NEANIAS
Novel EOSC services for Emerging Atmosphere, Underwater and Space Challenges

Deliverable Report

Deliverable: D3.1 Requirements, Specifications & Software Development Plan for Atmospheric Services

30/04/2020

NEANIAS is funded by European Union under Horizon 2020 research and innovation programme via grant agreement No. 863448

Document Info

Project Information			
Acronym	NEANIAS		
Name	Novel EOSC Services for Emerging Atmosphere, Underwater & Space Challenges		
Start Date	1 Nov 2019	End Date	31 Oct 2022
Program	H2020-EU.1.4.1.3. - Development, deployment and operation of ICT-based e-infrastructures		
Call ID	H2020-INFRAEOSC-2018-2020	Topic	H2020-INFRAEOSC-2019-1
Grant No	863448	Instrument	RIA
Document Information			
Deliverable No	D3.1		
Deliverable Title	Requirements, Specifications & Software Development Plan for Atmospheric Services		
Due Date	30-Apr-2020	Delivery Date	08-May-2020
Lead Beneficiary	UBIWHERE (6)		
Beneficiaries (part.)	NOA (20); MEEO (17); UNIMIB (11); ATHENA (2); NKUA (1);		
Editor(s)	Ricardo Vitorino (Ubiwhere); Konstantinos Karantzalos (NKUA); Spyridon Rapsomanikis (Athena)		
Authors (s)	Ricardo Vitorino, Francisco Cardoso (Ubiwhere); Konstantinos Karantzalos (NKUA); Spyridon Rapsomanikis (Athena)		
Contributor (s)	Diogo Oliveira, João Garcia, Carlos Cortinhas (Ubiwhere); Filippos Bodare (ATHENA); Olyna Olympia Gounari (NKUA); Alessandro Tibaldi (UNIMIB)		
Reviewer(s)	Ugo Becciani, Sergio Molinari (INAF)		
Workpackage No	WP3		
Version	V.1.0	Stage	Final
Version details	Revision: 31 . Last save: 2020-05-07 , 15:33 Pages: 46 . Characters: 64.846		
Distribution	Public	Type	Report
Keywords	EOSC; Services; Cloud Computing; Atmospheric Research;		

Change Record

Version	Date	Description	Editor(s)
V1.0	8/5/2020	Document version submitted to EC	R. Vitorino

Disclaimer

NEANIAS is a Research and Innovation Action funded by European Union under Horizon 2020 research and innovation programme, via grant agreement No. 863448.

NEANIAS is project that comprehensively addresses the 'Prototyping New Innovative Services' challenge set out in the 'Roadmap for EOSC' foreseen actions. It drives the co-design, delivery, and integration into EOSC of innovative thematic services, derived from state-of-the-art research assets and practices in three major sectors: underwater research, atmospheric research and space research. In each sector it engages a diverse set of research and business groups, practices, and technologies and will not only address its community-specific needs but will also enable the transition of the respective community to the EOSC concept and Open Science principles. NEANIAS provides its communities with plentiful resource access, collaboration instruments, and interdisciplinary research mechanisms, which will amplify and broaden each community's research and knowledge generation activities. NEANIAS delivers a rich set of services, designed to be flexible and extensible, able to accommodate the needs of communities beyond their original definition and to adapt to neighbouring cases, fostering reproducibility and re-usability. NEANIAS identifies promising, cutting-edge business cases across several user communities and lays out several concrete exploitation opportunities.

The European Union (EU) was established in accordance with the Treaty on the European Union (Maastricht). There are currently 27-member states of the European Union. It is based on the European Communities and the member states' cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice, and the Court of Auditors (http://europa.eu.int/).

Table of Contents

Tables of Figures & Tables

Document Figures

Document Tables

Abstract

The objective of Deliverable D3.1 is to report on requirements, specifications and software development plan of the Atmospheric Research Services. The present document describes the different end-user communities working within the Atmospheric theme and their research needs, the datasets and products used and needed by these communities, as well as the current status of employed software and services. Then it describes a mapping of requirements to the technical specifications and finally, the foreseen software development plan.

1. Introduction

1.1. Context

In the framework of WP3, NEANIAS will deliver cloud-based services that target specific challenges of the atmospheric thematic sector. WP3 embraces co-design under an agile methodology that spans all project activities, where service consumers and providers are engaged in a cycle of requirements, design, implementation/verification, evaluation and feedback. WP3 activities are amplified by engagement and onboarding activities, ensuring the potential of sustainability from the perspective of adoption and uptake. In particular, bringing together experts and know-how from Open Science, EOSC and Research Infrastructures, such as National and Kapodistrian University of Athens, Università Degli Studi di Milano Bicocca and Athena Research Centre, just to name a few, NEANIAS seeks not only to satisfy the needs of the atmospheric sector communities, boarding the cloud era, but also to identify and elevate opportunities that exploit the plentiful resources (data, CPU, storage, services) at reach, weaving innovative value chains and innovative ways to conduct e-science.

In such a context, the first Task of WP3 *i.e., 'T3.1 atmospheric research sector user requirements, service co-design and gap analysis'* has analysed information, data and requirements collected from the Atmospheric-related application domains and user communities to form the first deliverable of WP3, the present document: '*D3.1 atmospheric Research Services Report on requirements, Specifications & Software development Plan*'.

1.2. Contents and Rationale

The scope of Deliverable D3.1 is to document the preliminary efforts undertaken within the context of T3.1, containing, in particular, the requirements, specifications and software development plans for the Atmospheric Research Services.

In particular, to properly collect requirements (categorised below), the atmospheric end-user communities, the involved datasets and products and the current status of employed software and services are presented. A mapping of requirements to technical specifications has been performed and finally an adequate software development plan has been prepared.

1.3. Structure of the document

The deliverable is structured in five sections (chapters), being the first one responsible for introducing the report, briefly mentioning its content and how it is structured. Chapter 2 describes the Atmospheric end-user communities (section 2.1), their datasets and products (section 2.2), the current state of the software and services employed (section 2.3) and the requirements (section 2.4), while Chapter 3 details their co-design and specifications, including a gap analysis. Chapter 4 presents the software development plan and guidance, including the presentation of the preliminary architecture of the Atmospheric Services and the datasets to validate the services. Finally, Chapter 5 outlines the conclusions and future steps.

2. User Requirements for the Atmospheric Services

This chapter aims at defining the user requirements of the project. To do so, first, the target communities will be identified (who they are, their needs and how NEANIAS can help them), and then current existing products and datasets will be listed and described, which will help in identifying tools and sources that can be helpful in designing, implementing and testing NEANIAS. Finally, current software and services will be studied to understand which tools can help in the development of the project. With this, it will be possible to define the user requirements of the project.

2.1. Atmospheric End-user Communities

In this section, target end-user communities are identified, where it is described who they are, what they need (in the form of user stories) and how NEANIAS can help achieve such purpose. This is an important step in understanding what to do along the project, since the developed product must be useful and valuable to its end-users. The end-user communities for the Atmospheric Services are listed below, being analysed in more detail afterwards.

End-user Community	Brief Description
Urban air quality authorities	Monitor the air quality in cities or urban environments to track the performance of environmental policies and ensure the quality of life of citizens, interested in calculating the energy and anthropogenic energy fluxes as well as greenhouse gases (GHG) emissions fluxes.
Geologists	Study the solid, liquid, and gaseous matter that constitutes the Earth, interested in tracking and calculating gas fluxes from active faults, fractures and volcanic vents in relation to tectonic stresses.
Geohazards, civil protection	Focused on the protection of citizens from natural disasters, and using the principles of emergency operations (prevention, mitigation, preparation, response, or emergency evacuation and recovery), this community is interested on the assessment of dangerous areas regarding the air quality and gases emissions.
Meteorological services	Providing weather forecasts, warnings of hazardous weather, and other weather-related products to organisations and the public for the purposes of protection, safety, and general information, these are interested in calculating GHG emissions fluxes, to validate national inventories.
Energy and power sector and industrial air pollutant emitters	With the business of producing energy or resources through industry facilities, these communities need to calculate GHG emissions, specific pollutant fluxes as well as energy fluxes, to ensure compliance with regulations and benchmark their environmental and energy performance.
Ecologists	Study the interactions of organisms and their biophysical environments, by analysing their abundance, distribution and biodiversity, monitoring the ecological relationships and how they regulate the flux of energy, nutrients, and climate all the way up to the planetary scale. This community is interested in tracking and analysing the dynamic history of the planetary atmosphere's CO_2 and O_2 composition and how it has been affected by the biogenic flux of gases coming from respiration and photosynthesis, in relation to the ecology and evolution of plants and animals.
Rural/urban planners	Organising and monitoring socio-economic activities and initiatives to manage the spatial organisation of regions for efficient allocation of land use and improve standard of living, either in cities or rural areas. Depending on how these planning outcomes are applied, they can improve air quality in the long run, by strategic location of polluting sources and exposed population, and encouraging a structure that would minimise

	pollution emissions, hence the interest in tracking the GHG emissions fluxes, as well as energy fluxes.
Insurance, health	With the purpose of ensuring their customers or patients safest conditions, where the level of pollutants prescribed by regulations must not be exceeded during a given time in a defined area, these communities need to access atmospheric data about the GHG and pollutants emissions to validate incidents and cross-check with policies.

Table 1 – End-user communities for Atmospheric Services

2.1.1. Urban air quality authorities

Urban air quality authorities seek to better understand human activity's effect on air quality and the environment. Since anthropogenic energy fluxes and GHG emission fluxes have a big impact on several sectors like agriculture and health, it is essential to have the proper tools for monitoring greenhouse gas flux density as well as for air quality estimation, monitoring, and forecasting. To achieve that, it is necessary to automate and validate existing software modules for these two main ambitions.
The first one, monitoring greenhouse gas flux density is achievable using data obtained from specifically set meteorological stations, integrated as a novel cutting-edge service into EOSC. For that, an orchestrator will be built to select the optimal method or combination of methods, for processing, depending on the dataset, that can be uploaded by the user.
The second one, air quality estimation, monitoring and forecasting, is achieved with the novel integration of satellite, geospatial and in-situ data and a management platform implemented as a Software-as-a-Service (SaaS) platform. With web responsive layouts and REST interfaces for integration with external systems, the interfaces of the platform will be configurable, being able to give daily, weekly and monthly reports of all the data collected by the sensing stations.
With the previous objectives completed, it becomes possible for urban authorities not only to improve their analysis on the existent datasets and to achieve better results, but also to build and improve forecast and monitoring services.

2.1.2. Geologists, Geohazards and civil protection stakeholders

The geologists' community working on geohazards needs to better understand the relationships between different phenomena that can affect the civil society. A great concern is represented by gases released in volcanic areas (including radon, SO_2, CO_2) that have serious health negative effects, as well as by ashes produced by volcanic eruptions. The latter also has an impact on health, along with important problems to air traffic and other infrastructures.
Particularly, Radon emission depends on a series of other phenomena such as atmospheric pressure, temperature variation, and changes of stresses acting on the fractures from which gas is outpoured. The geological community has asked NEANIAS to release services capable of cross-correlating the time series of data regarding gas emission, atmospheric conditions and earthquake occurrence, possibly to be also applied to CO_2 and SO_2 series from multiscale database (e.g. spanning from satellite data to field-collected ones). The latter process is composed of several variables that comprehend the surface distance between the gas emission point and the earthquake epicentre location, the distance between the gas emission point and the depth of the earthquake hypocentre, the tectonic stress characteristics (fault kinematics trajectories of the principal stresses obtainable from the earthquake focal mechanism solution), and the Magnitude of the seismic event, and the spatial resolution of data. It is thus necessary to develop a service capable of cross-correlating all these parameters in order to distinguish which is the dominant process in dictating the variations of gas radon emission, and possibly also to CO_2 and SO_2.
The emission of gases and ashes from the summit crater of an active volcano such as Etna or Stromboli, are dependent upon the endogenous dynamic of the magma into the volcanic conduit, thus mainly related to the

pressure dependent on the three-phases fluid system, and upon the stress state of the host rock surrounding the conduit, which is to say the volcanic edifice. Earthquake occurrence can change the local stress state favouring gas flux by stress unclamping. The geological community is asking to NEANIAS to release services capable of cross-correlating the time series of data regarding gas emission, earthquake occurrence, and stress changes. Again, earthquake occurrence means that the service must be capable of correlating the distance of the earthquake epicentre and hypocentre location from the crater, the tectonic stress characteristics (fault kinematics trajectories of the principal stresses obtainable from the earthquake focal mechanism solution), and the Magnitude of the seismic event.
The software can be released in three phases addressing incrementally more complex datasets and scenarios towards analysing efficiently the variation in gas concentrations before during and after a seismic event. The software will be tested against the detection of changes in ash and gas concentration at regions with active volcanoes (e.g. Etna and Stromboli in Italy, Santorini in Greece), which constitutes a targeted application of significant importance for assessing geohazards and risks.
The geological community also needs the implementation of a service capable of calculating and visualize the stress trajectories in plan and 3D view. A basic C implemented software called "Lissage" program from Lee and Angelier (1994) can calculate stress trajectory in 2D (maps) but has several limitations, it is not user-friendly, does not consider different data sources (e.g. field data, focal mechanisms, earthquakes) and, finally, it does not provide a 3D or an automatic 2D-layered output compatible with common GIS workspace (e.g. ArcMap, QGIS, Google Earth) and Matlab/Excel.
The existing software comes from: [10] Lee, J. C., & Angelier, J. (1994). Paleostress trajectory maps based on the results of local determinations: the "Lissage" program. Computers & Geosciences, 20(2), 161-191.

