MINISTERIO DE CIENCIA E INNOVACIÓN

Cofinanciado por la Unión Europea

AGENCIA ESTATAL DE INVESTIGACIÓN

MEMORIA CIENTÍFICO-TÉCNICA DE PROYECTOS INDIVIDUALES
Convocatoria 2021 - «Proyectos de Generación de Conocimiento»

1 DATOS DE LA PROPUESTA – *PROPOSAL DATA*

IP 1 (Nombre y apellidos)**:** Antonio Vallecillo Moreno

IP 2 (Nombre y apellidos)**:** Carlos Canal Velasco

TÍTULO DEL PROYECTO (ACRÓNIMO): Desarrollo de aplicaciones para smart cities teniendo en cuenta a las personas (IPSCA)

TITLE OF THE PROJECT (ACRONYM): Including people in smart city applications (IPSCA)

2 ANTECEDENTES, ESTADO ACTUAL Y JUSTIFICACIÓN DE LA PROPUESTA – *BACKGROUND, CURRENT STATUS AND JUSTIFICATION OF THE PROPOSAL*

2.1 Context and research problem

Currently, there is a growing interest in the social aspects of smart cities, which are cities that use technology to provide useful services to their citizens and to solve urban problems. A smart city aims to improve transportation and accessibility, enhance social services, promote sustainability, or reduce energy consumption, by using digital technologies.

Despite its a priori relevance, the role of individual citizens has often been neglected: smart city applications typically focus on collecting and analyzing data coming from all kinds of sensors and other sources, including some related to human activities, but without really integrating individual users into the loop, just considering them collectively, i.e., as a *crowd*. Moreover, citizens have to hand over their personal information to the service providers, losing control of their data, which becomes the property of third parties. Once under their control, this information can be employed for any purpose, sometimes different from those it was generated for by the citizens – well-known examples are Google or Facebook, which use personal information to provide personalized advertisements and offers, among other things.

In this project we propose to integrate people into the picture, not only from the perspective of synthetic or purely statistical data (e.g., rate of arrival of pedestrians at a crosswalk, or the number of people that use a bus line a particular day) but by exploiting the information available about the people living in these cities, their behavioral habits and personal preferences. In addition, we want to build applications for smart cities that are decentralized and where the information is under the users' control rather than controlled by other entities. The goal is to allow users to have full control of their own data, including access control and storage location, as well as benefit from better services from the external companies and organizations to which they entrust the information. Finally, the individual opinions and levels of trust in a particular service or its providers should also be considered in their interactions and daily decisions.

To develop smart city applications, software engineers are starting to use the concept of Digital Twin City [DWZ21, DZS21], also known as *Urban Digital Twin* (UDT), which is a digital twin able to model specific aspects of a city: transportation, heat maps of density of people, air quality, etc. Like any other digital twin, a UDT allows two-way feedback between the model and the physical entities represented in it. UDTs may help cities not only to perform real-time remote monitoring, but also to recommend adaptation policies and leverage more effective decision-making processes, such as improved urban governance, smart healthcare, smart transportation, optimization of city resources, energy savings, and anticipation of problems.

Platforms and solutions such as FIWARE (https://www.fiware.org/about-us/smart-cities/), FLIR CITY (www.flir.com/applications/smart-cities/), MOCA (www.mocaplatform.com), TCS AT&T Smart Cities (www.business.att.com/categories/smart-cities.html), or Intelligent Urban Exchange (www.tcs.com/smart-city-solutions), among many others, provide a set of data models and APIs for developing smart city applications, and also urban digital twins, using interoperable data schemata and common service components. Although these solutions provide the basic building blocks for any smart city application, they present several shortcomings. First, they force users to hand over their information to them. This way, citizens lose control over their own data and the information they produce. Second, providing this information is not enough to make a city smart [Coh12]: a global and more integrated point of view should be adopted, including a greater implication from people, thus making citizens one of the main pillars in the development of smart cities [MCC+14]. This will enable better planning of city infrastructures, adapting them to the context of the people that use them and their preferences. Third, these platforms do not offer a unified mechanism for communicating with users or providing feedback to them: connections are clearly defined between sensed data and the applications, but not the reverse ones (from the smart city applications to the citizens) – mainly because there is no unified mechanism for communicating with people, since they are just treated as crowds.

In our previous project [COSCA] we realized how important the information from users was for developing distributed collaborative social computing applications, and how its access could be safely controlled and managed by the individuals themselves, who are, after all, the owners and producers of such information. We introduced the concept of Digital Avatar (DA) [MPM+21, BMP+20, PCP20], for representing the virtual profile of a person, to realize collaborative social computing applications. In that proposal, the DA of a user is able to record information about the user and her habits, combine it with data supplied by external sensors or open data providers, and interact with the environment and with the DAs of other users. The DA is also in charge of responding, on behalf of the user, to requests coming from other avatars or from other social computing applications, and of adapting its records and expected behavior accordingly. Thus, DAs enable the empowerment of users, allowing them to take control of the information and contents they generate, manage how all that information is accessed and exploited in a secure manner by third parties, and be in control of their changes and adaptations. As for drawbacks, DAs provide limited support for global decisions and comprehensive planning, which are key elements of any smart city application.

To obtain the benefits that both UDT and DA approaches individually provide, in this project we will propose a distributed architecture and set up a reference framework for the development of smart city applications, able to implement the Digital Twin architecture [TZLN19, MTV21] and to exchange data and interoperate seamlessly with the citizens' DAs. Therefore, users become both producers and consumers of information, and their individual preferences and habits can be taken into account, while their privacy and the access to the information they produce and consume are respected. The framework will provide the necessary infrastructure and appropriate services for the implementation, deployment, analysis and execution of these applications. We will also define models of the habits and behaviors of individuals so that these models can be combined with those of the UDTs to incorporate citizens into smart city applications, and fully exploit all the functionality and benefits of DTs for simulation, prediction and self-adaptation of urban environments. For example, models of the collective behavior of citizens can be more accurately developed from the models of the individual behaviors and behavioral patterns provided by their DAs.

Another essential aspect of this project is the incorporation of uncertainty into the information managed by the applications. Any application that interacts with physical elements and with humans is subject to uncertainty, vagueness, and imprecision. The representation, management, and propagation of uncertainty are required because of the inherent lack of precision in the measurement instruments that manage real data (e.g., GPS sensors or local clocks); lack of confidence in the data sources (due to, e.g., erroneous or incomplete data, or unreliable networks), wrong assumptions about the environment in which these applications operate, or the sometimes unpredictable behavior of humans. The explicit representation and management of such uncertain aspects in the information models cannot be neglected.

As part of the research project, we also plan to study the analysis and verification of the applications being developed, as well as some of their properties, which are critical in the domain of smart cities. In particular, we will address the analysis of their performance, scalability, privacy, security, and reliability properties. We plan to use different techniques, including static and dynamic ones, depending on the property of interest. Simulation methods and, specifically, statistical model checking, play a key role in the project to achieve analysis of complex systems with uncertainty. The automated synthesis of models and patterns of behavior [JM+20] (either of individuals or collective) from the traces obtained by the sensors or by the DAs is another chief aspect of the proposal. Self-adaptation policies and other application-dependent smart services for the UDTs will also be considered.

The project has a duration of 4 years and will be carried out in close contact with some of the companies and institutions interested in these topics, namely the Malaga City Council, the Malaga Public Transportation Company, and SoftCrits, together with the international community researching on issues related to Model-Based Software Engineering (MBSE), UDTs, and Smart Cities. Three exemplary applications (Sect. 2.3) will be used to prototype and validate the proposal, as well as to evaluate its scope, applicability, usability, and effectiveness. Existing solutions and services will be used, where appropriate, to improve reusability and interoperability, and to allow us to focus on the novel aspects of the proposal.

2.2 Supporting concepts and technologies

This section briefly describes the main concepts and technologies used in the project.

2.2.1 Smart city applications and infrastructures

Smart cities are complex entities, composed of myriads of elements of different nature (hardware, software, communication infrastructure, humans, physical elements) interacting among themselves in complex manners and using disparate technologies. The fact that they comprise physical elements and humans significantly increases their complexity. So far, the development of smart city applications has been mostly conducted in ad-hoc and proprietary manners, duplicating efforts and costs, and hindering reusability and interoperability.

There are some relevant initiatives that aim at tackling this problem, such as the aforementioned FIWARE, FLIR CITY, MOCA, TCS AT&T Smart Cities, or Intelligent Urban Exchange, among others. They provide similar services, only using different technologies and infrastructures for developing smart solutions, digital twins, and models to represent the data produced and consumed by a smart city, as well as a set of APIs to manage it, normally using standards such as NGSI-LD (https://en.wikipedia.org/wiki/NGSI-LD). In addition, most of them define reference architectures to support data management, acquisition, processing, analysis and visualization, as well as components that implement common functions such as interfaces to the Internet of Things (IoT) or identity and access management.

