# VILNIUS CITY DATA POLICY

| Compiled by: Stasys Savilionis, the Intelligent Solutions and Data Management Department of the Vilnius Municipal Enterprise Vilniaus Planas _____ 26/11/2021 | Verified by: _____ ……../2021 | Approved by: _____ ……../2021 | Valid since: _____ ……../2021 |
|---|---|---|---|

# Document version

Date and version of the document, list of amendments thereto.

| Document date | Version | Description of amendments |
|---|---|---|
| 26/11/2021 | V1.0 | First edition. |
| | | |
| | | |

CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# 1. GENERAL PROVISIONS

*The Vilnius City Data Policy* has been compiled in accordance with the *European Strategy for Data[1]*, in view of the Vilnius City Strategic Plan 2020-2030 and other normative and legal documents regulating the activities of the City of Vilnius and its subordinate enterprises. This policy supplements existing legislation and contributes to the priorities of the City of Vilnius in the development and implementation of the city's strategic vision, the provision of high-quality services to residents, attracting investment to the city, acting in a transparent, responsive and open manner to residents, visitors and investors.

*The Vilnius City Data Policy* has been compiled based on the needs and insights of the City of Vilnius, its subordinate enterprises and related institutions, as well as taking into account the global best practices in the development of data platforms, and the requirements of international standards of data processing, handling and security applicable to them.

*The Vilnius City Data Policy* sets out the objectives and tasks, the main fields of the policy implementation, the content of the policy requirements, the benefits, the concerned parties, their roles and responsibilities, and the process and sequence for monitoring its implementation.

*The Vilnius City Data Policy* applies to the City of Vilnius, its subordinate enterprises and institutions, it is recommended for natural persons and legal entities (including research and scientific institutions, innovation developers, developers of advanced technologies and investors) planning to carry out or carrying out their activities in Vilnius, as well as for the residents of Vilnius, in order to make the Data Policy useful for all, the aim is to involve a broader range of concerned parties in the development of the Data Policy, including the general public.

*The Vilnius City Data Policy* has been compiled by the staff of the City of Vilnius and its corporate group, taking into account the content of the data collected, with a focus on the sustainable development and integrity of a smart city and a smart society. The responsibility for the proper implementation of the policy and the cases in which it applies to natural persons and legal entities operating in Vilnius and residents of the city lies with the City of Vilnius, its subordinate enterprises and institutions.

# 2. BENEFITS FOR THE CITY

Benefits of the *Vilnius City Data Policy* for the City of Vilnius and its corporate group:



**DIGITALIZATION FOR VALUE CREATION**

*Intelligent digital solutions*

High-quality and value-creating data generated by the digitization of the city and advanced solutions are managed. Priority is given to data enrichment, the benefits to the city as a whole are measured, so each project generates data as a result.

**PARTNER ECOSYSTEM**

*Method of implementation – adapt and make use*

---

[1] European Strategy for Data:

https://eur-lex.europa.eu/legal-content/LT/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066

A sufficiently flexible model is being developed to allow for the adaptation and make use of the platforms and data the city already has. Ability to use data creating value from systems available to municipal enterprises, state-owned enterprises, other IT companies and to use the most effective solution to develop a solution for a growing partner ecosystem. Strategic planning and feasibility studies to improve infrastructure and accessibility of services for the city's residents, project development and other needs of the city's residents, business entities, research and academic institutions, are carried out by the City of Vilnius and its subordinate enterprises.

## ANALYSE DIFFERENT DATA SOURCES
*Data analytics*
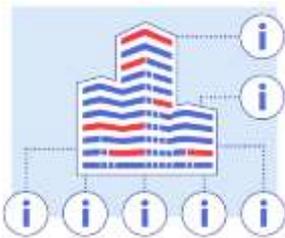
Data from the City of Vilnius and its corporate group should be analysed using automated solutions, i.e. self-contained, appropriately configured software that can quickly operate and analyse large volumes of data from different sources and formats. Without losing sight of its usefulness for the role of analysts, who can see and manage the information they need.

## OPEN API
*Easy integration with other applications*

Based on widely used standards, APIs are developed easily to integrate with other existing applications or new applications under development, and to retrieve data using standard protocols.

Innovative solutions are being developed to facilitate the use of platforms and applications at a lower cost on the city scale.

## ENSURING SECURITY
*Data security*

The City of Vilnius and its corporate group should be a trusted service provider that guarantees the security of the processed data through independent entities responsible for carrying out a "penetration test", which consists of searching for vulnerabilities in the systems and providing a technical solution to any security problems detected, thus ensuring its robustness.

## BETTER EVERYDAY DECISIONS
*Manage and monitor city services*

The resulting data is aggregated, analysed and displayed in a dashboard. It allows to produce reports in different sections, to define critical value thresholds and to display KPI's in an easy to understand way, which helps managers responsible for the provision of the respective services to make daily decisions in a safe and balanced way.

**EASY TO UNDERSTAND**

*City governance*

An indicator expression, which is the sum of all the city's services and performance indicators.

Vilnius should become a model where business and public sector entities can use comprehensive real-time and historical data to make more informed strategic decisions that affect the quality of life of residents, the performance of businesses and the quality and accessibility of public services.

To achieve this, the City of Vilnius will be able to rely on a strong legal framework for data protection and cybersecurity.



*Figure 1 – Benefits of Data*

# 3.   CONCEPTS, DEFINITIONS AND ABBREVIATIONS

Key concepts and abbreviations used in the Vilnius City Data Policy.

