

DIORAMA: Digital twin FOR sustAinable territorial Management

Andrea Bianchi, Giordano d'Aloisio, Andrea D'Angelo, Antinisca Di Marco, Alessandro Di Matteo, Jessica Leone, Giulia Scoccia, Giovanni Stilo and Luca Traini

Department of Engineering and Information Sciences and Mathematics, University of L'Aquila, Italy

Abstract

Territorial Management is a challenging task that goes beyond the concept of smart city mainly because it involves decisions taking into consideration several aspects including the urban planning, risks assessment, disaster recovery, as well as health, social and economical management of a specific area. Territorial Management and related decisions are still little supported by software and IT systems even if a lot of information and data are already available, as stored in institutional databases or released as open data. In this paper, we present DIORAMA, a digital twin for sustainable territorial management, that leveraging on the results of Territori Aperti and SoBigData projects, aims to provide citizens, associations and institutions a system that supports decision-makers to take sustainable decisions in the field of territorial management guaranteeing citizens participation and operating under the FAIR and Open Science principles.

Keywords

Territorial Management, Digital Twin, Sustainability

1. Introduction

Territorial management is always challenging, given the several domains involved, sometimes in contrast to each other (e.g., health, economy, regulations, urban planning, risk assessment and reduction, etc.). Managing a territory after a disaster is even more complex due to the damages suffered by public/private buildings and infrastructures like bridges, roads, gas pipeline, sewerage, water pipeline and electricity network [1]. Effective management of resources (such as funding, time, data, software, and so on) is required to improve the efficiency of operations. In addition, the population's wellness must be considered to increase the effectiveness of the management. The combination of effective management of resources, quick decisions and wellness of the population comprises our definition of *sustainability*. We argue that the management of the territory, especially under disaster recovery, must be *sustainable*. However, decision-makers and public authorities often lack a comprehensive view of the territory and a

ITADATA2022: The 1st Italian Conference on Big Data and Data Science, September 20–21, 2022, Milan, Italy

✉ andrea.bianchi@graduate.univaq.it (A. Bianchi); giordano.daloisio@graduate.univaq.it (G. d'Aloisio); andrea.dangelo6@student.univaq.it (A. D'Angelo); antinisca.dimarco@univaq.it (A. D. Marco); alessandro.dimatteo1@student.univaq.it (A. D. Matteo); jessica.leone@student.univaq.it (J. Leone); giulia.scoccia@student.univaq.it (G. Scoccia); giovanni.stilo@univaq.it (G. Stilo); luca.traini@univaq.it (L. Traini)
🆔 0000-0002-6046-4355 (A. Bianchi); 0000-0001-7388-890X (G. d'Aloisio); 0000-0002-0577-2494 (A. D'Angelo); 0000-0001-7214-9945 (A. D. Marco); 0000-0002-2092-0213 (G. Stilo); 0000-0003-3676-0645 (L. Traini)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

data science tool is required to perform sustainable, efficient and effective management.

The availability of big data coming from heterogeneous sources (such as open data repositories, geographic information systems, smartphones, and social media) enables a complete digital representation of the territory covering each aspect of its management, including disaster recovery. In this paper, we want to make a step ahead and present DIORAMA, a *Digital twin fOR sustAinable territorial MAnagement* [2] we are developing under the Territori Aperti project. Territori Aperti is a documentation, training, and research center for sustainable territorial management with a particular focus on disaster recovery [3]. DIORAMA represents the core of Territori Aperti since it leverages on all the work and research we are performing in the project. As it happens for Territori Aperti, DIORAMA also leverages on SoBigData++ to implements the open science principles.

Through a digital twin (DT), it is possible to model any type of physical entity [4], therefore DTs are increasingly used to model entities of any domains, such as in medical, corporate and manufacturing domains. In our particular case, we propose to model through the DT concept the full sustainable territorial management. Papers such as [5], discuss the principles underlying the construction of DT and give references to the main applications of DT to territories. It highlights how, falling the topic within themes such as smart city development and sustainable urban development, most of DT's creations for the territory are confined to the city level [6, 7]. There are also several reviews of city-level DT construction types and applications of this type [8, 9]. In general, there are several projects that aim to develop DT at urban level, at different levels of abstraction [10, 11]. We recall that in the development of DT for more general territorial management, we must consider different and sometimes opposite goals. For this reason, since several representations of a territory are already available (satellite images, GIS systems, etc), we have to rely on the integration of data from different sources and involve different stakeholders from different domains. All the previously cited papers develop DT to model processes, conditions and other various concepts related to territorial physical entities. None of them is able to model the entire sustainable territorial management, considering also non-physical concepts and by developing applications and analysis useful for users in territorial management. With DIORAMA we aim to overcome this lack, by proposing DT for the sustainable territorial management.

