Zbornik 24. mednarodne multikonference

INFORMACIJSKA DRUŽBA

Zvezek H

Proceedings of the 24th International Multiconference

INFORMATION SOCIETY

Volume H

IS 2021

Delavnica URBANITE 2021

URBANITE Workshop 2021

Uredniki • Editors:

Sergio Campos, Shabnam Farahmand, Nathalie van Loon, Erik Dovgan

8. oktober 2021 Ljubljana, Slovenija • 8 October 2021 Ljubljana, Slovenia • http://is.ijs.si

Zbornik 24. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2021

Zvezek H

Proceedings of the 24th International Multiconference

INFORMATION SOCIETY – IS 2021

Volume H

Delavnica URBANITE 2021
URBANITE Workshop 2021

Uredniki / Editors

Sergio Campos, Shabnam Farahmand, Nathalie van Loon, Erik Dovgan

http://is.ijs.si

8. oktober 2021 / 8 October 2021
Ljubljana, Slovenia

Uredniki:

Sergio Campos
TECNALIA Research & Innovation, Spain

Shabnam Farahmand
Forum Virium Helsinki, Finland

Nathalie van Loon
City of Amsterdam, Netherlands

Erik Dovgan
Department of Intelligent Systems
Jožef Stefan Institute, Ljubljana

Založnik: Institut »Jožef Stefan«, Ljubljana
Priprava zbornika: Mitja Lasič, Vesna Lasič, Lana Zemljak
Oblikovanje naslovnice: Vesna Lasič

Dostop do e-publikacije:
http://library.ijs.si/Stacks/Proceedings/InformationSociety

Ljubljana, oktober 2021

Informacijska družba
ISSN 2630-371X

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID **85953795**
ISBN 978-961-264-221-1 (PDF)

PREDGOVOR MULTIKONFERENCI INFORMACIJSKA DRUŽBA 2021

Štiriindvajseta multikonferenca *Informacijska družba* je preživela probleme zaradi korone v 2020. Odziv se povečuje, v 2021 imamo enajst konferenc, a pravo upanje je za 2022, ko naj bi dovolj velika precepljenost končno omogočila normalno delovanje. Tudi v 2021 gre zahvala za skoraj normalno delovanje konference tistim predsednikom konferenc, ki so kljub prvi pandemiji modernega sveta pogumno obdržali visok strokovni nivo.

Stagnacija določenih aktivnosti v 2020 in 2021 pa skoraj v ničemer ni omejila neverjetne rasti IKTja, informacijske družbe, umetne inteligence in znanosti nasploh, ampak nasprotno – rast znanja, računalništva in umetne inteligence se nadaljuje z že kar običajno nesluteno hitrostjo. Po drugi strani se je pospešil razpad družbenih vrednot, zaupanje v znanost in razvoj. Se pa zavedanje večine ljudi, da je potrebno podpreti stroko, čedalje bolj krepi, kar je bistvena sprememba glede na 2020.

Letos smo v multikonferenco povezali enajst odličnih neodvisnih konferenc. Zajema okoli 170 večinoma spletnih predstavitev, povzetkov in referatov v okviru samostojnih konferenc in delavnic ter 400 obiskovalcev. Prireditev so spremljale okrogle mize in razprave ter posebni dogodki, kot je svečana podelitev nagrad – seveda večinoma preko spleta. Izbrani prispevki bodo izšli tudi v posebni številki revije Informatica (http://www.informatica.si/), ki se ponaša s 45-letno tradicijo odlične znanstvene revije.

Multikonferenco Informacijska družba 2021 sestavljajo naslednje samostojne konference:

- Slovenska konferenca o umetni inteligenci
- Odkrivanje znanja in podatkovna skladišča
- Kognitivna znanost
- Ljudje in okolje
- 50-letnica poučevanja računalništva v slovenskih srednjih šolah
- Delavnica projekta Batman
- Delavnica projekta Insieme Interreg
- Delavnica projekta Urbanite
- Študentska konferenca o računalniškem raziskovanju 2021
- Mednarodna konferenca o prenosu tehnologij
- Vzgoja in izobraževanje v informacijski družbi

Soorganizatorji in podporniki multikonference so različne raziskovalne institucije in združenja, med njimi ACM Slovenija, SLAIS, DKZ in druga slovenska nacionalna akademija, Inženirska akademija Slovenije (IAS). V imenu organizatorjev konference se zahvaljujemo združenjem in institucijam, še posebej pa udeležencem za njihove dragocene prispevke in priložnost, da z nami delijo svoje izkušnje o informacijski družbi. Zahvaljujemo se tudi recenzentom za njihovo pomoč pri recenziranju.

S podelitvijo nagrad, še posebej z nagrado Michie-Turing, se avtonomna stroka s področja opredeli do najbolj izstopajočih dosežkov. Nagrado Michie-Turing za izjemen življenjski prispevek k razvoju in promociji informacijske družbe je prejel prof. dr. Jernej Kozak. Priznanje za dosežek leta pripada ekipi Odseka za inteligentne sisteme Instituta "Jožef Stefan" za osvojeno drugo mesto na tekmovanju XPrize Pandemic Response Challenge za iskanje najboljših ukrepov proti koroni. »Informacijsko limono« za najmanj primerno informacijsko potezo je prejela trditev, da je aplikacija za sledenje stikom problematična za zasebnost, »informacijsko jagodo« kot najboljšo potezo pa COVID-19 Sledilnik, tj. sistem za zbiranje podatkov o koroni. Čestitke nagrajencem!

Mojca Ciglarič, predsednik programskega odbora
Matjaž Gams, predsednik organizacijskega odbora

FOREWORD - INFORMATION SOCIETY 2021

The 24th *Information Society Multiconference* survived the COVID-19 problems. In 2021, there are eleven conferences with a growing trend and real hopes that 2022 will be better due to successful vaccination. The multiconference survived due to the conference chairs who bravely decided to continue with their conferences despite the first pandemic in the modern era.

The COVID-19 pandemic did not decrease the growth of ICT, information society, artificial intelligence and science overall, quite on the contrary – the progress of computers, knowledge and artificial intelligence continued with the fascinating growth rate. However, COVID-19 did increase the downfall of societal norms, trust in science and progress. On the other hand, the awareness of the majority, that science and development are the only perspectives for a prosperous future, substantially grows.

The Multiconference is running parallel sessions with 170 presentations of scientific papers at eleven conferences, many round tables, workshops and award ceremonies, and 400 attendees. Selected papers will be published in the Informatica journal with its 45-years tradition of excellent research publishing.

The Information Society 2021 Multiconference consists of the following conferences:

- Slovenian Conference on Artificial Intelligence
- Data Mining and Data Warehouses
- Cognitive Science
- People and Environment
- 50-years of High-school Computer Education in Slovenia
- Batman Project Workshop
- Insieme Interreg Project Workshop
- URBANITE Project Workshop
- Student Computer Science Research Conference 2021
- International Conference of Transfer of Technologies
- Education in Information Society

The multiconference is co-organized and supported by several major research institutions and societies, among them ACM Slovenia, i.e. the Slovenian chapter of the ACM, SLAIS, DKZ and the second national academy, the Slovenian Engineering Academy. In the name of the conference organizers, we thank all the societies and institutions, and particularly all the participants for their valuable contribution and their interest in this event, and the reviewers for their thorough reviews.

The award for lifelong outstanding contributions is presented in memory of Donald Michie and Alan Turing. The Michie-Turing award was given to Prof. Dr. Jernej Kozak for his lifelong outstanding contribution to the development and promotion of the information society in our country. In addition, the yearly recognition for current achievements was awarded to the team from the Department of Intelligent systems, Jožef Stefan Institute for the second place at the XPrize Pandemic Response Challenge for proposing best counter-measures against COVID-19. The information lemon goes to the claim that the mobile application for tracking COVID-19 contacts will harm information privacy. The information strawberry as the best information service last year went to COVID-19 Sledilnik, a program to regularly report all data related to COVID-19 in Slovenia. Congratulations!

Mojca Ciglarič, Programme Committee Chair
Matjaž Gams, Organizing Committee Chair

KONFERENČNI ODBORI
CONFERENCE COMMITTEES

International Programme Committee

Vladimir Bajic, South Africa
Heiner Benking, Germany
Se Woo Cheon, South Korea
Howie Firth, UK
Olga Fomichova, Russia
Vladimir Fomichov, Russia
Vesna Hljuz Dobric, Croatia
Alfred Inselberg, Israel
Jay Liebowitz, USA
Huan Liu, Singapore
Henz Martin, Germany
Marcin Paprzycki, USA
Claude Sammut, Australia
Jiri Wiedermann, Czech Republic
Xindong Wu, USA
Yiming Ye, USA
Ning Zhong, USA
Wray Buntine, Australia
Bezalel Gavish, USA
Gal A. Kaminka, Israel
Mike Bain, Australia
Michela Milano, Italy
Derong Liu, Chicago, USA
Toby Walsh, Australia
Sergio Campos-Cordobes, Spain
Shabnam Farahmand, Finland
Sergio Crovella, Italy

Organizing Committee

Matjaž Gams, chair
Mitja Luštrek
Lana Zemljak
Vesna Koricki
Mitja Lasič
Blaž Mahnič
Klara Vulikić

Programme Committee

Mojca Ciglarič, chair
Bojan Orel,
Franc Solina,
Viljan Mahnič,
Cene Bavec,
Tomaž Kalin,
Jozsef Györkös,
Tadej Bajd
Jaroslav Berce
Mojca Bernik
Marko Bohanec
Ivan Bratko
Andrej Brodnik
Dušan Caf
Saša Divjak
Tomaž Erjavec
Bogdan Filipič
Andrej Gams
Matjaž Gams
Mitja Luštrek
Marko Grobelnik
Nikola Guid
Marjan Heričko
Borka Jerman Blažič Džonova
Gorazd Kandus
Urban Kordeš
Marjan Krisper
Andrej Kuščer
Jadran Lenarčič
Borut Likar
Janez Malačič
Olga Markič
Dunja Mladenič
Franc Novak
Vladislav Rajkovič
Grega Repovš
Ivan Rozman
Niko Schlamberger
Stanko Strmčnik
Jurij Šilc
Jurij Tasič
Denis Trček
Andrej Ule
Boštjan Vilfan
Baldomir Zajc
Blaž Zupan
Boris Žemva
Leon Žlajpah
Niko Zimic
Rok Piltaver
Toma Strle
Tine Kolenik
Franci Pivec
Uroš Rajkovič
Borut Batagelj
Tomaž Ogrin
Aleš Ude
Bojan Blažica
Matjaž Kljun
Robert Blatnik
Erik Dovgan
Špela Stres
Anton Gradišek

KAZALO / TABLE OF CONTENTS

Zbornik 24. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2021

Zvezek H

Proceedings of the 24th International Multiconference

INFORMATION SOCIETY – IS 2021

Volume H

Delavnica URBANITE 2021
URBANITE Workshop 2021

Uredniki / Editors

Sergio Campos, Shabnam Farahmand, Nathalie van Loon, Erik Dovgan

http://is.ijs.si

8. oktober 2021 / 8 October 2021
Ljubljana, Slovenia

PREDGOVOR

Delavnica URBANITE bo forum za predstavitev najsodobnejših mobilnostnih rešitev v urbanem okolju s poudarkom na prebojnih tehnologijah, kot so umetna inteligenca, sistemi za podporo odločanju, analitika velikih podatkov in napovedni algoritmi, ki se uporabljajo pri analizi podatkov o mobilnosti, napovedovanju dogodkov in podpori javnim upravam pri sprejemanju strateških odločitev.

Delavnica je aktivnost projekta URBANITE. Vabimo prispevke akademskega sveta, industrije in oblikovalcev strategij na področju mobilnosti in pametnih mest.

Delavnica se bo osredotočila na naslednje teme v okviru mobilnosti v pametnih mestih:

- Umetna inteligenca
- Inteligentni sistemi
- Strojno učenje
- Podatkovno rudarjenje
- Sistemi za podporo odločanju
- Analitika velikih podatkov
- Dejavnosti soustvarjanja
- Socialni vidiki
- Urbana preobrazba

FOREWORD

The URBANITE Workshop will be a forum for presenting the state-of-the-art solutions for the urban mobility with the focus on disruptive technologies such as artificial intelligence, decision support systems, big data analytics and predictive algorithms, which are applied in mobility data analysis, eventualities prediction, and supporting public administrations in making policy-related decisions.

The workshop is an activity of the URBANITE project. We welcome papers from the academia, the industry, and the policy makers in the mobility and smart cities fields.

The workshop will focus on the following topics within the scope of mobility within smart cities:

- Artificial intelligence
- Intelligent systems
- Machine learning
- Data mining
- Decision support systems
- Big data analytics
- Co-creation activities
- Social-related aspects
- Urban transformation

PROGRAMSKI ODBOR / PROGRAMME COMMITTEE

Sergio Campos Cordobes (co-chair)

Shabnam Farahmand (co-chair)

Nathalie van Loon (co-chair)

Denis Costa

Yury Glickman

Maria José López

Giuseppe Ciulla

Maria Suberbiola

Dino Alessi

Massimo Villari

Matjaž Gams

Maj Smerkol

Erik Dovgan (local chair)

How Disruptive Technologies can Strengthen Urban Mobility Transformation. The Experience of URBANITE H2020 Project

Giuseppe Ciulla
Research & Development Laboratory
Engineering Ingegneria Informatica
Palermo, Italy
giuseppe.ciulla@eng.it

Roberto Di Bernardo
Research & Development Laboratory
Engineering Ingegneria Informatica
Palermo, Italy
roberto.dibernardo@eng.it

Isabel Matranga
Research & Development Laboratory
Engineering Ingegneria Informatica
Palermo, Italy
isabel.matranga@eng.it

Francesco Martella
Engineering Department, University of Messina - ALMA Digit S.R.L.
Messina, Italy
fmartella@unime.it

Giovanni Parrino
Engineer - Municipality of Messina
Messina, Italy
nanni.parrino@gmail.com

Shabnam Farahmand
Forum Virium Helsinki Oy
Helsinki, Finland
shabnam.farahmand@forumvirium.fi

ABSTRACT

URBANITE (Supporting the decision-making in URBAN transformation with the use of dIsruptive TEchnologies) is an H2020 project (started in April 2020) investigating the impact, trust and attitudes of civil servants, citizens and other stakeholders concerning the introduction and adoption of disruptive technologies (e.g. AI, Decision Support Systems, big data analytics) in decision-making processes related to the planning and management of urban mobility. The project experiments and validates its approaches and tools in the context of four real use cases in the cities of Amsterdam (NL), Bilbao (ES), Helsinki (FI) and Messina (IT). This article summarises the main findings matured during the first half of the project in the four cities, their main mobility issues and how disruptive technologies can play a role in supporting the decision-making process to solve them. Despite the four cities face different kinds of mobility issues and are characterised by different levels of IT maturity, we identified a chain of three categories of technologies that can improve the efficiency and effectiveness of decision-making processes in all four cities: data access and harmonisation, data analysis and data visualisation.

KEYWORDS

Urban transformation, disruptive technologies, urban mobility, URBANITE project, decision making, data access, data analysis, data visualisation.

1 INTRODUCTION

Today's cities are facing a revolutionary era in urban mobility; this is due to different factors, among the others their continuous growth and the concentration of human activities. To prevent and solve problems related to mobility such as traffic congestion and air pollution (for instance due to $PM_{2.5}$) and its potential link with other risk factors (e.g. Covid-19 spread, as envisaged in recent studies [3], [4]), cities are in continuous search of adequate mobility solutions to satisfy the demand of the growing population, both living in or moving around the cities every day. As a result, decision-makers have to face more and more complex challenges when managing and planning mobility, combining new forms of mobility, that must coexist in the urban structure of modern cities, in compliance with the well-being of citizens and protection of the environment.

The concrete adoption of disruptive technologies in the decision-making processes can represent the pivoting point for a paradigm change in the management of mobility. Decision Support Systems, Artificial Intelligence, predictive algorithms, simulation models, Big Data analytics, etc. offer the opportunity to analyse the current mobility situation, identify present and future trends allowing to predict potential future mobility scenarios [6], [9].

Our investigation focuses on four European cities distributed in four different countries: Amsterdam, Bilbao, Helsinki, and Messina. Each of them offers a different perspective on urban mobility, in terms of characteristics, offered services and challenges. Section 2 presents the four cities, their general characteristics, the specific urban mobility issues they are currently facing, and which kind of disruptive technologies (e.g. artificial intelligence, decision support systems, big data analytics, predictive algorithms, simulation engines) can improve the decision-making processes and how. Final considerations and conclusions are reported in Section 3.

2 URBANITE CITIES

2.1 Amsterdam

Amsterdam, the capital of the Netherlands, in recent years has been growing rapidly in terms of inhabitants and visitors; this growth leads to increased mobility and traffic issues. The city has complex traffic streams with massive amounts of bicycles combined with cars and public transport; this drives the need for finding solutions that can conciliate the ever-growing use of bikes with the other means of transportation (from public transportation to private cars) resulting in more sustainable mobility for the whole city. Part of this view is a strategy tending to increase the appeal of bikes as the main mobility option [5]. This strategy goes through the improvement of the city network of bike lanes and of the overall cycling experience within the city, encouraging virtuous behaviours (e.g. respect of traffic lights) to avoid potential discomfort.

What Amsterdam is aiming for. To reach these objectives the city of Amsterdam would like to align the mobility policies to the real needs of bike mobility, realise a data-driven decision

Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia

mechanism, strengthen the safety and comfort of cycling, and encourage citizens to make sustainable mobility choices.

The role of disruptive technologies in Amsterdam. From a broader perspective, a unique point to access data coming from different sources can support the decision-makers in the identification of the required information, reducing the time spent to search it and speeding up the decision-making process. Since different departments of the municipality (i.e. civil servants) are involved in decision-making, the possibility to easily share information among them (such as data, results of analysis/simulations, map layers, charts, graphs) would improve collaboration and overcome inefficient communication and silos, allowing at the same time the reduction of policy fragmentation and the subsequent uncertainties. From a more specific perspective, data analysis tools can support decision-makers in understanding different aspects of bike mobility (through the analysis of bike-related data) and in identifying dependencies among factors that could influence directly or indirectly bike mobility and its adoption. In this sense tools and models to simulate how decisions and policies can potentially impact on traffic and mobility would offer predictions and the possibility to compare different scenarios. This would allow decision-makers to make choices with minimal negative impact and to minimize related costs. Finally, effective visualisation of information is essential; a dashboard offering map layers, charts and graphs that summarise the status of bike mobility in the city would allow decision-makers to have, in a single view, the overall and relevant information they need to gain new insights about bike mobility in the city (e.g. type of road infrastructure/ bike paths, road safety level, traffic mix/sources, congested routes, cleaner routes in terms of air quality, greener routes, faster routes).

