TOWARDS AN INTELLIGENT PLATFORM
FOR BIG 3D GEOSPATIAL DATA MANAGEMENT

N. Mazroob Semnani[1], P. V. Kuper [1]*, M. Breunig [1], M. Al-Doori [2]

[1] Geodetic Institute, Karlsruhe Institute of Technology, Germany - (mazroob, kuper, martin.breunig)@kit.edu
[2] Dept. of Electrical and Computer Engineering, American University in Dubai, United Arab Emirates – maldoori@aud.edu

Commission IV, WG IV/7

KEY WORDS: Geo-Data Management, Big Geospatial Data, 3D Geodatabase, BIM, Intelligent Geospatial Data Analytics

ABSTRACT:

The use of intelligent technologies within 3D geospatial data analysis and management will decidedly open the door towards efficiency, cost transparency, and on-time schedules in planning processes. Furthermore, the mission of smart cities as a future option of urban development can lead to an environment that provides high-quality life along stable structures. However, neither geospatial information systems nor building information modelling systems seem to be well prepared for this new development. After a review of current approaches and a discussion of their limitations we present our approach on the way to an intelligent platform for the management and analysis of big 3D geospatial data focusing on infrastructure projects such as metro or railway tracks planning. three challenges are presented focusing on the management of big geospatial data with existing geo-database management systems, the integration of heterogeneous data, and the 3D visualization for database query formulation and query results. The approach for the development of a platform for big geospatial data analysis is discussed. Finally, we give an outlook on our future research supporting intelligent 3D city applications in the United Arab Emirates.

1. INTRODUCTION

The use of 3D intelligent technologies within planning processes will decidedly lead to efficiency, cost transparency, and on-time schedules. The mission of smart cities as a future option of urban development is to build an environment that provides high-quality life along a favourable, stable structure. To improve and simplify the decision-making processes, multiple technologies work collaboratively within the construction of a city. Different vendors of diverse fields such as IT, energy, and infrastructure provide more and more collaborative solutions to achieve steadily developments for better cities. This procedure leads to multiple systems and the collection of huge data sets, which demand a professional organized data management. The handling and integration of such big city data is a challenge especially for the design and maintenance of smart cities. Obviously the long-time data-availability for all participants within a construction project leads to an easier and earlier determining of risks and planning costs.

This paper is structured as follows: In section 2 we refer to related work in the field of big 3D geospatial data management and analysis, respectively, followed by the restrictions of current solutions in section 3. In section 4 challenges for the management and analysis of big geospatial data are discussed. First results on the management of big 3D data based on the geo-database management system PostGIS are presented. In section 5 our approach to big geospatial data analysis is presented. Finally, section 6 summarizes the paper and gives an outlook on our future research.

2. RELATED WORK: MANAGEMENT AND ANALYSIS OF BIG GEOSPATIAL DATA

Nowadays Geospatial Information Systems (GIS) can manage relatively small spatial data sets up to 2.5D, however the capabilities for handling big 3D data is much more challenging. Since some years, new achievements to support 3D objects and buildings with multi-patch feature classes are in development (Schön et al., 2009). Nevertheless, the sufficient support for the storage and management of big 3D geospatial data is still a major research topic (Breunig et al., 2016; Sugumaran et al., 2012; Breunig and Zlatanova, 2011; Schön et al 2009; Liu et al., 2009).

Striking open standards in the BIM and GIS domains, respectively, such as the Industry Foundation Classes (IFC) (IFC, 2018; buildingSMART International, 2018) and CityGML (CityGML, 2018) are providing a framework for the integration of objects from the built environment (Biljecki, 2017). Obviously these standards have to be adapted for the modelling of big geospatial data, having in mind that there is no obvious frontier to divide big data and regular data (Chen et al., 2014). It is well known that Laney characterized big data by his "3Vs model" including volume, velocity, and variety (Laney, 2001, Berman, 2013). For the first time, other characteristics rather than volume have been defined for big data. Afterwards, multiple studies added veracity (IBM Big Data & Analytics Hub, 2018), value and variability to the V model (McNulty, 2014). Because of their complexity, the handling of big data systems demands particular techniques and algorithms supporting data streaming and parallel computing (Amirian et al., 2014). A constructive building model with unified building information must provide a synchronized database capable to be accessed simultaneously by all participants of a construction

---

* Corresponding author

project (Breunig et al., 2017). Furthermore, the modelling and development of a mathematical representation for any complex surface in 3D space or any volume, e.g. to represent buildings, are challenging tasks. Also modelling the variations of a building can lead to benefits but also to new problems in urban planning, mapping, and visualization, emergency response, etc. (Volk et al., 2014).