Business intelligence services (also called *smart* services) are also part of most of these proposals. With the expansion of big data, integrating multi-source heterogeneous data through learning methods is also becoming commonplace for achieving a comprehensive multi-dimensional perception of the city and its citizens. The set of services in charge of these tasks is typically called the “City Brain” or the “Urban Brain” [DZS21]. The city brain platform is the intelligence hub for the operation of cities, responsible for the analysis, prediction, and intelligent intervention needed for urban governance and city management.

2.2.2 Smart cities: the role of people

Local institutions in smart cities are currently offering open data collected from different sources, including real-time information about air quality and pollution, availability and schedules of city buses and other public transportation methods, etc. Their goal is to enable more efficient management of the resources and services offered by the city, as well as to improve the quality of life of its citizens [NP11]. This is the case, for instance, of the Malaga City Council (https://malagasmart.malaga.eu/), which is actively working in this area.

To achieve this goal, we do not only need to obtain statistical data about how citizens use their services, but also information about the context in which they use them, and their user experience. This will enable better planning of city infrastructures, adapting them to the context of the people that use them. In previous works (e.g., [PC17, GMB+14, NCD+18]) we proposed the People-as-a-Service (PeaaS) model for IoT environments. It allows smartphones to collect and analyze information about their owners, in order to obtain information on their habits and routines, and to use this information for interacting with nearby IoT devices. We used this model for social computing applications in the COSCA project, where the information belongs to the individuals generating it. We now plan to use this model in the context of smart cities, incorporating individual users into the urban digital twin models. This will allow changing the relationship between citizens, their personal data, and the applications used in smart cities.

2.2.3 Digital Avatars

Societal-scale systems (such as Google or Facebook) and most social computing applications use a server-centric model where all the individual information is transferred to the servers of large companies that employ this information to perform all the computations – including those of interest to the users, but also to others, which allows obtaining benefits to those large companies, sometimes even misusing the users' information. This is the model used now by the majority of applications for smart cities. One exception is the Solid project (https://solidproject.org/), which lets people store their data securely in decentralized data stores called *Pods*. Pods are like secure personal web servers for data that allow controlling which people and applications can access it. However, pods are just data stores, with no associated behavior or learning capabilities. To avoid handing over (i.e., donating) all our information to large companies, in our previous project COSCA we proposed the use of Digital Avatars and a peer-to-peer and collaborative approach to social computing.

DAs are digital entities that reside in a user's smartphone or tablet and are able to record information about the user and their habits, combine it with external data provided by external sensors or with open data, and interact with the environment and with the DAs of other users [MPM+21,BMP+20,PCP20] to perform social computing applications. In these interactions, the DA is responsible for ensuring the levels of privacy and security dictated by its owner. In short, a DA serves as a proxy for its owner in the context of collaborative computing applications.

DAs comprise structure and behavior. The structure determines the information relevant both to the owner and to the applications she participates in, and will be persistently stored in the user's mobile phone. The behavior dictates how the DA reacts to external requests and stimuli (e.g., external events, environmental conditions, or context changes) and are specified in terms of Complex Event Processing (CEP) rules. Such rules can also be parameterized using values that can be changed, and that form part of the information stored and maintained in the avatar. The set of rules that define the behavior of a DA can be modified during its lifetime, for example when the DA is engaged in a new application, or when the user decides to incorporate a new behavior. Analogously, rule parameters can be automatically updated by the applications, for example, to adapt that DA behavior to the user's observed practices and current preferences. Finally, they are also in charge of learning from the user behavior about their habits and routines, detecting situations of interest from external stimuli or deviations from behavioral patterns, adapting such patterns according to the actual behavior of the DA owner.

2.2.4 Digital Twin systems

A *Digital Twin* (DT) is a comprehensive digital representation of a real system, service or product (the *Physical Twin*, PT), synchronized at a specified frequency and fidelity [DTC21]. The DT includes the properties, conditions, and behavior of the physical entity through models and data, and is continuously updated with real-time data about the PT's performance, maintenance, and health status throughout its entire lifetime [BCE+20, MML19, HA18]. The exchange of data between the digital and the physical twins takes place through bi-directional data connections. This is why many authors have argued that a DT system contains three dimensions: physical, digital, and the connections between them [Gri14].

Additionally, other authors consider that a DT system should also comprise a set of *Services* that permit exploiting the data exchanged by the two twins in different ways [TZLN18, TZLN19]. Examples of such services include, among others: dashboards for visualizing and displaying the data in different formats; machine learning (ML) components to provide decision support and alerts to users, or to predict changes in the PT over time, e.g., to achieve preventive maintenance; or algorithms to analyze the available data to improve the PT performance or implement self-adaptive mechanisms.

This architecture is shown in Fig. 1 [TZLN19]. It comprises the physical entity (PE), the digital twin or virtual entity (VE), the connection between them (DD), and a set of services (Ss). The communication between them is achieved through a *Data Lake* (DD in Fig. 1). As defined in [HQJ21], a data lake is "a flexible, scalable data storage and management system, which ingests and stores raw data from heterogeneous sources in their original format, and provides query processing and data analytics in an on-the-fly manner." The rest of the components of the architecture use the DL to write data and obtain information in a loosely coupled and asynchronous manner. Basically, this implements a Blackboard architectural pattern [BMR+96] in which the information is stored in raw format (i.e., as produced by the sources).

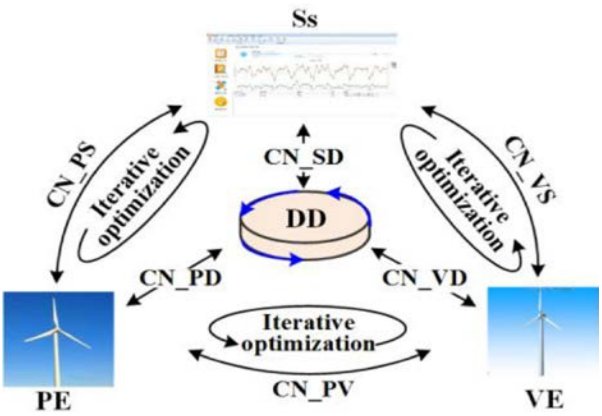

Fig. 1. A digital twin system architecture (from [TZLN19]).

To implement the distributed architecture that we propose in this project, and to avoid users having to send all their information to a centralized data lake (using a push model whereby data is continuously uploaded to the data lake by the users), we will develop a distributed data lake that follows a pull model instead. Thus, the information will reside in the DAs of the citizens and will be requested, on-demand, by the smart city applications. This reversal of flow will ensure secure, controlled access to data, and that each application only uses the information it needs from users, and no more.

2.2.5 Urban Digital Twins

An Urban Digital Twin (UDT) is a virtual representation of a city's physical assets, using data, data analytics and machine learning to help simulation models that can be updated and changed in real-time as their physical equivalents change. A digital twin describes the reality (and the history of it), while it is the additional applications that bring the real intelligence and help create the common picture of reality that is the added value of an urban digital twin.

A UDT can also provide a risk-free testing environment that increases the precision of long-term predictions, improves monitoring and impact assessment of certain decisions that are part of a city's ecosystem. UDTs like the ones developed in, e.g., Singapore, Turkey, Australia, New Zealand, Amsterdam, Cambridge, or Paris [UE20], can bring cost and operational efficiencies, better crisis management, more openness, better-informed decision-making, more participatory governance and better urban planning. In sum, UDTs can change the way cities are planned, operated, monitored and managed [DUET21].

There are, however, a number of challenges related to governance, ecosystem data management (data availability and sharing, data governance, shared data models and standards), cybersecurity and privacy, interoperability and skills, or gender inequalities. Some also believe that while twins for some areas can be very effective, they may be less effective in modeling complex social challenges at city-scale, e.g., social inequalities or disparities in pollution impacts, or implementing personalized services to the citizens [EU20]. In turn, a number of helpful enablers can be used for the replication of urban digital twins in other cities, such as the interoperable urban digital platforms (https://living-in.eu/), the forthcoming data ecosystem for climate-neutral and smart communities, and other initiatives such as EIF4SCC, the European Interoperability Framework for Smart Cities and Communities (https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/news/connecting-eif-smart-cities-communities-eif4scc).

2.2.6 Uncertainty

Uncertainty is unavoidable in any system that comprises or interacts with physical elements or with humans. Uncertainty is defined as the quality or state that involves imperfect and/or unknown information. It applies to predictions of future events, estimations, physical measurements, or unknown properties of a system. Uncertainty can take different forms. For example, *measurement uncertainty* refers to the inability to know with complete precision the value of a quantity. It can be due to different causes, such as unreliable data sources and communication networks; tolerance in the measurement of the values of the physical elements; estimates due to the lack of accurate knowledge about certain parameters, or the inability to determine whether a particular event has actually happened or not. *Belief uncertainty* refers to the inability to decide whether something is true or not, and depends on the individual agent stating such uncertainty.