2.2 Bilbao

With an area of 41,60 km^2 and around 355,000 inhabitants, Bilbao is the heart of a metropolitan area that extends along the estuary of the Nervioén River with a population close to 1 million people. In the last 25 years, Bilbao has suffered an important urban transformation from an industrial economy to a city based on a service economy. This has helped to balance the city and provide a friendly environment for pedestrians with wider pavements, reduction of on-street car parking in the city centre, traffic light control system to cater for pedestrians and promenades for walking and cycling. Today, 65% of internal movements are produced on foot. In this context, the Sustainable Urban Mobility Plan (SUMP) [8] in Bilbao plays a significant role; its main objectives are:

- Reducing air and noise pollution.
- Improving safety by reducing accidents and fatalities.
- Guaranteeing universal accessibility.
- Improving energy and transport (passengers and goods) efficiency.
- Contributing to improve the attractiveness and environmental quality of the city.

Of particular interest is the “Pedestrian mobility strategy” aiming to promote non-motorized modes of transport (especially pedestrian displacement) since these best suit the sustainable mobility objectives. Part of this strategy is the transformation of Moyuéa plaza, for exclusive use of public transport, pedestrians, and cyclists, prohibiting private traffic.

What Bilbao is aiming for. To reach its objectives, the city of Bilbao aims to obtain a global vision of the city in terms of sustainable mobility, to take decisions based on updated data (predicting the impact resulting from applied measures), follow a more agile decision-making process (facilitating communication between stakeholders involved in the definition and development of the SUMP), translate measures impact into health and life quality indicators and access data coming from scattered sources that is automatically collected and integrated.

The role of disruptive technologies in Bilbao. In the context of Bilbao, it is essential that decision-makers can easily access the most updated data; in this sense tools that facilitate the connection of data sources and the data harmonisation (leveraging common and well-defined data models) would support decision-makers in their daily activities. Once data is collected, a data catalogue (as a unique point of access to the data) would offer the capabilities to search data considering different criteria; among them, the possibility to filter the available data by for example the “transport mode" would allow the decision-makers to reduce the time they spend to identify the data they need. Facilitated setup and execution of simulations (for instance, to forecast impact on traffic, mobility patterns or SUMP’s KPIs resulting from a measure/policy applied) would support the decision-making process reducing the time spent in performing those simulations. Tools to create charts and graphs that summarise the status of mobility in the city from the sustainability point of view would allow the decision-maker to have, in a single view, the overall and relevant information to globally monitor the mobility in the city. On the other hand, the possibility to define and create customised KPIs and indicators would allow the decision-makers to fine-tune the dashboards with all the relevant information that they need to take into account in the planning of the mobility in the city. To this aim, checking if the data is updated would allow the creation of analyses and simulations based on correct information that represents the real status of the city, whereas pre-processing of collected data would reduce the time needed to setup the analysis and simulation for decision-making processes.

2.3 Helsinki

Helsinki, the capital of Finland, is a continuously evolving and developing city. In this sense a particular example is represented by the Jaétkaésaari area. The shore area of Jaétkaésaari, literally meaning “Dockers’ Island”, was previously used for industrial and harbour purposes; now it has gradually transformed itself into a residential area offering workplaces and services. At the same time, Jaétkaésaari is also a growing passenger and transport harbour due to its location (right adjacent to the centre of Helsinki). The harbour is the main connection between Helsinki and Tallinn, with growing mobility and opening of a new terminal in 2017. Annually 1 million private cars travel on the connection where a single main road leads in and out of Jaétkaésaari. This road feeds directly to the largest car commuting junction (70.000 cars daily) from the city centre to the western suburbs of Helsinki, creating interference. The Jaétkaésaari area is emblematic of the overall development Helsinki is facing, in particular, concerning mobility. In this context, to correctly cope with this evolution, the City of Helsinki’s traffic planning and traffic management need up-to-date and high-quality traffic information to support data-driven decision making. In addition, proactive and forward-looking approach is needed as the population of the metropolitan area grows and traffic situation changes.

What Helsinki is aiming for. In this context, the City of Helsinki aims to check the status of traffic and its development, analyse how traffic could evolve, perform traffic forecasts, simulate traffic planning and land use, check the development and implementation of new infrastructures and policies, develop a master plan for city development (e.g. land use, mobility, housing). To reach these objectives it is essential to establish a unique view and understanding among traffic planning and urban planning, allowing the exchange of information among different departments (overcoming information silos). In doing so, the city of Helsinki faces some issues related to the availability of different map layers with different information representations moving from a department to another, the lack of people with competences for demanding analysis, the lack of time to get deep understanding of data and problems related to obtain raw data to be analysed with external tools.

The role of disruptive technologies in Helsinki. A data catalogue as unique point of access that brings under the same umbrella the data produced by different departments would simplify the discovery and access of needed data, avoiding complications caused by scattered repositories managed by different departments of the same organisation. The data catalogue could leverage tools for the integration with existing ICT software and applications. This would allow on the one hand, the automatic check of information (e.g. automatic detection of inconsistencies in the data, such as missing mandatory fields, infringement of time constraints about updates) and on the other hand, the automation of repetitive tasks (e.g. extract relevant information and provide it in a more usable manner). Leveraging the data made accessible it would be possible to define pre-packaged simulations that need only minor operations to be executed (e.g. few parameters and/or initial input data). This would simplify the use of this kind of technology by personnel without specific competencies and skills who would be able to set up an entire simulation from scratch, and reduce the time needed and the acceptance of this technology, since the personnel will not spend too much time to learn how to use it.

2.4 Messina

The metropolitan area of Messina is one of the most extended urban areas in the south of Italy and the first in Sicily and counts over 620.000 citizens. In the city of Messina alone, there are over 250.000 inhabitants and most of them are commuters between Sicily and Calabria regions. Geographical peculiarities (the geographical shape of the city of Messina is stretched for 32 km beside the Tirrenian sea, and tight between its hills and the sea) and its role of main connection point between Sicily and the Italian peninsula have a huge impact on mobility in the city of Messina. The local transport system of the city consists of sea transport (hydrofoil and ferry boats fleets) and land transport (buses, tramway and rail transports network), operated by public and private companies. One of the main issues that affects both kinds of services (sea and land transport) is the lack of interoperability among the different departments of the Municipality that are involved for different reasons in the management of the mobility.

What Messina is aiming for. Concerning mobility, the main challenge of the city of Messina for the upcoming years is twofold: on the one hand, to build mobility services able to fulfil the needs of citizens, dwellers, commuters and visitors, allowing them to move around and through the city seamlessly; on the other hand, the challenge consists in optimising the management and interaction among the different mobility and monitoring systems and services available in the urban area of the city of Messina reducing the waste of resources and costs for the Public Administration. A particular attention is paid on light mobility (e.g. extension of the cycle network with new bike-lanes and links between the centre and suburbs zones of the city to spread the use of bicycle mobility [2]) and pedestrians (definition of an integrated system of pedestrian areas and paths).

The role of disruptive technologies in Messina. The different Departments of the Municipality would benefit of a unique data-access point to their data, avoiding the complication generated by the need of accessing scattered data sources (for instance, in the case of data hosted and managed in different repositories for the different departments). This would simplify the discovery of and access to the data needed by the decision-makers. In this context, tools to facilitate the connection to data sources (also from third parties) are vital. Data is the fuel of any activity related to analysis, simulation and the more information is available (not only in terms of amount but also in terms of variety), the more accurate and precise can these analysis and simulations be. In this context, advanced smart devices and virtual devices [7] (abstracted component characterized by specific high-level functionalities) offer the chance to access the needed information with the most appropriate frequency and accuracy, avoiding information overload and allowing a more efficient computation. In the management of urban mobility, analysis and simulations would support decision-makers in the identification of potential solutions (such as multimodal paths and possible intervention to increase public safety) [1] and hidden problems (such as related to public transportation and for planning maintenance interventions of road and public transportation vehicles). Customisable dashboards to represent the information a decision-maker needs would allow to obtain a clearer view of the status of mobility, supporting the decision-making process in the most appropriate manner. Finally, the possibility to share information (such as data, results of analysis/simulations, map layers, charts, graphs) with people working in the same or a different department would improve the collaboration and the efficiency of the decision-making process, overcoming inefficient communication and information silos.

3 CONCLUSIONS

Despite their specific peculiarities such as organisational approaches and mobility needs to be satisfied, the cities of Amsterdam, Bilbao, Helsinki and Messina have some commonalities in terms of potential application of disruptive technologies that can help their decision-making processes. The main aspect that emerged is related to the need of data, as a vital element to perform any decision-making activity; in this sense it is important to underline that here the need is related to the easiness of accessing the data, that in most of the cases is scattered, or represented using different data structures with non-uniform standards. Uniform access to the data drives to another common point among the four cities, that is the exploitation of the possibilities offered by simulation tools, in particular to forecast and predict the impact of decisions taken on traffic and mobility (such as the building of a new road, the creation of a LTZ). This kind of technologies would allow the decision-makers to better design mobility solutions and policies, giving the possibility to tackle complex problems and to evaluate the implications of new policies. The third common point is the data visualisation. Accessed data and results obtained from simulations and data analysis must be visualised in an easy-to-understand manner, this includes not only the data visualisation per se, but also the possibility of creating customisable dashboards in which the decision makers can arrange the information they need and represent it according to their preferences. From the result here summarised, it is possible to clearly identify a chain of needs with their corresponding solutions. The first link of the chain is the need of **accessing data**. Here tools facilitating the connection to data sources and the integration with existing IT systems can offer a valuable solution to overcome information silos and to build a unique data-access point to available data, allowing also the harmonisation of the data thanks also to common and well-defined data models and highlight the relevant information reducing the time to find it. The second link of the chain is the **analysis of the data** made accessible through the previous step and the execution of simulation. Here it is important to highlight that beyond the possibility to perform analysis and simulation, availability of tools that simplify and reduce the time needed to set them up play a key role. In this sense, pre-packaged simulations ready to use, that guide the users in their setup, and tools, that allow the creation of customised KPIs and indicators, represent an advantage for the decision-makers. The third and final link of the chain is the **data visualisation**. Here, tools (e.g. Wizards) guiding the users in the creation of charts, graphs, map layers, etc. offer the opportunity to speed up the decision-making process by reducing the time of interpreting and understating the information. At the same time, the possibility to visualise different data in the same view through customisable dashboards offers the chance of obtaining a bird's-eye view on the information that is relevant for each decision-maker, according to their specific needs. Considering the reported results, a final consideration can be made; even if cities could be characterised by a different IT maturity level, the most suitable way to effectively improve mobility decision-making processes is not a single technology, but a combination of disruptive technologies, that glued together unlock their respective potentialities and benefits.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 870338. This work was co-financed by the European Union - FSE, PON Research and Innovation 2014-2020 Axis I - Action I.1 "Dottorati innovativi con caratterizzazione industriale"

REFERENCES

[1] Lorenzo Carnevale, Antonio Celesti, Maria Di Pietro, and Antonino Galletta. 2018. How to conceive future mobility services in smart cities according to the fiware frontiercities experience. *IEEE Cloud Computing*, 5, 5, 25–36. DOI: 10.1109/MCC.2018.053711664.

[2] Alessio Catalfamo, Maria Fazio, Francesco Martella, Antonio Celesti, and Massimo Villari. 2021. MuoviMe: secure access to sustainable mobility services in smart city, (September 2021).

[3] Silvia Comunian, Dario Dongo, Chiara Milani, and Paola Palestini. 2020. Air pollution and covid-19: the role of particulate matter in the spread and increase of covid-19's morbidity and mortality. *International Journal of Environmental Research and Public Health*, 17, 12. ISSN: 1660-4601. DOI: 10.3390/ijerph17124487. https://www.mdpi.com/1660-4601/17/12/4487.

[4] Chiara Copat, Antonio Cristaldi, Maria Fiore, Alfina Grasso, Pietro Zuccarello, Santo Signorelli, Gea Conti, and Margherita Ferrante. 2020. The role of air pollution (pm and no2) in covid-19 spread and lethality: a systematic review. *Environmental Research*, 191, (August 2020), 110129. DOI: 10.1016/j.envres.2020.110129.

[5] 2019. *CYCLING MATTERS 2019, How Bicycles Power Amsterdam.* CITY OF AMSTERDAM.

[6] Alina Machidon, Maj Smerkol, and Matjaž Gams. 2020. Urbanite h2020 project. algorithms and simulation techniques for decision – makers. *Proceedings of the 23rd International Multiconference INFORMATION SOCIETY*, A, 68–71.

[7] Francesco Martella, Giovanni Parrino, Giuseppe Ciulla, Roberto Di Bernardo, Antonio Celesti, Maria Fazio, and Massimo Villari. 2021. Virtual device model extending ngsi-ld for faas at the edge. In *2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, 660–667. DOI: 10.1109/CCGrid51090.2021.00079.

[8] 2018. *Plan de Movilidad Urbana Sostenible (PMUS) 2015-2030 de la Villa de Bilbao, Fase II. Propuesta.* Ayuntamiento de Bilbao, Área de Movilidad y Sostenibilidad.

[9] Maj Smerkol, Žan Počkar, Alina Machidon, and Matjaž Gams. 2020. Traffic simulation software in the context of mobility policy support system. *In Information Society 2020.*

An Overview of Transport Modelling Approaches – A Use Case Study of Helsinki

Shabnam Farahmand
Forum Virium Helsinki
Unioninkatu 24, 00130
Helsinki, Finland
shabnam.farahmand@forumvirium.fi

ABSTRACT

In this paper a general view to transport planning approaches have been articulated with a focus on the simulation models. To this end, different analytical methods have been investigated with regard to the scope of target policies, geographic scales, and modelling techniques. The paper also provides an overview to the transport planning approaches which are specifically applied in the City of Helsinki in close relation to the land use policies. Besides, further discussions have been included to shed light on the approach URBANITE project is seeking. Although there is still a need for overcoming the challenges regarding data-driven decision-making, we see a potential in the project's approach to foster the use of disruptive technologies for accelerating the uptake of the evidence-based policies.

KEYWORDS

Transport planning, scales of analytics, policy-making, transport modelling in the City of Helsinki, simulation

1 INTRODUCTION

Transport planning plays a major role in defining the way public resources such as funds and spaces are used. Transport plans are mainly applied to understand the strategic capacity and consequences of high-level democratic decisions. Hence, it is important to consider the political and societal preferences of relevant stakeholders including citizens [1]. This also explains the urge for developing transparent, open-source, and simplified solutions in order to evoke citizen engagement and public participation [2]. Moreover, the advantage of transport planning models most probably lays in the fact that the scope of identified solutions by these models are inherently geographic [3]. Geographic analysis and tools speed up the uptake of new technologies due to the power and potential to provide evidence for interventions in transport planning [4].

In the following, the different approaches to tackle transport problems based on analysis levels will be addressed. In section 3, a schematic framework for transport planning approaches is suggested with the focus on analytical and simulation techniques. Furthermore, the transport planning techniques applied specifically by the City of Helsinki is included here. Section 4 discusses URBANITE project's global view and argues the advantages and challenges ahead of mobility decision makers.

Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia

2 TRANSPORT PLANNING APPROACHES

There are different approaches to analyze characteristics of a transport network and to evaluate the outcomes of the strategic and/or ad-hoc interventions with the transport. Ni [5] considers the geographic scales of transport planning models and proposes a framework which can enable multiscale traffic modelling which can be seen in Figure 1. In another study, Vassili [6] compares the transport analysis tools based on the scope and complexity of research area and highlights the importance of distinguishing between Analysis, Modelling, and Simulation (AMS) tools. Some of the tools for each scale of geographic analysis are already suggested in Figure 1. In addition to the geographic scale, the purpose of policy making processes to tackle a specific problem is also an important criterion in defining the right approach. Larger geographic scale of analysis can be chosen to support policy making with less data granularity [7]. However, it is reasonable to opt for micro-scale analysis when dealing with ad-hoc interventions in a specific area. This, on the other hand, becomes demanding on obtaining more detailed and comprehensive data.

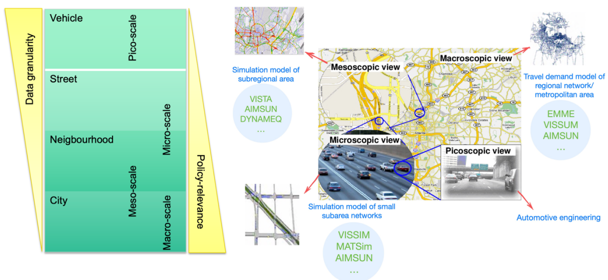

Figure 1. Scales of Transport Planning Approaches; Tools & Solutions

3 TRANSPORT PLANNING APPROACHES

De Dios Ortúzar and Willumsen [8] structured the transport planning approaches into five main stages as problem formulation, data collection, modelling and analysis, evaluation, and implementation of the solutions. In this paper, a new schematic framework is formulated based on Dios Ortúzar and

Willumsen's approach in Figure 2. The framework is modified in accordance with the approach of Helsinki Region Transport (HSL) and URBANITE's global view to provide a clear understanding of current applied techniques as well as a basis for the comparison of the two approaches.

Australian Road Research Board [9] categorizes the problem-solving techniques into analytical and simulation techniques. The research implicates that the analytical techniques are sort of closed form mathematical equations which provide statistical results such as forecasts and predictions. On the other hand, simulations are physical mathematical models, the results of which is to project objects moving around in a transport network. It is also possible to check the network state at different time stamps [9] & [10].

