have a very low modelling depth with respect to classical spatial applications, since its degree of abstraction is very low. The degree of realisation is very high, therefore it is realistic, although the focus is still on technical processing and presentation. However, this is a limiting factor insofar as it falls below the limit of perception in terms of closeness to reality. For a large part of the applications, however, this is technically conceivable, but irrelevant, since it does not contribute to better presenting or solving a problem [6]. At this point, the different approaches of the aforementioned scientific disciplines become clear, although both use the example of the virtual city model to work on the same application. This corresponds to the observation of a continuum that reflects iconicity [11]. Both disciplines approach each other bipolarly, but always with different emphases on modelling and "look-a-like". However, there are points of contact that form the basis of this research. The application, which is based, among other things, on geodata, can certainly benefit from the complementary combination of different concepts, techniques and processes [6]. However, this offers the possibility to branch evaluations and analyses into specific assets for an improved root cause analysis and a multitude of new use cases for improved manufacturing processes. Digital twins help companies outperform previous performance levels in various aspects of manufacturing, from the product design phase to supply chain management. The main tasks and advantages of a digital twin are summarised:

– Monitoring (Unsurpassed level of monitoring)
– Maintenance (From reactive to predictive approach)
– Training (Education and guidance)
– Communication (Analytical information generation and accessibility)
– Strategy (Testing and improvements and traceability)

All different components of a physical object that are emulated, identified and described in a structured way with a digital twin are part of a multi-dimensional model. An accurate and comprehensive virtual model allows for visualising how the physical real object functions, behaves and changes at the moment. The biggest challenge for a digital twin of an urban city is that one model does not fit all. In other words, a model is needed that is as generic as possible and as domain-specific as necessary. Conversely, a digital twin would be needed for each individual image of an urban space. This is due

to the fact that each urban or metropolitan region has functional spatial structures that can be expressed in socio-economic and socio-demographic differences. A generic digital twin is thus a modelled, abstract and generalised instance of the real world [12]. It is therefore a model of reality and no longer a digital twin, although the basic principles of the twin are maintained. However, the basic population of objects and data patterns that characterises an urban Middle Eastern agglomeration are the greatest common denominator and testify to the fact that a transparent transfer to different regions and spatial structures is at least possible.

5   Technical Objectives of a Geospatial Database

During the last decade of the century, database systems such as the open source PostGIS have developed into reasonably usable media that can be used effectively with the problems described in dealing with geodata. Today, however, most software developers assume, as they did ten years ago, that the design of scalable geodatabases is a simple task. As it turned out, at the beginning of 2000, there was not even the computer science necessary to design such things. Typical methods for representing dynamic geodata, from R trees or hyper dimensional hashing to space-filling curves, were all developed using techniques from the 1980s or earlier [13]. Geodatabases, at the level of the basics of computer science and implementations, are not linked to conventional databases at all. The superficial similarities hide countless design challenges that are specific to geodata models. All architectural differences must therefore be regarded as lessons to be mastered and manifested as critical faults in real applications. Predominant algorithms in the scientific discipline of computer science apply, with few exceptions, only those properties that are unique to one-dimensional scalar data models. This means that an abstraction to the integer already takes place in the data type definition. This circumstance is problematic, since multi-dimensional data are also represented reductively in this way in predominant database systems. A data model that is not scalar in its nature can only be represented with some difficulties using current methods. There are only a few operatively used systems that can handle non-scalar data types like paths, vectors, polygons and other elementary aggregations of scalar coordinates used in spatial analysis [13].

The reason for this is relatively profane: Computational relations today are always topological and not graphic. Geodata types include interval data types as well as some other common data types. An interval data type cannot be represented with less than two single values of any dimensionality, such as the boundary of a hyper rectangle. Interval data types differ from scalar types in two important aspects: Sets have no meaningful linearisation and intersection relationships are not synonymous with equality relationships. The algorithms that exist in the literature for interval data are simply not applicable for the use in the context of geographical issues due to the existing discrepancy. For example, there are no dynamic sharding algorithms that produce even distributions of interval data. Furthermore, there is no general partitioning function that distributes n-dimensional interval data evenly in n-dimensional space. Hash and range partitioning, which in this context is essentially just a distributed quad tree, have been shown to be wrong-headed decisions, but are still popular in naive

