Towards First Urban Data Space in Bulgaria

Vassil Vassilev
London Metropolitan University &
GATE Institute – London & Sofia
v.vassilev@londonmet.ac.uk

Dessislava Petrova-Antonova
GATE Institute - Sofia University
Sofia, Bulgaria
dessislava.petrova@gate-ai.eu

Tarun Bali, Sorin Radu
GATE Institute - Sofia University
Sofia, Bulgaria
{tarun.bali, sorin.radu}@gate-ai.eu

Viktor Sowinski-Mydlarz,
Dion Mariyanayagam
London Metropolitan University &
GATE Institute – London & Sofia
{w.sowinskimydlarz,d.mariyanayagam}
@londonmet.ac.uk

Evgeny Marinov, Petar O. Hristov
GATE Institute - Sofia University
Sofia, Bulgaria
{evgeniy.marinov,petar.hristov}
@gate-ai.eu

Sabin Nakarmi, Monika Rabka
GATE Institute - Sofia University
Sofia,Bulgaria
{sabin.nakarmi,monika.rabka}
@gate-ai.eu

***Abstract*— Smart city is no longer fiction, which targets the future – it is now around us and requires immediate action in different directions: environment, transport, energy, social and cultural life, healthcare, etc. This article presents the current efforts of GATE Institute at Sofia University for building an urban data space based on data from a variety of sources in the Bulgarian capital. It has an open architecture, based on a private cloud, which allows the integration of diverse data and provides different data processing capabilities and services necessary to build integral data spaces. The pilot implementation currently performs monitoring and analysis of the environmental factors in Sofia using a set of bespoke components, which implement data management and data analysis algorithms from simple filtering and correlation to data analytics and prediction using historical data, static modelled data and dynamic data from environmental sensors in real-time. It serves as a basis for data enrichment based on different sources and cross-domain analysis using a variety of methods. This new opportunity has a huge potential and will have a significant impact on urban life – from planning the infrastructure and managing the communal services to the personalization of social services for the citizens.**

Keywords— AI-powered smart city services, Data platform, Urban data space, Environment monitoring, Urban analysis and simulation

I. INTRODUCTION

The use of digital services in an urban environment has been around for many years on different levels and for a variety of reasons. It has been used for monitoring and control of the environment and the urban systems as a whole, for optimization of utilization of resources and investments, for improving the development process and decision-making practices, etc. The progress in information technologies over the last twenty years (esp. Cloud Computing and Internet of Things), the maturity of the software engineering methodologies (component software, service-oriented architectures and agile development) coupled with increasing the potential of academic research for innovation created the completely new picture. It provides opportunities for both horizontal and vertical integration of components and applications which were not feasible before. GATE Institute at Sofia University was established to meet these challenges both locally, within Bulgaria and regionally, in the Balkans. It operates also on a wider European scale so it can support the recent move towards building common European digital spaces. A central role in its strategy plays the GATE Data Platform, which provides an environment for the development and deployment of data services across several application areas. Amongst them, Future Cities plays an important organising role. This article presents the current state of development of GATE Data Platform in the context of the Future Cities application area. It has now entered its second phase, which aims at integrating the resources involved and the results achieved up to the moment as a preparation for the establishment of the first urban data space in the country. It aims at serving local authorities, businesses and citizens by providing easy-to-use, secure and reliable urban services for data sharing. A number of new concepts and recent technologies have been utilized in the research in GATE – city digital twin, hybrid AI combining sensor data with meta-data and ontological information, parametrization of ML engines for automatic configuration and easy orchestration on the cloud, etc. We will briefly review them in the next section. After that, we will introduce the GATE Data Platform as a technical backbone of the research in the Future Cities application area. We will illustrate it with the pilot for monitoring and analysis of the air pollution in Sofia. Further, we will review the current research at GATE Institute on different aspects of urban data processing, which will feed into the development of more urban data services, as well as will provide technical support for the future Urban Data Space.

II. SMART CITIES AND BEYOND

A. *Smart Cities vs. Future Cities*

The concept of "smartness" is a marketing term which has been brought to life largely by the industry to promote its products (see [1] for a typical example). The academic response to this move is the creation of a reference model which pretends to give a more abstract and universal formulation of the potential solution, largely considered to be a standard [2]. However, since it focuses on the urgent needs of the urban establishment to cope with the complexity of urban life, particularly in large metropolitan areas, it needs substantial extensions as the technologies to develop and frequent updates as the practices evolve [3]. Our approach is purely pragmatic here – we are trying to provide services to the city as it is and this way, to solve some of its real needs at this moment in time.

B. *Urban Data Services*

The classical service-oriented architecture (SOA) is still an adequate metaphor for dealing with both software and data services. The recent fashion to shift the attention from operations to data (Computer Science *vs.* Data Science) does not change the need for orchestrating the services, which can be done the same way for both software and data services. The main difference is in the flows of data between the different services since this can be done in several different ways: in one piece, as a continuous stream of potentially infinite data, through periodic messaging, and in batch instalments. The powerful tools for dealing with this variety offered by the commercial cloud platforms such as Amazon AWS, Google GCP and Microsoft Azure make an attractive offer to both academia and business the use of the cloud. At the same time, the public clouds contain a hidden trap for both the service providers and the service consumers due to the financial burden on both sides, the lack of control over the data for the providers and the steep learning curve for the consumers. Because of this GATE Institute has adopted the idea to build its own data platform on a private cloud instead as presented in the position paper [4]. The current paper presents the actual implementation of this concept using a public domain software stack and the pilot implementation of the main services needed for building a data service provider for the future Urban Data Space.