2.1.3. Meteorological services

It is usually asked from national meteorological services to obtain, calculate GHG emissions according to the guidelines given by the EC directive 2003/87/EC, as amended by the 2007/589/EC. Not only can it be a monotonous job, when performed manually, literature has proven that the obtained data are inaccurate. On the other hand, flux determination by measurements are also expensive and not practical for a whole country. There are two ways of bypassing this impasse: the first is to separate these fluxes in economic sectors and legislate the requirement to obtain these fluxes from them, and the second is to validate fluxes calculated from space satellites and improve their algorithms. These approaches will be followed to provide a service that provides weather services with GHG emissions calculations for proper monitoring and evaluation.

2.1.4. Energy and Power sector and industrial emitters

Once the requirement to obtain GHGs fluxes from the sectors Energy, Power and other Industrial emitters has been legislated, their requirements to submit GHGs emissions will emerge. This requirement will not be based on inventories of material used and accompanying algorithms but on "point" and "area" source measurements, therefore the atmospheric services shall be able to receive as input or query parameters the location (a tuple of coordinates, a street name, a polygon or area or other geography type of data), besides the time intervals, so industry operators become able to benchmark their services and perform audits and confirm they comply with the regulation and policies.

2.1.5. Ecologists

Our services described from 2.1.1 to 2.1.4 will help them to establish the validity of the above sector's presentation of results, i.e. leveraging gases fluxes monitoring to evaluate the atmosphere composition and compare with the ecosystem's evolution.

2.1.6. Rural/urban planners

Planners of urban or rural areas need a reliable service that provides them with sustainable indicators of the environment such as Fine particulate (PM2.5) and particulate (PM10) concentration, as well as greenhouse gas emissions measured in tonnes per capita, so as to correlate with the demographic information available. Moreover, and especially considering the pandemic faced from late 2019 to 2020, due to COVID-19, it is important to monitor the impact of these particulate and spread of respiratory pandemics, as some studies alert to. Therefore, the Atmospheric services must be able to provide standard Key Performance Indicators about these conditions for a certain area and period of time.

2.1.7. Insurance and health

Similar to 2.1.6, the services will allow insurance or health communities to validate incidents and cross-check with policies, therefore, the services shall support thresholds as query parameters to request all data of a certain area that was under or above a given threshold.

2.2. Relevant Atmospheric Datasets and Products

In this section, various atmospheric datasets and products are identified, which will be helpful to understand what is available that can be useful within the project, be it data that can be used to test the project or products that can be either used to provide some sort of service to the project, or serve as a guideline to define some feature.

2.2.1. Data and Products for Urban air quality authorities

The city of Porto makes air quality data openly available through their urban data platform's portal[1], where the data is free and open, following FIWARE data models and open standards. It uses Monitar's Air Quality and Weather stations[2], which follow the directive 2008/50/EC. The air quality stations monitor air quality gaseous parameters with constant flow sampling and particulate filtration and can also integrate other air quality sensors. The communication is done using the 3G/4G network or Wi-Fi networks, when available.

Figure 1 – Monitar Air Quality Monitoring Station (left) and Weather station (right)

The following tables represent some of the solution's main characteristics regarding the parameters that the air quality stations and the weather stations can measure.

[1] http://history-data.urbanplatform.portodigital.pt/v2/ui/#/queries
[2] https://www.monitar.pt/

Parameter	Range
O3	10 – 1000 ppb
NO2	0.05 - 0.5 ppm
CO	0 – 1000 ppm
CO2	0 – 5000 ppm
PM1	0 –100 $\mu g/m^3$
PM2.5	0 – 100 $\mu g/m^3$
Volatile organic compounds (VOCs)	1 – 1000 ppm

Table 2 – Monitar's air quality parameters

Parameter	Range	Accuracy
Temperature	-40 – 125 ºC	±0.3 ºC
Relative humidity	0 – 100 %	±2 %
Wind direction	0 – 360 º	± 3º±5%
Wind speed	0 – 89 m/s	0.5 m/s or ±5%
Rain fall	0 – 999.8 mm	± 0.2 mm or ±4 %
Solar radiation	0 – 1800 W/m^2	± 5%
UV radiation	0 – 16 index	± 5 %

Table 3 – Monitar's weather parameters

Regarding the products for Urban air quality authorities, the following software solutions are available:

2.2.1.1. Urban Platform

The Urban Platform[3] consists of a multitenant software platform capable of collecting, storing, analysing and publishing (as open data, when needed) multi-domain information, provisioned by different sources such as sensor networks, Open Data catalogues as well as information systems and mobile applications.

Figure 2 – Urban Platform real-time map

The Urban Platform is a solution that provides in real-time a holistic view of a city or region, with an interactive map displaying georeferenced information about the different verticals/use cases (traffic, parking, public transports, air quality, energy, street lighting, waste management, water leakage, and flow level, electric charging, energy, street lighting, waste management, water leakage and flow level, electric charging,

[3] Urban Platform - https://urbanplatform.city

and noise level), plus the capability of focusing on or zooming in specific pins and filtering the data per vertical solution or region in an always-visible element, offering an easy way for users to search for the relevant information. All information is represented in a rich way with multiple criteria such as colour, dimension and icon types, providing an easy way to understand the overall status of the city at any given moment.

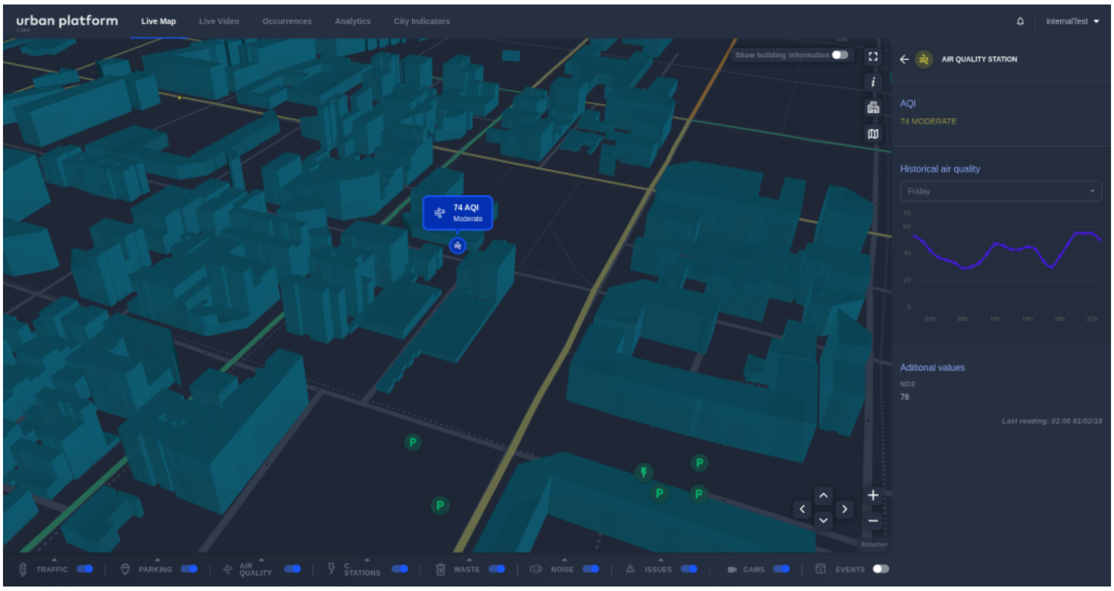

Figure 3 – Urban Platform Air Quality analysis from an in-situ monitoring station

Despite all the domains that the Urban Platform can accommodate, the most relevant for this case is the Weather and Air Quality. The two images (the one above and the other below) show two scenarios for different locations where it is possible to see a detailed view of the air quality measured by a station and the weather parameters measured.

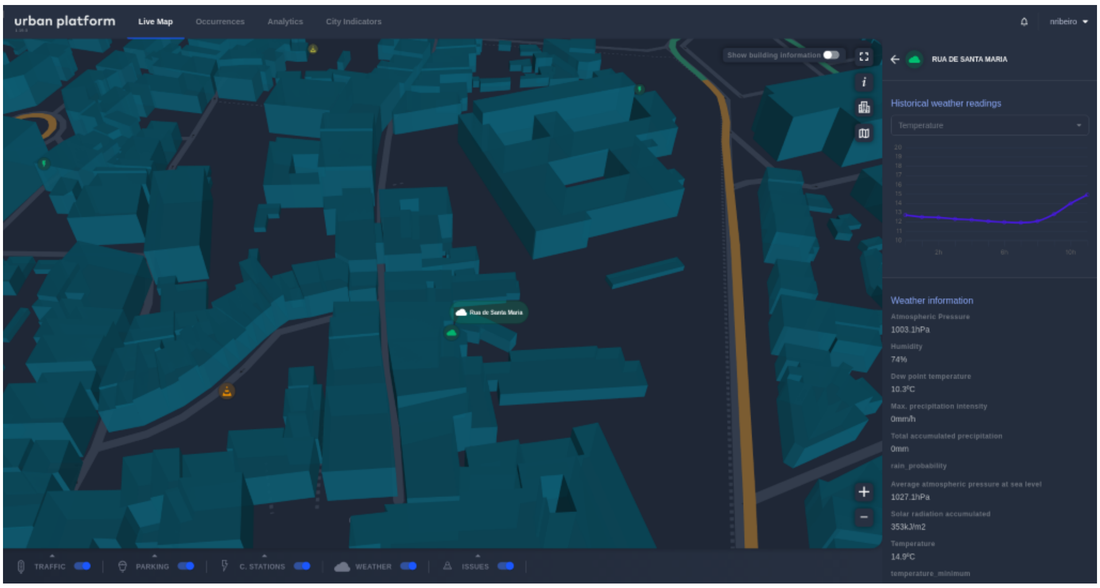

Figure 4 – Urban Platform Weather analysis from an in-situ monitoring station

Besides the visualisation of real-time and historical information, the Urban Platform also calculates some insights, as it is possible to see in the following image, by correlating information such as traffic congestion with air quality index for a specific area or region, but also to have some insights based on historical data, for example, the noise level observations per hour and per weekday.

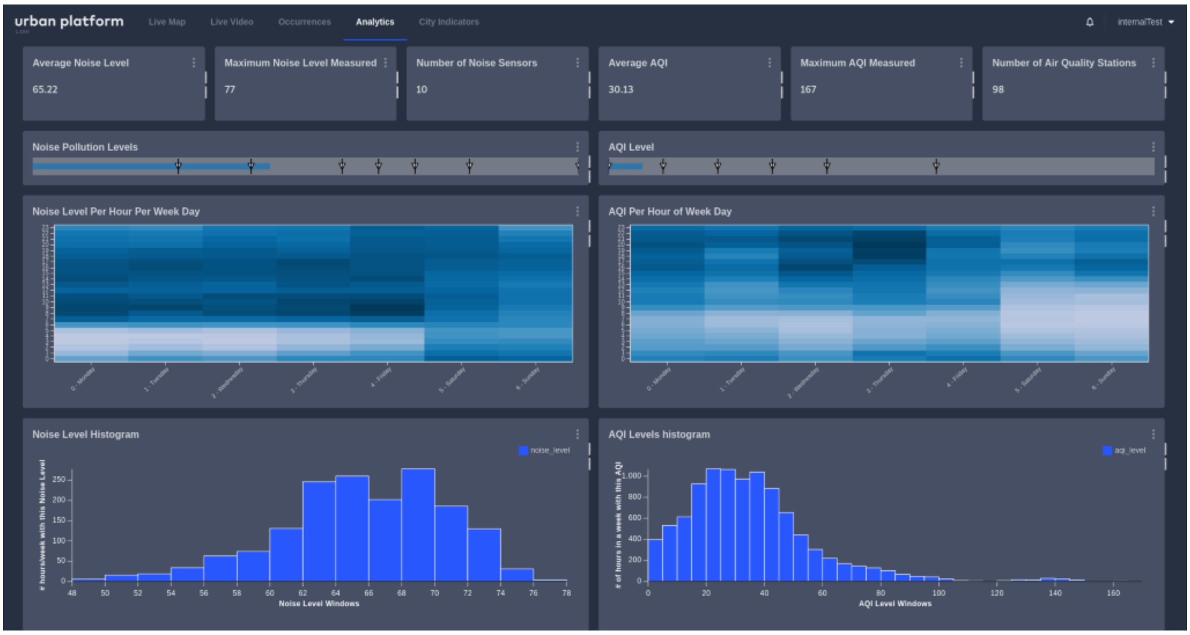

Figure 5 – Urban Platform Air Quality analysis with customizable dashboard and charts

2.2.1.2. Smart Air Quality

Smart Air Quality is a cloud solution for smart cities that focuses on air quality monitoring, which allows public authorities to have a detailed view of all the city air quality stations and allows to take some insights into the overall air quality. It collects and stores the data provided by air quality stations (IoT devices) that communicate in real-time through wireless networks. The solution then makes the data available through open and interoperable REST APIs (so they can be shared with the community) and responsive user interfaces that adapt to the device's screen dimensions, to be used by expert staff from the smart city.

