A distributed data-mining software platform for extreme data across the compute continuum

# D1.1 Use-case requirements

## Version 1.0

## Documentation Information

| | |
|---|---|
| Contract Number | **101093110** |
| Project Website | www.extract_project.eu |
| Contractual Deadline | 30.06.2023 |
| Dissemination Level | PUBLIC |
| Nature | Report |
| Author | OBSPARIS, VEN |
| Contributors | MATHEMA, IBM, BSC, LRI |
| Reviewer | LRI |
| Keywords | Use case scenarios, requirements, analytics |

The EXTRACT Project has received funding from the European Union's Horizon Europe programme under grant agreement number 101093110.

# Change Log

| Version | Description Change |
|---|---|
| **V0.1** | File created |
| **V0.2** | Overall structure of the document + ToC |
| **V0.3** | Outline of TASKA use case added |
| **V0.4** | Outline of PER use case added |
| **V0.5** | Digital Twin Technologies - Background and requirements analysis Description / PER datasets listed |
| **V0.6** | TASKA + PER stakeholders' analysis + functional and non-functional requirements tables added |
| **V0.7** | Conclusions |
| **V0.8** | Cleaning up |
| **V1.0** | Final Review, BSC |

# Table of Contents

# 1. Executive Summary

The purpose of this deliverable is to investigate and define the requirements for the EXTRACT project, which will serve as the foundation for the development of the EXTRACT platform/system.

The first step in defining the requirements was to identify the stakeholders and system actors involved in EXTRACT, along with their respective roles within the project. Additionally, the various application environments in which the EXTRACT solution is expected to operate were identified.

To define the user requirements, a comprehensive strategy was implemented, including the use of questionnaires filled out by stakeholders and regular meetings between end-users and the consortium. These efforts aimed to gain a deeper understanding of the users' needs and expectations, ultimately shaping the EXTRACT solutions accordingly.

Building upon the stakeholder requirements, a set of system requirements was derived and presented in this deliverable. However, it is important to note that these requirements are not final and will be updated as the system architecture is further developed and finalized.

## 1.1. Relation with other deliverables and tasks

This report plays a significant role in fine-tuning the use-case functionalities, as defined in Section 1.2.3 From Lab to Market: The EXTRACT Use-cases of the EXTRACT DoA. The emphasis on analyzing and fulfilling the use-case's KPIs directly shapes further tasks and deliverables in the project, and in particular: WP2 (DE2.1), WP3 (DE3.1), WP4 (DE4.1).

## 1.2. Structure of the document

Deliverable D1.1 is structured into several main sections to provide a comprehensive understanding of the EXTRACT project.

Section 1 serves as an introduction to the document, outlining its purpose and structure. In Section 2, a thorough and analytical overview of the project's scope and challenges is presented. This section defines the stakeholders, system actors, and their respective roles within the project, focusing on the identified use cases. Section 3 begins by describing the methodology employed for gathering and analyzing user requirements specific to the EXTRACT project's context. It then provides an analytical overview of the questionnaire distributed to project stakeholders and discusses the outcomes derived from their responses. These outcomes played a crucial role in defining and specifying the EXTRACT user requirements. The definition and prioritization of the EXTRACT stakeholders' requirements are detailed in Section 4, emphasizing their relative importance. Section 5 outlines the functional and non-functional requirements of the project through a series of lists encompassing user requirements and general requirements. Finally, Section 6 offers concluding remarks summarizing the key findings and insights presented in the document. Additionally, an annex is included, featuring the questionnaire utilized during the user surveys.

# 2. Acronyms and Abbreviations

- CA – Consortium Agreement
- D – deliverable
- DoA – Description of Action (Annex 1 of the Grant Agreement)
- EB – Executive Board
- EC – European Commission
- GA – General Assembly / Grant Agreement
- HPC – High Performance Computing
- IPR – Intellectual Property Right
- KPI – Key Performance Indicator
- M – Month
- MS – Milestones
- PM – Person month / Project manager
- UDT –Urban Digital Twin
- WP – Work Package
- WPL – Work Package Leader
- PER – Personal Evacuation Route
- TASKA - Transient Astrophysics using the SKA pathfinder

# 3. EXTRACT Overall Description

## 3.1. Project Scope and Extreme Data Use Case Contextualization

In the digital era, data has become the cornerstone of transformative processes across diverse sectors. However, conventional data mining strategies, although adept at handling specific data needs, often struggle when the data attributes venture into extreme territories. The emergence of extreme data characteristics presents a formidable challenge that necessitates innovative, comprehensive solutions. These solutions need to foster the design, deployment, and streamlined execution of data mining workflows across a diverse, secure, and energy-efficient computing landscape, while addressing the unique demands of extreme data.

The EXTRACT project sets out to tackle this technological void by introducing a data-driven, open-source software platform. This platform amalgamates state-of-the-art technologies, promoting the development of trustworthy, accurate, fair, and eco-friendly data mining workflows that generate superior, actionable insights. Targeting the entire lifecycle of extreme data mining workflows, the EXTRACT platform prioritizes enhancements in performance, energy efficiency, scalability, and security while holistically addressing the intricate aspects of extreme data.

The platform's effectiveness and capabilities will be examined through **two real-world scenarios**, each **with distinct extreme data requirements**:

**The Personalized Evacuation Route (P.E.R.)** This use-case unifies data from European services like Copernicus and Galileo, 5G localization signals, and IoT sensors embedded in smart city infrastructures, focusing on civilian-centric crisis management.

Disasters in this scenario trigger sudden and large data spikes. For example, a scenario such as a large fire in an urban context and the consequent customized collective evacuation mechanism generates a data flow that can be considered extreme in terms of volume, heterogeneity and dynamics.

The fundamental challenge lies in the real-time processing and analysis of this data surge to produce effective evacuation strategies.

**Transient Astrophysics using the SKA pathfinder (TASKA)** - This scenario processes extreme volumes of data from radio-telescopes for real-time solar activity assessment, enabling further scientific exploration.

This use-case constantly handles vast amounts of data, making it "extreme" due to its substantial size and continuous need for real-time interpretation. The primary challenge with TASKA is to manage, process, and extract meaningful insights from these uninterrupted, colossal data streams.

To meet these challenges, the EXTRACT platform will integrate a spectrum of computing technologies, from edge to cloud to high-performance computing (HPC), forming a unified, secure computational continuum. It incorporates enhanced data infrastructures, AI and big data frameworks, novel data-driven orchestration, and distributed monitoring mechanisms, providing a unified continuum abstraction and ensuring cybersecurity and digital privacy across all its layers.

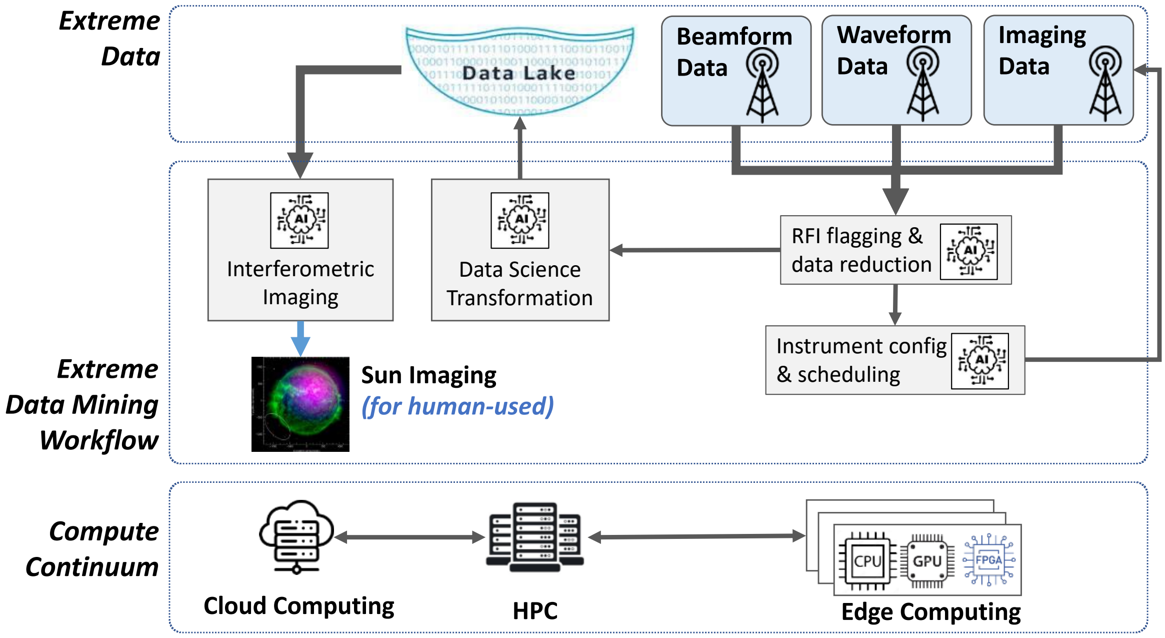
# 4. Use Case Definition

The pivotal task in the EXTRACT project is defining use-cases that can present realistic scenarios dealing with extreme data characteristics. This section will primarily focus on detailing the context, narrative, challenges, and enabling technologies of the TASKA and P.E.R. use cases.

## 4.1. TASKA Use Case

At the horizon 2030, the Square Kilometre Array (SKA), a worldwide project covering the frequencies 50 MHz – 30 GHz, will lead to major science breakthroughs. In Europe, the LOFAR ERIC is the main low frequency SKA pathfinder. Since 2008, France has developed the concept of NenuFAR, a giant compact array with as many low frequency antennas in Nançay as in the whole LOFAR (~2000 antennas), optimized to cover the 10-85 MHz range with a sensitivity 2x to 8x higher than LOFAR, and compatible with LOFAR (the signals can be correlated, NenuFAR playing the role of a second LOFAR super-station). In autonomous mode, the very high instantaneous sensitivity of NenuFAR allows it to address in an innovative way about fifteen scientific themes, from the detection of the cosmological signal of the Cosmic Dawn to the exoplanetary radio emissions. In the LOFAR-super-station mode, NenuFAR will significantly improve the sensitivity of LOFAR for low frequency imaging with high angular resolution.

The improved sensitivity of these new generation telescopes enables the access to a large range of transient radio sources. They usually cover all energy; physical and temporal scales and the associated processing is always challenging. This is itself a cutting-edge science topic, dealing with remote energetic sources (e.g., Fast Radio Bursts) as well as powerful local solar emission (e.g., type III bursts) that could directly affect human activities. It has an undeniable societal impact that takes the form of a “space weather” monitoring service with many applications. Signal processing techniques range from simple threshold detection of events to deep learning classification and restoration of faint variable sources. TASKA has the ambition to prepare and develop the necessary framework and tools that will help scaling by addressing the current challenges of: (i) gigantic dataset transfer and ingestion and (ii) heavy signal processing and (iii) workload and workflow management. In TASKA, we focus on the real-time detection and observation triggering of solar radio events that is a key asset for space weather modeling and alert systems.




The NenuFAR raw dataflow ranges from 600 MB/s (beamforming mode), 102 * 5.5 Gb/s (raw waveform mode per antenna, 5-second snapshots only, or real-time FPGA processing) and up to 5.6 TB/h (imaging mode). The data acquisition strategy is currently set before the observation. Due to the radio frequency interference environment, high resolution acquisition is necessary, and the first processing step is a semi-automated RFI flagging software, allowing to clean the data flow before reducing it by temporal and spectral integration. Transient phenomena also require high temporal and spectral resolutions. Furthermore, the astronomical signals are often much fainter than the local RFI environment. In addition, the imaging mode produces a set of interferometric visibilities, which contains a sparse representation of the observed sky, including instrumental and environmental distortions. The NenuFAR instrument is thus producing a high-speed data flow, with complex intrinsic signatures, and an extreme dynamical range. Its imaging mode produces a sparse sampling of the sky, possibly including non-linear distortions. Finally, the many operating modes and science objectives of NenuFAR make the instrument scheduling highly challenging.

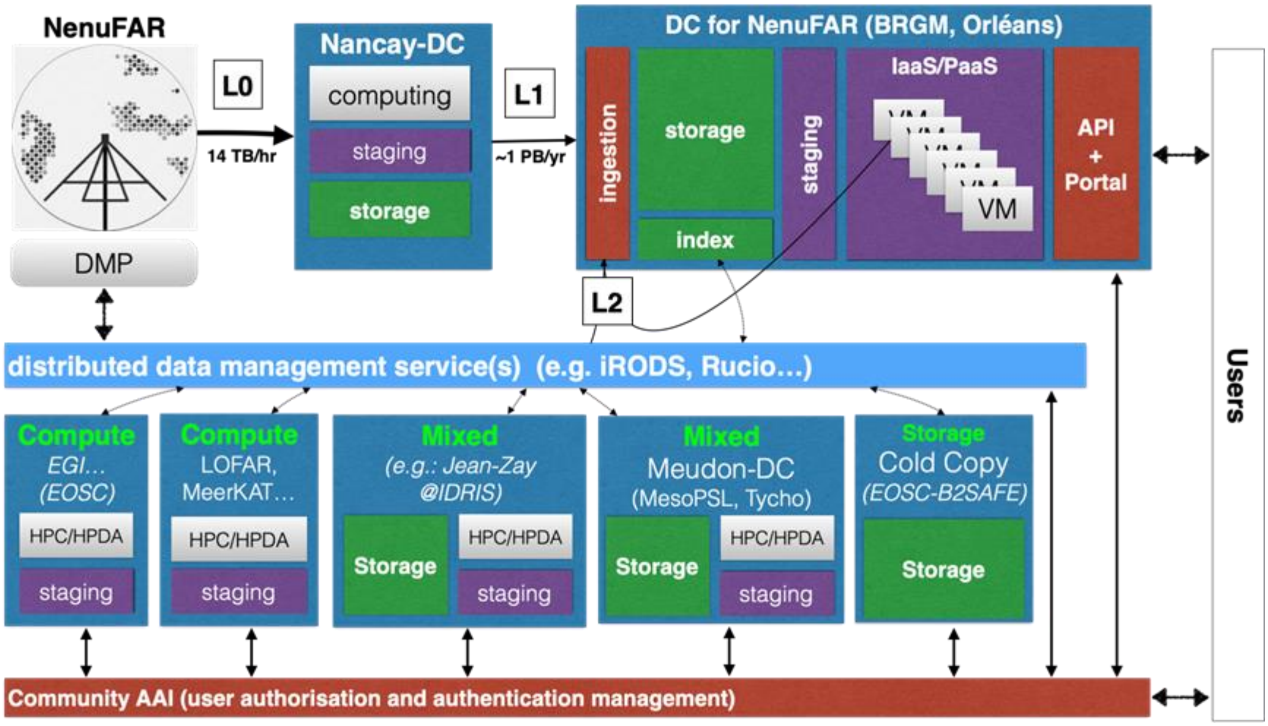
**Use Case A**: Real-time processing supporting the data acquisition, data flagging or transient detection would greatly enhance the efficiency of NenuFAR observation and usability of the NenuFAR data. The NenuFAR team is studying AI pipeline proof of concepts allowing to detect dispersed transients, fine structures, as well as flag radio interferences, leading to data acquisition triggering, or data acquisition cadence selection. The very high data rate currently limits the length of full-band, full-resolution observations with NenuFAR. Using an online pipeline on the raw data stream could lead to radio event detections that could trig a change of measurement mode and receiver parameters (e.g., if a burst is detected, then the frequency and time resolution increases) to record extracts of high value data slices at high resolution. When no event is detected, the receiver will record the data with moderate resolution while still monitoring the arrival of a new event. Such a feature will help selecting only the relevant data and get rid of massive data that do not contain useful scientific information. Within the EXTRACT project, the team will implement a real-time implementation of such prototype, fed by real-time NenuFAR data, producing data acquisition triggers, and trained on existing data collections. In order to further improve the quality of the input raw data stream, an online input data stream validator will be developed allowing to detect faulty antenna on the fly and blank their signal before aggregating the final raw data stream.

**Use Case B:** In the original proposal, smart observation scheduling was proposed but considering the planned infrastructure for EXTRACT and the fact that observation planning is completely in the EDGE part, Case B was considered as unfit for TASKA during semester 1. However, it has been replaced by Case E which is founded on the combination of UCs A, C and D.

**Use Case C**: NenuFAR will produce 1 to 2 PB of reduced imaging data per year. These data need to be ingested, processed and imaged using dedicated software developed by the radio astronomy community. To help in this task, the NenuFAR Data Centre (NenuFAR-DC) will open up late 2023. It will be located in Orleans (France) and developed by BRGM (hosting institution) and Observatoire de Paris. NenuFAR-DC is designed with cloud technologies (OpenStack, CEPH, Kubernetes, etc.). The challenges in interferometric imaging data post-processing require setting up a datalake, distributed on several data centers (of regional, national, or European scales) offering storage and computing capabilities. With EXTRACT, a distributed data processing framework will be developed allowing orchestrated workflows for interferometric image reconstruction, with automated software container staging, data staging, parameterized computing and feedback of data products into NenuFAR-DC. The workflows also include selection of adequate computing facility, authorization, and access granting, as well as data quality and validation. It will implement astronomy interoperability standards where applicable. The developed system should also raise the challenge of the scalability to ingest and process the future SKA data.