The main contributions of this paper are: *i*) to present the DIORAMA digital twin and its high-level architecture; *ii*) to describe how Territori Aperti contributes to DIORAMA by describing how its achievements will be embedded in DIORAMA.

The paper proceeds as follows: section 2 describes more in detail the Territori Aperti project and DIORAMA. Sections 3, 4, and 5 are then respectively related to in-depth descriptions of the analysis, visualization, and application projects we are conducting in Territori Aperti and which comprises DIORAMA. Finally, section 6 describes some future works and concludes the paper.

2. DIORAMA Description

In this section, we describe the DIORAMA digital twin, we will develop under the Territori Aperti project [3] by assembling all the applications, analyses, and visualizations developed within the project. DIORAMA aims to create a complete representation of the territory integrating

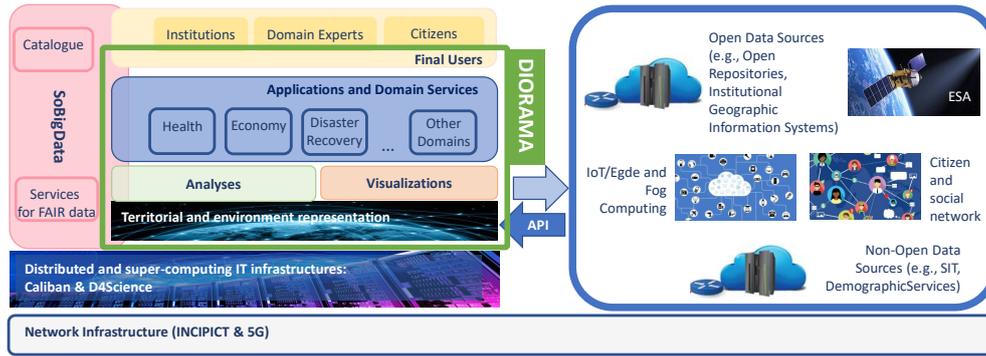


Figure 1: Overview of DIORAMA

different existing data sources each covering a specific domain of the territorial management (i.e., health management, economy management, and so on), helping institutions, citizens and domain experts in its sustainable management.

Figure 1 reports a high-level layered architecture of DIORAMA which represents the follow-up of *Territori Aperti* and is founded on a *Territorial and environment representation* derived from data. Using this constantly updated representation of the environment, we conducted in *Territori Aperti* several research activities related to analyses and visualization. These research activities have resulted in approaches and methodologies that we will deploy in the *Analyses & Visualization layer* of DIORAMA digital twin. Such approaches and methodologies are then used in the development of several applications addressing one or more domains related to the territorial management. These applications are used by stakeholders (such as, institutions, domain experts, citizens, and so on) to actively participate to, to be informed of and to effectively take decisions about a territory.

Data is the primary resource of DIORAMA and drives the analyses, visualizations, and applications we will develop. In *Territori Aperti* we collected information through APIs coming from different sources (e.g., online data sources¹, ESA information systems [12], citizen and social networks, and so on). We will reuse them in DIORAMA as first step and, as done in *Territori Aperti*, all the resources produced in DIORAMA (such as data, methods, applications, and so on) will be publicly shared following the Open Science and FAIR principle.

We aim to support the DIORAMA implementation on the distributed and super-computing IT infrastructures *Territori Aperti* leveraged on: Caliban Cluster and D4Science. The Caliban Cluster supercomputer has a power computing of about 5.0 Teraflops and allows the execution of extensive and computational complex analyses. It is a very important infrastructure for a broad audience of researchers and students from various disciplines in the Department of Information Engineering, Computer Science, and Mathematics of the University of L'Aquila. D4Science is, instead, the IT distributed platform managed by National Council of Research on top of which *Territori Aperti* gateway and SoBigData EU research infrastructure [13] are deployed. Being one exploratory of SoBigData, *Territori Aperti* (as well as the future DIORAMA) exploits SoBigData and all the resources it provides, including the public IT platform that implements