Intelligent buildings, smart infrastructures, and smart cities require 3D data for more convenience. For example, the application of BIM technologies at an airport terminal prior to the construction may support the simulation and prediction of directions for passengers within the building. Decision makers may examine how a building operates and may come to accurate assessments of the building plan. With the increasing use of information modelling in Architecture Engineering Construction (AEC), various governments push for the obligatory use of BIM to improve the quality of buildings and to reduce cost (Digital Built Britain, 2015; Federal Ministry of Transport and Digital Infrastructure in Germany (BMVI), 2015; Building Information Modeling (BIM) e-Submission, 2018).

There are several attempts to store and manage big BIM data: Software products such as MapReduce® and Bigtable® are used in several studies as DBMS solution (He et al., 2008; Ekanayake et al., 2008; Schatz, 2009; Zhao et al., 2009; McKenna et al., 2010; Lin et al., 2010; Taylor, 2010; Seo et al., 2010; Xiaoqiang et al., 2010; Yu et al., 2012). Project participants and engineers develop BIM models with the help of commercial BIM software and utilize data centres such as CloudBIM® to share their 3D models and parsed building information with subcontractors and co-workers. BIMServer® is a popular open source model server supporting IFC (Beetz, et al, 2010). There are several examples of frameworks for the storage and analysis of massive BIM data. Hung-Ming Chen and others developed a BIM data centre that can be accessed by multiple users and is able to manage massive BIM data by using cloud technologies. A web-based visualization using Web 3D technologies is available by Chen et al. (2016). Chen addressed cloud (network) computing technologies as a solution to resolve the limits of stand-alone systems. Since the web-based user interface and the display of an individual query do not need high-performance hardware, they still can be applied simultaneously by multiple users. Solihin et al. (2016) observed trends of the continuous growth of BIM models, e.g. caused by the integration of building sensor data that lead to additional data represented as point clouds. Multiple frameworks such as the SOA4BIM® framework (Jardim-Goncalves et al., 2010), the conceptual framework proposed by Amarnath et al. (2011), the BIM visual system developed by Chuang et al. (2011), and the framework proposed by Wu et al. (2012) were developed for the storage and management of 3D BIM data in general.

However, to provide solutions for the operative management of big 3D building information models, data storage issues have to be considered by experts of both the BIM and the GIS domain. The cooperation with users in real projects is essential to find practical solutions. Furthermore, building information has a continuous information chain in the life-cycle of objects. Within this life-cycle the planning and in-situ data are changing continuously and they must be stored and retrieved accordingly. To process both in a unified way, BIM has to be integrated with other technologies such as GIS, point cloud processing as well as virtual (VR) and augmented reality (AR). E.g. the project 'Future City Pilot' ran by the Open Geospatial Consortium (OGC) observes the possibility of the integration between GIS

and BIM in urban planning projects (Open Geospatial Consortium, 2016).

The Analysis of Big Data requires the utilization of parallel processing on groups of servers, see Figure 1. To accomplish this, it is important to monitor all the components of the inquiry and to combine the outcomes into a dataset. Google® has introduced several tools to accomplish this task. Most likely, the best-known Big Data instrument available is the Apache Establishment's Hadoop®. The essential part of this tool's calculation is to deal with the coordination and management of all the distributed analysis processes working in parallel, also to deal with adaptation to internal failure and excess. The analysis of big geospatial data results in the classification and sorting of source data into streams of data which are then passed to a tree of specialist servers dedicated to deploy a suite of big data analytics. These servers will pass their outcome back to the main server, which will consolidate all results of data mining tasks carried on source data into a final outcome to the original query utilizing spatial predicates. Hadoop can be utilized to run examination on greatly vast volumes of information on server groups with any number of hubs. However, Hadoop is not designed to utilize Geospatial predicates on inquiries or applications.
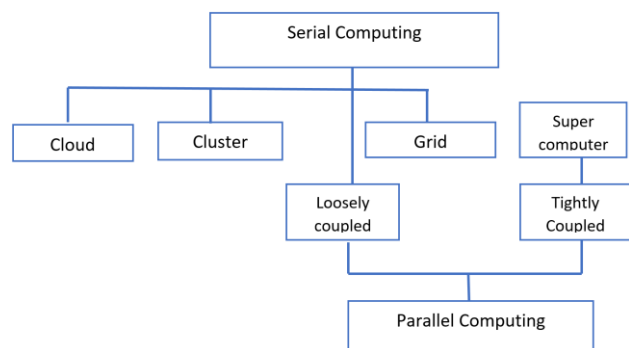

Figure 1. Computational arrangement in clusters, grids, clouds, and supercomputers (adapted from (Karimi, 2014))

