POLITECNICO
MILANO 1863

iAPP4P

Requirement Analysis and Specification Document

Lorenzo Carlassara

Angelica Iseni

Emma Lodetti

Virginia Valeri

Software Engineering for Geoinformatics

Academic year 2022/23

Title:	Requirement Analysis and Specification Document
Authors:	Carlassara L. (10601118), Iseni A. (10668862), Lodetti E. (10619244), Valeri V. (10639607)
Date:	June 18, 2023
Professor:	Giovanni E. Quattrocchi

Contents

1 Introduction

The purpose of this document is to describe the requirements for the development of an interactive client-server application that enables users to access, query and visualize air quality data retrieved from existing public digital archives. The system consists of three main components: a database to ingest and store the data, psycopg2 and SQLAlchemy packages for querying the databases and retrieve the data through a python script, and a jupyter dashboard to provide means for requesting, processing (descriptive statistics), and visualizing data (maps, dynamic graphs).

1.1 Context and motivations

The European Commission's agreement on electronic fuels will enable the sale of heat-powered vehicles beyond 2035, when the ban on gasoline and diesel cars comes into effect. This is provided that these vehicles run on synthetic, climate-neutral fuels. Additionally, European energy ministers have approved the regulation to phase out gasoline and diesel engines by 2035 through a majority vote.

Public authorities collect air quality and weather observations in near-real time from ground sensor stations and store them in digital archives. Ground sensors data are composed of long time series of observations and sensors metadata, including coordinates, type of measured variable, etc. Often, observations from different sensors and/or providers require different patterns for data accessing, harmonization, and processing. Interactive applications and dashboards, capable of facilitating such tasks, are key for supporting both public authorities as well as ordinary citizens in data processing and visualization [1].

Our app, **iAPP4P** (interactive Application for Air Pollution monitoring) aims to provide people with a clear and scientifically based view of the current state of knowledge about air pollutants and their potential environmental impacts on human health.

1.2 Definitions, acronyms, abbreviations

- **Common citizens (CC):** refers to an average person who is a member of a particular community, society, or country and who is not distinguished by any special status, profession, or achievement. In the context of government and politics, the term is often used to describe the majority of the population who are not involved in the decision-making process or who do not hold any formal political power.
- **Geoinformatics/Environemental Engineer(GEO):** is a specialist who integrates scientific knowledge, engineering principles, and advanced geospatial technologies to analyze, model, and address environmental challenges, with a focus on sustainable practices and minimizing the impact of human activities on the environment.
- **Computer Science Engineer(CS):** is a specialist who applies computational principles and technological expertise to design, develop, and optimize computer systems and software solutions.
- **Mathematical Engineer(MTM):** is a skilled professional who applies mathematical theories, computational techniques, and algorithmic approaches to design, analyze, and optimize engineering systems, enabling efficient problem-solving and innovative solutions in various domains.
- **Air quality:** refers to the degree to which the air is clean, clear, and free from pollutants.
- **Ground sensors:** refers to the sensors installed on the ground that monitor air quality and weather conditions.
- **Public digital archives:** refers to online repositories or databases of digital information that are publicly accessible and maintained by governmental or public entities.

- **Integrated Development Environment (IDE):** refers to a software application that provides a comprehensive set of tools and features for software development, including code editing, debugging, and project management.

1.3 Solution overview

In our project, the application consists in a set of function that allows user to manipulate and visualize data.

- **Data processing:** the user can obtain the descriptive statistics of the data of interest.
- **Data visualization:** the user can visualize data through maps, indexes, tables and graphs.
- **Work with updated data:** registered users can access to updated data.

1.4 Scope and limitations

Observations from different sensors require different patterns for data accessing, harmonization, and processing. This can make it difficult to access and process data from different sources. Additionally, the system may not be able to handle large amounts of data or complex queries. It is also important to note that the system relies on existing public digital archives to retrieve air quality data. If these archives are not up-to-date or do not contain all necessary data, the effectiveness of the system could be limited. The cities available for the analysis belongs to the following countries in Europe [2]:

Andorra	Austria
Belgium	Bosnia
Czech Republic	Denmark
Estonia	Finland
Germany	Spain
Sweden	Switzerland

Table 1: European Countries with dataset of interest

This cities will be evaluated in order to discern the differences in the main pollution indices:

$PM_{2.5}$: Particulate matter 2.5, refers to tiny particles or droplets in the air that are two and one half micron or less in width. Emissions from combustion of gasoline, oil, diesel fuel or wood produce much of the PM2.5 pollution found in outdoor air, as well as significant proportion of PM10.

