CONVERGE

view-to-communicate and communicate-to-view

CONVERGE - Telecommunications and Computer Vision Convergence Tools for Research Infrastructures

D1.1: Requirements and use cases

Work package	WP1	**Task**	Task 1.1, Task 1.2
Deliverable number	D1.1	**Due date**	31/07/2023
Editor	Helder Fontes (INESC TEC)		
Internal Reviewers	Paulo Marques (ALLBESMART LDA), Filipe B. Teixeira (INESC TEC), Filipe Conceição (INTERDIGITAL), Walid Dabbous (INRIA), Sérgio Silva (Adapttech)		
Version	1.0		
List of contributors (Alphabetically ordered, grouped by organisation)	Ana F. Nogueira (INESC TEC), Ana F. Sequeira (INESC TEC), André Coelho (INESC TEC), Filipe B. Teixeira (INESC TEC), Helder Fontes (INESC TEC), Hélder Oliveira (INESC TEC), Kamran Shafafi (INESC TEC), Luís Pessoa (INESC TEC), Manuel Ricardo (INESC TEC), Nuno Pereira (INESC TEC), Paula Viana (INESC TEC), Pedro Carvalho (INESC TEC), Rui Campos (INESC TEC), Antti Pauanne (UOULU), Miguel Bordallo Lopez (UOULU), Jorge García Vidal (BSC CNS), Maria-Cristina Marinescu (BSC CNS), Nadia Tonello (BSC CNS), Victor-Alejandro Ortiz (BSC CNS), Dirk Slock (EURECOM), Raymond Knopp (EURECOM), Serge Fdida (SORBONNE), Damien Saucez (INRIA), Okan Yurduseven (QUB), María García-Fernández (QUB), Guillermo Álvarez-Narciandi (QUB), Mondeep Saikia (QUB), Luis Pereira (ALLBESMART LDA), Paulo Marques (ALLBESMART LDA), Tiago Alves (ALLBESMART LDA), Ahmad Shokair (GREENERWAVE), Youssef Nasser (GREENERWAVE), Diogo Lopes (Adapttech), Frederico Carpinteiro (Adapttech), Sérgio Silva (Adapttech), Filipe Conceição (INTERDIGITAL), Remun Koirala (INTERDIGITAL), Juha Kela (Finwe Oy), Minna Rahkola (FinCloud Ltd.)		
Dissemination Level	PU ***PU:** Public, fully open, e.g., web (Deliverables flagged as public will be automatically published in CORDIS project's page)* ***SEN:** Sensitive, limited under the conditions of the Grant Agreement*		

Co-funded by the European Union

The CONVERGE project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101094831, including top-up funding by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee.

CHANGE REGISTER

Version	Date	Editor	Organization	Changes
A	20.03.2023	Paulo Marques	ALLBESMART LDA	Initial Deliverable structure
B	24.04.2023	Paulo Marques	ALLBESMART LDA	Table of Contents adjustments
C	22.05.2023	Helder Fontes	INESC TEC	Review contributions up to Chapter 4
D	19.06.2023	Helder Fontes	INESC TEC	Review Chapter 5 and Chapter 6 contributions
E	17.07.2023	Paulo Marques Filipe B. Teixeira	ALLBESMART LDA INESC TEC	Internal review
F	24.07.2023	Helder Fontes	INESC TEC	Quality Assurance review
G	27.07.2023	Filipe Conceição Walid Dabbous Sérgio Silva	INTERDIGITAL INRIA Adapttech	Review of additional round of inputs in Introduction, Tool 4 description and requirements, and UC1.4
H	31.07.2023	Luis Pessoa	INESC TEC	Document formatting ad minor corrections

EXECUTIVE SUMMARY

The main objective of the CONVERGE project is the development of an innovative toolset combining radio and vision-based communications and sensing technologies to enable an emerging area of research aligned with the motto "view-to-communicate and communicate-to-view", and through the integration of this toolset, advance the state of the art of a set of RIs to the greatest extent aligned with the ESFRI SLICES-RI.

This deliverable reports the use cases to be addressed by the CONVERGE project. The report starts by describing the set of 4 tools proposed in the project as well as the research questions that can be addressed by a researcher using each tool. Then, it describes the target user groups of the tools according to the classification of SLICES-RI and then identifies a relevant set of use cases addressable by the project, along 5 main vertical markets: telecommunications, automotive, manufacturing, media and health. For each use case, we describe its context and relevance, identify the different tools that may be used to address the use case and how they can help solving specific needs, and describe the data types involved. Finally, we present the proposed reference architecture for the implementation of the proposed toolset, describing their individual requirements, generated data types and associated interfaces. Finally, we discuss possible alignment opportunities with the ESFRI SLICES-RI.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AI	Artificial Intelligence
AoA	Angle-of-Arrival
AoD	Angle-of-Departure
BS	Base Station
C2V	Communicate to View
CNN	Convolutional Neural Network
CSI	Channel State Information
CV	Computer Vision
DoA	Direction-of-Arrival
E2E	End-to-End connectivity
ESFRI	European Strategy Forum on Research Infrastructures
FoV	Field-of-View
IMU	Inertial Measurement Unit
IoT	Internet of Things
LiDAR	Light Detection and Ranging
LIS	Large Intelligent Surface
LoS	Line-of-Sight
LSTM	Long Short-term Memory
MAC	Medium Access Control
ML	Machine Learning
NFV	Network Functions Virtualization
NLoS	Non Line-of-Sight
O-RAN	Open Radio Access Network
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RF	Radio Frequency
RI	Research Infrastructure
RIS	Reconfigurable Intelligent Surface
RRM	Radio Resource Management
RSSI	Received Signal Strength Indicator
SDR	Software Defined Radio
SLAM	Simultaneous Localization and Mapping
SLICES-RI	Scientific Large-scale Infrastructure for Computing/Communication Experimental Studies – Research Infrastructure

TDoA	Time-Difference-of-Arrival
UAV	Unmanned Aerial Vehicle
UE	User Equipment
VA-LIS	Vision-aided Large Intelligent Surface
V2C	View to Communicate

1. INTRODUCTION

Joint communications and sensing (JCAS), also referred in the literature as Integrated Sensing and Communication (ISAC) [Liu2022], has established itself as key trend towards 6G. The main reasons behind this convergence are associated with the evolution of both sensing and communications systems towards higher frequency bands (smaller size) and larger antenna arrays, leading to increasingly similar approaches in terms of hardware architectures, wireless channel parameters, and signal processing algorithms. Future wireless networks are hence expected to go beyond the current communications only paradigm and provide environment sensing services [Saad2019], making an efficient usage of wireless resources for both functions. Reconfigurable Intelligent Surfaces (RIS), also known as Large Intelligent Surfaces (LIS) are also expected to play a key role in 6G, potentially in combination with ISAC, mainly as a way of enabling the wireless environment to become reconfigurable while also leading to benefits in the energy/spectrum efficiency of wireless communications and better coverage, by providing new Line-of-Sight (LoS) sensing links to areas where Non Line-of-Sight (NLoS) or even a coverage hole was previously observed. This in turn greatly increases accuracy and resolution of radar sensing in those areas [Liu2022]. The large bandwidths available in higher frequency bands coupled with large antenna arrays provide improvements in communications capacity and allow for high spatial multiplexing of users with low interference. On the other hand, these characteristics also enable better sensing performance through improved radar range and angular resolution, leading to a more accurate environment mapping or discrimination between physical world objects/targets, enabling not only device localisation (active devices), but also mapping and even imaging of the surrounding environment and its objects (passive or device-free objects) [Don2022]. This research theme is stimulating investigations into applications in both communication-assisted sensing, as well as sensing-assisted communications.

This trend towards increasingly higher operation frequencies spanning into the millimetre wave and up to the sub-THz band intensifies the requirements for operation in line-of-sight conditions. This exposes an additional opportunity which has been little explored so far, namely the employment of computer vision solutions. In fact, while telecommunications and computer vision have evolved as separate scientific areas, this is envisioned to change with the advent of wireless communications with radios operating in increasingly higher frequency bands, characterised by line-of-sight operating ranges.

The evolution of communications both in terms of requirements and their applications to new use cases leads to increasingly stringent demands in terms of data-rates, latency and power consumption, creating a necessity to surpass them with low hardware complexity (for low power consumption) but also with almost instant response times (for real-time interaction). Thus, associating computer vision with communication can be a potential solution for many challenges communication systems face and vice-versa. On the one hand, computer vision by using visual data can obtain the spatial information of the communications surroundings, allowing to estimate communications parameters such as propagation directions, wireless equipment received power and blockage status [Xu2023]. This allows, for example, to do beam alignment and tracking [Xu2023, Li2023], to obtain the target radiation pattern for beam steering [Los2023], potentially minimising road accidents by recommending the speed to the drivers based on traffic information gathered from videos of monitoring systems [Sne2023] or create an optimized scheme of flight trajectory and power allocation for mmWave Unmanned Aerial Vehicle (UAV) communication systems [Hua2023]. On the other hand, wireless radio communications, especially in the sub-6 GHz frequency range, are particularly immune to blockages/occlusions, privacy concerns or bad illumination conditions improving computer vision methods in tasks like pose estimation and tracking [Lee2023, Zho2023] or action recognition [Com2023]. This demonstrates the complementary nature of the two domains. Therefore, continuous research is needed for enabling advanced wireless communications systems leveraging on sensing and vice-versa, therefore it is essential to have adequate Research Infrastructures (RI) and toolsets available for further developments which are the core goals of the CONVERGE project.

The main objective of the CONVERGE project is the development of an innovative toolset combining radio and vision-based communications and sensing technologies to enable an emerging area of research aligned with the motto "view-to-communicate and communicate-to-view", and through the integration of this toolset, advance the state of the art of a set of Research Infrastructures (RIs) to the greatest extent aligned with the ESFRI SLICES-RI (Scientific Large-scale Infrastructure for Computing/Communication Experimental Studies). This new area of research departs from the traditional and isolated research in each of the fields and aims at creating new knowledge and discoveries at the intersection of wireless communications, computer vision, sensing, and machine learning.

This toolset is a world-first and consists of 4 main tools: 1) vision-aided large intelligent surface, 2) vision-aided fixed and mobile base station, 3) a vision-radio simulator and 3D environment modeler, and 4) machine learning algorithms for multimodal data including radio signals, video streams, RF sensing, and traffic traces. An integrating part of this toolset will be the user interface, aimed at managing the user access and supporting toolset configuration, namely for configuring the experimental chamber or room environment where the toolset is physically installed. This toolset will be deployed into 7 RIs mostly aligned with the ESFRI SLICES-RI aiming at improving their competitiveness. The toolset and associated RIs will be used by the researchers and industry to study several scenarios in different market verticals of relevance for Europe including telecommunications, automotive, manufacturing, media and health as further described in the next sections of this document.

The main goals of this document are the following:

- Identify a relevant set of use cases addressable by the project, along 5 main vertical markets: telecommunications, manufacturing, automotive, media and health.
- Identify the different tools that may be used to address each use case and how they can help solving specific needs.
- Describe the data types involved in each use case.
- Gather a set of research questions that can be addressed by a researcher using each tool.
- Describe the target user groups of the tools according to the classification of SLICES-RI.
- Analyse the main technical requirements involved in the implementation and integration of the CONVERGE research infrastructure.
- Present the CONVERGE reference architecture, detailed individual components, data types and associated interfaces.

2. CONVERGE TOOLS

The toolset to be developed by CONVERGE, operating under the motto “view-to-communicate and communicate-to-view”, aims at being totally or partially deployed in research infrastructures, primarily within a controlled room or chamber available within the RI, such as an anechoic chamber, which we name here as the vision-radio experimental chamber. When totally deployed, the toolset may have the configuration shown in Figure 1.

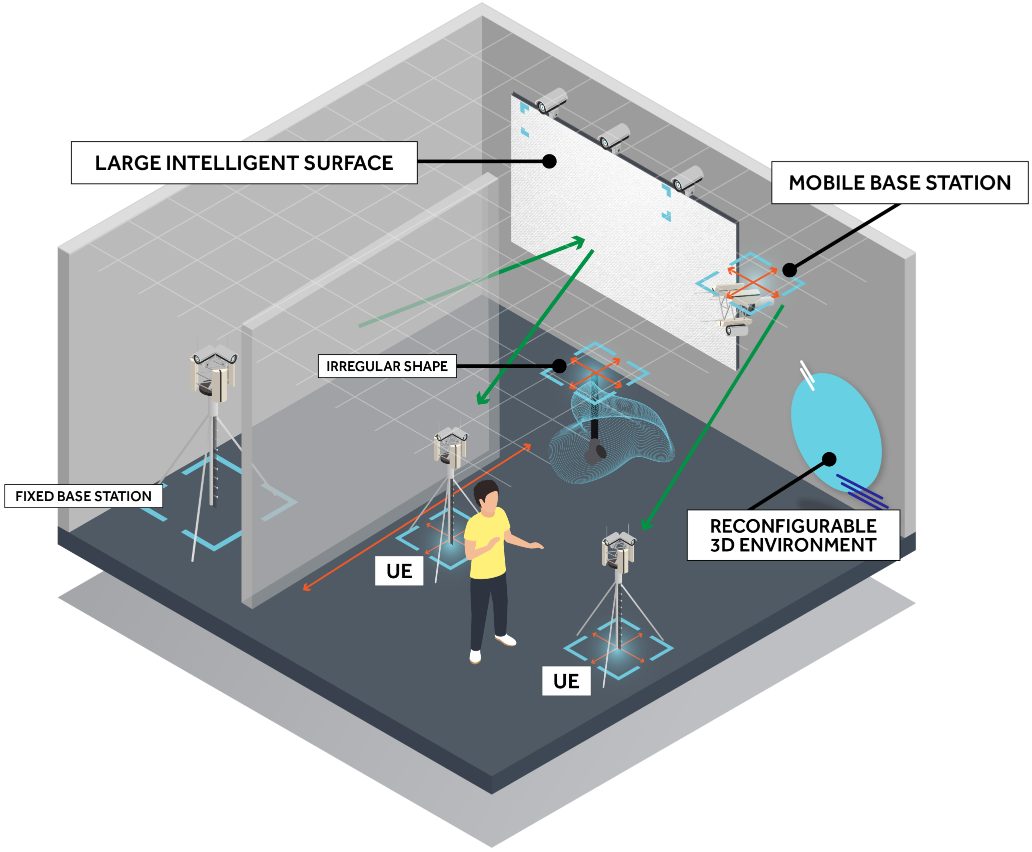


Figure 1 - High-level concept of the proposed vision-radio experimental chamber.

The toolset to be developed by the CONVERGE project will be a world-first, consisting of the four tools introduced in the following sections.

2.1 Tool 1: Vision-aided Large Intelligent Surface

CONVERGE Tool 1 – vision-aided large intelligent surface (VA-LIS) aims to allow experimentation, in a controlled room environment, of massive MIMO wireless communications, high precision 3D positioning and environment sensing, including human sensing and microwave holography, through the combination of a smart programmable meta-surface antenna with a video camera array (see Figure 2). Two LIS versions will be developed, addressing both sub-6 GHz and mm-Wave frequency bands, enabling research based on multi-frequency data.

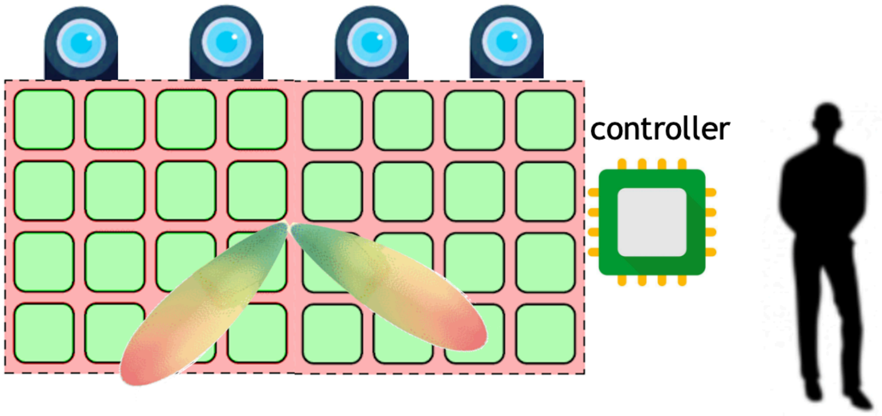
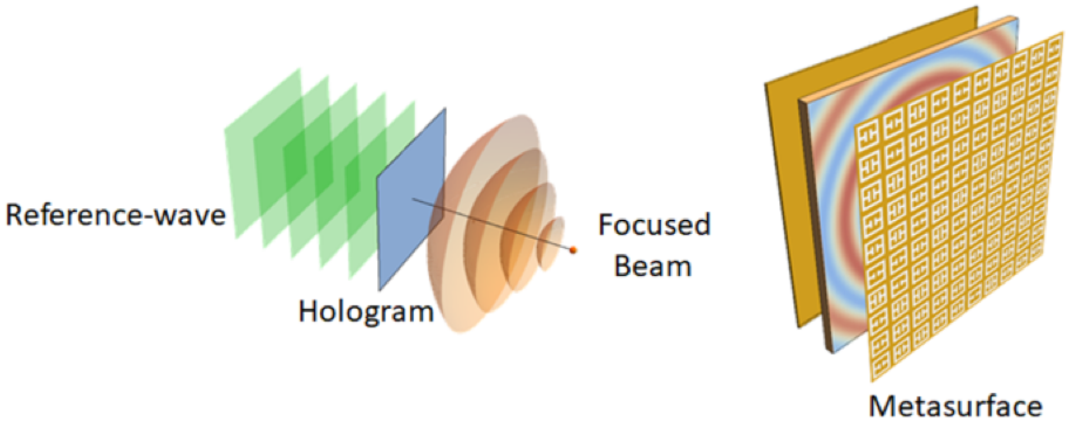


Figure 2 - Vision-aided Large Intelligent Surface.

The tool consists of a large antenna array articulated with a large video camera array. The antenna array is achieved in a modular fashion through the combination of sub-array tiles, which can be assembled in any 3D form beyond a flat surface (e.g., concave, cylindrical) depending on the specific focus of the research to be supported and on the availability of a space for its installation. Both active and passive modes of operation may be envisioned for the antenna array. The active mode corresponds to the capability of generation and reception of radio signals (within the array) which enables massive MIMO communications and beamforming with arbitrary control of the number of beams and beam focus in the 3D space, and the possibility of controlling multiple sub-arrays simultaneously. The passive mode corresponds to the capability of controlling the reflection phase of the incident electromagnetic wave at each antenna element, therefore enabling the same functionality of the active mode but using signals generated outside of the array.

An advantage of the intelligent surface technology is that it offers a significant number of degrees of freedom in electromagnetic wave control, for example by means of allocating different frequencies to different functionalities. By controlling the complex weights of the meta-atoms, the radiation pattern can be dynamically scanned to provide coverage in an arbitrary direction or synthesise quasi-random radiation patterns. In Figure 3, a software-controlled intelligent surface is shown to radiate spatio-temporally varying, quasi-random radiation patterns to probe the scene for imaging [Yur2016a, Ima2020]. Leveraging the compressive-sensing theory, probing the scene information using a random subset of these measurements enables an estimate of the scene to be reconstructed, a promising indication for unlocking a plethora of communications and sensing applications at different parts of the electromagnetic spectrum.

The interesting design of this LIS is the number of degrees of freedom that it offers to multi-target operations through the possibility to control each of its unit cells phases. Its design can be also adapted in size, in number of phases per element, in its type (active vs. Passive) and many others. It can be easily coupled with the cameras.



(a) (b)

Figure 3 - Microwave holography applied to intelligent surfaces: (a) a hologram converts a reference-wave to a focused beam (b) an example metasurface leveraging a guided-mode as input (i.e., reference-wave).

As an integrating part of this tool, an array of multiple cameras will be used allowing to create an aggregate virtual camera with greatly improved performance. Although one could design optical systems that ensure a common centre of projection for all of the cameras, these systems become costly and complex as the number of cameras grows. Instead, we pack the cameras as closely as possible to approximate a single centre of projection and compensate for parallax in software. By tightly packing the cameras and placing them with abutting or partially overlapping fields of view, it is possible to create an equivalent high-resolution video camera. This configuration is helped by the use of existing techniques from the image mosaicking literature, which makes it possible to register and blend the images to create a single image of high resolution. One advantage of using multiple cameras is that we can meter them individually. This allows us to capture scenes with a greater dynamic range than individual cameras. For scenes in which the local dynamic range exceeds our sensors' capabilities, we can trade resolution for dynamic range by increasing the overlap of the cameras' fields of view, so that each viewing ray is observed by multiple cameras with different exposure settings. This camera system enables one to gather a-priori estimation of the location coordinates of a constrained field of view that contains the object of interest in a cluttered environment for the experimentation of low computational complexity computational imaging at microwave frequencies using the intelligent surface [Yur2016b, Gol2017]. The combination of an intelligent surface with computer vision technology has also been shown for the first time in 2021 [Los2021] as a promising approach for interference reduction in future wireless networks.

2.1.1 Examples of research questions that Tool 1 helps to address

- How can the vision aided LIS identify and estimate the position of the UEs and the obstacles (both range/depth and direction) from the vision information?
- How can LIS deal with UE/obstacle tracking aided beam management?
- How can beamforming be optimized with the vision data to improve communication performance?
- How can the obstacles be classified as interfering UEs/scatterers (permeable/non permeable)?
- Can the information about interfering UEs aid in reduced network interference?
- What is the required system design perspective for joint communications and sensing enabled by a vision aided LIS and how can the related design parameters be optimized?
- What are the potential ways to link the antenna design/analysis phase with the physical model simulating a controlled EM environment, to study communications and localization scenarios?
- Based on the CONVERGE selected scenarios, what are the required system parameters to ensure acceptable performance (i.e., the electrical size of the LIS, modulation technique, reconfiguration speed, gain, etc.)?