¹In *Territori Aperti* we established conventions with some institutions to also have data that are not open

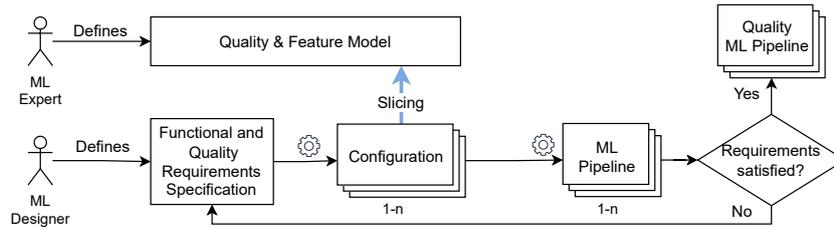


Figure 2: High-level architecture of MANILA

Open Science and FAIR principles.

Note that DIORAMA must be intended as an ecosystem of high-level services and end-user applications. Thus it must not be confused with other mid-level technologies such as Data Lakes, Federated DB, and Polystores [14, 15] which are part of the core components on which DIORAMA can be implemented.

As final remark, due to the high amount of data DIORAMA could exchanges, a suitable network infrastructure should be used. DIORAMA can count on two enabling technologies: the fiber optic ring implemented within the INCIPICT project and the 5G infrastructure being tested in the city of L'Aquila.

In the following sections, we describe first the Analyses and Visualization techniques we implemented in Territori Aperti (in Section 3 and in Section 4, respectively), and then we present the applications developed on top of such approached (in Section 5). Each application can be located in one or more of the domains depicted in figure 1.

3. Analyses and Research Projects

This section describes the analyses techniques and methodologies we have developed in Territori Aperti. Together with visualization approaches, they represent the foundations for the development of DIORAMA. In particular, we describe the following: *MANILA* (section 3.1), *BeFairest* (section 3.2), and *COMPASS* (section 3.3) approaches.

3.1. Model bAsed developmeNt of ml ppeLines with quAlity (MANILA)

In order to be sustainable and useful for end-users, machine learning-based (ML) data analysis systems² must be accurate in their predictions and must satisfy *quality attributes (QA)*. We consider as crucial QA: fairness [16], computational complexity [17], interpretability [18], explainability [19], and privacy [20]. In order to build analysis tools based on ML pipelines satisfying such QA, a data scientist must have a good knowledge of the underlying ML domain.

To simplify the implementation of quality ML pipeline, we are developing MANILA, an innovative model-driven framework that guides data scientists in developing ML pipelines assuring quality requirements [21, 22]. Figure 2 depicts the high-level architecture of such a framework. In particular, ML pipelines are made of typical phases [23] that embed a set of

²In this context, we define ML systems as a set of one or more ML pipelines.

standard components identified by the system's functional requirements and a set of variability points. Product-Line Architectures, specified by Feature Models [24], represent a suitable model to formalize ML pipelines with variability. But, they miss adequate means to specify quality attributes and requirements, thresholds and metrics. To address this issue, in MANILA approach (see Figure2), we extend the feature models meta-model to enable: *i*) the creation, by the ML expert, of a Quality & Feature Model (as done in [25]) *ii*) the specification of functional and quality requirements by the data scientist. In particular, the data scientist specifies a set of functional and quality requirements compliant with the defined meta-model. These requirements are used to automatically generate, from the extended feature model provided by the ML expert, a set of ML pipeline configurations able to satisfy the defined functional requirements (the *Configuration* boxes in the Figure). The configurations are defined by removing from the feature model all the components (and their relative specification) not suitable to meet the specified requirements. The derived configurations are used to generate a set of python scripts each implementing a ML pipeline possibly satisfying the posed constraints. These pipelines are then tested to verify if the quality constraint is actually satisfied.

The framework returns the set of Quality ML Pipelines, satisfying quality constraints, if any, or demands the data scientist to relax quality requirements and repeat the process.

3.2. BeFairest

Fairness represents one of the critical quality attributes for ML systems, as it ensures that they are unbiased and do not apply discrimination among groups in the input dataset. Its importance motivated the joint effort of studying the *Bias and Fairness* problem in ML through the approach named *BeFairest*. In particular, we developed and tested the *Debiasser for Multiple Variables (DEMV)*, a novel preprocessing algorithm to improve fairness in binary and multi-class classification problems with any number of sensitive variables [26].