In [TMBV21] we identified and classified different types of uncertainty, and analyzed how they were represented in software models, and used in the context of MBSE. We also defined an extension of the UML and OCL datatypes to deal with uncertainty [BMBV20] and implemented these types in a library and also as part of the software modeling tool USE [MBOV20]. We plan to extend these results so that measurement and belief uncertainty [J16] can be considered in the information managed by the UDTs and in the patterns that determine the users' behavioral models, too, with the goal of enriching the UDT models and the applications, which can now represent and process information in a more realistic way. Opinions by individuals, trust, and other subjective aspects that are relevant to interactions with and among users will also be implemented in our proposal.

2.2.7 Model synthesis, qualitative and quantitative analysis, and simulations

The high-level architectural models of smart city applications may be used for their formal analysis. Given synthesized models of DAs, and models of UDTs and applications, we will have abstract descriptions of the entire ecosystems on which we will be able to carry on different types of analyses. Specifically, given the particular properties of the individual behavioral patterns plus the novelty of the different alternatives for the design of smart city applications, we are particularly interested in the analysis of their privacy, performance and reliability. The use of probabilistic models for the specification of the behavioral models leads us to consider the use of probabilistic and statistical model checkers for the analysis of such properties. In particular, we plan to provide analysis mechanisms to be able to determine whether a specific service (attending to their QoS requirements and the environment capabilities) will be needed, and also its required configuration for providing optimal service.

Automata and rewriting logic have been extensively used in the past for the formalization of complex systems (see, e.g., [Mes12] for a survey). Multiple techniques are available for the verification of properties on such systems, including static analysis (such as confluence or termination) and different forms of model checking. In this regard, the main challenge is related to the representation of the DAs and UDTs of a city as a family of models, possibly using different formalisms, representing different abstraction levels.

Among the variety of model checkers available, we plan to use MultiVeStA, which is based on simulations to perform statistical model checking. Alternatives such as PRISM and Storm, will also be considered, given the expertise and knowledge by the project members [CPGS18, MCGS15]. Although the formalisms supported by these tools are a priori less expressive, the probabilistic computation tree logic (PCTL) supported by PRISM is very attractive, and Storm presents very good performance metrics (both in terms of computation time and memory).

Some models may be provided by experts, while others may be synthesized from the execution of the physical elements of the city or its citizens. This raises several interesting research questions, related to the synthesis of the models and to the relations between the various models of the DAs and DTs and of these with the physical twin. Although the synthesis of models has been studied for some time now [BF72], there are several recent works (see, e.g., [WTD16, JM+20]) that make the automatic synthesis of models really useful in the context of UDTs, mainly because of their generality and precision.

2.2.8 Smart services

Key elements of any digital twin are the set of smart services that it provides, beyond the faithful simulation of the physical city and its elements. Examples of such smart services include those mentioned above, namely, visualization of data and indicators; machine learning techniques for decision support and prediction of costs and energy consumption, or even undesirable situations such as traffic congestions or overcrowded places; or algorithms to analyze the available data to improve the city performance, suggest changes or implement self-adaptive mechanisms.

We have started to work on improving the performance and accuracy of data mining and ML methods, by using preprocessing techniques [RMV21]. These can be extended in this project to extract useful information about the behavior of individuals. Moreover, there are well-known frameworks for forecasting that include time series analysis [D06] and Long Short-Term Memory networks [HS97], which have been employed in the context of smart cities, e.g., to forecast air pollution [LZ21] and energy consumption [CMV20]. We have also employed these frameworks for predictive self-adaptation in our prior research [MCGS15, CMV20] and plan to adapt them to detect and forecast trends observed in the UDTs. Finally, our work on proactive self-adaptation that has been applied to complex IT [MCGS15] and IoT [CMV20] systems is a good candidate to be adapted and extended to implement smart services for the application case studies that will serve to demonstrate and showcase our proposal in this project.

2.3 Application scenarios

To illustrate the scope and implications of this project, and to validate the project results, three scenarios have been selected. They cover different aspects of the kinds of applications that our framework will support. Each one focuses on one aspect of a smart city (intelligent transport systems, smart buildings and outdoor activities) and exhibits different features and characteristics. The motivation and relevant features of each scenario are highlighted below.

2.3.1 Scenario 1: Urban transport

David is the head of the Public Transport Department of the city of Malaga, which is part of the digital transformation plan to make it a smart city. He now has access to a large amount of real-time information about the use of buses and metro lines in the city, the rental bikes and the electric scooters: occupancy of the urban multimodal transport network, traffic status, punctuality of individual buses and metro trains, etc. An important part of this information is available as open data: transport lines, stops and schedules, buses and metro trains arrival times, bikes and e-scooters pickup and dropping places, incidents, etc. David can also know how citizens use buses and metro lines with their travel cards, either regular (e.g., going to work) or occasional (e.g., attending a concert, or any special event that happens in the city).

To improve their services, David has recently installed a digital twin that allows his department to use the existing information about the transport network to make predictions about peak occupancy hours using time series forecasting algorithms, identify usage patterns, and even plan the resources to optimize their use. The digital twin can also monitor the current state of the network and react to unexpected peaks or incidents by suggesting the diversion of bus routes or the addition of extra metro trains, for example.

However, David believes that his department could provide a better service to citizens if it had more information about them and knew, for example, how they individually use the transport network, their habits and routines in relation to transport, or their behavioral patterns to go to work or to move around the city. This would allow for more accurate predictions, made on an individualized, rather than statistical basis. More importantly, communication with users could be made more effective because the system could individually alert them about events that will affect their plans, something its system cannot do now. This way, the citizens would be able to get personalized services, instead of the generic ones currently offered to them.

In addition, the personal preferences of individuals could be taken into account to suggest alternative transportation methods more suited to their likings and needs. Tourists visiting the city could also benefit from these services while letting the system know that they should be taken into account when estimating the transport needs. Finally, knowing the preferences of citizens could be very useful in deciding about the overall transport solutions to be implemented. If the metro is the preferred option for most of the people who have to go to work at the Malaga Techpark, knowing how many of them would use it if it were available would help David's department to decide if the costs of extending the metro line pay off or not.

Currently, some social-scale systems such as Google or Facebook can know most of these details, but a public administration must protect the privacy of its citizens' data and information. One way to address the trade-off between access to personal information and users' control over the data they own and generate is to ask them for the required information and let them decide whether or not they want to share it, and with whom. Trust and security are two key issues in this context as well, as well as the analysis of gender inequalities.

Connecting the DAs to the city's urban digital twin would solve these problems, as the UDT would have more detailed information about service recipients when making predictions and managing exceptional situations, while users would be able to benefit from more personalized services, tailored to their particular preferences, and deciding the level of information they want to share with each service provider.

2.3.2 Scenario 2: Smart buildings

Ana, Bob and Cam are engineers working in the Ada Byron Lab, a smart building in the University of Malaga that has decided to use our technology. The building has sensors in its rooms that measure their temperature and ambient light, as well as actuators that control the air conditioning and lighting systems in each room. It is also possible to know whether the building parking slots are free or occupied.

Engineers working in that building use their DAs, which among other information record data about their daily habits and routines. For example, Ana and Bob tend to arrive early, and they like parking their cars always in the same slots. Cam is often late, and parks her car in a different place every day. On Mondays and Fridays they meet for weekly project briefings together with the rest of the department. Each person has a preference about the temperature of the room, and its illumination.

Current intelligent buildings regulate their environmental conditions to the presence of people in them, even including the elaboration of usage and occupancy patterns to improve their predictions on energy costs and building space occupation (offices, rooms, parking lots). However, we can go a step further if we consider individual details and preferences of the users of these buildings. For example, the building can adjust the AC and ambient light systems of each room to match the median (or the mode) of the individual preferences of each room's occupants. Knowing the individual habits of people, as reported by their DAs, the building can recommend changing some of the meeting times in order to avoid excessive occupancy of some of the rooms, or peaks of energy consumption. It is also easy to predict, based on individual user preferences, the energy needed by the AC systems to maintain the required temperature in all rooms for the rest of the year, and to change accordingly the building rule that calculates the optimal temperature in each room to reduce overall costs.

The building can also develop models of the collective behavior of its users more accurately when it considers the individual behavioral patterns provided by each DA. For example, it can take into account when Bob is on leave and will not use his office and parking slot, or when members of one department start staying later in the building (due, for example, to an upcoming project deadline) to adjust the system calculations

Single individuals can also benefit from an application that uses the smart services of a digital twin. In addition to being able to help decide the room temperature based on their individual preferences, they can be alerted before leaving home if there is no free parking space equipped with an electrical charging outlet, or recommend Cam to work from home next Tuesday because the building is expected to be busy and her DA knows she has no meetings scheduled that day. Similarly, Bob can enjoy a more comfortable working environment when the temperature and illumination of his office are according to his preferences but are also set in accordance with the overall policies of the building about energy savings and respect the cost limitations that the company budget allows.