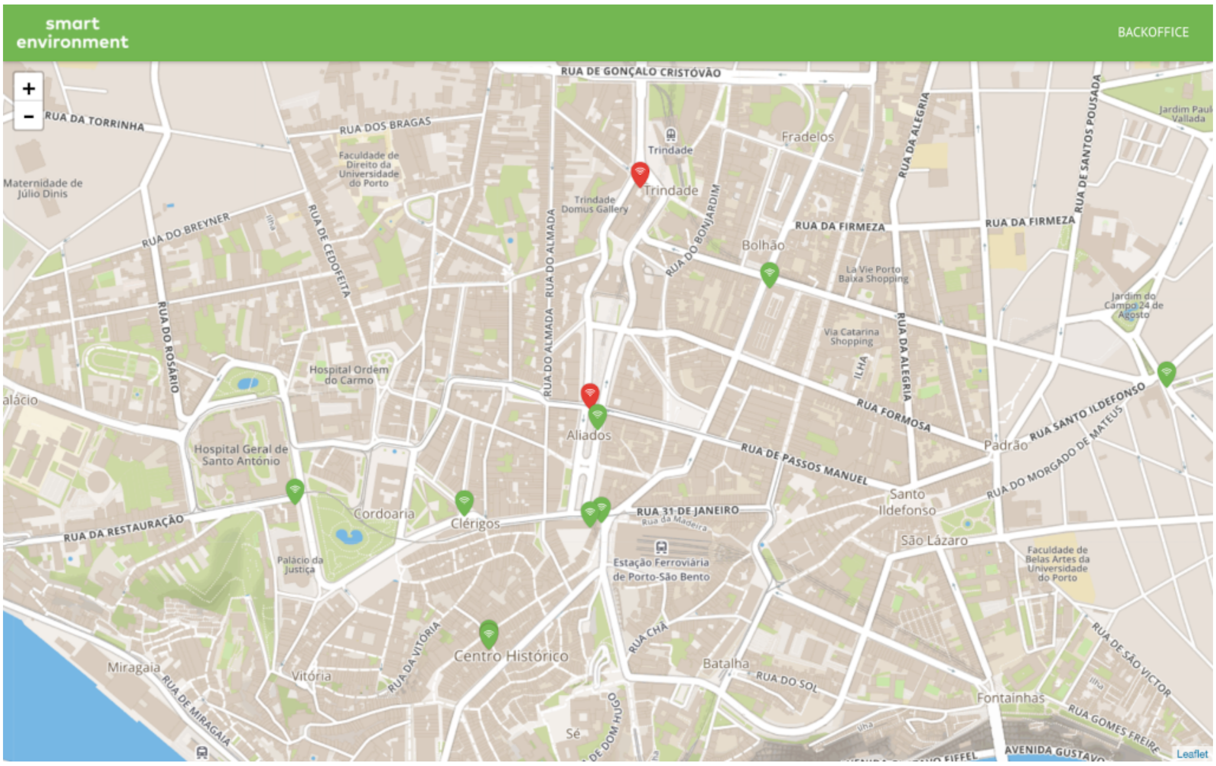

Figure 6 – Smart Air Quality real-time map

Besides allowing the end-users to identify and access air quality measurements from individual stations, it provides a system for asset and event management, to control and manage each station individually and the data collected by it.

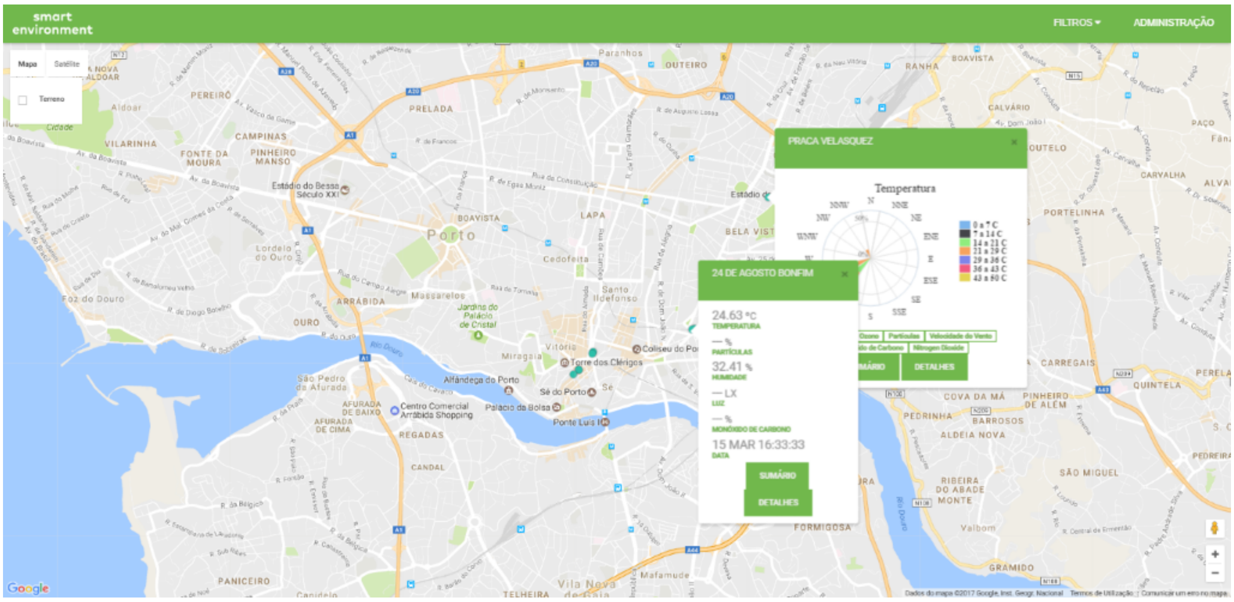

Figure 7 – Smart Air Quality individual analysis from an in-situ monitoring station

Besides this, it provides operational features where administrators can see the operational status of devices and dispatch teams to fix the ones not working properly or technicians to integrate new ones.

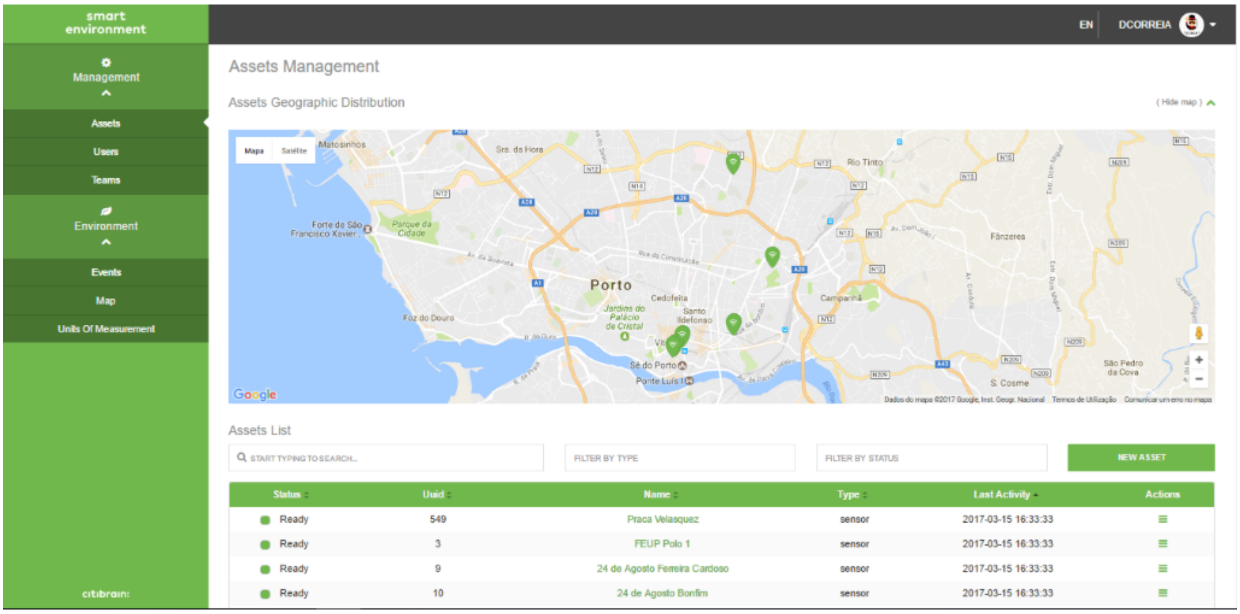

Figure 8 – Smart Air Quality asset management (stations and teams).

2.2.1.3. Dataset for Urban air quality authorities

The following table describes the datasets that urban air quality authorities shall get access to with the thematic service A3:

DATASETS – A3			
id	Name of Dataset	Type	Source
1	Porto Air Quality	REST API for real-time and historical data on air quality monitored from 5 stations (O3, NO2, CO, CO2, PM10, PM2.5, Volatile organic compounds)	Porto Urban Data Platform (Monitar)
2	Porto Weather	REST API for real-time and historical data on weather monitored from 7 stations (Temperature, Relative humidity, Wind direction, Wind speed, Rain fall, Solar radiation, UV radiation)	Porto Urban Data Platform (Monitar)
3	Porto Weather forecast	REST API for weather forecast for a given day	IPMA & OpenWeather
4	Porto Air Quality Index	REST API for air quality index calculation	AQICN
5	Satellite Air Quality	REST API providing: Ozone (O3), Nitrogen Dioxide (NO2), Sulphur Dioxide (SO2, Carbon Monoxide (CO), Methane (CH4), Formaldehyde (HCHO), UV Aerosol Index	MEEO – Sentinel-5P Data

Table 4 – Dataset for Urban air quality authorities

While Porto's Air Quality and Weather data has already been described before, the two other datasets are described in a later section. The dataset from MEEO was defined in a joint effort with WP4 and WP6.

2.2.2. Data and Products for Geologists

The following table describes the dataset required for geologists for thematic service A2:

DATASETS – A2			
id	Name of Dataset	Type	Source
1	Gas radon from Mt Etna faults	Excel table of values of gas radon emitted from the faults of Etna volcano	National Institute of Geophysics and Volcanology, Italy
2	Gas and ash plumes from Etna crater	Excel table of values of various gases and ashes emitted from the summit vents of Etna	National Institute of Geophysics and Volcanology, Italy
3	Gas from Nea Kameni volcano, Greece	Excel table of values of various gases emitted from the fractures on Nea kameni volcano, Santorini	Surveys by researchers of various universities
4	Map of all Etna faults	Database containing the georeferenced data of all faults on Mt Etna	National Institute of Geophysics and Volcanology, Italy
5	Seismicity of Mt Etna	Database containing all foci and M of earthquakes that occurred at Etna in the last tens of years	National Institute of Geophysics and Volcanology, Italy
6	Meteorological data of each site	Excel tables of meteorological data	
7	Thermal anomalies at selected volcanoes	CSV table	MODIS/MODVOLC
8	SO2 columns from satellites at various volcanoes	CSV table	Global Volcanism Program

Table 5 – Dataset for Geologists

The above listed datasets represent information about the two main components of processes that are related to the geological needs represented in the NEANIAS project: the atmospheric component and the solid Earth component. Regarding the atmospheric component, the team has already collected relevant data series on the gas emission from various volcanoes, namely Mt Etna in Italy and Nea Kameni in Santorini, Greece. The data series are more complete for Mt Etna having been collected in the last years each 20 minutes by installed instruments that were continuously monitoring the gas radon flux. On Nea kameni the measurements of gases (mainly CO2) have been done instead manually with a consequent more limited in time data collection. Other atmospheric datasets comprehend the meteorological data, mainly pressure and temperature. Finally, we collected data on thermal anomalies and SO2 columns from satellites at some selected volcanoes. Regarding solid Earth, the gas data will be compared with other datasets represented by the distribution of faults on Mt Etna and of fractures in the smaller Nea Kameni island, and with the datasets on the distribution of seismicity in the same time frame. The aforementioned datasets will serve the need of the following potential users and business cases (associate datasets and results with real world examples): Geohazard organizations and consultancies, civil protection, geologists, civil aviation agencies.

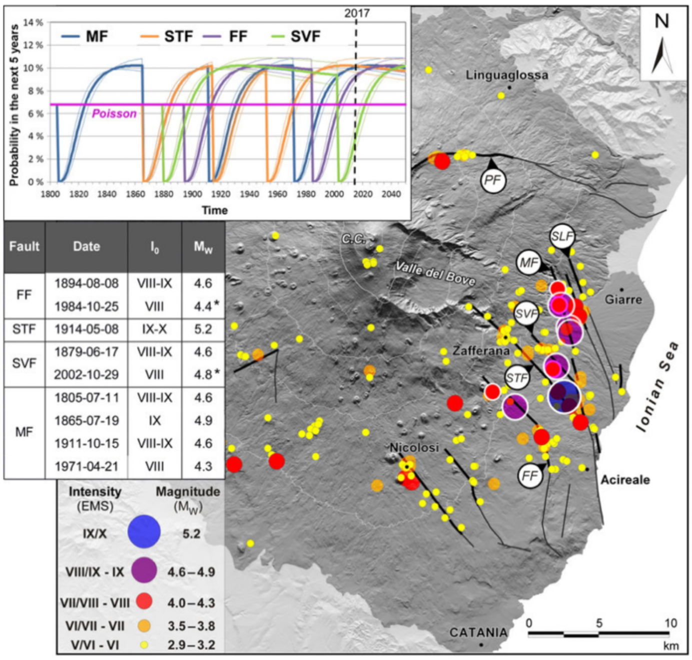

Figure 9 – Main active faults of Mt Etna and main historic seismic events.