C. *Urban Data Space*

The concept of a data space came out to free both the data service providers and the data service consumers from their dependence on the "middleman" in service provisioning. This is vital in many supply chains of B2B operations and was initiated in Germany first to serve the needs of the European industry [5]. The data spaces utilize distributed data processing model implemented using open architecture, which consists of adapters, brokers, bridges and other software components, which expose APIs allowing them to communicate with each other, and share data and data services in the public space of the Internet. The participants in the data space supply their own data for public consumption, while at the same time they can leverage the data of other suppliers. This is particularly attractive in an urban setting where often the businesses require a combination of data from multiple data sources – transport, logistic and communal services, environment and healthcare institutions, business, commercial and industrial organisations [6]. GATE Institute endorsed this idea and embedded it into the development of its own Data Platform by incorporating several components which provide support for it. The pilot application in the Future Cities area implements a number of data services for the future Urban Data Space.

III. GATE DATA PLATFORM AND AIR POLLUTION PILOT

The technical backbone of the future Urban Data Space is the GATE Data Platform, currently under development at GATE Institute [4]. In this section, we will give a brief introduction to the platform and will illustrate its use for monitoring the air pollution in the city (both outdoor and indoor) together with some analytics services for processing urban data currently available on the platform.

A. *Data Platform on Private Cloud for Small and Big*

GATE Data Platform utilizes the concept of a private cloud where multiple data sources feed in data and the platform data services can be exposed for external use through the mechanism of virtualization. The technological stack used to build the platform is shown in **Fig. 1**. Core of the system architecture of the platform is the containerized engines and applications which run under the control of two additional management layers on top of the operating system: the container management system, based on Kubernetes, and the virtualization system, based on Docker containers.

A number of different data repositories are available on the platform, allowing to accommodate both fast-growing data, typically transported in a continuous flow, large volumes of static data, uploaded in separate files, as well as homogeneous collections of data records transported periodically on demand or in an event-driven manner. The platform allows collecting data from multiple different data sources (sensor measurements, messages, event logs, files and repositories) using a variety of communication protocols (text-based as well as binary, synchronous as well as synchronous). Based on the number of data repositories installed on the infrastructure the platform supports several different data formats: SQL, JSON, RDF, CityGML as well as files in plain text, hypertext and binary format.

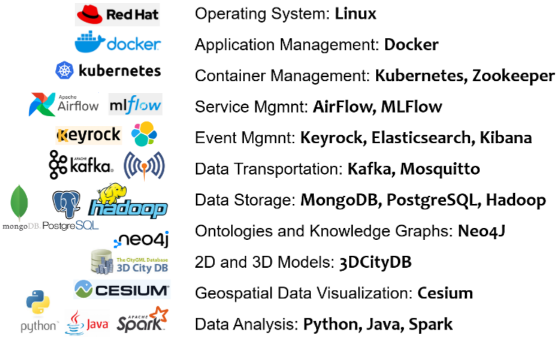


Fig. 1. Technological Stack of GATE Data Platform

The platform has been built using entirely free software from the public domain or free versions of commercial products which are not only attractive financially for in-house development but also important because it supports an open architecture with a high level of compatibility and extendibility. Over the last two years GATE Data Platform has been used in two pilot applications, developed in partnership with London Metropolitan University: for real-time security analytics of network packets using algorithms for correlation, recognition and trend analysis [7], and for assessing the air quality in Sofia, reported for the first time here. **Fig. 2** shows the software components installed on the platform in support of the application for controlling the air pollution in Sofia, which are the core of the support for the future Urban Data Space.

An important enhancement of the platform which is coming in the second half of the year is the replacement of the existing Apache version of Hadoop with the powerful Cloudera CDH software with its petabyte memory space, which will substantially increase the capacity for storing and manipulating truly Big Data on the platform.

B. *Monitoring of Air Pollution in Sofia*

Due to its rapid development after years of stagnation, Sofia is facing serious problems related to the quality of the air in the metropolitan area. Increased traffic, chaotic development and weakened control mechanisms caused a significant increase in air pollution from both organic and non-organic origins. Sofia was even penalized by the EU authorities as the European capital with the worse quality of the air. This, together with many areas which contribute to, or are affected by the environmental situation, like the increase of respiratory diseases, the transport gridlocks and the enormous losses reported along the urban networks, caught the attention of many authorities, including GATE Institute for which Future Cities and Digital Health are amongst its priority application areas.

The research reported here has been supported by GATE project, funded by the EU H2020 WIDESPREAD-2018-2020 TEAMING Phase 2 programme under Grant Agreement No. 857155, by BG Operational Programme Science and Education for Smart Growth under Grant Agreement No. BG05M2OP001-1.003-0002-C01 and by the Scientific fund of Sofia University under agreement no. 80-10-11/10.05.2022.

