POLITECNICO
MILANO 1863

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 16, 2023
Professor:	Giovanni E. Quattrocchi

Contents

1 Introduction **2**
1.1 Context and motivations 2
1.2 Definitions, acronyms, abbreviations 2
1.3 Solution overview 3
1.4 Scope and limitations 3

2 Requirements **5**
2.1 Stakeholders 5
2.2 Actors 5
2.3 Domain Assumption 5
2.4 Uses cases 5
2.5 User stories 10
2.6 Requirements 11
2.6.1 Non-Functional Requirements 11
2.6.2 Functional Requirements 12
2.6.3 Technical Requirements 12
2.7 Constraints 12

Bibliography **13**

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 selected data, a web server (backend) to expose a REST API for querying the database, and a dashboard to provide means for requesting, processing (descriptive statistics and forecasting), and visualizing data (e.g. maps, dynamic graphs, etc.).

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].

This tool 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:** refer to online repositories or databases of digital information that are publicly accessible and maintained by governmental or public entities.

1.3 Solution overview

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

- **Data harmonization:** the user has the possibility to manipulate different formats of data.
- **Data preprocessing:** the user has the possibility of querying only interested data area.
- **Data cleaning:** data are shown to the user without outiliers.
- **Descriptive statistics:** the user can obtain mean, mode, median, range, standard deviation, variance, interquartile range of the data of interest.
- **Forecast:** the user can obtain predictions of air pollution in the next hours.

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	Bulgaria	Croatia
Czech Republic	Denmark	Estonia
Finland	France	Germany
Gibraltar	Greece	Hungary
Ireland	Italy	Lithuania
Luxembourg	Malta	Netherlands
Norway	Poland	Portugal
Romania	Serbia	Slovakia
Slovenia	Spain	Sweden
Switzerland	United Kingdom	

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).

O_3: Ozone, an highly reactive gas composed of three oxygen atoms. It occurs both in the Earth's upper atmosphere and at ground level.

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
- Public Authorities
- Computer Science Engineers

2.2 Actors

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

- **Common citizens:** every common person interested in air pollution can accede to our application and visualize data and available functions.
- **Mathematical Engineers:**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 utilize the data for analysis and the development of custom models.
- **Geoinformatics/Environmental Engineers:**This is an Environmental Engineer 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 can register, going on the website, clicking on 'sign-in'

Flow of the events: The events follow this order:

1. A user opens the web-app
2. Clicks on the Log In button
3. User puts in the name and password
4. 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 already exists in the database

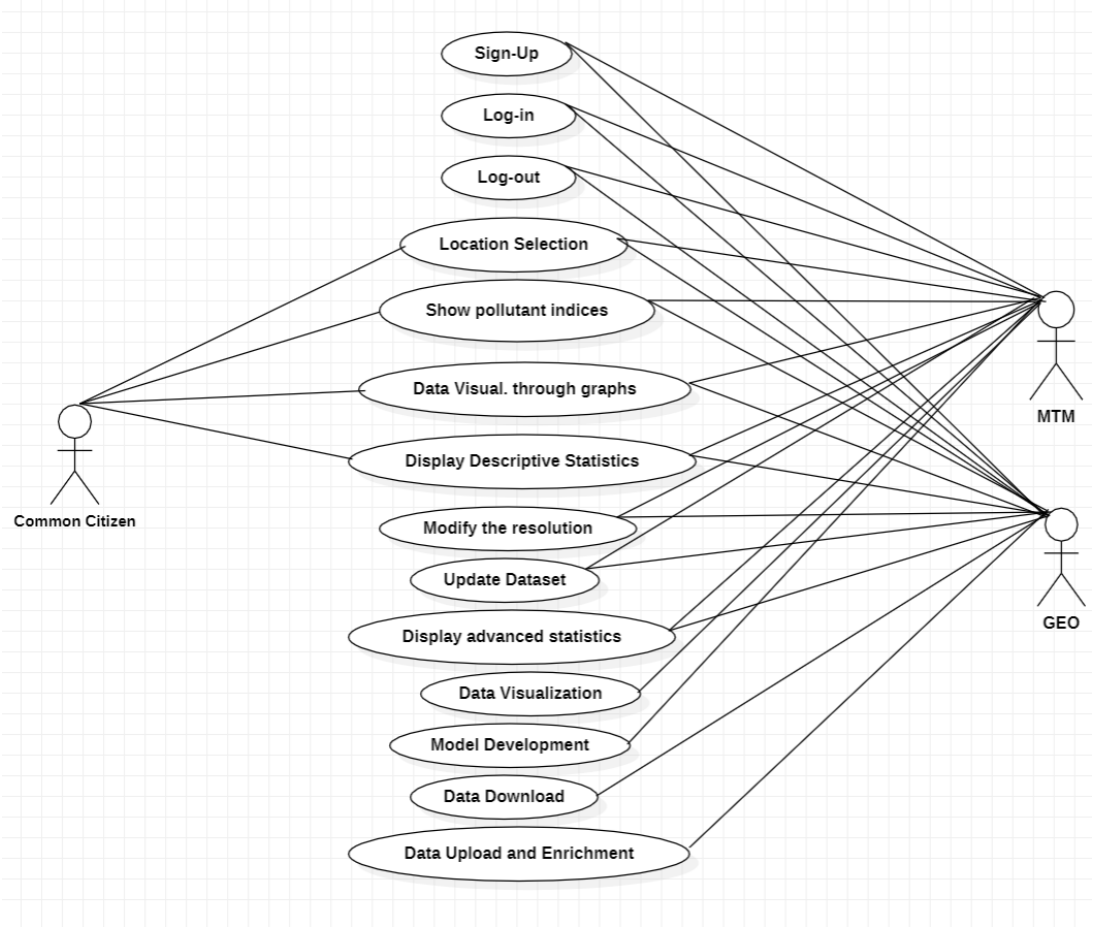

Figure 1: Uses cases diagram

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. A user opens the application through a browser
2. The user clicks on the "Log In" button.
3. User puts hers/his credentials in the “Name” and “Password” fields
4. The user clicks on the “Log In” button
5. The user successfully login and the system automatically redirect the user to the homepage.