**Use Case D**: Furthermore, interferometric imaging pipelines are currently facing two challenges: (a) the sparsity of the visibility's domain (all algorithms in production are re-gridding the visibilities before any processing); (b) the transient (or time-varying) radio sources, which are not easily processed by classical imaging pipelines. Low frequency solar system radio sources are non-thermal plasma instabilities, with intrinsically sporadic underlying drivers, and high dynamical ranges. The detectability of such powerful sources is guaranteed with the Sun. However, when the transient source is faint and not resolved, the detection is a critical issue. On the one hand, high temporal resolution data lead to noisier images that could prevent faint transient radio detection (a.k.a. the "detection" problem). On the other hand, a longer time integration reduces the noise but could insufficiently sample the temporal behavior of the transient source and dilute it (a.k.a. the "dilution" problem). The development proposed in EXTRACT is therefore two-fold: (a) applying Deep Learning techniques to detect and keep track of time varying structures in the interferometric field of view; and (b) developing a trained interferometric imaging technique that could handle spatial and temporal nature and variations of the radio source and produce image cubes accounting for the 3-D structure of the signal. Another improvement of this tool could be to explore the possibility of directly imaging the visibility data without having to grid them in an image. This could help build the fastest possible imaging method that keeps the full resolution of the data. The prototype will be deployed on the infrastructure developed in task C and will be trained using realistic mock data as well as real data generated in task C. The algorithm developed in this task will lead to dramatically improved Space Weather solar imaging capabilities that would guarantee a "fast response" thanks to the fast image reconstruction products.

**Use Case E**: While UC A is focusing on real time processing of high-resolution temporal-spectral data streams, near the telescope, and UCs C and D are dealing with flexible and dynamic interferometric imaging, there is a high scientific interest to go a step further, which is reconciliating high resolution data produced through UC A, and the spatially resolved images produced with UC C and D. One way is to use the DynSpecMS pipeline, developed by Cyril Tasse for reanalyzing LOFAR and NenuFAR data imaging survey, searching for transient signatures of exoplanets or flaring stars. Within EXTRACT, this pipeline will be adapted to be applicable to observations in MS format processed by TASKA. A pattern matching algorithm will also be studied to map the UC C and UC D outputs, possibly processed by DynSpecMS, with the high temporal-spectral resolution produced by UC A. Each feature on the UC A data output would then be spatially enhanced thanks to the UC C and D output. This would provide a unified dataset of very high scientific added value, merging simultaneous products from the two main receivers of NenuFAR.




**Machine Learning Analytics:** The various cases of TASKA, such as TASKA-A (event detection) and TASKA-D (faint sources) are expected to require machine learning (ML). Both cases may require ML analytics such as a combination of a classifier model with a time-series feature. To that end, there are several technologies that will be made available in the context of EXTRACT, allowing both big-data computation and either ML training or serving, in a single workflow. For the sake of performance, i.e., meeting the timing requirements explained further below, it is currently expected that there would be separate workflows for training and serving the models, with a model repository that captures the most up-to-date versions of the models and allows fallback to earlier versions in case of detected errors. Models should be considered and handled as data, and possibly be tracked for consistency – for example, be able to tell which model was trained based on which datasets.

## 4.2. Personalized Evacuation Route (PER) Use Case

### 4.2.1 Background

Urban regions frequently face a wide array of potential disasters such as earthquakes, fires, floods, and acts of terrorism. These adversities represent significant challenges not only for emergency services but also for urban planning and engineering. The need for effective and equitable evacuation plans as part of a comprehensive strategy for community resilience has become increasingly apparent. Decision-makers now acknowledge the importance of preparedness, flexibility, and adaptability in the face of such threats.

From the development of Urban Digital Twins (UDTs), the notion of digital preparedness in emergency scenarios is extending to handheld mobile applications. Existing evacuation mobile apps like the FEMA App, the Emergency: Alerts App by the American Red Cross, and the 112: *Where ARE U* App have provided beneficial features enhancing real-time coordination and safety during emergency evacuations.

According to Westrum, resilience in such scenarios relies on three conditions:

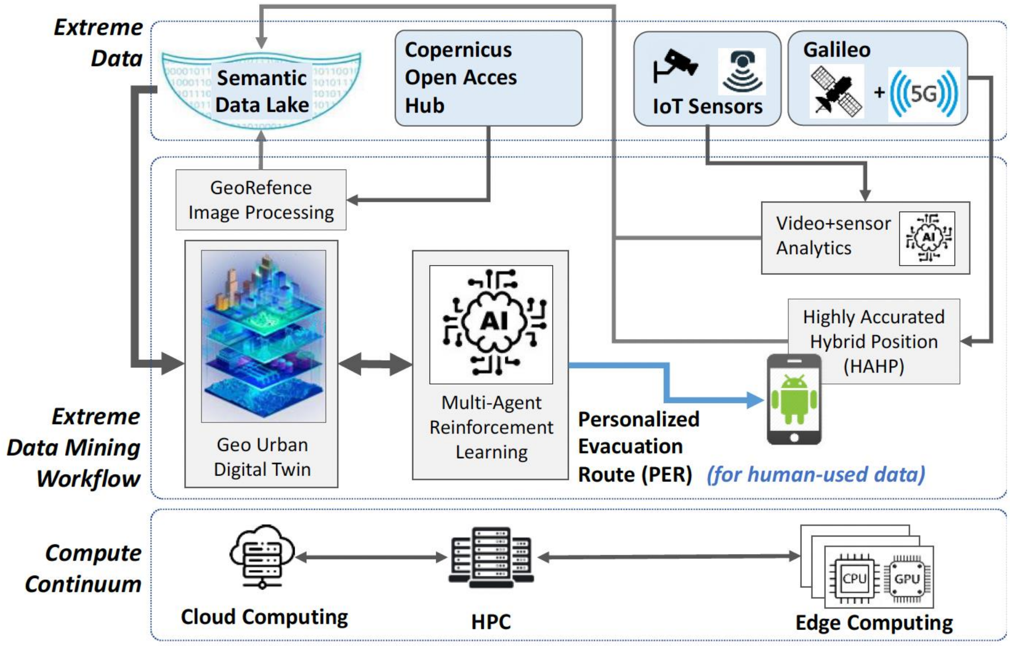
- **Preparedness**: The capability to prevent potential disasters.
- **Flexibility**: A critical characteristic for survival under diverse conditions and compromised situations.
- **Adaptability**: A necessity for rapid recovery from disruptions and restoration of desired performance.

At present, evacuation plans are pre-set, designed based on risk analysis results, with any adjustments or changes reliant on human decision-makers. These individuals often find it challenging to understand the rapidly changing ground conditions and the evolving nature of the event.

Moreover, it's evident that socio-technical systems like a (smart) city heavily rely on a multitude of interconnections. Examples include physical interdependencies (energy, materials, or people flow from one infrastructure to another, often across extensive and diverse supply chains), cyber interdependencies (transmission or exchange of information), geographical interdependencies (close spatial proximity of infrastructure elements), and human interdependencies (distributed decision-making within structured organizational processes). Such interdependencies encompass stochastic mechanisms of early warnings, panic-like defensive behaviors, and their cascading effects on a broader public level. These interconnections necessitate detailed representation and modeling.

In line with these principles, the Personalized Evacuation Route (PER) use-case employs the EXTRACT platform to develop, execute, and maintain a data-mining workflow. This workflow integrates high-precision positioning, AI-based computer vision, deep reinforcement learning, and an Urban Digital Twin (UDT). The goal is to generate personalized evacuation routes for each individual, available via a mobile app (Evacuation Mobile App - EMA). This app processes and analyzes data from diverse sources, including Copernicus and Galileo satellite data, citywide IoT sensors, 5G mobile signals, and a semantic data lake consolidating all this information.

While the comprehensive approach outlined above is a significant step forward in evacuation planning, it's critical to note that this methodology has its limitations. The reliance on a multitude of data sources and the processing of extreme data volumes can pose challenges regarding data quality, consistency, and timeliness. Additionally, the adoption of high-level AI models, while promising, demands rigorous testing and validation to ensure their performance in dynamic and unpredictable disaster scenarios. It's also essential to consider the scalability and responsiveness of such systems across various urban contexts and disaster types. Therefore, continuous review and improvement of these systems are necessary for enhancing their effectiveness and resilience.




## 4.2.2 Description of the Dataset from the Smart Control Room and its Role in the Use Case Scenario

The data generated by the Smart Control Room (SCR), a Venice operation center run by the local police in cooperation with strategic city offices like local transport services, plays a pivotal role in enabling the functionalities of the EMA application under the P.E.R. use case. This dataset is derived from recent historical records within a specific timeframe and captures crucial information, including temporary elements such as construction sites or gangways, which might obstruct evacuation routes. that helps to model the city and evaluate the best evacuation paths.

The SCR dataset includes several datasets concerning the number of people present in Venice along with geo-referenced layers that, when overlaid, provide an accurate representation of the city's status. However, these datasets are not real-time due to their close ties to police activities and city management; therefore, the EXTRACT application will rely on simulations.

The dataset captures information from different sources such as pedestrian flow sensors located in the heart of Venice, cameras monitoring water traffic on the

Grand Canal, and data on presence collected by analyzing distinct mobile phone numbers within the city. It also contains georeferenced data layers including pedestrian areas, green areas, lagoon canals, buildings, house numbers, water zones, and street names.

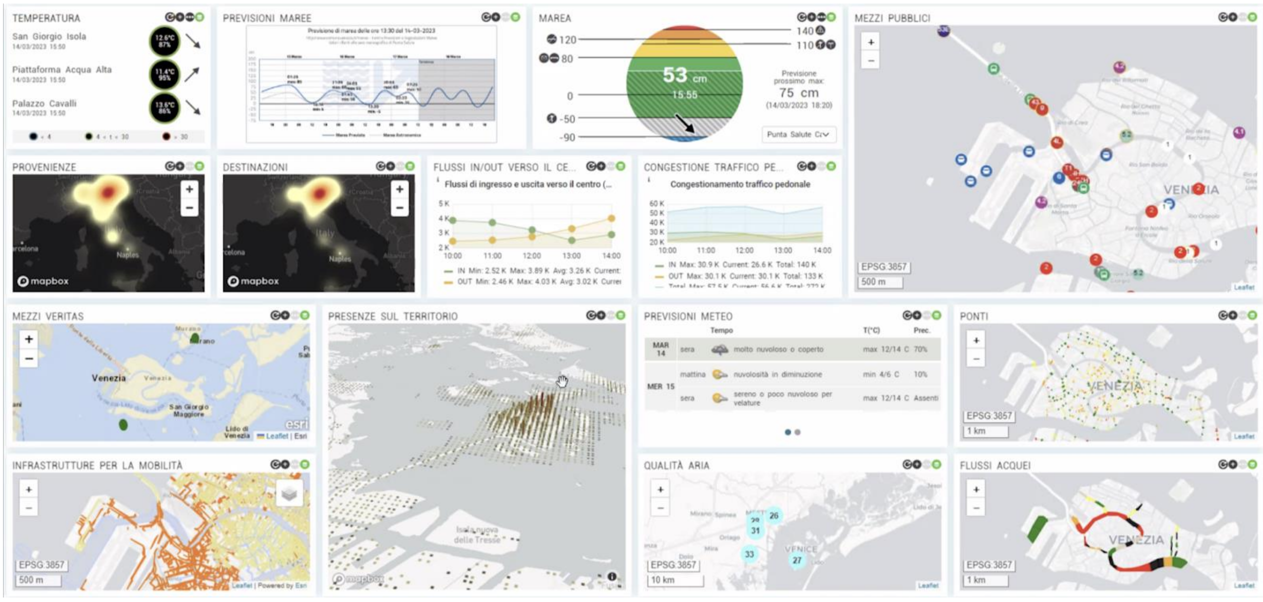



*Figure 1: The Smart Control Room Dashboard - A screenshot*

The SCR dataset thus contains:

1. Pedestrian flow data: More than 19,000 records from 10 sensors in the Rialto area, measuring the number of people entering or leaving the monitored area.
2. Water traffic data: Over 3,800 records from 2 cameras monitoring the Grand Canal, capturing information about types of vessels and their direction.
3. Presence data: By analyzing unique mobile phone numbers, the dataset estimates the number of people in the city, updated every 15 minutes and aggregated into 150m edge squares. More than 1 million records providing different insights of the data are available.
4. Geo-referenced data: Layered geo-referenced shapes, including pedestrian areas, green areas, lagoon canals, buildings, house numbers, water zones, and street names.

These data come in various formats: the georeferenced data are provided in .zip files containing "shapefiles" and other metadata files. The Presence data comes in several CSV files, each with a specific header structure chosen to best represent the detail level. The Pedestrian Flow data is provided in JSON format with data aggregated by sensor.

A distinctive feature of this dataset is its "extreme" nature, characterized by sudden spikes in data, such as a rapid increase in people's movement (flux) due to the evacuation or a sudden rise in hazard levels due to the spreading fire. These extreme data points are particularly crucial as they can trigger adjustments in evacuation strategies, making the data from the SCR instrumental in implementing efficient, dynamic, and safe evacuation routes.

### 4.2.3 Description of Urban Digital Twin Datasets and their Role in the Use Case Scenario

The concept of the Digital Twin, a virtual representation of a physical entity, has taken a leap forward in recent times. Initially introduced by Michael Grieves, Chief Analyst at the Florida Institute of Technology, in 2001 during a Product Lifecycle Management course, this concept has found extensive utility in industrial processes. In the context of Urban Digital Twins (UDTs), the digital counterpart isn't an exact replica of the city, but rather a sophisticated abstraction thereof. This reflects a classical dilemma in the field of modeling - models always maintain a certain level of abstraction. Implementing a UDT is a complex process that requires multidisciplinary expertise, including data science, urban planning, computer modeling, and software development. It also involves handling large volumes of data generated by a city. Successful implementation requires collaboration among a variety of stakeholders and domain experts. The broad steps include defining objectives, data collection and integration, model development, visualization and user interface development, simulation and analysis, stakeholder engagement, iterative development, data management and governance, scalability and integration, and evaluation and monitoring. UDTs have found diverse applications, and according to a study, the identification of decision-makers' informational needs is the first step in the data strategy to support an evidence-driven adaptive cycle.

### 4.2.4 UDTs in Emergency Preparedness and Response Strategies

In the context of the P.E.R. use case, we're interested in exploring instances where UDTs have played a crucial role in emergency preparedness and response strategies, including evacuation. Cities such as Singapore, Barcelona, New York, Helsinki, and Rotterdam provide compelling examples of how UDTs can be leveraged for effective evacuation management. In these cases, UDTs help simulate and optimize evacuation routes, identify potential bottlenecks, assess the effectiveness of emergency response plans, and more. They offer a dynamic and data-driven environment for decision-makers, aiding them in making informed decisions that ensure the safety and efficiency of evacuations during emergencies.

### 4.2.5 Fire Scenario Selection: Lessons from Venice's History

As a segue to the principles and objectives outlined for the PER use-case, we introduce the choice of our primary scenario for testing applications within the EXTRACT project - a fire hazard. This selection is not random; it draws from history and the compelling need to address immediate, visible threats inherent in fires, threats that have in the past posed significant risks to urban centers, notably Venice.

In 1996, Venice witnessed the tragic fire at the Fenice opera house, a prominent cultural and historical landmark. This catastrophe emphasized the city's vulnerability to fire hazards and the cascading effects of such events on its dense urban fabric, infrastructure, and populace. The hazardous factors associated with fires, such as dense smoke, can severely impair visibility and render navigation through narrow, smoke-filled, and unfamiliar streets exceedingly challenging.

*Figure 2: 1996 Fire at the Fenice Opera House*

These situations may elicit a range of human reactions, from paralysis induced by fear to hasty and uncooperative escape attempts.

In this scenario, the SCR dataset would capture the moment the fire breaks out and starts to spread. It would monitor and report the fire's location and severity in real time, enabling the EMA to define the danger zones dynamically. One of the critical elements in this dataset is the information about the movement patterns and average speed of people, both residents and tourists. These patterns, particularly during the 24 hours preceding the hazard, can provide valuable insight into potential evacuation routes.

Once a threat is detected, PER begins to calculate the safest evacuation routes, taking into consideration a variety of situational factors and constraints posed by the event. As a part of this strategy, real-time data on the evolving ground situation feeds into the AI module that guides route calculation.

With the EMA app, users are provided with these safe routes in a navigation mode, complemented by multi-sensory signals like sound, flashlights, and vibrations to guide them amidst the high-stress environment and minimize navigation errors.

The system has to continue to evaluate the situation, helping to refine the safety route, accounting for potential deviations by users or changes in the ground situation.

In creating an optimal safety route, the system estimates the number of people in different locations (start point, route, destination point) by considering people density per m2. This dynamic approach allows for adjustments as the event unfolds, thus honoring the principles of flexibility and adaptability in disaster response strategy.