- How can the LIS (and the overall communications system) benefit from sensor fusion (optical data, video cameras, LiDAR etc) for beamforming and sensing?
- How can the LIS deal with mobility in terms of vision aided UE and obstacle tracking (i.e., in a dynamically changing environment)?
- How are the functionalities of the LIS impacted by static vs dynamic environments and what are the hardware and software related solutions for mitigation? What trade-offs are involved?
- What are the signatures that can be learned directly from the measured LIS channel data?
- For sensing related tasks, can the reconstruction step be eliminated? Can LIS work with ML to help with joint communications and sensing?

2.2 Tool 2: Vision-aided base station

CONVERGE Tool 2 aims at enabling communications and experimentation with mobile terminals mostly related to beamforming, multi-user access and opportunistic scheduling by taking advantage of environment mapping made by video-cameras and of the LIS.

Figure 4 shows an overview of this this tool, consisting of both fixed and mobile versions of the base station that will be developed, the latter adding the possibility of controlled mobility (different positions along the time or predefined trajectories), while also taking advantage of video cameras and/or LiDAR and enabling its cooperation with the fixed base station and the LIS. Vision-aided mobile user equipment (terminals) will complement the tool, aimed at exchanging traffic with the base stations and obtaining multiple video perspectives of the environment.

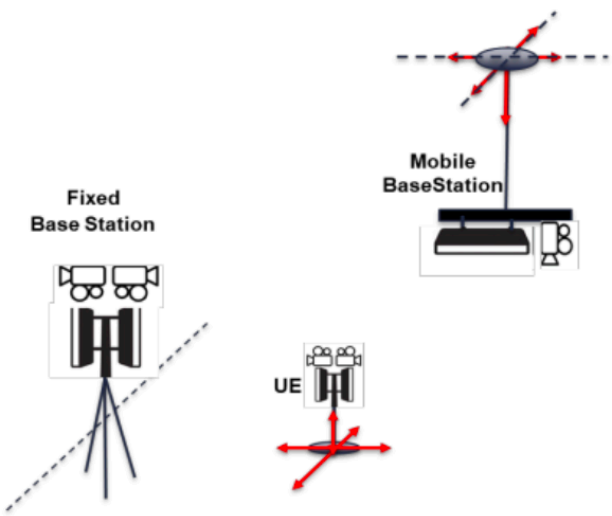


Figure 4 - Vision-aided fixed and mobile Base Stations, and complementary User Equipment.

The fixed base station will be similar to the one shown in Figure 5, which will still be moveable if needed, through a manual repositioning process. Nevertheless, repositioning of the base station should be done with high precision, or avoided, if possible, to ensure reproducibility of the experiments since even small placement or antenna positioning variations could lead to different RF and vision conditions. The mobile version of the base stations will be supported by a robotic arm or crane system, which will allow precise trajectory control along the time, enabling repeatable and reproducible experimentation in a mobile setting. The left-hand picture in Figure 5 shows the types of radio units that will be used while the right-hand picture in Figure 5 shows the servers and switch for running the base station (gNB-DU) equipment as well as the time-sensitive networking equipment (PTP switches) needed for synchronization of the radio and vision systems. The gNB and the 5G core network will run open-source OpenAirInterface code. Both mobile versions such as this one or the equivalent deployed in server rooms of INRIA/EURECOM will be used in the project.

Figure 5 - Example of a fixed Base Station, without the vision component represented.

2.2.1 Examples of research questions that Tool 2 helps to address

- How to detect the location of obstacles to signal propagation, interfering terminals, and terminals served by the vision-aided mobile base station?
- How to jointly optimize beamforming, position, and traffic schedule in the vision-aided mobile base station so that the Quality of Service offered to a set of terminals is maximized over time?
- Can visual information combined with other sensor data improve localization, Simultaneous Localization and Mapping (SLAM), and object tracking accuracy in dynamic environments, compared to traditional methods?
- How does incorporating visual information impact the QoE for UEs in terms of throughput, latency, and reliability?
- Which machine learning techniques are better suited to enable dynamic collaborative tracking by incorporating information from multiple base stations or cameras within a network, under variable environmental conditions or UE behaviour?
- How to learn from previous scenarios, similar or not?
- How to combine a mobile vision-aided base station operation with a LIS so that the communications Quality of Service (QoS) offered to a set of terminals can be maximized and sensing can be performed?
- How to take advantage of a digital twin of the anechoic chamber?
- What sensing strategy shall be used for complementing the visual information gathered by its video-cameras?
- How to form multiple sensing beams using radio sensing capabilities and process their echoes?
- How to position the mobile base station considering a multimodal visual capture system?

- How to combine images from radio and cameras? In particular, how to enrich traditional SLAM with image-based information, and how to perform image inpainting with information extracted from radio signals?
- Are new methods and procedures required to extract radio information from gNB receivers and transport them via O-RAN-like interfaces to an edge cloud? If so, how to represent this information efficiently?
- How to transport vision and radio information from User Equipments. What new signalling is required beyond standardized 5G procedures?
- Can we devise integrated sensing and communication protocols for such information?
- How to we jointly model synchronized radio and vision information to train ML algorithms at the edge?
- Can measured UE radio impairments be used to "fingerprint" UEs (i.e., different chipsets, different mobile integration platforms, etc.)?
- Based on the vision-based data from multiple UEs, how can the system merge the images?
- How to predict network demand to allocate resources and place the BS with the aid of vision?
- How to fuse vision information with estimated channel data for precoding?
- How can the BSs coordinate for seamless predictive handover aided by vision?

2.3 Tool 3: Vision-radio simulator and 3D environment modeller

CONVERGE Tool 3 is designed to generate digital 3D representations of environments, thereby enabling the simulation of observations from any location and the creation of geometric models appropriate for radio signal propagation simulations. These simulations will leverage data derived from material dielectric properties and antenna geometries, creating an electromagnetics-augmented 3D environment model for a comprehensive vision-radio simulator.

Figure 6 showcases an example of an indoor room to be modelled with the multiple visual and radio signal inputs, with Figure 7 showing a detail of a corner modelled as a diffraction edge. These inputs are combined in the tool to produce a 3D vision-radio environment. Central to this tool is a sophisticated ray-tracing software, capable of utilizing data from visible light and radio signal interactions with the environment.

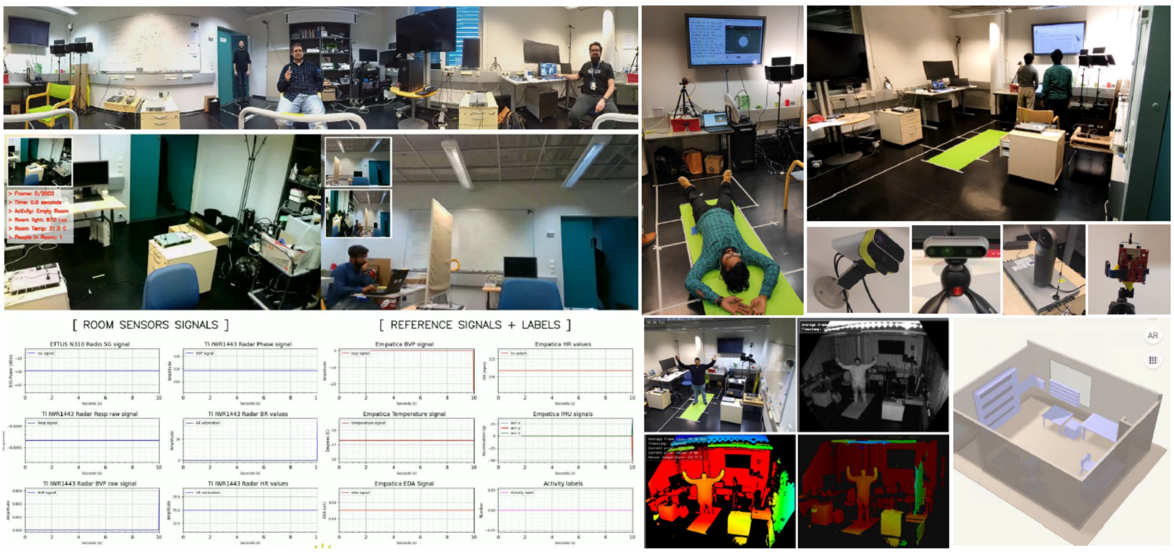


Figure 6 - Sensorized indoor room with multiple cameras (visual) and communication devices (radios) which form the input data for the environment modeler.

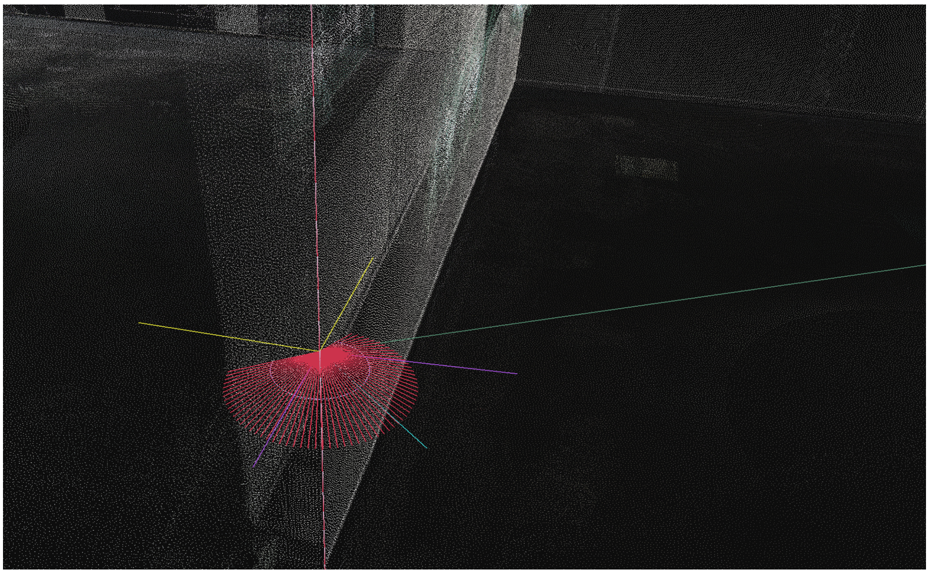


Figure 7 - A detail of a corner modelled as a diffraction edge.

The data provided by multiple cameras (images and videos) and wireless communication devices (distances and material properties) is then transmuted into a dynamic 3D model, which can be manipulated and observed from any perspective, offering an intricate and adaptable representation of the environment using the following modules:

1. **Triangle Mesh Creation**: This module allows for the creation of 3D triangle-mesh models to represent the environment accurately. Both explicit 3D (e.g., LiDAR-based) and implicit 3D (e.g., camera-based NERFs) models could be supported. By exploiting existing software packages, any viewpoint can be generated, with limited support for moving objects. The software integrates reflection, diffraction, and absorption models for light and radio signals, providing a comprehensive environment simulation.
2. **Mapping Radio Signals**: This module includes parallel implementation of radio path propagations, utilizing technologies such as CUDA/OptiX. It can accommodate various approximate models at different resolutions, including the number of triangles and ray launches, thereby facilitating a range of research requirements. This feature enables the tool to model beamforming strategies efficiently, contributing to more accurate signal propagation simulations.
3. **Sensor Data Flow Modelling**: This module is designed to construct environment models based on realistic data flows from cameras, LiDARs, and radios. It can handle heterogeneous sources of data, accounting for different spatio-temporal distributions of data. The software builds geometric representations through algorithms that support data-flow inputs, ensuring the accurate and efficient conversion of sensor data into 3D geometric models.

The simulator is designed to not only create environmental models but also to enhance our understanding of the accuracy required from geometric and electromagnetic models to achieve and surpass the performance of statistical channel models in radio communications. This is accomplished by providing a platform for testing and refining these models, thereby aiding the optimization of radio signal propagation, and resulting in improvements in the efficiency and reliability of wireless communications. Leveraging on the geometric and electromagnetic models, bit and error rates will be estimated for cellular communications happening in the environment. Cellular communication systems permanently change their communication scheme to adapt to the conditions via dynamic update of their modulation. Because modulation is used for communications, amplitude, frequencies, and phases have time dependencies, meaning that these parameters are not constants in the system but instead dynamic with non-null derivatives. Moreover, the simulator will consider the PHY and MAC used in 5G for radio spectrum scheduling. This means that short time control loop (i.e., 1ms) will be considered, hence the need of efficient simulation schemes to keep simulation tractable. The simulator can be used to feed machine learning algorithms as it allows to generate a plethora of cases and compute their behaviour. It also allows evaluate special cases that would be complex to make in practice (e.g., large number of devices, interactions with humans, etc.).

Figure 8 provides an example of a snapshot of a possible 3D model generated by the environment modeller and simulator tool. The image highlights the tool's ability to create a detailed and accurate representation of an environment, using data from both visible light and radio signal interactions.

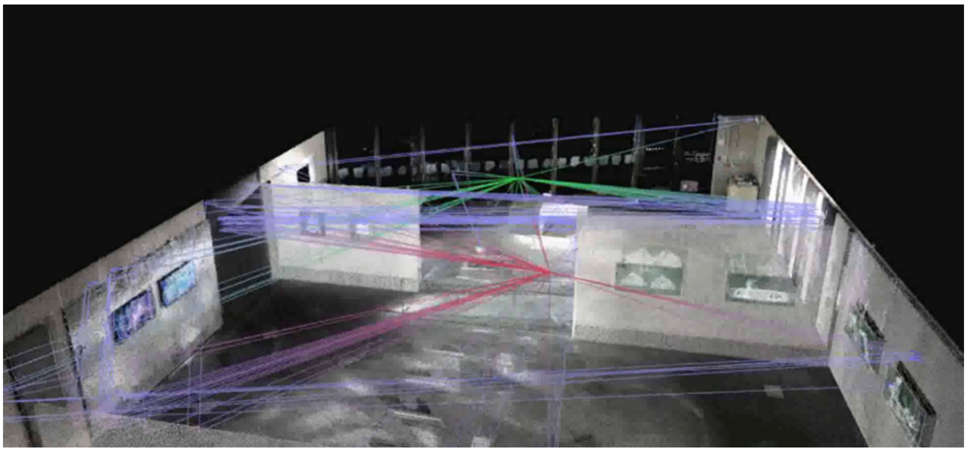


Figure 8 - Snapshot of a vision-radio 3D environment with radio propagation paths.

This tool will offer a platform for researchers to gain valuable insights into the complexities of radio signal propagation in different environments. This is anticipated to lead to advances in wireless communication technology, with the tool representing a significant step in the ongoing efforts to improve the efficiency and reliability of wireless communications systems. The tool can be used to simulate a wide range of scenarios and environments, and can provide valuable insights into the behaviour of radio signals and wireless networks. Complementary, the data provided by the simulation of communication devices can improve computer vision tasks such as depth perception, 3D reconstruction or object classification and tracking.

2.3.1 Examples of research questions that Tool 3 helps to address

- How will radio signals propagate in a particular environment, and what factors will affect the strength, quality, and performance of the signal? How accurate are existing statistical channel models compared to the accuracy of the 3D environment modeller and vision-radio simulator tool?
- How will different antenna designs perform in a specific environment, and how can they be optimized to improve performance? What are the best beamforming strategies (including statistical and learning-based) to use in a specific environment, and how can they be optimized to improve performance?
- What is the optimal placement of wireless sensors in a particular environment, and how will their transmissions affect the load of the network and the quality of observations made by other sensors?
- How can researchers plan experiments that use an experimental chamber, and how can they simulate those experiments if they do not have physical access to the chamber?
- How to detect obstacles affecting the communications between the mobile base station and terminals? How to combine a vision-aided communication operations with a large intelligent surface so that the communications QoS offered to a set of terminals is guaranteed?

- How to simulate the indoor, outdoor, and mobile environment in the chamber? How to account for the mobility of UE? How to merge the images from multiple cameras to form the visual representation?
- What information should be gathered from each measurement? How can information with high data volume be transmitted? Should there be any compression on the images for reducing computational complexity?
- What is the computational complexity of the radio frequency (RF) and optical-related tasks? How to integrate the reconfigurable intelligent surface (LIS) into the simulator, and what are the relevant system parameters for this? What is the desired system resolution and the corresponding aperture size to achieve this resolution?
- What is the best way to deal with RF and optical vision data in terms of computational complexity and maximizing the benefit of sensor fusion?
- How will view-to-communicate and communicate-to-view scenarios perform in a specific environment, and what are the communications, computing, and energy costs associated with these scenarios? How can machine learning methods of CONVERGE Tool 4 be integrated in the simulator (Tool 3) and be used to improve the accuracy of view-to-communicate and communicate-to-view scenarios? What experimental campaigns can be designed in simulation so they can be conducted to support this?
- If the vision-aided mobile base station (Tool 2) is equipped with radio sensing capabilities, can we simulate different sensing strategies to be used for complementing the visual information gathered by its video cameras? How can we simulate the formation of multiple sensing beams and process their echoes?
- How can we simulate visual information using an accurate 3D model of the environment, and how can these images be used for image classification? How should the information from the 3D model be associated with the information from the images? How can we simulate the effects of imperfect synchronization of the temporal information from LiDAR and the multiple cameras?
- How to optimize the location of the base station by simulating multiple positions and beamforming strategies? How to deal with the mobility of the base station and/or variations in the environment during data acquisition? How to leverage the optical data to simulate and reduce the complexity for electromagnetic (EM) propagation and related tasks? How can we simulate differences in the synchronization the EM and vision data both in time and space and what is the effect of error in synchronization between different sensors?
- Research question related to the implementation of the Tool 3, and its integration within the CONVERGE toolset: Is the trade-off between a full edge computing approach and an edge/cloud approach? What are the best metrics/approaches needed to evaluate the fusion mechanisms? If using machine learning approaches, what are the characteristics needed for the datasets and how to collect/create them?

2.4 Tool 4: Machine Learning (ML) algorithms

CONVERGE Tool 4 aims at facilitating the processing of heterogeneous data made available by the above tools and external datasets, including videos from cameras and RF sensing signals from the large intelligent surface, as well as timestamped and space-referenced communications traces, including received powers, signal-to-noise ratios, antenna radiation patterns, objects positioning, and traffic performance indicators such as bitrates or delays.

The goal of this tool is to be a generalized solution for V2C & C2V (view-to-communicate and communicate-to-view) data analysis. The tool aims to be a *generic* modular solution and to provide a *pipeline of functional interactive*[1] *modules* that users may choose from and interconnect, depending on:

a. The input type (e.g., image or structured numerical data) and format.
b. The nature of the problem (e.g., incremental vs batch, object recognition, pattern prediction, etc.).
c. The desired output (and format).

Tool 4 will provide transversal support for the rest of the tools in the project and is designed to be used by two different kinds of users, depending on their permissions:

- The non-expert Data Analyst users, for which we provide some default configurations and visual tools in each data processing stage to easily get started with data analysis. This functionality allows anyone to configure and use the tool to execute standard data analysis.
- The experienced Machine Learning users, for which we provide fine-tunning capabilities to suit more advanced needs and extend the existing capabilities of the standard analysis tool. Advanced users will be able to:
  - Modify the existing functionalities to better adjust them to their data processing needs.
  - Add new functionalities, such as different data transformation or synchronization algorithms, new training algorithms, additional evaluation metrics, new visualization methods, etc.
  - Dismiss modules that do not add value to their analysis to reduce the overhead of the tool.

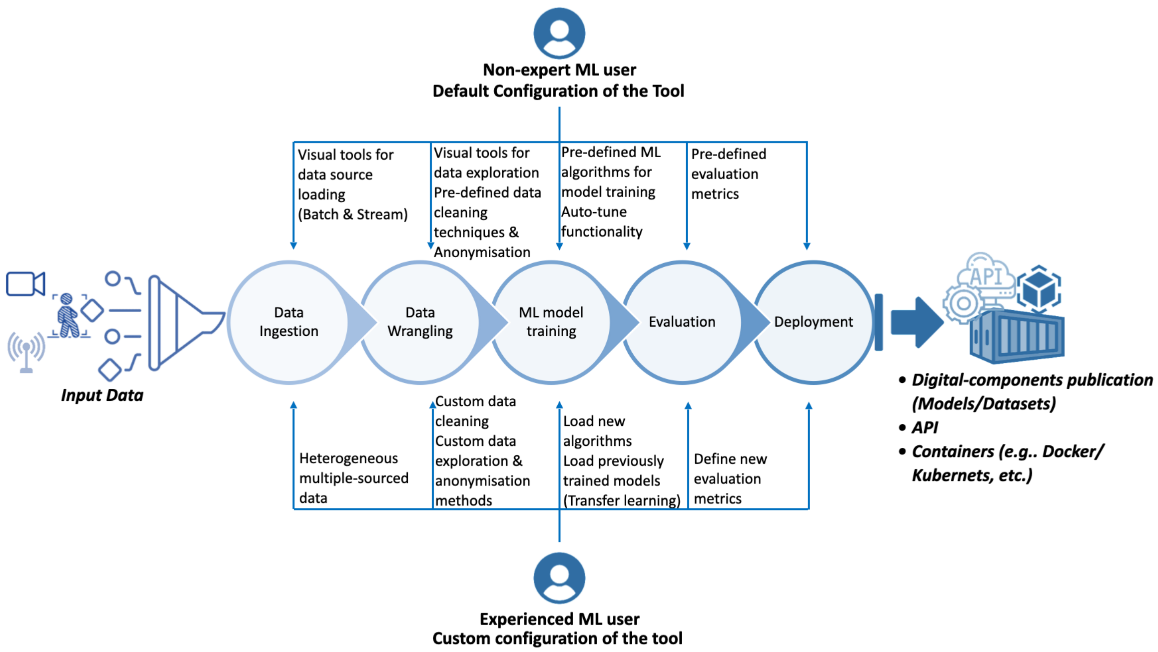


Figure 9 - ML Tool for V2C and C2V data analysis.