**Figure No 1 – Concepts, Definitions and Abbreviations**

| No | Concept | Definition |
|---|---|---|
| 1. | **"2D"** | means a two-dimensional space (plane) used to create graphical primitive objects such as curves, shapes etc. |

| 2. | **"3D"** | means a three-dimensional space used to create sets of points connected by lines, curves, planes, etc., which are used to create volumetric bodies. 3D objects can be displayed as: objects created in real space (displayed live); three-dimensional optical (holographic) spatial images; two-dimensional images simulated by computers of 3D object. In the BIM methodology, 3D is a building modelling, i.e. an object-based information model of a building which information sample is limited to the geometry of the elements of the building defined at three levels of detail. |
|---|---|---|
| 3. | **"LLPPD"** | means the Republic of Lithuania Law on Legal Protection of Personal Data. |
| 4. | **"API"** | means Application Programming Interface that is an open technical interface, compliant with the requirements of the EBA/RTS, for communication between different systems and software. |
| 5. | **"Personal Data"** | means any information relating to an identified or identifiable natural person ("data subject"); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. |
| 6. | **"BIM"** | means the process of developing and managing a building information model throughout its lifetime. It is usually carried out using object-oriented modelling software to increase the efficiency of building design and construction. During the process, the Building Information Model (BIM) is obtained with all information on building geometry, display of spatial relations and nodes, quantities and properties of building elements. |
| 7. | **"DU"** | means salary. |
| 8. | **"Data"** | means general information and personalised data managed by the City of Vilnius or its enterprises, which will be consolidated and used for the purposes set out in the Policy. |
| 9. | **"Data and Information Security"** | means ensuring the confidentiality, integrity and availability of data and information. |
| 10. | **"Data and Information Management System"** | means the part of a management system that aims to identify, apply, manage, monitor, review, maintain and improve the security and management of data and information, based on the results of a risk analysis. |
| 11. | **"Description of Data Structure"** | means a document that grants access, reuse and distribution rights to users of open data with a few or no restrictions. |
| 12. | **"Data Subject"** | means the natural person to whom the personal data belongs. |

| 13. | **"Processor"** | means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller. The processor shall act in accordance with the controller's instructions. |
|---|---|---|
| 14. | **"Processing"** | means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction. |
| 15. | **"Controller"** | means the natural or legal person, public authority, agency or other body which determines the purposes and means of the processing of personal data; |
| 16. | **"EBA/RTS"** | means the European Commission Delegated Regulation for strong customer authentication and common and secure open standards of communication. |
| 17. | **"EBITDA"** | means the company's profit before interest, tax, depreciation and amortisation. |
| 18. | **"ETL/ELT"** | means a three-step process – extract, transform, load – in which selected data is transferred from the original source to a new repository by modifying, merging or otherwise processing it according to specified rules. This abbreviation also refers to the software tools used to carry out this process – e.g. ETL software. |
| 19. | **"IaaS"** | means an Infrastructure as a service, where the data centre provides hardware infrastructure capacity. |
| 20. | **"Enterprise"** | means an enterprise, institution or organisation subordinate to the City of Vilnius. |
| 21. | **"ISO"** | means the International Organisation for Standardisation. |
| 22. | **"ISP"** | means the Information Security Management Policy. |
| 23. | **"IT"** | means information technologies. |
| 24. | **"ITG"** | means a description of open data structure describing data that has already been opened and made publicly available. |
| 25. | **"ITIL"** | means the Information Technology Infrastructure Library, i.e. business management theory focusing on work optimisation and quality ensuring in IT companies. |
| 26. | **"ISDC"** | means data maturity levels set out in the Open Data Guide of the ISDC (https://atviriduomenys.readthedocs.io/) |
| 27. | **"Head of the ISDC"** | means the Information Society Development Committee – a budgetary body under the Ministry of Transport and Communications of the Republic of Lithuania, responsible for the state policy on the development of information and communication technologies in Lithuania and coordinates its implementation. |
| 28. | **"PaaS"** | means a Platform as a service, where a software platform for developing software |

| | | and not hardware capabilities is leased. |
|---|---|---|
| 29. | **"Policy"** | means guidelines, objectives and directions for the overall management of information assets formally established by management. |
| 30. | **"VAT"** | means a value added tax. |
| 31. | **"SA"** | means a system administrator. |
| 32. | **"SaaS"** | means a Software as a service, where special software is installed in a data centre and distributed across many servers, allowing it to serve a very large number of customers at once. |
| 33. | **"Third Party (External Party)"** | means a natural or legal person involved in accessing, processing and managing information assets. |
| 34. | **"Web"** | means the part of the Internet, the resources that can be accessed on the Internet using URLs (Uniform Resource Identifiers). Web resources can be used by an application called a browser. |

# 4. THE MAIN OBJECTIVES AND TASKS OF THE VILNIUS CITY DATA POLICY

*The Vilnius City Data Policy* is intended to provide guidelines for the development of an integrated database for the City of Vilnius and its subordinate enterprises, to ensure the secure and efficient management of data and related processes and provision of services, in accordance with global best practice and international standards. The policy describes a data architecture that enables the integration of different types of data, the establishment of unified and common data architecture standards and principles, and, most importantly, empowers the Vilnius City Administration to analyse and act on the basis of insight-based data covering different fields, processed according to common methods and standards. At the development stages of the integrated database of the City of Vilnius and its enterprises, only depersonalised data and other information of a general nature will be used, from which it will not be possible to identify a specific data subject or group of data subjects.

## 4.1. OBJECTIVES:

● Enable the City of Vilnius and its subordinate enterprises (concerned parties) to consolidate and exchange data using the central Vilnius API and to display them using the data platform developed by the Vilniaus planas, in compliance with data security requirements;

● Determine and define the future data platform architecture and related documentation (standards, methods, principles, functionalities, tools, technology base, processes, etc.);

● Determine and define the data platforms to be used and the related technology components;

● Create an innovative, friendly environment in Vilnius and increase added value for the city's residents;

● Achieve a more efficient use of the city's financial resources;

● Successfully implement cooperation between the City of Vilnius and its corporate group for the practical application of the policy;

● Determine and define the necessary data sources and determine the requirements for data exchange.

## 4.2. TASKS:

● Create an Economic and Service Unit template including geo-location and data parameters based on the 10 strategic directions of Vilnius City:

- Education;
- Social security;
- Health, Wellness and Sport;
- Culture;
- Mobility;
- Economic Development and Tourism;
- Environment and City Development;
- Safety and Security;
- Services;
- Management Excellence and Planning.

● Create a unified statistical template for the Population activity profile, including basic data and other related information, for example by taxpayer number. The initial data of the data model could include data from the defined 10 fields of Vilnius city.

● Create a unified template for the Production (Enterprise) activity profile, for example by taxpayer number. The initial data of the data model could include data from the defined 10 fields of Vilnius city.