ubiquitous implementations. Interval indexing algorithms in the literature are either not scalable or not general [13]. The R-Tree family of algorithms, the traditional choice for small datasets, cannot scale even in theory. The quad tree algorithm family is pathological to interval data. Since R-trees were invented to replace quad trees for this reason, the quad tree algorithm family is not suitable for interval data. Sophisticated variants of raster indexing (tiling) can be scaled for static interval data, but are unsuccessful for dynamic data, and most software engineers use variants that are even more naive in practice because they are quasi standard. The basic phenomenon and problem are that interval data sets do not have a usable order. The assumption that data is sortable is so widespread in computer science that many design patterns are either useless or simply wrong when applied to interval data, despite supposed results. For example, which is the best storage format for an analytical geodatabase is not obvious. Many storage formats, such as those used in columnar databases, make compromises that require the sortability of scalar data types [6].

5.1   Geospatial Databases in Times of Big Data

The problem described here increases in times of Big Data. The original use of most geographic databases was primarily to create maps, such as geospatial models, which can be displayed as image tiles or paper products. Mapping databases evolved in an environment where the data sets were small, rarely changed, the production of a finished edition could take days, and updating and continuation could be cyclical. However, modern spatial applications are increasingly based on real-time spatial analysis and contextualisation of data from the IoT, sensor networks and mobile platforms. These workloads have little in common with the original use cases. In fact, these workloads are different from all the workloads studied in database literature [14]. The basic problem is that there is no existing design that could be copied or that could support this use case.

For a modern spatial question and application, an optimised database engine must be designed and implemented according to the basic principles, which requires an increased level of capabilities and the lay user, who is mainly the user, faces unsolvable problems. Many geospatial data sources continuously generate data at extremely high rates. Not only are complex geospatial data analysed, indexed and stored, but low-latency queries against incoming and historical data that cannot be aggregated are also possible. Also much praised in-memory databases are useless here, since the volume of online data, for example, is quickly too large. Although technically feasible, virtually no databases have yet been developed for these applications. However, this again reflects the typical organisation of scalar data models and is therefore difficult to implement. Geographical applications are typically highly distorted and shift unpredictably across the data model [9]. Traditional skew and hotspot mitigation strategies, based on a priori assumptions about workload distribution and hard-coded in software such as those used for processing time series data, are largely useless in the geospatial domain. The adaptive reduction of severe and unpredictable hotspots is therefore a novel architectural requirement.

5.2   Database Engineering Missed Geospatial Requirements

Geospatial operators are inherently built around computer-aided geometry primitives. Looking up a description of the Vincenty algorithm, for example, is simple. The implementation of non-Euclidean operators, on the other hand, which are generally correct, precise and fast, exceeds the knowledge of most programmers. Implementations developed for cartographic and GIS application cases typically do not have the performance and precision required for analytics [9]. Analytics require the creation of responses in a time frame that counts with tiny and accurately characterised error limits. In practice, performance is often so miserable that commercial providers provide analytical results that are delivered in weeks for datasets that might not fit into memory. The reliable generation of correct and precise calculation results from complex geometry operations is known to be difficult to achieve, especially on a scale of urban agglomerations heterogeneity. This may be sufficient to produce maps, but it is almost useless for modern sensor analysis applications in times of digital twins [12]. The physical world is non-Euclidean. This creates horizon effects even within a single large building. The simplest geometric surface that still applies to most calculations is a geodetic ellipsoid. The computationally simpler approximations used by many common platforms under the hood work for maps, but not for spatial analysis. Geodata analysis requires the ability to perform polygon sections quickly. Polygon section algorithms, such as relational joins, are quadratic in nature. Naively intersecting polygons with thousands of nodes, as are common in some industries, can require millions of high-precision geodetic operations to evaluate a data set. This can be quickly multiplied by billions of data sets and the query cannot be completed in a month. The computational geometry must be precise when it comes to analysis. For trivial algebraic geometry, it is not difficult to guarantee error limits in floating-point calculations. However, most programmers do not realize that the transcendental functions implemented in their CPUs have significant precision losses in a variety of edge cases that need to be considered. On the scale of these data sets, edge cases are quickly and frequently tested [6].