Figure 9 illustrates the different data analysis phases that the tool covers, as well as functionalities that each of these may include for the two types of users, where the default functionality is provided for the non-expert ML user, and the experienced ML practitioner can customize according to their needs.

- **Data Ingestion.** Refers to the process of obtaining data from a single data source or from multiple heterogeneous data sources.

[1] Interactivity in the sense that different available modules should be able to interact between themselves, e.g., one module of batch image processing should be able to interact/use the anonymization module for images.

  - As a default functionality, visual tools to load batch datasets and to connect streaming data sources will be provided. The extension to load and synchronize heterogeneous data from multiple independent sources (by time, event, or geo-localization) is possible for the expert users at a coding level, by allowing implementing and integrating external software packages with existing available modules.
- **Data Wrangling**. Covers the process of exploring the characteristics of proprietary data and transforming it into formatted data appropriate for ingestion by other modules of the tool.
  - Visual tools for standard Data Exploration and Analysis, including techniques that apply to specific data formats (sensor data, images, video steaming, etc).
  - Visual pre-defined basic Data Cleaning methods will be available as a default. Expert users will be able to extend or modify such methods by developing and integrating additional modules.
  - Standard data transformation techniques, such as outlier detection, data imputation, data structuring, data augmentation, amongst others.
  - Standard anonymization methods and privacy preserving capabilities for the datasets.

  All previously mentioned methods can also be modified and extended by expert-users by accessing the source-code. After the Data Wrangling process, if the user´s only intention is publishing the dataset, one can do this directly by going to the deployment process.
- **ML model training.** This phase defines the overall approach of the learning task.
  - Machine Learning algorithms of different types (Online learning, Bio-inspired, Supervised, Semi-supervised and Reinforcement Learning) are provided for the user to select from. An Auto-tune functionality is available for the user to easily find the optimal set of hyper-parameters for the algorithm of choice.
  - Expert users can extend and optimize those functionalities by defining custom training approaches, loading pre-trained models for Transfer Learning and/or specifying their desired set of hyper-parameters.
  - Simulators are efficient solutions to feed ML models in the training phase.
- **Evaluation**. Uses standard evaluation metrics provided by the tool as default configuration, such as accuracy, precision, and recall.
  - Non-expert users are able to assess the model performance using a test set, using Hold-out or Cross validation methods.
  - Expert users can further define different methods or metrics to evaluate the strengths and weaknesses of their model.
- **Deployment.** Publication of digital components such as pre-trained ML models for transfer learning, ML trained models as a service (API) or as an image (docker/Kubernetes), and datasets as endpoint or download files.

2.4.1 Examples of research questions that Tool 4 helps to address

- How to detect and/or classify objects present in RF images and RF-optical images? What are the best ML algorithms to do this, given the size and characteristics of the data set? How to identify users to detect unauthorized users to a facility? How to do motion prediction and implement an object tracking module based on visual and RF data?
- Based on data from cameras or LiDAR sensors, how to best predict signal blockage and trigger a vision-aided proactive handover? What type of training and fine-tuning will be necessary to achieve good predictions for action and emotion recognition? Examples of actions are: walk on stairs, walk on flat, walk on ramp, turn right, etc. Emotion recognition refers to facial expressions.
- How to detect anomalous situations such as objects out of context or unusual actions? How to identify gait patterns as a way to detect people in a more privacy-preserving manner?
- How to synchronize and fuse data of different types from different sources (e.g., sensing and vison devices). How to predict the evolution of the environment, e.g., whether blockages will occur, what is the direction or velocity of objects, etc.

- How can ML help improve the resolution of objects in the reconstructed images and minimize the distortions created by moving obstacles?
- Given the computational complexity, memory, or latency requirements, how should the ML algorithm be structured and where should they run?
- Given that actual experiments are time-consuming and may need specific on-site interventions, what is the trade-off that defines the minimal subset of experiments to actually perform in the physical chamber (with tool 3) to enable reasonable performance of the ML tool?
- Given that there is some flexibility in where to locate some of the sensors (radio and video), how can the ML tool help in defining the best locations of the different components, for each particular case?
- How can the tool leverage on sensing data semantics to achieve the target use case goals?

3. CONVERGE TARGET USER GROUPS

CONVERGE target user groups are the communities that should benefit from the CONVERGE infrastructure. It is of utmost importance to clearly identify the different groups, engage a dialogue with them and monitor their demand. This is essential in order to maximize the value of the CONVERGE tools. Note that the CONVERGE target user groups will benefit but also extend the ESFRI SLICES user groups CONVERGE clearly enrich and extend the SLICES-RI test platform.

CONVERGE target user groups as initially identified are listed hereby. Table 1 represents the target audience for the CONVERGE tools.

Table 1 - Target audiences for the CONVERGE tools.

Beneficiaries	Main objective	Method/Material
Scientific academic community at universities and research centres.	Promote CONVERGE both to raise awareness and also to attract users for the test platforms. The education dimension will also be discussed.	Flyers, workshops and questionnaires. National and European dedicated workshops will be considered.
Industry in the domains covered by CONVERGE.	To promote the use of CONVERGE to industrial user communities. To onboard the industrial demand on the platforms.	Questionnaires and liaison to identify and raise awareness of the industrial community
Funding agencies and related stakeholders.	Disseminate CONVERGE platform as a solution to be supported and used in national and European calls. Support the SLICES-RI development and contribute to the DIGITAL WG of ESFRI.	Engaging with stakeholders at the national and European level. Presentation of the access policy and costs models. Dialogue with ESFRI-RI.
National authorities (Government, Ministries, dedicated agencies).	Raise awareness about the strategic value and potential impact of CONVERGE, including industrial competitiveness, societal challenges and capacity building.	High-level materials such as mission statement, slide-deck, brochure for policymakers.
National and EU regulators as Policymakers.	To explore how CONVERGE could be used from different user communities. Issues regarding access to FAIR research data and GDPR will be considered. Service to society will also be addressed when the targeted verticals should have an impact.	Specific report on the technical and regulatory barriers that limits the use of this particular experimental facilities for research in Europe. Questions related to societal impact.

Members of 6G IA (SNS).	Networking in international initiatives related to testbeds supporting the evolution towards 6G.	Posters, presentations, contributions to white papers, etc.
European initiatives supporting research in wireless, vision and their vertical sectors etc.	Promote CONVERGE as the way of providing validation at scale on the different programs.	Promotion of CONVERGE in the different groups related to the domains covered by CONVERGE.
Standardisation organisations at the global level.	Raise awareness about CONVERGE and establish links with relevant international organisations that support research work addressed by CONVERGE.	Specific report on the technical and regulatory barriers that limits the deployment of SLICES-RI. Invitation to assist to the different project results meetings and workshops.
Non-European agencies or institutions, in particular with the USA.	To promote cooperation and to ensure alignment with other initiatives.	Invitation to workshops, direct communications. Joint activities.

To be more specific, we identify in Table 2 the main groups that matches the CONVERGE tools.

Table 2 - Targeted communities per CONVERGE tool.

CONVERGE tools	Community	Support
Tool 1: Vision-aided Large Intelligent Surface. and Tool 2: Vision-aided base station.	Signal processing, wireless communications, computer vision, vehicular technology (radar, lidar), RF and electromagnetics, data science.	IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) IEEE International Conference on Communications (ICC) IEEE Global Communications Conference (GLOBECOM) IEEE Vehicular Technology Conference (VTC) IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) European Signal Processing Conference (EUSIPCO) IEEE International Conference on Image Processing (ICIP) IEEE Transactions on Signal Processing IEEE Transactions on Wireless Communications IEEE Transactions on Machine Learning in Communications and Networking, IEEE Communications Magazine, IEEE Transactions on Multimedia,

		IEEE Transactions on Signal and Information Processing over Networks, IEEE Open Journal of Signal Processing, IEEE Journal of Selected Topics in Signal Processing, IEEE Signal Processing Magazine, IEEE Trans. on Computational Imaging, IEEE Trans. on Big Data, IEEE Sensors Journal.
Tool 3: Vision-radio simulator and 3D environment modeller.	Computer vision, ubiquitous computing, IoT, multimodal sensing, radio propagation, VR, 3D modelling.	International Conference on Computer Vision (ICCV) IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR) International Conference on 3D Vision (3DVconf) European Conference on Antennas and Propagation (EuCAP) IEEE Transactions on Antennas and propagation, IEEE Transactions on Pattern analysis and Machine Intelligence, IEEE Conference on Virtual Reality European Conference on Computer Vision (ECCV) ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)
Tool 4: Machine Learning (ML) algorithms.	CV, AI, wireless communications	International Conference on Computer Vision (ICCV) IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR) European Conference on Computer Vision (ECCV) International Joint Conference on Artificial Intelligence (IJCAI) Journal of Machine Learning Research (JMLR) ACM Transactions on Knowledge Discovery from Data (TKDD) Neural Computing and applications, Journal of web semantics

CONVERGE will work towards developing the tools and activities to promote its services and solutions in the most effective way towards building a strong community. This objective will be tightly associated to the promotion and dissemination of CONVERGE. Strong links will be developed with SLICES-RI in order to properly articulate both activities and gain a mutual benefit.

4. CONVERGE VERTICAL USE CASES

The CONVERGE toolset and associated research infrastructures will be used by the researchers and industry to achieve new technological solutions, scientific advancements, and discoveries, with significant innovative potential for industrial exploitation and benefits for the society in a number of use-cases and vertical markets relevant for Europe including telecommunications, automotive, manufacturing, media and health.

This will enable new classes of applications having an enormous business potential in the verticals of Telecommunications (enabling the control of the electromagnetic response of the environment and allowing for ubiquitous communications and high accuracy localization and high-resolution sensing services), Automotive (improved perception of vehicle's surroundings and external conditions that may affect the driving quality), Manufacturing (improved understand of the factory floor), Media (lower computational complexity imaging of objects in cluttered environments), and Health (improved fall-detection and noncontact detection of vital medical signature, automatic assessment of subject posture and prosthetic device alignment in physical rehabilitation).

4.1 Telecommunications vertical market

Following the trend initiated in 5G systems, 6G will continue to develop towards even higher frequency ranges, wider bandwidths, and massive antenna arrays. In turn, this will enable sensing solutions with very fine range, Doppler and angular resolutions, as well as localization to cm-level degree of accuracy. Moreover, new materials, device types, and reconfigurable surfaces will allow network operators to reshape and control the electromagnetic response of the environment. At the same time, machine learning, machine vision and artificial intelligence will leverage the unprecedented availability of data and computing resources to tackle the biggest and hardest problems in wireless communication systems. 6G systems will be intelligent wireless systems that will not only provide ubiquitous communications but also empower high accuracy localization and high-resolution sensing services. They will become the catalyst for this revolution by bringing about a unique new set of features and service capabilities, where localization and sensing will coexist with communication, continuously sharing the available resources in time, frequency, and space. Therefore, the CONVERGE toolset will provide telecommunication network operators, vendors, and other stakeholders an opportunity to develop new service types and applications that benefit from the capability of imaging, sensing and high accuracy localization including real-time digital twins, network optimisation tools and capacity provisioning to hot spots with moving cells. In 6G, intelligent context-aware networks will be capable of exploiting localization and sensing information to optimise deployment, operation, and energy usage with no or limited human intervention. Aspects which can be studied by the CONVERGE toolset include the following: 1) obstacle avoidance to maintain radio line-of-sight; 2) massive MIMO and beam-forming optimization in a mobile setting; 3) proactive handover/routing for link obstruction avoidance, based on mobile obstacles trajectory prediction; 4) usage of high-quality datasets related to high resolution synchronised traces of video, RF environment and mobility patterns, which can be used to train ML models; 5) usage of distributed antennas to improve network robustness and coverage, considering the cooperation of multiple base stations, large intelligent surface, mobile communications relays, integrated access and backhaul (IAB) for cellular applications and new vehicular and non-terrestrial network scenarios.

4.1.1 UC 1.1: Coverage Enhancement for Wireless Communications

Engineering electromagnetic wavefronts to control the propagation environment can play a crucial role for wireless networks. As an enabling technology, beamforming is of significant importance to establish a wireless communication link between the end user and the base station(s). In this scenario, shadow regions are inevitable due to the high probability of the direct signal path between the base station and the end user being obscured by various obstacles in the environment. Especially, with the increasing demands for high data rates (and, thus, for higher operating frequency bands), the reliance of future wireless communication networks on various electromagnetic wavefront manipulation techniques is growing at unprecedented rates.

A promising method to deal with this challenge can be to exploit the advanced beamforming concepts and integrate them into existing base station technologies. Beamforming can especially be useful when used in conjunction with Large Intelligent Surfaces (LISs). Within this framework, the LIS architecture can operate as a programable structure and adapt the electric and magnetic properties of its surfaces to control the signal propagation within the wireless network. This makes them work in a different way compared to other multi-element transmission technologies, e.g., phased-arrays and multi-antenna relays. A significant advantage of the LIS technology lies in its capability of achieving all-electronic beamforming, illuminating shadow regions without any additional phase shifting equipment, significantly reducing transmission overheads. Another attractive feature of the LIS architecture may be the ease with which their placement and orientation can be optimized based on the targeted coverage requirements.

Leveraging these significant technological advantages, the LIS can be used to manipulate the electromagnetic waves and provide coverage in blind spots or regions where no RF Line of Sight (LoS) link is present, i.e., Non Line of Sight (NLoS) scenarios, as shown in the example provided in Figure 10. In particular, the LIS can do all-electronic beam-scanning through obstacles to ensure connectivity between the base station and the end user within the controlled environment. A vision-aided LIS architecture can leverage a hybrid-approach, making use of an optical camera to identify the location of the receiver within the controlled environment and track it in real-time. The optical data can then be fed to the RF chain to relay the direction-of-arrival (DoA) information to the LIS. Once the LIS has the DoA information, the LIS aperture can be modulated to reflect the incoming waves in the direction of the receiver, potentially in real-time. The optical data can also be used to identify the location of the transmitter. This can be particularly useful in scenarios where the transmitter location changes over time and an accurate modelling of the incoming waves upon the LIS is needed to manipulate the EM environment.

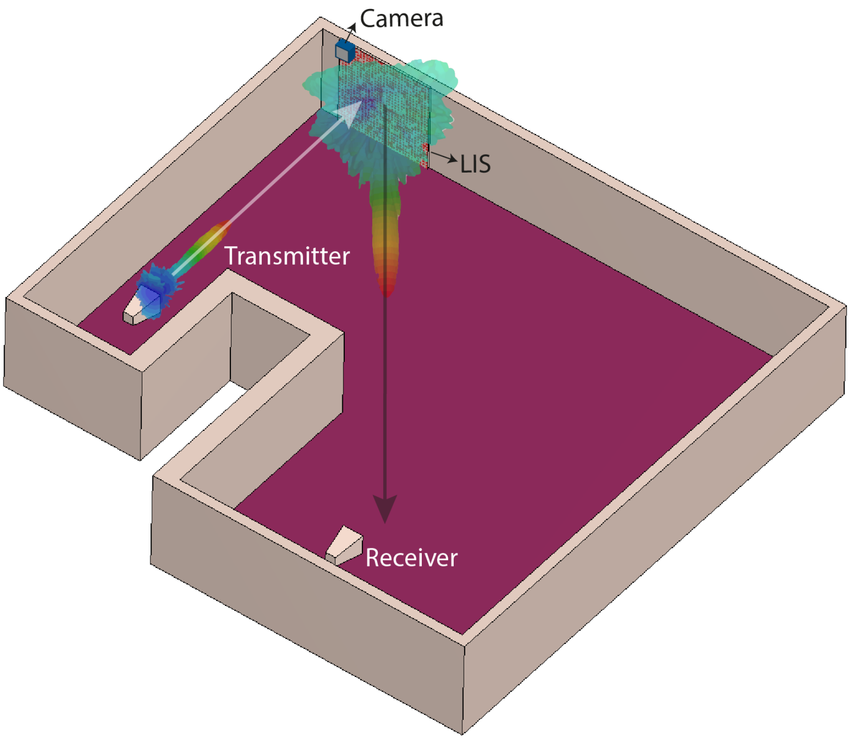


Figure 10 - UC1.1: An example scenario using LIS for coverage enhancement in NLoS environment.

Data involved in the use case:

- Direction-of-Arrival (DoA) data for sub-6 GHz and millimetre-wave sources (end users) to identify beamforming requirements.
- Beamforming, electric field and scattering-parameter (S-parameter) data to verify the illumination of blind spots and dynamic beam scanning through the LIS.

Tools involved in the use case:

- CONVERGE Tool 1 (Vision-aided Large Intelligent Surface) is the main tool involved in this use case. The LIS (at both sub-6 GHz and mm-Wave bands) will play a crucial role in providing coverage enhancement for blind spots within the controlled environment.
- CONVERGE Tool 2 (Vision-aided Base Station) will also be used to illuminate Tool 1, and provide communications endpoints (i.e., the Base Station itself and the UEs) whenever needed.
- CONVERGE Tool 3 (Vision-radio simulator and 3D environment modeler): A numerical model of the LIS will be embedded within the numerical model of the controlled environment to study and optimize the LIS performance.

4.1.2 UC 1.2: Blockage Prediction and Proactive Handover

In 5G evolution, mmWave signals will play an important role thanks to their abundant bandwidth. Because of the strong propagation losses, these networks mainly rely on Line-of-Sight (LOS) links. However, these signals are highly directional and vulnerable to blockages, such as moving pedestrians, vehicles, and so forth. Therefore, rapid link blockages highly compromise the reliability of the 5G network. In fact, when the LOS link is blocked, the network typically needs to hand off the user to another LOS gNB, which may add latency, especially if a search over a large codebook of narrow beams is needed (predefined beam directions). Hence, predicting the occurrence of blocked and non-blocked channels, that is LoS and NLoS, is crucial in ensuring reliable connectivity.

Traditionally, handover strategy is formed based on current RF conditions (e.g., channel state or received power); however, RF conditions are not necessarily informative to forecast sudden transitions between LoS and NLoS conditions. Predicting such events using past RF signals is extremely challenging while consuming spectral resources.

To obviate this problem, visual data such as RGB-D images and 3D point clouds from LiDAR sensors that capture a variety of hidden features in wireless environments (e.g., object locations and mobility patterns) can be exploited. In so doing, one can accurately predict future mmWave channel conditions without consuming RF resources to probe and estimate the radio propagation channels. Based on cameras or LiDAR sensors deployed at the gNBs a machine-learning algorithm can predict signal blockage and trigger a vision-aided proactive handover to another gNB. This requires the development of machine learning models that learn to predict link blockages proactively, i.e., transitions from LoS to NLoS. Figure 11 shows an overview of this use case, exemplifying the prediction of the blockage event and its impact on the observed bitrate.

The major difficulty in dealing with LiDAR data is that the sensor produces unstructured data in a point cloud containing typically around 100 thousand 3D points per 360-degree sweep, generating a throughput of 100 Mbit/s (example for a LiDAR with 16 beams). With the large amount of data coming from multiple LiDARs, vision-added 5G infrastructures would require tremendous computation capability. This poses a large computational challenge for real-time detectors and is not scalable. Processing the data at the network edge can save a significant amount of communication bandwidth and satisfy the low latency requirements. This further enables Artificial intelligence (AI) applications, which train and deploy powerful machine learning models at these edge nodes.

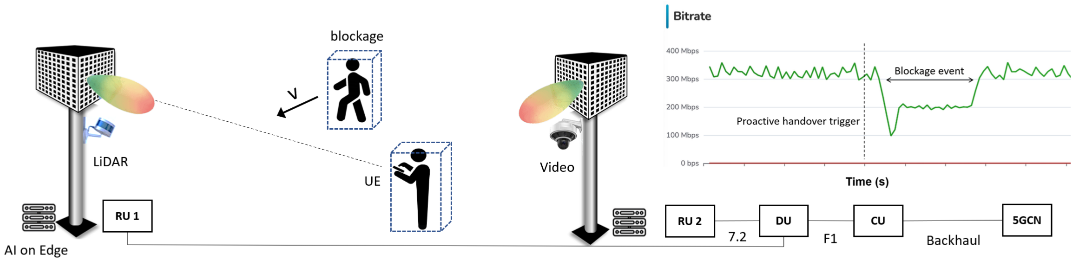


Figure 11 - UC1.2: Blockage prediction and proactive handover high-level scenario.

<u>Data involved in the use case:</u>

- 3D LiDAR point clouds from multiple LiDARs co-located with the gNBs. LiDAR data is commonly delivered in two standard file format types: .LAS and .LAZ.
- Video data sources from cameras (e.g., security cameras).
- Position and movement direction of obstacles (velocity, displacement angle).
- 5G physical layer parameters and handover duration.
- QoS and QoE metrics to evaluate the impact of proactive handover on end-to-end connectivity.

<u>CONVERGE Tools involved in the use case:</u>

- CONVERGE Tool 2 (Vision-aided Base Station) can be used together with the vision-radio experimental chamber to evaluate the proposed solution for both blockage prediction and proactive handover in an indoor scenario.
- CONVERGE Tool 4 (Machine Learning algorithms for dataset analysis) can be used in the context of this UC for training and fine-tuning different ML techniques to detect and tracking of objects (e.g., VoxelNet, PointNet, Complex Yolo, PointPillars) and predict blockage events. The ML algorithms will run in Edge devices close to the gNB (e.g., NVIDIA Jetson Nano, Jetson Xavier).