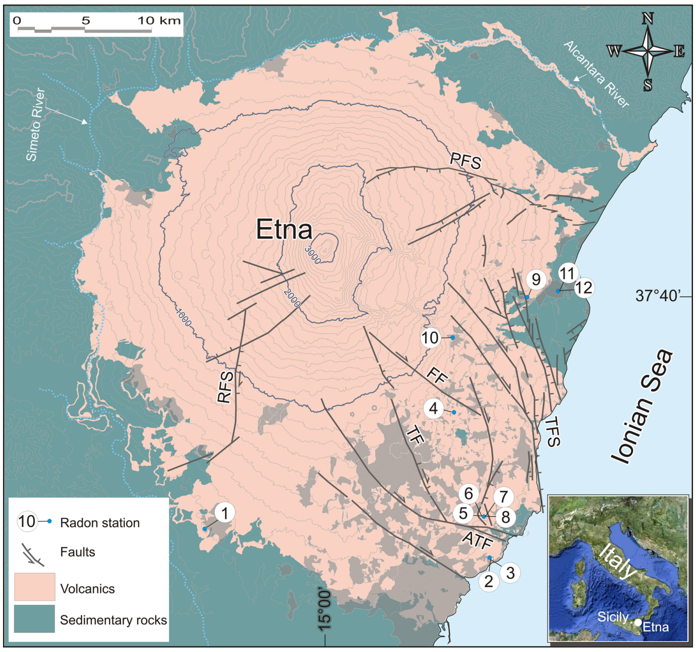

Figure 10 – Simplified geological map of Mt Etna with main active faults and the sites of indoor radon measurements. The urbanized areas are shown in light gray. PFS, Pernicana fault system; RFS, Ragalna fault system; TFS, Timpe fault system; FF, Fiandaca fault; TF, Trecastagni fault; ATF, Aci Trezza fault.

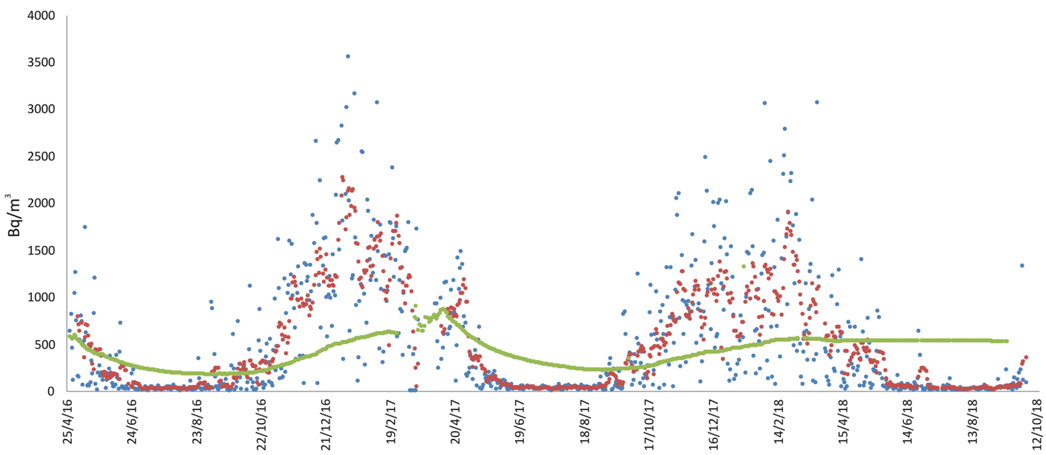

Figure 11 – 2016-2018 Radon measurements recorded at Mt Etna in the Giarre municipality. Radon concentration are presented as daily (blue line), weekly (red), and long-term (green) averages. No data interruption occurred during the acquisition period.

2.2.3. Data and Products for Meteorological services

For developing and validating the Atmospheric Service (A1), we will employ the following datasets:

DATASETS – A1			
id	**Name of Dataset**	**Type**	**Source**
1	ATHENS THERMOPOLIS	Eddy covariance	ENTA

2	ATHENS THERMOPOLIS	AIRCRAFT- three level VARIANCES METHOD	ENTA
3	VOA-Gr STATION	Eddy covariance	ENTA and at www.unitus.it
4	VOA-Gr STATION	Gradient method at 4 heights	ENTA
5	ATHENS THERMOPOLIS	Gradient method at 5 different heights	ENTA
6	Copernicus DATABASE	various	ECMWF

Table 6 – Dataset for Meteorological Services

The data obtained in the "Athens Thermopolis" and financed by the ESA, will be made available so that users can try our algorithms to obtain fluxes of GHGs, energy and momentum above a city. From our publications one can also obtain an algorithm that indicates anthropogenic fluxes of energy, in an urban environment. The use of tried algorithms for the calculation of GHGs, energy and momentum from aircraft flights will also be supported as a service to users. Eddy Covariance and gradient method from met towers algorithms will also be provided. The Copernicus database, refers also to fluxes but data are sporadic and of unknown validity. Below are some filtered, processed and explained flux data from our Climate Watch Station VOAGr in northern Greece. This station is part of and provides data to the European Fluxes Database Cluster (http://www.europe-fluxdata.eu/) with data deposited at "database@unitus.it".

2.2.3.1. Flux densities examples from VOAGr

Spectral flux density is the quantity that describes the rate at which energy is transferred by electromagnetic radiation through a (real or virtual) surface, per surface area and per wavelength. The Climate Watch Station VOAGr (installed in northern Greece) shown in the picture below provides flux densities data that will be used in NEANIAS, which need to be processed through calculation methods.

Figure 12 – VOAGr MET Tower.

Below one can find in graphical form flux densities from our VOAGr MET TOWER using different flux densities calculating methods. The most commonly used method is Eddy Covariance since it requires the vertical velocity datum covarying with the concentration of a scalar. Hence flux densities calculations are relatively straightforward. However, these parameters have to scan a number of frequencies, a fact that requires fast response sensors, not always available. The aerodynamic gradient method and its derivatives is based on the fact that scalars vary with height according to the prevailing atmospheric stability. Determination of horizontal wind velocities and "slow" determinations of scalars suffice to calculate fluxes, based for example on the "Monin-Obukhov" similarity theory.

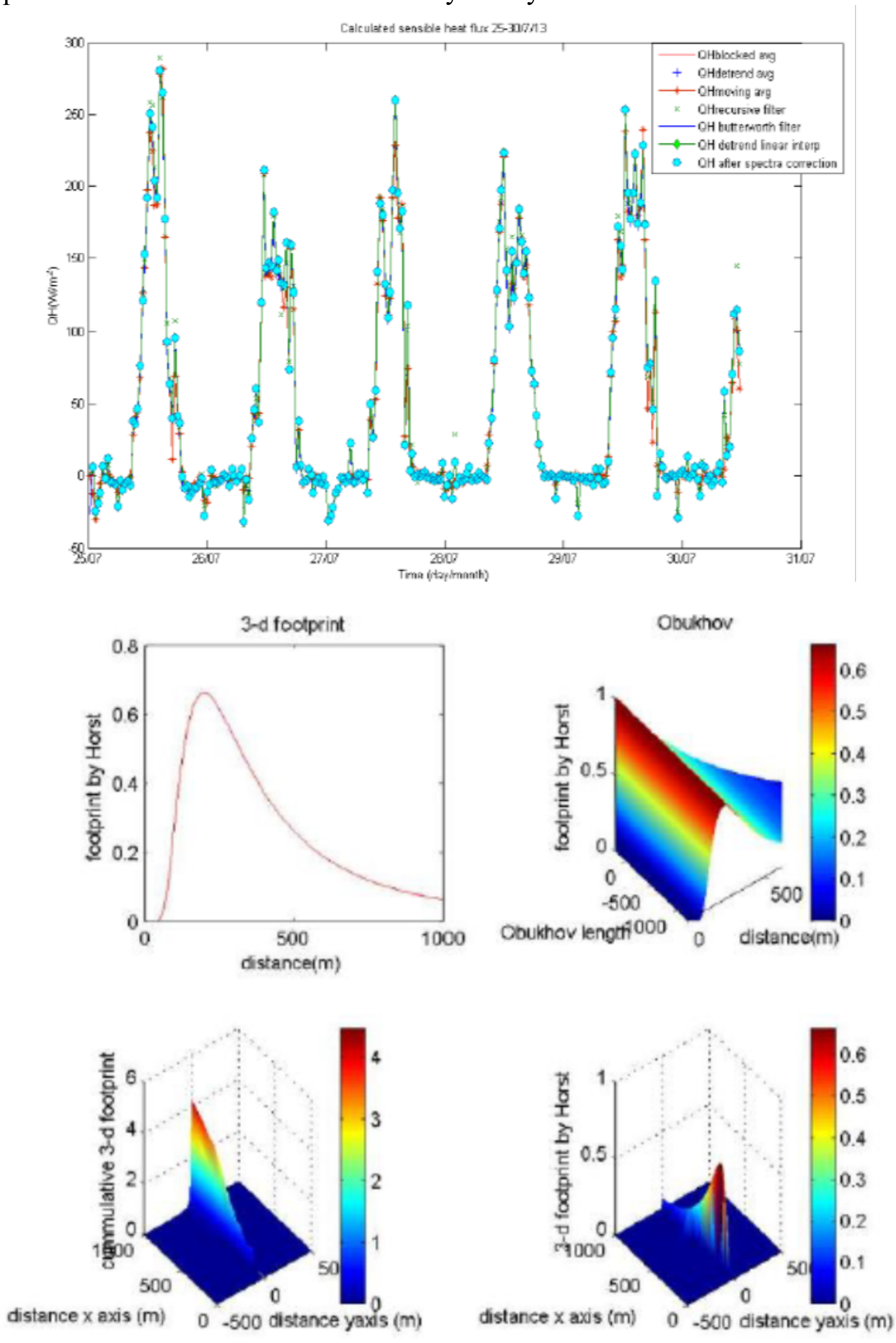

Figure 13 – VOAGr MET TOWER graphical form flux densities using different flux densities calculating methods.

2.2.3.2. Open data sources

There are different services capable of providing useful information regarding meteorological information, with some of them providing publicly available open data, listed below.

2.2.3.2.1. *OpenWeather*

OpenWeather[4] is an online service that provides historical, current and forecast data. Depending on the data/service that is intended to use, it can be free or paid. The complete list of the available parameters and units that the service provides can be found at https://openweathermap.org/current, with the following table giving a brief summary of the available data:

Name	Available
Temperature (min/max)	Yes
Humidity	Yes
Pressure (sea / ground level)	Yes
Wind (speed / direction)	Yes
Clouds	Yes
Precipitation	Yes
Snow	Yes
Visibility	Yes
Sunrise / Sunset	Yes
UV Index	Yes

Table 7 – OpenWeather Dataset from its service

Regarding the API, the following table provides some insights about it:

Characteristic	Details
Calls per minute	60
Search by city Name/ID/Coordinates	Yes
Current weather	Yes
Forecast	5 Days, 3-hour intervals
Calls per day (historical)	5000 (paid)
Historical weather	1 month back (paid)

Table 8 – OpenWeather API conditions

2.2.3.2.2. *IPMA*

IPMA[5] is the Portuguese service holding the biggest national network of meteorological infra-structures that provides current weather and forecast data[6]. The last one can be aggregated by the day (until 3 days) or local (until 5 days). The following table gives a summary of the available data:

Name	Available
Temperature (min/max)	Yes
Humidity	No
Pressure (sea / ground level)	No
Wind (speed / direction)	Yes (by classes)
Clouds	Yes * (inferred by weather classes)
Precipitation	Yes * (inferred by weather classes)
Snow	Yes * (inferred by weather classes)
Visibility	No

[4] https://openweathermap.org/
[5] http://www.ipma.pt/pt/index.html
[6] https://api.ipma.pt/

Sunrise / Sunset	No
UV Index	No

Table 9 – IPMA Dataset from its service

Regarding the API, the following table gives some insights about it:

Characteristic	Details
Calls per minute	Infinite
Search by city Name/ID/Coordinates	No/Yes/No
Current weather	Yes
Forecast	No (current weather is a forecast)
Calls per day (historical)	Yes, 5 days
Historical weather	No

Table 10 – IPMA API conditions

2.2.3.2.3. *DarkSky*

DarkSky[7] provides current and forecast (minutely, hourly and daily) weather and alerts. The complete list of the available parameters and units that the service provides can be found at https://darksky.net/dev/docs. The following table gives a summary of the available data:

Name	Available
Temperature (min/max)	Yes
Humidity	Yes
Pressure (sea / ground level)	Yes (only one, not distinguished)
Wind (speed / direction)	Yes
Clouds	Yes
Precipitation	Yes
Precipitation probability	Yes
Precipitation Intensity/accumulation	Yes
Snow	Yes
Visibility	Yes
Sunrise / Sunset	Yes
UV Index	Yes
Temperature High/Low times	Yes
Dew Point	Yes
Apparent Temperature	Yes

Table 11 – DarkSky Dataset from its service

Regarding the API, the following table gives some insights about it:

Characteristic	Details
Calls per day	1000 (free)
Search by city Name/ID/Coordinates	Yes
Current weather	Yes
Forecast	Yes (next 48h)
Historical weather	Yes

Table 12 – DarkSky API conditions

[7] https://darksky.net/

At the time of writing this deliverable, the service has already been acquired by Apple[8], which will certainly impose restrictions to the usage of the service, therefore the consortium will continuously monitor the service's conditions to assess its practicality for EOSC and NEANIAS.