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

User logout

User: GEO,MTM

Condition: The user has previously successfully logged in

Flow of the events: The events follow this order:

1. The user clicks on the "Log Out" button
2. The user successfully logout and the system redirects the user to the homepage

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

Location selection

User: GEO,MTM, CC

Condition: The user has previously open the homepage

Flow of the events: The events follow this order:

1. The user clicks on "location" bar
2. The user selects one of the pre-defined locations

Exceptions: none

Show single/multiple pollutant index/indices

User: GEO,MTM, CC

Condition: The user should have selected a location.

Flow of the events: The events follow this order:

1. The user clicks on the button "pollutants indices"
2. The user can select the voice correspondent to the name of the wanted pollutant index/indices

Exceptions: the user has not selected a location.

Data visualization through graphs

User: GEO,MTM,CC

Condition: The user should have previously select a location and the desired pollutant indeces

Flow of the events: The events follow this order:

1. The user clicks on the button "data"
2. The user can select the voice "show graphs" that allows the user to see the plot of the area of interest showing the content level of one or more pollutant indices

Exceptions: the user has not previously selected the desired location and the desidered pollutant indeces

Display of some descriptive statistics

User: GEO,MTM, CC

Condition: The user should have selected the desired pollutant indeces and also the desired location

Flow of the events: The events follow this order:

1. The user clicks on the button "descriptive statistics "
2. The user can select the voice corresepondent to the wanted descriptive statistic: mean, mode, median, interquantile range, min, max, range

3. The user will see a window in which are computed all the wanted descriptive statistics

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

Data visualization through tables

User: GEO,MTM

Condition: The user should have previously logged in

Flow of the events: The events follow this order:

1. The GEO and MTM users clicks on the button "data"
2. The GEO and MTM users should make query to the database to extract the data of its interest (SQL Language)

Exceptions: the user has not successfully log-in into the application, there is some syntax error in the query

High resolution map/graph

User: GEO;MTM

Condition: The user should have previously logged in

Flow of the events: The events follow this order:

1. The GEO and MTM users clicks on the button "resolution"
2. The GEO and MTM users selects on "lower resolution" or "higher resolution" in order to decrement or increment it

Exceptions: the user has not successfully logged-in in the application

Update Dataset

User: GEO,MTM

Condition: The user should have previously logged in

Flow of the events: The events follow this order:

1. The GEO and MTM users opens the section "Python Script"
2. The GEO and MTM users clicks on the button "run the code"
3. The GEO and MTM users obtains the download of new updated data

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

Display of advanced statistics

User: GEO,MTM

Condition: The user should have previously logged in (GEO,MTM), the user should have selected a location and a temporal range(max 6 hours)

Flow of the events: The events follow this order:

1. The user clicks on the button "advanced statistics"
2. The user selects the desidered temporal range (max 6 hour)
3. The user selects on of the available voices (prediction, uncertainty quantification)
4. The user can see the desired statistics in a graphical format

Exceptions: the user has not successfully logged-in in the application (GEO,MTM), the user has not selected a location or the temporal range.

Data Visualization

User: MTM

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

Flow of Events: The events follow this order:

1. The user selects the "Data Visualization" option.
2. The user chooses the desired data set from the available options.
3. The system generates visual representations (charts, graphs, etc.) of the selected data.
4. The user can interact with the visualizations to explore and analyze the data.

Exceptions: The user is not logged in or there is no data available for visualization.

Model Development

User: MTM

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

Flow of Events: The events follow this order:

1. The user selects the "Model Development" option.
2. The user specifies the variables and parameters for the model.
3. The system performs numerical, computational, or statistical calculations to develop the model.
4. The user can validate and refine the model based on the results.

Exceptions: The user is not logged in or there are missing variables or parameters for the model.

Data Download

User: GEO

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

Flow of Events: The events follow this order:

1. The user selects the "Data Download" option.
2. The user searches for specific data sets or filters data based on criteria (e.g., time, location).
3. The system retrieves the requested data files.
4. The 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.

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 user selects the "Data Upload and Enrichment" option.
2. The user uploads new data files to the system, providing necessary metadata.
3. The system processes and integrates the uploaded data with the existing dataset.
4. The 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 user I can sign up to the application so that I can access to its functionalities.

User login

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

User logout

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

Location selection

- As a user, I can select the location of my interest so that I can access to its pollution data.

Show single/multiple pollutant index/indices

- As a user, I can visualize one or more pollutant indices so that I can extrapolate information on pollution at a given location.

Data visualization through graphs

- As a user, I can visualize data in graph so that I can get an immediate view of pollutant indices trends related to a certain location.

Display of some descriptive statistics

- As a user, I can visualize some values that represents my desired descriptive statistics.

Data visualization through tables

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

High resolution map/graph

- As a user, I can visualize graph in a lower resolution so that I can have a general overview with a faster and lighter processing.
- As a user, I can increase the graph resolution so that I can obtain a more detailed result for a higher acceptable quality level.

Update dataset

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

Display of advanced statistics

- As a user, I can visualize some advances statistics (forecasts on a selected temporal range and the related degree of uncertainty) in a graphical format.

Data Visualization

- As a Mathematical Engineer, I want to visualize available data on the dashboard, so that I can analyze and gain insights from the 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 scalable and 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