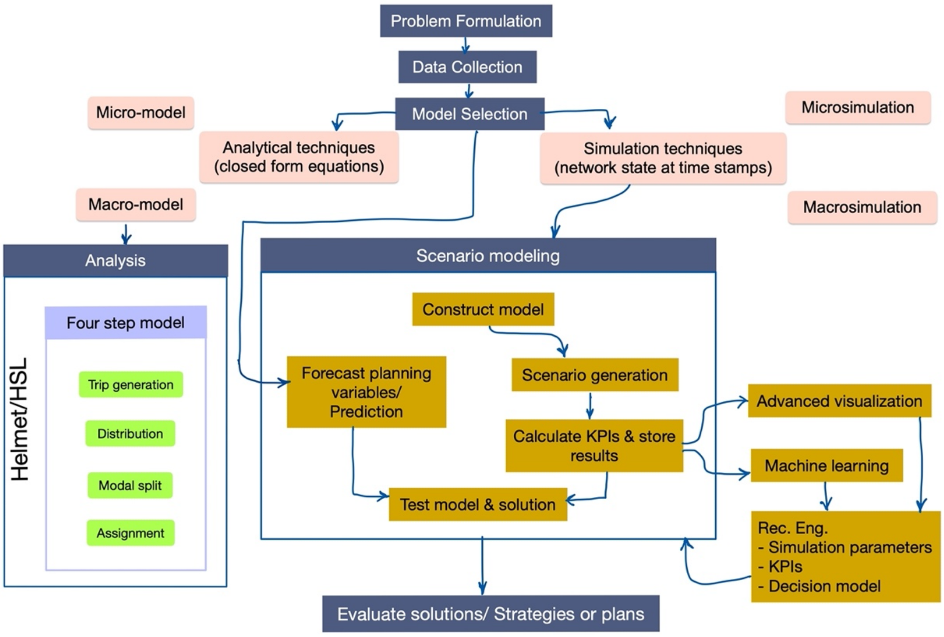

Figure 2. Proposed Schematic framework for Transport Planning[1]

4 TRANSPORT PLANNING – Use Case of Helsinki

The techniques used by the Helsinki Region Transport (HSL) follow an analytical approach to enable strategic transport and land use planning for the city. The model is called "HELMET" and is built with the help of proprietary tool EMME[2]. The statistical mathematical models in the field of transport models are usually referred as travel demand models when considered on a macro-level. These models have Four Step Transport Model (FSM) as the basis although they have evolved to more advanced levels to encompass the intelligence of models' agents [11]. The last version of HELMET model is therefore considering agent-based modelling (ABM) approach when it comes into trip chains analysis [12].

Helsinki Region Transport (HSL) developed its Sustainable Urban Mobility Plan (SUMP) for the City of Helsinki in 2015[3]. In particular, this plan focuses on 1) strengthening the strategic capacity and effectiveness, 2) integrating transport and land use, and 3) clarifying transport policy choices as well as the roles of different modes of transport.

According to the SUMP of Helsinki and on the basis of interviews performed with the City stakeholders, the interrelation between transport planning strategies as well as land use policies has been come into our focus frequently. Stover and Frank [13] suggested that development of transport and land use affect each other continuously in a cycle which is illustrated in Figure 3.

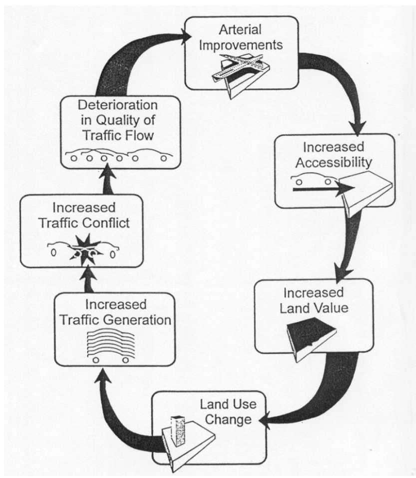

Figure 3. Transportation Land Use Cycle

[1] In blue: the main stages of transport planning processes; in yellow: URBANITE's global view
[2] https://www.inrosoftware.com/en/products/emme/
[3] http://www.bsr-sump.eu/good-example/helsinki-region-transport-system-plan-hlj-2015

Bearing this in mind, the proposed use case scenarios aim to find out the outcomes of the following decisions:

1. Intervention with the traffic network e.g., building a tunnel on the west harbor's junction to lead the main stream of heavy-duty vehicles caused by the arrival of ferries
2. Interventions with the land use in the area as it has been undergoing a lot of changes due to the constructions to turn the harbor into a dense residential area

The results of such analysis will help with understanding the causes of congestions and bottlenecks in the west harbor and serve as a tool for measuring the impacts of different policies on air quality and noise levels. Finally, the results will contribute to comprehending situational and statistical awareness which is one the main pillars of the City's Intelligent Transport System Development Programme 2030[4].

5 Discussions and Future Directions

URBANITE project aims to build microsimulation models which can help cities find out the outcomes of certain policies by applying new technologies and advanced techniques. Building transport models is demanding in terms of costs, time, data, and computation space requirements. However, URBANITE aims to take advantage of machine learning techniques as well of decision support systems to overcome these challenges. Hence, the models will be trained by the results obtained from simulations' input-output space exploration. Additionally, a recommendation engine will be built to provide decision makers with the relevant policies and KPIs tailored for their needs.

The approach facilitates data-driven decision making and will be fundamental in enabling real-time implementation and evaluation of solutions. Although there are still a lot of challenges regarding available data sources whether on the level of required infrastructure for gathering data or the quality of the available data. Recognition of the most relevant data sources and opening the data is a crucial step for the cities if they aim to realize evidence-based decision-making. The other challenge depends on the ability to include the benefits of all stakeholders esp. citizens in building technological solutions. In this regard, cities should come up with the ways to consider interests of all relevant beneficiaries and move towards participatory approaches.

ACKNOWLEDGMENTS

The URBANITE project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870338text here. Insert paragraph text here.

REFERENCES

1. Legacy, C., *Is there a crisis of participatory planning?* Planning theory, 2017. **16**(4): p. 425-442.
2. Peters, M.A., *Citizen science and ecological democracy in the global science regime: The need for openness and participation*. 2020, Taylor & Francis.
3. Lovelace, R., *Open source tools for geographic analysis in transport planning.* Journal of Geographical Systems, 2021: p. 1-32.
4. Jäppinen, S., T. Toivonen, and M. Salonen, *Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach.* Applied Geography, 2013. **43**: p. 13-24.
5. Ni, D., *Multiscale modeling of traffic flow.* Mathematica Aeterna, 2011. **1**(1): p. 27-54.
6. Alexiadis, V. and C. Systematics, *Integrated Corridor Management Analysis, Modeling and Simulation (AMS) Methodology*. 2008, United States. Joint Program Office for Intelligent Transportation Systems.
7. Allacker, K., et al., *Energy simulation and LCA for macro-scale analysis of eco-innovations in the housing stock.* The International Journal of Life Cycle Assessment, 2019. **24**(6): p. 989-1008.
8. de Dios Ortúzar, J. and L.G. Willumsen, *Modelling transport*. 2011: John wiley & sons.
9. Bennett, D., et al., *Guide to traffic management part 3: traffic studies and analysis.* 2009.
10. Shone, F., *City Modelling.* Medium, 2020.
11. McNally, M.G., *The four-step model*. 2007: Emerald Group Publishing Limited.
12. Parunak, H.V.D., R. Savit, and R.L. Riolo. *Agent-based modeling vs. equation-based modeling: A case study and users' guide*. in *International workshop on multi-agent systems and agent-based simulation*. 1998. Springer.
13. Stover, V.G. and F.J. Koepke, *Transportation and land development.* 1988.

[4] https://www.hel.fi/static/liitteet/kaupunkiymparisto/julkaisut/julkaisut/julkaisu-16-19-en.pdf

URBANITE: Messina Use Case in Smart Mobility Scenario

Francesco Martella
Engineering Department, University of Messina - ALMA Digit S.R.L.
Messina, Italy
fmartella@unime.it

Giovanni Parrino
Engineer - Municipality of Messina
Messina, Italy
nanni.parrino@gmail.com

Mario Colosi
Engineer - Municipality of Messina
Messina, Italy
colosimario96@gmail.com

Giuseppe Ciulla
Research & Development Laboratory
Engineering Ingegneria Informatica
Palermo, Italy
giuseppe.ciulla@eng.it

Roberto Di Bernardo
Research & Development Laboratory
Engineering Ingegneria Informatica
Palermo, Italy
roberto.dibernardo@eng.it

Marco Martorana
Research & Development Laboratory
Engineering Ingegneria Informatica
Palermo, Italy
marco.martorana@eng.it

Roberto Callari
Computer Engineer
Palermo, Italy
roberto.callari@outlook.it

Maria Fazio
Department MIFT, University of Messina - ALMA Digit S.R.L.
Messina, Italy
mfazio@unime.it

Antonio Celesti
Department MIFT, University of Messina
Messina, Italy
acelesti@unime.it

Massimo Villari
Department MIFT, University of Messina - ALMA Digit S.R.L.
Messina, Italy
mvillari@unime.it

ABSTRACT

The urban transformation and the changes that the world is undergoing lead, today more than ever, to the need to make faster and more timely choices in the field of mobility management. Technology is therefore essential for providing decision support tools that help managers and politicians to better manage cities. The European project URBANITE (Supporting the decision-making in URBAN transformation with the use of disruptive Technologies) aims to put in place a sustainable mobility with the support of disruptive and innovative technologies for this sector. The proposed study describes the URBANITE project with reference to the technologies and the strategies implemented in the city of Messina. As a partner and pilot use case, in the municipality of Messina, software tools have been created starting from a series of local data regarding traffic and public transport tracking. These tools allow technicians to quickly view traffic status or bottlenecks for public transport on a map.

KEYWORDS

Urban Transformation, Disruptive technologies, Urban mobility, URBANITE project, Decision making, Data Access, Data Analysis, Data Visualisation.

1 INTRODUCTION

In the context of Smart Cities it is crucial to pay attention to issues relating to mobility. Today Smart Mobility allows people to optimize their travels by reducing the stress associated with them, while Sustainable Mobility helps to protect the environment by improving the quality of life in Smart Cities. Institutions around the world are implementing policies that allow to decrease CO2 emissions. The issues of mobility and its optimization are therefore protagonists in the identification of these policies. In particular, the European Commission encourages projects in the field of Smart Mobility and Sustainable Mobility with H2020, Horizon Europe and the Next Generation EU programs. The URBANITE project was financed within the H2020 funding program. Among the objectives of URBANITE the main one is to promote the use of disruptive technologies in the nascent Smart Cities in technological terms through the use and analysis of Big Data, AI algorithms, etc. An innovative element, however, is that related to the promotion of innovative tools for participatory decision-making processes such as the Laboratory Social Policy (SoPoLab). The aim of the project is to provide the Stakeholders of the project with a series of innovative technological tools in order to support the decision-making processes of managers of public administrations and companies. Within the project there are four pilot cities: Amsterdam, Bilbao, Messina and Helsinki. In each of the pilots, the needs are studied and analysis tools developed which will then be applied to each of them. As regards the city of Messina, analysis were conducted on traffic and its effects on local public transport. This work describes the reference scenario and the actions implemented for the municipality of Messina within the URBANITE project regarding the purely Information Computer Technology (ICT) aspect. In particular, in section 2 the state of the art of the technologies studied and applied to achieve the objectives is described. Section 3 introduces the reference scenario. In section 4 the tools implemented will be illustrated, while in section 5 the final considerations and future developments are reported.

Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia

2 STATE OF THE ART

In [1] a case study concerning the home-office mobility of the University of Messina staff is discussed. The home-work commuting of public employees in the city of Messina is one of the main critical issues related to daily life. Traveling at particular times of the day causes both traffic congestion and pollution. Authors analyze different performance indicators to be used for the design and development of Smart Mobility services by adopting FIWARE technologies. After analyzing the travel habits of workers at the University of Messina, authors described how FIWARE can lead to an agile development of Smart Mobility services capable of minimizing traffic congestion, fuel consumption and CO2 emissions. In [2] authors describe the results of a Sustainable Mobility project in Messina. The presented application aims to encourage citizens to use low-impact vehicles instead of private cars. Through a partnership between different stakeholders a digital application to assign citizens electric bikes was developed, free of charge for a limited time period. Authors describe cyber security issues, both in terms of secure authentication for citizens that access the service and tracking of the whole assignment process. The flow is described from the user's request to the e-bike restitution. The adopted solution uses two-factor authentication (2FA) and Blockchain as the main technologies in the implementation phase. Innovative and advanced smart devices and virtual devices are described in [6]. Authors have designed, for one use case in the city of Messina, an abstracted component characterized by specific high-level functionalities. The system offers the chance to access the needed information with the most appropriate frequency and accuracy, avoiding information overload and allowing a more efficient computation. In this case it is important the access control and the security of the data. An interesting work for this purpose is described in [5]. In [3] authors show the use of customized generic Edge devices to carry out multiple activities at the same time, also focusing on how the proposed solution can lighten the work of cloud infrastructures. The presented concepts were implemented and tested in a real use case in the city of Messina by means Function as a Service (FaaS) paradigm. The proposed work allows users to perform multiple tasks on the same device. Applications such as vehicle counting, license plate recognition, object identification, etc. are proposed. In the considered use case two cameras were connected to a Raspberry PI 4 and the performance was compared. It is possible to connect different sensors to the proposed Edge devices and imagine each sensor as a different service. In [8] authors introduce a tool for studying mobility data. The basic principle is that technological innovation has led to the spread of various data tracking systems. The data are accumulated and can be used in various applications such as the analysis of mobility, urban planning and transport engineering. It is possible to use the data to extract information in matters relating to rough space-time trajectories, or by relying on statistical "laws" governing human movements [4]. However, authors do not neglect the attention to user privacy [7]. From the study and development comes an interesting Python library used in URBANITE for the analysis of mobility data in particular in Messina use case. From the state of the art it emerges that the city of Messina has been the subject of various scientific studies that have found practical application. Various national and European grants made it possible to achieve relevant innovations in the field of mobility. It is not clear how the data collected can be useful to administrators and managers in the decision making phase. This paper, therefore, want to synthesize and demonstrate how, thanks to URBANITE project, it is possible to put together what is already present in the systems of the city of Messina, creating the basis for the creation of new useful decision-making tools.

3 REFERENCE SCENARIO

The URBANITE project was created to provide communities with a long-term sustainable ecosystem model. Through a co-creation strategy we want to bring stakeholders (civil servants, citizens, etc.) closer to the use of disruptive technologies in the field of mobility. This model is supported with a data management platform and algorithms for data-driven decision making in the field of urban transformation. Furthermore, the model is validated by pilot mobility use cases in the context of the proliferation of sharing services. The URBANITE platform encapsulates the experiences of four pilot cities and acts as a junction point to create a unique analysis model for cities. Thanks to the platform it will be possible to have information regarding mobility that can be as a support in order to take serious technical and practical decisions.

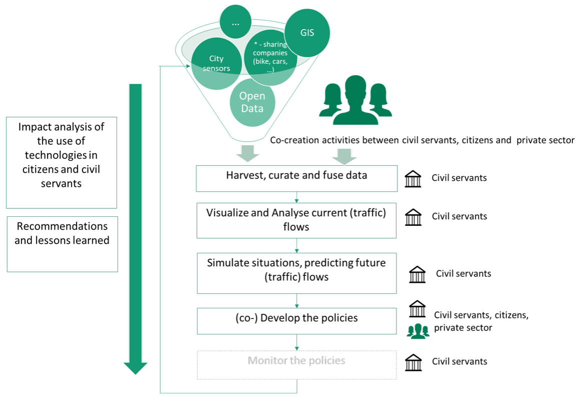

Figure 1: Urbanite Approach

In each pilot the data, useful for the mobility analysis, were analyzed and collected. The data considered funcional are collected on a single data storage. Thanks to different visualization and AI techniques/algorithms, the data were processed and made possible to create decision making tools that currently need validation (Figure 1). The use case regarding the city of Messina is described below.

3.1 Briefly on Messina Use Case

The metropolitan area of Messina is one of the most extended areas of the south of Italy, the first in Sicily and counts over 620.000 citizens. The city counts over 250.000 citizens and most of them are commuters between Sicily and Calabria. The local transport of the city of Messina consists of both sea transport (hydrofoil and ferry boats fleets) and land transport (buses, tramway and rail transports network). They are managed by public and private companies. The main issue that affects both kinds of services (sea and land transport) is the lack of facilities that can permit interoperability between different departments of the municipality and the communication with citizens and stakeholders. In order to overcome this problem, the Municipality of Messina is investing in intelligent infrastructures and services for the city and citizens. In particular, the main activities are focused on vehicle access detection in LTZ (Limited Traffic Zone) and pedestrian areas, centralised traffic management based on smart lights, traffic flows and analysis, incentives to use public transportation and video surveillance. URBANITE, for the city of Messina, is focusing on light and pedestrian mobility. Concerning the light mobility there are two main action lines:

(1) the extension of the cycle paths and the spread of bike mobility (but the main goal is to promote the use of bicycles and to offer better services to citizens)
(2) create new bike-lines and links between the centre and suburbs zones of the city.

Regarding pedestrian mobility, the objective is the definition of an integrated system of pedestrian areas and paths. Furthermore, from a wider perspective concerning public transportation, the city of Messina aims to extend the transport network in urban and extra-urban areas. The use case scenario in Messina (Figure 2) aims to evaluate the effects of the extensions of the public transportation services in terms of frequency, itineraries and stops on traffic and multi-modal transportation. In particular, a comparison of the impact on traffic between the different version of the public transportation network was performed. Moreover, the scenario includes an analysis of the suburban roads around the city of Messina (that represent an important connection with the surrounding towns) in terms of traffic congestion and connection with public transport network.

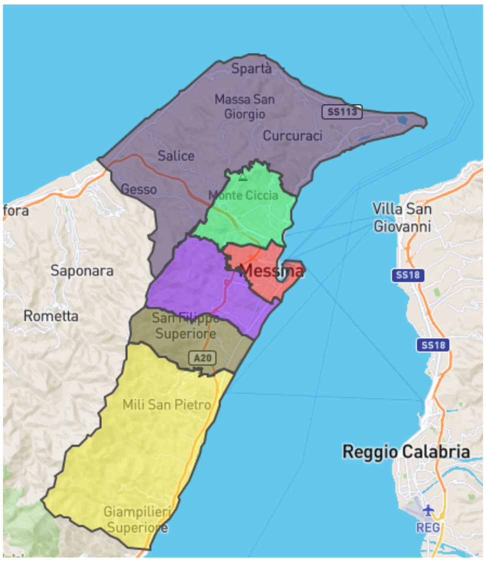

Figure 2: City of Messina

3.2 The URBANITE Architecture

The architecture created within URBANITE is made up of several abstract components that interact with each other. Thanks to the interaction between the different components, it is possible to provide all the tools necessary to achieve the objectives of the project. In Messina this architecture has been enriched by building new dedicated components, at the Edge level, which fully integrate with the existing Cloud ecosystem as shown in Figure 3, in which these components are highlighted.

In particular, for the Messina Edge Components, a local component called *Messina Data Storage* has been added. This component acts as a support for the parent component *Data Storage & Retrieval* (reported in URBANITE Cloud Components) through the *Data Harvester & Preparation* and is filled with data by the *Data Importer*. The *Data Processor* allows both to expose the data via Restful API and to process them ensuring correct formatting. Finally, within the *Urbanite UI*, three new specific components for the Messina use case have been built: *Messina Traffic Evolution*, *Messina Traffic Flows*, *Messina LPT Critical Areas*.