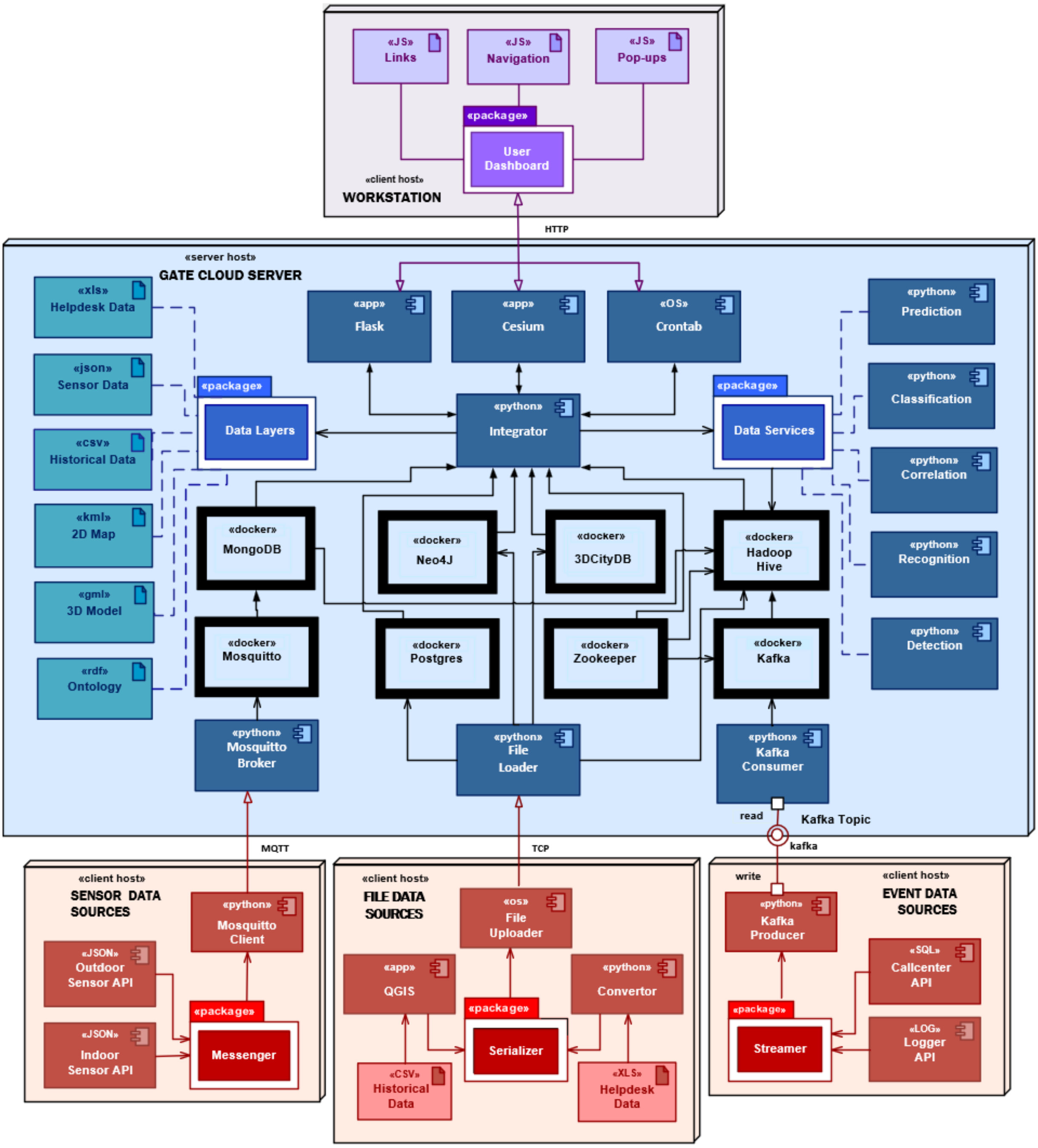
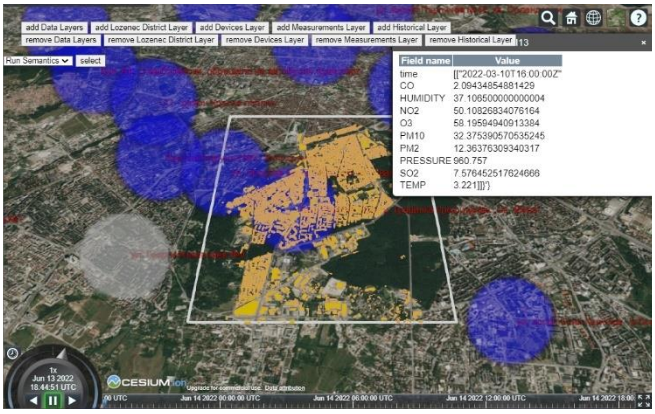


Fig. 2. Software Components for Air Pollution Assessment on GATE Data Platform

GATE currently provides a comprehensive picture of the air quality in Sofia straight on its private cloud, fuelled by GATE Data Platform (**Fig. 3**). The map shows the Sofia urban area covered by the sensor stations of AirThings platform of Sofia Municipality, measuring CO, NO2, O3, PM10, PM2, and SO2 as well as the temperature, humidity and pressure. The data are extracted via an API which updates the sensor measurements on an hourly basis. The data is then transported to the GATE Data Platform where it undertakes some processing before being shown on the web. **Fig. 3** shows the most important pollutant coloured darker on a map. The application can also show the historical data, gathered over a period and available from the same stations.