A solution emerged from Esri® to this problem is 'GIS Tools for Hadoop', a toolbox using Esri Geometry API for Java® to provide Hadoop with vector geometry tasks. Furthermore, Spatial Framework® for Hadoop, which empowers Hive Query Language® (HQL) to utilize spatial information tasks and Geoprocessing Tools for Hadoop®, is an arrangement of geoprocessing devices for ArcGIS® that empowers clients to move their information all through Hadoop to execute work processes. Utilizing these devices, it is conceivable to take information held in a spatial data mine, bundle and transfer it into a Hadoop bunch. Complex investigation can at that point be performed on the information, and the aftereffects of the investigation downloaded straightforwardly into ArcGIS for Desktop where additional detailed examinations can be performed. This toolbox provides a suitable approach for handling the analysis of big geospatial data, as large volumes can be reduced into a more manageable subset on which definite spatial examination can be performed. For instance, it could be used to perform the starting point / target examination on a high volume of traffic data by handling and breaking down a large number of GPS directions over the course of a day. The GPS instructions could be reduced to a suitable size applicable to a particular region by transferring the spatial determination queries to Hadoop. The chosen subset can then be brought over

into ArcGIS work areas where a more specific system investigation has to be performed.

3. LIMITATIONS AND ISSUES OF CURRENT SOLUTIONS

The current handling of 3D geospatial data related to the BIM and the GIS domain rises multiple challenges and issues that need to be addressed. One of these is the integration of BIM and GIS data. Both domains show similarities and connections when affecting infrastructure and buildings. The data integration between GIS and BIM is therefore very valuable for upcoming 3D city modelling projects. It has been suggested that the elaborated BIM data can provide the fine details that the city models of the GIS domain are usually lacking (Ohori et al., 2017). Biljecki (2017) discussed the level-of-detail problem in 3D city models. However, there are some critical problems that need to be addressed. Ohori et al. (2017) studied geometric and topological issues such as bad georeferenced BIM data as well as geometric and topological problems in BIM models from the perspective of GIS. In this process, they discovered a suitable method for the transformation of data between the IFC format and the CityGML format. They also mentioned further problems that are common in the GIS domain. For example, intersections between objects and self-intersection as well as different objects that are shown as one object or vice versa. Additionally, some non-planar faces occur as flat surfaces in GIS applications. Some of these errors can lead to errors during advanced spatial analysis in the GIS field.

One of the major drawbacks in using commercial BIM servers such as Autodesk Revit BIM Server® (Autodesk Revit, 2018), Graphisoft ArchiCAD BIM Server® (Graphisoft ArchiCAD, 2018) or Bentley ProjectWise Integration Server® (Bentley, 2018) seems to be that they operate on a single computer only. With these servers, it is necessary to download an entire BIM file to view or query the model. However, a dynamic model that changes over time and continually expands with the continuous monitoring of a BIM project, must be stored and managed effectively and accurately presented. Jiao et al. (2013) designed a cloud approach to solve the big-data management problem in AEC applications. But there are still multiple obstacles in hosting and managing large amounts of BIM data. In most applications it is necessary to process big data with the help of parallel algorithms and in heterogeneous networks. Therefore, ordinary databases are usually not compatible with big data solutions (Correa, 2015).

The service developed early by Dean et al. (2008) uses Apache Hadoop® to stabilize multiple distributed servers for a BIM data centre and utilizes MapReduce® for the parallel processing of big dynamic BIM datasets. They outline several reasons that reliable management of large BIM datasets can be realized with CloudBIM®. Two major drawbacks of using such systems, however, are the poor computing resources and the access restriction by a single user.

4. CHALLENGES FOR THE MANAGEMENT AND ANALYSIS OF BIG GEOSPATIAL DATA

In our research we identified three major challenges concerning the processing of big 3D geospatial data.

4.1 First challenge: Managing big geospatial data with existing geo-database management systems

The first challenge we target is the management of large amounts of geospatial data with existing geo-database management systems. For a typical application, e.g. a railway track planning and construction project, we expect an amount of ~50TB of thematic and geospatial data. Such an amount of data cannot be handled by a database working on one single computer. Therefore we need to use distributed databases that are scalable and work on clusters. However, such distributed database solutions often do not support spatial or even spatio-temporal database queries.

Also the PostGIS/PostgreSQL community advances solutions for a fast access on mass data by extending existing multi-dimensional access methods such as the R-Tree and the more abstract access method GiST (Generic Index Structure), PostgreSQL supports a new index type, called BRIN. BRIN stands for Block Range Index and enables a generic indexing taking slight physical memory. It obviously is useful for large tables. BRIN splits the table into block ranges and compresses entities into min and max values. It supports the horizontal partitioning of a database and can improve the handling of large datasets in terms of writing and querying. One of the limitations of the GiST-based index implemented in PostGIS is the size of the RAM that is available. Thus it does not work smoothly, once the available amount of RAM exceeds. The continuous update of object values or update operations decrease the efficiency of the GiST-index. I.e. the index must be updated manually by calling the method REINDEX. With BRIN, the computed minimum bounding boxes of the managed objects are mutually exclusive. Therefore, the resulting index will be smaller, and the index structure can be more efficient than GiST in cases we deal with overlapping data. In such cases the GiST index – in PostgreSQL implemented on top of the R-Tree – is disadvantaged as sub-trees of the R-Tree must be searched multiple times. However, as our research showed, for practical real data sets this is not necessarily an issue and it is worth to use a GiST-based R-Tree index. Comparing both index structures, the BRIN technique takes less time for creating compared to the GiST structure (Mazroobsemnani, 2017). Furthermore, parallel query execution is provided (PostgreSQL, 2018).