It is clear that there are currently a large number of problems with the existence and use of databases in the context of spatial applications. However, these problems are not trivial and there is a small trend towards solving these problems that are not contained in public informatics literature. With the advent of mobile, IoT and sensor analytics as high-value markets, any database platform is in the process of at least enabling spatial analysis [15]. However, as described above, it is challenging to add adequate support for powerful spatial analysis to databases that have not been developed specifically for such use. There is ample evidence that many database companies have tried and failed, or that the products are being mistakenly presented as a geospatial database and as standard. From a database architecture perspective, the cause of the problem is that the internals of a database system that has been developed for traditional text and number data models are virtually everything but have a very limited link to a database system that is needed for spatial (real-time) data models. No amount of software duct tape bypasses this impedance mismatch, but most laymen try to build a geographic database. Geodatabases are difficult to create at best. They are almost impossible to build,

as long as it is naively and predestined that they are essentially the same as traditional databases with a certain amount of geographical dispersion over them [8].

6 A Multi-usable Database for Virtual City Models

Addressing this issue, a comprehensive multi-usable database is built, embedded in a web-based, modular geospatial software environment that meets the requirements of seamless processing, analysis and visualisation of multidimensional geodata. Built to the principles of service-oriented computing, the geospatial database and the service platform is accessed via a string of web services targeted to non-expert users as well as experts of different application domains. By using the respective services, the lay audience will have restricted access to the system allowing them to select, process, analyse and visualise all data required for a particular application problem, such as location analysis of a community building. Experts will have full access to all system functionalities to perform more complex processing and analysis tasks, such as advanced network analysis of public transport. This geospatial service platform is designed for the use of stakeholders, both professional and non-professional, in urban agglomerations of the Middle East with particular attention to the communities of East Jerusalem. The desired modular components are shown in Fig. 4.

6.1   The Conceptualised Digital Twin of an Urban Environment

The first step in creating a digital representation (digital twin) is to characterise all components of an existing physical object in a virtual proxy. Given the complex nature of the Middle Eastern urban agglomeration that is digitally replicated, there are thousands of different components that represent the style and structure of the model. Therefore, a comprehensive basic data set is derived from stereophotogrammetric high-resolution image data for the entire study area. Using the intermediate product of point clouds, a first robust version of the digital twin sketches the objects of the building type exemplarily and stores these in a database. For example, a multi-dimensional model of a building in the study area can show how a Middle Eastern building is divided in its functional structure. Furthermore, on the basis of geostatical analyses, it can be recognised or deduced how certain facts affect the environment or whether improvements need to be made [15]. The digital representation therefore serves as a test medium for the real world, as long as the model is fed with up-to-date data. Due to the granular, highly varied nature of the objects constituting the urban fabric, processing needs to deal with different planimetric and vertical spatial as well as semantic dimensions and accuracies. This poses a major problem for precise photogrammetric image processing and subsequent construction of high-resolution 3D city models in oriental cities such as East Jerusalem.

The fundamental question that arises, however, is which form of representation (degree of iconicity) is sufficient to address a spatial question. In most cases, it is only a simplified geometry that offers an added value through semantic enrichment and can be used for decision support. Following this assumption, it is also sufficient to enrich a simple LOD1 geometry model with semantic data in order to add it to the database.

Any analysis that requires the spatial location as a base can be served in this way. The degree of detail, or conversely, the abstract level of the geometry has fewer effects. There are unique object primitives that are necessary. These object primitives are assigned geometries and semantics as characteristics, which form the basis for a digital representation. Different attribute classes, types and values can be added to the fundamental object of a building, for example the categories administration and public order, traffic and transport, service, business, residential, and culture and leisure. The first steps are therefore:

– Adding existing data to the database: geometries, reconstructed objects, integration of OpenStreetMap data
– Initiation of procedures for checking or securing data harmonisation, data fusion, data quality, data completeness, semantic enriching, plausibility of contained data
– Use of 2D/3D visualisation and geoanalysis for regularly update, monitor, process and disseminate relevant key indicators such as
  • Demographic trends: housing, and infrastructure
  • Socio-economic data: extent of poverty, unemployment, and food insecurity
  • Availability, accessibility and adequacy of health, training, transport and wastewater disposal (i.e. drinking water, wastewater and waste management)

6.2 Service-Based Process Chain with Encapsulated Modules

The geospatial software environment provides tools to define, compose, map, and execute workflows. Which allows for a cost-effective and easy to customise development, accessibility anywhere and to a range of devices. The web-based implementation also ensures increased interoperability, which is also reflected in simple commissioning and maintenance. The service-based approach is also characterised by increased security due to encapsulated components that can be flexibly composed and assembled [8].