Field name	Value
time	[["2022-03-10T16:00:00Z"
CO	2.09434854881429
HUMIDITY	37.106500000000004
NO2	50.10826834076164
O3	58.19594940913384
PM10	32.375390570535245
PM2	12.36376309340317
PRESSURE	960.757
SO2	7.576452517624666
TEMP	3.221]]}'}

Fig. 3. Outdoor Air Pollution on GATE Data Platform

Several components of GATE Data Platform are operating along the pipeline for processing environmental data from its source to the destination on the web:

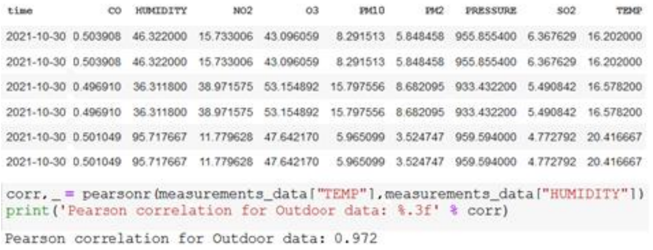
- **Data Extractor** captures the sensor measurements using the station API, performs pre-processing of the rough data and formats it in JSON format suitable for further processing on the cloud.
- **Data Transporter** uses Mosquitto API to transport the measurements to a cloud server over MQTT protocol.
- **Data Broker** server-side counterpart of Mosquitto MQTT server, responsible for collecting the JSON objects encapsulating the sensor measurements from the stations.
- **Data Collector** buffers the sensor data to allow real-time processing of the incoming data and maintains a queue of buffered data for storing in permanent storage.
- **Data Analyzer** performs real-time data analysis of the buffered data (in the pilot it correlates the humidity and the temperature).
- **Data Accumulator** periodically flushes the data buffers into the persistent MongoDB database in JSON for further analysis.
- **Data Integrator** combines the data retrieved from various repositories, graphic and ontological models, historical data and sensor measurements) and prepares several layers of data for visualization in KML format.
- **Data Visualizer** is based on Cesium and it uses the KML data layers and the 3D model of the area from OpenStreet to produce the live map of the air quality which goes viral on the web.

All the above components are written in Python. They work with the respective software products from the technological stack, which are installed on the server within separate Docker containers. The visualization is within a standard web browser.

C. *Analysis of Environment Data*

The environment data provides valuable information for decision-makers, but in order to maximize the added value of having it gathered in one place, it is even more interesting to perform analysis using more powerful methods. Apart from purely statistical calculations, the easiest analytics which can be performed in real-time using efficient recurrent algorithms is the correlation analysis. As a proof of concept, we have implemented a simple correlation analysis of the outdoor data using the standard Pearson method. Table I shows buffered data and a result of correlating the temperature and humidity over a limited period of time.

TABLE I. CORRELATION ANALYSIS OF OUTDOOR DATA



time	CO	HUMIDITY	NO2	O3	PM10	PM2	PRESSURE	SO2	TEMP
2021-10-30	0.503908	46.322000	15.733006	43.096059	8.291513	5.848458	955.855400	6.367629	16.202000
2021-10-30	0.503908	46.322000	15.733006	43.096059	8.291513	5.848458	955.855400	6.367629	16.202000
2021-10-30	0.496910	36.311800	38.971575	53.154892	15.797556	8.682095	933.432200	5.490842	16.578200
2021-10-30	0.496910	36.311800	38.971575	53.154892	15.797556	8.682095	933.432200	5.490842	16.578200
2021-10-30	0.501049	95.717667	11.779628	47.642170	5.965099	3.524747	959.594000	4.772792	20.416667
2021-10-30	0.501049	95.717667	11.779628	47.642170	5.965099	3.524747	959.594000	4.772792	20.416667

```
corr,_ = pearsonr(measurements_data["TEMP"],measurements_data["HUMIDITY"])
print('Pearson correlation for Outdoor data: %.3f' % corr)
```

```
Pearson correlation for Outdoor data: 0.972
```

There are many interesting cases of potential correlation between the sensor measurements, which can give valuable insight into the impact of the different parameters on the state of the environment. They can relate different factors in the same data stream (like in the above case), compare the dynamics of the same factors in different streams (i.e., correlating the level of particles in indoor and outdoor measurements) and look for more complex correlations by relating historical data to current measurements.