● Develop initial data models by grouping tax, enterprise and population groups (according to a specific classification) and controllers.

● Develop an initial analytical data model including data of Economic and Service Unit, Population Activity Profile, Production (Enterprise) Activity Profile, with geo-locations and geo-location relations.

● Produce visualisations to display processed data, for example:
  - 3D map with 3D geolocation polygons and 3D dashboards capable of presenting the results in colour and dimension (different sizes in x, y, z axes), numbers and symbols, graphs by Economic and Service Unit, Population Activity Profile and Production Unit, Production Activity Profile, with their geolocation from the smallest segment – a street, an eldership or an eldership up to the whole city or a specific territory;
  - The map should show coloured polygons of values at the selected detail according to the value of the layer: simple view, show all financial data, income and expenses, income only, expenses only;
  - The visualisations provided should allow to distinguish between types and categories of activities, so that changes can be planned, new policies can be amended or developed to improve the efficiency of a particular sector in the selected field;
  - Review the dynamics of changes over the previous period, based on historical data, and understand whether the changes have had an economic effect. Simulate and execute forecasts in a 3D environment by changing the values of financial and economic data;
  - Graphically map 3D and present the economic dimensions in 3D expression of how much and what data has been collected in each segment from which the calculations are made, and the

visualisations of all data of Economic and Service Unit, Population Activity Profile and Production Unit, Production Activity Profile data to their geolocations and the results of the calculated values (distances, quantities, metrics, areas, heights, types and other criteria.

- Develop tools to analyse interrelationships and dependencies in order to reform the segment's economic management, financial management, effective decision-making and implementation of change policies in the city, and to reform the entire administration of internal and external services. To enable planning for the development of city services, production and implementation of innovations.

# 5. KEY COMPONENTS AND REQUIREMENTS FOR THE POLICY IMPLEMENTATION

## 5.1. STAGES OF IMPLEMENTATION:

1. **Analysis and Assessment** – Analysis and assessment of the current situation – Inventory of data managed by the City of Vilnius and its subordinate enterprises;

2. **Data Integration** – Contracting, data preparation, development of tools for integration and integration of internal and external data sources.

3. **Data Analytics and Domain Visualisation** – Development of tools for data analytics and visualisation.

4. **Operation and Improvement of the Data Platform** – Ensuring the continuous operation, availability and accessibility of the platform, and the continuous improvement of the data platform. Monitoring and development of improvement measures.

## 5.2. IMPLEMENTATION REQUIREMENTS

In order to avoid harmful inter-institutional fragmentation, the Vilnius City Data Policy should be implemented in accordance with the principles of cross-sectoral (horizontal) access to data and means of their use.

# 6. DATA PLATFORM ARCHITECTURE

## 6.1. COMMON DATA ARCHITECTURE PLATFORM

Based on currently available technologies and their development directions, the core of the data platform should be based on a massively parallel processing (MPP) type architecture, which allows for the partitioning of data storage and processing into distinct, independent components, enabling the implementation of processes working in parallel with individual data blocks. These types of architectures are fully adapted to work with large and constantly growing volumes of data and provide end-users with fast turnaround times for data queries and the ability continuously to select and expand both data storage and processing resources.

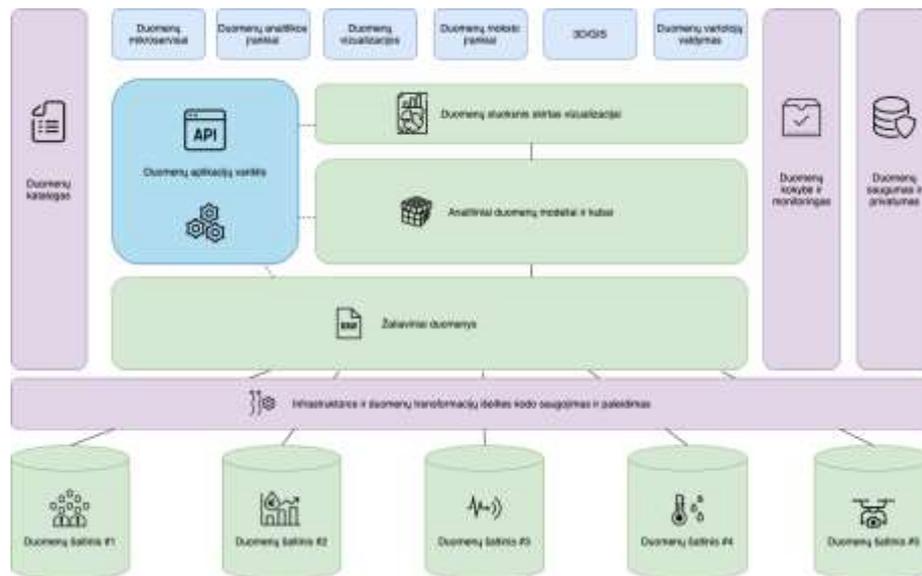The figure below shows the architecture of the data platform components:

*Figure 2 – Data Platform Architecture*

## 6.2. COMPONENTS OF THE DATA PLATFORM, THEIR COMMUNICATION AND PURPOSE

- **Data Sources** are a wide range of data sources and their types, including but not limited to structured data, CSV, XML, PSV files, direct logins to databases using various libraries (e.g. ODBC/JDBC), APIs, data and message queues, and unstructured large volume data, e.g. video and voice files, 3D/2D images, etc. It is essential that the data platform has the ability to work with as many different sources as possible and does not create artificial constraints on data load.

- **Data Storage Layers** (raw data, analytical data models and cubes, data for visualisation) are layers of data storage for implementation of different tasks. The raw data layer is designed to store data that is not modified and mirrors the data stored in the data source. In the analytical data models and cubes layer, the data is transformed, taking into account the business logic and its applicability to analytical tasks. The highest abstraction data layer for visualisation is composed of aggregated data that are adapted to the visualisation tools and to work efficiently with them, without moving the data transformation to the visualisation tools.

- **Data Application Engine** is a layer for implementation of data processing and transformation. It accesses all data layers and can efficiently implement ETL/ELT data transformation logic for aggregation, business logic implementation, merging, etc.