For any ML system to be *fair* means to comply with several fairness metrics (for instance, *Statistical Parity* [27]) within strict thresholds. DEMV re-balances the various combinations of *sensitive variables*, each embodying one particular unprivileged group of samples, and by doing so is capable of significantly improving the dataset's fairness while keeping the inevitable accuracy losses of the classifier down to indiscernible percentages.

DEMV efficiently manage multiple sensitive variables, both binary and categorical, making it highly flexible for any use. This is a considerable improvement concerning the baselines described in Fairness literature, as they are often limited to one sensitive variable (e.g., *Exponentiated gradient* [28]) or only binary labels (e.g., *Reweighting* [29]). Thanks to these properties, DEMV can be used for many of the application domains depicted in figure 1.

The implementation of DEMV is available at the Territori Aperti RI.

3.3. Improving prediction by integrating explainable cOMputational Models based on heterogeneous data (COMPASS)

In complex scenario, such as the medical domain, data contain variability on the types and in the formats and they usually comes from different sources. Due to issues related to high redundancy, missing data, untruthfulness and having been created for several purposes, it

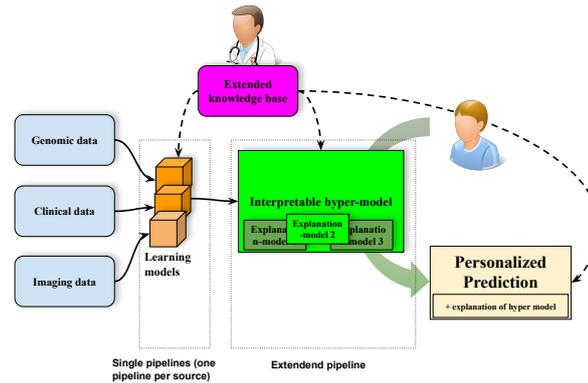


Figure 3: High-level visual description for the system to analyze in the CVD domain.

is difficult to integrate these heterogeneous data to meet the business information demand. In addition, although there are approaches that bring together different types of data and from different sources (neural networks trained and which fuse together heterogeneous and/or multi-sources data), in complex and sensible scenarios this is not always possible, because they are critical domains protected by different regulations possibly related to ethics and data privacy issues. To better clarify our goal, let us refer, as real case study, to cardiovascular risk (CVD) prediction where we can apply the process mentioned above and summarized in Figure 3. Suppose there are different departments within a hospital (the one responsible for the genetic data, the one responsible for the clinical data, the one responsible for the images) that do not share the data in their entirety. In these cases, it is necessary to proceed for local learning and subsequently aggregate their individual predictions and tuning parameters to create a new extended learning model. The aim of COMPASS is to define a system that generates accurate predictions, exploits heterogeneous data and aims to be interpretable and explainable, also using a pre-defined domain (in the considered scenario, medical) knowledge to assist the intelligent learning models used in the system. In this case study, starting from single pipelines, we define a new extended model, called Hyper Model (HM). The HM will be equipped with a knowledge of application (e.g., medical) domain and it will be composed by components capable of defining the explainability and interpretability, essential in the (medical) domain to bring trust in the results of the predictions obtained by the single learning pipelines. Since HM is driven on the predictions and on the parameters/weights of the individual local pipelines, it does not directly access to raw data and hence solves the problems related to data privacy issues. Note the difference with respect to architectural models such as federated learning [30], central learning or swarm learning [31], which, although similar and while solving privacy problems in the same way, do not solve intrinsic prediction issues, such as interpretability and explainability modules.

4. Visual Analytics

The analyses within Territori Aperti and hence in DIORAMA are driven by vast sets of data gained by different sources. To derive meaningful insights from them, it is necessary to employ robust and scalable analytics that goes beyond the available data management systems enabling the views of small portions of data. Data visualization plays a crucial role in digital twin analysis and interpretation [4]. According to Vrabič et al. [32], a digital twin holds information that is continuously visualised in a variety of ways to predict current and future conditions. Research has shown how seeing and understanding together a large amount of data enables humans to gather deeper knowledge and insights. Thus, approaches that integrate the exploration capacity of experts with the enormous processing power of computers are winning for the realize a powerful knowledge discovery tool that harnesses the best of both worlds [33]. Such kind of approaches are called Visual Analytics. The typical steps of the Visual Analytics process can be summarized as follows: *i*) Data Pre-processing (i.e. cleaning, transformation, integration); *ii*) data analysis; *iii*) simple data Visualization; *iv*) Users insightful knowledge generation through human perception, cognition, and reasoning activities; *v*) Users make new hypotheses and integrate the newly generated knowledge into the analysis and visualization through interactions; *vi*) Regenerate an updated visualization based on the interactions to reflect the user's understanding of the data.