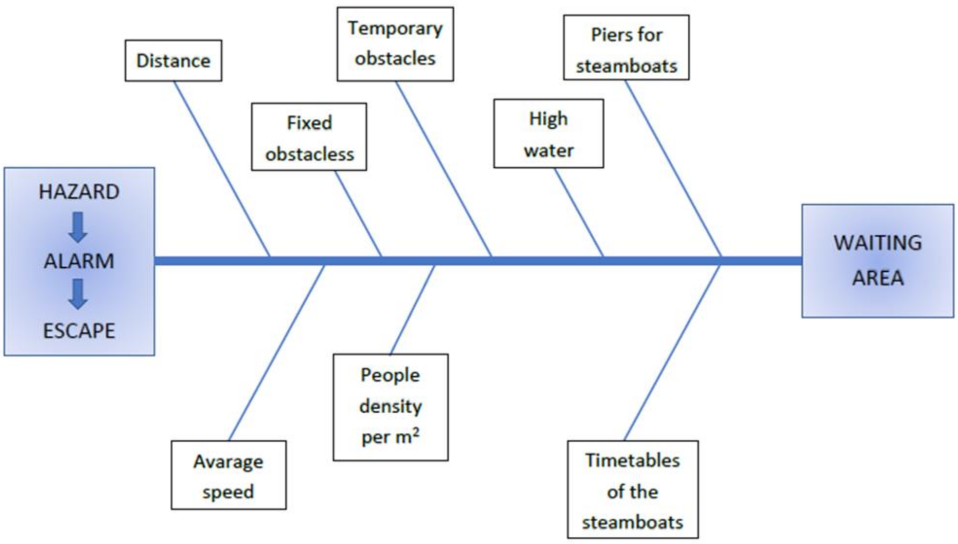
With its comprehensive, dynamic, and adaptable design, the PER use-case presents an innovative solution to urban evacuation challenges posed by extreme events such as fires, building on historical events and contemporary technology to protect urban centers and their populations.

### 4.2.6 The Ishikawa Diagram and its Application to the Venice Use-Case

as the cause-and-effect diagram, the Ishikawa diagram is an important tool in the management of brainstorming sessions. Its primary function is to identify possible causes and effects of a problem, a technique that is highly relevant to our P.E.R. use-case.

In the context of Venice, the application of the Ishikawa diagram to P.E.R. allows for an examination of a multitude of factors that may impact the successful evacuation of individuals from hazard zones to safe areas. The diagram gives an explicit depiction of the complexities involved in creating a P.E.R. during emergencies.

It serves as an essential aid to pinpoint areas that need more data or predictive modeling, potential bottlenecks, and aspects that need special consideration for successful evacuation.



### 4.2.7 Emergency Management: Evacuation Challenges for Venice

Venice's unique characteristics pose significant hurdles for effective evacuation. These challenges can be grouped under six main headings.

**Aging Infrastructure**

Venice's intricate network of narrow streets, historic buildings, and canals are aged and fragile. These structures may lack modern fire prevention and suppression systems, posing fire safety challenges.

**Limited Accessibility**

The narrow streets and bridges of Venice pose a hurdle for quick accessibility by emergency vehicles. This limitation can impede firefighting efforts and slow down the evacuation process during an emergency.

**Water Supply Issues**

Surrounded by water, Venice depends on its lagoon for firefighting water supply. However, the fluctuating water level in the lagoon can affect the availability of water for fire suppression.

**Unique Urban Layout**

Venice's densely built urban environment, interconnected buildings, and narrow streets contribute to the rapid spread of event consequences, such as fires. The city's layout poses significant risks to its residents and visitors.

**Limited Escape Routes**

The city's complex network of canals and limited number of bridges restrict escape routes during an evacuation. This makes it challenging for individuals, especially those unfamiliar with the city’s layout, to reach safety quickly.

**Overcrowding**

Venice hosts over 23 million tourists each year, leading to a constant state of overcrowding. In a crisis, these unfamiliar tourists are likely to follow crowd flow, complicating crowd management and evacuation processes for the P.E.R. strategy. A significant number of city users (e.g. employees or students that live in the mainland but work and study in the historical center) is always present in the city and while they are familiar with it, they contribute to the overall overcrowding.

**Evacuation challenges**

The foundational study for the P.E.R. use-case within the EXTRACT project was the EU-funded eVACUATE[1] project, even if solution provided by EU projects as SAVE-Me and RESOLUTE have been on also considered. eVACUATE laid the groundwork for designing evacuation strategies in pedestrian-only, medieval cities heavily visited by tourists. eVACUATE provided invaluable insights into integrating various technologies for effective evacuation management, especially in cities like Venice with unique infrastructure and crowd heterogeneity. This foundation offers invaluable background for the development requirements.

The evacuation model can be classified into three categories: macroscopic, mesoscopic, and microscopic. Macroscopic models primarily offer moderate

[1] https://cordis.europa.eu/project/id/313161/reporting

estimations for safe egress time and rescue durations without taking into account individual behavior or other human-related factors during emergency situations.

These models are predominantly utilized for evacuation planning, proving valuable for high-level analysis of large-scale events such as hurricanes, flooding, or accidents at nuclear power plants. Conversely, mesoscopic models have been employed in evacuation planning to represent congestion conditions and temporal effects more accurately than macro models, albeit requiring coverage over a wider geographical area. Therefore, mesoscopic planning models have developed as a response to the need for detailed information required by microscopic simulation programs and the analytical precision not achievable with macroscopic models.

Finally, microscopic models enable accurate modeling at the individual level, that facilitates more advanced monitoring capabilities. In PER use case we adopted microscopic models to manage personalized evacuation.

Despite advancements in evacuation models concerning optimization (real-time) and effectiveness (optimal rescue plans, optimal evacuation routes), limitations persist in incorporating algorithms that handle multiple sources and destinations, as well as real-time adaptation to the evolving nature of disastrous events due to various factors, such as final decisions made by individuals and dynamic changes in properties affecting the evacuation plans for both rescuers and individuals.

Moreover, evacuation models hold a dual purpose as both engineering and research tools. Within an engineering framework, these models contribute to the safety assessment of buildings. They focus on estimating the Required and Available Safe Egress Time (RSET/ASET), predicting congestion levels, and highlighting potential design complications. On the research front, evacuation models foster an improved comprehension of human behavior during emergencies. They empower researchers to generate new hypotheses that can later undergo empirical testing. The more we understand human behavior in emergency scenarios, the more effective evacuation models become.

When conceptualizing the P.E.R. use-case, the Metropolitan City of Venice's research team identified a link between the literature on crowd behavior, the modelling and simulation of pedestrian flows, and studies on evacuation processes. This use-case adopts a forward analysis perspective, integrating requirements that address both "dangers-critical elements" and "opportunities-favorable elements."

"Danger-critical elements" are integral to the P.E.R. use-case. The application needs to pinpoint and map these elements, subsequently designing a route that circumvents them. Examples of potential dangers could include high crowd density on the streets, among other critical factors. Determining crowd density levels from "not dangerous" to "very dangerous" is essential. This includes considerations such as the proximity to safe waiting areas, the physical condition of the individual, and the escape velocity if a person is slow.

Conversely, some factors can be categorized as "Opportunities-favorable elements." Aspects that may seem detrimental in some situations can be advantageous in others. Examples of this include a fast individual escape velocity or a person being in good health. Additional favorable elements might include safe waiting areas offering respite, medical treatment, and a temporary refuge before individuals can return home, and temporary extraction points, such as water bus stops with connecting services.

Understanding crowd dynamics is pivotal to managing the ebb and flow of crowd movement. A key area of focus is the study of individual movements within a crowd, fundamental to forecasting emergency evacuation scenarios. However, empirical studies on this subject remain limited.

It's important to discern significant drops in crowd flow before access or egress from large-scale events. This will allow for the prediction and management of "stop and go waves," a phenomenon that could lead to a dangerous domino effect of escalating crowd density and "crowd turbulence." For this reason, ensuring pedestrian facilities operate below their maximum capacity can significantly improve crowd safety.

In light of available literature, we propose dividing the issue of "crowd movement within the narrow streets of an ancient medieval city" into two factors: crowd density in the streets (the number of people per square meter), and the speed of an individual (per meter/second).

Various authors have described crowd dynamics as the interplay between density and speed. Four distinct dynamics emerge from these relationships, each representative of different crowd densities and speeds:

1. "Fluid Movement Effect" - A scenario with 1-2 people per square meter, capable of moving at a speed from 0.8 up to 1.2 meters per second.

2. "Stop and Go Movement" - A situation with 3-4 people per square meter, moving at speeds ranging from 0.4 to 0.8 meters per second.

3. "Congested Effect" - Here, 5-6 people occupy each square meter, moving at a speed of 0.4 to 0.5 meters per second.

4. "Turbulence Effect" - In this most critical scenario, more than 6 people per square meter are moving at up to 0.4 meters per second. However, movement is non-linear as people seek any escape route, even moving backward, leading to turbulence in escape trajectories

Understanding these crowd dynamics is a critical part of the requirements gathering process for developing the P.E.R. application. A practical and reliable evacuation app must account for all these dynamics, adjusting in real-time as circumstances change. In the most severe case, the "Turbulence Effect," the app must be capable of quickly identifying non-linear movements and suggesting alternative escape routes, thereby minimizing the potential for harm.

In summary, for the creation of an optimal safety route one must be able to determine the number of people in all places (at the starting point, in the safe route itself and at the arrival point). This can be calculated by multiplying the people density per m2 - according to some research, the maximum/optimal limit to “feel fine” corresponds to four people per m2. It is possible to design all of this, as mentioned above, especially considering the average speed, which must be already identified in the 24 hours preceding the hazard.

## 4.2.8 Obstacles to be Considered

From the starting to the arrival points, account must be taken of the distance separating them and, therefore, of all kinds of obstacles that can be encountered, including those 1) fixed, such as bridges and bottlenecks, and those 2) temporary, such as gangways and construction sites. These datasets are also georeferenced by the Civil Protection, which could turn out to be useful to be implemented in the map of the application as well.

## 4.2.9 Multi-Agent Reinforcement Learning in PER

To determine the optimal escape route in the Venice scenario, we will develop a MultiAgent Reinforcement Learning Model (MARL). Each agent will be associated with a person's smartphone and consider factors such as escape route and environment condition, agent's velocity, group size and route bandwidth. Personalized best paths will be provided to each agent, which may change dynamically due to evolving input data.

We consider the problem of finding the optimal route for all the residents of a city to a safe zone where an extreme emergency has occurred (e.g., a fire). Given a digital twin and satellite images of the city and GPS positions of all its residents, the routing must be calculated.

Reinforcement Learning (RL) is a framework in which an agent interacts with an environment via trial and error with the goal of learning a policy that maximizes cumulative reward, a measurement of the agent's performance. This framework has shown promising results, such as its application in robotics [2], or dynamic ambulance redeployment[3].

Multi-Agent RL (MARL) is the extension of RL with multiple agents interacting simultaneously with the environment. There is a clear parallelism between cooperative MARL, where the agents cooperate to achieve the same goal, and the proposed emergency routing.

The environment could be represented by the digital twin of the city. In this case, the agents would represent the city's different people (or groups of people). A reward function could be built that considers the following elements:

1. Arrive in the minimum amount of time possible.
2. There must be no casualties.
3. Consider the attributes of each person (age, disabilities, etc).

The different agents would interact with the digital twin to learn a joint policy that maximizes such reward.

In the case of a catastrophe in real life, a digital twin of the city should be created by comparing it to the satellite images (Copernicus) and populate it with the residents according to GPS locations (by means of Galileo and 5G) in their devices. Then the

[2] Kober, Jens, J. Andrew Bagnell, and Jan Peters. "Reinforcement learning in robotics: A survey." The International Journal of Robotics Research 32.11 (2013): 1238-1274.

[3] Ji, Shenggong, et al. "A deep reinforcement learning-enabled dynamic redeployment system for mobile ambulances." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3.1 (2019): 1-20.

MARL solution would be trained. We identify two possible paradigms for this training:

1. Centralized training and decentralized execution. The training in the digital twin will be centralized i.e., containing all the other residents' locations, then the agents will be sent to their respective phone devices and an inference will be executed in the same device without considering others' people position.
2. Centralized training and execution. In this paradigm, all the agents' actions will be computed in the cloud both in inference and training. This solution is more reactive than the first as there will be communication with agents in real-time. A mix of both paradigms can be implemented, for instance sending the location of all residents to each device periodically.

Moreover, the solution should be delivered as fast as possible. To reduce the significant training times that RL training needs, we could leverage two available solutions:

- High-performance capabilities to provide the required speed of such a large-scale training.
- Meta-learning or learning to learn. The combination of meta-learning and RL trains on different variations of the same problem such that in a new unseen variation, the training would be sped up.

Meta-learning would be paramount, as it would also be necessary if the city changes its structure during the evacuation. For instance, some roads may become blocked during the routing. Such environments are known as non-stationary environments and one of the possible solutions is meta-learning such that the multiple agents can adapt fast enough.

### 4.2.10 Integration of Galileo HAS and 5G

The synergistic integration of Galileo's High Accuracy Service (HAS) and 5G cellular technology is central to achieving high precision user positioning in the context of the P.E.R. use case. The Galileo HAS, as part of the Galileo Open Service, provides real-time Precise Point Positioning corrections that significantly enhance user positioning accuracy worldwide. This is made possible through access to raw signal data from the Galileo satellites, accessible via the Android Raw Global Navigation Satellite System (GNSS) API.

Despite this high accuracy offering, complications arise when navigating "urban canyons", urban environments with densely built-up areas, high structures, and narrow streets, where a clear line of sight to multiple satellites is often not possible. Venice, with its uniquely narrow lanes and densely packed architecture, exemplifies such an environment. The obstructive urban landscape introduces significant positioning errors, as buildings in proximity cause distortions and reflections in the satellite signal paths.

To overcome the problems deriving from the partial or non-reception of satellite data, 5G technology is exploited, in order to mitigate and improve position information. The integration of 5G technology becomes indispensable for a solid and continuous position awareness even in contexts where other data sources lose reliability. 5G technology, with its superior data transmission capabilities, contributes to precise positioning by introducing new positioning methodologies. These include multi-cell round-trip time (multi-RTT), multiple antenna beam measurements (UL AoA), and downlink angle of arrival (DL AoD) estimates. These methodologies have been proposed and tested for their potential to significantly improve positioning accuracy in urban landscapes.

The coalescence of GNSS and 5G technologies brings forth the prospect of delivering accurate, secure, and resilient Position, Navigation, and Timing (PNT) services, which are indispensable for the successful implementation of the P.E.R. use-case. This joint operation not only enhances accuracy but also increases the security and resilience of the navigation system, thereby making the system more reliable for the users.

Within the EXTRACT project, this dual-technology approach is operationalized with up to 10 Android mobile phones that will fuse Galileo+5G data for precise positioning. To simulate a larger number of users interfacing with the P.E.R. system and properly stress-test the EXTRACT platform, the project will also leverage a simulator tool developed by ERI. This tool will generate simulated 5G position data, enabling the creation of truly extreme datasets that will feed the “semantic data lake”, thereby ensuring the platform's robustness and preparedness for real-world application scenarios.

The integration of 5G technology within the EMA will take place by exploiting the possibilities made available by the Android operating system: once access to the 5G network has been obtained via the hardware component of the smartphone, it will be possible to interface with the data obtained from the network via APIs internal to the operating system. Access to information and their management can take place both through the tools natively available by the operating system and possibly through modules developed ad hoc based on operational needs.

### 4.2.11 Expected Results of the Venice Use-Case

The primary objective of the PER use-case is to create a robust evacuation system capable of adapting to Venice's unique characteristics. The system should not only offer a set of best individual safety routes, but it should also dynamically adjust these routes based on real-time data, including changes in hazard intensity, crowd density, and physical obstacles.

An effective evacuation system would function as a smart and dynamic safety network. It would provide multiple evacuation safety routes leading to different waiting areas to distribute people more effectively, thus minimizing the risk of overcrowding at any single location.

The framework should be able to process large volumes of data from various sources and apply advanced machine learning techniques, specifically a MultiAgent Reinforcement Learning Model (MARL), to determine the optimal escape paths. This involves consideration of Venice's physical layout, crowd movement dynamics, and other potential variables.

Furthermore, the system should be compliant with strict data privacy regulations, ensuring the protection of personally identifiable information, even while georeferencing all individuals in Venice at the moment of hazard detection.

In sum, a successful outcome for the PER use-case would be a responsive, adaptable, and secure evacuation system that optimizes citizen safety during emergencies while respecting privacy laws. The true measure of success, however, would be the system's effectiveness in actual emergency situations, which is something that would be evaluated post-implementation.

## 4.3. Stakeholders, system actors and roles

### 4.3.1. Stakeholders based on the Use Case Scenarios

The following stakeholders are defined in accordance with the two use cases.

In the context of EMA - P.E.R. Use Case, developed under the EXTRACT project, a range of stakeholders could be significantly interested due to the potential benefits offered in managing complex evacuation scenarios. This robust solution promises to streamline emergency response, improve safety, and enhance overall disaster management. Here's a snapshot of key stakeholders and their interests.