PM_{10}: Particulate matter 10, refers to tiny particles with a diameter of 10 micron or less .PM10 also includes dust from construction sites, landfills and agriculture, wildfires and brush burning, industrial sources, wind-blown dust from open lands(pollen and fragments of bacteria).

SO_2: Sulfur Dioxide, a heavy, colourless, and poisonous gas composed of sulfur and oxygen. It is produced from the burning of fossil fuels(coal and oil) and the smelting of mineral ores.

NO_2: Nitrogen Dioxide, a gaseous air pollutant composed of nitrogen and oxygen. This gas forms when fossil fuels (coal, oil, gas,diesel) are burned at high temperatures .

CO: Carbon Monoxide, a colorless,odorless gas that can be harmful when inhaled in large amounts. The greatest sources of CO to outdoor air are cars,trucks and other vehicles or machinery that burn fossil fuels.

These pollutants can have negative impacts on human health, as they can be inhaled deep into the lungs and bloodstream and cause respiratory problems as asthma, bronchitis and cardiovascular problems. Long-term exposure can increase the risk of chronic diseases such as lung cancer and heart diseases. Furthermore,they can also affect visilibility, air quality, and the health of surronding environment.

2 Requirements

2.1 Stakeholders

- Common Citizens
- Automotive Production Companies
- Computer Science Engineers

2.2 Actors

In our idea, different actors can access to different levels of our interactive dashboard.

- **Common citizens (CC):** this profile is dedicated to every common person interested in air pollution, who can accede to our application and visualize data and available functions.
- **Mathematical Engineers (MTM):** this profile is dedicated to Mathematical Engineers or professionals such as Data Scientists who can use the dashboard to visualize available data and develop numerical, computational, or statistical models. This user has the ability to use the data for analysis and the development of custom models.
- **Geoinformatics/Environmental Engineers (GEO):** this profile is dedicated Geoinformatics or Environmental Engineers who, in addition to having access to the dashboard, can also download existing data and, most importantly, upload new data. This user has the ability to contribute to the enrichment of available data.

2.3 Domain Assumption

- The user can find the source code on GitHub.
- The ground sensor stations are distributed geographically and collect data on a regular basis.
- The air quality and weather observations collected by the ground sensor stations are stored in databases.
- The data collected by different ground sensor stations may have different formats, structures, and levels of quality.

2.4 Uses cases

Sign up

User: GEO, MTM

Condition: GEO and MTM users should be registered in the app, clicking on "Sign up" button.

Flow of the events: The events follow this order:

1. The GEO or MTM user opens the iAPP4P
2. The GEO or MTM user runs the "Sign up" cell
3. The GEO or MTM user inserts his name and password
4. The GEO or MTM user confirms his username and password by clicking on the "Sign up" button.