- **Data Microservices** is an engine to create an integration layer for easy and simple creation of APIs or data knowledge-based independent applications that can be integrated with each other and with third parties.

- **Data Analytics and Data Science Tools** are applications for working and experimenting with data using popular programming, data processing and machine learning libraries.

- **Data Visualisations and 3D/GIS** are tools that are integrated with the data visualisation layer and visually display the desired information in a variety of formats – both classic graphs and 3D visualisations.

- **Data Catalog** is a critical component of the data platform for describing data and monitoring its transformation interfaces. A well-organised and complete data catalog makes it easier for organisations to understand their data, its provenance and quality, leading to comprehensive insights and global application of data in everyday processes and performance monitoring.

- **Data Quality and Monitoring** is the ability to monitor and ensure that the data on the data platform is up-to-date and of appropriate quality, and that in the event of any deficiencies in the data, an incident is automatically raised and the relevant specialists are informed so that they can resolve the data failures.

- **Data Security and Privacy, User Management** is a set of functions and indicators that allows to control the availability and use of data, as well as to monitor all data activities and ensure that data handling complies with all legal and security requirements (e.g. **GDPR**).

Storing and running source texts is a component that allows to store, version and run source texts in the relevant data platform environments. All work performed on the data platform should be done using this component and it is essential that any change is made using this environment and can be reverted to the previous version if necessary.

# 7.   DATA PLATFORM AND ARCHITECTURE REQUIREMENTS

The data platform and architecture is based on the following fundamental principles (guidelines):

- **Openness** - the data platform should be open to all data sources, formats, structures, types, and should not in any way be directed to working with data processed by only one manufacturer or technology;

- **Flexibility and usability** – the data platform should allow for rapid development of data processes and algorithms, ensuring high process quality. The platform should be constantly developed, updated and based on the idea that everything is constantly changing, so the functionality offered should be able to adapt as quickly as possible to changes and new needs;

- **Ease and simplicity** – the platform should require a minimum support and allow data engineers, analysts, scientists and other users easily to develop and adapt it to the needs and processes of the organisation;

- **Traceability and regulatory compliance** – the scope of the platform should include components and tools that allow easy identification and tracking of each user's actions, easy conduct of required audits and compliance with personal data protection and other applicable regulations.

- The data platform should provide completely isolated and separate technological environments for development, testing, teaching and production.

These principles should ensure the selection of various technological solutions and alternatives relevant to implementation that best meets all the needs of the organization (economic, technological, integrity, etc.). This document does not in any way exclude or otherwise restrict the organisation's choices or criteria

for equivalence of solutions, but clearly and unambiguously sets out the key guidelines for the organisation to follow when selecting a data platform architecture and available technologies and alternatives.

# 8. DATA PLATFORM FUNCTIONALITY REQUIREMENTS, TOOLS AND OPERATIONAL PROCESSES

## 8.1. DATA SOURCES, THEIR TYPES AND INTEGRATION

The data platform should have the functionality easily to integrate (without the need for additional tools) a wide range of data types, sources and structures using existing and established data integration mechanisms.

Below is a minimum list of data sources that should be supported for direct data integration, allowing data integration without the need for additional work to develop or parameterise the data integration software.

Files and file storage systems:

- Amazon S3
- Azure Blob Storage
- Google Cloud Storage
- HDFS
- FTP and SFTP

Databases:

- Microsoft Access
- Oracle
- MySQL
- MariaDB
- MongoDB
- Cassandra
- Microsoft SQL
- DB2
- PostgreSQL
- Google BigQuery

Protocols:

- SOAP
- Generic HTTP
- Generic REST
- Generic ODBC
- Kafka
- RabbitMQ

## 8.2. DATA STORAGE AND LAYERS

The data should be stored on the data platform using the traditional and widely used three-layer/level methodology, consisting of the following layers:

### 1. Raw data

This is a data layer that stores data in the same format as it is stored in the data source, using only the minimal transformations that are necessary to make optimal and efficient use of technological resources to store this data – e.g. a transformation or logic that allows to understand the changes that have taken place in the data source and only saves the changes in the raw data layer (delta data loading). Also, minimal transformations would include changes to the structure of the data that eliminate or remove duplicate or poor data.

It is important to note that this data layer includes the physical storage of data and is created by storing a real copy of data. This layer cannot be virtual, based on data stored in other systems or files, to avoid third parties being overloaded by further transformations.

### 2. Analytical data models and cubes obtained by transforming raw data

Analytical data models and cubes on the data platform will be created using the popular dimension and fact modelling methodology known as Kimball.

This data modelling methodology is simple and flexible, and its proposed data model design includes four steps:

1. (Business) process selection
2. Declaration of granulation or particularity
3. Identification of dimensions
4. Identification of facts

The main objectives of this data layer include the logic of integrating/merging different data sources, the display and transformation of historical data in order to structure it in the desired format. In many cases, data structuring involves the decomposition of data into individual data tables, therefore, such an image is not entirely convenient for end users, and a presentation or visualisation layer is used to meet their needs.

In this data model, data is also mostly physically stored, as faster data visualisation and transformation speed are achievable for obvious technical reasons, such as a lower need for RAM.

### 3. Presentation or visualisation layer created by aggregating data available in analytical data models and cubes

This data layer is the most abstract, but it is intended for end-users of data – managers, analysts, employees, etc., and is often used for integration with other tools or systems. This layer is usually used to store data in an intuitive, easy-to-understand format, ensuring a high quality of data accuracy.

Unlike other data layers, it often does not store data physically, but only virtually, by applying data transformation rules to data stored in other layers.

## 8.3.  DATA TRANSFORMATION AND DATA APPLICATION ENGINE

The data transformation and application engine should be based on a containerised software package orchestration engine for running applications, managing (increasing or decreasing) allocated resources, and other operations allowing to optimise the use of a large amount of CPU, RAM and hard disk resources deployed on more than one physical or virtualised workstation. It is the flexible, convenient and dynamic management of computing resources that is the key feature of the transformation and application engine, ensuring that the right amount of resources is dedicated to the desired tasks and avoiding situations where there is too much or too little computing capacity being acquired.