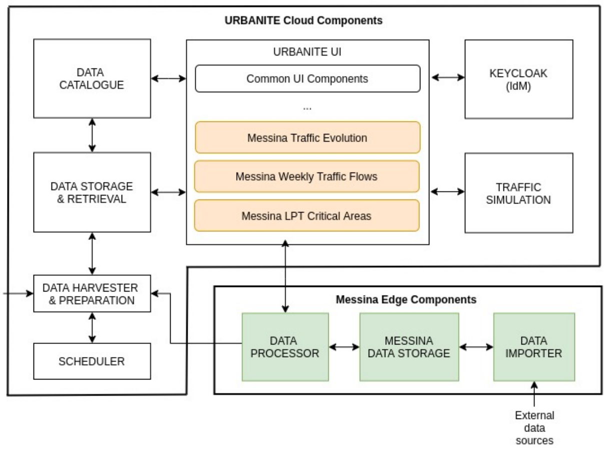

Figure 3: URBANITE Architecture - Messina

4 MESSINA IMPLEMENTATION

The use case scenarios described in Section 3 are accessible thanks to the functionalities provided by the URBANITE UI, the integrated URBANITE's framework at the UI level. The different analysis and visualizations provided aim to help the municipality's technicians in the extension of the current public transportation network. The tools allowing the users to interact with each visualization by filtering and querying the underlying data. Concerning the traffic congestion analysis for the municipality of Messina, Figure 4 depicts the temporal evolution of traffic flow on selected roads entering or leaving the city of Messina.

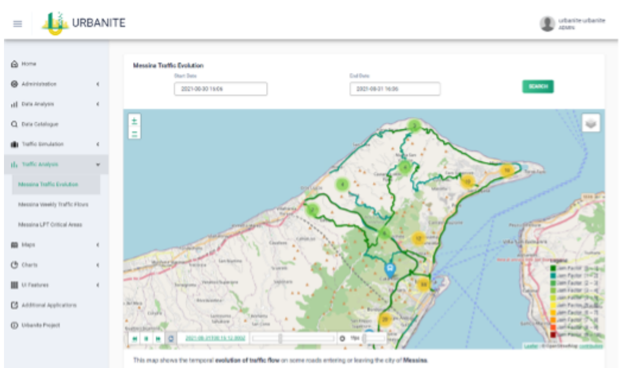

Figure 4: Messina Traffic Evolution

The traffic jam factor of each road, in a specific time of the day, is represented by the colour of the road itself, following the provided legend. Data used to this purpose are acquired and stored for real-time and historic analysis. Figure 5 illustrates the comparison analysis of the jam factors on two different roads of the city considering the time window of a week.

Figure 5: Messina Weekly Traffic Flows

The data source is the same of the previous analysis, but this time the purpose and the target users are people with a more technical background. For each road, if the road is bidirectional, the dashboard provides a chart for each direction using a different symbol for each one. The colors indicate the jam factor value. Finally, to identify areas of Messina where vehicles of public transportation are stationary for a certain time in a specific observation period, the heat-map analysis, depicted in Figure 6, is provided.

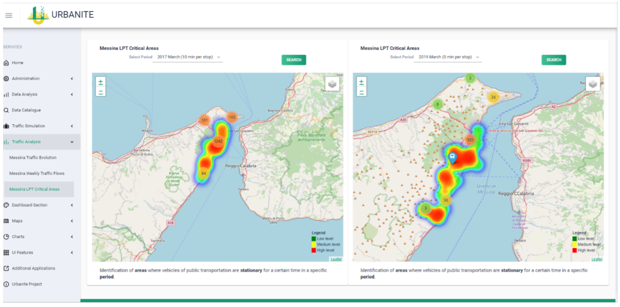

Figure 6: Messina LPT Critical Areas

To investigate if public transportation means use to be stationary in the same place for different time periods, the dashboard allows to compare two different time slots. In this case the data source is an historic database for the bus and tram position of the Local Transport Company. The data are elaborated with the scikit-mobility [8] Python library with the aim to obtain the heat-map visualization. In each described visualization, in order to have further information, the dashboard allows to visualize Points of Interest and Public Transport Stops on the map.

5 CONCLUSIONS

This paper describes the current state of the ICT systems put in place for the URBANITE project as regards the case of the Messina pilot. From the first results it is evident that, thanks to the use of data analysis and their appropriate visualization, it is possible to obtain information that is often difficult to understand. The visualization methods allow for immediate analysis and support decision-making policies. Thanks to the presented tools, in fact, it is possible to determine the effectiveness of the mobility policies used compared to the past, thanks to the historical harvested data, and possibly try to improve them. The next step will be to extend the functionalities. The scenario of each single pilot must be applied to all the case studies of the project. Moreover, it is necessary to improve smart algorithms in order to have responsive systems even in real-time. Finally, the system will make the APIs available for open-data, giving other scholars or stakeholders the possibility to carry out analysis or develop innovative solutions.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 870338. This work was co-financed by the European Union - FSE, PON Research and Innovation 2014-2020 Axis I - Action I.1 "Dottorati innovativi con caratterizzazione industriale". The authors thank the deputy Mayor of the Municipality of Messina Carlotta Previti and the administrative manager of the URBANITE project for the Municipality of Messina Dr. Placido Accolla for their support.

REFERENCES

[1] Lorenzo Carnevale, Antonio Celesti, Maria Di Pietro, and Antonino Galletta. 2018. How to conceive future mobility services in smart cities according to the fiware frontiercities experience. *IEEE Cloud Computing*, 5, 5, 25–36. DOI: 10.1109/MCC.2018.053711664.

[2] Alessio Catalfamo, Maria Fazio, Francesco Martella, Antonio Celesti, and Massimo Villari. 2021. MuoviMe: secure access to sustainable mobility services in smart city, (September 2021).

[3] Antonino Galletta, Armando Ruggeri, Maria Fazio, Gianluca Dini, and Massimo Villari. 2020. Mesmart-pro: advanced processing at the edge for smart urban monitoring and reconfigurable services. *Journal of Sensor and Actuator Networks*, 9, 4, 55. ISSN: 2224-2708. DOI: 10.3390/jsan9040055. http://dx.doi.org/10.3390/jsan9040055.

[4] Marta C. Gonzalez, Cesar Hidalgo, and Albert-Laszlo Barabasi. 2008. Understanding individual human mobility patterns. *Nature*, 453, (July 2008), 779–82. DOI: 10.1038/nature06958.

[5] Valeria Lukaj, Francesco Martella, Maria Fazio, Antonio Celesti, and Massimo Villari. 2021. Trusted ecosystem for iot service provisioning based on brokering. In *2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, 746–753. DOI: 10.1109/CCGrid51090.2021.00090.

[6] Francesco Martella, Giovanni Parrino, Giuseppe Ciulla, Roberto Di Bernardo, Antonio Celesti, Maria Fazio, and Massimo Villari. 2021. Virtual device model extending ngsi-ld for faas at the edge, 660–667. DOI: 10.1109/CCGrid51090.2021.00079.

[7] Yves-Alexandre Montjoye, Cesar Hidalgo, Michel Verleysen, and Vincent Blondel. 2013. Unique in the crowd: the privacy bounds of human mobility. *Scientific reports*, 3, (March 2013), 1376. DOI: 10.1038/srep01376.

[8] Luca Pappalardo, Filippo Simini, Gianni Barlacchi, and Roberto Pellungrini. 2019. Scikit-mobility: a python library for the analysis, generation and risk assessment of mobility data. (2019). arXiv: 1907.07062 `[physics.soc-ph]`.

Data commons in smart mobility – the road ahead?

Nathalie van Loon
Innovation Department
City of Amsterdam
Amstel 1, 1011 PN, Amsterdam, Netherlands
Nathalie.Loon@Amsterdam.nl

Rosalie Snijders
Faculty of Science, Information Studies
University of Amsterdam
Science Park 904, Postbus 94323, 1090 GH Amsterdam
Snijdersrosalie@hotmail.com

ABSTRACT

Mobility data collection and governance are mainly dominated by larger technology companies that gather all the data. Therefore, they also have exclusive control over what happens with the data. This calls for alternative data governance models. A viable alternative, introduced in recent years, is the data commons model. With this model, people can share their data on their own terms, while maintaining a certain amount of privacy. This model has been used with health data and scientific data, however, no viable example of a mobility data commons has thus far been found. This paper explores how local governments can facilitate a mobility data commons. And: is the commons a beckoning road for all of us?

KEYWORDS

Data governance, disruptive technologies, mobility data management, digital literacy, data commons, policy making.

1 INTRODUCTION

In the last decades, the concept of a smart city has grown in popularity both as a research subject and in government policies. Cities all over the world have started using technology to look for solutions that enable transportation linkages, mixed land uses, and high-quality urban services with long-term positive effects on the economy and sustainability of the city [1].

Smart cities are built on data. And one area where the generation and analysis of data have steadily increased is the mobility sector. App-based mobility services, like bike-sharing, scooter-sharing, peer-to-peer carsharing, and ride-hailing gather enormous amounts of information about how, when, and where people travel. And not only sharing apps, also other apps like weather apps or wayfinding apps generate data. Plus not only 'smart solutions' generate data but also 'regular' cars and bikes are becoming more and more mobile sensors in the city landscape by offering, to name just a few examples, 'tracking services' in case of theft, and cameras helping people to park.

In this context the City of Amsterdam aims to be a smart and mobile city, offering a large supply of mobility options; affordable, reliable, and accessible to everyone. However, most mobility data are enclosed by private companies, while the data generated by these services can be of great public value. As the city of Amsterdam is also part of the 'cities coalition for digital rights' and aiming to be a number one city in the protection of its citizens digital rights, Amsterdam is looking for good examples in the governance of data and cocreation of public value together with citizens, local stakeholders and SMEs.

Considering the context, and considering the role of the municipality, this paper explores the following question: can or should a local government organize a data commons in order to enable parties to share data in a trusted, fair and economic way, while observing privacy and security concerns? This paper therefore shortly explores the 'why' and 'how' and evaluates the applicability of a data commons as a disruptive technology and framework. This paper is based on existing literature and interviews with experts from the municipality of Amsterdam and is structured as follows: section 2 will start with some background information to support the research question. In section 3 the concepts of a smart city and data commons are explored, and section 4 will present the conclusions.

Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia

2 BACKGROUND

In the last couple of years, data have become a valuable asset to our economy. Some have claimed that the world's most valuable resource is no longer oil, but data [23, 49, 53]. A new form of capitalism has arisen where wealth is generated based on the accumulation, extraction, processing, and use of data.

The term Big Data has been on the rise since the start of the new millennium. Enabled by new and innovative technologies, companies can gather and analyse data from their customers or users and use it to their advantage. Digital data and information have become a critical economic, political, and social resource and most of this data is in the hands of just a few companies such as Amazon, Google, Facebook, and Apple [41, 43]. With this data, these few companies can have huge control and influence over human behaviour and societies. As a response, politicians, human rights movements and people in general have raised concerns about the misuse of their data. For many, it is not clear how much data these companies collect and what they do with it. As a result, people opt to not share anything with anyone and have started hoarding their data. However, data can be of great value for everyone if used in the right way. In the near future for instance, Artificial Intelligence will have to use data to play a role in the delivery of services [36]. If this data stay in the hands of big tech companies, the positive effects may never reach citizens.

As a digital rights city, therefore, it is of importance to look for new technologies that enhance public value and public benefit at the same time [43]. Citizens should have the power to decide on who they want to share their data with, under which rules, for what purpose and in a transparent manner. Data are (too) often regarded as a resource to be extracted for private profits, and technical developments have enabled technology firms to capture data from and about those who have not consented or have no viable alternatives. The view on data therefore must change from an asset that can offer a competitive advantage, to one of public infrastructure to ensure common welfare, which can be exchanged equally.

For this research, a smart city is defined as "A well defined geographical area, in which high technologies such as ICT, logistic, energy production, and so on, cooperate to create benefits for citizens in terms of wellbeing, inclusion and participation, environmental quality, intelligent development; it is governed by a well-defined pool of subjects, able to state the rules and policy for the city government and development". The city of Amsterdam has already become an example of how a smart city strategy can be implemented. With the above mentioned definition in mind, the main goal of smart cities is to improve the quality of life for its citizens in a sustainable way. At the same time, citizens also have the potential to be the main component of data acquisition. With the use of smartphones, the citizens can act as human sensors and help gather enormous amounts of data [29]. ICT can act as a platform to collect information and data to promote an improved understanding of how a smart city is functioning in terms of services, consumption, and lifestyle. Especially with mobility data, the input of citizens can be of great value [30].

While the potential of big data is explored on a daily basis in the development of new and possibly disruptive technologies, the potential societal disruption and ethical concerns attract less attention or even denial and/or apathy. This while multiple studies show that, with the creation of intelligent mobility systems in smart cities, the potential for intrusive surveillance is increased [31] and that the types of data used are privacy-sensitive [32]. Location history data, for instance, can act as an identifier of its users [33, 34]. Also bias in data can be a multiplier of societal injustice, as the Dutch 'toeslagenaffaire' [35] has shown, framing approximately 26.000 parents as possible fraudsters, based on their (second) nationality. Also multiple organizations may have multiple policies and rules regarding the protection of the data of their users. However, this is not always transparent - while it may lay in everyone's interest to share this data [36]. Therefore, one of the main challenges of the use of big data are privacy, transparency, and bias.

3 Data Commons

There are various definitions in use for commons and also for data commons. In general, the Nobel prize winning work on commons by Elinor Ostrom in 1990 is used as a reference for any such definition. Ostrom successfully described the commons as a governance model rather than open access to resources and introduced the commons as a framework to value various historical and contemporary social movements. In short one can define the commons as a commonly owned and managed (common pool) resource. More elaborate, Ostrom identified 8 design principles of stable common pool resource management in her groundbreaking work 'Governing the commons. The evolution of institutions for collective action.' [3, 6, 18, 19, 27].

3.1 Design principles data commons

Principles can be described as general rules and guidelines which a system architecture must follow to be as productive and cost effective as possible. Principles help guide the use and deployment of an architecture. Also principles may help identify concerns stakeholders might have that a system can address. Each principle should have a rationale and implication associated with it. This can help with promoting the acceptance and understanding of the principles [10, 25]. Here, we adapted and 'translated' 7 of Ostroms 8 design principles - in a first attempt - to rationales and implications for data commons [5].

(1) Define clear group boundaries:
- *Rationale:* Who can use the data should be clearly defined and should be easily identifiable
- *Implication:* An individual using the commons may require identifying information before allowing access to the commons. Additionally, the data sets should be easily identifiable. With this in place, poaching can be easily detected [23].

(2) Match rules governing the use of common goods to local needs and conditions:
- *Rationale:* The rules of governing the data commons should be matched to the local needs of the users. Since no data commons and its environment are the same.
- *Implication:* Setting up the rules and guidelines of the use of the commons should include the local users of the commons. Therefore, citizen participation is a crucial part of a successful commons.

(3) Ensure that those affected by the rules can participate in modifying the rules:
- *Rationale:* Both the data producer and user should be able to benefit from the data commons and be protected.
- *Implication:* All parties within a data commons should be able to change the conditions of the data commons, with agreement from all parties. The use and production in the data commons should always be in balance.

(4) Make sure the rule-making rights of community members are respected by outside authorities:
- *Rationale:* The rules and regulations of the commons should be respected by the local authorities, must be recognized as legitimate by the authorities.
- *Implication:* Local authorities shouldn't be able to change the rules without the consent of the parties involved.

(5) Develop a system, carried out by community members, for monitoring members' behaviour:
- *Rationale:* Monitoring of the data commons is needed to ensure that the data is used fairly.
- *Implication:* Unauthorized use of the data should be detected. In the case of a data commons, this could be a moderator, since the commons are not in a physical place. Ideally, this is done by the user community.

(6) Use graduated sanctions for rule violators:
- *Rationale:* Users and producers in the data commons that violate its rules should not be banned directly.
- *Implication:* A gradual system needs to be set up.

(7) Provide low-cost accessible means for dispute resolution:
- *Rationale:* When issues within the commons come up, the dispute would has to be resolved in an informal, cheap, and straightforward manner. This way problems are resolved, rather than ignored
- *Implication:* A process for conflict resolution should be created that is perceived as fair by all users of the data

commons. A mechanism for rule enforcement and for dealing with violators needs to be set up and discussed by all involved parties.

Concerns

The incorporation of the above-mentioned design principles can be a measure of success when organizing a data commons. But can they also be used to address the concerns the relevant stakeholders might have?

3.2 Citizen participation

Since citizen participation is a necessary step when organizing a datacommons and is essential for two design principles of a successful data commons, a major concern when it comes to a local government organizing or facilitating a datacommons is the participation of citizens. Is this a 'contradictio in terminis' or can and should the government act as a facilitator or incubator? Looking at the participation ladder by Arnstein [2] there is, indeed, a world to win, also calling for a different role of the government: a 'co-creating government' or 'co-city'.

Figure 1: Levels of participation

Transparency

Another important concern is transparency; in order to achieve a successful mobility data commons, the municipality needs to be transparent about every part of the data commons. To achieve full transparency, openness of all operations within the data commons is required, so that citizens if needed, can hold the consumers of the data accountable and are allowed to withdraw their consent [24]. However, measuring transparency within a data commons can be a tricky task. The question is not only how much information is available and under which terms, but is also a question of equality in the accessibility and usability of that information. Transparency is increased when the data within a data commons is given a proper context and, therefore, its users can use and understand the data without confusion. Transparency should cover all of these aspects of data access: physical access, intellectual access, and social access [13]. In the case of a data commons, physical access can refer to the ability to reach the content of the commons, social access is the ability to share the content of the commons and intellectual access is the ability to fully comprehend the content [7, 4], sometimes also referred to as 'digital literacy'. Not only in Amsterdam, but in more cities in the digital rights coalition, the Covid-19 pandemic and subsequent lockdowns showed that a lot of families don't have access to technology when public services like libraries and schools are closed. And how can Amsterdam residents take ownership over their data if they don't have access to technology, know where to access their data or how to object to their data being used? By introducing a 'digital agenda' [41] the city of Amsterdam is working on overcoming this divide and promoting and protecting digital rights, yet agency is complex and scattered. Also, the use of data and which algorithms are used should always be disclosed to the contributors of the data. Amsterdam has made a first step by introducing an 'Algorithm register' [16]. But can a commons be organized in such a way that no one has access to a contributor's data without their permission?