Since the support for 3D geospatial data types is still not fully evolved, there is a demand to improve the handling of huge amounts of 3D spatial data within spatial databases. Furthermore, robust topology structures for 3D objects, especially for various parts of a building, should be explored within existing geo-database management systems to improve the performance of the 3D query operations.

4.2 Second challenge: Integration of heterogeneous data

The second challenge we face is to handle heterogeneous geospatial data, e.g. raster images, 2D shapes and 3D solids simultaneously. Each model has its own requirements and needs to be handled properly. Furthermore, we need to reflect the links between such data, e.g. from raster images to 3D solids and vice versa. In the following we refer to a typical application scenario of a 3D railway tracks project consisting of planned and existing tracks. The data to be expected come in different data formats and primarily focus on raster data:

a) Data from helicopter flights
Large-area aerial survey of new routes and existing sections represented as Triangulated Irregular Networks (TIN) in formats such as *.ecw for the digital orthophotos or the raw data from laser scans in the format *.asc.

b) Data from drone flights
Elaborated shots from new tracks in the connection area between an existing and a new route represented as a digital terrain model in formats such as *.stl *.dwg, *.fbx, point clouds in formats such as *.e57, *.las, *.rcp, and digital orthophotos in formats such as *.tiff and *.jpg.

c) Data from stationary laser scans
Detailed recording of the existing bridges with digital terrain model in a format such as *.dwg, point clouds in formats such as *.las and *.rcp, raw laser scans data in the format *.asc.

d) 2D GIS Data
Various information and shape files of the network in the format *.shp and oracle dump *.dmp.

e) Track geometry
Laser scan data of the existing tracks, taken with a light measuring system in the format *.asc.

f) Traditional 2D plans
Plans of the existing network from archives in the formats *.dwg and *.pdf or as paper plans.

g) Data from already existing models
Based on existing 2D plans and 3D existing records, object-oriented models in formats such as: *.ifc, *.rvt, and *.pdf being created for the entire planning area.

h) Data from 3D, "4D", and "5D" planning models
For planning purposes, complex models have to be created, which usually consist of subject-specific sub-models, created by different specialist planners. The 3D, "4D" (3D + time), and "5D" (3D + time + costs) models are created using various applications such as Autodesk Revit®, Autodesk Navisworks®, and Autodesk Civil 3D®.

Obviously data integration coming from heterogeneous data sources can only be successful, if the data refer to the same examination area, if there are no semantic conflicts between different data, and if the geometric raster representations of the data are compatible to each other. It must be ensured that the data collection is unified and valid.

4.3 Third Challenge: 3D visualization of database query formulation and query results

For many DBMS the query language SQL is used as a well-established standard. Spatial extensions of SQL such as Geo-SQL® or Spatial SQL® are used to support the analysis of geospatial data. But when it comes to handle heterogeneous data with different data formats, an adjusted spatial access method is needed. Our approach to overcome this problem is to design and implement a spatio-temporal access method with a graphical 3D/4D interface that allows an appropriate visualization for the formulation and result-visualization of 3D/4D database queries.

To identify the necessary functionality needed for the visualization of the query formulation and database query results, respectively, the special requirements for 3D/4D queries have been identified. Typical requirements are:

- Changing the colours of single objects when being touched in the 3D cave;
- selecting sets of objects in space and time by defining rulers indicating spatial and temporal intervals:
- computing the distance between two objects by touching them in the 3D cave;
- selecting the topological neighbours of an object by touching an object and the "neighbours" or "distance" button;
- selecting a spatial and temporal minimal bounding box around an object as studied by Ohori et al. (2017). This supports the process of making explicit and discrete bounding representation geometries from the implicit and curved geometries. A main research question is how to support 3D and 4D database queries in a way that the users can intuitively examine the results of database queries.

The 3D visualization can also be well used as a preliminary step to examine the meta data before analysing the real data: due to the preparation cost of large amounts of data, referred to as big data, it is advantageous to view the data before downloading the files. The 3D viewer needs to support open and proprietary file formats, must guarantee the data visibility through open formats and allow users to select a specific format for download. We are aware that this solution may have limitations on uploading and downloading data.