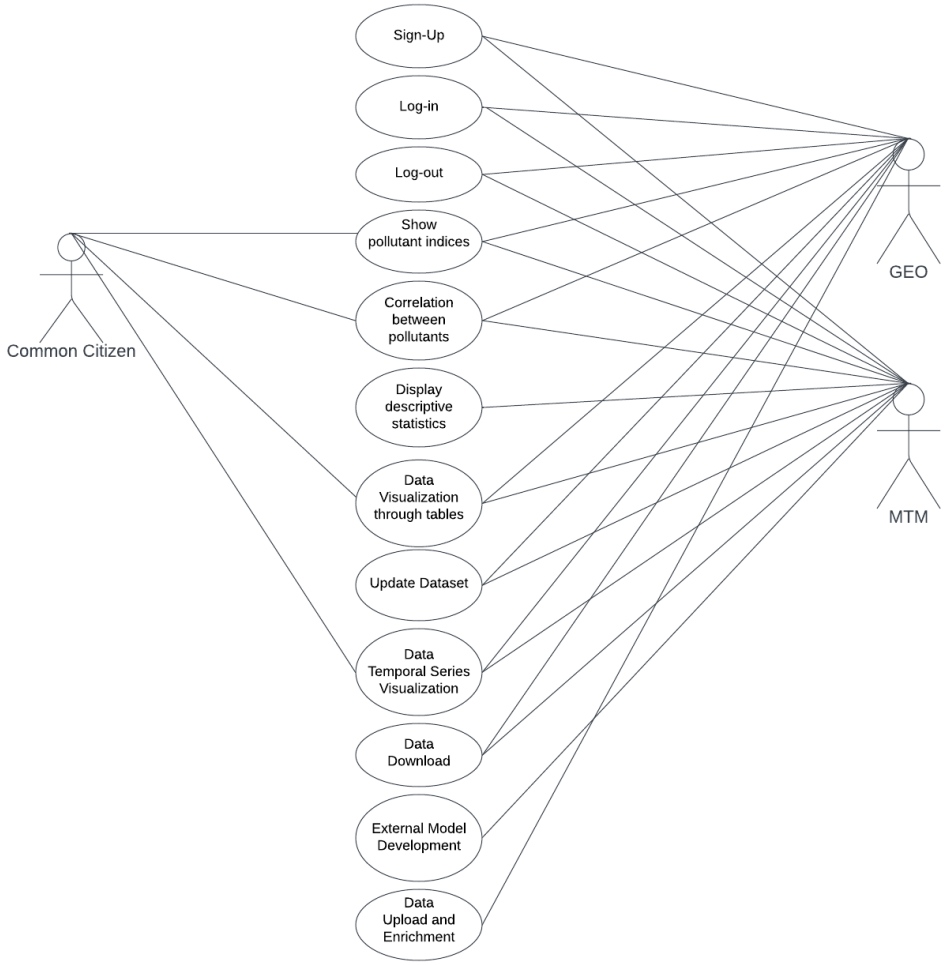

Figure 1: Uses cases diagram

5. After this procedure the user will be registered as a new user in the system and is redirected to the Log In page.

Exceptions: The user's name and password already exist in the database.

User login

User: GEO, MTM

Condition: GEO and MTM users have already sign-up without errors.

Flow of the events: The events follow this order:

1. The GEO or MTM user opens the application through IDE
2. The GEO or MTM user runs the "Log In" cell
3. The GEO or MTM user inserts username and password
4. The GEO or MTM user clicks on the "Log In" button to confirm
5. The GEO or MTM user has successfully login

Exceptions: The user is not registered yet, has put the wrong name or password, or he's already logged.

User logout

User: GEO, MTM

Condition: GEO and MTM have previously successfully logged in.

Flow of the events: The events follow this order:

1. The GEO or MTM user clicks on the "Log Out" button
2. The GEO or MTM has user successfully logout

Exceptions: The user is already logged out or he's not registered yet.

Show desired pollutant indices in all the countries

User: CC, GEO, MTM

Condition: None

Flow of the events: The events follow this order:

1. The user selects a day
2. The user selects a pollutant
3. The user can see the displayed map that reports the mean value of the desired pollutant in all the countries for the selected day

Exceptions: The user has not selected a day or a pollutant.

Correlation between pollutants visualization

User: CC, GEO, MTM

Condition: The data for the selected day should be available

Flow of the events: The events follow this order:

1. The user selects two pollutants
2. The user selects a day of interest on which the data are available
3. The user can see the plot showing the possible correlation between the two selected pollutants

Exceptions: The user has not selected the 2 pollutants or the day

Display of some descriptive statistics

User: MTM

Condition: The user should have registered and logged as MTM user

Flow of the events: The events follow this order:

1. The MTM user selects a city and a pollutant
2. The MTM user will see a window in which the descriptive statistics of min, max, average and the barplot of the selected data are shown
3. The MTM user can interact with the plots to explore and analyze the data

Exceptions: the MTM user has not previously selected the desired location or the desired pollutant indeces

Data visualization through tables

User: CC, GEO, MTM

Condition: None

Flow of the events: The events follow this order:

1. The users run the provided function
2. Some dataframes containing data are showed to the users

Exceptions: None

Update dataset

User: GEO, MTM

Condition: The user should have previously registered as GEO or MTM and logged in.

Flow of the events: The events follow this order:

1. The GEO or MTM user opens the section "Python Script"
2. The GEO or MTM user clicks on the button "Run the code"
3. The GEO or MTM user obtains the download of new updated data

Exceptions: The user has not successfully logged in to the application.