2.2.3.2.4. *Agromonitoring*

Agromonitoring[9] is part of the OpenWeather company but focuses on data suitable for agricultural applications, it provides current, forecast and historical weather base on satellite data for a specific area. The complete list of the available parameters and units that the service provides can be found at https://agromonitoring.com/api. It also provides NDVI (Normalised Difference Vegetation Index), EVI (Enhanced Vegetation Index) and accumulated temperature and precipitation, soil temperature and moisture. It has basically the same data and API limits as OpenWeather, it only adds the Satellite imagery for the agricultural applications. The historical data requires a paid subscription.

2.2.3.2.5. *AQICN*

AQICN[10] provides worldwide Air Quality Index for different locations and cities. This service only provides the index for the air quality (AQI) and not the pollutants considered for its calculation. Regarding the API, the following table gives some insights about it:

Characteristic	Details
Quota	1000 per second
Search by city Name/ID/Coordinates	Yes
Current AQI	Yes
Historical AQI	No

Table 13 – AQICN API conditions

2.2.3.3. Research Projects

In the following section, research projects that are relevant for NEANIAS regarding the Atmospheric services are analysed.

2.2.3.3.1. *Copernicus*

European Union's Earth Observation Programme, Copernicus[11], offers information services based on satellite Earth observation and non-space data. It delivers near-real-time data that can be used globally and regionally, served by a set of satellites and in-situ systems such as ground stations. The services that transform the gathered satellite and in-situ data into valuable datasets that ensure the monitoring of changes, allow better forecasts. The main areas where Copernicus focuses are Marine, Climate change, Security, Emergency, Land and Atmosphere[12].
The Copernicus Atmosphere Monitoring Service (CAMS) provides continuous data and information on atmospheric composition. It provides daily information on the global atmospheric composition, reactive gases, ozone, and aerosols. It also provides near-real-time and forecast (4 days) of the European air quality[13].

[8] https://www.businessinsider.com/apple-buys-dark-sky-app-grow-services-iphone-2020-4?op=1
[9] https://agromonitoring.com/
[10] https://aqicn.org/
[11] https://www.copernicus.eu/en
[12] https://www.copernicus.eu/en/about-copernicus/copernicus-brief
[13] https://www.copernicus.eu/en/services/atmosphere

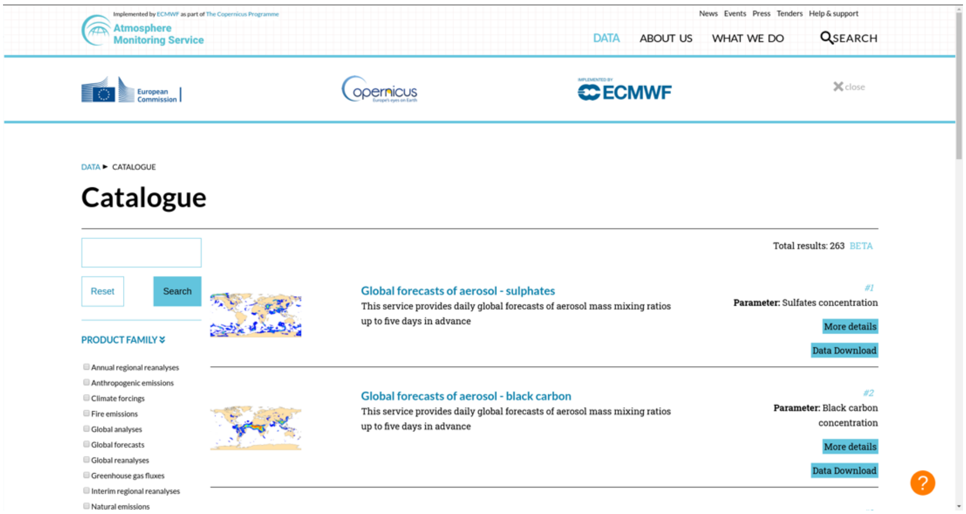

Figure 14 – Copernicus online catalogue of data.

For the Copernicus Atmosphere Monitoring Service, the data catalogue and daily analyses and forecasts can be found at https://atmosphere.copernicus.eu/data (shown in the above image).

2.2.3.3.2. *ACTRiS*

ACTRIS[14] (Aerosol, Clouds, and Trace Gases Research Infrastructure) is the European Research Infrastructure for the observation of Aerosol, Clouds and Traces Gases. Serves a vast community of users working on atmospheric research, climate and Earth system and air quality models, satellite retrievals, weather analysis, and forecast systems by offering high-quality data and research infrastructure services for atmospheric aerosols, clouds, and trace gases. This research infrastructure aims to contribute in resolving societal and environmental challenges, such as air quality, health, sustainability, and climate change, by providing a platform for researchers to combine their efforts and by providing open datasets of aerosols, clouds, and trace gases[15].

[14] https://www.actris.eu/default.aspx
[15] https://www.actris.eu/about/actris/whatisactris.aspx

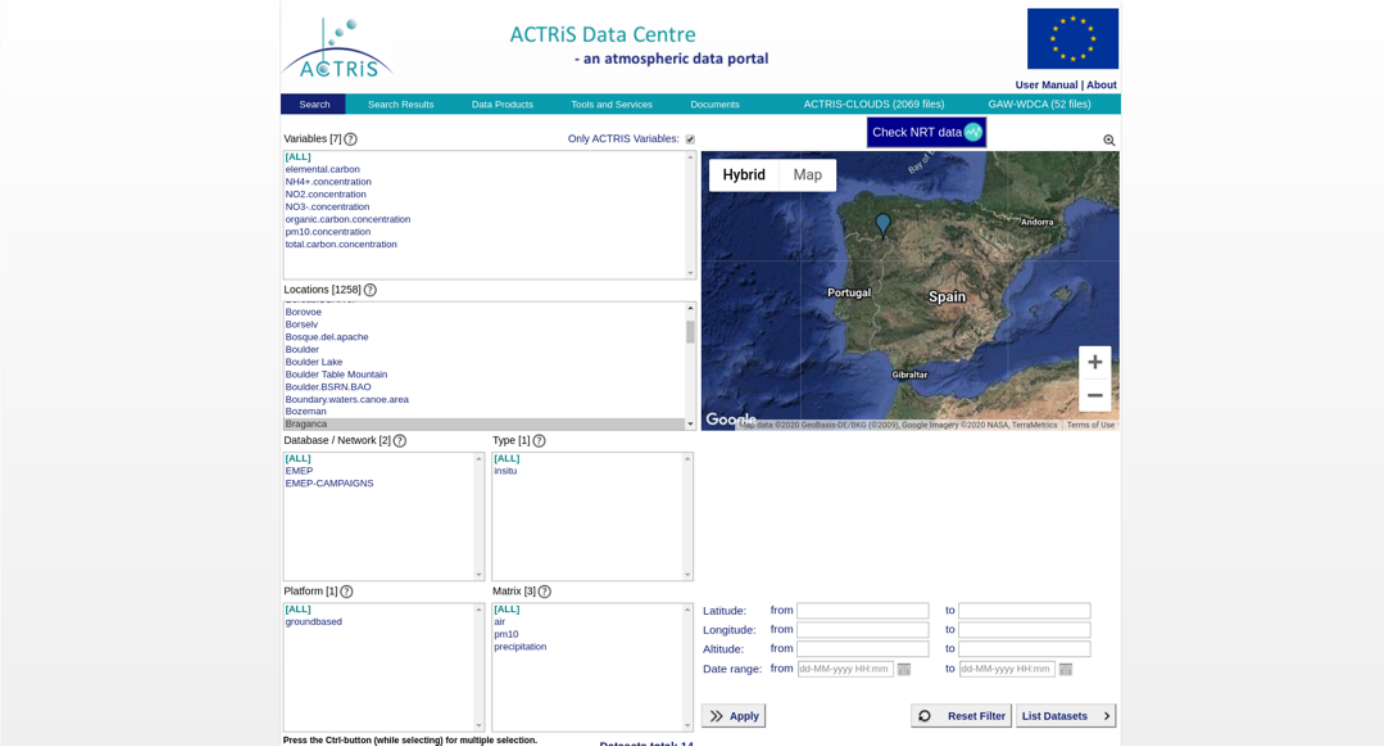

Figure 15 – ACTRiS Data Centre (an online portal for data).

The ACTRiS data portal can be found at https://actris.nilu.no/. It provides not only a way to explore the available datasets, but also a tool to plot the values for analysis, as seen below.

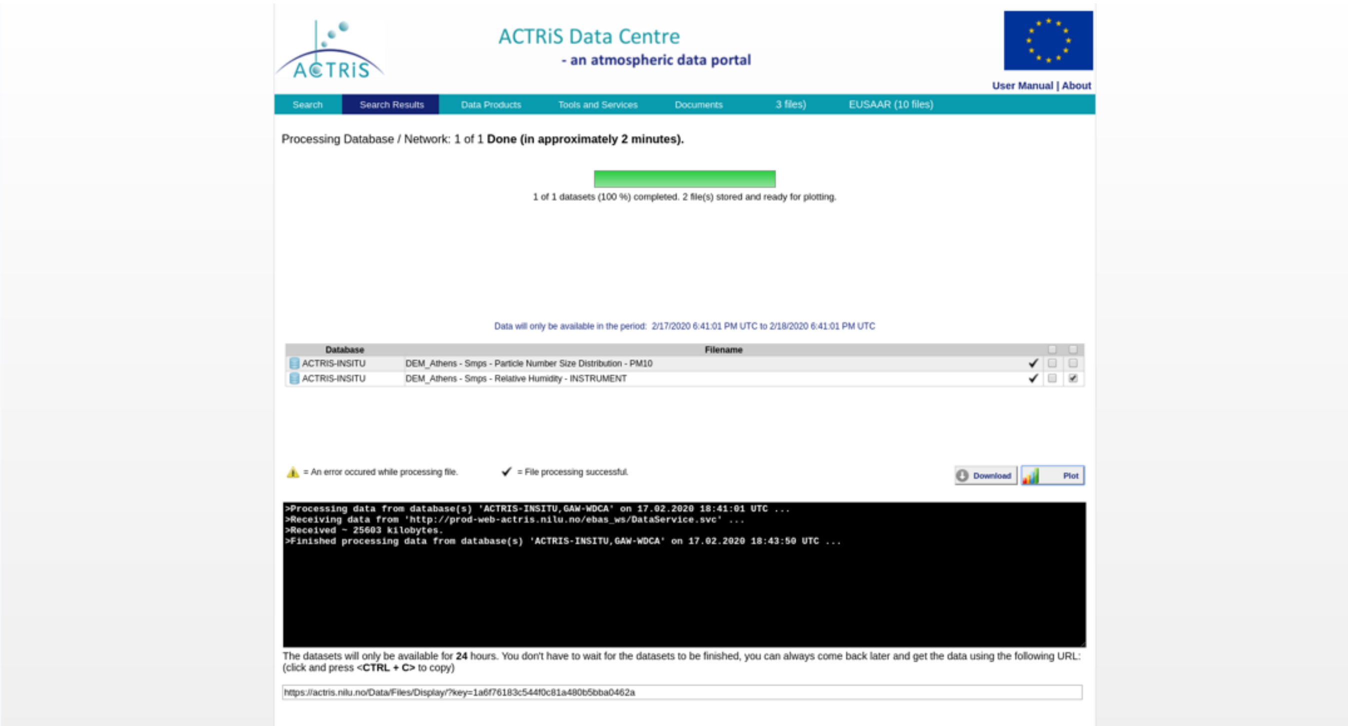

Figure 16 – ACTRIs processing data through its portal.

2.2.3.3.3. *symbIoTe*

The **symbIoTe** (symbiosis of smart objects across IoT environments, https://www.symbiote-h2020.eu/) project provided an interoperable mediation framework to enable the discovery and sharing of connected devices across different IoT platforms for rapid development of cross-platform IoT applications.

Within symbIoTe, the **Smart Mobility and Ecological Routing use-case** addresses the problem of inefficient transportation and poor air quality that many European cities face nowadays. This use case offers the ecologically most preferable routes for motorists, bicyclists and pedestrians based on the available traffic and environmental data acquired through various platforms. This scenario is extremely relevant for people who travel within the major European cities, since a constant exposure to pollutants can cause severe health problems. It is also of the interest of the municipalities' governing bodies that, by helping their citizens to avoid these health problems, they can reduce health care costs. Additionally, the use-case provided a way for users to search for Points of Interests, filtered by certain factors such as air quality, noise pollution and parking availability. symbIoTe empowered this use-case by providing platform interoperability, allowing for developers to easily access and handle data from different platforms and domains in the same manner.

symbIoTe Core Services are responsible for storing IoT meta-information about the platforms, their resources and domain models they are using. It also provides search, security, administration and monitoring features. Another layer, symbIoTe Cloud Services, provide a virtualization support for integrating IoT platforms and allowing unified and secured remote access to the observation data, services and actuators. The provided service is on the TRL 8, proposed to be included in the EOSC Portal[16].