5. FIRST STEPS OF OUR APPROACH

5.1 Tests with PostGIS/PostgreSQL

First, we studied the capabilities of PostGIS® for the management of big 3D spatial data focusing on issues such as spatial indexing, native partitioning, logical replication, and smart statistics for the query planner. We focused on the management of large amounts of point cloud data. However, PostGIS does not provide a dedicated data type and function for point cloud data. Thus the open source libraries PDAL® and pgpointcloud® were used to handle 3D point cloud data via PostGIS. Treating a point cloud merely as points can be challenging (Van Oosterom et al., 2015). With PDAL® and pgpointcloud® it is possible to define sets of points (point patches) to improve the data management performance. In order to support the access of a part of the point cloud, the data is retrieved by a 3D bounding box. PostgreSQL supports 3D features, data types and spatial indexes. However, a 3D point cloud query may have a bad performance for big database records even if it uses a proper index structure (Hinks et al. 2012; Van Oosterom et al., 2015). Several other DBMS face performance difficulties, while working with databases over 2GB (Zhang et al., 2016).

The real dataset we used was chosen from the spatial component of a learning LIDAR dataset provided by the Institut National de l'Information Géographique et Forestière (IGN, 2018) and the open data 3D building model from New York City (NYC, 2018). With the real dataset it takes approximately eleven times less time for creating the index with BRIN than with GiST, cf. Figure 2. The figure shows (f.l.t.r.) the total time for the index generation, the storage space that is needed by index and the average time that is needed for topological queries.
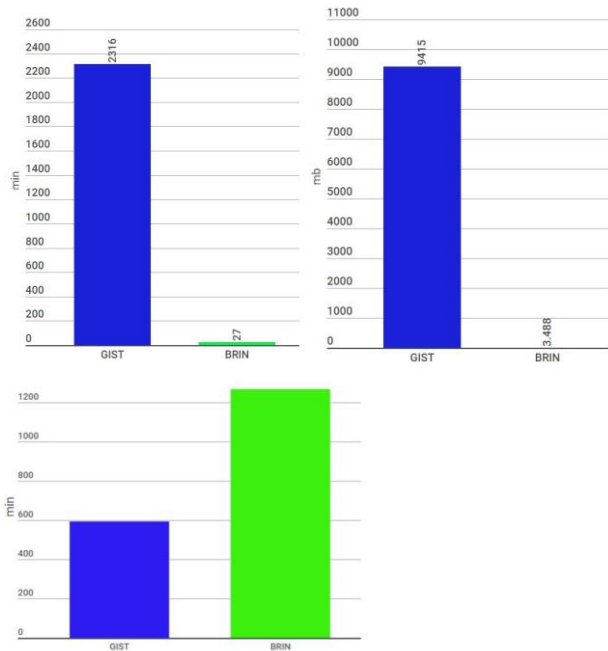
Figure 2. Comparison GiST vs. BRIN

In our previous work, within the DFG research group "Computer-aided collaborative subway track planning in multi-scale 3D city and building models" (Breunig et al., 2017), we focus on how 3D data can be queried and processed automatically from a collaborative data platform. This includes the support of transaction techniques for multi-scale modelling and the supply of spatio-temporal access methods such as examined in (Menninghaus et al., 2016). By scaling up solutions for the management of large amounts of 3D geospatial city and infrastructure data to be used in academia and practice, a prototype of a data integration platform for the management of large amounts of heterogeneous 3D spatial data will be developed. The primary target of the prototype is the support of data formats that are used for BIM processes in infrastructure planning, particularly railway tracks and tunnels.

5.2 Data integration and 3D visualization

Typical use cases to be covered by a data integration platform are the data import and the export, the preparation of the data for participating project partners, the 3D visualization of the data and the control and management of heterogeneous data types and formats.

An overview of the potential software architecture to be used for our data integration platform can be found in Figure 3.
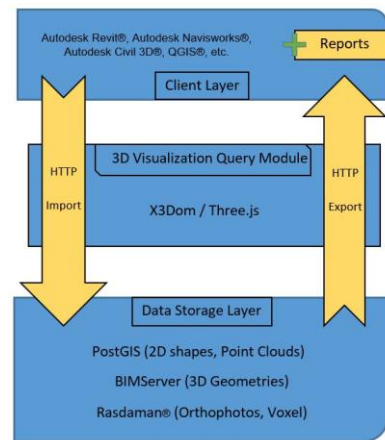

Figure 3. Software configuration to support visually supported 3D database queries

In the long run we propose to integrate the heterogeneous data by a 3D visualization interface (3D cave) that enables the user to "dunk" into the 3D model and to inspect the different data by defining metric, geometric and topological methods.

5.3 Intelligent platform for data analysis

Our research is targeted on the development of a smart platform to employ new methodologies that will facilitate interactive and collaborative queries, and at the same time support the fast nature of big geospatial queries and geospatial big data. This platform needs to include smart visual interfaces and analytical reasoning algorithms to enable the user to effectively interpret long-term analytical processes using complex spatio-temporal big data. There is a need to evaluating the ability of current visual methods in thematic GIS to support geospatial big data analytics. The intelligent platform will adapt GIS generalization principles and techniques to support visual geospatial big data analytics. It will combine computational methods and GIS best practices into an intelligent platform that is capable of proposing the sound design and prediction of real time decisions.