2.3.3 Scenario 3: Outdoor activities

Today, the number of people exercising outdoors in the city (walking, running, cycling) has increased considerably. The planning of these activities involves the combination of information from different sources, including the conditions of the city itself (traffic density, pollution levels), the weather (air quality, weather conditions) and the citizens' personal choices (preferences about whether to run near the seaside or in the sport centre, or the fact that you are training for the forthcoming marathon).

This is another example of an application that would benefit if a UDT of the city's sports and outdoor activities could take into account the personal preferences, routines and behavioral patterns of its individual citizens. Both the city and the citizens would gain from this. The city could more accurately plan the use of its resources, anticipate future demands, monitor the current state of these aspects of the city and propose changes when necessary, being able to adapt to exceptional circumstances and to let citizens know about the changes. In turn, citizens would be able to benefit from personalized services, know in real-time the current state of the traffic, air quality and occupation of their target training place, and also influence, by letting the city know their personal opinions and preferences, the planning of the city and its future evolution regarding outdoor activities. The UDT could also suggest collective activities to citizens with similar interests and preferences, while the individual DAs could help their owners arrange training groups or team activities.

2.4 References

[Ang17] M. Angelidou. “The role of smart city characteristics in the plans of fifteen cities.” J. Urban Technol., 24(4), 3e28, 2017

[BCE+20] F. Bordeleau, B. Combemale, R. Eramo, M. van den Brand, and M. Wimmer, “Towards model-driven digital twin engineering: Current opportunities and future challenges,” in Proc. of ICSMM’20, Springer, 2020, pp. 43–54.

[BCW17] M. Bambrilla, J. Cabot, M. Wimmer. Model-Driven Software Engineering in Practice: Second Edition, Morgan & Claypool Publishers, 2017.

[BF72] A. W. Biermann and J. A. Feldman, “On the synthesis of finite-state machines from samples of their behavior,” IEEE Trans. Comput. 21(6):592–597, 1972.

[BMBV20] M. F. Bertoa, N. Moreno, L. Burgueño, A. Vallecillo. "Incorporating Measurement Uncertainty into OCL/UML Primitive Datatypes." Sosym, 19(5):1163–1189, 2020

[BMP+20] M. F. Bertoa, N. Moreno, A. Pérez-Vereda, D. Bandera, C. Canal, and J. Álvarez-Palomo. “Digital Avatars: Promoting Independent Living for Older Adults”. Wireless Communications and Mobile Computing, 2020.

[CMV20] J. Cámara, H. Muccini, K. Vaidhyanathan: “Quantitative Verification-Aided Machine Learning: A Tandem Approach for Architecting Self-Adaptive IoT Systems.” Proc. of ICSA’20: pp. 11-22, 2020.

[Coh12] Cohen, B. What exactly is a smart city? Co. Exist 19, 2012. https://www.fastcompany.com/1680538/what-exactly-is-a-smart-city

[COSCA] Digital Avatars: A Framework for Collaborative Social Computing Applications. PGC2018-094905-B-I00. 2019-2022. http://atenea.lcc.uma.es/projects/COSCA

[CPGS18] J. Cámara, W. Peng, D. Garlan, et al. “Reasoning about sensing uncertainty and its reduction in decision-making for self-adaptation.” SCP 167:51-69, 2018.

[D06] P. A. Dinda. “Design, implementation, and performance of an extensible toolkit for resource prediction in distributed systems.” IEEE Trans. Par. Distr. Systems, 17(2):160–173, 2006.

[DTC21] Digital Twin Consortium, “Glossary of digital twins,” 2021. https://www.digitaltwinconsortium.org/glossary/index.htm

[DUET21] Living in EU. Urban Digital Twin:The European way of digital transformation in cities and communities,2021 https://living-in.eu/groups/solutions/urban-digital-twin

[DWZ21] L. Deren, Y. Wenbo, S. Zhenfeng. “Smart city based on digital twins.” Computational Urban Science 1(4):1-11, 2021.

[DZS21] T. Deng, K. Zhang, Z. Shen, “A systematic review of a digital twin city: A new pattern of urban governance toward smart cities,” Journal of Management Science and Engineering 6(2):125-134, 2021.

[EU20] European Commission. “Digital Twins for cities” October 2020. https://europa.eu/regions-and-cities/programme/sessions/1359_en

[FI20] FIWARE Reference Arquitecture for Smart Cities, 2020. https://www.fiware.org/wp-content/uploads/Smart-Cities-Brochure-FIWARE.pdf

[FZYM21] C. Fan, C. Zhang, A. Yahja, and A. Mostafavi, “Disaster city digital twin: A vision for integrating artificial and human intelligence for disaster management,” Int. J. Inf. Management 56: 102049, Feb. 2021.

[GMB+14] J. Guillén, J. Miranda, J. Berrocal, J. García-Alonso, J. M. Murillo, C. Canal. “People as a Service: A Mobile-centric Model for Providing Collective Sociological Profiles.” IEEE Software 31(2):48-53, 2014.

[Gri14] M. Grieves, “Digital twin: Manufacturing excellence through virtual factory replication,” Florida Institute of Technology, White Paper, 2014.

[HA18] S. Haagand, R. Anderl, “Digitaltwin—proof of concept,” Manufacturing Letters, vol. 15, pp. 64–66, 2018.

[HS97] S. Hochreiter, and J. Schmidhuber. “Long short-term memory.” Neural computation, 9(8), 1735-1780, 1997.

[HQJ21] R. Hai, C. Quix, and M. Jarke, “Data lake concept and systems: a survey,” CoRR, vol. abs/2106.09592, 2021.

[J16] A. Jøsang. “Subjective Logic - A Formalism for Reasoning Under Uncertainty.” Springer, 2016.

[JM+20] N. Y. Jeppu, T. F. Melham, D. Kroening, and J. O'Leary, "Learning Concise Models from Long Execution Traces," Proc. of DAC’20, pages 1-6. IEEE 2020

[LD20] H. Lehner and L. Dorffner, "Digital geoTwin Vienna: Towards a digital twin city as Geodata Hub". Photogramm., Remote Sensing Geoinformat. Sci., 88:63-75, 2020.
[LZ21] L. Liu, Y. Zhang, "Smart environment design planning for smart city based on deep learning", Sustainable Energy Technologies and Assessments, Vol. 47, 2021.
[MBOV20] P. Muñoz, L. Burgueño, V. Ortiz, A. Vallecillo. "Extending OCL with Subjective Logic." Journal of Object Technology, 19(3), 3:1–15, 2020.
[MCC+14] C. Manville, Cochrane, G., Cave et al. Mapping Smart Cities in the EU. European Parliament, 2014. http://www.europarl.europa.eu.
[MCGS15] G.A. Moreno, J. Cámara, D. Garlan, B.R. Schmerl, et al. "Proactive self-adaptation under uncertainty", In ESEC/FSE 2015, pp. 1-12, 2015.
[Mes12] J. Meseguer, "Twenty years of rewriting logic". J. Log. Algebraic Methods Program. 81(7-8): 721-781, 2012.
[MML19] A. M. Madni, C. C. Madni, and S. D. Lucero, "Leveraging digital twin technology in model-based systems engineering," Systems, vol. 7, no. 1, p. 7, 2019.
[MPM+21] P. Muñoz, A. P. Vereda, N. Moreno, J. Troya, and A. Vallecillo. "Incorporating Trust into Collaborative Social Computing Applications." EDOC'21, IEEE, Oct. 2021.
[MTV21] P. Muñoz, J. Troya, A. Vallecillo. "Using UML and OCL Models to Realize High-Level Digital Twins. " In Proc. ModDiT'21@MoDELS'21, IEEE, Oct. 2021.
[NCD+18] S. Nastic, J. Cubo, M. Donato et al. "SMART-FI: Exploiting Open IoT Data from Smart Cities". Internet of Everything, pp. 153-173, Springer, 2018.
[PC17] A. Pérez-Vereda, C. Canal "A People-Oriented Paradigm for Smart Cities." ICWE 2017: 584-591, 2017.
[PCP20] A. Pérez-Vereda, C. Canal, and E. Pimentel. "Modelling digital avatars: A tuple space approach". Science of Computer Programming, 2020.
[RMV21] A. Ramírez, N. Moreno, A. Vallecillo. "Rule-based preprocessing for data stream mining using complex event processing." Expert Systems, 38:8, Dec 2021
[SEI21] SEI "Architecting the Future of Software Engineering: A National Agenda for Software Engineering R&D" Software Engineering Institute, Nov. 2021. https://resources.sei.cmu.edu/library/asset-view.cfm?assetID=741193
[SH20] G. Schrotter and C. Hürzeler, "The digital twin of the city of Zurich for urban planning," PFG, J. Photogramm., Remote Sens. Geoinf. Sci., pp. 1-14, 2020.
[TMBV21] J. Troya, N. Moreno, M. F. Bertoa, A. Vallecillo. "Uncertainty representation in software models: A survey." Sosym, 20(4):1183-1213, 2021
[TZLN18] F. Tao, M. Zhang, Y. Liu, and A. Nee, "Digital twin driven prognostics and health management for complex equipment," CIRP Annals, 67(1):169–172, 2018.
[TZLN19] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," IEEE Trans. Ind. Informatics, 15(4):2405–2415, 2019.
[WTD16] N. Walkinshaw, R. Taylor, and J. Derrick, "Inferring extended finite state machine models from software executions," ESE 21(3):811–853, 2016.