The above pipelines operate in real-time. This makes the choice of suitable method for data analysis heavily dependent on the frequency of updating the measurements and the number of stations generating the sensor data. As a result, it allows only a relatively simple analysis. In order to conduct more thorough analysis over larger datasets, like analysis of the current trends or prediction of the future dynamics, and to combine the environment data with other sources like healthcare, for example, the data needs to be organised accordingly. For this purpose, on GATE Data Platform we configured two Hadoop data nodes for the distributed repository of heterogeneous data. A separate **Data Bridge** component on the platform (**Fig. 2**) is transferring the data from a MongoDB database to the data nodes of Hadoop, where we run SPARK for performing more deep analytics using Machine Learning techniques. This way we can integrate the data from different sources, in different formats and for different purposes. The Hadoop nodes will be replaced soon by a large multi-node cluster running Cloudera Hadoop.

IV. ENVIRONMENTAL RESEARCH AT GATE INSTITUTE

A. *Capturing Indoor Data*

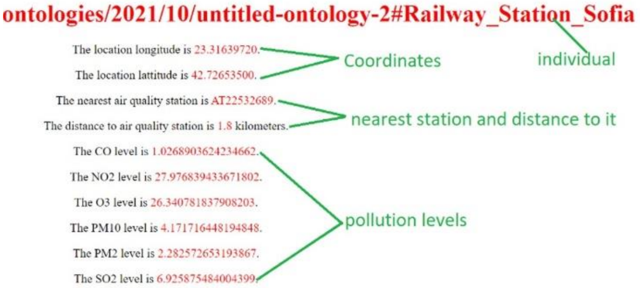
After successful completion of the outdoor air quality assessment pilot on the GATE Data Platform, the team embarked on a new journey - assessing the air quality inside designated buildings. For this purpose, the team engineered its own sensor station equipped with CO2, TVOC, PM1.0, PM2.5 and PM10 sensors in addition to the standard sensors for temperature, humidity, pressure and altitude. The station uses a cheap ESP32 controller combined with an Arduino toolkit, which provided the necessary flexibility to conduct experiments with the indoor air quality (**Fig.4**). Since there is not much difference in terms of the outdoor and indoor data, there was no need to change the existing pipeline to process the indoor data. The only difference is the location of the stations since in multi-story buildings 2D reconstruction is not sufficient to point at their location but as a first approximation, we used a point-based location on the 2D map, similar to the outdoor visualization. This allowed us to correlate outdoor and indoor data and thus, analyze the impact of outdoor air pollution on indoor air quality.

Fig. 4 Capturing Indoor Data in the building

Currently, we are working on adding sensors for organic compounds, which would allow a more interesting analysis linked to the presence of organic processes within the closed spaces inside the buildings. This would be very valuable in regard to the health and safety regulations for working in offices, the living conditions in old people's homes, the public spaces as well as private accommodations in general.

B. *Integration of Ontological and Geospatial Data*

CityGML provides a standard format for modelling the urban infrastructure and it has been used within the GATE Data Platform to represent the 3D reconstruction of the urban area in focus. Its version 3.0 provides extended support for integrating ontological information in RDF format with the geospatial representation of the urban infrastructure in CityGML [10]. Based on the standard ontologies we are currently working on the integration of the semantic information with the visual representation based on Cesium Ion. **Fig. 5** shows a fragment of the representation that mixes information coming from the sensor measurements with information coming from the ontological model of the area within a single stream of data passed to the visualization component. Cesium software which has been used for this purpose allows combining the graphical rendition of geospatial information with pop-up windows to annotate the navigation points.



ontologies/2021/10/untitled-ontology-2#Railway_Station_Sofia

The location longitude is 23.31639720.

The location lattitude is 42.72653500.

The nearest air quality station is AT22532689.

The distance to air quality station is 1.8 kilometers.

The CO level is 1.0268903624234662.

The NO2 level is 27.976839433671802.

The O3 level is 26.340781837908203.

The PM10 level is 4.171716448194848.

The PM2 level is 2.282572653193867.

The SO2 level is 6.925875484004399.

Fig. 5. Integrating Ontological Information with the Sensor Data

Such integration would enable more complex analytics which utilize semantic information from ontological models of the city infrastructure for making a logical inference. For example, in the case of outdoor pollution, we perform analysis by using the information about distances between different types of buildings. This allows us to identify the main polluter in the area, as well as to plan the location of new buildings in areas with a lower level of pollution. We are also investigating the opportunity to use the ontological model of the buildings for supporting a VR-style of navigation to show the precise location of the indoor sensors and a heat map of the air pollution inside closed spaces.

C. *Analysis of Environment Data*

The sensor data is small but fast-growing. This allows applying both real-time methods, which use efficient statistical algorithms, as well as more elaborate analysis of historical data, which employs statistical methods and machine learning. In the two pilot projects we completed on the GATE Data Platform we implemented several statistical and ML models for solving different analytics tasks: pattern recognition, classification and prediction, including classical regression, support vector machines (SVM), XGBoost regression. **Fig. 6** presents the prediction of the CO levels based on the measurements of the sensor stations located in the south-west area of Sofia (Kopitoto, Pavlovo and Hipodruma) over a period of one month. The prediction uses Auto Regressive Integrated Moving Average method (ARIMA) which gives sufficient precision. For discrete event-based sequences, much better results were obtained by using XGBoost regression.