The intelligent platform will incorporate methods that will embody the volume of geospatial big data and deploy the accurate methodology to the selection of the appropriate choice of geospatial big data overviews and create those at real time. The new platform will demonstrate new interfaces capable of handling the complexity and interpretation of spatio-temporal big data and its change over time. It is also necessary to address in new intelligent innovations the need for new methods to enhance the management of predictive analytics of dynamic events with visualizations. These visualizations will leverage geospatial big data by easing the process of handling such data for users. It is also necessary to address the issue of the velocity of the data so that intelligent applications will identify changes and the significance of the level of change dynamically from real time dynamic geospatial big data sources. Furthermore, the reliability of geospatial big data sources has to be assessed using sound data modelling techniques and assessing the certainty of geospatial big data.

Last, but not least, new innovations will also focus on incorporating analytical methods such as Post Markov Assumption, Estimate Neighbour Relationship from geospatial big data, and Place based Ensemble Models to address spatial heterogeneity. By this, an intelligent suite of algorithms will be

provided to facilitate the analysis of geospatial big data with ease.

6. CONCLUSIONS AND OUTLOOK

After having presented the state of the art and the limitations of existing approaches to manage and analyse geospatial data, we presented three challenges and consequently our approach on the way to solve the deficits and to develop an intelligent platform for the management and analysis of big 3D geospatial data. We focused on data typically used in metro and railway projects. The challenges considered in this paper concern the management of big geospatial data with existing geo-database management systems, the integration of heterogeneous data, and the 3D visualization of database query formulation and query results. Single steps towards the development of a platform for big geospatial data analysis have been discussed.

In our future research we will focus on the development of new methods for parallel geo-databases supporting the parallel execution of metric, geometric, and topological database queries used for data analytics. Furthermore, we intend to extend the intelligent platform for the visualization of spatial 3D operators and 3D query results. Also the integration of BIM and geo-referenced 3D GIS data will be an issue. Finally, it is our goal to apply the developed methods for intelligent 3D city applications in the United Arab Emirates.

ACKNOWLEDGEMENTS

REFERENCES

Amarnath, C. B., Sawhney, A., & Maheswari, J. U., 2011. Cloud computing to enhance collaboration, coordination and communication in the construction industry. In: *Information and Communication Technologies (WICT)*, 2011 World Congress on IEEE, pp. 1235-1240.

Amirian P., Basiri A., Winstanley A., 2014. Evaluation of Data Management Systems for Geospatial Big Data. In: *Computational Science and Its Applications – ICCSA 2014*, Lecture Notes in Computer Science book series (LNCS), Vol. 8583, pp. 678-690.

Arroyo Ohori, K., Biljecki, F., Diakité, A., Krijnen, T., Ledoux, H., and Stoter, J., 2017. Towards an integration of GIS and BIM data: What are the geometric and topological issues. In: *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, Proceedings of ISPRS 12th Geoinfo Conference, Melbourne, Australia, pp. 26-27.

Autodesk Revit, 2018. www.autodesk.com/products/autodesk-revit-family/overview (12 May 2018).

Building Information Modeling (BIM) e-Submission, 2018. https://www.corenet.gov.sg/general/building-information-modeling-(bim)-e-submission.aspx (12 May 2018).

Beetz, J., van Berlo, L., de Laat, R., & van den Helm, P., 2010. BIMserver.org–An open source IFC model server. In: Proceedings of the 27th International Conference on Applications of IT in the AEC Industry, CIB-W78, Cairo, Egypt, pp. 1-8

Bentley, 2018. ProjectWise https://www.bentley.com/en/products/brands/projectwise (12 May 2018).

Berman, J. J., 2013. Principles of big data: preparing, sharing, and analyzing complex information. Morgan Kaufmann, 288 p.

Biljecki, F., 2017. Level of detail in 3D city models, PhD thesis, Delft University of Technology, 353 p.

Digital Built Britain, 2015. Level 3 Building Information Modelling – Strategic Plan. https://www.cdbb.cam.ac.uk/Resources/ResoucePublications/bis15155digitalbuiltbritainlevel3strategy.pdf (12 May 2018).

Breunig, M., & Zlatanova, S., 2011. 3D geo-database research: Retrospective and future directions. In: *Computers & Geosciences*, Vol. 37(7), pp. 791-803.

Breunig, M., Kuper, P. V., Butwilowski, E., Thomsen, A., Jahn, M., Dittrich, A., Menninghaus, M., 2016. The story of DB4GeO–A service-based geo-database architecture to support multi-dimensional data analysis and visualization. In: *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 117, pp. 187-205.