4.1.3 UC 1.3: RF Mapping for Imaging and Coverage Extension

Radar imaging has become a key technology in multiple application fields due to its ability to retrieve RF images or maps of the inspected areas. This is particularly useful in complex scenarios, such as in NLoS conditions, thanks to the capability of electromagnetic waves to penetrate through certain optically opaque materials. Furthermore, another important advantage of radar systems is their ability to operate regardless of weather conditions.

LIS can be leveraged to improve the capabilities and/or performance of radar imaging and communication systems. In particular, a LIS might be used to enhance the coverage in NLoS scenarios, improving the propagation channel conditions. This would in turn contribute to increasing the signal to noise ratio of the radar signals backscattered by the targets or, alternatively, the communication signals arriving at an end-user device or at a base station. In addition, as the reflection and refraction characteristics of the LIS can be dynamically manipulated, the direction towards which the LIS is providing enhanced coverage might be also dynamically and instantaneously modified. Thus, the LIS-assisted radar (or communication system) might be able to adapt itself to changing environments and conditions.

The LIS can be also designed to ensure joint communications and sensing through appropriate wave control. For instance, the states of some of the unit cells can be configured appropriately for the sensing while others can be used for communications (including reflections). The design of the LIS for RF imaging is also dependent on the level of sensing precision required.

RF images retrieved by the LIS-assisted radar might also be combined with optical images and LiDAR maps in order to build RF-optical images. These sensors provide complementary information and, as a result, radar and vision fusion might provide enhanced capabilities, facilitating the detection of targets of interest and obstacles in challenging scenarios, even in NLoS conditions.

As aforementioned, radar systems can be used to "see" through certain optically opaque materials, especially at lower frequencies. This can be used, for instance, to detect targets hidden behind a wall, such as the example provided in Figure 12. In this regard, the LIS-assisted radar might be used to improve the channel conditions and the radar main beam might be dynamically redirected towards areas of interest where potential targets are located.

Figure 12 - UC1.3: An example scenario using a LIS-assisted radar for RF mapping of targets hidden behind a wall.

As part of this use case, additional value can be brough by the employment of a technique called image inpainting, which consists in modifying an image in an undetectable form [Ber2000], [Gui2013]. The goals and applications of inpainting are numerous, from the restoration of damaged paintings and photographs to the removal/replacement of selected objects. Image inpainting has been widely exploited in the field of computer vision and image processing. The main purpose of image inpainting is to produce visually plausible structure and texture for the missing regions of damaged images. In the past decade, the success of deep learning has brought new opportunities to many vision tasks, which promoted the development of a large number of deep learning-based image inpainting methods [Xia2023]. This use case may take advantage of radio-aided image inpainting and image-aided radio mapping. This will require converting image between various formats. In classical imaging, this may involve images from RGB cameras, infrared cameras, plenoptic, etc. [Iso2017], [Mal2019]. However here this needs to be extended to conversion between images and radio (radar) maps. The fusion of images of different formats may also lead to images with multiple components, such as e.g., RGB-D (where D stands for Depth)[2]. With Time-of-Flight measurements (radar), one can obtain depth information.

Data involved in the use case:

- Sets of raw multifrequency radar measurements (conventional and LIS-assisted radar).
- RF images created from the radar data.
- RF-optical images obtained from sensor fusion (radar, LiDAR and optical sources).
- Estimation of the target's position.

Tools involved in the use case:

- CONVERGE Tool 1 (Vision-aided Large Intelligent Surface) is the main tool involved in this use case. The LIS (at both sub-6 GHz and mm-Wave bands) will play a crucial role in the retrieval of RF images or maps, considering the proposed LIS-assisted radar technology. Besides, the fusion of the RF images with LiDAR/optical data might provide complementary information.
- CONVERGE Tool 2 (Vision-aided Base Station) is another important tool, providing a BS antenna array and cameras.
- CONVERGE Tool 4 (Machine Learning algorithms for dataset analysis) could be used to automatically detect targets in the RF images and/or in the RF-optical images. It might also be used to decide "on-the-fly" how to modify the behaviour of the LIS to enhance the detection capabilities of the LIS-assisted radar.

4.1.4 UC 1.4: Transparent Sensing for Enablement of Enhanced Radio Resource Management

Mobile network terminals are equipped with a diverse set of sensing equipment, built to measure and capture sensing data pertaining to different physical characteristics. This data can be used to map the physical environment, and the CONVERGE project aims at addressing and providing tools and enablers to realize this vision. However, when considering sensing capabilities for the above-mentioned purpose, user equipment can play a very important role in an end-to-end communication system setup. The fact they are equipped with sensing technologies for different physical properties, the coming enablement of RF sensing techniques in future mobile networks, their portability, and of course, the fact they may directly serve human users like in the case of smartphones, are some of the reasons why investigation is required on how to leverage the rich information they can provide to enable terminal oriented support for communication systems.

We use the terminology of 3GPP-based sensing for RF sensing signals and related measurements, and non-3GPP sensing for data pertaining to sensors such as cameras, LiDAR, etc.

By performing measurements such as time-difference-of-arrival (TDoA), angle-of-arrival (AoA), angle-of-departure (AoD) measurements, Received Signal Strength indicator (RSSI), etc., different conclusions can

[2] https://www.e-consystems.com/blog/camera/technology/what-are-rgbd-cameras-why-rgbd-cameras-are-preferred-in-some-embedded-vision-applications/

be drawn from the physical environment.

Figure 13 depicts an example of N terminals operating in bi-static operation, where sensing signals are emitted from an Access Network node, and the performed measurements are reported back to the same network.

By analysing a video stream or LIDAR reflection data, more information can be extracted such as the position of an object (absolute or relative), corresponding distances, speed, rotation information, etc.

This kind of data can be extremely useful in support for many applications, them being related to Radio Resource Management (RRM) or others.

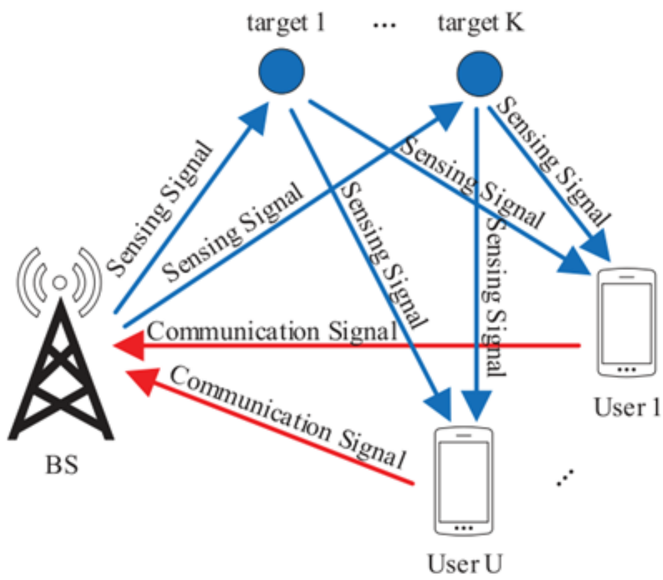


Figure 13 - UC1.4: Sensing in mobile communication systems - User terminals reporting on sensed signals to an Access Network [3GPP2023].

It is, however, a particularly difficult challenge to handle sensing data from a terminal's perspective. First, the integrity of the data needs to be addressed and ideally, guaranteed. This means that some form or measurement of the trustworthiness of this data needs to be in place to guarantee its accuracy. Secondly, the validity of the sensing data needs to be addressed. The usage of sensing data for purposes of RRM, under the 'sensing to communicate' moto, will need to consider validity as a fundamental principle. This is because there are many RRM procedures that can be enabled or enhanced with this data, but the challenge relies on the latency budget for its usage. Particularly in the case of terminal generated data, if the RRM mechanisms proposed for improvement are controlled by an Access Network, this network would require access to this data for monitoring and decision-making purposes. Due to the latency that a sensing data collection process would incur into, added together with processing of the integrity of that data, any decision-making process and finally, a decision instruction towards a user equipment, we conclude that certain RRM features would be excluded from possible sensing enabled enhancements, I.e., its validity period may be too short.

The transparent sensing for RRM enablement targets the study of enhanced functionalities that are leveraged by sensing information collected by network nodes, with a particular focus on user equipment. A three-tier RRM functionality enhancement leveraging on sensing data framework will be considered, where the proposed enhancements will leverage on data located on end user equipment, radio access nodes and on a server-based policy maker supervisor for access nodes, such as a Core Network. The three-tier approach is well aligned with the CONVERGE architecture, where sensing capabilities are considered at both CONVERGE's chamber and Core.

Data involved in the use case:

- Direction-of-Arrival (DoA), Direction of departure (DoA), Time difference of arrival (TDoA), positioning measurements (e.g., GGNS, x,y,z coordinates), speed, orientation, rotation data.

Tools involved in the use case:

- CONVERGE Tool 1 (Vision-aided Large Intelligent Surface) can be used as a sensing data generator.
- CONVERGE Tool 2 (Vision aided mobile base station) can be used as a sensing data generator.
- CONVERGE Tool 3 (Vision-radio simulator and 3D environment modeller) can be used to model the environment and help in the filtering of data when considering its integrity.
- CONVERGE Tool 4 (Machine Learning algorithms for dataset analysis) will be used for the evaluation of data integrity and validity.

4.2 Automotive vertical market

The CONVERGE toolset and associated research infrastructures can bring value for the automotive case. It will enable addressing current open challenges such as the urge to improve the sensorial capacities of autonomous vehicles whether by monitoring in-vehicle or its surroundings. As the intelligent mobility solutions evolve, there is a need for automated driving solutions as well as for robust ways to monitor its drivers and occupants. The proposed infrastructure will aid the development of precise, robust and safe vision of vehicles addressing the demand of integral perception of vehicle's surroundings and external conditions that may affect the driving quality (e.g., meteorological conditions) to ensure a better and more advanced perception, compared to that of the human being. Technologies such as mm-wave and vision aided sensing have the potential to provide continuous, real-time information via dynamic, contactless measurements for future vertical businesses. 6G simultaneous localization and mapping (SLAM) methods will not only enable advanced cross reality (XR) applications but also enhance the navigation of autonomous objects including vehicles and drones. In convergent 6G radar and communication systems, both passive and active radars as well as LiDARs and cameras will simultaneously use and share information, to provide a rich and accurate virtual image of the environment. In CONVERGE, vehicles may be, for instance, emulated by mobile vision-aided user equipment interacting with multiple vision-aided base stations, and a large intelligent surface equipped with RF sensing capabilities. The vision-signal propagation simulator interacting with vision-captured environments and enabling the simulations up to the mm-wave range will also be of great value. The solutions developed can also be additionally validated in the outdoor testbeds available in the CONVERGE research infrastructures.

4.2.1 UC 2.1: In-vehicle Sensing

A lot of work has been done in the field of autonomous driving. Vehicles equipped with several sensors that enable detecting obstacles, other vehicles and driving conditions have been tested to enable delivering self-driving cars. To have a better knowledge of these surroundings and improved navigational capability vehicles can be also connected with a satellite and can communicate and exchange information with other vehicles.

Although the research will continue to make autonomous vehicles more intelligent having enough capability to drive in any type of surrounding environment, sensing the interior of the vehicle is crucial for creating safe conditions in several situations. In-vehicle-sensing will contribute to increase the security of its occupants, for example, in situations of car sharing in a scenario of autonomous vehicles. It has also an important role in driver monitoring, for example, for fatigue identification.

The CONVERGE mobile base station, located in a car, train or bus (see Figure 14), can contribute to enhance driving assistant systems by providing communications to the vehicle users, enabling communicating with a central control station or other vehicles and send alarms in situations of danger. Simultaneously, using its RF sensing and vision capabilities, it can capture the vehicle interior including the attitudes of driver and passengers. Additionally, the mobile base station also sees and senses the vehicle exterior and communicates with other similar vehicles to exchange its views and sensed information.

Information collected by cameras will enable activity identification, emotion monitoring, identification of undesired behaviour and other supervision tasks. RF sensing will enable complementing the vision data and help overcoming problems related, for example, to the fast-changing conditions of illumination and background and occlusion. The 3D environment modeller also plays an important role for enabling virtualising the environment and setup several scenarios.

In large public vehicles, like buses and trains, multi-object multi-camera tracking will benefit from using RF sensors to accurately detecting and tracking a person across multiple non-overlapping camera views in a robust and efficient manner. This can rely on regular communication hardware available on site. The use of Large Intelligent Surfaces (LISs) to sense the environment contributes also to better sense movements and track persons enabling high-precision 3D positioning.

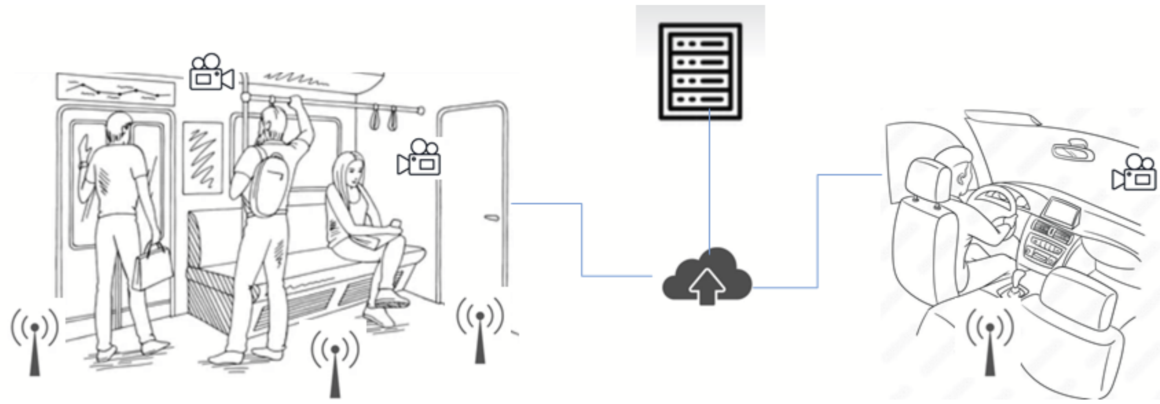


Figure 14 - UC 2.1: In-vehicle sensing and reasoning scenario.

<u>Data involved in the use case:</u>

- Raw data coming from RF sensors.
- Sensing data coming from LISs.
- Raw images captured by cameras.

<u>Tools involved in the use case:</u>

- CONVERGE Tool 1 (Vision-aided Large Intelligent Surface) complements the data coming from the vision systems and other RF sensors obtain accurate people and environment information.
- CONVERGE Tool 2 (Vision aided mobile base station) can be used as a vision and RF sensor for sensing the environment and communicating with the outside. This tool can also be used to illuminate Tool1, leveraging its potential.
- CONVERGE Tool 3 (Vision-radio simulator and 3D environment modeller) can be used to model the inside environment and contributing to better isolate passengers from the surrounding

environment. This can contribute to reduce the impact of identified challenges created by outside changing conditions.

- CONVERGE Tool 4 (Machine Learning algorithms for dataset analysis) can be used in the context of this UC for training and fine-tuning different ML techniques for action and emotion recognition and anomaly detection (e.g., abandoned objects, violence, etc.). Coordinated sensing and data fusion is required to complete information coming from the different sources. The ML algorithms will run mostly in the edge, although cloud-based communication and computing is also foreseen for collecting data to complement existing datasets and to re-train models with this additional.

4.2.2 UC 2.2: Collaborative Modelling of the Environment

Modern vehicles are equipped with several sensors used to reconstruct knowledge about their surroundings to simplify operations (e.g., rain and light sensors), automate movements and prevent collisions (e.g., radars, LiDARs, satellite geo-positioning, stereoscopic cameras). In addition, modern vehicles are equipped with communication systems to allow tele maintenance, emergency calls, but also exchange of information with other vehicles in the surrounding or the infrastructure itself.

Obtaining an accurate knowledge of the environment is essential for autonomous driving. Even though critical safety actions will always only rely on pure local information, exchange of information between the vehicles and the infrastructure is of great help to optimise safety, comfort, traffic flows, and energy consumption. In dense areas, optimising communication is essential, and we must answer the following two questions to be as efficient as possible (for whatever definition of efficiency): 1) on the one hand, we must determine what information to send, when and to who. For example, depending on their accuracy, LiDAR may generate hundreds of megabytes of traffic per second, bandwidth that is not sustainable in practice; 2) on the other hand, the communication schemes must be selected wisely to guarantee that the spectrum is used correctly. For example, modulation can depend on the density of emitters and their speed.

To build a temporal 3D environment from vehicles moving around in a smart city (as depicted in Figure 15) is a quite challenging task as it requires both accurate and timely information. A first aspect to consider is the need for synchronization to build an accurate model of the environment. We shall consider two types of synchronization: spatial and temporal. The spatial synchronization is needed as different observers (e.g., different vehicles) may have a different perspective on the same event (e.g., a moving obstacle) or even have a partial or no vision of the event because of different type of sensors that are used. The temporal synchronization is needed as the information about a particular event may not always be transmitted at the same moment because of issues with the transmission medium for example. In addition, the transmitted data coming from different types of sensors is expected to have partial and contradictory information to be combined to build accurate model of the environment.

Figure 15 - UC2.2: Collaborative modelling of the environment in autonomous driving scenarios.

Data involved in the use case:

- 3D LiDAR point clouds from multiple LiDARs co-located with the vehicles and fixed location (e.g., gNBs).
- Video data sources from cameras embedded in vehicles and facing outside.
- Video data sources from cameras located in fixed locations (e.g., gNBs) and facing the infrastructure.
- Position and movement direction of vehicles and environment (location, velocity, displacement angle).
- Electromagnetic field map of the environment, covering the different spectrums used for communications.

CONVERGE Tools involved in the use case:

- CONVERGE Tool 2 (Vision-aided base station) can be used to optimise communication schemes to enforce quality of service and spectrum efficiency.
- CONVERGE Tool 3 (Vision-radio simulator and 3D environment modeller) can be used to study potential scenarios (e.g., accidents, road closed) to determine the best actions and location of sensing and communication relays.
- CONVERGE Tool 4 (Machine Learning algorithms for dataset analysis) can digest data from coordinated sensing and data fusion to generate models of the environment for synchronisation and predictions of environment evolutions, such as the location with time of the different emitters, receivers, and obstacles. More precisely, it is expected to use the tool to construct the mobility model of the elements in the environments (e.g., cars, pedestrians, etc.) according to the way information is transmitted in the infrastructure. For example, depending on the timing of a traffic light congestion can happen or not. Or in case a car does not obtain the information in time it may not be rerouted on an adequate path or avoid a collision.

4.3 Manufacturing vertical market

Industry accounts for over 20% of the EU economy, directly employs around 35 million people and contributes to over 80% of exports. Although European industry still has a competitive advantage on many high value-added products and services, it is under a significant transformation towards a greener and more digital industry to ensure its competitiveness and decarbonization. A key factor in this transition is to achieve a drastic reduction of the time between the occurrence of an event and the implementation of an appropriate response. This agility allows companies to adapt to new situations promptly. This means that, for example, unexpected changes in the factory floor, captured, propagated and understood in due time, can be considered during the manufacturing process of a product, because the company guarantees the agility to adapt to the new situation. Being able to understand the factory floor is the key issue, which includes the monitoring and tracking of the mobility of people and objects, production lines, forklifts, and other intelligent robots. For example, [Vac2021] has recently shown with success the potential of using a LIS as a sensing platform to determine whether a robot has deviated from its predefined route. By leveraging LIS technology, companies can capture and propagate unexpected changes on the factory floor in real-time, providing valuable insights into the manufacturing process. This enables manufacturers to make informed decisions, optimize workflows, and ensure smooth operations.

On the other hand, computer vision has been used more extensively for this purpose [Des2020], but the combination of these two approaches is yet to be explored by the research and industrial communities and has the potential for bringing additional benefits to the manufacturing sector in terms of use of resources and energy efficiency.

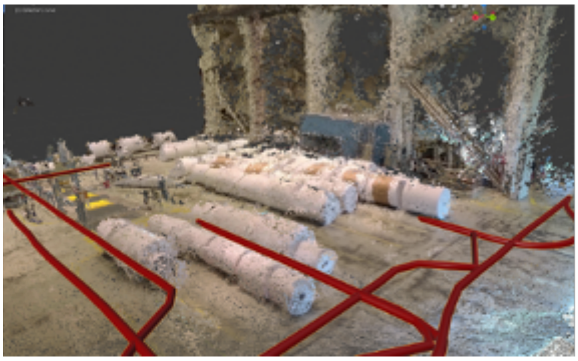
The integration of automated assembly lines with advanced technologies such as LIS, computer vision, privacy-preserving human identification through RF sensing, and digital twins further revolutionizes the manufacturing landscape. Automated assembly lines consist of interconnected robots working in synchronization to produce complex products. With the deployment of private networks leveraging 6G technologies, these assembly lines can achieve seamless communication, real-time data transmission, and ultra-low latency, ensuring efficient coordination between robots and other equipment. By embracing these advanced technologies, manufacturers can achieve greener and more digital operations, enhance resource utilization, and strengthen their competitiveness in an evolving global market.