- **Transportation Officials**: Interested in efficient crowd management during emergencies, ensuring a smooth flow of evacuees and minimizing traffic disruptions.
- **Metropolitan City Councils**: They could utilize the EMA to bolster city safety measures, providing citizens with reliable evacuation plans and improving overall urban resilience.
- **Social Groups**: Community organizations and nonprofits focused on public safety could adopt the EMA to enhance their disaster response initiatives and protect vulnerable communities.
- **Decision-makers (Mayors, County Commissioners)**: They would value the evidence-based approach to disaster management to inform policy, ensure public safety, and reduce potential damage.
- **First Responders (Police, Fire Rescue, Emergency Medical Staff)**: The tool's real-time updates and efficient route planning could assist in effective and rapid response during emergencies.
- **Public and Private Companies, Organizations, Institutions**: As part of their disaster management strategies, these entities could leverage the EMA to ensure the safety of their personnel and protect their assets.
- **Service Providers**: Businesses providing essential services such as utilities, telecommunications, healthcare, and food and water supplies can utilize EMA in their continuity planning to minimize disruptions and maintain their service delivery during and after the evacuation.
- **Embassies and Consulates**: They could use this tool to safeguard their staff and visitors, while also using the gathered data to inform their countries about safety protocols in the event of a disaster in Venice.
- **Ministries (Italian Embassy, Foreign Ministry - La Farnesina, Ministry of Tourism - Tourist Heritage Agencies, Ministry of Cultural Heritage - Heritage Emergency Task Force)**: These institutions could utilize the benefits of the EMA to optimize their emergency protocols, ensure the safety of cultural and tourist sites, and provide evidence-based recommendations for heritage protection.

In summary, the range of stakeholders for EMA spans multiple sectors, all united by the common goal of optimizing emergency response and improving public safety. This broad interest underscores the potential of the EMA to transform evacuation strategies in complex urban environments.

**TASKA**: the core group of stakeholders for the TASKA use case are academic/higher education and research/governmental organizations supporting the development and operations of NenuFAR.

**Research Institutions and Universities**: Researchers and scientists can greatly benefit from the comprehensive data and models provided by TASKA, enhancing their understanding of asteroid trajectories and aiding in the development of mitigation strategies.

**Astronomy Community**: Both professional and amateur astronomers would find TASKA beneficial for tracking and studying asteroids. It can provide accurate, up-to-date information that supports observational efforts and furthers our knowledge of near-Earth objects.

NenuFAR allows for spectro-dynamic studies of variable radio sources such as pulsars, magnetized exoplanets, storm flashes in the atmosphere of giant planets of the Solar System, erupting stars or other transient phenomena, such as Fast Radio Bursts (FRB) or radio emission associated with atmospheric spray phenomena caused by very high energy particles or radiation. In imaging mode, one of the most exciting topics of the instrument is the mapping of the power spectrum of the 21-cm signal produced by the intergalactic medium during the dark ages of the universe, at the time of the formation of the very first stars and galaxies.

NenuFAR is an ambitious radio astronomy project for exploring the lowest frequencies observable from the ground with the most efficient technologies currently available. Radio astronomy at low and medium frequencies is a field in full explosion, with instruments as LOFAR and NenuFAR in France and in Europe, or as MeerKAT, ASKAP, MWA in South Africa and Australia. The exploitation of these radio telescopes by the astronomical community demonstrates very clearly that the radio astronomy of the XXI century is no longer the exclusive business of radio astronomers. The provision of pre-processed radio data and calibrated de facto involves the entire astronomical community.

**Higher Education Institutions**: TASKA can serve as a valuable educational tool, providing data and models that aid in the teaching of astronomy, physics, and related subjects. It offers a practical resource for students to engage with real-world data and improve their understanding of asteroid tracking and prediction.

**Service Providers**: Companies providing cloud-based services, high-speed data processing, and advanced computational tools can find value in partnering with TASKA, leveraging their technology to facilitate complex calculations and data analysis.

**Non-profit Organizations**: Non-profits focusing on science education and space exploration can use TASKA as a resource to foster interest and learning in these fields. They can also leverage the tool in fundraising efforts aimed at supporting scientific research and education.

The main laboratories supporting the development are LESIA (Observatoire de Paris - PSL / CNRS / Sorbonne Université / Université Paris Diderot), the scientific unit of Nançay - USN (Observatoire de Paris - PSL / Université d’Orléans/ CNRS) and LPC2E (Université d’Orléans / CNRS / CNES).

The development of NenuFAR has been supported in terms of personnel and funding by the Radioastronomy Station of Nancay, CNRS, the Observatoire de Paris - PSL, the University of Orléans, the Observatoire des Sciences de l’Univers en région Centre, the Région Centre-Val de Loire, the DIM-ACAV and DIM-ACAV+ of the Île-de-France Region, and the Agence Nationale de la Recherche.

Additionally, NenuFAR is a LOFAR "superstation". LOFAR consists of approximately 50 antenna arrays (or "stations") distributed throughout Europe connected by high-speed internet to a calculator in the Netherlands. NenuFAR is – together with LOFAR – labeled "National Research Infrastructure" by the French Ministry of Higher Education, Research and Innovation.

Finally, the SKA Organization has officially recognized NenuFAR as a Pathfinder Project of the SKA telescope. NenuFAR helps building the French SKA users' community and building experience at very low radio frequencies. In 2021, the French government announced the decision of France's entry into the SKAO, which has integrated the 2021 National Roadmap for Research Infrastructure in the category of the Scientific Intergovernmental Organizations.

**Public Sector Entities (Space Agencies, Disaster Management Authorities)**: These organizations can use TASKA's predictive capabilities to enhance their readiness to space weather events such as bursts in the solar emission, helping to inform public policy and disaster response planning.

In summary, the range of stakeholders for TASKA spans multiple sectors, all united by the common goal of understanding and mitigating the risk related to space weather. This broad interest underscores the potential of TASKA to make significant contributions to both scientific research and public safety.

## 4.4. Applications environments

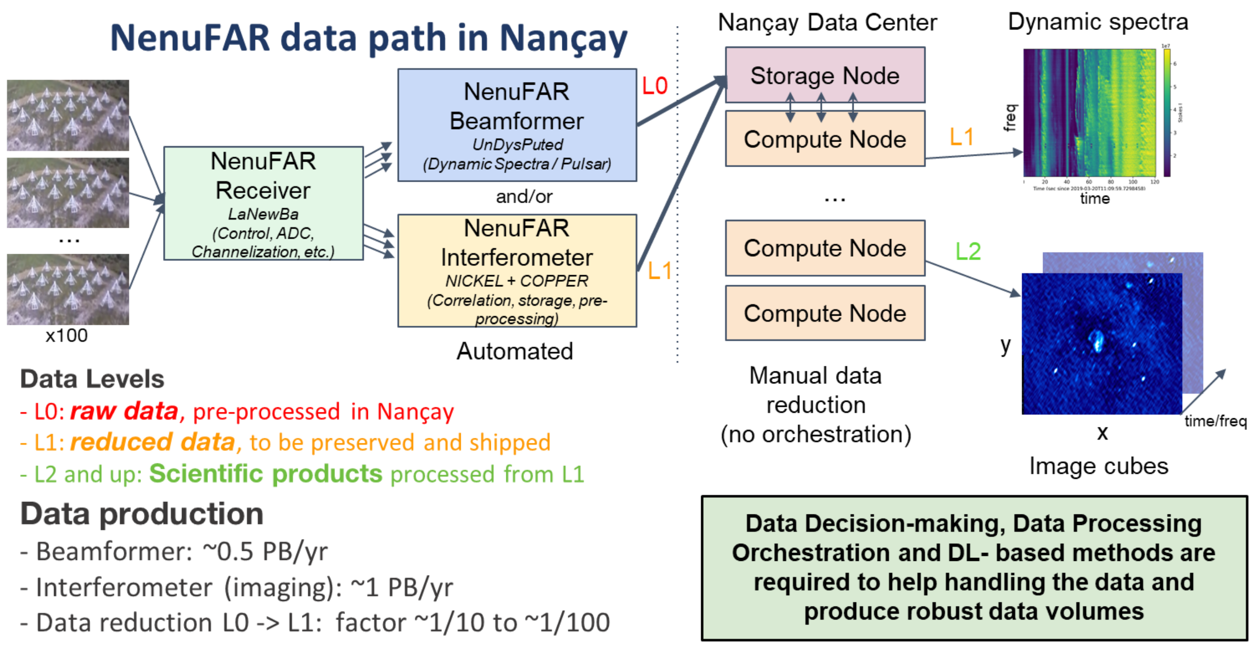
### 4.4.1. TASKA

**TASKA**: The TASKA use-case will be deployed and executed on a heterogeneous continuum composed of: (1) A set of edge devices (30 FPGAs,10 GPU, and a several CPUs counting 100 cores in total) located next to the radio-antennas ; (2) a storage space of 500 TB hosted at OBSPARIS; and (3) a HPC and a cloud infrastructure implement in the NenuFAR Data Centre (DC), hosted in Orleans (France). The NenuFAR-DC, that will open up early 2023, is being developed by OBSPARIS. Moreover, the EXTRACT continuum infrastructure will be part of the European Open Science Cloud (EOSC) [EOSC] environment to support EU science, on which the NenuFAR and the future European SKA Regional Centre frameworks will rely on. In that regard, EOSC distributed resources like AAI (Authentication and Authorization Infrastructure) to manage access to data by EU researchers will be included into the EXTRACT platform.

We recall below the main characteristics of the data considered in TASKA, as per DoA:

| Source | Volume | Speed | Variety | Dispersed | Complex. | Incompl. |
|---|---|---|---|---|---|---|
| **Beamform** | up to 16 TB /obs. | 600 MB/s | | | Mixed with RFI, no image, sources in secondary lobes | |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Waveform** | 350 GB for a 5s buffer | 102 * 5.5 Gb/s | | 102 spatially dispersed arrays | Mixed with RFI, no image, sources in secondary lobes | |
| **Imaging*** | 45 TB /obs. | 5.6 TB/h | | | Mixed with RFI, no image, sources in secondary lobes | Sparse sampling (spatial/ temporal) |
| **Data Lake** | 1-2 PB/year | | Several obs. modes, many obs. targets | Distributed storage and computing | Calibration and Image/Video restoration | Sparse sampling (spatial/ temporal) |




*The imaging data follow the "Measurement Set" (MS) standard created by the NRAO. Each MS contains multi-dimensional arrays (time, frequency, correlation) as well as metadata giving the context of the observation. Each MS is a self-contained data unit that should not diverge from the standard.

### 4.4.2. P.E.R. Use Case

The P.E.R use-case will be implemented and run on a heterogeneous continuum encompassing:

(1) A network of edge devices, comprised of smartphones running the EMA application, and an array of IoT devices spread across the city contributing high-dynamic data; (2) a central storage system housing critical city data such as GIS information, Copernicus

data, and emergency response protocols, hosted at the local data center; and (3) a cloud-based computational infrastructure that supports the Urban Digital Twin and the AI-driven route calculation system.

Similar to the TASKA case, the PER use case operates within the EXTRACT continuum infrastructure, which forms part of the European Open Science Cloud (EOSC) environment [to support EU-wide disaster management efforts. In this context, distributed resources from EOSC, such as the Authentication and Authorization Infrastructure (AAI), will be incorporated into the EXTRACT platform to manage data access by EU researchers and emergency response personnel]

We recall below the main characteristics of the data considered in PER, as per DoA:

| Source | Volume | Speed | Variety | Dispersed | Complex. | Incompl. |
|---|---|---|---|---|---|---|
| **Copernicus** | Avg. 1TB per image | 600 MB/s | | | Dense information and reach data structure | |
| **Galileo + 5G** | | 10Gbps 1ms/ latency | | Spatially dispersed users | | |
| **IoT sensors** | TB day | 15/20 fps 1 read/sec | Multiple data formats, models | Spatially dispersed sensors | Multi-vendors, multi-data formats, different rate, quality, signal meanings, weak standardization | Sparse sampling (spatial/ temporal) |
| **Data Lake** | Avg 5-10TB/day | Very low latent spatio-temporal semantic query | Intensive ETL processes for ontology-based transf. | Multiple data sources and formats | Abstraction, semantic do-main representation, implicit references | Semantic Data Lake |

# 5. Stakeholders' Requirements Investigation

## 5.1. Methodology

There are a lot of methodologies to use when eliciting requirements. The one followed here is the Volere methodology [1], taking into account the relevant legal frameworks issued from both the national and the EU side, as well as the use cases described by the stakeholders and the user needs.

For that, a four-phase process was used to analyze and specify the EXTRACT requirements, which are the following:

1. **Project objectives document**: This is the first phase in which the project's objectives, outcomes and impact are described.
2. **Stakeholders' requirements gathering**: This is the second phase in which background information about the stakeholders/users and processes takes place. According to Preece and Sharp, in their book "Interaction Design: Beyond Human-Computer Interaction" [2], data gathering can be done using the following methods: use-cases scenarios, observations, interviews and questionnaires, etc.
3. **Requirements analysis and evaluation**: Is the third phase which encompasses tasks that include determining the needs or conditions to meet for running a new project, taking account of the possibly conflicting requirements of the various end-users, analyzing, and documenting, validating and managing stakeholder requirements. Once the user requirements are gathered (phase 2), potential techniques for analyzing stakeholder needs are "brainstorming" (taking into consideration guidelines and standards), "presentation scenarios", "prototyping" etc. [3].
4. **User and General requirements specification:** This is the final phase in the general process for stakeholders' requirements analysis.

## 5.2. Methods for Stakeholders' Requirements Gathering

Multiple methods for identifying requirements can work harmoniously, each augmenting the others, to produce comprehensive and effective results. This section outlines the approaches we used to discern the specific requirements for the EXTRACT project and its two principal use cases, P.E.R. and TASKA:

**Face-to-face or Virtual Meetings**: We engaged with stakeholders through both face-to-face and virtual meetings, seeking to understand their perspectives, needs, and requirements. This included considerations for both Functional and Non-Functional requirements, tailored to the specifics of each use case.

**Questionnaires**: As a complement to our direct meetings, we utilized expertly designed questionnaires to gather further information from all stakeholders. Specifically, we elicited expert judgment from professionals in the field, which significantly enriched our understanding and bolstered the requirements gathering process for the use cases.

**Literature Review**: In-depth study and analysis of existing literature provided a wealth of contextual information and insights into potential requirements.

**Meetings with Retired Policemen Associations**: Consultation with retired law enforcement personnel offered invaluable first-hand insights into practical aspects of crisis management and response, directly influencing the development of P.E.R. use case requirements.

**Meetings with Neighborhood Watch Associations**: Engagements with these local groups provided a ground-level view of community needs and potential challenges, particularly pertinent to the P.E.R. use case.

**Action-Research with Citizen Groups**: Active fieldwork with these groups helped us understand real-world needs and potential issues from the citizens' perspective, a key consideration in designing systems for both the use cases.

**Engagement with Research Institutions and Universities**: We met with researchers and scientists in the field of astronomy and astrophysics. These discussions helped us understand the specific data needs, analytical tools, and model requirements for tracking and predicting asteroid trajectories.

**Engagement with the Astronomy Community**: Meetings with professional and amateur astronomers provided valuable insights into the practical needs and constraints of those who would be using the TASKA system for tracking and studying asteroids.

**Engagement with Higher Education Institutions**: We liaised with educators to understand how TASKA could be used as an educational tool. Their input informed our requirements for user-friendly interfaces, data accessibility, and integration with educational curricula.

**Engagement with Service Providers**: We consulted with companies providing cloud-based services, high-speed data processing, and advanced computational tools. Their input informed our understanding of the technical feasibility and requirements for TASKA's data processing and analytical capabilities.

**Engagement with Non-profit Organizations**: We interacted with non-profits focusing on science education and space exploration to understand how TASKA could best serve their educational and outreach goals.

**Engagement with Public Sector Entities (Space Agencies, Disaster Management Authorities)**: Consultation with these organizations offered insights into the strategic needs for space weather events prediction and disaster response planning, helping to shape the requirements for the predictive capabilities of TASKA.

## 5.3. Analysis and results

### 5.3.1 PER Use Case

According to the scenario envisioned in section 4.2 and the result of the questionnaire dispensed to the experts, a number of high-level requirements (HLR) have been identified.

| HLR | Definition |
|---|---|
| HLR_1 | Hazard should be identified in due time. The UDT should be able to provide relevant information about the status of the system and support |

| | |
|---|---|
| | operators in detecting anomalies. The operator, through the UDT interface, is in charge of launching the alarm triggering the evacuation procedure. |
| HLR_2 | User's behavior (citizens, tourists and first responders) should be tracked using the EMA. The EMA should calculate a reliable position combining multiple signals on board and should communicate it to the UDT. |
| HLR_3 | Citizens should be reached through a 4R communication strategy (right person, at the right time, in the right place, with the right message) exploiting the pervasiveness of the smart technologies. This means that the safe route should be calculated at least considering the current position of the user, its movement capability, the status of the city on the ground and the movement of the other users at certain time |
| HLR_4 | Personal information must not be disclosed when the EMA interacts with the UDT. |
| HLR_5 | The decisions should be immediately translated into actions, while the feedback about the effects should be immediately detected through urban sensors (including humans acting as passive or active sensors), so that adjustments aiming to reduce unwanted consequences (e.g., people escaping in a wrong direction) can be actuated in due time.<br>This implements a real-time control loop enabled by smart technologies to support Evidence Driven Decision Making process at city level. |
| HLR_6 | All the physical element of the city that are relevant for the route calculation (e.g., waiting areas) should be georeferenced and the related information used for the calculation. |
| HLR_7 | Need to integrate heterogeneous data coming from different sources into a comprehensive semantic aware data model. This is necessary to overcome the limits of the silo approach in urban data management. Moreover, the possibility to model interdependencies of urban entity with semantic relationships allows context-aware information retrieval. |

In alignment with the specific needs and strategies determined by our stakeholders, we've identified multiple classes of datasets that could significantly improve situational awareness and guide evacuation procedures within the PER. use case. These datasets have been meticulously selected to support the event contextualization, adhering to the requirements set forth by the operators, and catering to the specific informational needs that elevate the quality of operator decisions.