With this type of transformation and application engine, all the necessary transformation or application logic steps can be expressed as software source code, combined into a container and run as a task or process to perform the specified tasks.

Application containers should be based on the Docker operating system-level software virtualisation mechanism, which allows software packages to be isolated and interconnected by means of detailed methods. The data transformation and application engine should be able to work with Docker software registries and should be able fully to manage the running of registry packages and resource allocation.

Typical data transformations include the following functionalities and the implementation of the functionalities listed below are the minimum requirements for the application and transformation engine:

- Merging and combining arrays
- Data filtering
- Data aggregation
- Managing and filling empty values
- Creation of new data fields
- Data categories and characteristics

This data transformation and application engine is also used for the development and implementation of the data microservices described below, and to access open data of the City of Vilnius and its subordinate enterprises. Open data requirements and guidelines are defined in the Open Data Policy of the City of Vilnius and its subordinate enterprises.

## 8.4.  DATA CATALOG

Data catalog plays a key role in a modern data platform, as its integrated functions allow different users to detect data, understand its structure, identify its uses and understand its provenance, including transformations.

Data catalog should be able to describe a wide range of data, beyond its location, structure or source software, and should allow intuitive user interfaces for searching using technical data definitions or business sense information.

Key functional requirements for data catalog:

- Search by phrase and compiled dictionary
- Automated data crawling
- Data classification
- Technical data quality
- Business Data Log
- Metadata provenance path
- Documenting and linking policies and business logic to data
- Assignment of data owner, processor and other roles
- Defining the sequence of data processing activities
- Data security management and access control
- Implementation of data security rules and rights management

## 8.5. DATA MICROSERVICES

The data transformation and application engine described above should also enable the creation of various microservices. One of the most common microservices of this type are REST API applications that can perform data transformations, filtering, implement security requirements and enable the data contained on the data platform to be applied to other software solutions according to certain criteria.

Data microservices are standalone virtualised software packages using Docker technology, which can be programmed according to the organisation's needs using the following popular programming languages:
- Java
- Python
- Node.JS
- GO
- PHP
- .NET

## 8.6. DATA ANALYTICS TOOLS

One of the main tasks of the data platform is to enable data analysts to work qualitatively, simply and efficiently with established data analytics tools in the market, allowing to perform a variety of data transformations and provide the organisation with unique and useful insights.

Given the market trends, the top three programming languages for data analytics are Python, SQL and R, so it is only natural that the data platform should fully support all of these programming languages and allow to make use of various data analytics libraries (e.g. Pandas).

## 8.7. DATA VISUALISATION AND BUSINESS ANALYTICS

Data visualisation and business insight tools are often one of the most widely used components in organisations, as the resulting insight reports become a useful performance measurement and planning tool for many managers and employees. The data platform tools purchased should allow data to be accessed, visualised (using various types of graphs) and analysed at different scales.

Key requirements for data visualisation tools on the data platform:

- Publishing, sharing and communicating the reports produced among the members of the organisation using the chat or comment functions
- Specialised reporting software is installed on the computer of the data analyst or other data platform developer
- Access to all data on the platform and stored data
- Breakdown of pages within a report
- Accessibility and mobile-friendliness of reports
- Ability to update the data from the data source and select automatic update frequency
- Using AI-based visualisations (text analytics, image recognition)
- Data security settings and data encryption
- Ability to work with external microservices and APIs

## 8.8.    DATA SCIENCE AND TOOLS

Along with data analytics tools and programming languages, the data platform should enable its users and programmers to work with artificial intelligence, machine learning algorithms and other data science solutions. Based on current trends, data science should be provided with a Jupyter based interactive tool that allows real-time work with data, programming languages, various libraries and other tools needed for the work. It is important to note that the tool used for data science on the data platform should be web-based and should not require additional software on the data scientist's computer.

It is important to note that a Jupyter based tool on the data platform should support different libraries for adaptation, visualisation of data science and work with data.

## 8.9.    3D/GIS COMPONENT

In order to ensure the widespread use of the platform, it is essential to ensure that the data platform has a GIS component, which includes tools to work with 2D and 3D spatial data, the integration of BIM, and the visualisation of all this data on a common platform.

Requirements/recommendations for GIS component tools, work with 2D and 3D data:
- Ability to scan vector (e.g. Google Keyhole Markup Language (KML/KMZ), Esri Shapefile), raster (e.g. GeoTIFF, ERDAS Imagine (IMG)), LiDAR (e.g,  ASPRS LiDAR Data Exchange Format: .LAS, .LASD, .LAZ; Point Cloud XYZ), CAD (e.g. Autodesk Drawing: .DWG, .DWF, .DXF), Web Services (e.g. Web Feature Service (WFS),  Esri ArcGIS Online Web Services), 3D (e.g. COLLADA: .DAE) data;
- Ability to provide statistical information (minimum, maximum, total values) on data quantity and attributes;
- Ability to link data elements through a common denominator (e.g. administrative unit);
- Ability to select data based on location and attribute information;
- Ability to group data;
- Ability to conduct spatial analysis (e.g. density, distance calculation);

- Ability to conduct surface analysis (e.g. calculate slope gradient);
- Ability to edit 2D and 3D elements both manually and automatically;
- Ability to measure length, width and height on the map, as well as automatically to count such information in the fields of the attribute table.

Requirements/recommendations for 3D/GIS data visualisation:
- Visualisation of elements by a specified attribute;
- Visualisation of elements by grouped values;
- Visualisation and configuration of element headers;
- Visualisation and configuration of element pop-ups;
- Visualisation of elements in both 2D and 3D maps;
- Use of realistic symbolism for 3D elements (e.g. trees, buildings);
- Visualisation of BIM models (IFC, Revit), ability to disable/enable separate building layers.

## 8.10. DATA QUALITY AND MONITORING

In order to ensure that the platform can be used and adapted to improve and enhance business processes, it is essential to ensure that the data platform has a mechanism to control and measure the quality of data. Trust in the quality of the data contained on the data platform is one of the most important, but not obvious, factors determining the usability of the data platform and willingness of users to use it in their daily activities. Technically speaking, this functionality includes the calculation of metrics across all data layers, according to the required logic (e.g. population), and storage, visualisation and alarm messaging of these metrics.