3 OBJETIVOS, METODOLOGÍA Y PLAN DE TRABAJO – *OBJECTIVES, METHODOLOGY AND WORK PLAN*

3.1 Overall goals and objectives

This project aims at producing contributions at different levels.

- At the **conceptual level**, the project will provide methods and tools for the development of smart city applications and UDTs that allow incorporating the information, habits and preferences of individuals, as well as the definition of the supporting software architecture and the set of common services that these UDTs will use. A new business model will be developed between the users and the smart city applications where the producers of the information will remain in control of it, and for the exchange of information, both parties (consumers and producers) will have to obtain benefits. Other scientific contributions of the project will be the definition of models for the specification and exchange of such information (both structural information, e.g., the user data; and behavioral information, i.e., the models and rules that define the user patterns of conduct); languages for querying and managing such information; and APIs for interacting with the DAs, so that their information can be effectively used from other DAs or from UDTs. In addition, a distributed pull-based data-lake model will be defined, in contrast to the existing centralized, push-based data lakes currently used by many DTs. Finally, reasoning with the information about individual users represents a new challenge for smart city applications. Smart services should be defined for the UDTs based on this information as part of the project, e.g., data visualization, prediction services, as well as self-adaptation policies and mechanisms.

- At the **platform level**, the project will contribute in several ways. First, to the extension of the models that current smart city frameworks (such as FIWARE or MOCA) use to store, manage and exchange the data, incorporating models about individual users' information, preferences, habits and behavioral patterns. Second, to the representation, management and propagation of uncertainty in the data handled by the applications and in the users' behavior. Subjective information such as trust, reputation, and preferences should be explicitly represented and managed too, using mechanisms such as trust networks [J16]. Third, to the definition and implementation of the appropriate mechanisms and tools for the exchange of information between the users' DAs and the smart city UDTs, using decentralized data lakes.
- At the **analysis level**, the project will provide methods and tools for the analysis of several aspects of the applications that are developed using this project's framework and libraries. In particular, we are interested in using qualitative and quantitative techniques for the analysis of the performance, reliability and privacy of DAs-UDTs applications. For this we will mainly rely on probabilistic and statistical model checking. Another interesting research topic will be the semi-automated synthesis of behavioral models of people from their traces, routines, and expected behavior. The prediction and analysis services defined at the conceptual level will be implemented and studied at this level.
- At the **engineering level**, the project will address the stringent requirements that the use of mobile phones imposes on the applications and on the technologies they can execute, in terms of memory, battery lifetime, network bandwidth, or communications reliability, among others.

Given that these goals are too wide to be tackled in general, in this project we will focus on certain application scenarios, which are representative systems of the kinds of applications our framework will support. These applications are the ones described before in Section 2.4.

The ***specific objectives*** derived from the above overall goals are the following:

[O1] To develop an architectural framework and a set of common components to support incorporating individuals in smart city applications using UDTs, including the internal architecture of the UDTs and skeleton implementations for them.

[O2] To define modeling languages for the specification of individual information, preferences, habits and behavioral patterns, and the appropriate mechanisms, processes and tools for the synthesis, storage, modification and exchange of these models between the UDTs and the users' DAs.

[O3] To provide mechanisms and tools for the representation, management and propagation of uncertainty by such smart city applications, including measurement imprecisions, vague or incomplete information, subjective opinions and trust, unpredictable human behavior, and unreliable communication networks and data sources.

[O4] To develop methods and tools for the analysis of the performance, privacy and reliability of such smart city applications.

[O5] To define smart services for the visualization of the UDT information, recognition of behavioral patterns of citizens and predictions of potential situations of interest, as well as the implementation of corrective measures for realizing the required self-adaptation mechanisms.

[O6] To develop exemplar applications using the project that can serve to showcase the proposal and validate its results.

3.2 Methodology

Our project will develop an architectural framework to provide all the necessary infrastructure and common services to design, develop, analyze and deploy the applications targeted by the project, according to the project objectives mentioned above. The project work plan, schedule and methodology are described in this section.

3.2.1 Project Workplan

The project plan has been broken down into five technical work packages (WP1-WP5), plus one more (WP0) dedicated to project management, and another devoted to dissemination and exploitation (WP6). The description of these work packages follows, including goals, tasks, responsible persons, and deliverables for each of them.

WP 0: Project Management

Goals: This WP is responsible for coordinating the joint efforts of the researchers during the execution of the project, ensuring the smooth progress of the work plan by monitoring its progress towards milestones and deliverables. Periodical monitoring of each WP task advance towards the objectives will be carried out in the context of this WP. Also, risk and quality assessment, and the interim and the preparation of final project reports belong to this WP.

Tasks

- Task 0.1. Project progress monitoring: This task will be devoted to ensuring that the project progresses adequately, following the established methodology, and applying corrective measures in the case of issues not considered in the project plan.

WP Leader: A. Vallecillo, C. Canal (Project PIs)

Deliverables:

- D0.1 Interim activity report (M12)
- D0.2 Interim activity report (M24)
- D0.3 Interim activity report (M36)
- D0.3 Final project activity and management report (M48)

WP 1: Architectural framework

Goals: This WP is aligned with project objective O1, and will be devoted to developing an architectural framework and a set of common services to support the effective development of the type of smart city applications targeted by the project, including the internal architecture of the UDTs and skeleton implementations for them.

Tasks

- Task 1.1. Core concepts definition and requirement analysis: This task will be aimed at defining the features that UDTs (both at the structure and at the behavioral level) have to exhibit in order to fulfill the requirements obtained from the scenarios considered in WP5.
- Task 1.2. Implementation of the reference architecture and basic infrastructure of the UDTs that fulfill the previous requirements. We will rely on existing DT reference architectures, e.g., [FI20, DZS21, MTV21, SH20, TZLN19], adapting them to implement a decentralized and distributed data lake that uses a pull model to collect the information from users.
- Task 1.3. Prototypical implementation of basic services: An implementation will be provided in order to effectively use a minimum functionality to deal with these UDTs.

WP Leader: E. Pimentel

Deliverables:

- D1.1 Architectural description of the UDT (M12)
- D1.2 Architectural framework specification and prototype implementation (M24)
- D1.3 Prototypical implementation of a basic set of services (M36)
- D1.4 Testing and evaluation of the framework and the basic services (M48).

WP 2: Models and modeling languages

Goals: This WP will design and develop the appropriate models and modeling languages for the specification of the UDTs and, especially, the information about individuals and their behavioral patterns, as well as its synthesis, storage, modification and exchange between the UDT and the users' DAs. For this, the group has experience in the definition of modeling languages. This WP is aligned with objectives O2 and O3.

Tasks

- Task 2.1. Information modeling: This task will be devoted to the design and development of the appropriate models and modeling languages for the specification of individual information, preferences, habits and behavioral patterns. We will study the suitability of various MDE tools for constructing the languages, such as Xtext or JetBrains MPS.
- Task 2.2. Interaction between the UDT and the DAs: This task focuses on the analysis and design of the appropriate mechanisms, processes and tools for the synthesis, storage,

modification and exchange of these models between the UDT and the users' DAs, using a pull strategy.
- Task 2.3. Data management: implementation of the mechanisms defined in the previous task, in order to ensure the appropriate scalability, performance, privacy and reliability in contexts of reduced memory and low computing and network resources. We will study mechanisms to reduce memory consumption and execution times.
- Task 2.4 Uncertainty representation and management: This task is devoted to implementing the mechanisms and tools for the representation, management and propagation of both measurement and belief uncertainty in all the models and modeling languages defined in the tasks above.

WP Leader: J. Troya

Deliverables:
- D2.1 Information models for storing and managing the data about users by UDTs (M12)
- D2.2 Exchange models and mechanisms between the UDT and the DAs (M12)
- D2.3 Synthesis and analysis of models (M24)
- D2.4 Libraries for modeling and implementing uncertainty in UDTs (M12)

WP 3: Analyses

Goals: This WP will be in charge of defining the analyses that will be supported by our proposal, as well as the appropriate methods and tools for performing them. Given specifications of the architecture of a UDT application and of its components (including services offered, resources used, and interactions), a deployment plan, and models of the devices used (including smartphones and cloud/fog servers) and the provided infrastructure, our tools will allow the study of its performance, reliability and privacy. These properties are different in nature and will require different information in the models for the analysis. We will adapt existing tools and methods on which the project team has extensive experience for the development of the tools and methods tailored to the infrastructure and nature of smart city applications built using the planned infrastructure. This WP is aligned with objective O4.