Visualization techniques can be classified according to [34]: *i*) The type of data to be visualized: one-dimensional data such as temporal data, two-dimensional data such as geographic maps and relational tables, text and hypertext, hierarchies and graphics, etc. *ii*) the Visualization techniques: bubble chart, histogram, scatter plot, parallel coordinate, infographic, etc. *iii*) the Interaction techniques: zooming, linking, overview+detail, fisheye etc. The above dimensions can be considered orthogonal: any visualization technique can be used in conjunction with any interaction technique as well as any type of data. In addition, a specific system can be designed to support different types of data and can use a combination of multiple visualization and interaction techniques. This allows you to quickly create different types of views that help to dig deeper into the data. In Visual Analytics, the phases of querying, exploring and visualizing data come together in a single process, helping to interpret data more easily and thus make analytics easier for non-experts. The data are displayed interactively and graphically, users can discover insights into the data without having to know how build charts and other visualizations and without being proficient in analytical techniques, and therefore make smarter decisions faster. DIORAMA will implement a novel Visual Analytics that supports human thinking, fast data exploration and iteration, stakeholders collaborations and their insights sharing.

5. Developed Applications

In this section, we describe the applications we developed in Territori Aperti and that can be implemented in DIORAMA. These applications are made on top of the outcomes obtained from the analyses and visualization techniques described respectively in sections 3 and 4. In this paper we present the following applications: *TA-Analytics* (section 5.1), *Territori Aperti Toolkit*

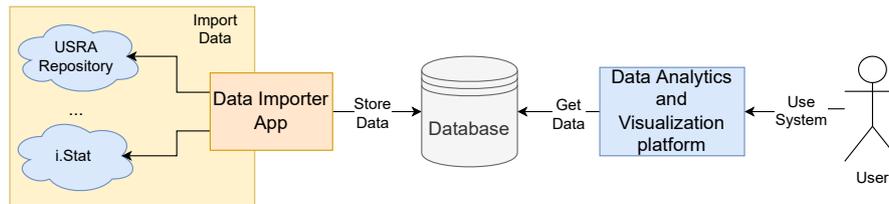


Figure 4: High-level architecture of TA-Analytics

(section 5.2), *Evacuation and Reconstruction Planning* (DiReCT approach, in Section 5.3) and *SismaDL* (section 5.4).

5.1. TA-Analytics

Following the Open Science and FAIR principles, each end-user of DIORAMA should be able to access and use the data collected, if restrictions do not bind them. TA-Analytics is an application for the collection and analysis of Open Data that provides an interface for building interactive dashboards shareable among all users.

TA-Analytics is made of two principal components, which are highlighted in figure 4. The first is the *Data Importer (DI)* component, which is responsible for automatically downloading datasets from different Open Data repositories using the services (i.e., APIs) exposed by them. The component can interact with different web services using different protocols. After downloading the datasets, DI stores the imported data inside a database. This process of downloading and storing data is repeated periodically to have the most updated data available. At the time of this paper, DI collects data from two Open Data repositories: *i.Stat*³, and *USRA*⁴.

The second principal component is the *Data Analytics and Visualization (DAV)* application which interacts with the database to retrieve the downloaded datasets. DAV is the entry point for the end-users to all the datasets and services offered by TA-Analytics, allowing them to build dashboards comprising analyses and charts. DAV offers the users a graphical interface to interact with the data and make interactive dashboards. Each TA-Analytics dashboard can include different datasets and visualizations and analyses. Finally, the implemented dashboards can be shared among users and embedded in other applications.

TA-Analytics represents one of the main results of the Territori Aperti project, embodying the Open Science and FAIR philosophy by allowing users to build and share analyses using Open Data from every domain of interest. Finally, TA-Analytics is extensible since it can embed new methods and visualization techniques, as well as, new data the users wants to share.