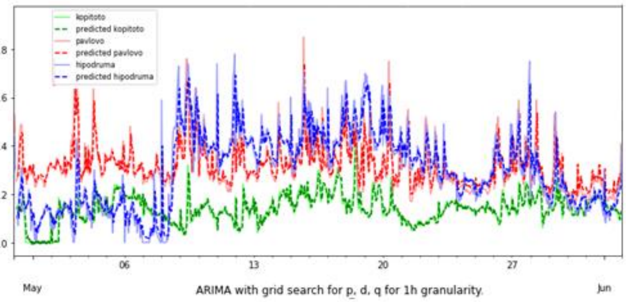


Fig. 6. Prediction of CO levels in three outdoor stations in Sofia

D. *Parametric Urban Planning*

GATE Institute collaborates with Sofiaplan, a municipal enterprise responsible for the spatial and strategic planning of Sofia Municipality, on a use case of parametric urban planning. An automated tool providing a data-driven decision support on urban planning based on preliminary defined neighbourhood indicators is developed. The indicators are structured into 10 groups. The current solution considers 3 of them: Demography including population number, density, and age groups; the built environment, including all types of buildings – residential, office, kindergartens, schools, etc.; and the potentiality, which is related to the territory of the city unit and includes indicators such as single lots, totals empty lots and occupied lots with buildings. The proposed solution is based on a grid, which is constructed on top of the whole free area, where the cells can obtain different land-use functions (**Fig. 7**). It is automated using a non-dominated sorting genetic algorithm (NSGA-II), which produces. Pareto-optimal set with multiple urban designs.

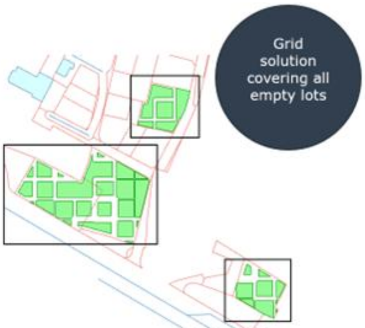


Fig. 7. Parametric Urban Planning at Sofia Municipality

Since parametric planning is a multi-objective planning problem it is computationally demanding. In addition, many of the indicators require information from the urban data model. Because of this GATE Data Platform provides an ideal environment for the automation of this task.

E. *CFD Simulations*

As a follow-up of the generated urban designs, the team constructed Computational Fluid Dynamics (CFD) models capable of simulating the pedestrian wind and thermal comfort and the dispersion of air pollutants. A core of the CFD simulation is the mesh model of the study area. The input data for its generation includes a digital terrain model (DTM) and building footprints. The building footprints are cleaned of short edges, sharp angles, small gaps, intersections, and duplicates. Where possible, contacting objects are combined into a single solid to reduce the number of model elements. **Fig. 8** illustrates the pipeline of the steps for mesh generation.

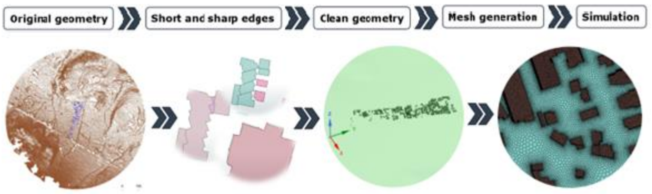


Fig. 8. Mesh Generation from 2D maps

Unfortunately, the creation of an automated workflow applicable for different urban designs and suitable for different analyses is possible only to a certain extent. Critical engineering judgement is essential at every step to assess the validity of the results produced by automated procedures and apply corrections where necessary.

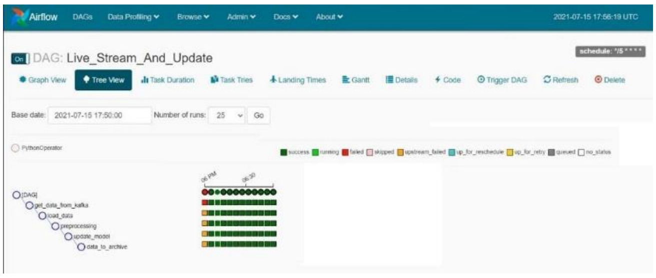
V. ON THE ROAD TO SOFIA URBAN DATA SPACE

Building Urban Data Space is an important target for GATE Institute – both because of its significance for the city of Sofia and because of its value for GATE itself. Because of this GATE Institute initiated and will continue actively to participate in the development of the space both as a provider of data services as well as a consumer of data for enrichment with new features. As a preparation for this, we are already working on incorporating adequate support for the requirements of the future Urban Data Space directly into GATE Data Platform. In this section, we will briefly review some of the most important components which we use for this purpose.

A. Parametrization, Worlglows and Orchestration

Although the data space may leverage from utilizing even a single data service, in order to maximize the potential effect of organizing several services in a data processing pipeline we need a mechanism for linking the data flows and for controlling the operations during their execution. In a service-oriented architecture, we can use tools which implement workflow control, similar to Business Process Execution Language (BPEL) processes. However, the business processes here are executed within containers which requires the control mechanism to operate in close interplay with the container management system.