Typically, a time-series database is used to store and display such data, and a complementary set of visualisation tools is used to allow flexible and virtually unrestricted computation, display and monitoring of a wide range of data quality parameters, which are heavily dependent on the data source and data modelling logic.

## 8.11. DATA SECURITY

Modern data platforms should provide various functionalities to classify data by assigning different security parameters and identifying different security classes or categories. The identification of different types and levels of data allows different data security measures to be applied to them and the following parameters to be set:

- Assign different access to each user or group depending on the type of data security, using a centralised user authorisation mechanism
- Encrypt data both during transmission and storage – the data platform should have the functionality to configure, activate and manage data encryption and keys according to the needs (it should be possible to use your own encryption keys)
- Monitor all actions performed on data and receive notifications on the use and transfer of high security data

In addition, the security parameters covering the entire data platform infrastructure should contribute to data security:

- The infrastructure provider should have a 24-hour security centre with specialists proactively working on infrastructure security
- DDoS protection should be integrated into all services
- The provider should provide clear and transparent information on security processes, compliance with standards, investments, liability model and incidents

Below is a model of the data security responsibilities of the data platform provider:



*Figure 5 – Service Provider Data Security Responsibility Model*

## 8.12. USER MANAGEMENT

The users of the data platform should be managed through the centralised user platform – Microsoft Active Directory, which currently already contains information on all users and can easily divide users into different groups to provide access to the data platform components and data. The data platform should fully integrate with Active Directory, provide for the management of different users and groups, guarantee the possibility to implement 2-factor authorisations, establish a common password management policy and other necessary security and monitoring functionalities.

In addition to user management, the data platform should include the following functionalities related to user management:

- Ability to detect user login anomalies and inform data platform administrators

- Enable users securely to store passwords and encryption keys
- Allow control of access levels and user access to individual components of the data platform
- Ensure additional login restrictions based on IP addresses, locations, etc.

## 8.13.    SOURCE TEXT STORAGE AND RUNNING

A Git based repository with code integration, running and testing abilities should be offered for storing and running the source texts on the data platform. This repository will store all the source text related to both the infrastructure of the data platform (all components and their configuration) and the data transformations and models (all data transformation rules, program code, data training engines, etc.).

It is important to ensure that the proposed source text repository has the ability to integrate with all other components and/or needs of the platform.

## 8.14.    DATA PROCESSING

Data of the City of Vilnius and its subordinate enterprises are processed and stored in accordance with the FAIR principles (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable). Data should be findable, accessible, interoperable and reusable.

**15 FAIR principles**[2]**:**
- **F- indable**
    1. F1: (Meta) data are assigned globally unique and persistent identifiers;
    2. F2: Data are described with rich metadata;
    3. F3: Metadata clearly and explicitly include the identifier of the data they describe;
    4. F4: (Meta)data are registered or indexed in a searchable resource.
- **A** - **ccessible**
    5. A1: (Meta)data are retrievable by their identifier using a standardised communication protocol;
    6. A1.1: The protocol is open, free and universally implementable;
    7. A1.2: The protocol allows for an authentication and authorisation where necessary;
    8. A2: Metadata should be accessible even when the data is no longer available.
- **I** - **nteroperable**
    9. I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation;
    10. I2: (Meta)data use vocabularies that follow the FAIR principles;
    11. I3: (Meta) data include qualified references to other (meta) data.
- **R** - **eusable**
    12. R1: (Meta)data are richly described with a plurality of accurate and relevant attributes;
    13. R1.1: (Meta)data are released with a clear and accessible data usage license;
    14. R1.2: (Meta)data are associated with detailed provenance;
    15. R1.3: (Meta)data meet domain-relevant community standards.

Usually, data should be integrated with other data. In addition, data should be interacted with applications or workflows for analysis, storage and processing.

---

[2] Guidelines on Fair Data Management in Horizon:

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

# 9. GENERAL REQUIREMENTS FOR PROCESSING PERSONAL DATA

The City of Vilnius and its enterprises, in the exercise of their autonomous and state-delegated functions, process and use personal data in accordance with their competence for the performance of their administrative functions and provision and development of public services.

When processing personal data, the City of Vilnius and its enterprises should comply with the GDPR, the LLPPD and other requirements set out in the Republic of Lithuania and in the internal legal acts of the City of Vilnius and its enterprises related to the processing and protection of personal data.

If none of the conditions for lawful processing of personal data and lawful purpose of processing of personal data set out in the GDPR are met, personal data cannot be collected and processed by the City of Vilnius and its enterprises.

# 10. OPERATIONAL AND SUPPORT PRINCIPLES AND REQUIREMENTS

Data platform operations, support and releases of all platform components should be included in the overall data platform package. According to the ITIL (Information Technology Infrastructure Library) methodology, the support and maintenance services for the data platform shall include the full provision of these three processes, including the release of the software for all components:

- Incident management
- Changes management
- Release management

A 24x7 centralised helpdesk is mandatory for these services:

- The helpdesk service uses specialised software to record all requests and incidents within the organisation and view the status of their resolution, generate reports, etc.
- The helpdesk should communicate in Lithuanian by email, in writing or orally, and all requests can be recorded by email, fixed-line and mobile phone, through the WEB interface, and automatically through the API interface from the data platform components.
- The helpdesk should operate on one-stop shop basis and should also be able to manage incidents, requests or changes related to third parties – manufacturers, raw material supply systems. Third party management includes monitoring service levels, monitoring and escalating agreed deadlines.

# 11. IMPLEMENTING STANDARDS AND LEGISLATION

The data platform described in this document is based on the following requirements set out in international standards and legislation governing the management of IT services and technologies and information security:

- **ISO 20000** – an international standard for IT service management, largely based on best practices and ITIL methodology;

- **ISO 27001** – an international standard for information security management, covering all stages of the information security management cycle;

- **ISO 14001** – an international standard that defines the process of controlling and improving the environmental performance of companies.

- **Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)** – a European Union regulation that sets out detailed requirements for business entities and organisations on the collection, storage and management of personal data, which are mandatory for all organisations whose activities are aimed at EU citizens;

- **EU Cloud Code of Conduct** – a voluntary set of requirements for cloud and management service providers, describing their operating principles and management.