2.3. Atmospheric Sector: Current Software and Services

In this section, software and relevant services are identified, in order to understand what is out there that can be used to reach the goals of the project. Currently the most popular software used within the atmospheric sector is Matlab, where the following algorithms are integrated and tested:

- **Direct eddy covariance**. The data are subjected to despiking by computing the mean and standard deviation for a series of moving windows of various lengths, [8] (Vickers and Mahrt, 1997). Any point which is more than 3.5std from the window mean is replaced by linear interpolation. If the consecutive spikes are more than three, they are considered as missing values. Since steady state conditions must be considered, the data are flagged when they fail to pass the respective test of Foken and Wichura (1996) [2]. They are also tested for skewness and kurtosis [8][6] (Vickers and Mahrt, 1997; Lee et al. 2004). To ensure that the data correspond to turbulent development, their integral turbulent characteristics are compared to model ones based on the flux variance similarity, [1] (Foken, 2008). Lag correlation analysis is also performed for the specific humidity and temperature for their correlation with vertical velocity. To fix the orientation of coordinate frame for the sonic anemometer wind measurements, both the 2-d coordinate rotation and the planar fit method [9] (Wilzcak et al., 2001) were applied separately to investigate the influence of the methods to the results. The post processing includes the determination of fluctuations and the separation of the turbulent signals from trends or low frequency components imposed either by instrumental drift or as a result of changes in meteorological conditions. The approaches used ([5] Moncrieff et al, 2004) are, time averaging and removal of a standard or a moving mean, detrending by subtracting the trend from each averaging period using the least square method, filtering the datasets with a recursive filter or a combination of pre-filtering with a low Butterworth filter and time averaging. The comparison of the results leads to the most suitable technique for each case study. Corrections for buoyancy fluxes according to Schotanus et al. (1983) [7] were also applied, while ogive tests determine the more suitable averaging period. Spectra corrections are thence considered for estimating flux losses in the high frequency

[16] https://marketplace.eosc-portal.eu/services/symbiote

range. In order to estimate the source areas for the convective flux measurements various footprint models are applied like the Eulerian analytical model suggested by Horst and Weil (1992) [3] and the footprint model of Hsieh et al. (2000) [4].

- **Modified Bowen Ratio Method**. Data corresponding to friction velocity below the limit of 0.07m/s are excluded.
- **Gradient method** based on Monin Obukhov similarity theory is also applied for calculating turbulent heat fluxes, momentum and scalar fluxes from the slow response measurements taken at four levels on the tower. The iterative method was used after imposing the correction function for stability conditions.
- **Bulk approach**. Calculating turbulent heat fluxes, momentum and scalar fluxes from the slow response measurements taken at 2 levels on the tower. Flux variance method based on power law and correlation methods to estimate scalar fluxes and sensible heat fluxes under free convection conditions.

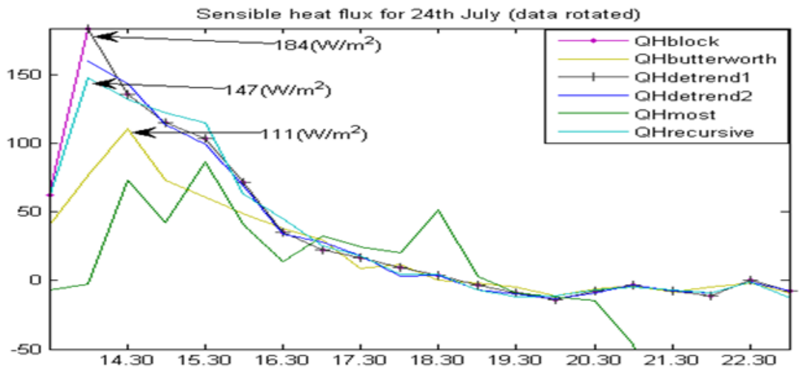

Figure 17 – Example of fluxes on the 24th July 2013 with different data filtering.

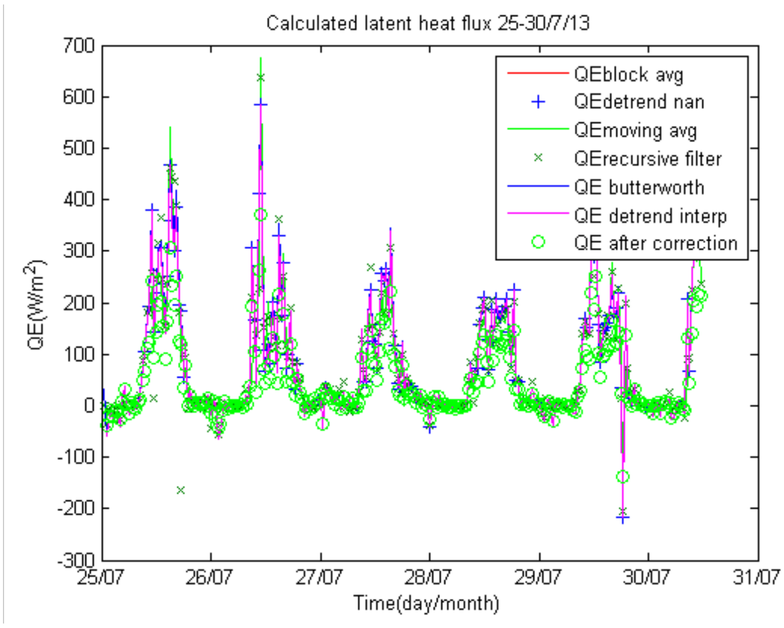

Figure 18 – Example latent heat flux density.

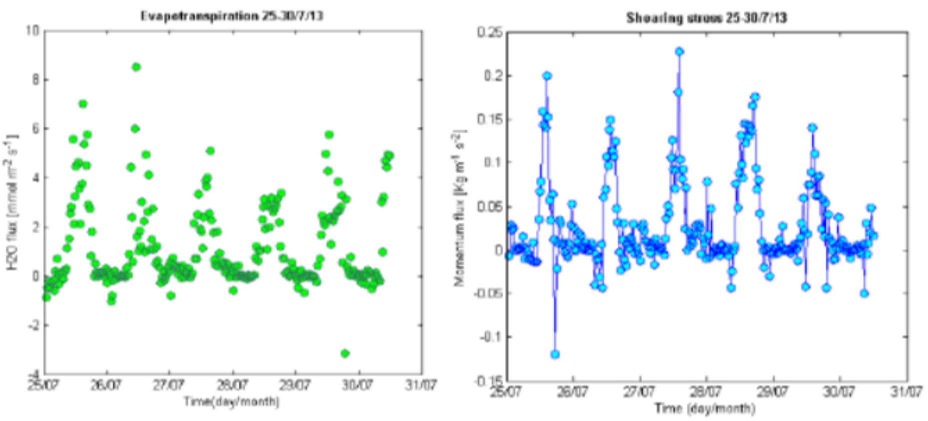

Figure 19 – Example of determined Evapotranspiration on 25-30 July 2013.

2.3.1. [A1] – Greenhouse gases flux density monitoring service

As mentioned above in section 2.2.3.1 and in the section 4.2 below, all software for the flux density implementation will be available in the form of a Matlab executable with a number of subroutines for filtering data, calculating fluxes with a number of methods and inter comparing these methods (see figure below). The format of data inputs to Matlab is ASCII. Furthermore, according to the EU Directives 2008/50/EU and 2015/1480/EU there should be a national reference point to validate these calculations. However, as it is common practice among atmospheric (and other) scientist "round-robin" exercises will reveal errors, accuracy and repeatability of the results. Adjustments can be easily implemented, once the "basic structure" of the software is in place.

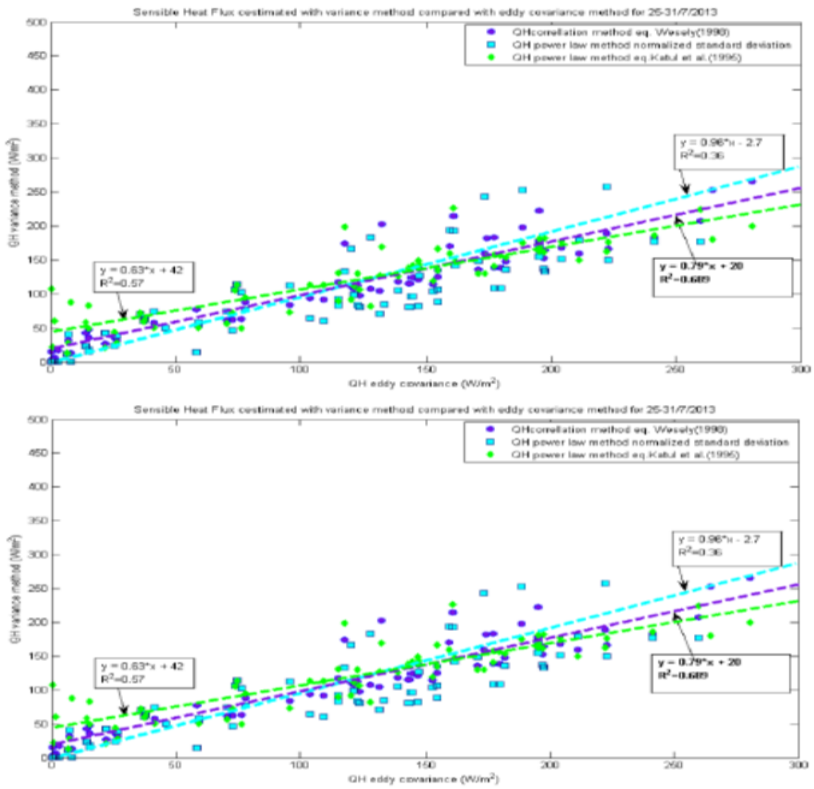

Figure 20 – Comparison of the determination Sensible Heat fluxes by Eddy Covariance and Variances methods

2.3.2. [A2] – Monitoring atmospheric perturbations and components in active tectonic regions service

In active tectonic regions, especially in volcanic areas, gas distribution is dependent on atmospheric conditions and also on changes of the stress forces acting on fractures. For these reasons, monitoring the atmospheric components, with special attention to radon emissions, represents an opportunity for geologists to obtain information about the tectonic activity of the area. In order to achieve this goal, the geological community is asking to the NEANIAS project to release a specific software, that currently is not yet available, in order to correlate gas emissions with earthquake parameters and atmospheric conditions. The considered earthquake parameters are: the earthquake Magnitude, the distance between the epicentre and the gas emission site, the distance between the hypocentre and the gas emission site, and data and time of the seismic event. This service is focused especially on radon, but it could be possibly used also for other gases emitted in volcanic areas, like SO_2 and CO_2. The correlation is thus multiparametric and considers the comparison between all the aforementioned data with the variation of gas emission, atmospheric pressure and temperature at the site of gas emission.

The second service the geological community is requesting to the NEANIAS project consists in the implementation of a software that can rebuild the regional stress field of a study area, in order to make it visible in 2D and 3D view. An existing software, called "Lissage" (by Lee and Angelier, 1994) can already reconstruct a 2D map of the stress trajectories starting from local stress tensors, which in turn are reconstructed from the collection of structural data.
The "Lissage" software can use two interpolation methods in order to obtain the resulting map, which consists of a gridded map and the trajectories of the stress field (see figures below):

i. a linear model which uses a bivariate polynomial function to obtain the distribution of the regional stress vectors. The smoothness degree of the trajectories, in this case, is controlled by the user which is asked to select the polynomial degree k.

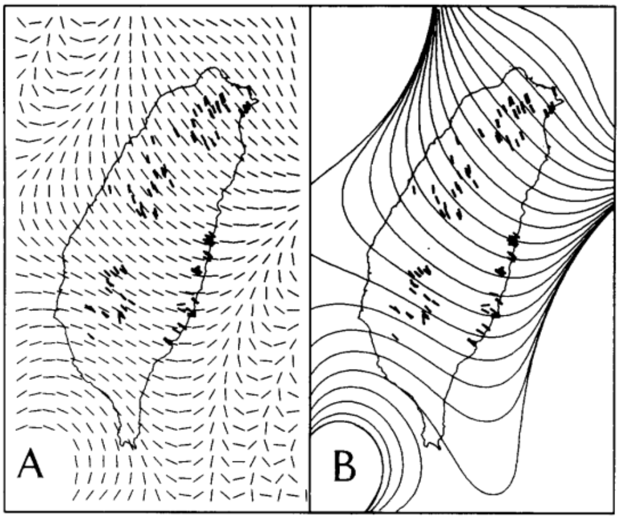

Figure 21 – Algorithm of polynomial function (k = 3). Thick bars illustrate local orientations of maximum horizontal stress axes. A) Gridded map of calculated stress field. Thin bars illustrate calculated maximum horizontal stress axes on the grid nodes. B) Corresponding calculated stress trajectory map. Figure from (Lee and Angelier, 1994)

ii. an inverse-distance weighting function which calculates the stress direction in any point of the study area. The smoothness degree of trajectories is controlled by the user by selecting the distance limit R to determine the weights values and the power value p of the weighting function.