4.3.1 UC 3.1: Automated Assembly Lines

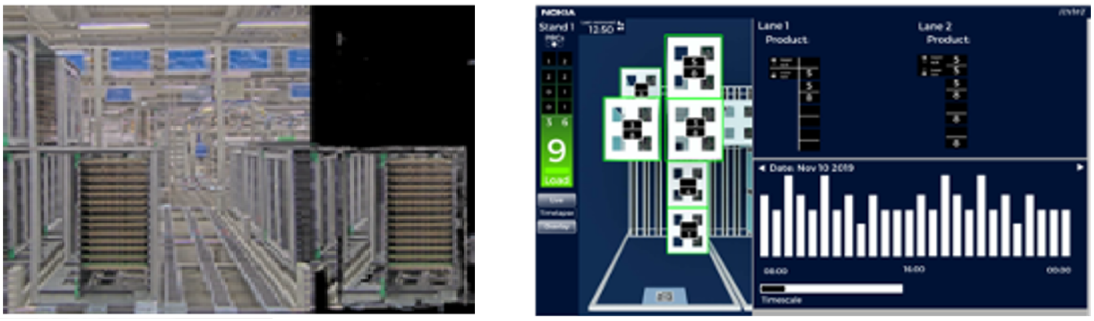
In a factory, robots work together on an automated assembly line to produce complex products. Let us assume a factory consisting of automated production lines with interconnected robots. Produced objects are flowing between synchronized production benches following well defined lines. Synchronization is performed by a central controller. All communications within the factory are wireless (private *G cells). These private networks ensure secure and low-latency connections, enabling real-time data transmission between the various components of the assembly line. The high data rates, ultra-low latency, and reliable connectivity offered by these advanced wireless networks enable real-time monitoring and control of the assembly process. This real-time responsiveness allows for faster decision-making, predictive maintenance, and adaptive manufacturing, ultimately leading to improved product quality, reduced costs, and increased production throughput.

To avoid the limitations of traditional wired connections and enable flexible reconfiguration of production lines, LIS are strategically deployed throughout the facility to guarantee communications. By leveraging LIS technology, the factory can maintain seamless wireless connectivity while accommodating changes in the layout or configuration of the assembly lines.

To optimize the efficiency of the assembly line and the placement of wireless sensors, cameras and antennas, 3D environment modeler and vision-radio simulator are employed. The 3D modeler and simulator allow for finding the best possible configurations and placements in a changing environment. The 3D environment modeler and vision-radio simulator tool can also be used to simulate and model the flow of materials and products through the factory, and to optimize the placement of wireless sensors and antennas to ensure efficient communication between robots and other equipment. In addition, 3D modeler could be used to refine the overall configuration of the assembly line. This approach helps in achieving the most efficient communication between robots, equipment, and other components, leading to enhanced productivity, and minimized downtime. Through wireless connectivity, LIS deployment, and advanced simulation tools, factories can achieve seamless communication, efficient operations, and optimized configurations, thereby maximizing productivity and competitiveness in the rapidly evolving manufacturing industry. Figure 16 shows two examples of such production lines.



a.



b.

Figure 16 - UC3.1: a) Accurate robot and asset tracking in production line using 360° photogrammetry b) Real time 3D inventory tracking in Nokia production line with multiple cameras.

<u>Data involved in this use case:</u>

- Sensing data from the LIS.
- Video streams from the assembly line and factory floor (potentially also from cameras mounted on the robots, if available). Robot location based on the video information from multiple sources.
- Sensing data from RF devices.
- Data from the assembly line production, if available.

<u>CONVERGE Tools involved in the use case:</u>

- CONVERGE Tool 1 (Vision-aided Large Intelligent Surface) will be used for the retrieval of RF data and image/video streaming capture.
- CONVERGE Tool 2 (Vision-aided base station) can be used to optimise communication schemes to enforce QoS and spectrum efficiency in the assembly line.

- CONVERGE Tool 3 (Vision-radio simulator and 3D environment modeller) can be used to study potential scenarios to determine the best actions and location of sensing and communication relays and potentially to optimize the configuration of the assembly line.
- CONVERGE Tool 4 (Machine Learning algorithms for dataset analysis) can digest data from coordinated sensing and data fusion to generate models of dynamically changing production environment and assets. It will also assist on the placement of wireless sensors, cameras, and antennas.

4.3.2 UC 3.2: Digital Twinning

The use of the CONVERGE 3D environment modeler with vision simulator can aid in manufacturing by creating a digital twin of the product and allowing tweaking of various components and testing of various iterations of the product. The product can also be subject to further predictive analysis for fault checking or any other issues it can possess. It can not only then be used by designers and engineers on the field with the product but also remotely.

The vision-radio simulator can serve as a digital-twin model of the controlled environment to study physical system limits, and their impact in sensing and communication related tasks. The simulator can play a key role in understanding the system parameters and the limits imposed by them for the planned use case scenarios. This can be achieved through systematic analyses of the environment and these studies are not feasible to be carried out by means of pure experiments. In particular, the vision-radio simulator can be used to evaluate the impact of considering different LIS aperture sizes in the achievable resolution and the overall system performance, or to assess the effect of the LIS and base station positions within the environment and that of its complexity (e.g., number of obstacles).

Moreover, the vision-radio simulator can act as a digital platform for evaluating different sensor fusion algorithms leveraging radiofrequency data, positioning information, or vision-based inputs to name a few. In particular, the simulator can play a key role in understanding how different processing and fusion schemes perform under exactly the same conditions, enabling their optimization. Furthermore, the simulator may be leveraged to plan actual tests to experimentally assess the algorithms and systems developed by the partners. Figure 17 presents an example of a meshed scenario of a manufacturing environment that could be used for the vision-radio simulator.

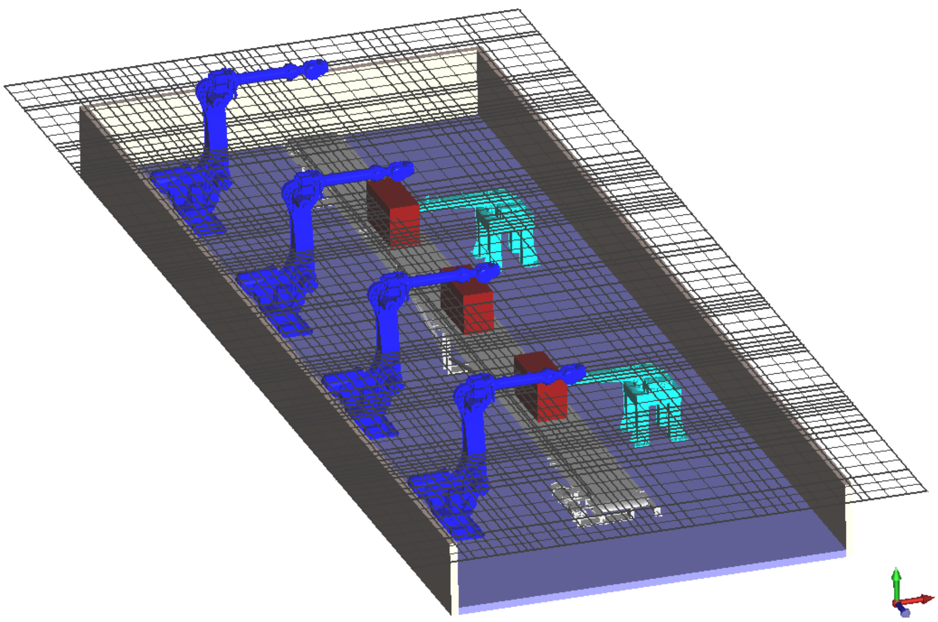


Figure 17 - UC3.2: An example of a meshed scenario for the vision-radio simulator.

In addition, the simulator can also be useful to generate a significant amount of data to train some of the ML algorithms in Tool 4. ML algorithms require a significant number of training data in order to adequately learn from the dataset without overfitting. However, in some cases the acquisition of such amount of data is not feasible or is too time-consuming. To overcome this issue, the vision-radio simulator might be used to generate synthetic data resembling a real scenario.

The vision-radio simulator can potentially play a key role in the manufacturing vertical market as industrial environments usually exhibit a high-degree of complexity. Therefore, the simulator may help to optimize deployments and systems used in this context.

Data involved in the use case:

- LIS radiation patterns as a function of different system parameters, such as BS location, LIS location, etc.
- RF images created from simulated radar data.
- Large set of RF images and raw channel data for ML and learning-related tasks.

CONVERGE tools involved in the use case:

- CONVERGE Tool 3 (Vision-radio simulator and 3D environment modeler): A numerical model of the LIS will be embedded within the numerical model of the controlled environment to optimize the LIS performance, study & optimize the BS location, consider sensor fusion and how the LIS can benefit from it and investigate coverage enhancement and RF mapping scenarios.
- CONVERGE Tool 4 (Machine Learning algorithms for dataset analysis): The digital model of the controlled environment will be used to generate large sets of training data that can be used for several purposes, including (but not limited to): learning from raw channel data to achieve detection/classification of objects without doing an image reconstruction, improving the resolution in the reconstructed images, and minimizing the distortions caused by motion of the BS and RF obstacles.

4.3.3 UC 3.3: Privacy-preserving Human Identification through RF Sensing

This use case aims at exploring the possibilities and limitations of performing human identity recognition by mining the unique patterns of individuals hidden in wireless sensing signals. The understanding of the factory floor includes the monitoring and tracking of the mobility of people. These actions are crucial to the manufacturing activities as they allow for instance the access control to facilities and goods without harming the preservation of privacy of the individuals.

In existing human identification systems, the most commonly used sensors are cameras and radars. On one hand, the static features hidden in biometric traits such as fingerprint, iris, or face are extracted from images and can be used for identity recognition, on the other hand, dynamic characteristics of the human body, such as gait, are captured by radars or videos and provide unique features for identification. The use of radars in real world applications of human identification systems is limited by its high cost. Ordinary cameras are inexpensive and easy to deploy but have a high risk of privacy leakage. In comparison with these abovementioned sensing techniques, there is a clear advantage in using wireless sensing as it does not require special sensor equipment, allow the control of privacy aspects disclosure, is more robust to variations in the environment lightning, and constitutes a valuable support for achieving ubiquitous sensing.

The general problem of uniquely identifying an individual from a large user population in any physical setting is arguably very challenging. In this scenario, we are considering a setting where the goal is to uniquely identify a person from a group of N people. For instance, a simple yet common scenario for an office setting, wherein a single person (from a group of N people) walks through a room or through a corridor towards a door and the goal is to uniquely identify this person so as to either permit or deny access.

The traditional camera-based approaches have been consistently deployed and achieve good accuracy in identifying humans, however, they may be considered to be too intrusive and not privacy respecting for use in offices and homes. Motivated by several guidelines and regulations that enforce methods that comprise privacy-by-design characteristics, recently many works propose to use wireless radio for human body sensing and identity recognition. Diverse approaches can be found in the literature that actively try to promote wireless sensing technologies for human identification. Undoubtedly, RF is a must needed component for us in our day-to-day life and present in most of the spaces, from our home to offices. The RF devices all have in common the fact they fill the air with radio frequency signals. Thus, the identification of humans is possible to perform with the help of RF signals because when a human walks in between those RF signals they affect the propagation of those signals in a unique manner, since every human has a unique walking style. One possible approach is to examine the RF signals from existing telecommunications hardware (e.g., 5G) using CSI (Channel State Information) leading to the identification of every human uniquely. The image captured by optical cameras will be use as ground-truth to assess the performance of the developed methods. Figure 18 shows an example scenario for the privacy preserving human identification through RF sensing.

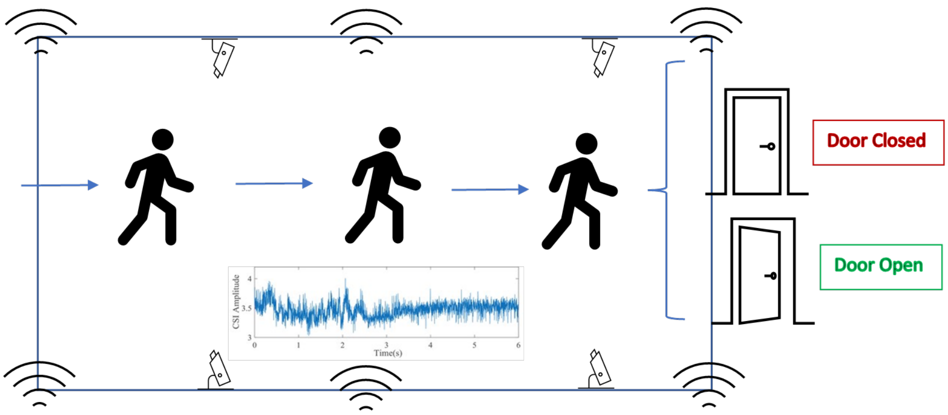


Figure 18 - UC3.4: An example of a scenario for the privacy preserving human identification through RF sensing.

<u>Data types involved in the use case:</u>

- Raw data coming from RF sensing functionality provide by the communication systems (e.g., 5G Sounding Reference Signals for the uplink, and Reference Signals for the downlink).
- Raw images/Video streams captured by cameras.

<u>CONVERGE Tools involved in the use case:</u>

- CONVERGE Tool 1 (Vision-aided Large Intelligent Surface) will be used for the retrieval of RF data and video streaming capture.

- CONVERGE Tool 2 (Vision-aided base station) can be used to generate the radio signals that illuminate the LIS.
- CONVERGE Tool 4 (Machine Learning algorithms for dataset analysis) can be used in the context of this UC for training and fine-tuning different ML techniques to achieve human identification (CNN and LSTM combined with attention mechanism, for instance).

4.4 Media vertical market

Leveraging sensor fusion technology with multi-mode imaging capabilities can offer significant advantages in imaging of objects in cluttered environments. The developed toolset will enable the scientific community to explore unique sensing scenarios by leveraging microwave holography, intelligent surfaces, and vision cameras.

By taking advantage of a camera, it is possible to gather *a-priori* estimation of the location coordinates of a constrained field of view (FoV) that contains the object of interest for imaging in a cluttered environment. Once this information is gathered, the holographic intelligent surface can be activated to perform computational imaging at microwave frequencies over the constrained FoV. Such an approach can substantially reduce the computational complexity of the imaging problem by processing only the constrained FoV for image reconstruction. Because the intelligent surface operates at microwave frequencies, the backscattered data from the imaged object is light-independent. The microwave-based backscattered data from the imaged object remains unaffected by variations in lighting conditions, providing robust imaging capabilities in diverse lighting scenarios.

Beyond telepresence, the integration of intelligent surfaces and vision-aided base stations opens up new opportunities for bandwidth optimization in crowded areas such as festivals or large events. In these scenarios, temporary high-bandwidth networks can be established to support the increased data demand from attendees. Vision-aided base stations, deployed through Unmanned Aerial Vehicles (UAVs) or other means, act as temporary network nodes, delivering a high-bandwidth network to the crowded areas.

The integration of telepresence capabilities through intelligent surfaces, coupled with human and object tracking in indoor settings and vision-aided base stations for bandwidth optimization, transforms the media landscape. It enables remote collaboration, immersive teleconferencing, and interactive experiences that bridge physical distances. Additionally, in crowded event scenarios, this technology ensures reliable connectivity and enhances the overall attendee experience.

These advancements hold significant potential in various media-related applications, such as virtual events, remote education, virtual meetings, and interactive entertainment experiences. By harnessing the power of intelligent surfaces, vision-aided base stations, and robust tracking capabilities, the media industry can create engaging, seamless, and immersive experiences that transcend traditional boundaries and redefine the way we connect, collaborate, and consume media content.

4.4.1 UC 4.1: Outdoor Vision-aided Mobile Base Station

Temporary crowded events, such as fairs, parades, and music festivals, pose unique challenges for providing reliable and high-capacity wireless connectivity. Traditional wireless access infrastructures may struggle to offer adequate coverage and sufficient network capacity when faced with thousands of potential network users in confined spaces for limited periods. This makes it difficult to meet the sporadic high throughput, ultra reliability, and low latency requirements associated with bandwidth-intensive communications, including holographic-type applications, live video-streaming, augmented and virtual reality, and real-time social media sharing.

Moreover, outdoor events typically involve unpredictable crowd movements. For instance, in a music festival, people may move from stage to stage according to the dynamics of the event, resulting in large and variable traffic demand over time. This creates increased challenges for ensuring ubiquitous wireless connectivity. The challenges are exacerbated by the presence of obstacles, including physical structures (e.g., stages and buildings), vehicles, and terrain features, which can compromise the Line of Sight (LoS) and degrade the Quality of Service (QoS) provided to network users. In this context, novel communications solutions are required to address these challenges and ensure a seamless wireless connectivity experience for attendees, performers, and event staff.

On-demand wireless networks, including airborne networks consisting of mobile base stations carried by Unmanned Aerial Vehicles (UAVs), are emerging as a promising solution for providing wireless connectivity in different scenarios (e.g., Figure 19). Airborne networks leverage the UAVs' 3D positioning ability to enhance wireless coverage and network capacity anywhere, anytime. Despite extensive research on UAV placement, existing solutions consider free-space scenarios (obstacle-free) and have not considered the impact of obstacles and the potential LoS blockage they create in non-obstacle-free environments.

Addressing the challenges of providing obstacle-aware, on-demand networks requires considering users' positions, the presence of obstacles potentially affecting LoS, and meeting heterogeneous QoS requirements of different services. A solution that jointly considers these factors has not yet been achieved, motivating the need for ongoing research to advance the state of the art in this field. To support such research, the community can take advantage of the CONVERGE's tools to envision obstacle-aware on-demand wireless networks.

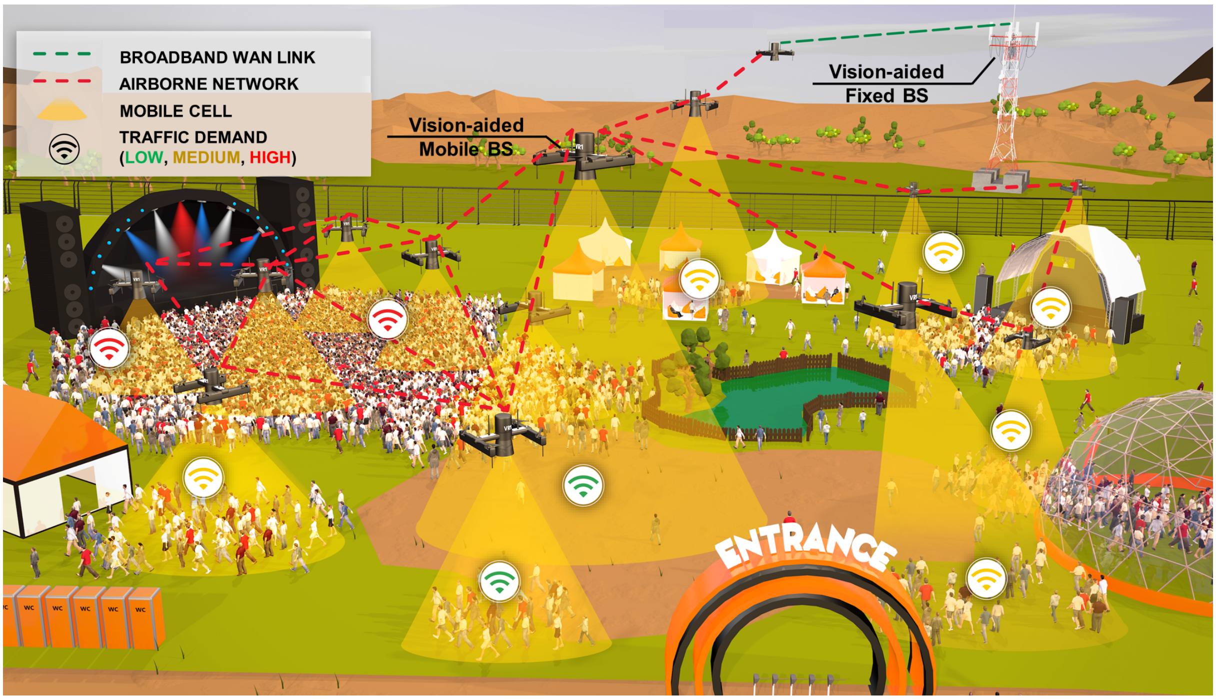


Figure 19 – Example of a crowded event that could benefit from multiple vision-aided mobile base stations.

The exchange of data between the CONVERGE's tools will promote continuous learning from collected information and adaptation to environmental changes. This will ensure the conditions for validating and evaluating novel on-demand wireless networks while considering existing obstacles and users' positions. Furthermore, it will enable large-scale validation and evaluation through simulations carried out under realistic conditions, taking into account a variable number of network users, which can be defined according to typical crowded events. Achieving these conditions in real-world controlled environments is challenging.

<u>Data involved in the use case:</u>

- Raw video data: This includes video frames captured from the environment by LIS, which can be used for user and obstacle identification and tracking.
- Radio Frequency (RF) sensing data: This involves capturing RF signals, such as received power levels, Signal-to-Noise Ratio (SNR) values, and antenna radiation patterns, which can be used for radio signal propagation estimation and network performance evaluation.
- Timestamped and space-referenced communications traces: This data captures the timing and spatial information of network communications, aiding in network performance analysis.
- Geographical coordinates: These are the coordinates of potential network users and obstacles, enabling precise positioning and spatial awareness.
- Machine Learning (ML) algorithms: ML algorithms can use the mentioned data types as input for training and inference, enabling object identification, network optimization, and prediction tasks.