5.2. Territori Aperti Toolkit

Natural disasters have a significant impact on the population and must be handled promptly by competent administrations. However, the management of a disaster is not a simple activity. It

³<http://dati.istat.it/>

⁴<https://bde.comuneaq.usra.it/bdeTrasparente/openData/openDataSet/>

includes a series of critical aspects that must be considered and carefully managed, as demonstrated by the two experiences of seismic craters of 2009 and 2016, which highlighted a series of critical issues in the reconstruction process and emergency management.

For this reason, we have developed the *Territori Aperti Toolkit*⁵ that we believe will improve sustainability of post-disaster recovery procedure. This dynamic tool provides recommendations to organizations, institutions, and citizens to better manage every critical aspect of a disaster. The Toolkit is implemented as a website that maps every critical aspect of managing a disaster to a series of good and bad practices. Recalling figure 1, the *Territori Aperti Toolkit* can be located in the *Disaster Recovery* domain since its primary users can be identified in institutions and citizens involved in managing a disaster (e.g., municipalities, special offices, civil protection, and so on). The Toolkit is made of cards, classified by phases and sectors of application, each with a common set of fields highlighting several aspects of disaster management. Among the main features of the Toolkit, in addition to the consultation of the various data sheets, there is the possibility of filtering them by *Phases*, *Sectors* and searching them using keywords, titles, or names of the entities involved. All this happens through a dedicated search panel. It is also possible to generate a PDF of all the cards currently shown, possibly filtered through the defined conditions.

5.3. Disaster Recovery: the Direct Approach

Natural disasters can cause widespread damage to buildings and infrastructures and kill thousands of living beings. These events are difficult to be overcome both by the populations and by government authorities. Two challenging issues require in particular to be addressed: find an effective way to evacuate people first, and later to rebuild houses and other infrastructures. An adequate recovery strategy to evacuate people and start reconstructing damaged areas on a priority basis can then be a game changer allowing to overcome effectively those terrible circumstances. In this perspective, in [35] we present DiReCT, an approach based on i) a dynamic optimization model designed to timely formulate an evacuation plan of an area struck by an earthquake, and ii) a decision support system, based on double deep Q Network, able to guide efficiently the reconstruction the affected areas. The latter works by considering both the resources available and the needs of the various stakeholders involved (e.g., residents social benefits and political priorities). The ground on which both the above solutions stand was a dedicated geographical data extraction algorithm, called “GisToGraph”, especially developed for this purpose. To check applicability of the whole approach, we dovetailed it on the real use-case of the historical city center of L’Aquila (Italy) using detailed GIS data and information on urban land structure and buildings vulnerability. Several simulations were run on the underlining network generated. First, we ran experiments to safely evacuate in the shortest possible time as many people as possible from an endangered area towards a set of safe places. Then, using DDQN, we generated different reconstruction plans and selected the best ones considering both social benefits and political priorities of the building units. The described approaches are comprised in a more general data science framework delved to produce an effective response to natural disasters. DIORAMA will embed the two services belonging to DiReCT.

⁵<https://toolkit.territoriaperti.univaq.it/>

5.4. SismaDL

The emergency caused by a natural disasters must be tackled promptly by public institutions. In this situation, Governments enact specific laws (i.e., decrees) to handle the emergency and the reconstruction of destroyed areas. As it happened in 2009 and 2016 when the Italian Government issued several, very different, decrees to face respectively the earthquakes of *L'Aquila* and *Centro Italia*. In this work, we implemented SismaDL⁶ [36], an LKIF [37] based ontology, that models the laws in the domain of natural disasters. SismaDL has been used to model the aforementioned laws to build a knowledge base useful to reason about why one regulation is less effective and efficient than the other. In particular, SismaDL extends the LKIF ontology to add entities and properties specific of the 2009 and 2016 regulations. Using this ontology is it possible to analyze the differences between the two regulations concerning, for instance, the Legal Model, the Social Measures, or the Financing Mechanism. Recalling figure 1, SismaDL can be located under the *Disaster Recovery* domain, since it is mostly focused on understanding and analyzing the regulatory context that manages a natural calamity.

6. Conclusion and Future Works

In this paper, we described DIORAMA, a digital twin for sustainable territorial management we will develop as the follow-up of Territori Aperti project. It will integrate, by leveraging on SoBigData research infrastructure, all the research results and applications developed in Territori Aperti by promoting the Open Science and FAIR principles. We have also described in more detail the analyses and the visualization research we are conducting and the main applications developed in Territori Aperti, each belonging to at least one of the domains depicted in figure 1. To the best of our knowledge, DIORAMA, with its innovative data science approach and tools, goes beyond state of the art in the field of territory management because it realizes innovative and efficient services for the sustainable territorial management targeting several involved stakeholders by reusing existing data, providing novel data analysis and powerful visualization techniques.