For this purpose, we use Apache *Airfow* [11], which is fully integrated with Kubernetes. It uses external configuration files to specify the parameters of the workflows in the form of acyclic graphs (DAG files) and controls the data processing operations executed within Docker containers. **Fig. 9** shows a pipeline of data processing operations executed under the control of *Airflow*.



DAG: Live_Stream_And_Update

Base date: 2021-07-15 17:50:00

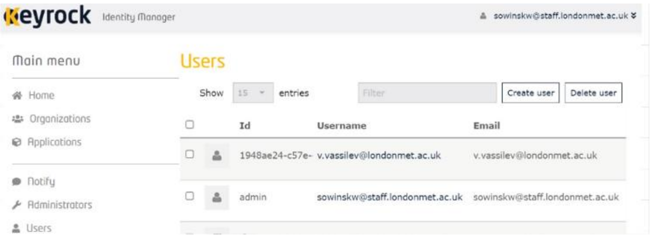
Number of runs: 25

Fig. 9. Orchestration of the Containers using Airflow

Airflow has its intricacies and constraints, which limit its use for specifying more complex workflows and complicate the possibility to automate the generation of DAG files. Still, its use is straightforward and relatively easy to work with, which makes it a good starting point for extending the service provision from single services to sequential pipelines.

B. Identity Management and Access Control

Identity Management is an important requirement for any platform which supports data spaces not only because of the security with respect to data services exposed by it, but also because it is a starting point for developing and managing the *trust* between the parties. Furthermore, since the identity of the principal initiating the operations is part of the session management, this component often provides information necessary for further analysis of the operations, such as tracking/debugging and measuring/billing. **Fig. 10** shows how the FIWARE *Keyrock* component [12] tracks the work of the principal by enhancing the low-level mechanisms for identity management by the server infrastructure. A disadvantage of using *Keyrock* is the relative complexity since it plays multiple roles within the enterprise architectures.



☐	Id	Username	Email
☐	1948ae24-c57e-	v.vassilev@londonmet.ac.uk	v.vassilev@londonmet.ac.uk
☐	admin	sowinskw@staff.londonmet.ac.uk	sowinskw@staff.londonmet.ac.uk

Fig. 10. Identity Management using Keyrock

C. Log Analysis and Event Management

During platform operation, several log files generated at runtime contain information pertinent to the data processing: Operating System, Container Management System, Containers, Software Components, and Services.

This information can be used for different purposes during the development, deployment and execution of the data services. Instead of retrieving and processing the information from the different logs individually, on the GATE Data Platform, we use *Elasticsearch* [13], which builds a single point of management of all log files. Thus, it allows centralizing the entire event processing in a single component. The data collected is then made available in JSON format to any component which needs it on the data space level. For system purposes, we currently also use the sister product *Kibana* [14], which allows inspecting this information directly within a web browser both locally and over the Internet (**Fig. 11**).

137 hits

May 13, 2022 @ 08:14:06.885 - Jun 12, 2022 @ 08:14:06.885 Auto

```
Time      _source
> Jun 12, 2022 @ 08:11:27.762  @version: 1 path: /home/keyrock_logs.txt host: message: GET /console/ 200 12.746 ms - 13749 08:11:27.762 _id: iZgKWIEB_3MwpFSGbz0D _type: _doc
                               _index: keyrock_logs _score: -
> Jun 12, 2022 @ 08:11:27.762  @version: 1 path: /home/keyrock_logs.txt host: message: GET /console/ 200 12.746 ms - 13748 08:11:27.762 _id: jZgKWIEB_3MwpFSGbz0G _type: _doc
                               _index: keyrock_logs _score: -
> Jun 12, 2022 @ 08:11:27.761  @version: 1 path: /home/keyrock_logs.txt host: message: GET /console/ 200 12.746 ms - 13747 08:11:27.761 _id: kZgKWIEB_3MwpFSGbz0G _type: _doc
                               _index: keyrock_logs _score: -
> Jun 12, 2022 @ 08:11:27.761  @version: 1 path: /home/keyrock_logs.txt host: message: GET /console/ 200 12.746 ms - 13747 08:11:27.761 _id: iJgKWIEB_3MwpFSGbz0D _type: _doc
                               _index: keyrock_logs _score: -
> Jun 12, 2022 @ 08:11:27.761  @version: 1 path: /home/keyrock_logs.txt host: message: GET /console/ 200 12.746 ms - 13747 08:11:27.761 _id: jJgKWIEB_3MwpFSGbz0G _type: _doc
                               _index: keyrock_logs _score: -
> Jun 12, 2022 @ 08:11:27.761  @version: 1 path: /home/keyrock_logs.txt host: message: GET /console/ 200 12.746 ms - 13747 08:11:27.761 _id: lZgKWIEB_3MwpFSGbz0I _type: _doc
                               _index: keyrock_logs _score: -
> Jun 12, 2022 @ 08:11:27.760  @version: 1 path: /home/keyrock_logs.txt host: message: GET /console/ 200 12.746 ms - 13747 08:11:27.760 _id: kJgKWIEB_3MwpFSGbz0G _type: _doc
                               _index: keyrock_logs _score: -
```

Fig. 11. Log Analysis using Elastic Search and Kibana

All above components are added to the GATE Data Platform and work with the selected software products, which have recommended themselves during years of maturing within the open source community. Thanks to their containerization such components can be easily embedded in the platform to support most of the necessary tasks typical for the functions of the data space – *configuring*, *debugging*, *securing*, *monitoring*, *auditing*, *measuring* and *billing*.