CONVERGE Tools involved in the use case:

- CONVERGE Tool 1 (Vision-aided Large Intelligent Surface) enables high-precision 3D positioning and environment sensing, allowing for the estimation of geographical coordinates of potential network users and obstacles to LoS.
- CONVERGE Tool 2 (Vision-aided mobile base station) enables communications and sensing, while illuminating the LIS. The robotic arm/overhead crane system, located in the CONVERGE's anechoic chamber, will enable the replication of the real UAV positioning within an airborne network in a secure manner. The airborne network can be combined with the static vision-aided base station, in order to validate novel solutions for optimized network performance. Additionally, a representative number of vision-aided mobile User Equipment (UE) devices can be configured in the CONVERGE's anechoic chamber to exchange traffic with the mobile and fixed vision-aided base stations, considering characteristic traffic patterns and the presence of irregular and movable shapes representing obstacles to LoS.
- CONVERGE Tool 3 (Vision-radio simulator and 3D environment modeler) creates a digital 3D representation of the environment, including geometric models suitable for accurate radio signal propagation estimation, based on RF sensing and vision data gathered the LIS. The vision-radio simulator and 3D environment modeler will ensure enhanced realism for subsequent experiments and enable reproducibility and replicability of experiments under the same exact conditions, without requiring physical access and real-time interaction with the equipment in the CONVERGE's anechoic chamber.
- CONVERGE Tool 4 (Machine Learning algorithms) assist in identifying the environmental conditions based on timestamped and space-referenced RF and video captures, including obstacles and network users, while learning incrementally from data collected over time. Moreover, the Machine Learning algorithms will allow determining the optimal position for the UAV, replicated by the robotic arm/overhead crane system carrying the mobile vision-aided base station, in order to meet the users' QoS requirements.

4.4.2 UC 4.2: Motion Tracking and Object Recognition

In the context of smart home or office spaces, the deployment of various types of wireless sensors, such as radio frequency (RF) sensors, infrared sensors, and ultrasonic sensors for human movement tracking and object recognition has become an increasingly prevalent practice. These sensors collect a plethora of data such as signal strength, time of flight, and phase shift, which can be processed to determine the location and movement of people and objects in the environment. However, these environments are dynamic and non-stationary, with furniture, appliances, and even the human occupants frequently moving or adjusting their positions. This constant change in the environment introduces variability in the RF propagation characteristics due to factors like multi-path fading, shadowing, and Doppler shifts. These factors can significantly affect the accuracy and reliability of the motion tracking and object recognition systems.

Traditionally, manual calibration and trial-and-error approaches have been used for sensor placement, resulting in suboptimal performance and inefficient resource usage. Optimizing the placement of sensors and antennas is a crucial task that should involve determining their best locations and orientations to maximize their coverage and minimize signal obstruction or interference. Such a complex task involves the consideration of the RF propagation characteristics in the environment, the sensor's specifications (such as its range and field of view), and the anticipated movements and locations of the humans and objects in the environment.

To handle the high dimensionality and complexity of this task, advanced techniques such as machine learning and optimization algorithms may be employed. For instance, machine learning algorithms can be used to learn the typical patterns of human movement and object placement in the environment, which can then inform the sensor placement strategy. On the other hand, optimization algorithms can be used to find the optimal sensor placement that maximizes coverage and minimizes signal obstruction or interference, given the current state of the environment. In this context, the CONVERGE's suite of tools could support this use case through the simulation of the propagation of wireless radio signals within a realistically modelled environment, providing valuable insights for optimizing the sensor and antenna placement. The tools' ability to simulate the remodelling allows for flexibility in accommodating changes in the environment, such as different furniture placements.

Through the integration of the CONVERGE tools, this use case aims to enhance the effectiveness of motion tracking and object recognition systems in dynamic smart environments (e.g., Figure 20), leading to more efficient and reliable smart home or office solutions.

Note that, although this use case is introduced under the Media vertical, it can be also applied to the Manufacturing vertical. The information that is gathered from the manufacturing environment can be used to improve the communications (view to better communicate) by, for example, predicting Line-of-Sight (LoS) blockages between IoT devices such as robots and the network infrastructure, and optimizing beamforming and LIS configuration.

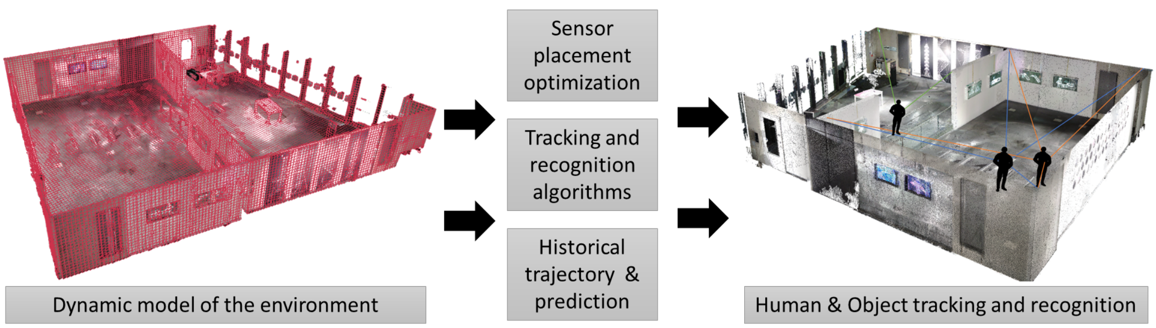


Figure 20 - UC4.2: Motion Tracking and Object Recognition in Smart Environments. High-level scenario.

<u>Data involved in the use case</u>:

The data used in this use case includes information from wireless sensors that track human movement and recognize objects. Specific data points would include:

- Trajectories of human movements and activities.
- Position and orientation of objects (furniture, appliances, etc.).
- Data on the layout of the environment.

This data is stored historically, which allows for retrospective analysis in various contexts, such as assisted living or space optimization. The data is used to optimize the placement of sensors and antennas for accurate motion tracking and object recognition. It is also stored on a historical basis for further analysis in multiple contexts, such as assisted living.

CONVERGE Tools involved in the use case:

- CONVERGE Tool 1 (Vision aided LIS) and Tool 2 (Vision-aided Base Station) can supplement the data collected by wireless sensors with additional visual information. This holistic approach allows for a more comprehensive and accurate system for motion tracking and object recognition.
- CONVERGE Tool 3 (Vision-radio simulator and 3D environment modeler) can be utilized to simulate the propagation of wireless radio signals within the environment. This assists in optimizing the placement of sensors and antennas, considering the layout of the environment and possible changes in furniture placement.
- CONVERGE Tool 4 (Machine Learning algorithms for dataset analysis) can be leveraged to analyse the collected data and improve the accuracy of motion tracking and object recognition. By identifying patterns in the data, the tool can learn to distinguish different types of human movements and recognize specific objects. Additionally, based on historical data, it can predict future movements or behaviours, enabling proactive adjustments to sensor placements and system configurations.

4.4.3 UC 4.3: Immersive Telepresence

Skyport [Mat2022], Project Starline [Law2021], or Holoportation [Ort2016] are just a few examples of research that recently explored the idea of 3D immersive telepresence, where users interact with a remote environment in real-time as if they were physically present. While the concept might be traced back to the 60s [Iva1965], recent advances in computer graphics, sensing, algorithms for advanced processing pipelines, and the availability of better Augmented Reality (AR) and Virtual Reality (VR) devices are significantly contributing to enhance the user-experience when operating these systems Applications range from telemedicine, robotic surgery, remote learning, virtual tours, real estate, managing remote machinery or hazardous environments to video gaming and other entertainment scenarios. The ultimate goal is to make the interaction as realistic and seamless as possible, reducing the barriers of distance and time.

A critical component of 3D telepresence systems is the capability to sense and represent the environment and people in it using sensors and cameras to track and replicate the user's movements in the remote environment, including head movements, hand gestures, or body position. The success of deep learning techniques in performing visual recognition tasks, such as object and people detection, pose detection, and people reconstruction, contribute to these systems' increased practicality. However, vision-based systems struggle with challenging lighting conditions, occlusions, scale variation across multiple cameras, and background clutter. Alternative sensing modalities to overcome these limitations are essential for immersive 3D telepresence. These can replace or complement vision systems to sense people and the surrounding environment, for example, by fusing the information for different sensors (see, e.g., [Mil2021]).

The integration of various sensing modalities, including vision-based systems and complementary sensors, unlocks new possibilities for achieving truly immersive 3D telepresence experiences. By combining the strengths of different sensing technologies, such as depth sensors, thermal cameras, or wearable haptic devices, researchers can overcome the limitations associated with individual modalities and create a more comprehensive and accurate representation of the remote environment. This fusion of sensor information facilitates a more natural and intuitive interaction between users and their virtual counterparts, making the telepresence experience feel more authentic and lifelike.

CONVERGE is expected to contribute to enhance 3D virtual telepresence systems where a combination of vision and Large Intelligent Surfaces (LISs) is used to sense the environment. Figure 21 depicts the proposed system, where two different locations are equipped with cameras and LISs to form a virtual representation of the environment and people in it that can be used to allow remote users wearing AR headsets to interact as if they share the same physical location. On the right, we can see the view of users with an AR headset, combining the people and objects at both locations. The capturing solution can be also enhanced with additional sensors to support emerging novel display technologies such as HoloProto[3].

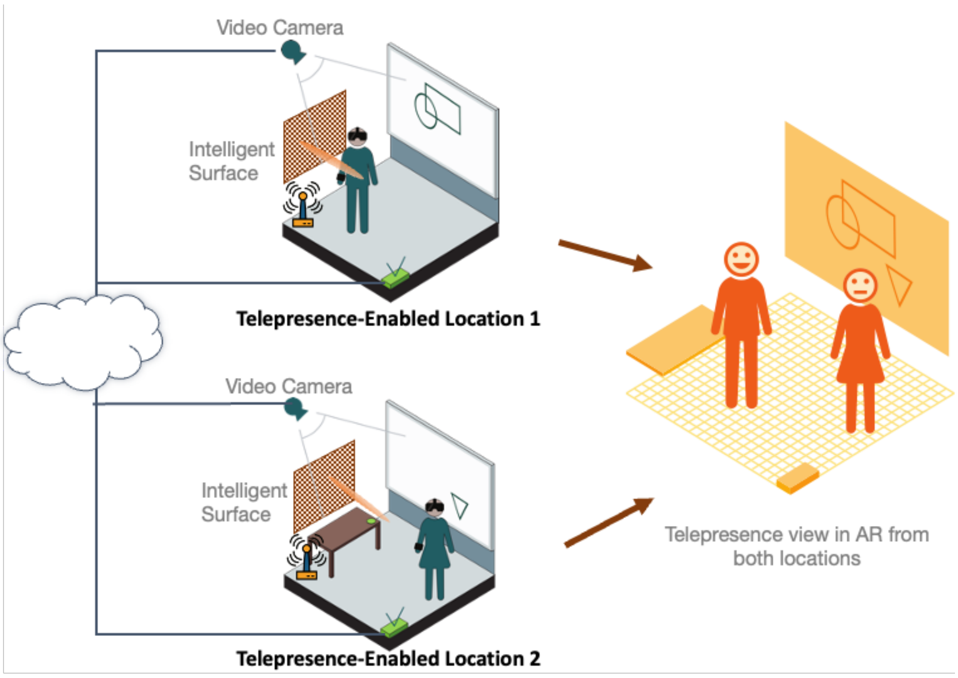


Figure 21 - UC4.3: Immersive 3D Telepresence Scenario.

Data involved in this use case:

- Video streams coming from cameras in the environment.
- Pose and tracking information from a visual processing pipeline.
- Sensing data from the LISs.
- Sensing data from RF devices.

CONVERGE Tools involved in the use case:

- CONVERGE Tool 1 (Vision-aided Large Intelligent Surface) is a critical piece of this use case, complementing the data coming from the vision systems. Algorithmic advances developed with

[3] https://protohologram.com/about/

the support of Tool 1 will be required to process the LISs sensing data in real-time and obtain people and environment information.

- CONVERGE Tool 2 (Vision-aided base station) can be used to generate the radio signals that illuminate the LIS.
- CONVERGE Tool 3 (Vision-radio simulator and 3D environment modeller) can be used to model the environment the users will be placed in.
- CONVERGE Tool 4 (Machine Learning algorithms for dataset analysis) can be used in the context of this UC for training and fine-tuning different ML techniques to detect and tracking objects and persons. It will also be the basis for training ML models that process RF data from the LIS.

4.5 Health vertical market

Radio Frequency and Microwave based imaging can be advantageous for making computer vision applications more robust against occlusion and failure as well as making them light-independent. Such instruments could be a complement to the traditional devices in a multimodal scenario. Cameras based on this type of technology could result in portable, safe, non-invasive and inexpensive devices for multiple usages. For instance, in health, many medical devices used for radiological-based diagnosis, such as computerised tomography, which are based on ionising radiations, could be replaced by this type of technology. Another example, would be the recognition of humans, objects and scenes in 3D environment, making use of the capability of creating depth maps from the acquired signals (like a sonar) and working in light-independent conditions, namely with no lighting (like an infra-red device), with the advantage of improving the resolution and accuracy, for example using multiband RF, or improve the efficiency of detection of objects in near mode in comparison with IR devices. Besides imaging, by making use of Doppler signal processing, the holographic intelligent surface can also be used for medical purposes, particularly for fall-detection and non-contact detection of vital medical signatures, such as heart rate and respiration rate. Such a statistical analysis of backscattered signalstoo without reconstructing an image can be done in real-time.

A potential use case of the CONVERGE toolset would be the ability of creating subject models that would allow automatic assessment of subject posture and prosthetic device alignment. By having a user with a prosthetic limb walking inside the chamber it would be possible to assess walking posture and radio signal propagation from the sensors in the socket. This would allow the improvement of the sensor antenna design used to transmit the data to the processing and storage unit. An initial 3D Model can be used to measure and compare posture, gait and device alignment changes supported by the Machine Learning algorithms that will combine the visual information with Inertial Measurement Unit (IMU) data, in order to improve socket and sensor design/location. The CONVERGE toolset could be used to develop a clinical focused product to be used in a clinical setting focused on looking into patient movement, gait, posture and device alignment. By recording and comparing models and using machine learning algorithms using both IMU and vision data, the optimal clinical outcome can be determined, and the progress of the patient can be tracked and assessed by physicians. The impact of such application will be considerable, not only due to the benefits to the patients, but to the healthcare community since this is something that doesn't exist in the industry. This could have a significant impact in the rehabilitation market, not only within the orthotics and prosthetics (O&P) industry but also more generally into any physical rehabilitation facility.

4.5.1 UC 5.1: Image Sensing and RF Sensing Patient Monitoring

RF sensing can be used to gather both visible and invisible information about a patient's posture and gait. Radio signals and patient posture are filmed while the patient moves through the camera to create a virtual patient for gait analyses. The simulation of radio propagation signals can improve antenna placements and antennas that are used to collect daily activity while being embedded in the patient socket. This is extremely useful since sockets are made of materials that can deflect and change the radio propagation reducing the signal efficiency thus increasing the battery consumption. Sensor data taken from IMU sensors can be merged with vision data to improve gait detection algorithms. The creation of new algorithms can improve daily activity which is extremely important to decide on the correct fitting and in the patient activity level used by physicians to prescribe the correct socket and prosthesis components.

Tagged Image data can be used to improved Gait detection using IMU data, merged with pressure values. Usually at least two IMU's are needed to identify all 5 gait phases. Using Machine learning with Tagged Image from the user walking inside the chamber (see **Error! Reference source not found.**) can be used to i mprove and identify in the IMU and pressure values the different gait phases. It can also be used to identify obstacle transposing, walking on ramps, walking in steps among other activities that can be simulated inside the chamber. The Machine Learning can identify and tagged these activities automatically to allow then to be feed to our current Machine Learning pipeline.

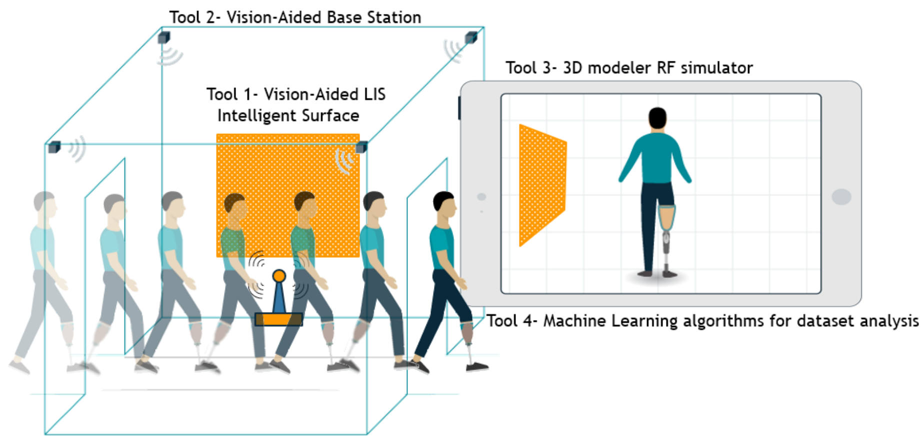


Figure 22 - UC5.1: Example of patient monitoring through image and RF sensing.

<u>Data types involved in the use case:</u>

- Raw data coming from RF sensing provided by existing communications systems (e.g., 5G BS).
- Raw video streams captured by cameras.
- Raw data coming from IMU.

<u>Tools used in this case:</u>

- CONVERGE Tool 1 (Vision-aided LIS) can be used to perform RF sensing of the patients by retrieving RF data from existing communications hardware, such as 5G.
- CONVERGE Tool 2 (Vision-aided base station) can be used to perform Vision sensing of the patients.
- CONVERGE Tool 3 (3D environment modeler and vision-radio simulator) can be used to simulate the propagation of wireless radio signals in the environment, and to optimize the placement of sensors and antennas.
- CONVERGE Tool 4 (Machine Learning algorithms for dataset analysis) can be used for the fusion

of IMU sensor data with vision data to improve Gait detection algorithms, as well as for the creation of new algorithms for object/activity detection. The extracted trajectories of humans and their poses can be stored on a historical basis for further analysis.

4.5.2 UC 5.2: RF Signals to Extract Breathing and Heartbeats

Health monitoring equipment is increasingly prevalent as public health receives greater attention. The primary purpose of most health monitoring devices is to detect physiological signals in humans, including heart rate and respiratory rate. These signals serve as indicators of overall health, encompassing both mental and physical well-being, and can also aid in categorizing sleep stages. By monitoring respiration rate, heart rate, and sleep stages, healthcare professionals can assess health conditions and diagnose disorders. Numerous health issues, such as depression, insomnia, obesity, and other ailments, can potentially benefit from this technology. The advent of RF technologies has ushered in an era of smart health and medical care, which offers distinct advantages, particularly within the healthcare domain.

Chronic conditions like heart disease affect many individuals, while sleep problems are increasingly prevalent. Sleep quality is closely linked to certain health issues like sleep apnea, asthma, and chronic insomnia. Therefore, it is crucial to monitor human physiological signals, such as heart rate and respiratory rate, both during sleep and wakefulness, in order to track overall health, sleep status, and critical conditions like preterm infants. Consequently, public awareness of the progress in smart medical care and AI biomedical technology has been heightened. With the rapid growth of the IoT and AI, the issue of individuals being unable to monitor their health and sleep quality has been partially alleviated. Researchers have developed various wearable devices for health monitoring. However, due to concerns regarding long-term discomfort, researchers have started designing smart healthcare devices that utilize common appliances for monitoring purposes. For instance, basic physiological parameters such as heart rate and respiration rate can be monitored using RF sensing.

An application example relevant to this context involves the monitoring of preterm infants, represented in Figure 23. Contact-free monitoring technologies are needed for critically ill preterm infants to eliminate discomfort and potential harm caused by adhesive electrodes, temperature sensors, and saturation sensors. Atypical or irregular respiratory frequency is considered an early marker of physiological distress. Furthermore, monitoring this vital parameter plays a significant role in diagnosing respiratory disorders and detecting sudden infant death syndrome at an early stage. However, current measurement methods require sensors to be attached to the patient's body, leading to discomfort and stress. In this particular use case, our focus will be on non-contact physiological monitoring of infants using RF sensing using existing communication systems (e.g., 5G), offering a less invasive monitoring approach. Additionally, as a traditional way to solve this issue, the video streams will serve as a benchmark for validation of the solution.

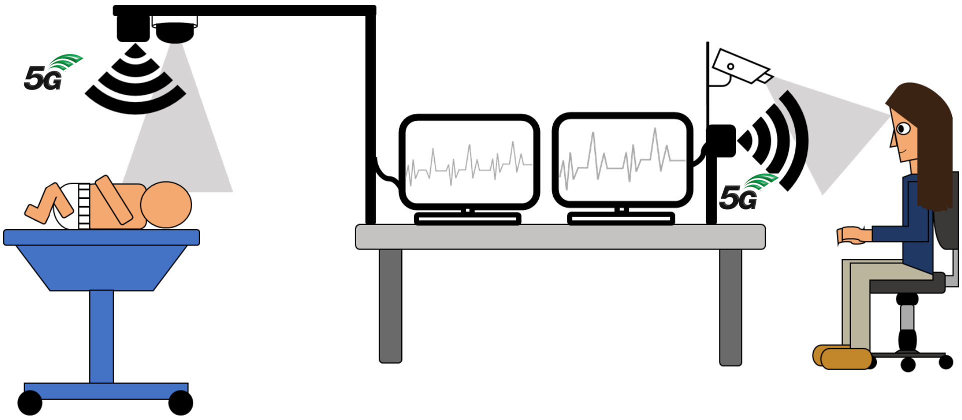


Figure 23 - UC5.2: Example of contact free extraction of breathing and heartbeat information through RF sensing based on existing communications systems, such as 5G.

Data types involved in the use case:

- Raw data coming from the RF sensing functionality provide by the communication systems (e.g., 5G Sounding Reference Signals for the uplink, and Reference Signals for the downlink).
- Raw video streams captured by cameras.