One key component of the service platform is a powerful, eventually distributed database in which all geospatial data will be stored, harmonised, updated and managed. Along with the classical dimensions of space and time all available semantic attributes of the geospatial objects will be mapped. The basic data sets are acquired from high-resolution stereo aerial imagery and ancillary data from various sources. All data are collected and processed to the requirements of a data model specifically designed for that purpose and aligned to the EU INSPIRE conform schemes. An overview of the modular system component is given in Fig. 4.

For the generation of a virtual clone of an oriental city it is necessary to initially use a consistent reference data set. Since this is usually a problem, this approach also provides a stereophotogrammetric processing of aerial data sets. As an intermediate medium, a so-called point cloud is created, as described above, which is analysed using relatively simple algorithms in order to extract buildings in a first iteration. The parameters of area and angle are the most important in a neighbourhood environment. The point cloud is classified by spatial filter operations. In further steps, classified point clusters are simplified and aggregated. Finally, building outlines are reconstructed using polygon sections. Then building models are constructed according to the scheme of a Level-of-Detail 1 block. These raw buildings serve as a base geometry for a unique
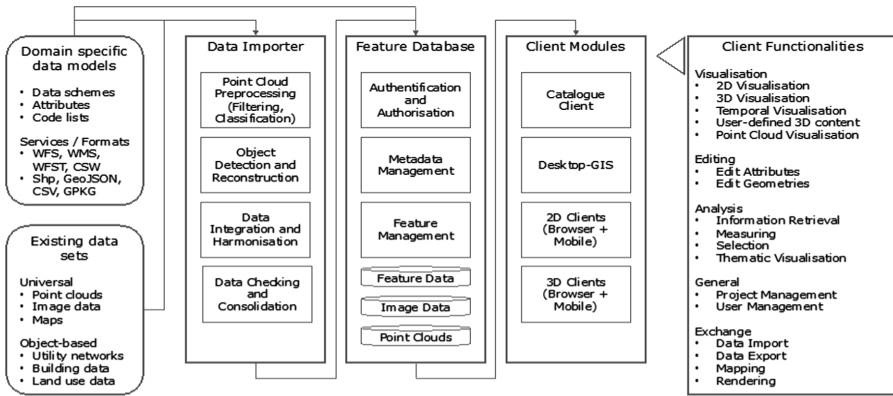
Fig. 4. Conceptual overview of relevant system components.

object, which can be semantically enriched in the following. The point cloud as such is also available as another highly accurate visual representation medium. The database act double as a central interface and anchor point for all project-relevant geodata. Robust and easy-to-use basic GIS functionalities are implemented for the provision, analysis and cartographic visualisation of this data. The modular encapsulated architectural and functional building blocks of the system are integrated into the geospatial software environment that support the decision-making-process. To be more precise: Standard interfaces are used to allow access to geodata from and within the spatial database and other tools operating on it that allow users to build 3D maps, e.g., for virtual reconstructions and future planning, to navigate the 3D environment, to access objects of interest and to request thematic data referenced by the objects. The system use means for the design, generation, and use of analytics maps, which represent thematic data such as urban functions, building and infrastructure conditions, where all is mapped to graphics attributes. Stakeholder-specific and application-specific map designs, including colour encodings and signatures, are implemented to allow users to produce their own maps according to their need and lastly to explore, compare, and assess large quantities of multi-dimensional geodata, its structures, and relations by projecting geospatial data to a non-geospatial reference frame using concepts of information cartography and advanced analytics functionalities such as functions to analyse demographic trends per community, ratios relating health services to number of people, housing scarcity, and plans for future housing development amongst other functions. An automated data processes can update the database content and trigger the generation and delivery of thematic maps as meaningful side-effect.