Breunig, M., Borrmann, A., Rank, E., Hinz, S., Kolbe, T., Schilcher, M., Mundani, R.-P., Jubierre, J. R., Flurl, M, Thomsen,A, Donaubauer, A, Ji, Y, Urban, S, Laun, S, Vilgertshofer, S, Willenborg, B, Menninghaus, M, Steuer, H, Wursthorn, S, Leitloff, J, Al-Doori, M, and Mazroob Semnani, N, 2017. Collaborative Multi-Scale 3d City and Infrastructure Modeling and Simulation. In: *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 42, pp. 341-352.

BuildingSMART International, IFC: Industrial Foundation Classes, 2018. www.buildingsmartalliance.org (12 May 2018).

Chen, M., Mao, S., & Liu, Y., 2014. Big data: A survey. In Mobile networks and applications, 19(2), pp. 171-209.

Chen, H. M., Chang, K. C., & Lin, T. H., 2016. A cloud-based system framework for performing online viewing, storage, and analysis on big data of massive BIMs. In: *Automation in Construction*, Vol. 71, pp. 34-48.

Chuang, T. H., Lee, B. C., & Wu, I. C., 2011. Applying cloud computing technology to BIM visualization and manipulation. In: *Proceedings of 28th International Symposium on Automation and Robotics in Construction*, Vol. 201, No. 1, pp. 144-149.

CityGML. http://www.citygmlwiki.org (12 May 2018).

Correa, F. R., 2015. Is BIM Big Enough to Take Advantage of Big Data Analytics. In: *ISARC Proceedings of the International Symposium on Automation and Robotics in Construction*, Vol. 32, Vilnius Gediminas Technical University, Department of Construction Economics & Property, pp. 1-8.

Dean, J., & Ghemawat, S., 2008. MapReduce: simplified data processing on large clusters. In: *Communications of the ACM*, 51(1), pp. 107-113.

Ekanayake, J., Pallickara, S., & Fox, G., 2008. Mapreduce for data intensive scientific analyses. In: *eScience'08. IEEE Fourth International Conference on IEEE*, pp. 277-284.

Federal Ministry of Transport and Digital Infrastructure in Germany (BMVI), 2015. Stufenplan Digitales Planen und Bauen.
https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/stufen plan-digitales-bauen.pdf?__blob=publicationFile (12 May 2018).

Graphisoft ArchiCAD, 2018. http://www.graphisoft.com/bim-server (12 May 2018).

Guttman, A., 1984. R-trees: A dynamic index structure for spatial searching Vol. 14, No. 2, ACM, pp. 47-57.

He, B., Fang, W., Luo, Q., Govindaraju, N. K., & Wang, T., 2008. Mars: a MapReduce framework on graphics processors. In: *Proceedings of the 17th international conference on Parallel architectures and compilation techniques*, ACM, pp. 260-269.

Hinks, T., Carr, H., Truong-Hong, L., & Laefer, D. F., 2012. Point cloud data conversion into solid models via point-based voxelization. In: *Journal of Surveying Engineering*, Vol. 139(2), pp. 72-83.

IBM Big Data & Analytics Hub. The Four V's of Big Data. http://www.ibmbigdatahub.com/infographic/four-vs-big-data (12 May 2018).

IFC, 2018. IFC Wiki http://www.ifcwiki.org/index.php?title=IFC_Wiki (12 May 2018).

IGN, TerraMobilita 3D urban analysis, http://data.ign.fr/benchmarks/UrbanAnalysis/ (12 May 2018).

Jardim-Goncalves, R., & Grilo, A., 2010. SOA4BIM: Putting the building and construction industry in the Single European Information Space. In: *Automation in Construction*, Vol. 19(4), pp. 388-397.

Jiao, Y., Wang, Y., Zhang, S., Li, Y., Yang, B., & Yuan, L., 2013. A cloud approach to unified lifecycle data management in architecture, engineering, construction and facilities management: Integrating BIMs and SNS. In: *Advanced Engineering Informatics*, Vol. 27(2), pp. 173-188.

Karimi, H.A., 2014. Techniques and Technologies in Geoinformatics, CRC Press, 312 p.

Kang, H. S., & Lee, G., 2009. Development of an object-relational IFC server. In: Proc. of the 3rd International Conference on Construction Engineering & Management/6th International Conference on Construction Project Management, (ICCEM/ICCPM)

Laney, D., 2001. 3D data management: Controlling data volume, velocity and variety. META Group Research Note, 6(70), pp. 1-4.

Lin, J., & Dyer, C., 2010. Data-intensive text processing with MapReduce. Synthesis Lectures on Human Language Technologies, 3(1), 177p.

Liu, G., Zhu, Q., He, Z., Zhang, Y., Wu, C., Li, X., & Weng, Z., 2009. 3D GIS database model for efficient management of large scale underground spatial data. In: *Geoinformatics, 2009 17th International Conference on IEEE*, pp. 1-5.

Mandelbrot, B., 1982. The fractal geometry of nature. New York, NY: W.H. Freeman, 14 p.