Data temporal series visualization

User: CC, GEO, MTM

Condition: None

Flow of Events: The events follow this order:

1. The user selects the pollutant
2. The user selects the month
3. The system generates a plot of the temporal behaviour of the pollutant for all the locations

Exceptions: The user hasn't selected pollutant or temporal reference .

Data download

User: GEO, MTM

Condition: The user has logged in and accessed the dashboard.

Flow of Events: The events follow this order:

1. The GEO or MTM user selects the "Data Download" option.
2. The GEO or MTM user searches for specific data sets or filters data based on available criteria (e.g. time, location).
3. The system retrieves the requested data files.
4. The GEO or MTM user downloads the selected data files to their local device.

Exceptions: The user is not logged in or there are no data files available for download.

External model development

User: MTM

Condition: The user has successfully logged.

Flow of Events: The events follow this order:

1. The MTM user implements and trains a model.
2. The MTM user specifies the variables and parameters for the model.
3. The system performs numerical, computational or statistical calculations to develop the model.

Exceptions: The user is not logged in or he has not selected all variables or parameters needed for the model development.

Data upload and enrichment

User: GEO

Condition: The user has logged in and accessed the dashboard.

Flow of Events: The events follow this order:

1. The GEO user uploads new data files to the system, providing necessary metadata.
2. The GEO user can verify and contribute to the enrichment of available data.

Exceptions: The user is not logged in or there are issues with the uploaded data or metadata.

2.5 User stories

User sign up

- As a MTM or GEO user I should sign up to the application so that I can access to its functionalities.

User login

- As a MTM or GEO user, I can log in to the application so that I can access air quality data.

User logout

- As a MTM or GEO user, I can log out to the application so that I can close the work session started with the log in.

Show desired pollutant indeces in all the countries

- As a user, I can visualize the indeces of one pollutant so that I can extrapolate information on pollution at a given location.

Correlation between pollutants visualization

- As a user, I can visualize data in graph so that I can see if 2 pollutants have some kind of correlation

Display of some descriptive statistics

- As a MTM user, I can visualize some values that represents my desired descriptive statistics and analyze my data and extrapolate informations.

Data visualization through tables

- As a user, I can visualize data in dataframes so that I can get an overview of the data.

Update dataset

- As a user, I can update dataset so that I can access to most up-to-date data.

Data temporal series visualization

- As a user, I want to visualize available data on the dashboard, so that I can see what is the temporal behaviour of my data.

Model development

- As a Mathematical Engineer, I want to develop numerical, computational, or statistical models using the dashboard, so that I can make predictions and analyze complex systems.

Data download

- As a Geoinformatics/Environmental Engineer, I want to download existing data from the dashboard, So that I can analyze it and use it for my research or projects.

Data upload and enrichment

- As a Geoinformatics/Environmental Engineer, I want to upload new data to the dashboard, So that I can contribute to the enrichment of available data and expand its scope.

2.6 Requirements

The following requirements must be met by the system:

2.6.1 Non-Functional Requirements

- The interface will support English language only.
- The system should be available 24/7 with all service level agreements adhered to. Maintenance should be done during off-peak periods and back ups should be provided.
- The data should be displayed in a manner that is clear and easy to understand even for the common citizen.

2.6.2 Functional Requirements

- The application needs to be able to access and harmonize data from different sources and formats, using standard protocols and tools.
- The application needs to be able to perform basic data processing and analysis tasks, such as data filtering, aggregation, and visualization.
- The application needs to be able to handle large amounts of data, as well as multiple concurrent users.
- The application needs to be secure and protect sensitive data, such as personal information or confidential data.
- The application should easily let the users to log in/out

2.6.3 Technical Requirements

- The main backend implementation language must be Python

2.7 Constraints

- The system must comply with data privacy regulations and standards.
- The system should be developed using open-source software.
- The system must be developed using Python programming language.
- The system must be hosted on a cloud platform.

References

[1] G. E. Quattrocchi, “Se4geo project,” 2023.

[2] “Download of utd air quality data,” 2023-04-05. [Online]. Available: https://discomap.eea.europa.eu/map/fme/AirQualityUTDExport.htm