The information provided by these datasets serves a dual purpose. Firstly, they help construct the context, providing an essential foundation for decision-making. Secondly, they offer valuable insights that benefit all kinds of decision-makers, thus promoting cross-functional collaboration. Each dataset plays a pivotal role in enhancing the overall effectiveness and accuracy of the evacuation strategies in the event of a fire hazard, thereby directly contributing to the successful implementation of the PER use case.

The identified datasets, their roles, and the value they provide are outlined in the table below. This comprehensive map of data assets is critical to our understanding and deployment of the PER use case within the larger scope of the EXTRACT project.

| Contextual Information | Dataset type |
|---|---|
| Geographic based city structure/entities | • streets and related civic numbers<br>• street directions, type<br>• squares<br>• bridges<br>• pedestrian areas<br>• building<br>• green areas<br>• Public services (bar, hotels, restaurants, police...)<br>• admission areas<br>• waiting areas<br>• defibrillators (position and status) |
| Passengers/civilians: position, number | • water traffic flow sensors position<br>• traffic flow sensors<br>• people presence detection sensors |
| Weather information and alerting | • local weather forecast<br>• civil protection alerting |
| News/event (planned and unplanned) updates: | • Public events in the city – RT |
| Traffic information (public transport) | • Limited traffic area,<br>• traffic flow sensors position<br>• traffic flow sensors data<br>• mobility events (incidents) |

*Table 1 Contextual Information identified*

| **Specific Information needs** | **Datasets (IoT, GIS, etc.)** |
|---|---|
| Magnitude and dynamic of the event | • anemometers lat/long &measure<br>• rainfall lat/long & measure<br>• hydrometers lat/long &measure<br>• hygrometry lat/long &measure |
| Real time data | • # of people present on the ground and close to the event<br>• # of vehicles/boats present on the ground and close to the event<br>• People behavior (direction movement) RT<br>• Services status (e.g., open/close) – RT |

| | |
|---|---|
| | • Presences at school - periodical |
| Number of qualified operators available | • # of volunteers (civil protection) available in RT<br>• # of first responders available and displaced |
| Means availability | • Triage - emergency rooms status RT<br>• # of patients admitted to the emergency department<br>• # of civil protection vehicles available<br>• # of ambulances available in RT |
| Critical Event prediction | • weather forecast<br>• people present on the ground trend & forecast<br>• river basin level trend & forecast |
| Energy supply status | • Outage status - RT |
| Alternative transport means | • boats map, schedule, position RT<br>• taxi/boats stops and positions RT |
| Position and behavior dynamics of the people in the city | • traffic flow sensors position<br>• traffic flow sensors data<br>• people presence detection sensors- RT<br>• people concentration (actual+trend) |

Essential data for this purpose has been compiled from the "Smart Control Room", a collaborative operation center spearheaded by the local police alongside other strategic city departments in Venice, such as the local transport service.

The datasets gathered will enable the system to construct a comprehensive model of the city, critical for identifying the optimal evacuation route considering city congestion, available boats for escape, among other factors. It's worth noting that this data isn't gathered live due to its association with police activities and city management. Instead, these datasets have been drawn from recent historical records within a specific timeframe. However, in order to emulate the same live operational condition, the extracted and the simulated extreme data will be sent to the DT at the same collection frequency rate of the live system.

The datasets that have been actually collected are to be categorized as follows:

- **Pedestrian Flow**: Sensors strategically placed in the city's core provide data on the number of people entering or exiting specific areas. Over 19,000 records are available from 10 sensors in the Rialto zone.
- **Water Traffic**: Cameras continually monitor the Grand Canal, providing insights into water traffic. They can identify major types of ships passing through the surveillance zone and their directions. Data from two cameras yielded over 3,800 records.
- **Presence**: Analysis of unique mobile phone numbers provides an estimation of the city's population with an update rate of 15 minutes, aggregated into 150m per edge squares. This dataset contains over 1 million records.
- **Georeferenced Data**: Layers of georeferenced shapes are available, including pedestrian areas, green areas, lagoon canals, buildings (major and minor), house numbers, water zones, and street names.

Each dataset has its unique format, and the system should be capable of fully utilizing them. Georeferenced data are in .zip files, containing "shapefiles" and other files for associated metadata. The Presence dataset comprises several CSV files, each having a tailored header structure to best represent the detail level. The Pedestrian Flow and

Water Traffic datasets are in JSON format, with data aggregated by sensor and camera, respectively.

Summarizing, here are the primary components of the PER use case system:

- **Input**: The system utilizes a rich variety of data sources. This includes high-dynamic raw data samples from IoT devices and the EMA application. Additionally, static Geographic Information System (GIS) data and low-dynamic data from Copernicus significantly contribute to the overall data pool that feeds into the system.
- **Facilities**: The Urban Digital Twin serves as the primary visualization and interaction facility. It is complemented by an AI system responsible for calculating optimal evacuation routes, and the EMA application, which tracks individuals' positions and provides them with safe route information.
- **Analytics**: Real-time analysis is carried out on the extreme data flow generated by various inputs. This analysis is designed to be responsive and agile, ensuring that the system can react swiftly and accurately during emergency situations.
- **Output**: The output of this system is twofold. Firstly, it provides a personalized evacuation route for each connected user, taking into account their current location, the nature of the hazard, and the status of the city. Secondly, it produces a real-time status of the city, complete with a 3D representation. This visualization assists emergency response teams and decision-makers in understanding the scale and spread of the hazard, helping to guide their actions and strategies.

### 5.3.2 Triggering of Requirements Based on Data and Use Case Scenario

This section outlines the various requirements that are triggered based on the data gathered from the SCR and the urban digital twin. A more detailed list and explanation of these requirements are provided in Chapters 6 and 7.

**Functional Requirements**

- The system must be able to detect various hazards such as fires, earthquakes, terrorist attacks, etc., and trigger an emergency evacuation protocol.
- It should generate optimal escape routes for each individual in real time, taking into account the unique characteristics of Venice, such as its narrow streets and waterways.
- Implementation of advanced machine learning techniques is required, specifically a MultiAgent Reinforcement Learning Model to determine optimal escape paths.
- The system should interface with the Smart Control Room and other relevant local authorities for alarm triggering and coordination.

**Performance and Dependability Requirements**

- The system should be capable of georeferencing all individuals in Venice at the moment of hazard detection.
- It should be reliable and robust, functioning accurately and consistently, even in high-stress scenarios.
- The system must ensure that safe routes do not themselves become hazards due to overcrowding.

- It should include capabilities to monitor real-time changes in the number of people at each node, edge, and destination (waiting areas) and make dynamic adjustments.

**Security, Energy Efficiency, and Communication Requirements**

- The application should adhere to strict data privacy regulations, ensuring the protection of personally identifiable information.
- The application should facilitate reliable communication between EMA users and the UDT during emergencies.

**Execution Services Requirements**

- The system must provide APIs for easy integration with the UDT primary and Smart Control Room (to support the exploitation phase) and other relevant systems like the GPS and environmental sensors.
- Libraries should be available for processing large volumes of data and machine learning tasks.

### 5.3.3 TASKA Use Case

**TASKA**: F2F meeting in Venice has allowed the team to narrow down specific objectives per use cases for TASKA.

**Regarding case A,** the goal is to leverage ongoing development of AI/ML algorithm to identify science content during the post-processing and implement a real-time pipeline close to the receivers (UnDysPuted) to detect and classify events so as to enable adaptive resolution (see Fig. A1) set upon science content. UnDysPuted is currently built using the GPU technology thus appropriate programming models will have to be used for this development on the edge component. The end goal is to process data on the fly based on classification (clean and integrate) including short / medium/large scale events (including time-frequency feature polygon), radio interferences (RFI), quiet (background) regions, etc. A first MVP will be developed through an off-line pipeline (better for tracking pipeline parameters and reproducibility), with the goal to enable classification storage (analogue-to-information) potentially using an RDF database. The frontend system ("Σ" symbol on figure A1) will also be improved with an automated faulty antenna stream detector, in order to detect and blank faulty signals before they are merge into the raw data stream fed into UnDysPuted.

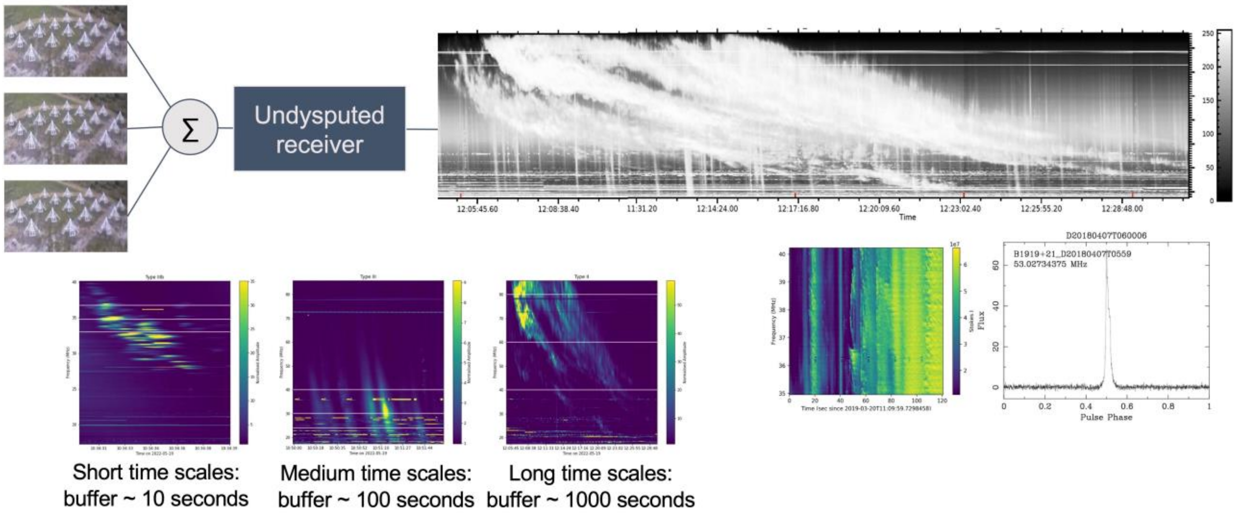



*Figure A1: Schematic representing the combination of radio sensors as a sum inside Undysputed receiver which transform the raw voltage into dynamic spectra (upper right). The frequency and time resolution will restrict the access to the physical structure of the emission (here Type II & Type III bursts). A dynamic monitoring that could modify the receiver parameters can preserve the access to fine structure at small scales (bottom left).*

**Regarding case C**, we recall below the typical data reduction and processing pipeline, including the various data levels (L0, L1 and L2). L0 is produced directly in the edge facility (i.e., the telescope), and creates the L1 data that will be served to the datalake. L2 is the next stage of the data after they have been calibrated for. The data format that stores the interferometric data is the "Measurement Sets" (MS) file format described in 4.4.1.

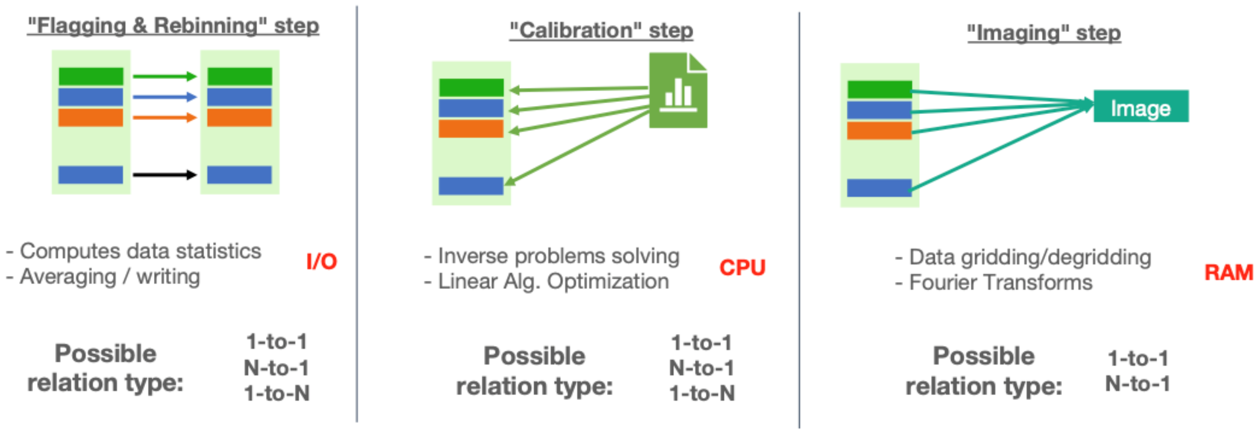



*Figure C1: Description of the three main data processing steps in radio interferometry: i) cleaning the data and reducing its volume (left), ii) calibrating the data (center) and iii) imaging the whole dataset (right). One dataset (light green) contains several MS files, each one corresponding to a particular frequency.*

From the recording to the final image cubes, the data have to go through several generic steps described hereafter: i) **pre-processing step** to transform the raw data into a "clean" version free of outliers and ready to undergo the next step, ii) **the calibration step** to correct the bias of the instrument and iii) **the imaging step** that converts the interferometric measurement (in Fourier space) into a discrete image space (direction, time, frequency) that is the scientifically exploitable end products. All steps must guarantee the integrity of the data and of the scientific content of the data. These steps are generic and are non-unique as the various scientific programs might use a different combination of them, multiple calls to them with various sets of parameters (e.g., coarse calibration and coarse imaging followed by refined calibration derived from the coarse image, and final imaging). Likewise, datasets files can interact together in a specific task and can have 1-to-1, 1-to-N, N-to-1 and N-to N relationship. Continuous access to all files in relevant dataset is therefore mandatory.

The processing "recipes" can vary a lot but can always be described in terms of building blocks which nature is one of the three steps describe below (shown as blue boxes in the figure):

**Step 1: Pre-processing**: the raw and heavy recorded data (L0) that is directly produced by the correlator undergo a statistical analysis over all its axes (a long time, frequency, baseline, antenna, direction axes etc.) and the detected and unwanted outliers (e.g., Radio Frequency Interferences) are flagged and discarded from the dataset (using the FLAG column). Once the bad records are no longer affecting the data, the latter can undergo a dimensionality reduction (e.g., averaging or lossy compression) over arbitrary axes (e.g., averaging 1s records to seconds records, averaging between frequency channels, etc.). This lighter data forms the "L1" data that are ready for instrumental calibration. This step can be decomposed into "Flagging" and "Rebinning" steps. TASKA use case C (and D) will always rely on L1 data as they are produced by the Edge facilities. However, further data flagging and averaging can be performed on L1 data therefore this capability has to be implemented.

MVP: Capability to ingest various sizes of input datasets (from few 100s of MB to several TB) and various number of files (from 1 large dataset of 1 file, to 100s of datasets containing 100s of files).

<table>
<tr><th>File type</th><th>Min size</th><th>"Max" size</th></tr>
<tr><td>1 MS file</td><td>Few 10s-100s MB</td><td>~100s GB</td></tr>
<tr><td>e.g. 1 observation (244 MS)<br>(T=1h – df=1 ch/SB – dt=8s)</td><td colspan="2">15 GB</td></tr>
<tr><td>e.g. 1 observation (244 MS)<br>(T=1h – df=64 ch/SB – dt=1s)</td><td colspan="2">~10 TB</td></tr>
<tr><td>e.g. 1 observation (244 MS)<br>(T=8h – df=1 ch/SB – dt=8s)</td><td colspan="2">~120 GB</td></tr>
<tr><td>e.g. 1 observation (244 MS)<br>(T=8h – df=64 ch/SB – dt=1s)</td><td colspan="2">~80-100 TB</td></tr>
</table>

*Table C1: Range of possible values for interferometric data volumes depending on typical observation parameters with NenuFAR.*

**Step 2: Instrumental calibration**

The cleaned data coming from step 1 is ready to undergo correction from the instrumental response. This step is the most important one to associate a meaningful scale and values that can be scientifically exploitable to the measured data. The conversion from the receiver measurement scale to a scientific scale constitutes the calibration step. The necessary ancillary inputs are a model of the instrument and/or a model of the observed sky. The model of the sky serves as a reference to deduce the correction factor for all antennas measurements. These corrections factors are called calibration tables (stored in external HDF5 format) and can be used for application to another relevant observation. Nonetheless, calibration is mandatory for each dataset at all times and all frequencies. The core of calibration is to solve a large set of linear equations using advanced algorithms. The calibration step will create a new column (CORRECTED_DATA column) in each MS to store the corrected values while keeping the original clean data (DATA column) untouched for future reprocessing. Input datasets can be split in two main types of data: scientific datasets and calibrator targets. Calibration can be performed on the calibrator dataset, and the corresponding calibration tables can be applied to the scientific datasets.

MVP: Capability to process multiple datasets in parallel, monitor their interactions and organize the results according to future usage.