Tasks
- Task 3.1. Performance: We will use the statistical model checkers VeStA and the one provided as part of PRISM for the predictive analysis of the smart city applications we will develop with this project. We will also consider the use of PRISM for analyzing self-adaptive systems, as well as Simulizar, a Palladio plug-in, and Storm.
- Task 3.2 Reliability: The same Maude specifications of the smart city applications, the infrastructure and the devices will allow us to perform the analysis of their reliability using the wide range of tools available in the Maude formal environment, which includes tools for reachability analysis and model checking. Simulation-based techniques and statistical model checking seem also appropriate in the context of uncertainty.
- Task 3.3. Privacy: In the context of UDTs, the verification of security properties is needed. Specifically, we plan to mechanically prove privacy guarantees on the developed infrastructure, and guarantees on the use of people's information as granted by the DAs through this infrastructure.

Task Leader: F. Durán

Deliverables:
- D3.1 Formal specification of the infrastructure and APIs (M12)
- D3.2 Methods and tools for the analysis of performance and reliability (M24)
- D3.3 Report on the privacy guarantees of the infrastructure (M36)

WP 4: Smart services

Goals: This WP will define and implement the methods and framework required to build the smart services for the UDTs developed for each application. Such application-dependent services include the visualization of UDT information, recognition of behavioral patterns of citizens and prediction of potential situations of interest, what-if scenario analysis, as well as the implementation of corrective measures and self-adaptation mechanisms. The analysis of possible gender bias and inequalities by the smart services will also be considered, verifying that our algorithms and results do not discriminate citizens based on their sex or gender, nor are they gender-biased. This WP is aligned with objective O5, and includes the integration of existing visualization and ML components into the UDTs.

Tasks

- Task 4.1. Behavioral pattern analysis and forecasting: The purpose of this task is to develop a library of reusable mechanisms to detect and forecast trends observed in the system. To do that, we will adapt and extend well-known frameworks used in our prior work [MCGS15, CMV20] that employ time series prediction and Long Short-Term Memory (LSTM) networks. The results of this task will serve as the basis for tasks 4.2 and 4.3.
- Task 4.2. Dashboard framework development: The purpose of this task is to develop a customizable framework for the development of dashboards for smart services that will facilitate situational awareness through the: (i) assessment of the current situation as reflected in the models of the UDT and visualized by the dashboard widgets (similar to those provided by platforms like MOCA and FIWARE), (ii) projection of how the situation is likely to evolve and what-if scenario analysis. These functionalities will be informed by the analysis techniques developed in Task 3.1, as well as by the forecasting capabilities resulting from Task 4.1 (in ii).
- Task 4.3. Policy generation and self-adaptation methods: This task will focus on methods to develop automated self-adaptation mechanisms that will proactively adapt aspects of the system to mitigate the effects of potential undesirable situations that might appear in the short term (according to the forecasting and analysis components). To do that, we will extend and adapt to the context of smart cities our work on proactive self-adaptation that has been applied to complex IT [MCGS15] and IoT [CMV20] systems.

Task Leader: J. Cámara

Deliverables:

- D4.1 Library of reusable forecasting mechanisms (M12).
- D4.2 Customizable framework for dashboard development (M24).
- D4.3 Methods and tools for developing proactive self-adaptation mechanisms (M36).

WP 5: Applications and Validation

Goals: In this WP we will validate the whole project approach, including methodology, tools and services by developing the three application scenarios, corresponding to objective O6. In this way, we will demonstrate the feasibility and soundness of our proposal, as well as the results of the project. This WP will serve also as a proof of concept showing how the proposal can be successfully applied to the development of some smart city applications.

Tasks

- Task 5.1. Requirement analysis: The purpose of this task is to gather and specify the requirements of the three scenarios described in Section 2.4. This task will be performed in close collaboration with Tasks 1.1 and 1.2, providing and getting feedback from them. Additionally, we will explore different technological alternatives that make it possible to build applications like those described in the scenarios.
- Task 5.2. Application design and development: This task will take advantage of the ongoing results of WPs 1, 2, and 3, tailoring them for developing the three application scenarios. It includes the definition and implementation of particular services and behavioral models of the users for each of the scenarios in the context of the UDTs, as well as defining specific interfaces both to the user and to other systems.
- Task 5.3. Deployment: In this task we will deploy our three applications and test them with controlled users, to validate and assess the feasibility of the proposal. This task will iteratively deploy, test, and refine the applications until stable versions are achieved.
- Task 5.4. Evaluation: Each scenario focuses on different aspects of the proposal: transportation, mobility and health, smart buildings. Hence, evaluating the implementation of the scenarios will serve both as a validation of the ongoing results of the different technical tasks and WPs and as a proof of concept of the whole approach.

WP Leader: C. Canal

Deliverables:

- D5.1 Requirement specification of the application scenarios (M9)
- D5.2 Prototype version of each application scenario (M18)
- D5.3 First stable versions of each application scenario (M30)
- D5.4 Refined and tested version of the application scenarios (M36, M42, M48)
- D5.5 Evaluation of the results of new opportunities of research and development (M48)

WP 6: Dissemination and Exploitation

Goals: Although this is a research-oriented project, it is expected that its results will be tested in real environments, and also have an impact on standards. Some of our industrial partners (Section 3.3) have expressed interest in the project and will help with the non-academic dissemination inside their own companies and in possible projects in collaboration with our group. Technology transfer will be an important objective of the project. The project will follow Open Access policies for all produced artifacts, results and publications.

Tasks

- Task 6.1. Academic Dissemination: The basic form of academic dissemination will be through high-quality journal and conference papers. Members of the project group have a previous record of publishing in top-notch journals. We also plan to organize international workshops about the topics of the project, using the relationship of this project with other projects and initiatives in which our group is also engaged. These events will allow us not only to disseminate the results of the project but also to collect and integrate information from external sources. The main achievements of the project and its progress will be documented in a dedicated project webpage.
- Task 6.2. Exploitation and relationship with industry: This task will be devoted to making the results of the project visible to the industry, and carried out in cooperation with some industrial partners. We also plan to patent some of the results of the project, given its expected novelty and applicability.
- Task 6.3. Standardization activities: Our group actively participates in different international standards organizations (such as ISO, UNE and OMG). It is important that the project efforts and results are fully aligned with existing and forthcoming international standards, and also that our results and findings can influence and contribute to new standards or the revision of existing ones. Participation and attendance to standardization meetings and venues are also an integral part of our dissemination plan for this project.

WP Leader: A. Vallecillo

Deliverables:

- D6.1 Project Webpage (M06)
- D6.2 Final Exploitation Report (M48)
- D6.3 One international patent or participation in one international standard (M48)

3.2.2 Project Chronogram

The proposed plan spans for 4 years. The chart in Fig. 2 displays the expected schedule and duration for each WP and task, as well as the project participants that will work on each one.

3.3 Personnel

The research team is composed of seven researchers who exhibit significant scientific productivity and have a recognized activity in different fields directly related to the proposal: Javier Cámara (JC), Javier Troya (JT), Carlos Canal (CC), Francisco Durán (FD), Nathalie Moreno (NM), Ernesto Pimentel (EP), and Antonio Vallecillo (AV). The first two joined the group recently, opening some research lines where they have previous experience. The rest are members of the COSCA project and have shown their ability to work together, combine their expertise, and obtain relevant results.

This project proposal includes one person in the working team, Paula Muñoz (PM), plus two contracted engineers (C1 and C2). Paula holds the FPI grant associated with project COSCA. The proposal is very intensive in the design and implementation of software. The development of the architectural framework and the common services; the design of the analysis tools; the implementation and testing of the digital twins for the case study applications, as well as their evaluation, will require a significant amount of effort that cannot be achieved without extra personnel. These activities require highly qualified staff, and preferably with experience in similar or related matters. This is why we have included in the project working plan two engineers (identified as C1 and C2 in the chart shown in Section 3.5) with similar profiles: Software Engineers, with previous experience in one (or more) of the following topics: MBSE, real-time distributed computing, open data, and mobile applications. The plan is to hire them for 40 and 30 months, respectively, when the development efforts start. The expected tasks and activities of these engineers within the scope of the project, as well as those of the rest of the participants, are those indicated in the chronogram in Fig. 2.