7   Conclusion

When fully operational, the web-based service platform allows for the guided composition and analysis of multidimensional, multifaceted virtual models of urban space by expert and non-specialist user groups. It will support both audiences to understand

and document the changing urban situation in East Jerusalem as well as to make informed decisions on urban planning and communal advocacy. It has been argued that the difficulties of a comprehensive database managing, and processing spatial data constitute a bottleneck in handling of spatial issues. This problem can be partially solved by elementarising several functions which are encapsulated in a modular way within a process chain embedded in a software environment. In an ongoing R&D project, the presented web-based approach will be further tested for its feasibility and advantages compared to previously applied limited methods.

Due to its historical development, East Jerusalem is facing dramatic development challenges that affect its urban fabric and community identifying a politically highly sensible environment. Urban agglomerations are evolving and abandoning early established urban concepts. As a result of inadequate urban planning with a strong focus on road infrastructure and housing, the study area also shows a weakly organised structure of neighbourhoods with different functions that lack meaningful distribution [1]. Urban development is poorly regulated as there are no development plans for the region. Some houses and buildings can be erected arbitrarily, which is observable. It is therefore extremely essential to establish a digital inventory of the region using modern technology and technical infrastructure. A first step has been taken with a modelled virtual representation of the study area in order to spatially assess and analyse future developments and to support the local society in their well-founded decision-making-process [8].

References

1. Loibl, W., Etminan, G., Gebetsroither-Geringer, E., Neumann, H.M., Sanchez-Guzman, S.: Characteristics of urban agglomerations in different continents: history, patterns, dynamics, drivers and trends. In: Urban Agglomeration. IntechOpen (2018)
2. Insee (Institut national de la statistique et des études économiques). https://www.insee.fr/fr/metadonnees/definition/c1501. Accessed 03 Apr 2019
3. Miller, H.J., Goodchild, M.F.: Data-driven geography. GeoJournal **80**(4), 449–461 (2015)
4. Mucksch, H., Behme, W. (eds.): Das Data Warehouse-Konzept: Architektur-Datenmodelle-Anwendungen. Springer, Heidelberg (2013)
5. Hart, G., Dolbear, C.: Linked Data: A Geographic Perspective. CRC Press, Boca Raton (2016)
6. Tolpekin, V.A. Stein, A.: The core of GIScience: a systems - based approach. ITC Educational Textbook Series; no. 2012. University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), Enschede (2012)
7. Burrough, P.A., McDonnell, R., McDonnell, R.A., Lloyd, C.D.: Principles of Geographical Information Systems. Oxford University Press, Oxford (2015)

8. Fricke, A., Döllner, J., Asche, H.: Servicification – trend or paradigm shift in geospatial data processing? In: Gervasi, O., et al. (eds.) ICCSA 2018. LNCS, vol. 10962, pp. 339–350. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-95168-3_23

9. Worboys, M.F.: A unified model for spatial and temporal information. Comput. J. **37**(1), 26–34 (1994)

10. Lü, G., et al.: Geographic scenario: a possible foundation for further development of virtual geographic environments. Int. J. Dig. Earth **11**(4), 356–368 (2018)

11. Galton, A.: Fields and objects in space, time and space-time. Spat. Cogn. Comput. J. **4**(1), 39–68 (2004)

12. Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., Sui, F.: Digital twin-driven product design, manufacturing and service with big data. Int. J. Adv. Manuf. Technol. **94**(9–12), 3563–3576 (2018)

13. Rigaux, P., Scholl, M., Voisard, A.: Spatial Databases: With Application to GIS. Morgan Kaufmann, San Francisco (2002)

14. Milz, D., Zellner, M., Hoch, C., Radinsky, J., Pudlock, K., Lyons, L.: Reconsidering scale: using geographic information systems to support spatial planning conversations. Plan. Pract. Res. **33**(3), 291–308 (2018)

15. Nama, G.F., Ulvan, M., Ulvan, A., Hanafi, A.M.: Design and implementation web based geographic information system for public services in Bandar Lampung City—Indonesia. In: 2015 International Conference on Science in Information Technology (ICSITech), pp. 270–275. IEEE, October 2015