Monitoring and validating

This also raises the question if local governments can organise the monitoring of the use and validation of data. A solution could be implementing an interoperable context-aware meta-databased architecture [15]. This type of architecture is context-aware and allows permissions and policies to be attached to the data. Additionally, due to its flexibility, trust norms can be changed and can account for increased transparency and accountability. This is an architecture that associates data with user permissions and policies which enables any consumer to handle the data in a way that is consistent with a contributor's wishes [21]. This is a method that could increase accountability in a decentralized data ecosystem like a data commons. However, this method does thus far not provide a way for community members to contribute to monitoring the behavior within the community.

Concerns

Interoperability is a practical, yet very prominent concern when organizing data commons [7, 8] since a data commons is not only about access to data, it is also a platform for data experimentation and interaction. Technically, a data commons is a repository of personal manifests that describes the access and usage rights of all data generated by an individual within a digital service. Therefore, the data commons must regulate relationships between the organizations and individuals that use and share ownership of the data. This way, data commons help citizens having a say in what data they want to share and under which conditions. Also data commons could provide users easy access to their own data, information about who has access to their data and what they could do with this information. However, for this to be successful also trust needs to be built between the different parties participating in the commons. As our last concern we raise the question on the definition and the narrative. The commons, although part of an important and impactful historical movement, that, amongst others, created the guilds in the Middle Ages, the common land movement in the UK and, more recently, knowledge commons Wikipedia [11], mutuals like 'broodfondsen' in the Netherlands and citizen energy communities in most European countries, are not part of our current, dominant, narrative. Although he European Union and Dutch government have legal frameworks in place for several types of commons - in housing and energy for instance- no real understanding of the potential public value or even clear definition of a data commons currently exists.

4 CONCLUSION

For now, the larger technology companies dominate the data collection in the area of mobility. As a result, these companies have exclusive control over what happens with the data the citizens of a city generate. In this paper we described how this ‘enclosure’ of data by big tech builds a powerful value driven case for cocreating and/or facilitating commons in mobility data as a local government. Although a clear pathway on how to organize a mobility data commons is not yet available, the road ahead is one of cooperation, building trust between participants and experiment. By taking it one step at a time, setting clear boundaries and rules that are understood by partners involved and, obviously, involving citizens in every step. However, considering digital literacy and other possible constraints for citizen participation, careful thought on how to involve citizens -for a longer period- is paramount. One suggestion would be to just ‘follow the music’: there is a vibrant movement of active citizens communities and SMEs in town, how can the local government cooperate towards the creation of a data commons in mobility as a spill-over effect from these efforts? This way data commons can prove to be an alternative for apathy and distrust in big tech, contributing to a strong and growing narrative on local cooperation.

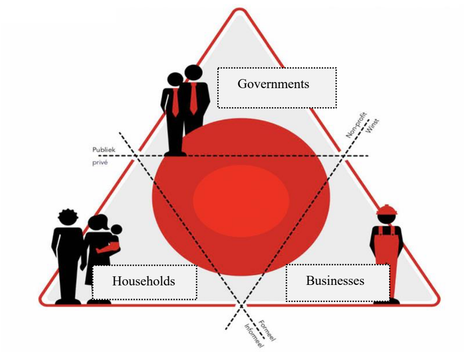

Figure 2: the third sector model [37]

ACKNOWLEDGMENTS

This paper is partly based on interviews with public servants and experts at the city of Amsterdam, a research done by Rosalie Snijders in Q2 of 2020. During this period Rosalie was an intern at the city of Amsterdam, writing her master thesis, supervised by Nathalie van Loon, working at the city of Amsterdam and Leon Gommans and Frank Nack, both working at the University of Amsterdam. This paper is an adaptation by Nathalie van Loon, written in the context of the Urbanite project: https://urbanite-project.eu/, under grant agreement #870338.

REFERENCES

[1] V. Albino, U. Berardi, and R. Maria Dangelico. Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, 2015.

[2] S. R. Arnstein. A Ladder Of Citizen Participation. *Journal of the American Planning Association*, 1969.

[3] B. J Birkinbine. Commons praxis: Toward a critical political economy of the digital commons. *TripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 16(1):290–305, 2018.

[4] G Burnett, PT Jaeger, and KM Thompson. The social aspects of information access: The viewpoint of normative theory of information behavior. Library & Information Science Research, 30(1):56–66, 2008.

[5] https://en.wikipedia.org/wiki/Elinor_Ostrom

[6] J.B. Fisher and L. Fortmann. Governing the data commons: Policy, practice, and the advancement of science. Information & Management, 47(4):237 – 245, 2010.

[7] R. L. Grossman, A. Heath, M. Murphy, M. Patterson, and W. Wells. A case for data commons: Toward data science as a service. Computing in Science Engineering, 18(5):10–20, 2016.

[8] R. Grossman. A proposed end-to-end principle for data commons, 2018 (accessed April 5, 2020).

[9] https://assets.amsterdam.nl/publish/pages/964754/agenda_digitale_stad_tussenrapportage_2019_-_2020.pdf

[10] Van Haren. TOGAF Version 9.1. Van Haren Publishing, 10th edition, 2011.

[11] I. Abaker T. Hashem, V. Chang, N. Badrul Anuar, K. Adewole, I. Yaqoob, A. Gani, E. Ahmed, and H. Chiroma. The role of big data in smart city. International Journal of Information Management, 36(5):748 – 758, 2016.

[12] C.Hess and E. Ostrom. Introduction: an overview of the knowledge commons. Understanding knowledge as a commons: from theory to practice., 2006.

[13] C. Humby. Data is the new oil. Proc. ANA Sr. Marketer’s Summit. Evanston, IL, USA, 2006.

[14] P.T. Jaeger and J. Carlo Bertot. Transparency and technological change: Ensuring equal and sustained public access to government information. Government Information Quarterly, 27(4):371 – 376, 2010. Special Issue: Open/Transparent Government.

[15] P. T. Jaeger and G. Burnett. Information worlds: Social context, technology, and information behavior in the age of the Internet. Routledge, 2010.

[16] S. Maguire, J. Friedberg, M. H.Carolyn Nguyen, P. Haynes. A metadata-based architecture for user-centered data accountability. Electronic Markets, 2015.

[17] https://ai-regulation.com/amsterdam-and-helsinki-launch-algorithm-and-ai-register/

[18] H. Mehr. Artificial Intelligence for Citizen Services and Government. Harvard Ash Center Technology & Democracy, 2017.

[19] E. Ostrom. Governing the commons: The evolution of institutions for collective action. Cambridge university press, 1990.

[20] E. Ostrom, R. Gardner, James Walker, J. M Walker, J. Walker. Rules, games, and common-pool resources. University of Michigan Press, 1994.

[21] F. Pasquale. From territorial to functional sovereignty: The case of amazon. Law and Political Economy, 6, 2017.

[22] E. Politou, Efthimios Alepis, and Constantinos Patsakis. Forgetting personal data and revoking consent under the gdpr: Challenges and proposed solutions. Journal of Cybersecurity, 4(1):tyy001, 2018.

[23] B. Prainsack. Logged out: Ownership, exclusion and public value in the digital data and information commons. Big Data and Society, 2019.

[24] N. Purtova. Health Data for Common Good: Defining the Boundaries and Social Dilemmas of Data Commons, pages 177–210. Springer International Publishing, Cham, 2017.

[25] S. Ranchordás. Nudging citizens through technology in smart cities. International Review of Law, Computers and Technology, 2019.

[26] E. Schlager. Common-pool resource theory. Environmental governance reconsidered: challenges, choices, and opportunities, pages 145–175, 2004.

[27] S. Spiekermann, Alessandro Acquisti, Rainer Böhme, and Kai-Lung Hui. The challenges of personal data markets and privacy. Electronic markets, 25(2):161–167, 2015.

[28] J. Yakowitz. Tragedy of the data commons. Harvard Journal of Law & Technology, 25(1):1, 2011.

[29] M. Srivastava, T. Abdelzaher, and B. Szymanski. Human-centric sensing. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370(1958):176–197, 2012.

[30] J. Vázquez Salceda, S. Álvarez Napagao, J. A. Tejeda Gómez, L. Javier Oliva, D. Garcia Gasulla, I. Gómez Sebastià, and V. Codina Busquet. Making smart cities smarter using artificial intelligence techniques for smarter mobility. In SMARTGREENS 2014: proceedings of the 3rd International Conference on Smart Grids and Green IT Systems, pages IS7–IS11. SciTePress, 2014.

[31] M. Büscher, P. Coulton, C. Efstratiou, H. Gellersen, D. Hemment, R. Mehmood, and D. Sangiorgi. Intelligent mobility systems: Some socio-technical challenges and opportunities. In Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 2009.

[32] Y. Alexandre De Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel. Unique in the Crowd: The privacy bounds of human mobility. Scientific Reports, 2013

[33] C. Bettini, X. S. Wang, and S. Jajodia. Protecting privacy against location-based personal identification. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2005.

[34] Hui Zang and Jean Bolot. Anonymization of location data does not work: A large-scale measurement study. In Proceedings of the 17th Annual International Conference on Mobile Computing and Networking, MobiCom ’11, page 145–156, New York, NY, USA, 2011. ACM

[35] https://nl.wikipedia.org/wiki/Toeslagenaffaire

[36] Lina Zhou, Shimei Pan, Jianwu Wang, and Athanasios V Vasilakos. Machine learning on big data: Opportunities and challenges. Neurocomputing, 237:350–361, 2017.

[37] V. Pestoff, Third sector and cooperative services -An alternative to privatization. Journal of consumer policy 15(1): 21-45, 1992.

URBANITE Mobility Data Analysis Tools

Ignacio (Iñaki) Olabarrieta[†]
TECNALIA, Basque Research and Technology Alliance (BRTA),
P. Tecnológico Bizkaia, Ed. 700, 48160 Derio, Spain
ignacio.olabarrieta@tecnalia.com

Ibai Laña
TECNALIA, Basque Research and Technology Alliance (BRTA),
P. Tecnológico Bizkaia, Ed. 700, 48160 Derio, Spain
ibai.lana@tecnalia.com

Urrotz Larrañaga
Bilboko Udala ITS Engineer
Ernesto Erkoreka Plaza, 12, 48007 Bilbo, Bizkaia
ularranaga@bilbao.eus

Sergio Campos
TECNALIA, Basque Research and Technology Alliance (BRTA),
P. Tecnológico Bizkaia, Ed. 700, 48160 Derio, Spain
sergio.campos@tecnalia.com

Raquel Gil
Bilboko Udala Mobility Management Deputy Director,
Ernesto Erkoreka Plaza, 12, 48007 Bilbo, Bizkaia
raquel.gil@bilbao.eus

Shabnam Farahmand
Forum Virium Helsinki, Unioninkatu 24, 00130 Helsinki, Finland
shabnam.farahmand@forumvirium.fi

ABSTRACT

The decision-making process in the policy making should rely on data driven evidence, in most of the cases the raw data needs to be processed to transform it into actionable information. For this purpose, several tools have been developed within the URBANITE project to transform urban mobility data into usable information. Specifically: (1) traffic prediction models based on historical data, (2) Origin-Destination (OD) matrix estimation models and (3) a methodology to analyse the locations visited in several trajectories.

KEYWORDS

Traffic prediction, Origin-Destination Matrix Computation, Data Analysis, Artificial Intelligence.

1 INTRODUCTION

URBANITE project goal is to provide tools for the decision-making in the urban transformation field using disruptive technologies and a participatory approach. These tools should aid the process of taking decisions guiding it on data evidence. The main features of the URBANITE architecture include:

- **Modularity**, i.e., each component provides specific functionalities and exposes clear interfaces,
- **Adaptability** to heterogeneous city and region contexts and ICT maturity levels, from complete implementation to complementary add-on components.
- **Interoperability**, i.e., vertical, and horizontal interoperability among modules and with existing systems.

And using the European standards as much as possible.

The main elements that URBANITE offers are the following:

[†] Corresponding Author

Information Society 2021: 24th international multiconference, *4–8 October 2021, Ljubljana, Slovenia*

Social Policy Lab – an environment to promote digital co-creation with methodologies and methods to support the communication among public servants, private companies, and citizens. The aim of the Social Policy Lab is to develop joint ideas and to produce co-creation guidance for policies. **Data Management Platform** – to provide automatic support to the whole data processing chain and its life cycle, starting with the collection process all the way up to the use of the data. **Decision-Making Support System** – powerful tools which combine multiple data sources with advanced algorithms, a simulation engine, a recommendation, and visualisation system. These tools include predefined analysis pipelines to be used by non-technical users, intuitive and understandable visualisations, and setups to perform simulations of new mobility policies and situations that allow their evaluation. URBANITE is implemented in four different use cases: Amsterdam, Bilbao, Helsinki, and Messina.

The analysis tools that are presented in this communication belong to the Decision-Making Support System. More concretely they belong to the set of algorithms designed to obtain information from the historical data stored in the URBANITE Data Management Platform. The results obtained from these algorithms can be used to understand better what is the state of the mobility at a given time, or, alternatively, they can be used as input for simulations of new policies.

Among all the tools within the Decision-Making Support System three components are explained in this communication, namely: traffic prediction, OD matrix estimation, and trajectory location analysis. These components are discussed in the following sections, sections 2-4. This communication ends with some concluding remarks in section 5.

2 TRAFFIC PREDICTION

Road traffic forecasting has been a topic of study since the sixties [1] when time series analysis methods where mainly used [2][3][4]. In the last two decades, heuristic machine learning methods [5] started being used allowing to find more complex relations within the traffic data. Nowadays traffic prediction component has become one key tool for any ITS system. The component developed within the URBANITE project can forecast what is the traffic flow that a sensor within the city would measure for a given set of features.

The set of features at the time of this communication include the day of the week and the time of the day but other ones are in the process of being incorporated. Some of these features include if the day of the forecast is a bank or school holiday, weather features (precipitation and temperature), the arrival of ferries to Helsinki port or sport events (soccer games) in Bilbao (using the method developed in [6]). Note that the approach within URBANITE is not to consider previous measurements as features since the data is not available in real time. Therefore, this approach can be considered as long-term prediction because the predicting horizon is only limited by the accessibility of external features (i.e., access to a weather prediction for example).

Figure 1: Integrated tool to perform traffic prediction showing the Helsinki use case.

The web portal to the integrated tool can be seen in Figure 1, the tool allows to train a new model, to perform a prediction and to visualize the results.

The process of performing the training implies that the user needs to choose the following:

- The regression model type, two options are available: random forest [7] and distribution inference [8] (only for features with discrete values).
- The number of features to consider: 1. considers only the day of the week, 2. also considers the time and so on.
- The time resolution, typically either 5 or 15 minutes. This is the aggregation period on which the individual counts of vehicles moving over the sensor are combined to produce a time series.
- The traffic sensor, this is chosen by selecting the available sensors within a map.
- The period of the training data, the period can be chosen from the available data within Data Management Platform, being able to change this period allows for instance to avoid choosing the anomalous period due to the restrictions due to COVID-19. In addition, a percentage of the training data can be reserved to test the goodness of the model, this percentage can also be specified.

Once a model is trained this can be used to perform a prediction, there are different ways to perform this, one way is to use the URBANITE web visualization tool to choose a given date and perform the prediction for either the following 24 hours or the following 7 days. Alternatively, specific set of features can be feed to the model using the REST Web service in JSON format to obtain a result at a given instant of time.

An example of the result can be seen in Figure 2 where in addition to the prediction (red line) the confidence interval is shown (orange band). The details of how to compute the confidence interval are explained in [9]. In the Figure the result of the prediction for a week is shown, where the peaks for the different days are clearly visible, including the difference in the pattern due to the weekend (fourth and fifth peaks in the series).

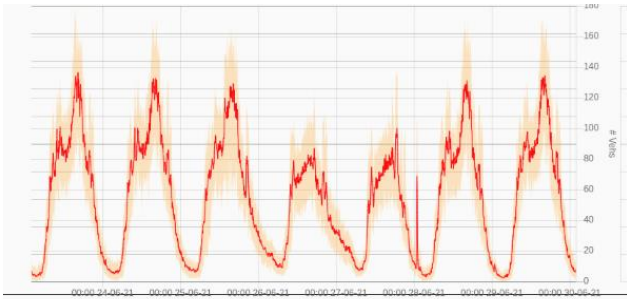

Figure 2: Detailed of the visualization for the result of the traffic flow prediction for 7 days including the confidence interval.

3 OD MATRIX ESTIMATION

The OD Matrix estimation works in a similar way than the prediction module. In this case we use data from bike rental city service, specifically we consider the origins and the destinations of each one of the rentals. These are both temporally and spatially aggregated by providing the time resolution (the same way as for the traffic prediction) and by providing a set of geographic areas where to aggregate the origins and the ends of each rental. These areas can be specified either via a GEOJSON or by specifying a set of points, the URBANITE web can be used to obtain the Voronoi areas [10] associated with those points and use those to perform the spatial aggregation.

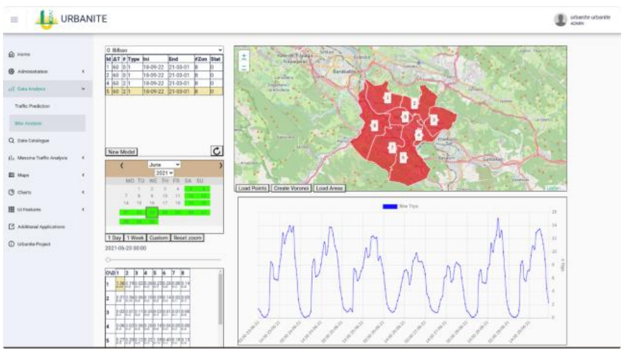

Figure 3: Integrated tool to perform OD matrix estimation for the Bilbao use case.

Training a model to perform OD matrix estimation implies choosing a regression model type, the number of features (in this case 0 can be chosen, which implies the use of only spatial information), the time resolution, and the period of training data.

The result of the estimation, at a specific instant of time, consists of a square matrix of size NxN where N is the number of different areas considered for the spatial aggregation. The web tool within URBANITE allows to compute and visualize these estimations for all the instants within a period (typically a day or a full week). In Figure 3 a detail of the web tool is shown where it can be seen the result at a given instant in the form of a matrix (lower left hand side) and the time evolution of one of the matrix components for a whole week (lower right hand side).

It is worth mentioning that this process to estimate OD matrixes, by means of the use of regression algorithms, have the capability to generalize the values measured obtaining results even in regions of the feature space where no values have been obtained yet.