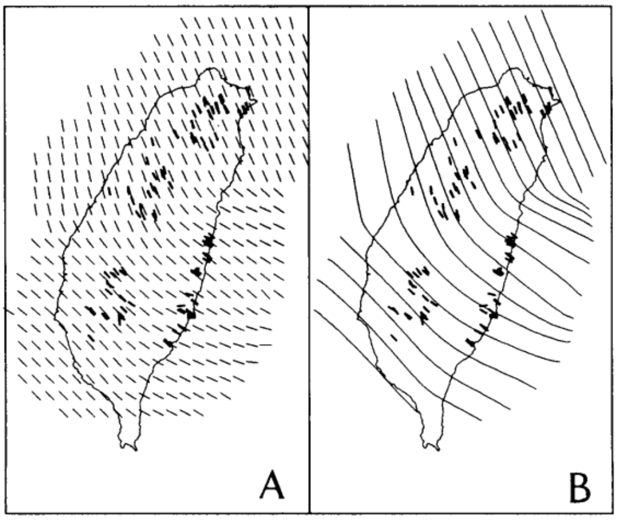

Figure 22 – Algorithm of distance-weighting method (R = 60, p = 2). Thick bars show local orientations of maximum horizontal stress axes. A) Gridded map of calculated stress field; thin bars indicate local maximum horizontal stress axes. B) Corresponding calculated stress trajectory map. Figure from (Lee and Angelier, 1994)

The Lissage software presents several limitations: first of all, it has been written using the Turbo C language, and therefore it does not work on modern PCs. Furthermore, it is not user-friendly and it has an obsolete graphic interface which can be improved. Additionally, it does not consider different data sources (e.g. field data, earthquake focal mechanisms, in-situ geotechnical stress measurements) and it cannot manage a large amount of data. Finally, it does not provide 3D or 2D outputs which can be opened in a GIS environment. Hence, the goal is to implement a service based on the "Lissage" software using Python, a high-level programming language which offers access to a large variety of libraries and modules for scientific computing and visualisation.
Moreover, this upgrade will provide to the geological scientific community, a tool that allows to export the computed products in a variety of file formats compatible with OGC standards (e.g. geopackage, klm, shp, etc.) accompanied by the corresponding georeference.

2.3.3. [A3] – Air quality estimation, monitoring and forecasting service

To achieve the expected goals, an aggregation system is needed, capable of collecting the air quality data from multiple sources. To do so, a system offering high levels of interoperability, allowing to connect different data sources, collect its data, process and make available, is needed. An example of a system like this is EMBERS[17], a EU-funded R&D project that was developed in the scope of the H2020 program by Ubiwhere, Sorbonne Université, Inria, Fraunhofer, FOKUS and Technische Universität Berlin.
The core of this project was to create a cloud mobility platform – the Mobility Catalogue - on the grounds of Ubiwhere's Smart Parking and Smart Traffic products. The Catalogue came to replace all-in-one systems, that required municipalities to purchases full sets of components from a single vendor. Allowing the city manager to purchase best-of-breed devices and apps developed by third parties instead, with the only condition being the interoperability with the back end through an open Smart City mobility API.
This project resulted in different software parts for standardized cross-domain data collection, harmonisation, and storage, that can be used in order to collect, harmonise and store air quality data to be used to build the forecast service.

[17] https://embers.city/

To help with the harmonisation process, it is necessary to find the proper way to store the data. To do so, it's necessary to achieve a data model capable of storing data from different data sources. FIWARE[18], that provides open-source components to accelerate the development of a smart solution, provides not only software components, but also data models, for different use cases. Relevant for this case, FIWARE, provides three data models, Weather Observed[19], Air Quality Observed[20] and Weather Forecast[21], designed for weather and air quality observation at a certain place and time and to provide weather forecast data. These three represent a very helpful beginning, but if they appear not to be adequate to the use case, considering user feedback, they will be adapted. NEANIAS aims at providing a service that considers the most suitable data model not only to handle and store the data but also to be used by other developers, users and researchers.

[18] https://www.fiware.org/
[19] https://fiware-datamodels.readthedocs.io/en/latest/Weather/WeatherObserved/doc/spec/index.html
[20] https://fiware-datamodels.readthedocs.io/en/latest/Environment/AirQualityObserved/doc/spec/index.html
[21] https://fiware-datamodels.readthedocs.io/en/latest/Weather/WeatherForecast/doc/spec/index.html

2.4. User Requirements for Atmospheric Services

Considering the feedback provided by the several user communities described before (urban air quality authorities, geologists, geohazards and civil protection stakeholders, meteorological services, industry and energy decision makers, ecologists, rural and urban planners, as well as insurance and health stakeholders), the following user requirements have been initially formed like ‘User Stories’. A User Story is an instrument used in Agile software development to capture a description of a software feature from an end-user perspective. The User Story describes the type of user, what they want and why.

Id	End-User	User-requirement	Reason
R1	All targeted	Upload, store and possibly publish datasets of various formats	To process end-user data with different services and be able to store them and make them available to other users
R2	All targeted	Use datasets provided by the services	To test a service and learn its functionalities
R3	All targeted	Visualize raw data as well as resulting products and reports	For a first glance, viewing the raw/ vendor data (map or mosaic etc.); inspect and evaluate the results
R4	All targeted	To access data from suitable on-line repositories and data sources	To have access to essential data for performing the tasks of each service and compute the corresponding results
R5	All targeted	Document the workflow	For quality assurance, traceability, reproducibility and backlogging
R6	All targeted	Export the results in various file formats	Facilitate the exchange of datasets between users
R7	Geologist	Correlate gas emissions (e.g. Radon) with earthquake events and atmospheric conditions in active tectonic regions	To study the dependencies between the monitored quantities
R8	Geologist	Produce maps of the regional stress field of a study area which are compatible with GIS environments	To study the relation of the stress field with the gas emissions in the area
R9	Urban air quality authority / Public health authority	Aggregate, harmonise and store cross-domain data from multiple sources (satellite, geospatial, in-situ)	To monitor/forecast the air quality of a region of interest
R10	Meteorologist	Aggregate, harmonise and store GHG flux related data from multiple sources (satellite, geospatial, in-situ)	To improve flux estimation and compute reliable GHG emissions on a regional / national scale
R11	Industry / Energy and power decision maker / Ecologist	Calculate the flux densities of GHG emissions	To monitor the emissions in a region of interest and formulate suitable management policies
R12	Rural / urban planner	Filter the GHG emissions and energy fluxes per area or location	To monitor the emissions in a region of interest and formulate suitable management policies
R13	Researchers/ governmental organisation	To upload their own measurements for computing flux densities	Produce their own flux densities with NEANIAS software
R14	Researchers/ governmental organisation	To employ available atmospheric data from NEANIAS for flux densities benchmarks and learning purposes	Produce their own flux densities
R15	Researchers/ Industry	To compare their own software with the NEANIAS developed one for flux densities estimation	Benchmark their own flux densities solution
R16	All targeted	Improve portability and reproducibility	To make methods / results / findings accessible to the public / scientific community

Table 14 – User Requirements for Atmospheric Services

Then, the aforementioned user stories/ requirements were linked with a level of priority and value based on end-user recommendations. Also, the manner to confirm their acceptance was also marked down.

Id	Priority	Value	Acceptance
R1	High	High	User Interface (UI) provides functionalities for uploading / downloading data, relying on a Data Transfer Service
R2	Low	Medium	Services provide pre-loaded datasets for testing their functionalities
R3	Medium	High	UI provides functionalities to visualize the data and the results
R4	Medium	High	Services provide access to selected external datasets and repositories which provide essential data for their operation
R5	Medium	Medium	Provide proper support to the logging, backlogging, auditing and accounting functionalities offered by the core services
R6	Low	Medium	Implement data serialization processes supporting several widely-used data formats.
R7	High	High	Be able to examine correlations between gas emissions and any earthquake related events/ time series
R8	High	High	Be able to produce maps of a regional stress field
R9	High	High	Be able to fuse data from multiple sources for air quality monitoring
R10	High	High	Be able to aggregate atmospheric and flux related data on a regional / national scale
R11	High	High	Provide the capability to calculate flux densities of GHG emissions
R12	Medium	High	Be able to ask for data for a certain area given a polygon of coordinates
R13	High	High	The service will provide the capability to upload external data for computing flux densities
R14	Medium	High	Be able to employ available from NEANIAS data for flux density estimation
R15	High	High	Be able to upload and process data and download the result.
R16	Medium	High	Provide related services and be able to store and publish the parametrization corresponding to specific workflows / data

Table 15 – Acceptance criteria for user requirements

3. Co-design and Service Specifications

Based on the requirements relevant to the atmospheric services, as they have been expressed by the users in the previous section, this section defines corresponding specifications for the software development cycle.

Firstly, the service requirements have been categorized according to their type. Five different types have been identified, namely, storage, computing, cloud, functional and quality requirements. The first two types regard the access and utilization of storage and computational resources available to the services. Cloud requirements depend on the features of the cloud infrastructure which will host the services. Functional requirements request for certain features to be made available from the services. Finally, quality requirements are enforced to monitor and evaluate the quality of the results produced by the services.

Another important factor in the process of co-designing the specifications of the atmospheric services is the ranking of each specification, based on the feasibility and the effort required to satisfy a specific requirement in relation to its value, as expressed by the users (Table 15). According to this, each specification defined to address a specific user requirement has been characterized as mandatory, convenient or optional.

Table 16 presents the specifications of the atmospheric services which have been defined to address the requirements expressed by the users. The table, besides the description of the requirement and the relevant aspects described above (type of requirement and ranking), provides the description of the specifications which have been defined for satisfying each requirement, and the services which should implement the specification. Almost all specifications related to the atmospheric services have been ranked as mandatory, with only one been ranked as convenient, namely, the interfacing with the NEANIAS visualization services.

Requirement			Ranking	Specifications		
Req#	Requirement	Type of Req. (Computing / Storage / Cloud / etc.)	Mandatory / Convenient / Optional /	Spec #	Description of the required functionalities, software specification	Addressed by the Atmospheric Service
R01	Upload, store and possibly publish datasets of various formats	Storage	Mandatory	S01	Interface with suitable Data Transfer NEANIAS/EOSC services for uploading / downloading data to medium / long storage locations through well-known protocols (FTP, WebDAV, etc.) and make data FAIR when applicable	A1, A2, A3
R02	Use datasets provided by the services	Storage	Mandatory	S01	Interface with suitable Data Transfer NEANIAS/EOSC services for accessing data made available by the service provider for testing and learning the functionalities of the service	A1, A2, A3
R03	Visualize raw data, resulting products and reports	Functional	Mandatory	S02	Each Atmospheric service implements task-specific reporting mechanisms	A1, A2, A3
		Functional	Convenient	S03	Interface with NEANIAS visualization services for visualizing raw data and results	A1, A2, A3
R04	To access data from suitable on-line repositories and data sources	Functional	Mandatory	S04	Implement task-specific services for accessing data provided through custom APIs	A1, A2, A3
		Functional	Mandatory	S01	Interface with suitable Data Transfer NEANIAS/EOSC services for accessing data made available by external repositories when provided through standard data protocols	A1, A2, A3
R05	Document the workflow	Cloud	Mandatory	S05	Atmospheric services implement suitable logging mechanisms and store meta-data describing the service parametrization	A1, A2, A3
		Cloud	Mandatory	S06	Interface with AAI, auditing and other relevant NEANIAS services for collecting user-related information	A1, A2, A3
		Storage	Mandatory	S01	Interface with NEANIAS data transfer and other relevant services for storing and publishing logs and the workflow related meta-data	A1, A2, A3

R06	Export the results in various file formats	Functional	Mandatory	S07	Each service offers option to serialize the produced data in widely used file formats	A1, A2, A3
		Storage	Mandatory	S01	Interfaces with NEANIAS data transfer services for storing the output data on suitable storage locations	A1, A2, A3
R07	Correlate gas emissions (e.g. Radon) with earthquake events and atmospheric conditions in active tectonic regions	Computing	Mandatory	S08	Compute correlation and other statistical quantities of interest between a set of relevant variables defined by the user	A2
		Storage	Mandatory	S01	Interface with NEANIAS/EOSC Data Transfer services for accessing data through standard data transfer protocols	A2
		Functional	Mandatory	S04	Interface with task-specific services for accessing data through custom APIs	A2
R08	Produce maps of the regional stress field of a study area which are compatible with GIS environments	Computing	Mandatory	S09	Compute stress field maps based on well-known interpolation methods (e.g. inverse distance weighting or bivariate polynomial function)	A2
		Functional	Mandatory	S07	Support export to data formats compatible with GIS environments (e.g. geoTIFF, KML, geoJSON, etc.)	A2
		Storage	Mandatory	S01	Interface with NEANIAS/EOSC Data Transfer services for storing data using standard data transfer protocols	A2
R09	Aggregate, harmonise and store cross-domain data from multiple sources (satellite, geospatial, in-situ) for air quality monitoring	Cloud	Mandatory	S10	Integrate Smart Air Quality's management platform as a SaaS platform	A3
		Storage	Mandatory	S01	Interface with NEANIAS/EOSC Data Transfer services for accessing and storing data through standard data transfer protocols	A3
		Functional	Mandatory	S04	Interface with task-specific services for accessing data through custom APIs	A3
R10	Aggregate, harmonise and store GHG flux related data on a regional / national scale	Computing	Mandatory	S11	Aggregate and harmonize cross-domain GHG flux data	A1, A3
		Storage	Mandatory	S01	Interface with NEANIAS/EOSC Data Transfer services for accessing and storing data through standard data transfer protocols	A1, A3
		Functional	Mandatory	S04	Interface with task-specific services for accessing data through custom APIs	A1, A3
R11	Calculate the flux densities of GHG emissions	Computing	Mandatory	S12	Compute flux densities of GHG emissions (dynamic gradient calculations, eddy covariance)	A1
R12	Filter the GHG emissions and energy fluxes per area or location	Functional	Mandatory	S14	EOSC service is able to filter data with geographies as input parameters	A3
R13	To upload their own measurements for computing flux densities	Storage	Mandatory	S01	Interface with NEANIAS/EOSC Data Transfer services for storing user-provided data using standard data transfer protocols	A1
R14	To employ available atmospheric data from NEANIAS for flux densities benchmarks and learning purposes	Quality	Mandatory	S13	Define default parametrization and example scenarios for testing / learning each provided service	A1, A2, A3
R15	To compare their own software with the NEANIAS developed one for flux densities estimation	Quality	Mandatory	S02	Each service reports suitable evaluation metrics which allow for comparison with other related methods / software	A1, A2, A3
		Storage	Mandatory	S01	Interface with NEANIAS/EOSC Data Transfer services for uploading user reference results for comparison	A1, A2, A3
R16	Improve portability and reproducibility	Cloud	Mandatory	S05	Atmospheric services implement suitable logging mechanisms and store meta-data describing the service parametrization	A1, A2, A3
		Storage	Mandatory	S01	Interface with NEANIAS data transfer and other relevant services for storing and publishing logs and the workflow related meta-data	A1, A2, A3

Table 16 – Co-Design Specifications

4. Software Development Plan and Guidance

4.1. [A1] – Greenhouse gases flux density monitoring service implementation

This task will implement the A1 through three incremental software releases. The service will develop a range of algorithms, conditions and optimised procedures so that the users can calculate and/or recalculate flus densities according to their understanding and with or without the provided algorithms. The whole optimisation will depend on the meteorological station instrumentation and the principles of its set up. Different procedures and data requirements will be addressed for the dynamic gradient calculations, to calculations via the "eddy covariance" method. Aircraft data will be also employed and integrated in the developed modules considering all prevailing weather conditions. Finally, software modules to process and integrate the ground-based sensed aerosol fluxes (e.g. lidar data) will be also developed. All software modules, workflows, intensive validation to several dataset will be delivered together with documentation and links to relevant large databases.