As lesson learned in Territori Aperti and in the definition of DIORAMA, we want to highlight two main aspects: i) a lot of quality data already exists and it can be exploited both for new research in territorial management and for the development of innovative and sustainable services and application for institutions, citizens and decision-makers; ii) by reusing and integrating research projects achievements we are able to implement a novel DT that also implement the open science principles and the data FAIR capabilities.

Future works are manifold. We need to still work on the foundational approaches described in sections 3 and 4 and on improving the applications reported in section 5. In the future, we aim to expand our work to other domains (such as urban planning) to make our DIORAMA more complete and valuable.

Acknowledgments. This work is partially supported by Territori Aperti a project funded by Fondo Territori Lavoro e Conoscenza CGIL CISL UIL and by SoBigData-PlusPlus H2020-INFRAIA-2019-1 EU project, contract number 871042.

⁶SismaDL is available in the Territori Aperti RI.

References

- [1] S. Tavakkol, H. To, S. H. Kim, P. Lynett, C. Shahabi, An entropy-based framework for efficient post-disaster assessment based on crowdsourced data, in: Proceedings of the 2nd ACM SIGSPATIAL International Workshop EM-GIS '16, 2016, pp. 13:1–13:8.
- [2] M. Grieves, Digital twin: manufacturing excellence through virtual factory replication, White paper 1 (2014) 1–7.
- [3] Territori Aperti website, 2019. URL: <https://territoriaperti.univaq.it/>.
- [4] A. Fuller, Z. Fan, C. Day, C. Barlow, Digital twin: Enabling technologies, challenges and open research, IEEE access 8 (2020) 108952–108971.
- [5] A. Suvorova, Towards digital twins for the development of territories, in: Digital Transformation in Industry, Springer, 2022, pp. 121–131.
- [6] F. N. Abdeen, S. M. Sepasgozar, City digital twin concepts: A vision for community participation, Environmental Sciences Proceedings 12 (2022) 19.
- [7] G. Schrotter, C. Hürzeler, The digital twin of the city of zurich for urban planning, PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science 88 (2020).
- [8] E. Shahat, C. T. Hyun, C. Yeom, City digital twin potentials: A review and research agenda, Sustainability 13 (2021) 3386.
- [9] T. Deng, K. Zhang, Z.-J. M. Shen, A systematic review of a digital twin city: A new pattern of urban governance toward smart cities, J. of Management Science and Eng. 6 (2021).
- [10] A. V. Mukhacheva, M. N. Ugryumova, I. S. Morozova, M. Y. Mukhachyev, Digital twins of the urban ecosystem to ensure the quality of life of the population, in: International Scientific and Practical Conference Strategy of Development of Regional Ecosystems “Education-Science-Industry”(ISPCR 2021), Atlantis Press, 2022, pp. 331–338.
- [11] G. Caprari, Digital twin for urban planning in the green deal era: A state of the art and future perspectives, Sustainability 14 (2022) 6263.
- [12] T. Hörber, The european space agency and the european union, EU space policy (2015).
- [13] V. Grossi, B. Rapisarda, F. Giannotti, D. Pedreschi, Data science at sobigdata: the european research infrastructure for social mining and big data analytics, International Journal of Data Science and Analytics 6 (2018) 205–216.
- [14] F. Nargesian, E. Zhu, R. J. Miller, K. Q. Pu, P. C. Arocena, Data lake management: Challenges and opportunities, Proc. VLDB Endow. 12 (2019) 1986–1989. URL: <https://doi.org/10.14778/3352063.3352116>. doi:10.14778/3352063.3352116.
- [15] J. Duggan, A. J. Elmore, M. Stonebraker, M. Balazinska, B. Howe, J. Kepner, S. Madden, D. Maier, T. Mattson, S. Zdonik, The bigdawg polystore system, SIGMOD Rec. 44 (2015) 11–16. URL: <https://doi.org/10.1145/2814710.2814713>. doi:10.1145/2814710.2814713.
- [16] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, A. Galstyan, A Survey on Bias and Fairness in Machine Learning, ACM Computing Surveys 54 (2021) 1–35.
- [17] S. A. Cook, An overview of computational complexity, Comm. of the ACM (1983).
- [18] D. V. Carvalho, E. M. Pereira, J. S. Cardoso, Machine learning interpretability: A survey on methods and metrics, Electronics 8 (2019) 832.
- [19] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, Explainable AI: A Review of Machine Learning Interpretability Methods, Entropy 23 (2020) 18.
- [20] B. C. M. Fung, K. Wang, R. Chen, P. S. Yu, Privacy-preserving data publishing: A survey