Mazroobsemnani, N., 2017. Towards the management of big 3D geospatial data for building information modelling (BIM) with PostGIS, Master Thesis, Chair of Geoinformatics, Geodetic Institute, Karlsruhe Institute of Technology, 64 p.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A.,Garimella, K., Altschuler, D., Gabriel, S., Daly, M. & DePristo, M. A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. In: *Genome research*, Vol. 20(9), pp. 1297-1303.

McNulty, E. H., 2014. Understanding Big Data: The seven V'S http://dataconomy.com/2014/05/seven-vs-big-data/ (12 May 2018).

Menninghaus, M., Breunig, M., & Pulvermüller, E., 2016. High-Dimensional Spatio-Temporal Indexing. In: *Open Journal of Databases (OJDB)*, Vol 3(1), pp. 1-20.

NYC, 3-D Building Model. http://www1.nyc.gov/site/doitt/initiatives/3d-building.page (12 May 2018).

Open Geospatial Consortium, 2016. Future City Pilot http://www.opengeospatial.org/projects/initiatives/fcp1 (12 May 2018).

PostgreSQL, 2018. Documentation PostgreSQL 10, How parallel query works, Chapter 15. www.postgresql.org

Rafiee, A., Dias, E., Fruijtier, S., & Scholten, H., 2014. From BIM to geo-analysis: View coverage and shadow analysis by BIM/GIS integration. In: *Procedia Environmental Sciences*, Vol. 22, pp. 397-402.

Schatz, M. C., 2009. CloudBurst: highly sensitive read mapping with MapReduce. In: *Bioinformatics*, Vol. 25(11), pp. 1363-1369.

Schön, B., Laefer, D. F., Morrish, S. W., & Bertolotto, M., 2009. Three-dimensional spatial information systems: State of the art review. In: *Recent Patents on Computer Science*, Vol. 2(1), pp. 21-31.
Seo, S., Yoon, E. J., Kim, J., Jin, S., Kim, J. S., & Maeng, S., 2010. Hama: An efficient matrix computation with the mapreduce framework. In: *Cloud Computing Technology and Science (CloudCom)*, 2010 IEEE Second International Conference, pp. 721-726.

Solihin, W., & Eastman, C., 2016. A Simplified BIM Model Server on a Big Data Platform. In CIB W78 2015.

Solihin, W., Eastman, C., & Lee, Y. C., 2016. A framework for fully integrated building information models in a federated environment. In: *Advanced Engineering Informatics*, Vol. 30(2), pp. 168-189.

Sugumaran, R., Burnett, J., & Blinkmann, A., 2012. Big 3D spatial data processing using cloud computing environment. In: *Proceedings of the 1st ACM SIGSPATIAL international workshop on analytics for big geospatial data*, ACM, pp. 20-22.

Taylor, R. C., 2010. An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics, In: *Proceedings of the 11th Annual Bioinformatics Open Source Conference (BOSC) 2010*, BioMed Central, 6p.

Van Oosterom, P., Martinez-Rubi, O., Ivanova, M., Horhammer, M., Geringer, D., Ravada, S., Tijssen, T., Kodde, M. & Gonçalves, R., 2015. Massive point cloud data management: Design, implementation and execution of a point cloud benchmark. In: *Computers & Graphics*, Vol. 49, pp. 92-125.

Vaquero, L. M., Rodero-Merino, L., Caceres, J., & Lindner, M., 2008. A break in the clouds: towards a cloud definition. In: *ACM SIGCOMM Computer Communication Review*, Vol. 39(1), pp. 50-55.

Volk, R.; Stengel, J.; Schultmann, F., 2014. Building Information Models (BIM) for existing buildings – literature review and future needs. In: *Automation in Construction*, Vol. 38, DOI: 10.1016/j.autcon.2013.10.023, pp.109-127.

Wu, Wei, and R. R. A. Issa., 2012. Leveraging cloud-BIM for LEED automation., In: *Journal of Information Technology in Construction (ITcon)*, Vol. 17.24, pp. 367-384.

Xiaoqiang, Y., & Yuejin, D., 2010. Exploration of cloud computing technologies for geographic information services. In: *Geoinformatics*, 18th International Conference, pp. 1-5.

Yu, B., Cuzzocrea, A., Jeong, D., & Maydebura, S., 2012. On managing very large sensor-network data using bigtable. In: *Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, IEEE Computer Society, pp. 918-922.

Zhang, X., Yi, Z., Yan, Z., Min, G., Wang, W., Elmokashfi, A., Zhang, Y., 2016. Social computing for mobile big data. In: *Computer*, Vol. 49(9), pp. 86-90.

Zhao, W., Ma, H., & He, Q., 2009. Parallel k-means clustering based on mapreduce. In: *IEEE International Conference on Cloud Computing*, Springer, Berlin, Heidelberg, pp. 674-679.