4 TRAJECTORY LOCATION ANALYSIS

Finally, the last component that we explain in this communication consists in a tool able to analyze not only the origin and the destination of trajectories but also what happens in between. More specifically, and to fix ideas, we can think this tool's goal to be obtaining the points more popular to visit in a trajectory. The processing consists in two different phases: the cleaning phase, and the aggregation phase.

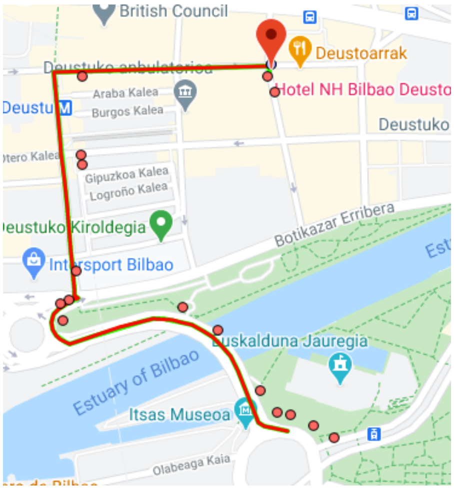

Figure 4: Result of the cleaning phase for a set of GPS points obtained from a single bike city rental in Bilbao.

The cleaning phase is a crucial phase when processing GPS data obtained from affordable, not very accurate sensors or in areas with tall buildings (urban environment) where the multi-path of the satellite signal can increase the noise of the measurements. The purpose of this phase is to align the obtained measurements with the navigational road network, i.e., the possible allowed positions for the vehicles. In URBANITE, Hidden Markov models [11] are used in this phase. Moreover, this process provides an additional result, which consists in the most likely points between measurements.

The second phase, the aggregation phase, compares the points obtained in the cleaning phase for all the trajectories. Probably the simplest of these aggregations is to compute the number of times a location is visited independently of the trajectory it belongs. The result of this process applied to the trajectories of the bike city service in Bilbao is shown in Figure 5.

Other types of aggregations can also be performed, as for example most likely points to be visited depending of the day of the week and the time of the day, the most popular chain of consecutive points visited, the longest route accomplished, etc.…

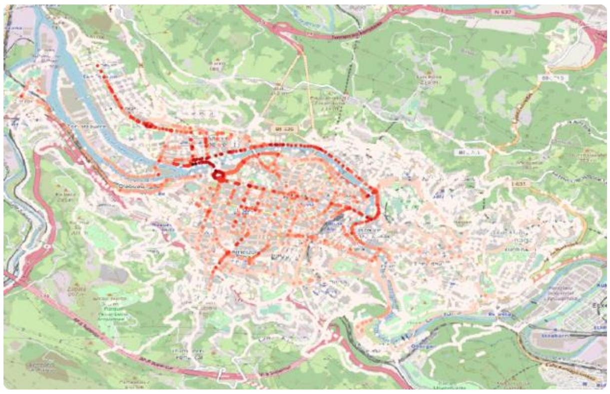

Figure 5: The most popular points corresponding to trajectory locations are labelled with darker color.

5 CONCLUSIONS

In this paper we have introduced three components developed within the URBANITE project to convert data into information. The first component is designed to obtain a prediction of the typical traffic flow at a particular sensor location given a set of features, the second component aims to produce OD matrixes from bicycle data and finally the last component consists in a methodology to analyse trajectory locations. These results have been achieved during funding project from the European Union's Horizon 2020 research and innovation programme under grant agreement #870338.

REFERENCES

[1] I. Laña, J. Del Ser, M. Velez, and E. I. Vlahogianni, (2018) *Road traffic fore-casting: Recent advances and new challenges*, IEEE Intelligent Transportation Systems Magazine, vol. 10, no. 2, pp. 93–109.

[2] M. S. Ahmed and A. R. Cook, (1979) *Analysis of freeway traffic time-series data by using Box-Jenkins techniques*, Transportation Research Record, no. 722, pp. 1-9.

[3] M. Levin and Y.-D. Tsao. (1980) *On forecasting freeway occupancies and volumes (abridgment)*, Transportation Research Record, no. 773.

[4] C. Moorthy and B. Ratcliffe, (1988) *Short term traffic forecasting using timeseries methods*, Transportation planning and technology, vol. 12, no.1, pp. 45–56.

[5] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias (2007), *Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks*, Computer-Aided Civil and Infrastructure Engineering, vol. 22, no. 5, pp. 317–325.

[6] I. Olabarrieta, I. Laña. (2020). *Effect of Soccer Games on Traffic, Study Case: Madrid*. 1-5. 10.1109/ITSC45102.2020.9294749.

[7] L. Breiman (2001). *Random Forests*. Machine Learning. 45 (1): 5–32. doi:10.1023/A:1010933404324.

[8] G. Upton, I. Cook, (2008) *Oxford Dictionary of Statistics*, OUP. ISBN 978-0-19-954145-4.

[9] I. Laña, I. Olabarrieta, J. Del Ser. (2021) Output Actionability of Traffic Forecasting Models: Measuring and Understanding Uncertainty of Forecasts to provide Confidence Levels (in preparation).

[10] G. Voronoi, (1908a). *Nouvelles applications des paramètres continus à la théorie des formes quadratiques*. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites" (PDF). Journal

für die Reine und Angewandte Mathematik. 1908 (133): 97–178. doi:10.1515/crll.1908.133.97. S2CID 116775758.

[11] P. Newson, J. Krumm. Nov. (2009*). Hidden Markov Map Matching Through Noise and Sparseness*. 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2009), November 4-6, Seattle, WA

Applicable European Regulations for Data-driven Policy-making

Sonia Bilbao
TECNALIA, Basque Research and Technology Alliance (BRTA),
P. Tecnológico Bizkaia, Ed. 700, 48160 Derio, Spain
Sonia.bilbao@tecnalia.com

Maria José López
TECNALIA, Basque Research and Technology Alliance (BRTA),
P. Tecnológico Bizkaia, Ed. 700, 48160 Derio, Spain
MariaJose.Lopez@tecnalia.com

Sergio Campos
TECNALIA, Basque Research and Technology Alliance (BRTA),
P. Tecnológico Bizkaia, Ed. 700, 48160 Derio, Spain
sergio.campos@tecnalia.com

ABSTRACT

Data-driven policy making aims to make optimal use of data and collaborate with citizens to co-create urban policies and, in general, to conduct a more reliable decision-making process. The European Commission considers data an essential resource for economic growth, competitiveness, innovation, and disposes as part of its strategy a set of regulations and guides, aiming to that more data becomes available for use while keeping the rights and trustworthy of the companies and individuals who generate and consume the data during the whole lifecycle. These regulations impact and open new challenges and opportunities when addressing the decision-making process in URBAN Transformation and specifically urban mobility as in the case of the URBANITE project.

KEYWORDS

Data, regulations, ethics, trustworthy, privacy, governance

1 INTRODUCTION

Urban mobility faces greater uncertainty and complexity in the long term generated by two main factors: the demand for growth in urban environments, the pressure for a more sustainable model of mobility in the face of the emergence of global warming. In general, we find that the social conscience is changing in favor of more equitable and sustainable ways, and the recovery of the space of the city for the people. On the other hand, the accelerated technological development in the transport modes and business models themselves, including innovations such as autonomous driving, micro-mobility, connected vehicles, electro-mobility, mobility as a service (MaaS), new vehicle ownership models, etc. mark specific challenges. in your deployment. These trends are changing the landscape of urban planning and mobility management in cities, incorporating new challenges. All of these require new advances in mobility planning processes and methods, with the aim of helping public administrations and policy makers to better understand this new context, supporting them in decision-making and policy definition. Policies should be discussed among the main actors in the new urban mobility scenario: citizens, service providers, public servants and political leaders.

Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia

This scenario can be built on two pillars: 1) co-creation sessions and 2) empirical analysis on stakeholder trust, attitude, impact, benefits and risks in the use of disruptive technologies. Now, traditional technological solutions are no longer valid in this situation, and disruptive technologies such as big data analysis or artificial intelligence emerge as a promising support to those responsible for formulating new policies. Data-driven policy making aims to make optimal use of existing heterogenous data and collaborate with citizens to co-create policy. This opportunity entails specific challenges to favor the acceptance by users of the results obtained through the application of these technologies and, first, to collect the relevant data from the different local stakeholders. These are some of the objectives of the URBANITE project, to face challenges, attitudes, confidence and opportunities in the use of disruptive technologies in public services in the context of urban mobility.

URBANITE identifies several key results: a Social Policy Lab – an environment to promote digital co-creation and methodologies and methods to support the co-design and co-creation for policies, a Data Management Platform – To provide automatic support to the whole data processing chain and its life cycle, starting with the collection process up to its use and a Decision-Making Support System – Powerful tools which combine multiple data sources with advanced algorithms, simulation, recommendation, and visualisation.

The project identifies different stages from the perspective of data and more specifically, its availability, openness and privacy:

- 1st Stage- Setup of participation labs and initial gathering:
  - Open data currently available
  - including identification and recruitment of participants, the preparation of an informed consent procedure to implement for individual participation
  - The register and use of the virtual participation platform as a complement of previous sessions
- 2nd Stage. The potential use of existing non-open data, personal and non-personal on the cities to the objectives of the project.
- 3rd Stage. The transfer of collected data from 3rd parties, defining a transfer agreement among both parties (company and city use case)

2 RELATED EUROPEAN REGULATIONS

On the other hand, the European Commission considers data an essential resource for economic growth, competitiveness, innovation, defining an European strategy for data aiming to ensure Europe's global competitiveness, a data sovereignty, that more data becomes available for use, while keeping the rights of the companies and individuals who generate and consume the data during the whole lifecycle. As part of this support, the Commission has proposed some material as part of its data strategy, disposing several normative and guides to conduct a success and trustworthy data management.

2.1 Data Governance Act

This initiative refers to the management of personal as well as non-personal data, therefore being linked at the legislative level with the General Data Protection Regulation (GDPR)[1] and the Directive on privacy and electronic communications [2]. The European Commission has implemented a solid and trustworthy legal framework for the protection of data, in order to promote a single data market, for which it must guarantee that data from the public sector, companies and citizens can be available and used in the most efficient and responsible way possible, while companies and The Data Governance Act [3] is the first of a set of measures announced in the 2020 European data strategy, aims to promote the availability of data for its use, increasing trust between the parties and strengthening data collection mechanisms throughout the European Union. The DGA will also support the establishment and development of common European data spaces in strategic domains, involving both public and private actors.

The framework addresses the following scenarios:

- The transfer of public sector data for reuse, in cases where such data is subject to the rights of third parties. It establishes a mechanism for the reuse of certain categories of protected data from the public sector, which is subject to respect for the rights of third parties.
- The transfer of personal data with the help of intermediaries, whose work will consist of helping providers to exercise the rights conferred by the General Data Protection Regulation (GDPR). The objective is to strengthen trust in the exchange of personal and non-personal data, and reduce the costs of transactions linked to the exchange of data between providers and their consumers, with neutral facilitators,
- The transfer of data for altruistic purposes (making data available to the common good, on a voluntary basis, by individuals or companies). Establish a registration and consent in order to reduce costs and facilitate data portability.
- The exchange of data between companies in exchange for some type of remuneration.

2.2 ETHICS GUIDELINES FOR TRUSTWORTHY AI

Despite the fact that AI technologies are mature enough, their adoption by companies is very uneven, and in general, much lower than one would expect. There are obstacles that hinder the widespread extension of AI technologies, both cultural and technical. AI technologies will not spread massively until the scientific community is able to develop reliable technology from the user and from the different data providers. On the other hand, the use of these technologies involves risks that must be managed appropriately. To ensure that we are on the right track, it is necessary to abide by a human-centered approach to AI, without losing the goal of improving human well-being. The concept of trusted AI addresses reliance on technology as a first step. The new guidelines are aimed at all parties involved who develop, apply or use AI, encompassing companies, organizations, researchers, public services, institutions, individuals or other entities.

According to the regulations "Reliable AI has two components: 1) it must respect fundamental rights, current laws and essential principles and values, so as to guarantee an« ethical purpose », and 2) it must be reliable and technically sound , since a little technological mastery can cause involuntary damages, although the intentions are good ".

Therefore, these Guidelines establish the framework of a reliable AI, guiding in three levels of abstraction, from the most abstract to the most concrete of aspects to be evaluated:

- Guarantee of the ethical purpose of AI, establishing fundamental rights, as well as the essential principles and values, which it must comply with.
- A series of guidelines, addressing both ethical purpose and technical soundness, listing the requirements for reliable AI, and providing a summary of technical and non-technical methods that can be used for their application.
- A concrete, but not exhaustive, list of aspects that must be evaluated in order to achieve reliable AI.

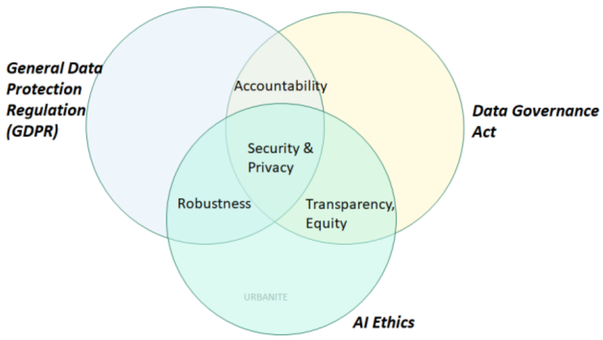

Figure 1: Relation and overlapping among regulations

According to these principles, the work package in charge of the definition, design and adaption of artificial intelligence and data analytics is ensuring that the applied methods meet the seven key requirements for Trustworthy AI:

1) human agency and oversight. URBANITE proposes a decision support methodology and supporting tool for policy creation, that combine and carefully balance different methods:
2) harvested historical data, GIS, expert knowledge, outputs of decision models, and others. The last word will remain in the hands of municipal experts, the platform being a tool to facilitate their decision. On the other hand, one of the pillars of the project is the implementation of a thoughtful space of discussion among the main actors of the new urban mobility scenario: citizens, service providers, public servants and policy makers.
3) technical robustness and safety. The work focused on the algorithms and simulations to be deployed, defines for as

objective metrics a pair of KPIs, refereed to the precision of predictions and the quality of recommended policies or procedures.

4) privacy and data governance, a fundamental right particularly affected by AI systems. Prevention of harm to privacy also necessitates adequate data governance that covers the quality and integrity of the data used. All the algorithms will work on the gathered data, just (pseudo) anonymised. If applied during the research stage on the algorithms, all the GRPR measures will be analysed and adopted to use personal data.
5) Transparency [7], the principle of explicability and encompasses transparency of elements relevant to an AI system: the data, the system and the business models: This implies traceability of decision along the whole cycle of data, from datasets, gathering, labelling and process, Explainability concerns will be considered for the methods applied, ensuring a better understanding of the underlying processes and related human decisions (e.g. xAI approach). In any case, the simulation and rules-based reasoning approaches, are well sited from the explainability point of view.
6) diversity, non-discrimination and fairness. URBANITE gather existing open data portals, geographical information systems (GIS), data coming from private data providers, the basis is any, comes in origin. SoPoLab sessions and use case evaluation support the feedback from stakeholders who may directly or indirectly be affected throughout its life cycle.
7) dedicated assessment of the algorithms during their design and use case deployment ensures the auditability.
8) environmental and societal well-being. In the last term, URBANITE provides a new decision-support system for Planning Sustainable Mobility and the early evaluation of urban policies. A Sustainable Urban Mobility Plans (SUMP), defines strategic plans based upon a long-term vision of transport and mobility, guaranteeing technical, economic, environmental and social sustainability.

3 ALGORITHMS ACTIONABILITY

Taking into account the previous regulations and based on previous experience in the context of Intelligent Transportation Systems [4], it is confirmed that aspects such as trust, precision and reliability, among other non-functional properties, are essential for predictive and analytical techniques to be practices in its use. We present the term Actionability, as the characteristic that any system based on data analysis or artificial intelligence must present to be implemented and used successfully in a real operating environment. This concept, in turn, identifies a series of desirable characteristics, which in URBANITE are contextualized in the field of urban mobility planning.

Data-based models are usually subject to uncertainty, involving non-deterministic stochastic processes, both in the learning, execution or training mechanisms / input data, and also present in the results. Once deployed, it is essential to provide an objective measure of the reliability and precision of the results, winning in terms of Trust. The need to explain and render the underlying analytical models interpretable is undoubtedly one of the research fields with the greatest impact, being considered under the concept of Explainable Artificial Intelligence [5][6] (xAI). This field of study comprises different techniques and methods, taking into account three fundamental factors: the nature of the model to be explained, from intrinsically transparent to completely opaque and unintelligible; the user of the algorithms; and finally, the way in which said explanation must be prepared and presented to the decision maker, which will depend on their degree of knowledge, as well as the intrinsic possibilities offered by the model to be explained in one way or another.

Adaptation is the reaction of a system, model or process to new circumstances, with the idea of maintaining its performance or reducing its loss, compared to the ideal conditions that were taken into account in its design and initial adjustment. The main problem in scenarios whose underlying phenomena change over time, without being addressed by the model itself, is that the conclusions, predictions or categorizations will not be reliable. This phenomenon is called concept drift [8][9].

Robustness refers to the ability of a system to maintain service when external incidents occur. In the case of urban planning, it will not be so critical, since the decisions to be made will not be made in real time; However, the data ingestion of the different data sets and available stores, if it must be operational, to minimize the loss of input data, in addition to being robust data algorithms in such loss situations, fluctuations in the frequency of the themselves, poor quality data, etc. In the URBANITE project, data quality is explicitly addressed through the implemented components associated with data preservation.

Stability means ensuring that there are no surprises for the user in terms of functionality. In general, the algorithms are worked in a specific geographic area and according to the available data sets. However, for their deployment in a real environment, it is necessary to project them to larger areas and volumes of data. This issue must be taken into account from the design stage of the algorithms, to optimize their algorithmic complexity, which represents the amount of resources (temporal, execution time and space, required memory) that an algorithm needs to solve a problem. This characteristic allows to determine the efficiency of this algorithm, not in terms of absolute measures but measures relative to the size of the problem. Currently, the availability of new technologies and paradigms of parallel and distributed processing of massive volumes of data, allows an escalation of the methods, obtaining adequate response times. However, its exploitation requires the adequate implementation and adaptation of the algorithms according to the architecture in which they will be deployed, as well as optimizing this deployment of analytical workloads in the different layers.

Another key feature is the compliance of the new methods with transportation engineering. Existing traffic and mobility engineering practices are well established, with a powerful knowledge base. A better understanding of the hybridization of data analysis and simulation methods, data-driven and model-driven approaches, by combining the strengths of each side, will help us improve the models by identifying more complex underlying assumptions. In general, the results are closely linked to the experiments carried out; transferability is a desirable characteristic for algorithms and any model, in order to present adequate performance and functionality in other contexts and starting data, different from those used in learning.