- of recent developments, *ACM Comput. Surv.* 42 (2010).
- [21] G. d'Aloisio, Quality-driven machine learning-based data science pipeline realization: a software engineering approach, in: *2022 IEEE/ACM 44th International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, IEEE, 2022, pp. 291–293.
 - [22] G. d'Aloisio, A. Di Marco, G. Stilo, Modeling Quality and Machine Learning Pipelines through Extended Feature Models, 2022. doi:10.48550/arXiv.2207.07528.
 - [23] S. Amershi, A. Begel, C. Bird, R. DeLine, H. Gall, E. Kamar, N. Nagappan, B. Nushi, T. Zimmermann, Software engineering for machine learning: A case study, in: *IEEE/ACM 41st ICSE, SEIP track*, IEEE, 2019, pp. 291–300.
 - [24] K. C. Kang, S. G. Cohen, J. A. Hess, W. E. Novak, A. S. Peterson, Feature-oriented domain analysis (FODA) feasibility study, Technical Report, Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst, 1990.
 - [25] M. Asadi, S. Soltani, D. Gasevic, M. Hatala, E. Bagheri, Toward automated feature model configuration with optimizing non-functional requirements, *Information and Software Technology* 56 (2014) 1144–1165.
 - [26] G. d'Aloisio, G. Stilo, A. Di Marco, A. D'Angelo, Enhancing fairness in classification tasks with multiple variables: A data-and model-agnostic approach, in: *International Workshop on Algorithmic Bias in Search and Recommendation*, Springer, 2022, pp. 117–129.
 - [27] M. J. Kusner, J. Loftus, C. Russell, R. Silva, Counterfactual Fairness, in: *Advances in Neural Information Processing Systems*, volume 30, Curran Associates, Inc., 2017.
 - [28] A. Agarwal, A. Beygelzimer, M. Dudik, J. Langford, H. Wallach, A Reductions Approach to Fair Classification, in: *Proc. of the 35th Int. Conf. on Machine Learning*, 2018, pp. 60–69.
 - [29] F. Kamiran, T. Calders, Data preprocessing techniques for classification without discrimination, *Knowledge and Information Systems* 33 (2012) 1–33.
 - [30] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, H. Yu, Federated learning, *Synthesis Lectures on Artificial Intelligence and Machine Learning* 13 (2019) 1–207.
 - [31] S. Warnat-Herresthal, H. Schultze, K. L. Shastry, S. Manamohan, S. Mukherjee, V. Garg, R. Sarveswara, K. Händler, P. Pickkers, N. A. Aziz, et al., Swarm learning for decentralized and confidential clinical machine learning, *Nature* 594 (2021) 265–270.
 - [32] R. Vrabič, J. A. Erkoyuncu, P. Butala, R. Roy, Digital twins: Understanding the added value of integrated models for through-life engineering services, *Procedia Manufacturing* 16 (2018) 139–146. doi:<https://doi.org/10.1016/j.promfg.2018.10.167>, proceedings of the 7th International Conference on Through-life Engineering Services.
 - [33] P. C. Wong, Visual data mining, *IEEE Computer Graphics and Applications* 19 (1999).
 - [34] D. A. Keim, M. O. Ward, *Visual data mining techniques*, 2002.
 - [35] G. Mudassir, E. E. Howard, L. Pasquini, C. Arbib, E. Clementini, A. Di Marco, G. Stilo, Toward effective response to natural disasters: A data science approach, *IEEE Access* 9 (2021) 167827–167844. doi:10.1109/ACCESS.2021.3135054.
 - [36] F. Caroccia, D. D'Agostino, G. d'Aloisio, A. Di Marco, G. Stilo, SismaDL: an ontology to represent post-disaster regulation (2021) 14.
 - [37] R. Hoekstra, J. Breuker, M. Di Bello, A. Boer, et al., The LKIF core ontology of basic legal concepts., *LOAIT* 321 (2007) 43–63.