**Step 3: Imaging**

The imaging step is the final step that takes cleaned and corrected data to inverse it and produce the result image cubes (L2 data) adapted for scientific exploitation. This step consists of two substeps: i) the data gridding and ii) data deconvolution. The first substep takes the corrected listing of visibility data from the Fourier space and project it to a sampled image space. Missing data due to the nature of interferometric data translates in an incomplete view of the sky containing a lot of distortion. The second step consists of correcting such distortion by means of image deconvolution. The radio astronomy community has developed modern tools to handle this task but relies on a heavy usage of RAM and access to a large number of MS that will be agglomerated together prior to imaging.

MVP: The imaging step should be able to access all desired calibrated dataset in a seamless way. Targets products range from a single image integrated over a time range and frequency range, but also be a large image cube containing a frequency axis and a temporal axis in addition to the 2-D image spatial axes. This product should be available quickly for inspection to possibly enable rerun with a different choice of data selection and imaging parameters.

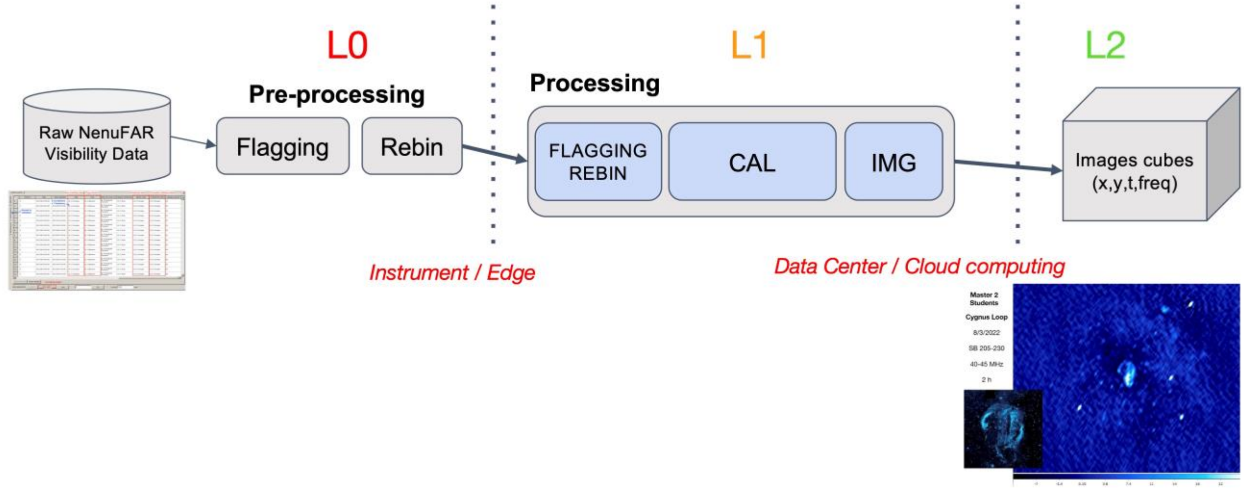



*Figure C2: Generic pipeline for reducing raw interferometric data (L0, in the edge facility) to the final scientifically exploitable multi-dimensional image cubes (L2) after cleaning, compressing and calibrating the data.*

In order to build a final flexible pipeline, we listed the milestones that all partners in TASKA Use Case C could follow to reach the use case goals.

**Milestone 1:**

DATA: First, a common lightweight data repository containing relevant datasets corresponding to real scientific programs but share the same processing needs. Second, we needed the teams involved in Use Case C to be acquainted with the MS data format and the associated data manipulation tools for chunking, grouping, rebinning the data, as well as computing statistics and flagging the data following the MS standard.

PROCESSING: The use case C members should provide a set of minimal demo datasets for which the final products are known and controlled (and provided for comparison). A minimal working processing recipe is also provided for each of them to evaluate the baseline performance and measure the impact of pipeline optimization developments on the products.

**Milestone 2**:

DATA: Build a long-term reference dataset repository with support documentation in the preparation of different scenarios considered for scaling up to the Square Kilometre Array. This includes ingesting heavy (TB) datasets and distributing them smartly in the available storage facilities in preparation for subsequent processing and imaging.

PROCESSING: During milestone 2, the processing blocks should all be identified and tested on the range of relevant selected datasets to explore the parameter space of the I/O, CPU and memory metrics. Milestone 2 will be achieved when all processing steps are mastered and run using the baseline parametrization before EXTRACT project. The main processing strategies have also to be laid out depending on the nature and quality of end-products to generate.

**Milestone 3**:

DATA: From the data sample and the realistic datasets involved in Milestone #1 and #2, part of the HPC team must provide feedback on the first hints of optimization for the storage part (data slicing, intelligence distribution, easy access for processing, transfer delays, typical response, etc.).

PROCESSING: Using the a priori control over the processing steps and typical recipes used for processing, the milestone will be achieved when the first hints of optimization of each processing steps (flagging, rebinning, calibration and imaging), as well as the strategy to use for each of them (containerization, parallel, etc.)

The total time spent on data transfer and end-to-end data processing is measured using the Processing-over-Observation time ratio (P/O ratio) which indicate the fraction of time devoted to the processing to create the scientific products compared to the observation time. Target value for P/O should not exceed P/O of 1 (e.g., 10-hour observations should be reduced within the following next 10 hours, without counting the transfer time from the edge facility).

**Milestone 4**:

DATA: The final milestone in the scope of TASKA is to propose a completely optimized pipeline that can Ingest new data from a real, heavy observation coming from NenuFAR Edge facility in Nançay.

PROCESSING: All processing steps should be optimized depending on their nature and resource consumption (I/O, CPU, RAM). All steps should be chained seamlessly and enable logging, quality assessment, possibility to rerun all or specific steps of the processing without breaking the data flow continuity. Alternate recipes could be explored by the end-user by "branching away" from default predefined recipes. This workflow editing must be flexible enough to conjugate presets, parameters and pipeline editing, looping specific steps while ensuring that each step is optimized and automatized. The Provenance standard (IVOA) can be used as an example for traceability and reproducibility. The final workflow should also be compatible with the FAIR principles for data generation, access and distribution.

**Regarding Case D**, as an extension of Use Case C, the goal of Case D is to implement and test a new dedicated method for imaging transient from Measurements Sets at the imaging step and using deep neural networks (DNN). From point sources, an approach has already been developed by the team (Chiche et al., PhD 2023, Chiche et al. accepted, Fig. D1). A dedicated approach needs to be developed for the case of extended / moving targets (see figure D2). We expect strong interactions with Use Case A in terms of information / data flow.

MVP: The baseline MVP would be to have a new trained network that provides scientifically validated images cubes containing unresolved and (as a target objective) resolved sources (e.g., either on simulated data or images of solar system bodies that have been observed with space probes). An additional possible target for this use case would be to add on top of this MVP, a classification network (faint sources vs. artifacts, light curve, etc.).

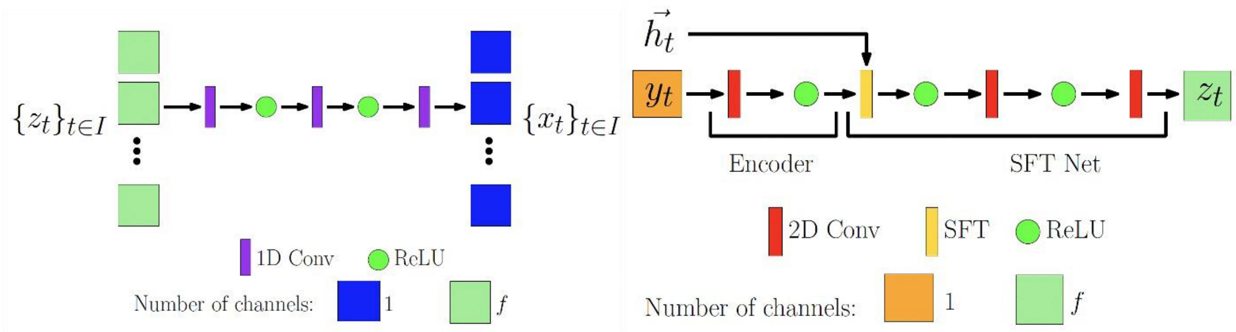



*Figure D1: Proof of concept from Chiche et al. (accepted in A&A) (left) 1D net to capturing the temporal structure of the source and (right) 2D net to solve the 2D deconvolution problem.*

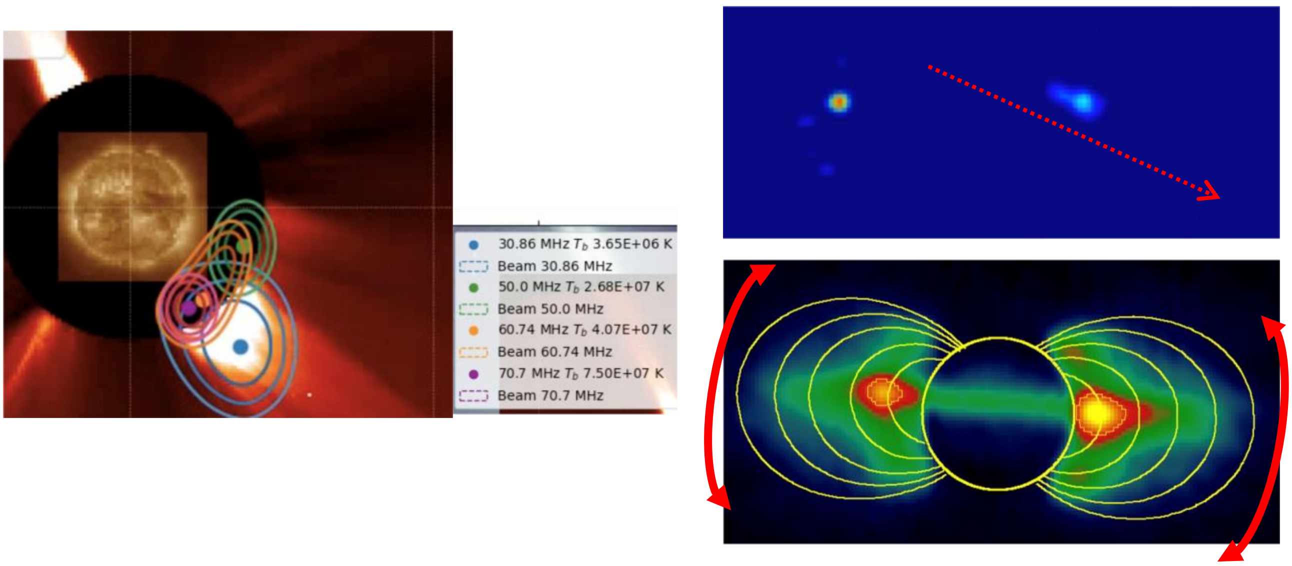



*Figure D2: (left) Partially resolved Solar bursts detected at low radio frequencies and traveling away from the solar disk. (right) Portion of the sky displaying a static radio source (left) and a partially resolved bean-like image of the Jupiter Van Allen belts. The apparent motion makes Jupiter travel in the sky reference frame. (bottom) Close-up view of the synchrotron emission associated with the radiation belts that rocks around the planet dynamic plane.*

**Regarding Case E**, recently, thanks to the DDFacet/KillMS tools which enable refined foreground subtraction, the image residuals can be exploited to their fullest. A small tool, DynSpecMS, enables producing the dynamic spectrum of any pixel in a residual image. This extra step can be very efficient in case of faint sources that are hidden in the noise and that displays sparse structure in the time frequency plane (like solar burst, exoplanetary radio emissions). By adding this extra step, it is possible to build a hybrid use case based on an extension of developments from case A, C and D. Pipeline sequencing is inherited from A and D, while data logistics is inherited from C. The goal is to extract dynamic spectrum from Measurement Sets for selected directions, from a large survey. Hence, the primary requirement would be to include DynSpecMS as analytics in case C. The science goal is to get location of components through the mapping of low-resolution features with high resolution beamform from case A and use astrometry for better calibration. Hence the second main requirement is to develop software to realize the spectral-temporal features mapping (i.e., image cubes from C or D to Beamform from A).

Finally, a feature matching algorithm has to be developed to map the UC A high resolution data temporal-spectral features with the dynamic spatial features detected in UC C, D and through DynSpecMS.

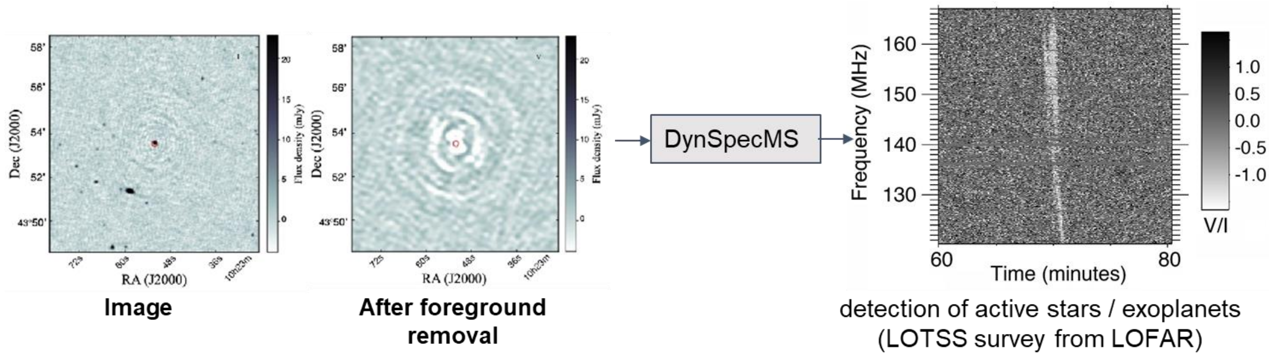



*Figure E1: (left) comparison before and after foreground removal by killMS. By selecting a direction (red circle) and using DynSpecMS, one can perform a spectroscopy analysis towards the direction to possibly reveal hidden signature of transient sources (right).*

We provide bellow, a summary for the various use cases of TASKA:

- Case-A:
    - input = raw beamformed data samples from NenuFAR
    - facility = Telescope + Beamformer (Undysputed) on the edge side
    - analytics = real-time statistical analysis on extreme data flow to detect and classify features seen in the dynamic spectra. Dynamic configuration switching between low-resolution and high-resolution will be controlled depending on the source detection.
    - output = information on the detected trigger + high-resolution data slice around the trigger that can be cleaned and reduced after recording.

- Case-C
    - input = pre-processed radio visibilities (MS format) (L1)
    - facility = NenuFAR datalake (cloud)
    - analytics = optimized sequence of operations for imaging pipelines for each considered scientific case.
    - output = multiple image cubes (spatial / spectral) obtained after a target minimum P/O ratio of 1.

- Case-D
    - input = pre-processed radio visibilities (MS format) (L1)
    - facility = NenuFAR datalake (cloud)
    - analytics = Image cubes containing unresolved transient sources (goal 1) and resolved transient sources (goal 2) (~video reconstruction)
    - output = cubes (spatial / spectral / temporal) from the DNN.

- Case-E
    - input = case-A output + case-C/D output
    - facility = NenuFAR datalake (cloud)

- analytics = morphological mapping between 2 datasets (high-res / low-res)
- output = matched features information + astrometry

# 6. Stakeholders' Requirements Definition

## 6.1. Requirements elicitation and prioritization approach

Following the consolidation of insights from the meetings and questionnaires outlined in the preceding section, the subsequent phase focuses on defining and assessing the significance of the requirements. A comprehensive analysis was conducted to identify a set of requirements linked to the use-case scenarios.