Finally, we cannot forget the contextual aspects identified by the EU for the definition of sustainable mobility policies, measures and solutions, as part of SUMP and any support tool, with the aim of contributing to urban regeneration, transport sustainability, social inclusion and social empowerment through active participation.

4 OPPORTUNITIES AND NEXT STEPS

Additionally, to the actionability requirements for our methods and algorithms, the new regulation and especially the Data Governance Act presents a set of topics or opportunities to explore from the different action lines. The following table presents some of them, according to the type of data to explore on the project: public, personal or altruistic data

Table 1: Challenges and Research Opportunities

	Public sector data for reuse	*Reuse of personal data*	*Altruistic purpouse*	*Data Sharing among business*
SoPoLab	Presentation of **Data Gobernance Act among local stakeholders**. **Engagement** of city departments, public transport service providers, collectives and final users on sharing data. **Identification of potential local "Data Sharing Providers"**.			
Data Management Platform	**Support of Data Gobernance functionalities** by the Urbanite Platform: - **Data exchange policies**. Establishment of policies regarding the exchange of data and events across the boundaries of an organization. - **Data ownership**. Opportunities or barriers to data and events between companies: IPRs, rights of access and exploitation, sensitivity, etc. - **Security and privacy**. Access and authentication, integrity, privacy. Etc.			
Decision support tools	**Explainable** IA-techniques. Other algorithms **avoiding the trasfer of raw data** between origin and the platform			

5 CONCLUSIONS

During the first period of the project, some relevant algorithms and data analysis have been identified based on the project's use cases. Having analyzed the different regulations around the exploitation, data management and artificial intelligence technologies, machine learning and advanced processing, the requirements that these new algorithms must present in the future have been identified. Finally, it introduces the concept of Actionability as a key property of any data-based modeling and treatment process to generate knowledge of practical value for decisioning. All these aspects open challenges and, also opportunities for URBANITE project.

ACKNOWLEDGMENTS

These results have been achieved during funding project from the European Union's Horizon 2020 research and innovation programme under grant agreement #870338.

.

REFERENCES

[1] DO L 119 de 4.5.2016, p. 1.

[2] DO L 201 de 31.7.2002, p. 37.

[3] https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-data-governance-data-governance-act

[4] Ibai Lana. Vlahogianni Elenni, Javier Del Ser, 2020, "From Data to Actions in Intelligent Transportation Systems: a Prescription of Functional Requirements for Model Actionability," arXiv, 2020.

[5] Alain Barredo, N. Díaz-Rodríguez, Javier Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, Francisco Herrera, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. arXiv:1910.10045," 2020.

[6] D. Doran, S. Schulz and T. R. Besold, 2017, "What does explainable AI really mean? a new conceptualization of perspectives. arXiv:1710.00794," 2017.

[7] A. Datta, S. Sen and Y. Zick, 2014, "Algorithmic transparency via quantitative input influence: Theory and experiments with learning system," in 2016 IEEE symposium on security and privacy (SP), 598–617, 2016.

[8] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy and A. Bouchachia, 2014, "A survey on concept drift adaptation," ACM computing surveys (CSUR), vol. 46, no. 4, p. 44, 2014.

[9] Jesús Lobo, Javier Del Ser, Nekane Bilbao Cristina Perfecto, 2018, "DRED: An evolutionary diversity generation method for concept drift adaptation in online learning environments," Applied Soft

Supporting Decision-Making in the Urban Mobility Policy Making

Erik Dovgan
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia
erik.dovgan@ijs.si

Maj Smerkol
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia
maj.smerkol@ijs.si

Miljana Sulajkovska
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia
miljana.sulajkovska@ijs.si

Matjaž Gams
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia
matjaz.gams@ijs.si

ABSTRACT

City mobility is changing rapidly due to population growth and disruptive technologies. To efficiently handle these changes, policymakers need advanced tools based on AI, including simulation, prediction, decision making, and visualization. In the URBANITE H2020 project, we are developing a decision support system (DSS) that is based on DEXi and enables the decision-makers to combine low-level mobility data obtained with simulation, into high-level attributes suitable for decision making and comparison of mobility scenarios. By providing the user preferences in advance, DSS can be also used in combination with machine-learning models to search for the best mobility policies automatically.

KEYWORDS

decision making, mobility, urban transformation

1 INTRODUCTION

The mobility in cities is changing rapidly. On one hand, the population in cities is growing which results in increased congestion and pollution. On the other hand, new and disrupting mobility modes are being introduced, such as vehicle sharing, hop on/off bikes, etc. The city policymakers thus face a very complex problem: how to improve mobility under growing congestion pressures, while considering new mobility modes [1]. Advanced tools that include artificial intelligence (AI) approaches can significantly help policymakers to select the most appropriate actions [4].

AI-based tools for city mobility typically include the city models and traffic simulation, which enables the users to simulate various traffic situations [3]. We are developing a system that will, besides city models and traffic simulation, include also a decision support system (DSS) [2] and a machine learning module. The decision support system will support the user, either human or algorithm, in selecting the best policy, while the machine learning module will aim at replacing human decision-makers with algorithmic ones. In this paper, we focus on the DSS of the URBANITE H2020 project [5].

Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia

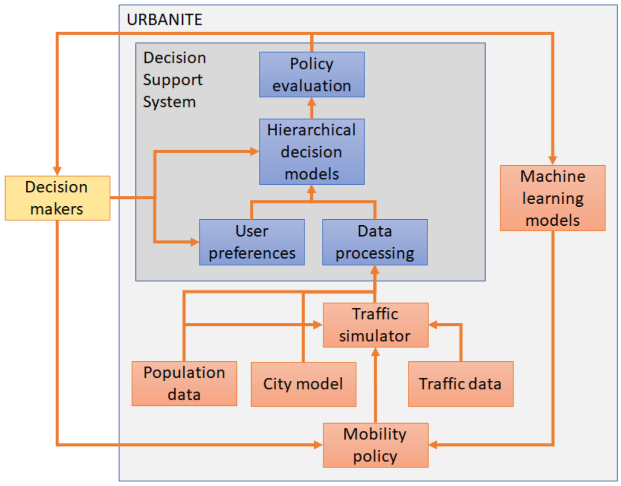

Figure 1: The architecture of the URBANITE system.

The developed DSS aims to enable the users to select the most appropriate policy actions based on data from simulations, population data, current and predicted traffic data, and user (citizen, decision-makers) preferences. By defining the user preferences in advance, it weights and hierarchically aggregates the basic data to obtain one or a few objectives, based on which the evaluated policies are compared and ranked. The policy ranking represents the key information for the final decision regarding which policy should be applied.

The rest of the paper is organized as follows. Section 2 presents the URBANITE system. The decision support system within URBANITE is described in Section 3. Finally, Section 4 concludes the paper with a summary and ideas for future work.

2 OVERVIEW OF THE URBANITE SYSTEM

The URBANITE system consists of several modules such as tools for the involvement of various types of stakeholders, including the general public. However, from the point of view of the presented decision support system, only the modules relevant for DSS are presented in Figure 1.

There are two main inputs to DSS. The first one consists of the expert knowledge, provided by the decision-makers. This knowledge is of key importance when building hierarchical decision models, as well as when defining user preferences (see Section 3 for details).

The second input consists of raw data including city models, population data, and evaluation results from the traffic simulator. Population data include the number of people in the urban area as well as their distribution between the districts. The city model consists of a map of roads, districts, public areas, etc. Finally, traffic simulation results are trip traces that include all the relevant data such as the (vehicle) positions, time, and pollution. These results are obtained by evaluating a mobility policy with the traffic simulator. To this end, the simulator processes the population data, the city model, and the past traffic data to emulate the characteristics of real-life traffic as much as possible.

The mobility policy consists of a set of actions to be applied within the urban area (such as closing a specific road for cars) and can be proposed either by decision-makers or by machine-learning models. Both take into account the policy evaluation, computed by the DSS. The main difference between the two approaches is the fact that decision-makers rely on expert knowledge and define the mobility policies by hand, while machine-learning models apply pattern-recognition approaches, process a possibly huge amount of data, and select mobility policies automatically.

Finally, the decision support system consists of several components that are described in Section 3.

3 DECISION SUPPORT SYSTEM (DSS)

Our DSS aims to evaluate mobility policies, i.e., for each policy produce one or a limited set of objectives that are easily interpretable and handled by the experts. Note that a baseline mobility policy is evaluated by the traffic simulator, but the evaluation provided by a standard simulator is very difficult to process by experts due to a large amount of data since the evaluation consists of traces of all the trips within the city. Therefore, the DSS aggregates evaluation data into meaningful high-level attributes to enable efficient and effective decision-making.

3.1 Components of the URBANITE DSS

The main component of the DSS is the hierarchical decision model (see Figure 1). A hierarchical decision model is defined by the experts/decision-makers based on their expert knowledge. It starts with the evaluation values, provided by the traffic simulator, and iteratively combines semantically similar attributes into higher-level attributes until only one attribute remains. This results in a tree structure in which the root represents the final evaluation of the policy. However, it is not required to always use the final evaluation during the decision-making process. In some cases, it is more appropriate to use several high-level attributes (e.g., pollution and congestion) to compare the policies in all the aspects that the decision-makers are interested in. In this case, the selected attributes are inner nodes of the tree structure.

To create the hierarchical decision model and to select the relevant attributes, user preferences have to be obtained. They are included in the module by experts/decision-makers. When creating the decision model, the preferences are used to weigh the attributes within the tree structure. More precisely, when combining attributes into a higher-level node/attribute within the tree structure, a utility function needs to be defined, which specifies how each combination of lower-level attributes transforms into the higher-level attribute. This is a preference-based process and typically involves combining qualitative attributes of various types.

Hierarchical decision models are not able to directly handle the city model data or the raw data obtained from the traffic simulator. Therefore, the baseline data need to be preprocessed and, if appropriate, aggregated. For example, if the city pollution is required as an input to the hierarchical decision model, it has to be calculated from all the trips within the city.

Finally, policy evaluation has to be executed. This is done by applying the preprocessed traffic simulator data within the hierarchical decision model. The resulting values of the high-level attributes, selected based on user preferences, are then sent to decision-makers or machine-learning models (see Figure 1). The hierarchical decision models, including their definition and execution, were implemented with DEXi [2].

3.2 Hierarchical Decision Model for Mobility Policy Evaluation

A new hierarchical decision model was developed by focusing on the needs and preferences of the URBANITE project [5], based on the user experience of four major EU cities. The model shown in Figure 2 was developed based on mobility policies that include building new roads, closing parts of the city like squares, setting up new lines of public transport including ferries, and other potential modifications of the city mobility. For a policy, three areas within the city were identified as relevant:

- Target area where the policy action is applied
- Nearby area that surrounds the target area and which is directly influenced by the applied policy
- The entire city

The attributes were divided into three categories:

- Road network
  These attributes measure the size of the city area where the policy action has a direct influence. They also consider the capacity of the affected roads and take into account both target and nearby areas.
- Population-related attributes including the type of the area and public transport data.
  Area type is defined with the position within the city (e.g., center, periphery), the district type (e.g. residential, commercial), and the population number. Public transport counts the available bus and underground stops, and the lanes of public transport. All these attributes are measured in both target and nearby areas.
- Policy impact
  It measures the change with respect to the baseline scenario when no policy action is applied. The following aspects are taken into account:
  – Change in air pollution
  – Change in the number of used private vehicles
  – Change in the number of used bicycles
  – Change in the number of used public transport
  – Change in the number of pedestrians
  In addition, it also takes into account congestion change. All the attributes are measured in both target and nearby areas, as well as in the entire city.

The developed model is intended to be used for both comparing the effects of applying a policy with the baseline as well as comparing the effects of various policies between themselves. As

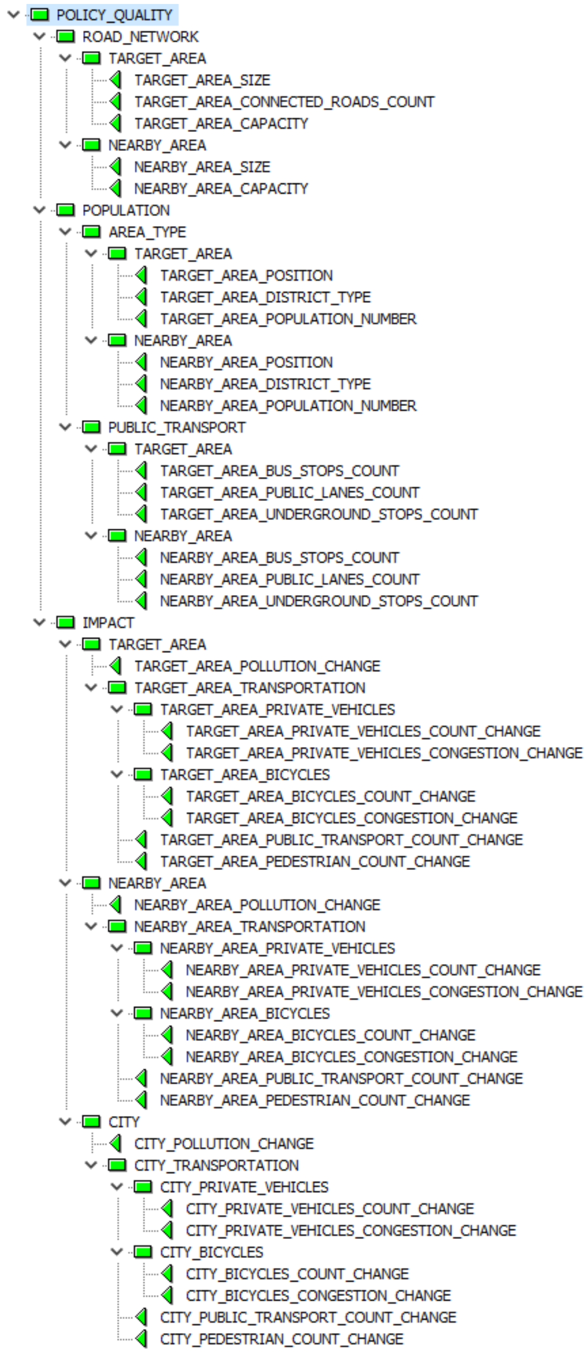

Figure 2: A hierarchical decision model for mobility policy evaluation.

a consequence, some attributes focus on comparison with baseline, while others focus on differences among various policies.

Selection of the attributes and their organization into the tree structure is only the first step when building the hierarchical model. The second step consists of defining the functions that aggregate the lower-level attributes into higher-level ones, i.e., utility functions. All the attributes in the inner nodes of the tree were defined as categorical from 1 to 5, which facilitated the utility function definition. The default scale defined the higher the better, except for the pollution where the the-lower-the-better scale was applied. An example of the utility function is shown in

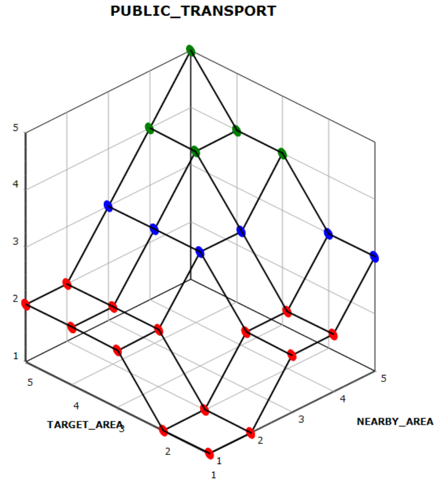

Figure 3: An example of the utility function for the selected attributes.

Option	WITHOUT_INTERVENTIONS	CLOSED_SQUARE
TARGET_AREA_SIZE	2	2
TARGET_AREA_CONNECTED_ROADS_COUNT	2	1
TARGET_AREA_CAPACITY	2	1
NEARBY_AREA_SIZE	2	2
NEARBY_AREA_CAPACITY	2	2
TARGET_AREA_POSITION	center	center
TARGET_AREA_DISTRICT_TYPE	residential	residential
TARGET_AREA_POPULATION_NUMBER	1	1
NEARBY_AREA_POSITION	center +districts	center +districts
NEARBY_AREA_DISTRICT_TYPE	residential	residential
NEARBY_AREA_POPULATION_NUMBER	2	2
TARGET_AREA_BUS_STOPS_COUNT	3-6	3-6
TARGET_AREA_PUBLIC_LANES_COUNT	3-6	3-6
TARGET_AREA_UNDERGROUND_STOPS_COUNT	1-2	1-2
NEARBY_AREA_BUS_STOPS_COUNT	7+	7+
NEARBY_AREA_PUBLIC_LANES_COUNT	7+	7+
NEARBY_AREA_UNDERGROUND_STOPS_COUNT	3-6	3-6
TARGET_AREA_POLLUTION_CHANGE	5% decrease - 5% increase	+20% decrease
TARGET_AREA_PRIVATE_VEHICLES_COUNT_CHANGE	5% decrease - 5% increase	+20% decrease
TARGET_AREA_PRIVATE_VEHICLES_CONGESTION_CHANGE	5% decrease - 5% increase	+20% decrease
TARGET_AREA_BICYCLES_COUNT_CHANGE	5% decrease - 5% increase	+20% increase
TARGET_AREA_BICYCLES_CONGESTION_CHANGE	5% decrease - 5% increase	+20% decrease
TARGET_AREA_PUBLIC_TRANSPORT_COUNT_CHANGE	5% decrease - 5% increase	5-20% increase
TARGET_AREA_PEDESTRIAN_COUNT_CHANGE	5% decrease - 5% increase	+20% increase
NEARBY_AREA_POLLUTION_CHANGE	5% decrease - 5% increase	5-20% increase
NEARBY_AREA_PRIVATE_VEHICLES_COUNT_CHANGE	5% decrease - 5% increase	5-20% increase
NEARBY_AREA_PRIVATE_VEHICLES_CONGESTION_CHANGE	5% decrease - 5% increase	5-20% increase
NEARBY_AREA_BICYCLES_COUNT_CHANGE	5% decrease - 5% increase	5-20% increase
NEARBY_AREA_BICYCLES_CONGESTION_CHANGE	5% decrease - 5% increase	5-20% increase
NEARBY_AREA_PUBLIC_TRANSPORT_COUNT_CHANGE	5% decrease - 5% increase	5-20% increase
NEARBY_AREA_PEDESTRIAN_COUNT_CHANGE	5% decrease - 5% increase	5-20% increase
CITY_POLLUTION_CHANGE	5% decrease - 5% increase	5-20% decrease
CITY_PRIVATE_VEHICLES_COUNT_CHANGE	5% decrease - 5% increase	5-20% decrease
CITY_PRIVATE_VEHICLES_CONGESTION_CHANGE	5% decrease - 5% increase	5-20% increase
CITY_BICYCLES_COUNT_CHANGE	5% decrease - 5% increase	5-20% increase
CITY_BICYCLES_CONGESTION_CHANGE	5% decrease - 5% increase	5-20% increase
CITY_PUBLIC_TRANSPORT_COUNT_CHANGE	5% decrease - 5% increase	5-20% increase
CITY_PEDESTRIAN_COUNT_CHANGE	5% decrease - 5% increase	5-20% increase

Figure 4: Definition of two test scenarios.