CONVERGE Tools involved in the use case:

- CONVERGE Tool 1 (Vision-aided Large Intelligent Surface) and Tool 2 (Vision-aided base station) will be used for the retrieval of RF data using the sensing capabilities from existing communications hardware.
- CONVERGE Tool 2 (Vision-aided fixed and mobile Base Station) will be used to generate the radio signals that will illuminate the LIS.
- CONVERGE Tool 4 (Machine Learning algorithms for dataset analysis) can be used in the context of this UC for training and fine-tuning different ML techniques.

4.6 Mapping of Use Cases in CONVERGE Tools

This section presents a map, in Table 3, listing all the use cases described in this deliverable, and the respective CONVERGE tools that are used.

Table 3 - Mapping of Use Cases in CONVERGE Tools.

Vertical markets	Use cases	Tool 1	Tool 2	Tool 3	Tool 4
1. Telecommunications	UC1.1 - Coverage Enhancement for Wireless Communications	x	x	x	
	UC1.2 - Blockage Prediction and Proactive Handover		x		x
	UC1.3 - RF Mapping for Imaging and Coverage Extension	x	x		x
	UC1.4 – Transparent Sensing for Enablement of Enhanced Radio Resource Management	x	x	x	x

2. Automotive	UC2.1- In-vehicle Sensing	x	x	x	x
	UC2.2 – Collaborative Modelling of the Environment		x	x	x
3. Manufacturing	UC3.1 - Automated Assembly Lines	x	x	x	x
	UC3.2 - Digital Twinning			x	x
	UC3.3: Privacy-preserving Human Identification through RF Sensing	x	x		x
4. Media	UC4.1 - Outdoor Vision-aided Mobile Base Station	x	x	x	x
	UC4.2 - Motion Tracking and Object Recognition	x	x	x	x
	UC4.3 - Immersive Telepresence	x	x	x	x
5. Health	UC5.1 - Image Sensing and RF Sensing Patient Monitoring	x	x	x	x
	UC5.2: RF Signals to Extract Breathing and Heartbeats	x	x		x

5. CONVERGE ARCHITECTURE AND REQUIREMENTS

This section presents the technical requirements of the CONVERGE infrastructure in terms of components from the different partners, the performance of those components, as well as the interfaces. The project partners provide their requirements for the integration of their components and solutions in their own and/or other partner infrastructure.

5.1 CONVERGE infrastructure architecture

The CONVERGE high-level architecture, depicted in Figure 24, consists of 3 main building blocks: the **CONVERGE Chamber**, the **CONVERGE Core**, and the **CONVERGE Simulator**. The CONVERGE Chamber represents the experimental chamber to be developed in the project to be deployed in some research infrastructures; it contains 3 types of equipment: the CONVERGE gNB, the CONVERGE UE, and the CONVERGE LIS. The CONVERGE gNB will have functionality for controlling the placement of gBN (PCg), radio communications functionality (RCg) defined for 5G gNB, O-RAN based, functionality for radio sensing (RSg), and functionality for vision sensing (Vsg). The CONVERGE UE will have functionality for controlling the placement of UE (PCue), radio communications functionality (RCue) as defined for 5G, and functionality for radio sensing (RSue). The CONVERGE LIS will have functionality for controlling the placement of LIS (PClis), radio communications functionality typical of a LIS (RClis), functionality for radio sensing (RSlis), and functionality for vision sensing (VSlis). The CONVERGE Chamber equipment will be connected and controlled by the CONVERGE Core building Block.

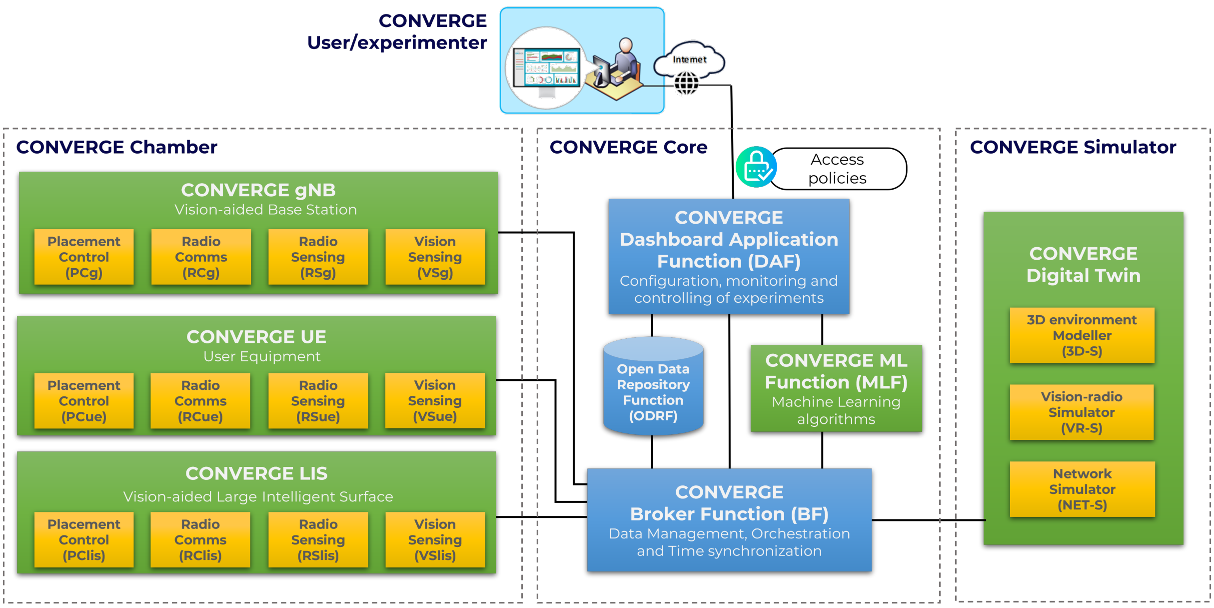


Figure 24 - CONVERGE high-level architecture.

The CONVERGE Core building Block is responsible for controlling the operation of the equipment located in the chamber, as well as the simulator, and it consists of several functions. The CONVERGE Dashboard Application Function (DAF) will interface with CONVERGE users, enabling them to configure, monitor and control experiments to be carried out inside CONVERGE. The management of a CONVERGE session will be carried out by the CONVERGE Broker Function (BF) that will orchestrate execution a CONVERGE session carried out by equipment or simulation tools, being also responsible for coordinating the storage of data resulting from a CONVERGE session. The Open Data

Repository Function (ODRF) will store data generated by CONVERGE, including open data sets. The CONVERGE Machine Learning Function (MLF) represents the functions made available by the machine learning tools developed in the project; these tools may be used by external users, CONVERGE equipment, or CONVERGE simulation tools.

The CONVERGE Simulator building block consists of 3 simulation tools. The 3D Environment Modeller (3D-S), the Vision Radio Simulator (VR-S), and the Network Simulator (NET-S). Together, these tools will enable recreating digitally a physical environment (Digital Twin), modelling the radio and vision communications in the environment by using ray tracing like capabilities, and generating simulated traffic considering the realistic physical channel models obtained by the previous two tools. Sessions in the CONVERGE Simulator are likely to be controlled also by the CONVERGE Broker Function and the simulations scripts and simulation results stored in the Open Data Repository Function.

The detailed description of the CONVERGE architecture, including the identification of data plane and control plane modules along with their interfaces is left for deliverable D1.2.

5.2 Requirements of individual components and interfaces

Starting from the agreed infrastructure architecture, the individual components are described in this section with regards to their technical requirements, and interfaces between each other in tabular form. Individual aspects of augmented RAN components will be respected, such as network deployment scenarios, coverage, desired number of beams as well as number and trajectories of users, base station antenna arrays, obstacles, reconfigurable intelligent surfaces, and cameras of different types. The software and hardware requirements will be elaborated as a function of radio frequencies and bandwidths. Other requirements concern exchangeable data formats and storage requirements. The protection goals of availability, integrity, authenticity, and confidentiality will be addressed.

5.2.1 Tool 1 Requirements - Large vision-aided intelligent surface

The vision aided large intelligent surface (VA-LIS) is intended to support experimentation with LIS-aided communications and sensing, including mapping, imaging, and localisation. The LIS will be equipped with a control unit designed specifically for this application, offering a seamless integration through a python API, providing flexibility for effortless testing of new algorithms. With its modular design, the LIS control unit enables easy experimentation with various applications, allowing customization and tailoring to specific needs. Moreover, through the API, the LIS will seamlessly integrate with the CONVERGE toolset, ensuring smooth and reliable communication between components.

Radio and vision requirements:

The VA-LIS is estimated to meet the requirements present in Table 4 and Table 5:

Table 4 - Requirements of the VA-LIS radio component.

Parameter	Value/characteristics
Dimensions (Length x Width)	1000 mm x 1000 mm
Mobility	1 axis – horizontal 2 axis – horizontal and vertical (optional)
Frequency range of operation	FR1: 3.4-3.8 GHz

	FR2: 27.0-29.5 GHz
Polarization	H-pol and V-pol
Beam Scan Range (Azimuth, Elevation)	+/- 60°
RF power dissipation at reflection, maximum	3 dB
Beam Steering rate	< 50 ms
Control software	Python-based
Near-to-far field operation mode	Yes (using horn antenna in the near field)
Modes of phase-profile configuration	Far-field Near-field (optional)
Control of phase profile	Random and pre-defined profiles Time-dependent control (for localisation and sensing applications)
Pre-defined phase profiles	Beam control: beamwidth control, beam direction Custom: controlled by the user
Unit cell phase discretisation	1-bit 2-bit (optional)
Capability of signal generation	Near field spatial feeding (single stream) Integrated (dynamic meta-surface concept, possibly with multiple streams, optional)

Table 5 - Requirements of the VA-LIS vision component.

Parameter	Value/characteristics
Video cameras	Large field of view cameras (180 degrees) Standard field of view cameras Multiple cameras distributed along LIS Multi-camera PTP synchronization enabling the fusion of multiple video streams

Interfaces

- Interface between the vision-aided LIS and the Converge Broker Function, enabling the configuration of the VA-LIS tool for a specific experiment, including the control of:
    - VA-LIS placement control in the horizontal axis.
    - Vision sensing configuration.
    - LIS phase profile configuration for radio-based communications.
    - LIS phase profile configuration for radio-based sensing.

- The configuration of the experiment parameters in the CONVERGE Dashboard Application Function should enable two types of control:
  - Manual control by the user directly at the DAF.
  - Automatic control through an API, enabling the user to link the control of the parameters to a specific algorithm under test, which may use input data from the experiment itself.

Description of the data generated by Tool 1

Depending on the mode of operation, the LIS may work either as a standalone antenna array (not connected to any radio transceiver), or as a hybrid system including one (or optionally more than one) radio transceiver chain combined with the LIS.

A standalone LIS does not allow for any data collection, but data can be collected by external devices such as those used as part of Tool 2, under the influence of the LIS. Such data can be in the form of real-time monitoring of gNB and UE network parameters: RSSI, RSRQ, SINR, PHR, CQI, RI, PMI, PUCCH SNR, PUSCH SNR, Downlink Bitrate, Downlink MCS, Downlink BLER, Uplink Bitrate, Uplink MCS, and Uplink BLER. Additionally, RF sensing data can be captured, as a result of using a pre-determined sensing protocol and commodity wireless signals (from the 5G standard). In the case of a hybrid system including one (or more) radio transceiver chains combined with the LIS, the above-mentioned data can also be captured at the radio transceiver connected to the LIS, namely by probing the used transceiver chains. Finally, video data streams will be generated by the video-cameras employed as part of this Tool.

5.2.2 Tool 2 Requirements - Vision-aided base station

The vision-aided base station is aimed at enabling communications and experimentation with mobile terminals mostly related to beamforming, multi-user access and opportunistic scheduling by taking advantage of environment mapping made by video-cameras. Both fixed and mobile versions of the base station will be developed, the latter adding the possibility of controlled mobility (different positions along the time or predefined trajectories).

Radio and vision requirements:

The vision-aided BS is estimated to meet the requirements present in Table 6 and Table 7:

Table 6 - Requirements of the vision-aided BS radio component.

Parameter	Value/characteristics
Cellular technology	5G NR SA FR2 O-RAN compliant, split option 7.2x
Protocol stack	OAI 5G CN and OAI gNB open source Beam management procedures compliant with 3GPP R16
Antenna technology	Multibeam operation with 2 independent steerable beams per O-RU. 8x8 antenna elements. Horizontal/Vertical beam scan angle –60 to 60 deg. Phase resolution of 5.6 deg.
Frequency bands	3GPP bands n257, n261 (26.5-29.5 GHz)

SCS	120 kHz subcarrier spacing
Bandwidth	Up to 400 MHz
UL/RL Duplex mode	TDD with a reconfigurable slot structure
Max gNB EIRP	49 dBm
Max. Throughput DL	>1 Gbps
Max. Throughput UL	>400 Mbps
Range	10 m (indoor) and 100 m (outdoor)
UE	COTS UE 5G FR2 OAI UE 5G FR2 connected to an USRP
Monitoring of real-time parameters	Real-time monitoring of gNB and UE parameters: RSSI, RSRQ, SINR, PHR, CQI, RI, PMI, PUCCH SNR, PUSCH SNR, Downlink Bitrate, Downlink MCS, Downlink BLER, Uplink Bitrate, Uplink MCS, and Uplink BLER.
Data export capabilities	gNB and UE data logs exported in JSON format.
Mobility	pitch, roll, yaw, and heading orientation capabilities, on top of the basic 3D (x, y, z) positioning ability inside the room/chamber

Table 7 - Requirements of the vision-aided BS vision component.

Parameter	Value/characteristics
Video cameras	360º video camera with embedded AI Multi-camera PTP synchronization enabling the fusion of multiple video streams
LiDAR	16 Channels Measurement Range: 100 m Range Accuracy: Up to ±3 cm Field of View (Vertical): +15.0° to -15.0° (30°) Angular Resolution (Vertical): 2.0° Field of View (Horizontal): 360° Angular Resolution (Horizontal/Azimuth): 0.1° – 0.4° Rotation Rate: 5 Hz – 20 Hz Integrated Web Server for monitoring and configuration

Interfaces

- Both fixed (but relocatable) and mobile versions of the base station will be developed, the latter adding the possibility of controlled mobility (different positions along the time or predefined trajectories) crane system.
- Interface between the vision-aided base station and the Converge Data Broker.
- Interface between the gNB DU and RU and the external LIS along with the scheduling/beamforming procedures to jointly control the beam-selection and direction

(transmit/receive) of multiple mmWave arrays and phase-shifting of the LIS for both communications and mmWave-based vision. O-RAN E2 interface extensions and xApps will be used to allow for signalling per-user measurements from sensing procedures in the BS.

Description of the data generated by Tool 2

Data that is generated by Tool 2 can be in the form of real-time monitoring of gNB and UE network parameters: RSSI, RSRQ, SINR, PHR, CQI, RI, PMI, PUCCH SNR, PUSCH SNR, Downlink Bitrate, Downlink MCS, Downlink BLER, Uplink Bitrate, Uplink MCS, and Uplink BLER. Additionally, RF sensing data can be captured, as a result of using a pre-determined sensing protocol and commodity wireless signals (from the 5G standard).

The mobile base-station also generates placement information (x, y, z), since its place can be dynamically reconfigured, even during an experiment.

Finally, video data streams will be generated by the video-cameras employed as part of this Tool (fixed and mobile Base Stations).

5.2.3 Tool 3 Requirements - 3D environment modeller and vision-radio simulator

Tool 3, the 3D environment modeller and vision-radio simulator, represents the CONVERGE digital twin which is designed to generate a simulated environment for testing and optimizing wireless sensor systems in an indoor or outdoor environment. The tool consists of a high-quality 3D modelling software capable of creating a realistic virtual environment with accurate dimensions and textures, as well a vision-radio simulator that can accurately model the propagation of wireless radio signals in the virtual environment. This tool has the requirements and interfaces described in Table 8:

Requirements:

Table 8 - Requirements of the 3D environment modeller and vision-radio simulator.

Parameter	Value/characteristics
3D environment modelling – supported methods for environment creation	Implicit methods Explicit methods Includes the representation of material textures
3D environment modelling – format representation	Point clouds Triangle meshes (optional)
Vision-radio simulator – simulation method	Ray tracing (channel modelling at the physical layer), including the interaction of the radio waves with the materials ns-3 (for packet level simulation, optional)
Vision-radio simulator – integration of external data	Integration with data from wireless sensors to provide partial (but accurate) representations of humans and objects Movement of humans and objects must be only partially supported (e.g., slow movement or placement with low degree of freedom)

Vision-radio simulator – layout reconfiguration capabilities	Configuration of different object and furniture placements and layout changes (to optimize the placement of sensors and antennas)

Interfaces:

- Integration with wireless sensors and camera data collected from real-world environments to generate a realistic representation of the layout of the simulated environment, that partially supports moving objects and humans.
- Interface with Machine Learning Function through the Broker function enabling the optimisation of the placement of sensors and antennas in the experimental chamber based on the data generated by the simulation.
- Integration with the CONVERGE toolset to ensure seamless communication with other tools used in the indoor/outdoor environment setting.

Description of the data generated by Tool 3

The data generated by Tool 3 includes a 3D virtual environment model that accurately represents the layout of an indoor environment (e.g., smart home or office) or an outdoor environment (e.g., street intersection), including object, vehicle, furniture, and appliance placement. This model also includes information on the propagation of wireless radio signals within the environment, allowing for the optimization of sensor and antenna placement. Additionally, the simulation generates data on human movement and object recognition, which can be used to improve the accuracy of wireless sensor systems. Overall, the data generated by Tool 3 is critical for the successful design and implementation of an optimized vision-radio environment.

5.2.4 Tool 4 Requirements - Machine Learning algorithms for dataset analysis

This tool will consist of a set of modules that provide different types of data analysis and will allow to connect several of them to implement data analysis pipelines, as depicted in Figure 25. The algorithms may take both sensing and vision data as input. The prediction functionality imposes several constraints for the data that is used for training, as well as computational requirements (generally with respect to CPU/GPU, memory, but possibly also energy that is consumed). Additional constraints may determine where these algorithms are run, e.g., restrictions on latency.

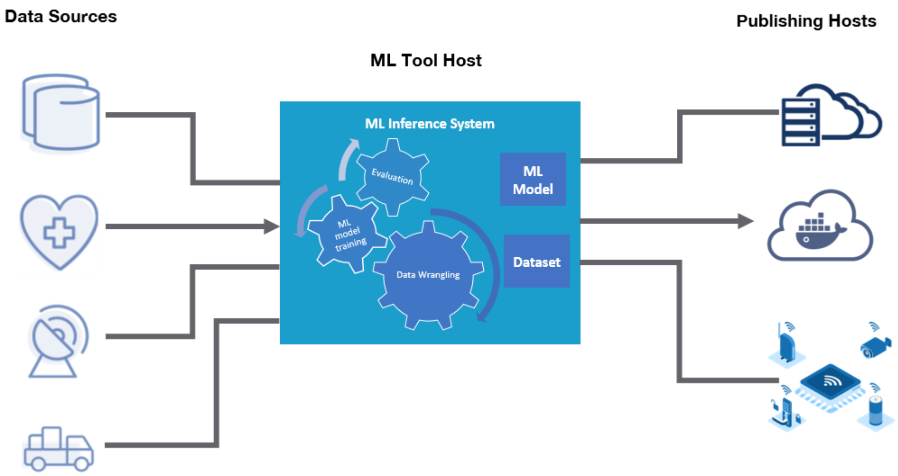


Figure 25 - Data Sources, Publishing Hosts and the components of the ML Inference System that runs in a ML Tool Host.

Requirements:

Concretely, Table 9 lists some of the requisites that are important to be able to provide reliable AI/ML algorithms.

Table 9 - Requirements of the ML algorithms for datasets analysis.

Parameter	Value/characteristics
Data quality	For images: resolution, lighting, angle, non-overlapping (or mostly non-overlapping/non-occlusion) of objects to be identified. For RF data: characteristics of frequency, data streams, nr of channels, bandwidth. General: The variety / fairness of the data in the training set. Cleanness of the data. Ground-Truth (for supervised approaches): correct labels and bounding-boxes (for images).
Data quantity	The number of data items for training purposes, to avoid overfitting. Enough instances for each target item. A good representation of data from all/most possible scenarios to maximize the chance of the algorithms to generalize well.
Ground Truth	The list of objects / actions / emotions to be recognized.
ML methods	Cover at least supervised and unsupervised learning methods.
Data preparation capabilities	Clean, normalize, and prepare the data for analysis. This includes integrity and validity checking (period of time during which the data is considered to be valid, hence usable, to a specific purpose) of sensing data.
The computational resources needed by the algorithms	Each use case will define its requirements in terms of latency, computation volume and speed, data types. These will generate concrete requirements for Memory and CPU/GPU usage, as well as the type of architecture that may be used to ensure e.g., low latency.

Data anonymization	Anonymization will be performed at the (data) source, before data transfer. All data must include a descriptive set of standard metadata to identify its authenticity, the content, the provenance, the confidentiality level, and a license allowing reuse. Confidentiality and licensing come with access and treatment permissions.
Data management	The datasets must be stored using standard formats, and identifiable with unique identifiers, which can permit reference and download. Data findability implies to have a catalogue (public or nor) where the user can explore the metadata and select datasets. Descriptions of data and selection of metadata standards should be well documented, and as broadly accepted as possible in the community, to allow interoperability.
Upload/download datasets	Allow users to upload their own datasets to which to apply data analysis.
Customization of the ML pipeline and (hyper)parameter setting	Expert users will be able to extend and optimize functionalities by defining custom algorithms, loading pre-trained models, specifying their desired set of hyper-parameters, define new evaluation metrics, effectively tuning all analysis phases.