## 6.2. Prioritized requirements

Table 3: Prioritized requirements

| ID | UR1 |
|---|---|
| Requirement type | User |
| Relevant use case | TASKA – Case A |
| Short title | Real-time classification on UnDysPuted (edge) |
| Rationale | Enable adaptive resolution set upon science content |
| Description | Real-time pipeline in "undysputed" receiver to detect and classify events close to the antennas.<br>Process data on the fly based on classification (clean and integrate) |
| Originator | OBSPARIS |
| Priority | High |
| Fit criterion | High (edge component) |
| Supporting material | Experience on typical observation preset from users |

| ID | UR2 |
|---|---|
| Requirement type | User |
| Relevant use case | TASKA Case C |
| Short title | Seamless performance optimization for interferometric data management |

| | |
|---|---|
| Rationale | Guarantee that all data processing steps are optimized depending on the evaluated I/O, CPU, RAM consumption with a motivation to reduce the P/O ratio to a minimum (target P/O ~1) |
| Description | Capability to ingest new data and prepare them for processing<br>Transparency and easy data representation & manipulation ("Data Objects")<br>Depending on the data volume, and the nature of the task of the processing, seamless data chunking and distribution to optimize the duration and load of the tasks.<br>Efficiency & time/load optimization of I/O, CPU, RAM for the processing tasks and subtasks |
| Originator | OBSPARIS |
| Priority | High |
| Fit criterion | High (Cloud component) |
| Supporting material | Existing standards (NRAO Measurements Sets, Casacore) and recipes (users' scripts) |

| ID | UR3 |
|---|---|
| Requirement type | User |
| Relevant use case | TASKA Case C |
| Short title | Flexibility definition and operation of the imaging pipelines |
| Rationale | Formulate a flexible way to automatize processing of imaging data from NenuFAR on the cloud while enabling partial reruns |
| Description | Using data prepared from UR2.<br>Automation of processing based on simple & complex "recipes" and external files (models, etc...)<br>User capability to rerun certain processes (e.g., imaging) with different choice of parameters |
| Originator | OBSPARIS |
| Priority | High |
| Fit criterion | High (Cloud component) |
| Supporting material | Existing standards (NRAO Measurements Sets, Casacore) and recipes (users' scripts) |

| ID | UR4 |
|---|---|
| Requirement type | User |
| Relevant use case | TASKA Case D |
| Short title | New AI methodology for imaging extended and moving targets |
| Rationale | Need for a Time-agile deconvolution accounting for the structure of extended sources on the time axis as well as on the direction axes. |
| Description | Develop an extension of framework from Chiche et al.[4], based on DNN, to reconstruct images of extended and moving targets from MS datasets |
| Originator | OBSPARIS |
| Priority | High |
| Fit criterion | High (Cloud component) |
| Supporting material | Existing methods (see above) and large training and testing datasets based on real data.<br>Link with classification methods developed from UR1 |

| ID | UR5 |
|---|---|
| Requirement type | User |
| Relevant use case | TASKA Case E |
| Short title | Enable morphological mapping (DynSpecMS) between data products of different nature (beamformed vs. image cubes) |
| Rationale | Faint source detection (e.g., exoplanets) and better source identification and positioning |
| Description | Develop a software to realize the spectral-temporal features mapping between MS datasets and raw data |
| Originator | OBSPARIS |
| Priority | High |
| Fit criterion | High (Edge to Cloud) |
| Supporting material | Existing DynSpecMS implementation |

[4] https://theses.hal.science/tel-03985254

| ID | UR6 |
|---|---|
| Requirement type | User |
| Relevant use case | TASKA Case E |
| Short title | Integrate morphological mapping |
| Rationale | extract dynamic spectrum from MS for selected direction, from a large survey |
| Description | include DynSpecMS as analytics on NenuFAR Cloud (link with UR2) |
| Originator | OBSPARIS |
| Priority | High |
| Fit criterion | High (Edge to Cloud) |
| Supporting material | Existing DynSpecMS implementation |

| ID | UR7 |
|---|---|
| Requirement type | User |
| Relevant use case | PER - Evacuation Mobile App (EMA) |
| Short title | Personal Escape Route retrieving |
| Rationale | At the moment only general alerting mechanism exists as GSM push notification (Civil Protection Alerting and IT-Alert Public Alert System" has been published - 2023). The content of the message is simple and generic for the entire population in the area, without specific relation with the distance from the event or movement capacity. Hence no 4R based personalized evacuation service is currently provided. |
| Description | The application receives the path relating to the PER calibrated on the user's profile and position, from the AI module, and provides for displaying them in the EMA according to the 4R approach. |
| Originator | MATHEMA |
| Priority | High |
| Fit criterion | High |
| Supporting material | 4R mobile messaging is a well-known technique well established in mobile social network |

| ID | UR8 |
|---|---|
| Requirement type | User |
| Relevant use case | PER - Evacuation Mobile App (EMA) |
| Short title | Evacuation Mobile App (EMA) should implement hybrid positioning onboard |
| Rationale | The Venice scenario poses several difficulties in terms of outdoor positioning. In fact, because of the very narrow streets relying on GPS only for mobile positioning is not reliable. To this end it is necessary to combine multiple sources to increase the current level of precision |
| Description | The EMA fuses signals coming from GPS, Galileo and 5G onboard in order to estimate a reliable position of the device during the evacuation. The position should be communicated to the central system close to Real Time to allow people movement tracking. |
| Originator | EMA |
| Priority | High |
| Fit criterion | High (Hybrid positioning) |
| Supporting material | Mobile Hybrid positioning is experimental, and few supporting materials exists. |

| ID | UR9 |
|---|---|
| Requirement type | User |
| Relevant use case | PER - Evacuation Mobile App (EMA) |
| Short title | Evacuation Mobile App should be able to send the position to the UDT in real time |
| Rationale | In order to support real-time reliable and informed decision making it is necessary to make data from the city available in real time. |
| Description | The application receives the data relating to the PER calibrated on the user's data from the MARL, and provides for displaying them in the application |
| Originator | MATHEMA |
| Priority | High |
| Fit criterion | High (EMA) |
| Supporting material | Mobile-server interaction through API is a well-known technology |

| ID | UR10 |
|---|---|
| Requirement type | User |
| Relevant use case | PER - Evacuation Mobile App (EMA) |
| Short title | EMA should display the path on the screen and support dynamic Guidance |
| Rationale | Evacuation and emergency app are based on GSM messaging or pre-defined routes that can be exploited on mobile map. |
| Description | Once the personal path is provided, EMA should monitor the progress of the movement and in case of deviation with respect to the planned path, EMA should send a message to the UDT + AI to obtain a new path according to the new starting position and the condition on the ground. |
| Originator | MATHEMA |
| Priority | High |
| Fit criterion | Evaluate/validate in pilots |

| ID | UR11 |
|---|---|
| Requirement type | User |
| Relevant use case | PER - Evacuation Mobile App (EMA) |
| Short title | Evacuation Mobile App must avoid to disclosing the personal information to unauthorized party |
| Rationale | Mobility Capacity can be considered sensible information. It is personal nature. Knowing this is crucial for personalized path calculation. |
| Description | The application should be able to send the position and the user profile intended as his/her mobility capacity to the central without disclosing any other personal information. |
| Originator | MATHEMA |
| Priority | High |
| Fit criterion | High (Edge-mobile) |

| ID | UR12 |
|---|---|
| Requirement type | User |
| Relevant use case | PER-Urban Digital Twin (UDT) |
| Short title | Gathering heterogeneous data from relevant data sources |
| Rationale | The heterogeneity of the data generated by the city should be collected and made immediately available to the system. The |

| | |
|---|---|
| | data to be collected includes user sensed data (Volunteer Geographic Information), environmental sensors, people movement sensors, traffic sensors and Copernicus data. |
| Description | The system should be able to collect data streams from heterogeneous sensors including EMA application, people movement |
| Originator | LOGOS, Venice City |
| Priority | High |
| Fit criterion | High (Cloud) |
| Supporting material | some tools to acquire data and generate RDF exists (e.g., TripleGeo, Kettle+Karma) |

| ID | UR13 |
|---|---|
| Requirement type | User |
| Relevant use case | PER-Urban Digital Twin (UDT) |
| Short title | Heterogeneous Data collected should be fused using a semantic aware data model |
| Rationale | The heterogeneity of the data generated by the city is currently organized in different databases according to its nature (e.g., GIS data are managed by a Geographic database, sensor data is managed by time series database, etc.). This means that the datasets (and in particular the entities responsible for the data generation) are not related to each other through a model. This condition generates the typical silo approach to data management preventing the possibility to extract complex datasets combining multiple sources at the same time. |
| Description | The Urban Digital Twin should be able to integrate data with a unified semantic aware data model in order to put in relation city entities on the base of their logical, physical and temporal connections. |
| Originator | LOGOS |
| Priority | High |
| Fit criterion | High (Cloud) |
| Supporting material | Existing ontologies to manage smart city are available such as KM4City |

| ID | UR14 |
|---|---|
| Requirement type | User |
| Relevant use case | PER-Urban Digital Twin (UDT) |
| Short title | Urban Data should be presented with a 3D dynamic model on the web |

| | |
|---|---|
| Rationale | The status of the city is presented through a dashboard managed by the security and safety operator where static as well as dynamic data are displayed. The dashboard presents different views on the base of the dataset to be represented. However, an integrated view where geographical data with dynamic data fused together in a 3D dynamic representation is missing. Thus, the operation may experience some difficulties in gaining a clear situation awareness |
| Description | The UDT should display the status of the city through an integrated 3D model of the city with dynamic data. This means that the web application should interact with the Knowledge base in real-time to update the status in |
| Originator | LOGOS, City of Venice |
| Priority | Normal |
| Fit criterion | High (Cloud) |
| Supporting material | Tools to render 3D model on browser exist such as CesiumJS |

| ID | UR15 |
|---|---|
| Requirement type | User |
| Relevant use case | PER-Urban Digital Twin (UDT) |
| Short title | Data should be made available to the AI module through a scalable API able to support high frequency complex semantic queries. |
| Rationale | The AI model has to evaluate the status of the situation during the event in order to evaluate the effect of the policy(actions) applied in the environment |
| Description | The API should support high frequency complex semantic queries and provide the status of the UDT to the AI module via standard data model (e.g., JSON). The data must be made available with a frequency compatible with the dynamics of the event |
| Originator | LOGOS, City of Venice |
| Priority | High |
| Fit criterion | High (Cloud) |

| ID | UR16 |
|---|---|
| Requirement type | User |
| Relevant use case | PER-Urban Digital Twin (UDT) |
| Short title | Knowledge base should model spatial, temporal, logical and functional inter-dependencies among the urban elements |

| | |
|---|---|
| | considered allowing performative queries and inferences in polynomial time |
| Rationale | Currently the inter-dependencies among the urban elements are not models and their potential interaction is managed on the base of expert judgment and implicit knowledge of the operators/experts |
| Description | The elements considered should be put in semantic relation among each element considered in the UDT. The inter-dependencies should be modeled according to a suitable description logic able to allow inferences in polynomial time (E.g., using OWL EL) |
| Originator | City of Venice, LOGOS |
| Priority | High |
| Fit criterion | High (Semantic layer) |
| Supporting material | Ontologies able to manage spatial-temporal aspects are available and integrable in a multi-domain ontology. |

| ID | UR17 |
|---|---|
| Requirement type | User |
| Relevant use case | PER-Urban Digital Twin (UDT) |
| Short title | Integrating simulation of urban elements |
| Rationale | Not all the elements needed to take decisions are freely accessible in the Smart City Control room due to security reasons. To this end elements of the city could be simulated in such a way to be similar to the reality in terms of data and dynamics. |
| Description | The UDT should integrate the simulation of the human movement of the ground. The simulator, on the basis of the provided datasets, should be able to:<br>▪ Simulate the number of human beings entering and exiting a specific zone of the city at a certain point in time during the day.<br>▪ Simulate which types of boat are present in the Grand Canal at a certain point in time during the day.<br>▪ Simulate the overall number of people present in the city during the day.<br>▪ Contextualize the crowding data with proper georeferenced data to evaluate population density and feasible escape routes. |
| Originator | MATH, LOGOS |
| Priority | High |
| Fit criterion | High (UDT) |

| ID | UR18 |
| --- | --- |
| Requirement type | User |
| Relevant use case | PER-Urban Digital Twin (UDT) |
| Short title | Trigger Evacuation procedure - procedure conclusion |
| Rationale | The evacuation procedure should be triggered by decision makers/security operators |
| Description | The UDT should integrate the simulation of the human movement of the ground. The simulator, on the basis of the provided datasets, should be able to:<br>▪ Predict the number of human beings entering and exiting a specific zone of the city at a certain point in time during the day.<br>▪ Predict which types of boat are present in the Grand Canal at a certain point in time during the day.<br>▪ Estimate the overall number of people present in the city during the day.<br>▪ Contextualize the crowding data with proper georeferenced data to evaluate population density and feasible escape routes. |
| Originator | MATH, LOGOS |
| Priority | High |
| Fit criterion | High |

# 7. System Requirements

The purpose of this section is to present the requirements of the EXTRACT project. The requirements fall into two main categories functional and non-functional. Functional requirements describe what the software system should do. On the other hand, non-functional requirements describe how the system should perform these certain tasks as well as restrictions that may apply. The requirements derived from the stakeholders' analysis and requirements definition presented in the previous sections.

## 7.1. Functional requirements

The functional requirements outlined here are the result of extensive stakeholder surveys, followed by an in-depth analysis to discern what the EXTRACT system should ideally accomplish. Both project partners and stakeholders were instrumental in gathering these requirements.

A user-centric approach was adopted to identify user requirements, leading to the distinction of both functional and non-functional requirements detailed in the subsequent sections.

The requirements derived from the stakeholder's analysis underwent further analysis, resulting in the key requirements presented in the following sections.

Tables in this section display the functional requirements for each use case. The tables use four columns: ID (a unique identifier for the requirement), Description (a succinct explanation of the requirement), Priority (the importance level derived from the above user requirements), and Relevant UR (indicating the User Requirement from which it is derived).

Table 3a: Functional requirements for TASKA

| ID | Description | Priority | Relevant UR |
|---|---|---|---|
| 1 | Real-time pipelines on Undysputed GPUs | High | 1 |
| 2 | Capability to ingest very large datasets formatted under the MS standard. | High | 2 |
| 3 | Adapt storage & computing strategy depending on science program (a lot of small dataset ≠ a single huge dataset) | Medium | 2 |
| 4 | Capability to offer flexibility for scheduling the data processing steps and analytics through dedicated scripts using standard building blocks (see table below) | High | 3 |
| 5 | New deconvolution method based on DNN to perform image restoration on time-varying extended sources | High | 4 |
| 6 | Extend DynSpecMS to map spectral-temporal features | High | 5 |
| 7 | Integrate DynSpecMS in building blocks from #4 | High | 6 |

Table 3b: Functional requirements for PER

| ID | Description | Priority | Relevant UR |
|---|---|---|---|
| 1 | Ingest extreme data from heterogeneous data sources into the KB and TSDB | High | UR9 |
| 2 | Integrate heterogeneous data into a common ontology able to model urban elements relations | High | UR10 |
| 3 | Capability to trigger the evacuation procedure | High | UR3 |
| 4 | Scalable API to extract city state using SPARQL | High | UR12 |
| 5 | Update the 3D city model on the web with the data collected | Medium | UR11 |
| 6 | Generate personalized evacuation route on the base of position, movement capacity of the user and resource available selected by the operator | High | UR3 |
| 7 | Continuous learning from the feedback received and personalized route refinement. | High | UR3 |

| | | | |
|---|---|---|---|
| 8 | Automatic Evacuation procedure completion | High | |
| 9 | Secure transmission of the data under HTTPS protocol | High | UR3, UR4 |
| 10 | Secure AUTH for use API REST | High | UR3, UR4 |
| 11 | EMA provides position to UTD | High | UR4, UR5 |
| 12 | EMA receives routes | High | UR4 |
| 13 | EMA guides the users to the destination using multi-sensorial approach | High | UR3 |

Table 4a: Software building blocks to be supported for TASKA

| ID | UR2 building blocks | License |
|---|---|---|
| 1 | Casacore (https://github.com/casacore/casacore) | LGPLv2.0 |
| 2 | python-casacore (https://github.com/casacore/python-casacore) | LGPLv3.0 |
| 3 | DP3 (https://github.com/lofar-astron/DP3) | GPLv3.0+deps |
| 4 | AOFlagger (https://gitlab.com/aroffringa/aoflagger) | GPLv3.0 |
| 5 | DDFacet (https://github.com/saopicc/DDFacet) | GPLv2.0+deps |
| 6 | killMS (https://github.com/saopicc/killMS) | GPLv2.0+deps |
| 7 | LoSoTo (https://github.com/revoltek/losoto) | LGPLv3.0 |
| 8 | LofarStMan (https://github.com/lofar-astron/LofarStMan) | N/A |
| 9 | Dysco (https://github.com/aroffringa/dysco) | GPLv3.0 |
| 10 | Nenupy (https://github.com/AlanLoh/nenupy) | MIT licence |
| 11 | DynSpecMS (https://github.com/cyriltasse/DynSpecMS) | TBD |

Table 4b: Software building blocks to be supported for PER

| ID | UR2 building blocks | License |
|---|---|---|
| 1 | Openlink Virtuoso | GPLv2.0 |
| 2 | CesiumJS (https://github.com/CesiumGS/cesium) | Apache 2.0 |
| 3 | TripleGeo (https://github.com/GeoKnow/TripleGeo) | GPLv3.0 |
| 4 | InfluxDB | MIT |
| 5 | Kettle (https://github.com/pentaho/pentaho-kettle) | Apache 2.0 |
| 6 | Karma (https://github.com/usc-isi-i2/Web-Karma) | Apache 2.0 |
| 7 | Grafana(https://github.com/grafana/grafana) | AGPL3.0 |

## 7.2. Non-Functional requirements

The following table (Table 4) enumerates the non-functional requirements, incorporating the priorities established through stakeholder requirement definitions and subsequent prioritization. The table comprises five columns: ID, Category, Description, Priority, and Relevant User Requirement.

1. **ID**: Assigns a unique identifier to each requirement.
2. **Category**: Organizes the requirements according to the classifications presented in Table 2.
3. **Description**: Provides a succinct explanation of the requirement.
4. **Priority**: Indicates the importance level for the development, implementation, and deployment of each non-functional requirement in the project.
5. **Relevant UR**: Identifies the original user requirement from which the non-functional requirement is derived.

Priorities are classified into High, Medium, and Low categories. High priority requirements should be prioritized and implemented as necessary due to their critical nature. Medium and Low priority requirements, while still important, should be addressed after fulfilling High priority ones. Each requirement's ID is a unique identifier, and the Relevant UR links each requirement to the corresponding user requirement.