# 12. STAGES AND GUIDELINES OF IMPLEMENTATION



*Figure 3 – Stages and Guidelines of Implementation*

# 13.   CONCERNED PARTIES, THEIR ROLES AND RESPONSIBILITIES

The City of Vilnius, its subordinate enterprises, education, health, science and study institutions, business entities, residents of the City of Vilnius, the District of Vilnius and other regions.

The Data Policy enables the City of Vilnius and its subordinate enterprises (concerned parties) to consolidate and exchange data using the central Vilnius API and display them using the data platform developed by the Vilniaus Planas, in compliance with data security requirements.

Role of concerned parties:
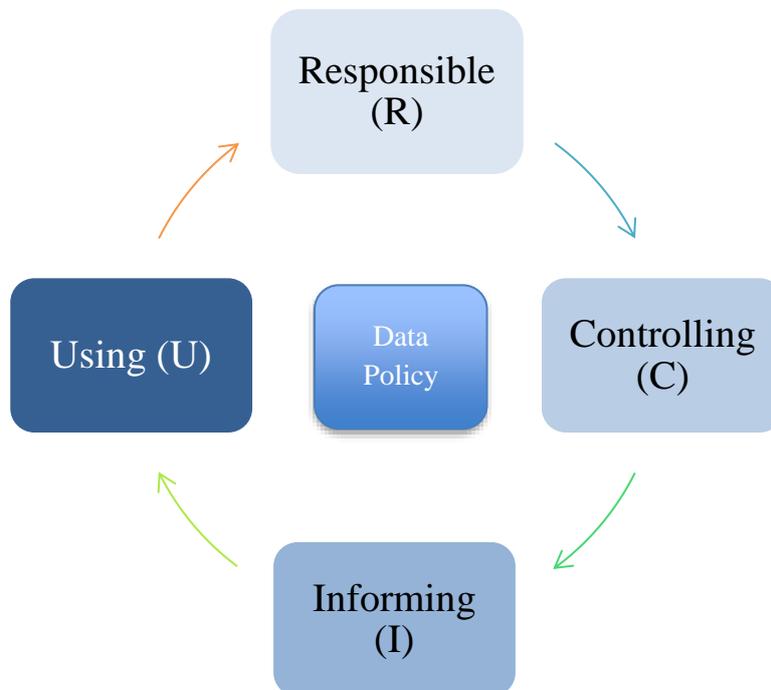


*Figure 4 – Role of Concerned Parties*

**Responsible (R)** – the departments of the City of Vilnius and its subordinate enterprises designated for the implementation of the data policy are required to implement the requirements of the policy for data consolidation, data exchange and display using the central Vilnius API and the data platform developed by the Vilniaus Planas, in accordance with the data security requirements. Responsible parties provide data, insights and proposals for the improvement of the Data Policy. Concerned parties are also responsible for developing the guidance, requirements and rules for the achievement of the intended results.

**Controlling (C)** – the role of controlling parties is mainly played by the Vilnius City Administration and administrations of the Vilnius City departments and enterprises and institutions managed by the City of Vilnius. The controlling parties monitor the course of implementation of the Data Policy in accordance with the calendar plans and periodic reports produced by the responsible parties, and compile and submit proposals to the responsible parties for improving the Data Policy.

**Informing (I):** Concerned parties who can influence or have an impact on, or pose risks or restrictions to the implementation of the policy, initiating supplements and adjustments to the Data Policy.

**Using (U):** Concerned parties who can use the results described and implemented in the policy.

The users are mainly the City of Vilnius and its subordinate enterprises, external parties, city's residents, business entities, research and study institutions.

# 14.   APPLICATION OF THE POLICY

The City of Vilnius and its subordinate enterprises carry out strategic planning, feasibility studies to improve the infrastructure and accessibility of services to the city's residents, to develop projects and meet other needs of the city's residents, business entities, research and study institutions.

# 15.   BENEFITS FOR SOCIETY

Over the past few years, digital technologies fundamentally changed the economy and society, affecting all sectors of activity and daily lives of all Vilnius residents. The decisive factor in this transformation, which is still in full swing, is the data. Data-based innovation will bring huge benefits to Vilnius residents, improving the accessibility of personalised services, providing new opportunities for mobility and other services, and contributing to the achievement of Vilnius City strategic goals by using data as a key tool for providing insights, making quick but reliable decisions, and reducing errors. When using data as a basis for decision-making, it is essential to ensure that data are collected and used in a way that puts the interests of the individual first and respects European, Lithuanian and Vilnius values, fundamental rights and rules.

Data is the engine of economic development. It underpins the development of new products and services, leading to increased productivity and resource efficiency in all sectors of the economy, more personalised products and services, better policy-making and modernisation of government services. Data is a key resource for developing products and services. Access to data is essential for training AI systems, as products and services are rapidly evolving from recognising patterns and providing insights to more sophisticated forecasting methods and therefore better solutions.

The socio-economic well-being potential provided by data is also worth taking advantage of because, unlike most economic resources, data can be reproduced at almost no cost and can be used by more than one person or organisation at the same time. This potential should be used to meet the needs of individuals, thereby creating value for society and economy. Better access to and responsible use of data is essential to realise these opportunities.

Vilnius City can become an example of a city where business and public sector entities can make more informed strategic decisions that affect the quality of life of its citizens, the performance of its businesses, and the quality and accessibility of its public services, through the use of detailed historical data, the monitoring of the current situation and accurate forecasts.

To achieve this, the City of Vilnius can rely on a strong legal system of data protection and cyber security. To become a leader in the data economy in Lithuania and across Europe, the City of Vilnius is taking action now and is working together with its subordinate enterprises and other entities to resolve the issues of data consolidation, processing and storage, and cybersecurity management.

In order to avoid harmful inter-institutional fragmentation due to inconsistency of actions in different sectors and enterprises, it is necessary to develop a comprehensive data system – a data exchange system – based on cross-sectoral (or horizontal) principles of access to and use of data, which is based on the principles of accessing and using data, and which is based on the principle of rationality and logical architecture aiming at the avoidance of duplication of data and/or systems, an innovative approach to integrations, as well as a rational planning, called by the City of Vilnius as "Green Data".