		Year 1				Year 2				Year 3				Year 4			
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
WP 0: Project Management	AV, CC																
T0.1. Project progress monitoring	all				D01				D02				D03				D04
WP 1: Architectural Framework	EP																
T1.1 Definition and requirement analysis	EP, CC				D11												
T1.2 APIs to access and manipulate DAs	EP, C1								D12								
T1.3 Prototype implementation of basic services	C1												D13				
T1.4 Testing and evaluation	C1																D14
WP 2: Modeling languages	JT																
T2.1 Information modeling	JT, PM				D21												
T2.2 Interacting between UDTs and DAs	JT, PM				D22												
T2.3 Data management	JT, PM								D23								
T2.4 Uncertainty representation and management	JT, PM				D24												
WP 3. Analyses	FD																
T3.1 Formalization of the proposal	FD				D31												
T3.2 Performance	FD, C2																
T3.3 Reliability	FD, C2																
T3.4 Privacy	FD, C2												D33				
WP 4: Smart services	JC																
T4.1 Behavioral patterns and forecasting	JC, NM				D41												
T4.2 Customizable framework for dashboards	JC, NM								D42								
T4.3 Self-adaptation mechanisms	JC, NM												D43				
WP 5: Applications and Validation	CC																
T5.1 Requirement Analysis	CC			D51	D51												
T5.2 Application design and development	CC,							D52			D53		D54		D54		D54
T5.3 Deployment	CC, all										D53		D54		D54		D52
T5.4 Evaluation	CC, all										D53		D54		D54		D55
WP 6: Dissemination and Exploitation	AV																
T6.1 Academic Dissemination	AV, all		D61														D62
T6.2 Exploitation and relationship with industry	EP, all																D62
T6.3 Standardization activities	AV, all																D63

* LEGEND: Deliverables shown in chart (see Section "Project Workplan"). Participants are named by their initials. C1 and C2 refer to the 2 contracted persons.

***Fig. 2**. Chronogram of the project.*

3.4 Labs and existing infrastructure

The group has a research laboratory in the premises of the ETSI Informática at the University of Malaga, and another laboratory in the premises of the recently inaugurated ITIS software institute. These laboratories would perfectly serve to host the new personnel that we are requesting in this proposal (2 contracts, 1 FPI). Most of the material required in this project is already in place, from previous and ongoing projects. However, to carry on the proposed tasks we will need to buy the new hardware and software licenses contemplated in the budget.

3.5 Methodological approach

The project will follow an iterative and incremental approach, where the WPs will overlap to allow for feedback and adaptation to the evolution of the project and the state of the art. We will use a fixed time span of six months for each iteration. A set of requirements will be assigned to the iteration following a risk-oriented approach. Each iteration will have the following phases:

- Risk analysis and requirement selection. Before starting an iteration, the project scope is analyzed in order to select those work areas where the focus should be put in that iteration.
- Iteration Planning. A detailed work plan for the time span will be developed, including only those work areas that can be developed during the time span.
- Development. A conventional analysis-design-implementation cycle will be followed during this phase. In the first iterations, the analysis and design phases will have greater importance, since the developments will be oriented to clarify requirements and to build proof-of-concept components.
- Validation. At the end of each iteration, a validation phase will be carried out. Validation activities will run in parallel to the development in such a way that possible problems can be discovered as soon as possible.

3.6 Coordination meetings and problem resolution

Given the physical proximity of the team members, there will be in continuous contact. In any case, and for coordination purposes, periodic follow-up meetings will be scheduled every three months. In these meetings, each WP leader will report on the status and level of achievements of his objectives, and corrective actions will be agreed upon in case of potential delays.

Generally, all issues and problems found will be dealt with at the WP level first, and then reported at the periodical coordination meetings if solved. WP leaders will also be in charge of coordinating with other WP leaders in case some problems are detected that may affect other WPs. The IP will always be informed of these issues, not only for coordination and project management purposes but also for recording all detected problems.

3.7 Relationship with industry and institutions

There are several companies and institutions that have expressed their interest in the results of the project. Although their interests vary depending on their domain, we have strong relationships with the following companies from past and ongoing projects:

- The **Malaga City Hall** is actively working on implementing the Smart Cities initiative, including different kinds of services and applications: https://malagasmart.malaga.eu/. They have also developed a portal with open data with useful information to Malaga citizens (https://datosabiertos.malaga.eu/), and are currently promoting the development of social computing applications that use that open data. They are also part of the FIWARE initiative. We will be working in close cooperation with them during the whole project.
- The **Malaga Area Metropolitan Transport Consortium** (https://ctmam.es/) is responsible for all public transportations in the province of Malaga, and in particular for the metro and bus lines of the city of Malaga. CTMAM is also part of the MalagaSmart initiative, working with the Malaga City Hall to digitize all its services and the relationship with citizens.
- **Telefonica** is one of the major Spanish Telecom companies and leads the FIWARE initiative in Spain. We have collaborated with Telefonica in the past, and they are interested in the use of FIWARE in this project.
- **ATOS Spain** is a global leader in digital transformation, and one of the founders of the FIWARE initiative. Our research group has extensive cooperation with this company, precisely on the usage of FIWARE to exploit open data in smart cities (SMART-FI).
- Software for Critical Systems S.L. (**SOFTCRITS**) is a technological start-up company from the University of Malaga with wide experience in sensor networks and critical infrastructure monitoring. They have installed and manage the smart elements of the building where the ITIS Software Institute (https://itis.uma.es), to which our research groups belong, is located. They have been our previous partners in several projects, and have shown their interest in participating with us in this project for improving the smartness of the ITIS building.

3.8 Risks and Contingency plans

The development of the project will follow a risk-oriented methodology whenever it is possible. Thus, the most critical activities have been planned in the initial stages of the process, in such a way that the appropriate measures can be taken as soon as possible if problems are found. Anyway, the following issues have been identified as primary risks for the project:

1 *The rapid evolution of underlying technologies*. Smart cities and digital twin technologies rapidly evolve. New systems, standards and tools will probably emerge during the next few years and they can affect the methodologies and tools developed in this project. The implication of some of the members of the project team in international standardization organizations and the close relationship of the group with leading industries in the area will try to minimize these effects.

2 *Inability to define models to represent and predict behavioral patterns.* The project incorporates leading-edge theoretical research, whose absolute success cannot be predicted. Although we strongly believe that it is entirely feasible to find abstract languages and models for the kind of concepts and mechanisms that we want to express, the intrinsic nature of the research objectives makes it impossible to assure that such languages and models can be found in the context of the project. If this happens, the results of the project will be adequately reported to the scientific community so that the community can benefit from the project. In order to reduce risk, we will try to reuse and adapt existing languages in related domains, trying to avoid creating completely new languages from scratch.

3 *Lack of tooling and technologies.* We may develop our own tools or extend already existing ones, which may represent a heavy overload for the project, due to the effort required to build tools. We expect to be able to reuse some of the existing tools and technologies, extending them accordingly, but the lack of maturity of this field represents a challenge that cannot be neglected. The experience of our group in building tools is a valuable asset that we expect can mitigate this risk.

4 *Difficulties for finding suitable candidates for the contracted personal positions*. We already have candidates for the two positions we need in the project. They are recent graduate students who have started collaborating with the group and would like to work on this project. Regardless of whether they are finally hired (people with better resumes may

apply), these two candidates will ensure that no position is left vacant if the project is approved and enough resources provided.

3.9 Integration of gender analysis in research

Gender inequalities are crucial to understand and take into account the different interests, needs, behaviors, roles, or stereotypes of women and men in terms of their access to resources, power, positions, activities, etc. People are essential elements of this project, so we will ensure that in all research methods and the smart services we develop, we explicitly consider the analysis of gender differences and similarities, treat men and women equally, the source data for our algorithms are not biased towards either women or men, and the results do not discriminate against citizens based on their sex or gender.

3.10 Research group's previous experience

The members of our group have been successfully collaborating together for the last years, joining our knowledge and experience in the design and development of distributed applications, service-oriented computing, MBSE, and formal methods. This combination has proved to be essential for enriching the overall group's expertise and achievements, and has leveraged the potential impact of the results and the international visibility of the group. The research team is a joint effort by two subgroups (SCENIC and ATENEA) of the much broader ITIS Software Institute (https://itis.uma.es/).

Previous research activities of our group are closely related to those proposed in this project. In fact, this project builds on our results from previous ones, including SEACLOUDS, on service-oriented computing; Smart-FI, on smart cities and open data; EU COST Action MPM4CPS, on multi-paradigm modeling of cyber-physical systems; POLYCIMS (TIN2014-52034-R), which introduced the use of CEP technologies for monitoring and analyzing streams of data coming from critical systems, and the representation and management of uncertainty; and MBT4CPS (P20-00067-FR) for model-based testing of cyber-physical systems.