Table 4: Non-functional Requirements Categories

| Category identifier | Requirement category |
|---|---|
| DPT | Data Protection Requirements |
| SEC | Data Security Requirements |
| TERSPEC | EXTRACT Specific Requirements |
| US | Usability Requirements |

Table 5: Non-functional requirements

| ID | Description | Priority | Relevant UR |
|---|---|---|---|
| 1 | Optimization of the Processing to Observation (P/O) ratio | Medium | 1,2 |
| 2 | Selective storage through efficient raw data cadence selection | High | 1,4,6 |
| 3 | Data available to observers with adequate access rights within 24hr | High | 2, 3, 4, 5, 6 |
| 4 | Fidelity of radio interferometric video restoration and calibration | High | 2, 3, 5, 6 |
| 5 | Data processing and result production replicability | High | 2,3 |
| 6 | Fast computation of new simulation and prediction of the status of the city if new data is given to the system | High | 17, 18 |

| | | | |
|---|---|---|---|
| 7 | Privacy by default (select the minimal dataset to harvest) | High | |
| 8 | Privacy by design (ensure data protection) | High | |

# 8. Conclusions

The preceding analysis, focused on identifying and refining the functional and non-functional requirements of the TASKA and P.E.R. use-cases, has provided significant insights into the key elements necessary for the success of each scenario.
The TASKA use-case, influenced by environmental, and infrastructure factors, requires a resilient system capable of efficient operation in an unpredictable urban landscape. The P.E.R. use case, faced with the unique challenges posed by Venice's distinctive structure and population dynamics, demands an innovative approach to effective emergency management.
The requirements specified in this report have been carefully considered, validated, and prioritized based on their alignment with the defined Key Performance Indicators (KPIs) of each use-case. Stakeholder input has been instrumental in this process, informing the understanding of user requirements and shaping the trajectory of the project.
As a critical component within the larger project structure, the insights derived from this analysis will directly impact subsequent tasks and deliverables, contributing to the strategic fulfillment of the KPIs.
Looking ahead, the conclusions drawn from this report will serve as a guiding principle in our development process, assisting us in aligning our efforts with the specified requirements, thereby ensuring the success of both the TASKA and P.E.R. use-cases within the EXTRACT project.

# 9. ANNEX: P.E.R. Survey Analysis - Expert Perspectives on Personalized Evacuation Route Mobile App

This survey was conducted to ascertain the needs and preferences for a Personalized Evacuation Route (PER) Mobile App, specifically through the lens of expert judgement. The survey garnered responses from a select group of 10 individuals, mainly composed of individuals with professional backgrounds in the police force or as first responders. This unique respondent profile added substantial value to the survey, as their extensive practical experience and knowledge in handling emergencies and orchestrating evacuations provided nuanced, expert perspectives. This level of expert judgement is often hard to obtain from a general audience, rendering the insights garnered particularly valuable for this application.

With a significant majority of respondents (80%) owning a smartphone, their collective feedback highlighted the critical role of modern technology in managing emergencies. Even though most respondents currently do not utilize mobile apps for safety or emergency management, they all agreed on the essential need for a personalized evacuation route during emergencies.

Key factors identified for this personalization included age, physical fitness, and disabilities, with additional considerations given to technological skills and specific needs like wheelchair access. These preferences reflect their expertise and deep understanding of the varying needs that can arise during an emergency situation.

The respondents, drawing on their experience, unanimously highlighted the paramount importance of receiving real-time updates about hazards and safe zones during emergencies. Most respondents preferred to receive these updates via push notifications, indicating an understanding of the speed and accessibility of this form of communication in emergency scenarios.

When exploring technical features, the experts provided invaluable feedback. Most found the potential for the app to function offline, include a panic button feature, and enable hazard reporting to be useful. However, some expressed uncertainty about certain features, suggesting a need for careful consideration and user-guided development, likely based on their understanding of real-world emergency scenarios.

In a demonstration of their comfort with the necessity of certain trade-offs in emergency situations, all respondents were comfortable with the app using their personal data, such as location, to provide personalized evacuation routes. Furthermore, they unanimously stressed the extreme importance of the app's ease of use during emergencies, highlighting their understanding of the pressure and urgency of such situations.

Interface preferences leaned towards a simplified, schematic view of evacuation routes, likely reflecting the respondents' recognition of the need for clear, easily digestible information in high-stress environments. The language preference was primarily Italian.

The respondents' suggestions for the app to be intuitive and model it after the existing "112 Where are you app" (https://where.areu.lombardia.it/) reflect their expertise and practical understanding of emergency management tools. This expert judgement could prove to be an invaluable guide during the design and development phases.

In conclusion, while the survey has a small sample size, the high level of expert judgement provided by the respondents gives the findings significant weight. The nuanced insights gathered from these experienced individuals will undoubtedly provide a strong foundation for the app's development.

**Survey on Personalized Evacuation Route (PER) Mobile App Requirements**

Part 1: Personal Information

Do you own a smartphone?

Yes

No

Part 1: Personal Information

Do you currently use any mobile apps for safety or emergency management?

Yes

No

Part 2: Understanding Your Needs

How important is it for you to have a personalized evacuation route in case of an emergency?

Extremely important

Not at all important

Part 2: Understanding Your Needs

What factors should the app consider when generating a personalized evacuation route for you? (Consider factors like your physical fitness, any disabilities, etc.)

Part 3: Real-time Information and Alerts

How important is it for you to receive real-time updates about hazards and safe zones in your area during an emergency?

Not at all important

Extremely important

Part 3: Real-time Information and Alerts

How would you prefer to receive emergency alerts? (Choose all that apply)

Push notifications

SMS messages

Automated phone calls

Email alerts

Part 4: Technical Features

Would you find it helpful for the app to work offline (after initial setup)?

Yes

No

Unsure

Part 4: Technical Features

Would you find it helpful if the app included a panic button to immediately notify emergency services of your location?

Yes

No

Unsure

Part 4: Technical Features

Would you find it useful if the app allowed you to report hazards or blockages during an emergency?

Yes

No

Unsure

Part 5: Data Privacy

Are you comfortable with the app using your data (like location, personal information) to provide personalized evacuation routes?

Yes

No

Part 6: Usability and Interface

How important is the ease of use of the app during an emergency?

Not at all important

Extremely important

Part 6: Usability and Interface

Would you prefer the evacuation route to be displayed in:

A detailed map view

A simplified, schematic view

Both

Part 6: Usability and Interface

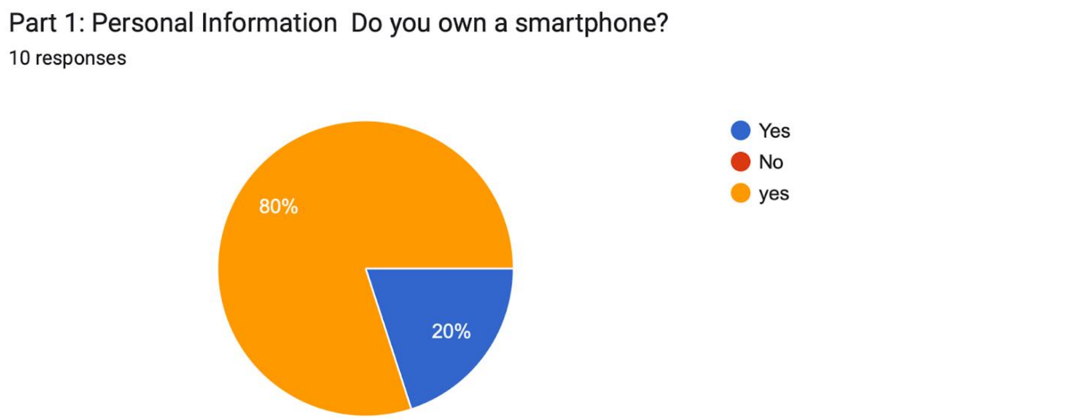
What is your preferred language for the app interface?

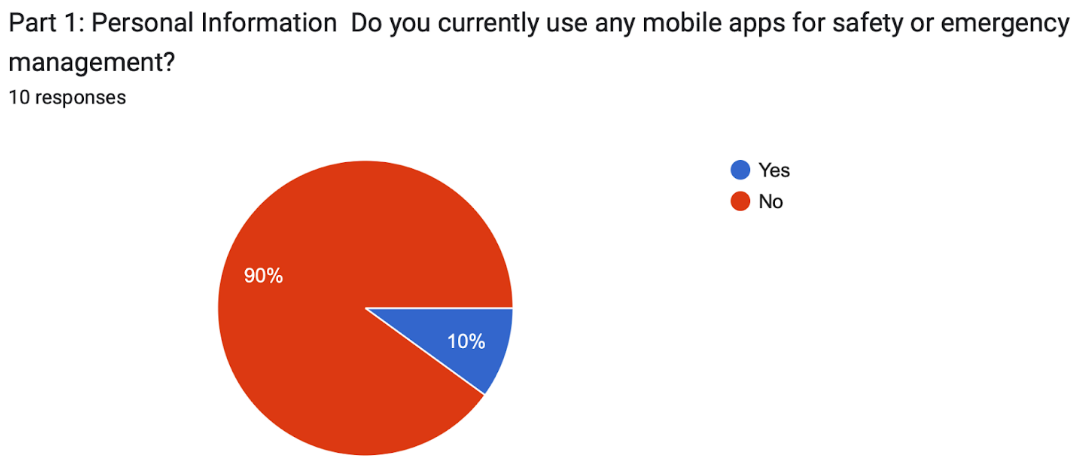
English

Italian

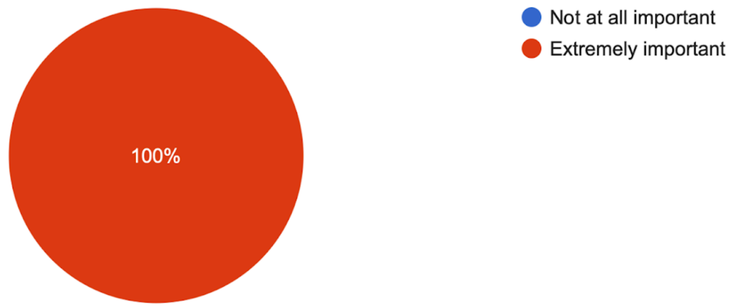
Part 7: Other Comments

Any other features you would like to see in the app or any other comments?



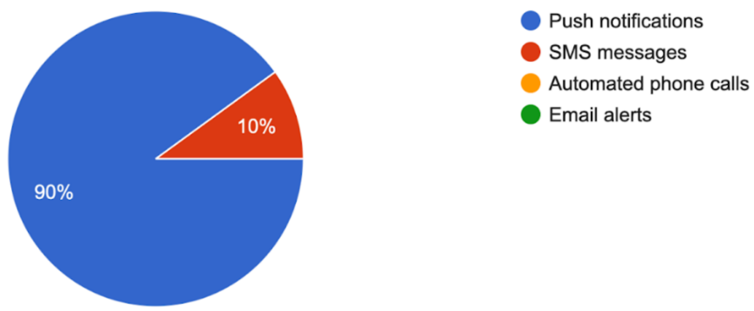



Part 3: Real-time Information and Alerts How important is it for you to receive real-time updates about hazards and safe zones in your area during an emergency?
10 responses



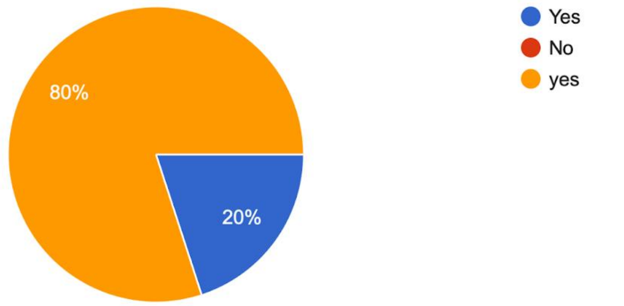

Part 2: Understanding Your Needs How important is it for you to have a personalized evacuation route in case of an emergency?
10 responses


Part 3: Real-time Information and Alerts How would you prefer to receive emergency alerts? (Choose all that apply)
10 responses



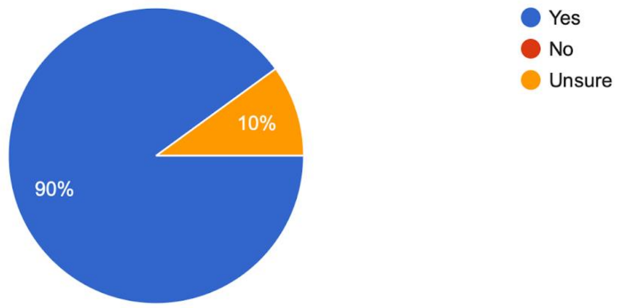
Part 1: Personal Information  Do you own a smartphone?

10 responses



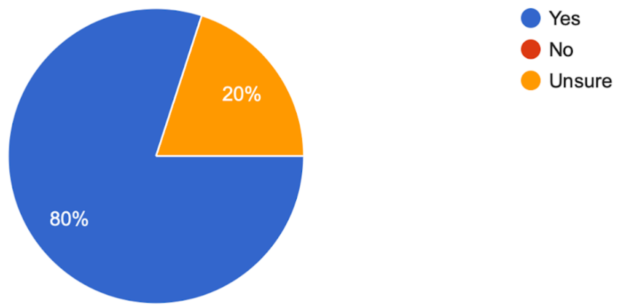

Part 4: Technical Features  Would you find it helpful for the app to work offline (after initial setup)?

10 responses



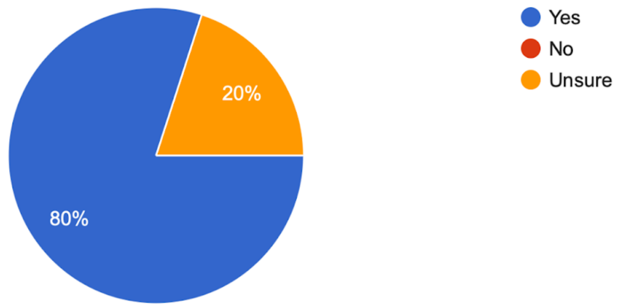

Part 4: Technical Features  Would you find it helpful if the app included a panic button to immediately notify emergency services of your location?

10 responses



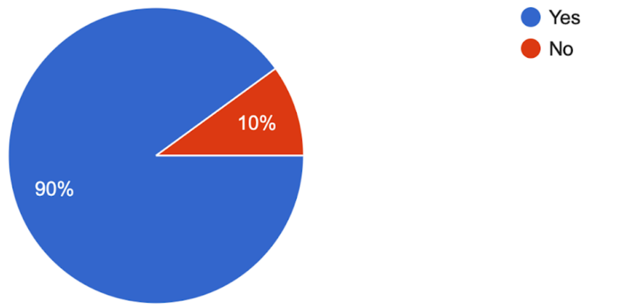
Part 4: Technical Features Would you find it useful if the app allowed you to report hazards or blockages during an emergency?

10 responses



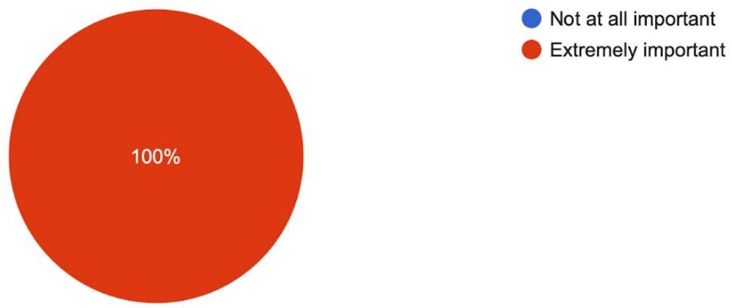

Part 5: Data Privacy Are you comfortable with the app using your data (like location, personal information) to provide personalized evacuation routes?

10 responses

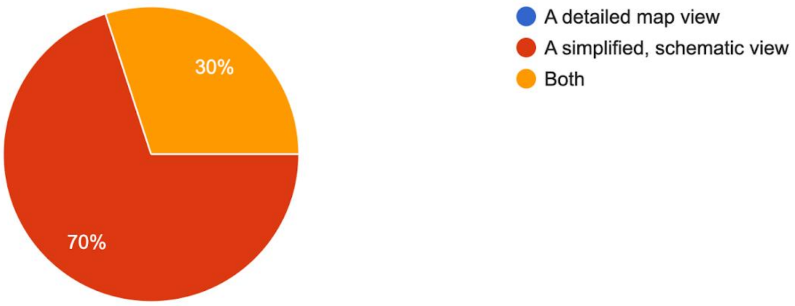



Part 6: Usability and Interface How important is the ease of use of the app during an emergency?

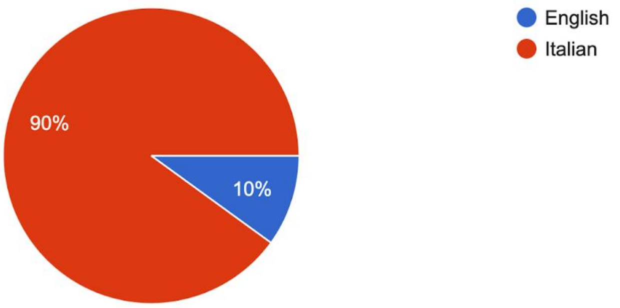
10 responses



Part 6: Usability and Interface  Would you prefer the evacuation route to be displayed in:
10 responses




Part 6: Usability and Interface  What is your preferred language for the app interface?
10 responses




Part 7: Other Comments  Any other features you would like to see in the app or any other comments?
10 responses