In line with this principle, the first priority in the implementation of the vision is to develop the right legislative framework for the management of shared data spaces in Vilnius City. Such management

structures should support decisions on which data can be used in which cases, facilitate the use/sharing of data and help prioritise sectoral and cross-sectoral interoperability requirements and standards, taking into account the need for sectoral authorities to identify sector-specific requirements. This framework will help to strengthen the structures of the City of Vilnius to facilitate the use of data for innovative business ideas, both at the sectoral and cross-sectoral levels. It will be developed on recent initiatives in EU Member States 35 and in individual sectors to address one or more of the following tasks:

- **strengthen** management mechanisms at the level of the City of Vilnius for cross-sectoral data use and data use in shared sectoral data spaces involving both private and public sector entities;

- **help** facilitate decisions on who can use what data for research purposes and how, without prejudice to the General Data Protection Regulation. This is particularly applicable for public sector databases containing sensitive data outside the scope of the Open Data Directive;

- **ensure** that those who so wish can more easily authorise to use the data they generate in the public interest (data altruism), without prejudice to the General Data Protection Regulation.

In implementing this policy, the City of Vilnius will ensure the conditions for:

- re-use of more public sector data of appropriate quality, in particular taking into account the opportunities it offers to SMEs;

- promotion of data sharing between business entities and public authorities for public interest purposes.

Mandatory access to data should be required only where specific circumstances so require, subject, where appropriate, to fair, transparent, reasonable, proportionate and/or non-discriminatory conditions.

# 16.  COMPATIBILITY

The Data Policy is being compiled in the context of the Vilnius City Strategic Plan 2020-2030, the Vilnius City Strategic Direction Vilnius 2IN, the Internet of Things Policy, the Vilnius City Master Plan, and the Smart Cities Initiative supported by the European Commission – the European Innovation Partnership on Smart Cities and Communities (EIP-SCC), which brings together cities, industry, small businesses, banks, research and others. The EIP-SCC aims to improve the city life through more sustainable integrated solutions and address city-specific challenges in different policies such as energy, mobility, transport and ICT development. The EIP-SCC is based on the involvement of the public, industry and other concerned groups in the development of innovative solutions and participation in the city governance.

# 17.  FINAL PROVISIONS

The Vilnius City Data Policy shall be approved by the Director of the Vilnius City Administration.

The Vilnius City Data Policy shall be reviewed annually and updated as necessary.

The Vilnius City Administration shall be responsible for concluding agreements with third parties on the receipt of data.

In order to comply with the requirements set out in Article 28(3) of Regulation (EU) 2016/679, the **rights** and **obligations of the controller and processor** shall be set out in the data exchange or data receipt agreements.

Data **consolidation** and **exchange** between the City of Vilnius and its subordinate enterprises (concerned parties) shall be carried out using the central Vilnius API.

Data shall be **displayed** using the data platform developed by Vilniaus Planas.

The implementation of the Vilnius City Data Policy shall be organised by the heads of the City of Vilnius and its subordinate enterprises, or other persons appointed by them.

When enterprises of the City of Vilnius provide data to third parties, the datasets provided that are not described in the *Open Data Policy* should be agreed with ITG and Municipal Enterprise Vilniaus Planas by informing them at e-mail addresses: atviras@vilnius.lt and duomenys@vplanas.lt.

Vilnius Municipal Enterprise Vilniaus Planas shall not provide any raw data to third parties. Only analysed and summarised statistical data shall be made publicly available by Vilnius Municipal Enterprise Vilniaus Planas.

The implementation of the Vilnius City Data Policy shall be monitored by the Director of Administration or another responsible person appointed by him.

# 18. ANNEXES

1. ANNEX NO 1 – STAGES OF DATA POLICY IMPLEMENTATION

Policy Development Working Group:

| No | Enterprise | Name | Surname |
|---|---|---|---|
| 1. | Vilniaus planas, SĮ | Stasys | Savilionis |
| 2. | Vilniaus miesto savivaldybės administracija, BĮ | Dovilė | Krekenaitė |
| 3. | Vilniaus miesto savivaldybės administracija, BĮ | Jonas | Pidkovas |
| 4. | Vilniaus miesto savivaldybės administracija, BĮ | Matas | Mačiulis |
| 5. | Vilniaus miesto savivaldybės administracija, BĮ | Aistė | Paludnevičiūtė |
| 6. | Vilniaus šilumos tinklai, AB | Anastasija | Cibulskienė |
| 7. | Susisiekimo paslaugos, SĮ | Mindaugas | Žironas |
| 8. | Grinda, UAB | Dalius | Kuliešius |
| 9. | Grinda, UAB | Dainius | Nemanis |
| 10. | Vilniaus atliekų sistemos administratorius, SĮ | Eglė | Grinkevičienė |
| 11. | Vilniaus vandenys, UAB | Simas | Sodys |
| 12. | Vilniaus planas, SĮ | Aurelijus | Deksnys |

**ANNEX NO 1** – STAGES OF DATA POLICY IMPLEMENTATION

| Data submission methodology | | | |
|---|---|---|---|
| **Stage** | **Annex** | **Result** | **Requirements** |
| **1** | (A) (K) (I) Data Inventory | List of data held by the City of Vilnius and its subordinate enterprises | The list should include:<br>1. Data owner;<br>2. Data (name of the data);<br>3. Primary data source;<br>4. Data storage format;<br>5. Frequency of data update/collection;<br>6. Legal basis for data collection. |
| **2** | (A)(K)(I) Data Preparation and Integration of Internal Data Sources | Data platform filled with internal data | List of internal system integrations;<br>Data exchange integration works. |
| **3** | (A)(K)(I) Integration of External Data Sources | Data platform filled with external data | List of external system integrations;<br>Data integration works. |
| **4** | (A)(K)(I) Analytical Tools | Description of analytical needs | Detailed description of analytical needs |

**Concerned parties:**
- **(R)** **Responsible**
- **(C)** **Controlling**
- **(I)** **Informing**
- **(U)** **Using**