For illustration purposes, our group produces, per year, an average of 8 (JCR) journal publications, 5 international conference papers, 4 workshops papers, and 3 national conference papers, all related to the topics of the project. We aim at being very active both at national and international levels. At the international level, the group maintains an active presence in conferences and journals, with participation in the editorial boards of journals (SoSym, JOT, STIOT, Computing, Services, Trans. on Internet and Inf. Systems), participation in the program committees of conferences, organization of conferences (e.g. MODELS 2013, ESOCC 2013, ICSOC 2017, ECOOP 2017, CEDI 20/21) and workshops, edition of special issues at various journals, etc. The complete list of publications and projects can be obtained from the individual CVs and web pages of the project participants.

Another strong value of our group is the collaboration with other research groups, both at national and international levels, as demonstrated by the authorship of our joint publications and by our international projects. Our group also works in different standardization bodies, namely ISO/IEC, UNE and OMG. With ISO/IEC and ITU-T we have worked on the Reference Model for Open-Distributed Processing (RM-ODP) framework (http://rm-odp.net/), on the representation using models of the framework languages, being editors of several international standards. In OMG we are part of the submission team on the new OMG standard on "Precise Semantics for Uncertainty Modeling" (PSUM, http://doc.omg.org/ad/2017-12-1). This project will be aligned with all these initiatives and standardization efforts, ensuring the corresponding technology transfer, dissemination, and internationalization activities.

Similarly, we have a strong connection with industry and other institutions, as demonstrated by our continuous collaboration with them. In particular, we have recently been collaborating with the Malaga City Hall in the development of mobile applications that make use of their open data portal, employing CEP techniques to analyze the data they provide and to extract useful information from it. As mentioned above, some of our industrial partners are interested in this project and are happy to participate in it by providing the case studies that will be used for demonstrating our proposal and results. We expect this project to make significant advances in our cooperation with them.

4 IMPACTO CIENTÍFICO-TÉCNICO – *SCIENTIFIC-TECHNICAL IMPACT*

The results of the project are oriented to two main types of impacts. First, at the research level, it is expected to produce significant advances in the state-of-the-art and state-of-the-practice in the domains of Digital Twins, Smart Cities and Software Engineering, in particular in the development of UDT systems for smart cities. Dissemination of the project activities will aim at promoting and publicizing the work achieved and the results obtained, with a direct influence on the final impact of the project. The basic form of academic dissemination will be through high-quality international journals and conference papers. Both national and international workshops about the topics of the project will also be organized, not only to publicize the project results and improve its visibility, but also to discuss and exchange information and experiences with other research groups working on related topics and projects, and to start collaborations that enable the development of joint European projects. In this respect, members of this project proposal will participate in the forthcoming Daghstul seminar on Model-Based Engineering of Digital Twins, to be held in September 2022. This shows the strong connection between the project members and the international community working on these topics.

Second, and more importantly, one of the most valuable contributions of the project is expected in terms of the definition of a new paradigm in the scope of smart cities, where the exploitation of people's information can be owned and controlled by the people themselves, instead of being used or commercialized by third parties — without neither transparent control nor clear benefits to the data owners.

Third, the project results will be fully aligned with existing and forthcoming international standards, and our results can influence and contribute to new standards or the revision of existing ones. Thus, we will coordinate our efforts with those currently in place within ISO/IEC and the OMG on Digital Twins and on Uncertainty Modeling, respectively.

At the national level, this project is also aligned with several of the 23 strategic lines defined by the Agencia Estatal de Investigación (AEI) in 2021, namely "Gemelos Digitales: modelización y diseño", "Estrategias inteligentes de movilidad urbana y metropolitana", or "Explotación y modelado de la complejidad en escenarios de previsión de riesgos."

And at the international level, the recent US report about "Architecting the Future of Software Engineering: A National Agenda for Software Engineering Research & Development" [SEI21] mentions "Engineering Societal-Scale Systems" as one of the three Advanced Architectural Paradigms main research focus. This is precisely the research area that our proposal targets. Similarly, "Digital Twins for cities" was included in October 2020 as a key driver for the European Commission [EU20]. These results will also give the group a perfect opportunity to be more competitive and visible at the international level.

Finally, to ensure open access to all project outcomes, as well as the reproducibility of the results, all produced artifacts, results and publications will be made openly available through the institutional repository of our University, the project's web page, and git repositories.

5 IMPACTO SOCIAL Y ECONÓMICO – *SOCIAL AND ECONOMIC IMPACT*

At the local level, we expect that our results become part of the Malaga City Hall initiative that aims at making Malaga a Smart city: https://malagasmart.malaga.eu/ as they are aligned with the strategic plan for the digitization of the city, its services and its relations with the citizens: https://malagasmart.malaga.eu/opencms/export/sites/msmart/.content/galerias/documentos/Plan_Estrategico_de_Innovacion_bajares.pdf

At the European level, the Intelligent Cities Challenge (ICC) is a European Commission initiative that gathers 136 cities to achieve intelligent, socially responsible and sustainable growth through advanced technologies. Among other goals, ICC aims to foster citizen participation and digitization of public administration, and also improve logistics and economics of mobility: https://www.intelligentcitieschallenge.eu/about-intelligent-cities-challenge. Our project is fully aligned with this initiative, and we expect our results to have an impact on it.

Finally, our collaboration with our partners on this project is expected to have a relevant impact on their applications within the next 3-4 years. We also expect some collaborations in H2020 EU projects based on the results of this project, particularly with Softcrits, since they have recently started working on digital twin technologies, and the results of our project, about incorporating humans in the loop within the digital twins' context, are of interest to them.

6 CAPACIDAD FORMATIVA – *TRAINING CAPACITY*

The research-oriented nature of the group provides the natural context for the development and supervision of PhD thesis. The list of recent PhD thesis supervised by the group members follows (including two theses to be imminently defended in early 2022):

1. Patrícia Araújo de Oliveira "A Flexible and Procedural Requirements-Driven Approach for Building Architecture Projects of Self-Adaptive Systems" ~Early 2022. Dir: E. Pimentel, F. Durán – Currently in Academia (Universidade Federal do Amapá).
2. Alejandro Rodríguez Tena "A multilevel modeling infrastructure for the definition, execution, and composition of domain-specific modeling languages" 20 January 2022 (Dir: A. Rutle, L. M. Kristensen, F. Durán).
3. José Carrasco. "Trans-Cloud: CAMP-TOSCA-based management of cloud applications" June 2021 (Dir. E. Pimentel, F. Durán). Currently in industry (The Workshop).
4. Gala Barquero Moreno, "Processing Structured Data Streams" Feb 2021 (Dir. A. Vallecillo, J. Troya) – Currently in industry (Zoconet S.L.).
5. Loli Burgueño. "On the properties of Model Transformations: Correctness and Performance". 2016 (Dir: A. Vallecillo, M. Wimmer) – Currently in academia (UOC)
6. Javier Troya. "On the Model-Driven Performance and Reliability Analysis of Dynamic Systems" 2013. (Dir: A. Vallecillo) – Currently in academia (U. Malaga)
7. Nathalie Moreno "WEI: Integración de aplicaciones externas en sistemas Web mediante modelos" 2012. (Dir: A. Vallecillo) – Currently in academia (U. Malaga)
8. José A. Martín Baena. "Secure Adaptation of Software Services" 2012 (Dir: E. Pimentel) – Currently in industry (VirusTotal, Google)

Members of our group actively participate in the Malaga University PhD programs, both in the past and now in the PhD Program of the University on "Tecnologías Informáticas." The following PhD thesis are currently in progress in our group:

- Alejandro Pérez-Veredas "New models for social computing" Dir: R. Hervás, C. Canal (defense expected in late 2022).
- Paula Muñoz. "Modelado y verificación de gemelos digitales" Dir: J. Troya, A. Vallecillo (defense expected in early 2023).

The present project proposal opens up new research lines that complement those previously tackled by the group, and require dedicated personnel. We believe, based on our previous records, that our group is able to successfully train one PhD and produce high-quality professionals (not only for academia but also for industrial positions).

Research plan, method and activities. The research plan for the requested PhD candidate assumes a three-year activity program, with a possible extension to four years if necessary. Design Science will be followed as a research methodology since our experience has shown that it provides an adequate scientific and systematic approach for conducting research in the areas of Software Engineering.

The research problem to be addressed is related to the analysis of emergent properties of complex applications that involve many different participating agents (humans, physical devices and communication networks) with uncertain environmental conditions. The applications designed and developed in this project are of that nature, and hence its suitability and the opportunity to develop a PhD within this project.

Concerning the research plan for the candidate, apart from the inherent research activities of any PhD, other activities to be performed by the candidate include: training courses both in general research subjects (research methods, scientific writing skills, etc.) and in specific topics related to her research theme (formal methods, analysis tools); short stays in research centers to collaborate with other researchers; and at least one longer stay (4 months) to work in a topic related to the thesis in another lab or a top-notch research center, probably with one of our current international contacts.

7 CONDICIONES ESPECÍFICAS PARA LA EJECUCIÓN DE DETERMINADOS PROYECTOS – *SPECIFIC CONDITIONS FOR THE EXECUTION OF CERTAIN PROJECTS*

N/A