Interfaces:

- Given that Tool 4 is horizontal and meant to provide AI services for the use cases, it will need to interface with all the rest of the tools via the CONVERGE Broker and with expert or non-expert users via the CONVERGE Dashboard Application.
- Non-expert users will be able to configure and use the tool to execute standard data analysis.
- Expert users will be able to use the DAF to customize the modular functionalities to adjust them to their needs, add new functionalities, or eliminate modules that are not necessary for their particular use case.
- The input data for which data analysis is sought will be accessed from the Open Data Repository. Users will be able to deploy the ML tools to fit their performance requirements, e.g., the capability of the devices where a specific algorithm needs to run.

Description of the data generated by Tool 4

Each functionality of Tool 4, presented in Figure 8, has different characteristics:

- **Data Ingestion** loads batch datasets and connects streaming data sources. No new data is generated.
- **Data Wrangling** transforms data into formats appropriate for ingestion by other modules of the tool. Existing data may be cleaned, augmented, structured, inputted, and/or anonymized.
- **Data modelling** applies known mathematical tools to assess the integrity and the validity of sensing data, and provides mechanisms for the inspection of sensing data error sources
- **ML model training** generates predictions and models.
- **Evaluation** generates values for standard evaluation metrics.
- **Deployment** publishes digital components such as pre-trained ML models for transfer learning, ML trained models as a service (API) or as an image (docker/Kubernetes), and datasets as endpoint or download files.

5.2.5 Considerations on requirements of sensing data

This section intends to address the aspects, characteristics and broader view requirements that relate to all the use cases in general, but particularly for the ones presented in this deliverable. There is a wide range of other use cases that could have made it to this document, but there are certain aspects that however seem broader, and are worth capturing.

The motivation includes considering strict new requirements for emergent communication systems like 6G, e.g., higher throughput requirements, which may easily be conflicting with sensing capabilities, or sustainable development. In the latter case, energy efficiency must be considered and embedded in the design of use cases, functionalities, and development of new tools.

System design for both communications and sensing purposes can enable enhanced performance in terms of coverage, throughput, service interruption during mobility, but comes at cost of increased complexity and energy expenditure. Hence, some energy efficiency principles need to be considered throughout the project lifecycle.

Below are some considerations that will be addressed in the context of the trade-off between sensing to communicate and communicate to sense paradigms.

Enablers for smart sensing

Use cases in the telecommunications vertical market, describe radio resource management enhancements that will be possible with the aid of sensing capabilities. The sensing requirements in these use cases should depend, however, on the current communication requirements. For example, it would be of little interest to improve coverage to an area with no devices in need to be served. This applies as well to the automotive and manufacturing verticals, where sensing is not required unless there is traffic or an assembly line in production. Similar to the telecommunications vertical, in the media vertical, sensing is required where there are users to be served.

This denotes a need for investigating broad level functionalities, where sensing capabilities need to be enabled and/or disabled, to serve the particular requirements of the use cases, it being on the setup or during execution of sensing to communicate principles.

Large volumes of data

Another common thread amongst use cases is the availability of several sources of sensing data or sensed information. Different sensing sources provide different information, and different levels of information. One simple example is positioning accuracy, which depends on whether one or several sensing sources are available, as well as on the sensing sources used. Therefore, principles of redundancy, reliability, accuracy, and others should be addressed in relation to the use case and/or market requirements.

Such mechanisms would address the large volumes of data that will be produced and transported over to the CONVERGE tools, by the majority of advanced use cases.

Sensing errors

Several are the metrics that can be used for sensing purposes and all of them are prone to errors. Any error from a sensing source is going to have an impact on the overall sensing integrity capability. In environments with multiple sources of data, considerations such as the origin of the error source, criteria for consideration as an error source, and mapping between an error source and the target functionality, are required.

Examples of sensing errors are ToA, Rx-Tx timing difference (time-based), and AoA or RSRP (angle-based). Such errors can result from multipath/NLoS channel/radio propagation environment, including multipath/NLoS channel itself as an error source.

It is therefore fundamental to address as well sensing errors as a foundational pilar for efficient sensing to communicate and communicate to sense, where identified sources of error can be identified and managed efficiently, so they contribute to the use cases and do not require network resources if unable to contribute.

5.2.6 Alignment of the CONVERGE toolset with the SLICES-RI

SLICES decided to adopt a sustainable and flexible model for the federation of next components of the test platform. The model is broken down in four phases, for ensuring the smooth evolution of the involved facilities to the sustainable model that SLICES RI already adopts [Fdi2022]. The phases of federation address several aspects around the operation of research infrastructures, not limited to only technical infrastructure, but also expands to data management, organizing the communities who benefit from the operation of the facilities, and providing training and education for them. On the technical side, control and user/experimental planes will follow a cloud-native approach, with open and well-established APIs, towards the efficient and automated sharing of resources among the different facility providers, and the integration of novel approaches for their management, AI-driven. The solution will adopt an agile architecture enabling the instantiation of multiple orchestration platforms on top, based on the needs of vertical service providers. In terms of data management, all the artefacts produced over the SUNRISE-6G platform will be annotated using models provided by SLICES [Dem2023] for complying with the FAIR principles and will ensure interoperability with the European Open Science Cloud (EOSC). Specifically, achieving phase 4 will allow the CONVERGE platforms to be aligned with the requirements of the long-term operation of SLICES within the ESFRI roadmap.

SLICES has therefore broken down the approach to different phases of federation that will be adopted by the various CONVERGE test platforms depending on their characteristics.

These are broken down to the following phases:

- **Phase 0 – Current status:** The different facilities operate independently and might use similar tools for their management. Each facility has its own purpose/goal of operation, and its own community of users/experimenters. Through the existing collaborations, facility owners discuss the adoption of different potential common components for the operation of the infrastructure.
- **Phase 1 – Loose Federation of facilities:** During this phase the different facilities federate in a relatively loose manner, only adopting and realizing a minimum set of common federation tools. Typically, these include common credentials for accessing different facilities, common resource descriptions across the different experimental islands, but would not eventually use a common API for the execution of experiments. Equipment dedicated to each facility would be used locally by accessing the facility only. The data produced over the facilities can be shared among them, complying with the FAIR principles, while any other testbed can federate with the existing testbeds, by implementing the respective APIs. The communities of the different facilities can hold joint training/educational activities. Past examples of such a federation paradigm are the OneLab and Fed4FIRE/Fed4FIRE+ projects.
- **Phase 2 – Convergence of Tools:** During this phase, the loosely federated facilities start to converge on the tools that are used locally and exposed to the federated setup. Common tools are adopted across the different sites, with standardized APIs for managing the individual resources. Sharing of data and experiment replicability is achieved over the different facilities, driven by appropriate packaging of experiments (e.g., the definition of Network Service Descriptors – NSD). The communities of the different facilities have joint training and education, while the federated facility can be exposed internationally through the standardized APIs of the control tools of each testbed. Past examples of such a federation paradigm include the ICT-17 and ICT-19 projects using project-specific multi-domain orchestration platforms.

- **<u>Phase 3 – Platform Clustering:</u>** During this phase, clusters of federated testbeds are created. Each cluster is seen from experimenters, logically, as a single facility, which transparently orchestrates resources of the testbeds in the cluster based on the experiment requirements. Data produced over different federations are annotated using similar metadata models, that enhance the findability, accessibility, interoperability, and repeatability/reuse (FAIR) of data across even multidisciplinary platforms. Such a process is followed by platform such as the European Open Science Cloud (EOSC) for enabling data interoperability and exchange.
- **<u>Phase 4 – Full Integration:</u>** For achieving such phase, the adoption of common control and user plane tools across the federated platforms is needed. Common orchestration tools are needed to orchestrate across different sites, as well as common experimental planes locally at the different sites. The overall federated platform is handled as a single facility from outside users, although beneath it comprises from different distributed islands. Data outputted from the integrated platform is following a standardized EOSC compliant format, and can be shared complying with the FAIR guidelines. SLICES implements this type of federation.

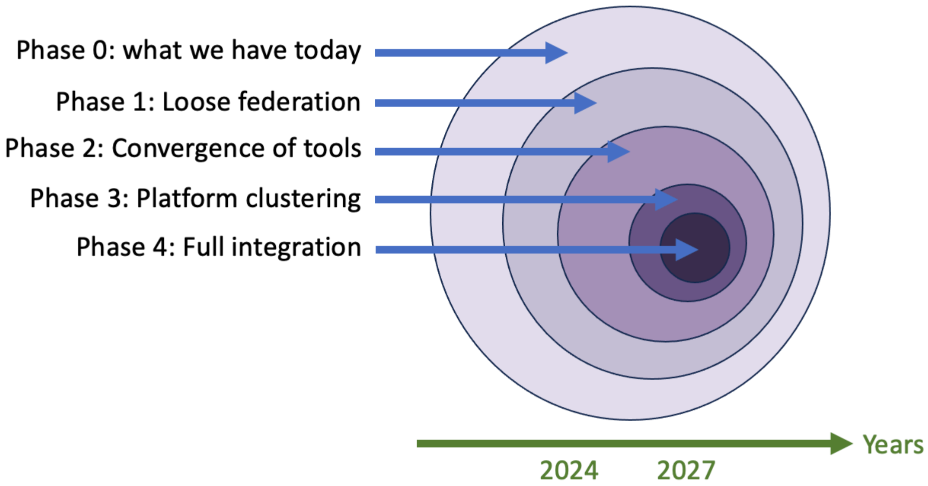


Figure 26 - Onion-model for the phases and timing of federation.

The different CONVERGE facilities are expected to evolve around these phases until the end of the project, according to the timeline (see Figure 26 - onion model). The SLICES partners possess the specific knowhow on how the facilities shall proceed towards achieving the final phase, which in turn will enable the CONVERGE platforms to achieve sustainability. For example, SLICES will guarantee sustainability of the testbed communities at Phase 0, by engaging the communities as stakeholders of the ESFRI initiative. Clearly, achieving Phase 4 will result in complete sustainability from the community, technical, data operation, management, and governance standpoints, which is therefore the (well-grounded) ambition of CONVERGE.

6. CONCLUSIONS AND FUTURE WORK

This document has identified the relevant set of use cases addressable by the project, with a total of 13 use-cases, included in 5 main vertical markets: telecommunications, manufacturing, automotive, media, and health, indicating how the CONVERGE toolset can be used to address each use-cases as well as the types of data involved. The document also described each of the 4 proposed CONVERGE tools, including their requirements, interfaces and types of data generated.

The CONVERGE target user groups and communities have also been presented, which include the scientific academic community, industry, funding agencies, national authorities, national and EU Policymakers, members of 6G IA and other European initiatives supporting research in wireless, vision and their vertical sectors, standardisation organisations and non-European agencies or institutions. It is of utmost importance to clearly identify specific users in the different groups, engage with them and monitor their demand, in order to maximize the value of the CONVERGE platform. In this context, the alignment with the ESFRI SLICES-RI is key, since CONVERGE will clearly enrich and extend the SLICES-RI test platform. Different levels of alignment between CONVERGE and SLICES-RI have been discussed in the document.

The document also introduces a high-level reference architecture for the CONVERGE research infrastructure, which builds on the CONVERGE Chamber and Simulator, and proposes the inclusion of a Core component aimed at controlling the Chamber and Simulator, aligned with the functions of the 5G standard, integrating key functionalities such as 1) Dashboard Application Function, 2) Open Data Repository, 3) Machine Learning Function, and 4) Broker Function.

Finally, while this D1.1 document aims at establishing the target use-cases, requirements and initial architecture, the detailed description of the CONVERGE architecture, including the identification of data plane and control plane modules along with their interfaces, the specification of the user interface as well as the toolset access modes and policies will be addressed in the upcoming deliverable D2.1. These two documents (D1.1 and D1.2) will constitute the baseline reference resulting from WP1 that will drive the developments to be made as part of WP2 and WP3. Figure 27 presents the inter-relations between work-packages, showing that the toolset requirements and specifications in WP1 feed both the toolset development (WP2) and integration (WP3).

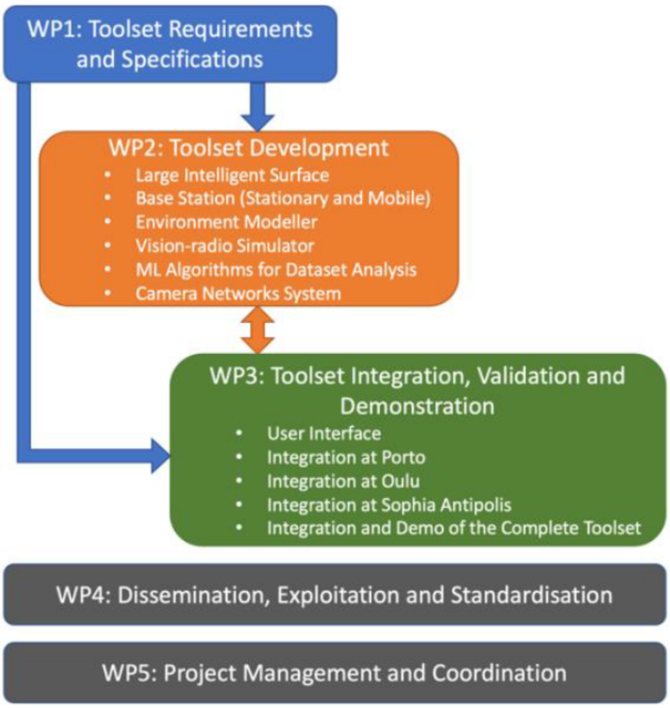


Figure 27 - Inter-relation between WPs.

At this early stage of the project, it is still inappropriate to define which use cases will be implemented as part of the planned demonstrations in WP3. Anyway, the high-level criteria that can be used for this selection later in the project include:

- Priority to use cases that use more than one Tool.
- Good representation of the different CONVERGE Tools in the selected use cases.
- Priority to use cases that are easily generalizable in terms of the use of Tool 4.
- Priority to use cases with higher potential to enable the development of new products and services that may benefit 1) the companies involved in the project and 2) the European R&D community and industry.

REFERENCES

[3GPP2023] 3GPP TR 22.837 V19.0.0 (2023-06) Feasibility Study on Integrated Sensing and Communication (Release 19), 3rd Generation Partnership Project; Technical Specification Group TSG SA. Accessed July 24, 2023.

[Ber2000] Bertalmio, M., Sapiro, G., Caselles, V., & Ballester, C. (2000). Image inpainting. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques, July.

[Com2023] M. Cominelli, F. Gringoli and F. Restuccia, "Exposing the CSI: A Systematic Investigation of CSI-based Wi-Fi Sensing Capabilities and Limitations," 2023 IEEE International Conference on Pervasive Computing and Communications (PerCom), Atlanta, GA, USA, 2023, pp. 81-90, doi: 10.1109/PERCOM56429.2023.10099368.

[Dem2023] Y. Demchenko, S. Gallenmüller, S. Fdida, P. Andreou, C. Crettaz and M. Kirkeng, "Experimental Research Reproducibility and Experiment Workflow Management," 2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS), Bangalore, India, 2023, pp. 835-840, doi: 10.1109/COMSNETS56262.2023.10041378.

[Don2022] Dong, F., Liu, F., Cui, Y., Wang, W., Han, K., & Wang, Z. (2022). Sensing as a service in 6G perceptive networks: A unified framework for ISAC resource allocation. IEEE Transactions on Wireless Communications.

[Fdi2022] S. Fdida, N. Makris, T. Korakis, R. Bruno, A. Passarella, P. Andreou, B. Belter, C. Crettaz, W. Dabbous, Y. Demchenko, R. Knopp. 2022. SLICES, a scientific instrument for the networking community, Computer Communications, Volume 193, 2022, Pages 189-203, ISSN 0140-3664, https://doi.org/10.1016/j.comcom.2022.07.019.

[Gol2017] J. N. Gollub, O. Yurduseven, O., et al. "Large Metasurface Aperture for Millimeter wave Computational Imaging at the Human-Scale," Scientific Reports, vol. 7, no. 1, pp.1-9, 2017.

[Gui2013] Guillemot, C., & Le Meur, O. (2013). Image inpainting: Overview and recent advances. IEEE Signal Processing Magazine, Jan.

[Hua2023] Z. Hua, Y. Lu, G. Pan, K. Gao, D. B. d. Costa and S. Chen, "Computer Vision-Aided mmWave UAV Communication Systems," in IEEE Internet of Things Journal, vol. 10, no. 14, pp. 12548-12561, 15 July15, 2023, doi: 10.1109/JIOT.2023.3251377.

[Ima2020] M. F. Imani, J. N. Gollub, O. Yurduseven et al., "Review of Metasurface Antennas for Computational Microwave Imaging," IEEE Transactions on Antennas and Propagation, vol. 68, no. 3, pp. 1860-1875, 2020.

[Iso2017] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition.

[Iva1965] Ivan Sutherland. 1965. The ultimate display. (1965).

[Law2021] Lawrence, J., Goldman, D.B., Achar, S., Blascovich, G.M., Desloge, J.G., Fortes, T., Gomez, E.M., Häberling, S., Hoppe, H., Huibers, A. and Knaus, C., 2021. Project Starline: A high-fidelity telepresence system

[Lee2023] S. -P. Lee, N. P. Kini, W. -H. Peng, C. -W. Ma and J. -N. Hwang, "HuPR: A Benchmark for Human Pose Estimation Using Millimeter Wave Radar," 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2023, pp. 5704-5713, doi: 10.1109/WACV56688.2023.00567.

[Li2023] Li, W., Ma, Q., Liu, C. et al. Intelligent metasurface system for automatic tracking of moving targets and wireless communications based on computer vision. Nat Commun 14, 989 (2023). https://doi.org/10.1038/s41467-023-36645-3

[Liu2022] Liu, F., Cui, Y., Masouros, C., Xu, J., Han, T. X., Eldar, Y. C., & Buzzi, S. (2022). Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond. IEEE journal on selected areas in communications, 40(6), 1728-1767.

[Los2021] Loscri, V., Vegni, A. M., Innocenti, E., Giuliano, R., & Mazzenga, F. (2021, June). A joint Computer Vision and Reconfigurable Intelligent Meta-surface Approach for Interference Reduction in Beyond 5G Networks. In 2021 IEEE 22nd International Conference on High Performance Switching and Routing (HPSR) (pp. 1-6). IEEE.

[Los2023] V. Loscrí, C. Rizza, A. Benslimane, A. M. Vegni, E. Innocenti and R. Giuliano, "BEST-RIM: A mmWave Beam Steering Approach Based on Computer Vision Enhanced Reconfigurable Intelligent Metasurfaces," in IEEE Transactions on Vehicular Technology, vol. 72, no. 6, pp. 7613-7626, June 2023, doi: 10.1109/TVT.2023.3243358.

[Mal2019] Mallat, K., Damer, N., Boutros, F., Kuijper, A., & Dugelay, J. L. (2019, June). Cross-spectrum thermal to visible face recognition based on cascaded image synthesis. In 2019 IEEE International Conference on Biometrics (ICB).

[Mat2022] Matviienko, A., Müller, F., Schmitz, M., Fendrich, M. and Mühlhäuser, M., 2022, April. Skyport: Investigating 3d teleportation methods in virtual environments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-11).

[Mil2021] Milani I, Bongioanni C, Colone F, Lombardo P. Fusing Measurements from Wi-Fi Emission-Based and Passive Radar Sensors for Short-Range Surveillance. *Remote Sensing*. 2021; 13(18):3556. https://doi.org/10.3390/rs13183556

[Ort2016] Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., Kim, D., Davidson, P.L., Khamis, S., Dou, M. and Tankovich, V., 2016, October. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology* (pp. 741-754).

[Saad2019] W. Saad, M. Bennis and M. Chen, "A vision of 6G wireless systems: Applications trends technologies and open research problems", IEEE Netw., vol. 34, no. 3, pp. 134-142, 2019.

[Sne2023] Snehil, P. K. S and A. Wadhwa, "RF And Computer Vision Based Warning System For Minimizing Road Accidents," 2023 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI), Chennai, India, 2023, pp. 1-5, doi: 10.1109/RAEEUCCI57140.2023.10134135.

[Vac2021] Vaca-Rubio, C. J., Ramirez-Espinosa, P., Kansanen, K., Tan, Z. H., De Carvalho, E., & Popovski, P. (2021). Assessing wireless sensing potential with large intelligent surfaces. IEEE Open Journal of the Communications Society, 2, 934-947.

[Xia2023] Xiang, H., Zou, Q., Nawaz, M. A., Huang, X., Zhang, F., & Yu, H. (2023). Deep learning for image inpainting: A survey. Pattern Recognition, 134, 109046.

[Xu2023] W. Xu, F. Gao, X. Tao, J. Zhang and A. Alkhateeb, "Computer Vision Aided mmWave Beam Alignment in V2X Communications," in IEEE Transactions on Wireless Communications, vol. 22, no. 4, pp. 2699-2714, April 2023, doi: 10.1109/TWC.2022.3213541.

[Yur2016a] O. Yurduseven, V. R. Gowda, J. N. Gollub and D. R. Smith, "Printed Aperiodic Cavity for Computational and Microwave Imaging," IEEE Microwave and Wireless Components Letters, vol. 26, no. 5, pp. 367-369, 2016.

[Yur2016b] O. Yurduseven, J. N. Gollub, A. Rose, D. L. Marks and D. R. Smith, "Design and Simulation of a Frequency-Diverse Aperture for Imaging of Human-Scale Targets," IEEE Access, vol. 4, pp. 5436-5451, 2016.

[Zho2023] Y. Zhou, H. Huang, S. Yuan, H. Zou, L. Xie and J. Yang, "MetaFi++: WiFi-Enabled Transformer-based Human Pose Estimation for Metaverse Avatar Simulation," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2023.3262940.