[2] Software (targeted TRL ≥ 8):

SOFTWARE – A1	
Current state	ALL FLUXES METHODS AVAILABLE in MATLAB language
Software development effort (next year)	(I) GRADIENT in MATLAB (II) AIRCRAFT VARIANCES METHOD in MATLAB (III) ALL ABOVE with wavelet corrections (IV) Eddy Covariance with OGIVE selection corrections (V) POSSIBLE INERTIAL DISSIPATION APPLICATION IN AVAILABLE DATA
Processing Pipeline, Software Inputs and Output	INPUT---DATA AT 10-20 Hz OF ALL SCALARS OUTPUT----FLUXES OF ENERGY (SENSIBLE AND LATENT HEAT), INERTIA, GREENHOUSE GASES, OTHER POLLUTANTS
User Interface*	NOT YET DECIDED UPON
Software Dependencies/ required libraries	MOMENTARILY MATLAB

Table 17 – Software plan for A1

* open issues to be discussed

4.2. [A2] - Monitoring atmospheric perturbations and components in active tectonic regions service implementation

This task will implement the A2 through incremental software releases and integrations. In fact, in order to achieve the A2 goals, the implementation of two service instances is foreseen:

- The first one will allow end-users to correlate statistically gas emissions over time, with a special attention on radon, and earthquake parameters, namely the earthquake Magnitude, the distance between the epicentre and the gas emission site, the distance between the hypocentre and the gas emission site, and data and time of the seismic event.

- The second one will calculate the stress field of a study area starting both from field and seismic data, releasing 2D and 3D map compatible with GIS environment as well as numerical outputs (e.g. spreadsheets).

SOFTWARE – A2	
Current state	The operations related to the first task are currently performed via statistical correlations performed by employing ad-hoc certain software libraries or mathematical statistical tools. The operations related to stress field calculation are performed using the LISSAGE software, with considerable limitations due to compatibility issues and restricted visualization capabilities.
Software development effort (next year)	The main effort will be concentrated in implementing a service for stress field trajectory computation, similarly to the LISSAGE software. Evaluation will be performed considering sample datasets of field and seismic data. Additionally, effort will be dedicated in providing Jupyter notebooks for the statistical correlation of seismic data and gas emissions data, provided via spreadsheets.
Processing Pipeline, Software Inputs and Output	Pipeline: 1) Data selection 2) Processing (interpolation/correlation) 3) Visualisation (Maps/Graphs) Inputs: Spreadsheets, API calls to related web services Outputs: Maps (2D images, GIS compatible products), Graphs (SVG, EPS, bitmaps, etc.), Spreadsheets, HDF5
User Interface*	Main Functionalities required: • Select task • Data selection and upload • Select relevant data fields • Control processing (start/pause/stop)

	• Visualise results • Produce and download reports and maps
Software Dependencies/ required libraries	✓ Python 3 ✓ Jupyter ✓ Numpy, Scipy ✓ Matplotlib ✓ scikit-learn ✓ scikit-image ✓ Pandas ✓ Docker ✓ …
Software Deployment In The Cloud	- The micro-services related to each task will be delivered in a containerized form - A RESTful API will be provided for managing and running the micro-services - A web interface will be used as a client for uploading the data, running the tasks and visualising/downloading the results

Table 18 – Software plan for A2

4.3. [A3] - Air quality estimation, monitoring and forecasting service implementation

This task will implement the A3 through incremental software releases and integrations. This service will be delivered by further optimizing, automating and validating software modules for air quality estimation along with the novel integrations of satellite, geospatial and in-situ data. The whole optimization and validation will depend on the available datasets.
Initially, it will be necessary to identify and analyse all the available datasets and integrate them into a software module capable of dealing and aggregate all the available datasets. After that, using, the available datasets, it will be possible to work on the air quality estimation, forecasting, and monitoring modules in order to achieve the goal of optimization and automation of these. Finally, the integration of monitoring software platforms that give the visualization module about all the collected data from sensing station.

SOFTWARE – A3	
Current state	- Operational weather forecasting models (such as WRF and MM5)
Software development effort (next year)	- Software modules to compute and integrate Lagrangian and Gaussian dispersion scenarios for the emission inventories of air pollutants
Processing Pipeline, Software Inputs and Output	- Validation against actual in-situ observation with real-time monitoring scenarios for measuring pollutant's concentrations
User Interface*	- A UI will be also developed letting the user to upload data and query data as well as obtaining the desired forecasts and reports. The user interface will communicate with the data analysis modules via REST APIs, while the validation will cover several application scenarios and multiple user communities (urban air quality authorities, meteorological services, energy and power generating sector, and industrial air pollutant emitters)
Software Dependencies/ required libraries	- Python, Django, Django REST Framework, pandas, sci-kit, PostgreSQL, PostGIS, Superset, ReactJS
Software Deployment In The Cloud	- Docker microservices, potentially in Kubernetes - A RESTful API will be provided for managing and running the micro-services - A web interface will be used as a client for uploading the data, running the tasks and visualising/downloading the results

Table 19 – Software plan for A3

5. Evaluation – Conclusions – Future Steps

This deliverable presented the requirements of the NEANIAS WP3 "Atmospheric Research Services" including a detailed description of the user communities, datasets and products involved as well as the development plan and validation for the three thematic services: A1 - Greenhouse gases flux density monitoring, A2 – Monitoring atmospheric perturbations and components in active tectonic regions, and A3 – Air Quality estimation, monitoring and forecasting.

Having analysed the data and computational requirements collected from the Atmospheric related application domains and user communities, this document served to WP6 for the definition of the needed architecture, design principles and specifications on core services (deliverable D6.1) especially for the task T6.5 C3 - AI services implementation and the task T6.6 C4 - Visualisation services implementation. Furthermore, it has been shared with the User Board and to the WP7 for planning the services delivery and operation.

The deliverable will be continuously updated once new requirements come out (both from user and technological perspectives), with a new deliverable D3.2, due on M25, entitled "Atmospheric Research Services Report on requirements and specification" collecting all the final service specifications and requirements. After defining the requirements, the team will identify the EOSC hub service modules to support the development of Atmospheric specific services and perform a gap analysis to examine and evaluate the desired/ required performance against the actual one.

6. References

[1] Foken T. 2008. Specifics of the Near-Surface Turbulence. In: Micrometeorology: Springer Berlin Heidelberg, 61-104.
[2] Foken, T., Wichura, B., 1996. Tools for quality assessment of surface-based flux measurements. Agric. For. Meteorol. 78: 83-105.
[3] Horst TW, Weil JC. 1992. Footprint estimation for scalar flux measurements in the atmospheric surface layer. . Boundary-Layer Meteorology, 59: 279±96
[4] Hsieh Cheng-I, Katul G, Chi Tze Wen. 2000. An approximate analytical model for footprint estimation of scalar fluxes in thermally stratified atmospheric flows. Advances in Water Resources, 23: 765-772
[5] Moncrieff J, Clement R, Finnigan J, Meyers T. 2005. Averaging, Detrending, and Filtering of Eddy Covariance Time Series. In: Handbook of Micrometeorology, Vol. 29,
[6] Lee X, Massman W, Law B, eds 2004: Kluwer A.P., Dordrecht Netherlands, 7-31.
[7] Schotanus P, Nieuwstadt FTM, de Bruin HAR. 1983. Temperature measurement with a sonic anemometer and its application to heat and moisture flux. Boundary-Layer Meteorology, 26: 81–93.
[8] Vickers D, Mahrt L. 1997. Quality Control and Flux Sampling Problems for Tower and Aircraft Data. Journal of Atmospheric and Oceanic Technology 14: 512-526.
[9] Wilczak J, Oncley S, Stage S. 2001. Sonic Anemometer Tilt Correction Algorithms. Boundary-Layer Meteorology, 99: 127-150.
[10] Lee, J. C., & Angelier, J. (1994). Paleostress trajectory maps based on the results of local determinations: the “Lissage” program. Computers & Geosciences, 20(2), 161-191.

List of acronyms

Acronym	Description
API	Application Programming Interface
AQI	Air Quality Index
ASCII	American Standard Code for Information Interchange
EC	European Commission
EOSC	European Open Science Cloud
EU	European Union
EVI	Enhanced Vegetation Index
GHG	Greenhouse Gases
GIS	Geographic Information System
H2020	Horizon 2020
IoT	Internet of Things
NDVI	Normalised Difference Vegetation Index
OGC	Open Geospatial Consortium
PC	Personal Computer
REST	Representational State Transfer
TRL	Technology Readiness Level
WP	Work Package

7. Appendix

List of End-users outside NEANIAS Consortium that their feedback on the requirements of the considered Atmospheric Services contributed to this deliverable.

#	Associated Entities/ Networks/ Initiatives/ User Communities	Letter[22]
A1	Marine Geosciences Group – Institut de Physique du Globe de Paris (**IPGP**) - CNRS - France	LoI
A2	Earthquake Planning & Protection Organization (**EPPO**) - Ministry of Infrastructure and Transport - Greece	LoI
A3	Geodynamic Institute (**GI**) of National Observatory of Athens (NOA) - National Tsunami Early Warning Center (NTEWC) - Greece	LoI
A4	Hellenic Plate Observing System (**HELPOS**) - NOA - Greece	LoI
A5	Square Kilometre Array (**SKA**) Organization - UK	LoS
A6	**GEANT** pan-European network (www.geant.org)	LoI
A7	Greek Research and Technology Network S.A. (**GRNET** S.A.) - Greece	LoI
A8	United Nations Educational, Scientific and Cultural Organization (**UNESCO**)	LoIn
A9	**DRASSM** - Ministère de la Culture - France	LoIn
A10	Greek Copernicus Relay Network - Space Cluster (**Si-Cluster**) Corallia - Greece	LoI
A11	United Nations - Sustainable Development Solutions Network (**UN-SDSN**) - Greece	LoI
A12	EIT **Climate-KIC** Hub Greece	LoI
A13	**LifeWatchGreece** Research Infrastructure - Greece	LoI
A14	**OpenAIRE** https://www.openaire.eu/	LoC
A15	International Lithosphere Program **(ILP)** https://www.scl-ilp.org/	LoI
A16	Democritus Research Center (**DRC**) - Greek representing for **ICOS** RI (Integrated Carbon Observation System Research Infrastructure) & **ACTRIS** (Aerosol, Clouds and Trace Gases RI)	LoI
A17	National Institute of Oceanography and Applied Geophysics (**OGS**) - Italy	LoI
A18	EDF Énergies Nouvelles (**EDF**) - France https://www.edf-renouvelables.com/en/	LoI

	External Advisors (EA) / Members of Extended Strategic Steering Committee (ExSSC)	Letter
1	**Dr Jorge Sanchez-Papaspiliou**, Member of the **si-Cluster** Steering Board, Chief Strategy & Financial Officer, Corallia, Greece	*yes*
2	**Dr. Yan Grange**, LOFAR Software Developer, member of **Science Data Centre** at ASTRON, the Netherlands Institute for Radio Astronomy	*yes*
3	**Dr Javier Escartin**, Head of the Marine Geosciences Group, Institute de Physique du Globe de Paris - **CNRS** - France	*yes*
4	**Prof. Paul D. Williams,** Professor of Atmospheric Science, University of Reading (**UoR**), UK	*yes*
5	**Dr. Gergely Sipos**, Customer and Technical Outreach Manager, **EGI** / member of the EOSC-hub Engagement team	*yes*
6	**Dr. Sofia Vallecorsa**, Computer Engineer, **CERN** openlab	*yes*
7	**Dr. Paolo Diviacco,** Coordinator of the **OGS DIAM-PROS** unit, Vice-director of the RIMA (marine research and technologies) department of the OGS.	*yes*
8	**Dr. Vicki Ferrini**, Marine Geology & Geophysics - Columbia University - USA	*yes*

[22] Letter of Intent (**LoIn**), Letter of Support (**LoS**), Letter of Interest (**LoI**), Letter of Commitment (**LoC**)
All letters can be found at the end of this Section