VI. CONCLUSION

In this article, we have shown how the GATE Data Platform can support the development and implementation of an urban data space. With its openness, extendibility and scalability GATE Data Platform is an ideal vehicle for providing inter-related data services using shared data. This synergy would enable the creation of the next generation of data infrastructure as foreseen by the European Gaia-X initiative: "an open, transparent and secure digital ecosystem, where data and services can be made available, collated and shared in an environment of trust" [15]. Built entirely using public domain software, GATE Data Platform is currently in a process of migration to a more powerful cluster infrastructure capable of manipulating petabytes of memory, which will run on similar industrial strength software. This new version will create an even better opportunity for serving the needs of the city across the entire urban life complex – *urban planning*, *transport management*, *social services*, *public healthcare*, etc.

ACKNOWLEDGEMENT

The GATE Urban Data Space project combines the efforts of two teams of GATE Institute: the researchers who have been working in the GATE Future Cities application area for more than two years, and the engineers from the Data Engineering division currently working on the GATE Data Platform. The research reported here has been also influenced by the collaboration with two partners of GATE Institute - Chalmers University in Gothenburg, Sweden and London Metropolitan University in London, UK. Although the work on this project is a collective effort involving more people the particular formulations are of the authors only and reflect their own experience and vision.

REFERENCES

[1] Siemens, "Products & Services" [Online: https://new.siemens.com/global/en/products.html]

[2] S. Zygiaris, "Smart City Reference Model", Journal of the Knowledge Economy, 2013, vol. 4(2), pp. 217-231.

[3] N. Tcholtchev, P. Lämmel, R. Scholz et al., "Enabling the structuring, enhancement and creation of urban ICT through the extension of a standardized smart city reference model", in: IEEE/ACM Int. Conf. Utility and Cloud Computing Companion (UCC Companion), IEEE, 2018.

[4] V. Vassilev, S. Ilieva, D. Petrova-Antonova, et al., "AI-based Hybrid Data Platforms", in: E. Curry, S. Scerri, and T. Tuikka, Tuomo, Eds., Data Spaces: Design, Deployment and Future Directions. Springer Nature (2022). ISBN: 3030986357

[5] I. Schieferdecker, L. Bruns, S. Cuno et al., "Urbane Datenräume - Möglichkeiten von Datenaustausch und Zusammenarbeit im urbanen Raum", Fraunhofer FOKUS, Berlin, 2019 [Online: https://publica.fraunhofer.de/entities/publication/6e843743-0e75-4904-aef4-01f50a0a91b0/details]

[6] International Data Spaces Association, "Reference Architecture Model Version 3.0", 2019. [Online: https://internationaldataspaces.org/publications/most-important-documents]

[7] V. Sowinsky-Mydlarz, J. Li, K. Ouazzane, and V. Vassilev, "Threat Intelligence Using Machine Learning Packet Dissection", in: Proc. 20th Int. Conference on Security & Management, 26-29 July 2021, Las Vegas, Springer (in print).

[8] E. Marinov, D. Petrova-Antonova, and S. Malinov, "Time Series Forecasting of Air Quality", Atmosphere, 2022, 13 (5), p. 788.

[9] E. Shirinyan and D. Petrova-Antonova, "Modeling Buildings in CityGML LOD1", ISPRS Int. J. Geo-Inf. 2022, 11(3), p.166.

[10] R. Billen, C. Zaki, M. Servières, et al., "Developing an ontology of space: Application to 3D city modeling". In: T. Leduc, G. Moreau, R. Billen (eds.), Usage, usability, and utility of 3D city models, Nantes, EDP Sciences, 2012, Vol. 02007 [Online: https://hal.archives-ouvertes.fr/hal-01521445]

[11] Apache Software Foundation, "Apache Airflow Documentation" [Online: http://airflow.apache.org/]

[12] FIWARE Foundation E.V., "Identity Manager – Keyrock" [Online: https://fiware-idm.readthedocs.io/en/latest/]

[13] D. Courcy, "Build seamless, search-powered customer experiences with Elastic 8.2" [Online: https://www.elastic.co/blog/whats-new-elastic-8-2-0]

[14] Elasticsearch B.V., "Kibana: Your window into the Elastic Stack" [Online: https://www.elastic.co/kibana/]

[15] F. Biegel, A. Biongers, R. Chidambaram, et. al, "GAIA-X: Driver of digital innovation in Europe", Federal Ministry for Economic Affairs and Energy, Berlin, 2021 [Online: https://www.bmwk.de/Redaktion/EN/Publikationen/gaia-x-driver-of-digital-innovation-in-europe.pdf?__blob=publication File&v=8]