Figure 3 that shows how the Target area attribute and the Nearby area attribute are combined into the Public transport attribute.

3.3 Evaluation of Mobility Policies

The hierarchical decision model, described in Section 3.2, was used to evaluate a test policy that prescribed that the main square of a test city should be closed. The effects of this policy were compared to the baseline, where no actions were taken.

First, both scenarios (no intervention and closed square) were simulated and the obtained results were preprocessed. Second, the data were inserted in DEXi as shown in Figure 4, where each column represents one scenario and colors represent the evaluation of single attributes (green: good, black: neutral, red:

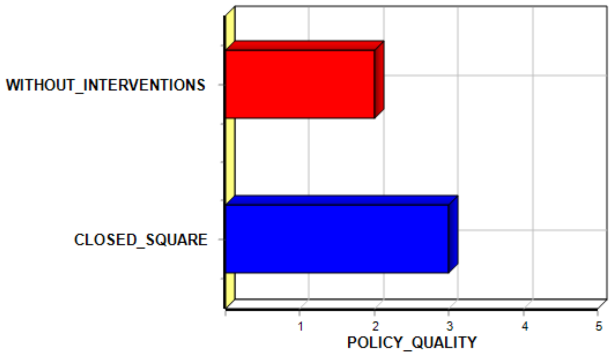

Figure 5: Comparison of the overall quality of the test scenarios.

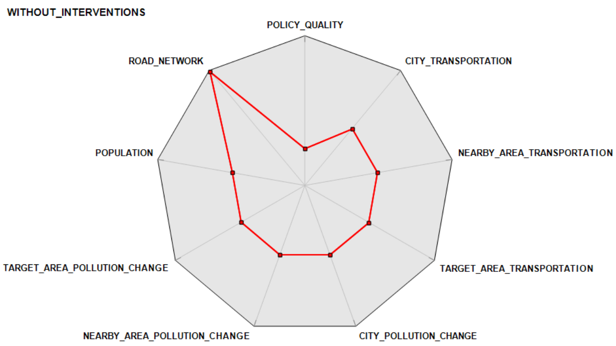

Figure 6: Evaluation of the scenario without interventions.

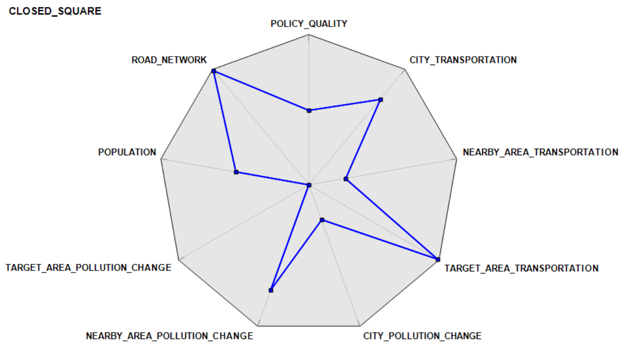

Figure 7: Evaluation of the scenario with closed square.

bad). This figure shows that the differences between the scenarios are in the target area roads and the impact attributes. The impact is negative only in a minority of attributes: in the nearby area and the congestion when observing the entire city (see the color change from black to red). On the other hand, there is a positive change in the majority of the impact attributes (black color to green).

The selected scenarios were evaluated both based on the overall quality and based on a set of the most relevant high-level attributes, i.e., inner nodes of the tree. The overall quality comparison is presented in Figure 5, while the comparison on the selected attributes can be found in Figures 6–7. These figures show that the overall quality of the closed-square scenario is higher in comparison to no interventions. As noted previously, for the pollution change in Figures 6–7, lower, i.e., near the center of the graph is better, while for other attributes, higher, i.e., near the edge of the graph is better. In these figures, we can observe a similar trend as in Figure 4. The difference is that in Figure 4 we compare the scenarios on basic attributes (leafs of the tree), while in Figures 6–7 we compare scenarios on the higher-level attributes (inner nodes of the tree). Finally, Figure 5 shows the comparison on the top-level attribute, i.e., the root of the tree.

4 CONCLUSION

Selection of the best mobility policy for a city is typically a complex task since the policy can influence a large variety of mobility aspects. In addition, simulation tools typically produce a large amount of data that needs to be appropriately preprocessed and aggregated. Consequently, a suitable approach for hierarchical aggregation of mobility attributes needs to be defined to get a low number of higher-level attributes that make the decision-making process feasible.

In this paper, we proposed to aggregate the mobility attributes with DEXi. DEXi applies hierarchical decision models that are defined based on expert/decision-maker knowledge. We developed a new hierarchical decision model that was then used for basic and multiobjective comparison of mobility scenarios.

This paper also presented a basic graphical interface for comparing the scenario outputs, while additional and more advanced GUIs are still under development. The evaluation of the developed decision model on a variety of mobility policies is ongoing and aims at determining whether the model is suitable for all the relevant scenarios. In case of discovered deficiencies, we will upgrade the model with additional attributes and/or attribute rearrangement.

ACKNOWLEDGMENTS

This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870338. The authors also acknowledge the financial support from the Slovenian Research Agency (research core funding No. P2-0209).

REFERENCES

[1] M. Batty, K.W. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis, and Y. Portugali. 2012. Smart cities of the future. *European Physical Journal-Special Topics*, 214, 481–518.

[2] Marko Bohanec. 2020. DEXi: Program for Multi-Attribute Decision Making, User's Manual, Version 5.04. IJS Report DP-13100. Jožef Stefan Institute, Ljubljana, Slovenia.

[3] Erik Dovgan, Jaka Sodnik, Ivan Bratko, and Bogdan Filipič. 2017. Multiobjective discovery of human-like driving strategies. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion* (GECCO '17), 1319–1326.

[4] Matjaž Gams, Irene Yu-Hua Gu, Aki Harma, Andres Munoz, and Vincent Tam. 2019. Artificial intelligence and ambient intelligence. *Journal of Ambient Intelligence and Smart Environments*, 11, 1, 71–86.

[5] The URBANITE Project. 2021. https://urbanite-project.eu/.

URBANITE Data Management Platform

Fritz Meiners
Fraunhofer FOKUS
Digital Public Services
Kaiserin-Augusta-Allee 31
10589 Berlin, Germany
fritz.meiners@fokus.fraunhofer.de

Sonia Bilbao
TECNALIA, Basque Research and Technology Alliance (BRTA), P. Tecnológico Bizkaia, Ed. 700, 48160 Derio, Spain
sonia.bilbao@tecnalia.com

Gonzalo Lazaro
TECNALIA, Basque Research and Technology Alliance (BRTA), P. Tecnológico Bizkaia, Ed. 700, 48160 Derio, Spain
gonzalo.lazaro@tecnalia.com

Giuseppe Ciulla
Research & Development Laboratory
Engineering Ingegneria Informatica
Palermo, Italy
giuseppe.ciulla@eng.it

ABSTRACT

This paper describes the Data Management Platform developed in URBANITE H2020 project. This platform provides automatic mechanisms to harvest, curate, fuse and visualize existing open and proprietary data coming from different sources related to urban mobility and transportation (e.g. traffic data from cars, public transport, bikes or ferries; air quality and noise; events, parking, and so on).

KEYWORDS

Data harvesting, data curation, DCAT-AP metadata, data storage

1 INTRODUCTION

One of the main goals of the research carried out in URBANITE H2020 projects, is to provide algorithms, tools and models to support decision-making processes in the field of urban planning and mobility. This support is based on the analysis of the current situation based on harvested and fused data, on data simulations and the prediction of future situations when changing one or more variables. Hence, the availability of good quality data coming from heterogeneous data sources and its interoperability for data aggregation and fusion is highly important.

The Data Management Platform (DMP) provides the components for data acquisition, aggregation and storage. These components are:

- Data Harvesting, Preparation and Transformation covering the entire process of fetching, preparing, transforming, and exporting data for storage
- Data Anonymization to transform datasets in conformity with data protection requirements for further data analysis.
- Data Curation which deals with enrichment and annotation of data.
- Data Fusion and aggregation. Data aggregation is the process of gathering data and presenting it in a summarized format, e.g. to hide personal information or to provide information in a synthetic form. Data fusion is the process of integrating multiple data sources to produce more consistent, accurate, helpful information and sophisticated models than those provided by any individual data source.
- Data Storage & Retrieval providing the means to store and retrieve datasets, DCAT-AP compliant metadata, and related data.
- Data Catalogue offering the functionalities to discover and access the datasets collected and managed by the components of the URBANITE Ecosystem.

Information Society 2021: 24th international multiconference, *4–8 October 2021, Ljubljana, Slovenia*

2 DMP ARCHITECTURE

Figure 1 represents the component diagram of the Data Management Platform (blue rectangle) and its interaction with the other modules in the URBANITE Ecosystem.

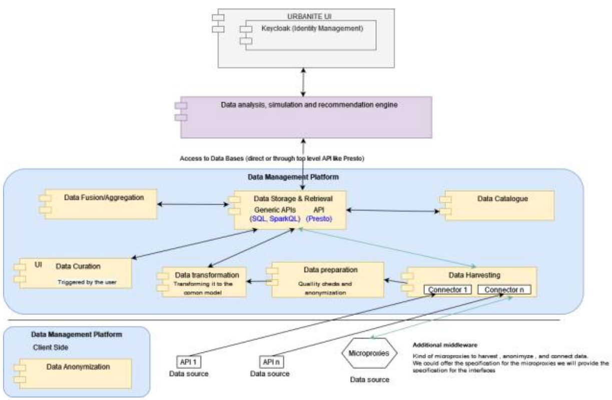

Figure 1. Component diagram of the DMP

3 IMPLEMENTATION

3.1 Data Harvesting, Preparation and Transformation

The process of fetching, preparing, transforming, and exporting data (from now on referred to as *harvesting*), i.e. providing a way to make heterogeneous data available in defined format and means of access, has been implemented following the form of a pipeline, as shown in Figure 2. This means that data is passed through the pipeline, and each component is agnostic of the other steps. This leads to lose coupling and improves flexibility allowing steps to be omitted if not necessary for a given data source. The pipeline has been implemented using the open source solution named Piveau Pipe Concept [1, 2].

Each of the components in the pipeline is implemented as a service that exposes a common RESTful interface. This way, services can be connected in a generic fashion to implement specific data processing chains. No central instance is responsible for orchestrating the services. A scheduler is in charge of launching the pipes and how services are connected together is specified in a JSON file known as the pipe descriptor.

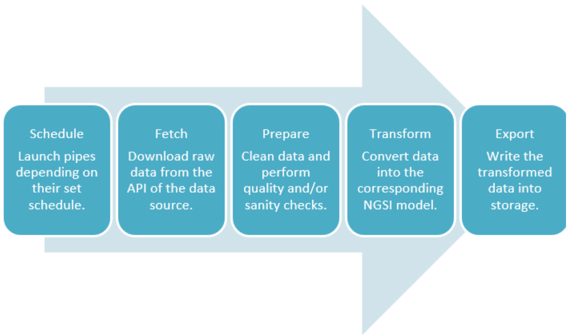

Figure 2. Harvesting process

In detail, the harvesting process would typically consist of the following steps:

1. The scheduler triggers a pipeline
2. The harvester retrieves the data from the source's API and forwards it into the preparation component.
3. After cleaning and validating, the preparation component forwards the data to the transformation component.
4. The data is transformed to the applicable NGSI data model and forwarded to the exporter.
5. Finally, the exporter writes the harmonized data into the data storage component.

The scheduler serves two main purposes: keeping track of existing pipe descriptors and managing triggers for these pipes. Each pipe descriptor is stored as a JSON file and contains a definition of components (endpoints, chronological order, specific configurations) that make up the processing sequence. Each processing chain is defined in one of these files. The scheduler reads these files to become aware of which pipes are available. These can then be assigned to a periodic trigger for recurring execution.

The data harvesting component is responsible for fetching data from a given API. It does not alter the data. It can be considered the entry point of the data into the pipeline. As such, a dedicated component is required for each type of data source. The harvesting component may implement pagination mechanisms for handling data in chunks. However, this does not impact the pipeline – each chunk is handled individually and does not depend on other chunks.

The data preparation component is responsible for performing initial cleaning and sanitation of the data provided by the harvesting component. This ensures a fixed level of data quality and integrity, which is required by the transformation component to operate flawlessly.

Data transformation is a key step in the harvesting pipeline. It cannot be expected that the municipalities provide their data in one of the common data models developed by FIWARE used in the URBANITE context. As such, the transformation of the heterogeneous data sources into common models is vital for frictionless processing of the data henceforth. For a flexible approach, the actual transformation instructions are loaded via scripts, either JavaScript for JSON based payloads or XSLT for XML based payloads. More engines can be added as pipeline modules at a later point in time.

An example of a pipe descriptor is provided in Figure 3. Each of the segments describes a service in the pipe. In the example there are three services, the first named *importing-bilbao-air-quality* that downloads Bilbao air quality data, the second named *transforming-js* that transforms the data according to a JavaScript file and the third one named *exporting-data-storage* that invokes the storage and retrieval component to store the data and its metadata in two dedicated repositories. The segment number field indicates the order in which the service should be executed in the pipe.

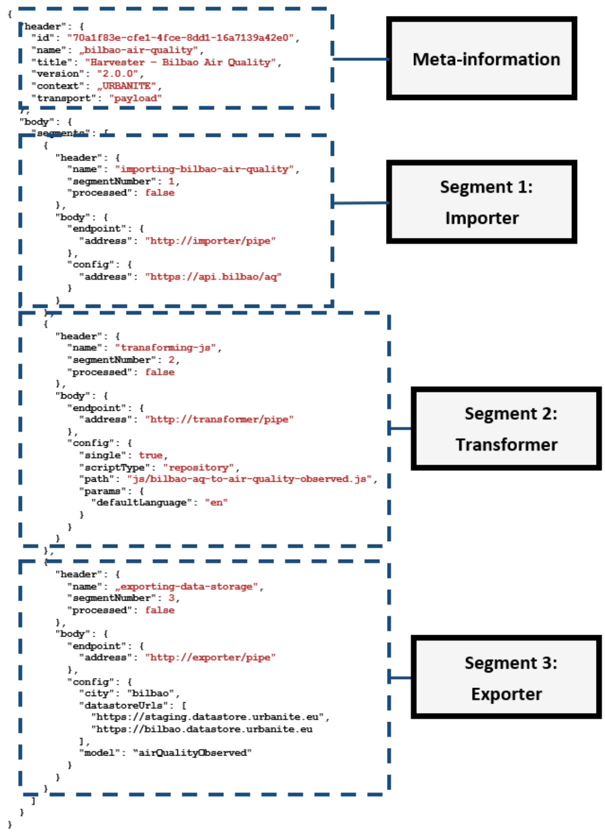

Figure 3. Example of a Piveau Pipe Descriptor

When a pipe is triggered the service in segment number 1 is called. Once finished, data that needs to be passed along the processing chain is written into a payload field of the next component in line. For smaller amounts of data this can happen directly, for larger amounts of data a pointer to an external datastore can be used.

3.2 Data Anonymization

The anonymization component is a RESTful microservice capable of transforming large datasets in conformity with data protection requirements for further data analysis. In order to achieve a certain degree of anonymization the user can mark specific attributes that are likely to reveal information about a person or a smaller group. Those identifiers are then transformed in a way that ensures a sufficient level of anonymization. Currently supported anonymization methods are suppression and generalization, which either delete attribute entries in a row or generalise them according to a fixed hierarchy, such as street -> zip code -> city.

3.3 Data Storage & Retrieval

The Data Storage & Retrieval component provides the means to store and retrieve datasets (transformed to URBANITE common data model compliant with FIWARE) and metadata (DCAT-AP). DCAT-AP [3] is used as the common metadata schema to describe datasets in URBANITE. Two repositories are used, one for the metadata and the other for the transformed data. This is shown in Figure 4.

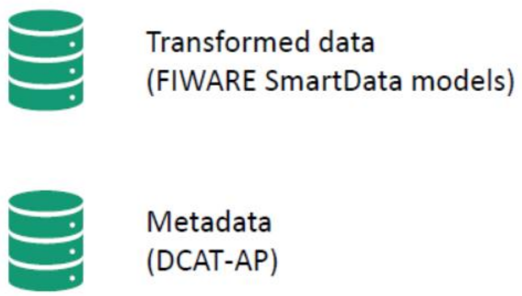

Figure 4. Data Storage & Retrieval repositories

The main concepts of the DCAT-AP model are catalogues, datasets and distributions. A catalogue represents a collection of datasets; a dataset represents a data collection published as part of a catalogue; and a distribution represents a specific way to access to specific data (such as a file to download or an API). This relationship is shown in Figure 5.

The concept dcat:Dataset informs about the title, description, access rights, creator, frequency, spatial/geographic and temporal coverage, spatial and temporal resolution, publisher, etc.

The concept dcat:Distribution provides metadata about the distribution, e.g. the property dcat:accessURL provides the information about how to access to specific data. Other important metadata related to the distribution are, for instance, the license, a description, the format of the data (e.g. CSV, JSON), etc.

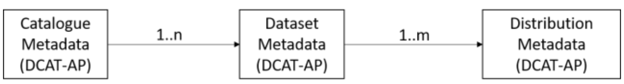

Figure 5. Simplified DCAT-AP Model

The Data Storage & Retrieval component provides REST APIs so that the Exporter of the harvesting process can store the transformed data. Besides, a new DCAT distribution is stored which is associated with the existing metadata of the dataset, with accessUrl equal to the API endpoint to access the transformed data. An example of this is shown in Table 1.

Table 1. Instance of a distribution in JSON-LD format

{
"@id" : "https://urbanite-project.eu/ontology/distribution/009b9f0e-e780-4e9d-8153-520dc8943195",
"@type" : "dcat:Distribution",
"description" : "Air Quality information for bilbao day 2021-05-01 in NGSI-LD representation",
"format" : "http://publications.europa.eu/resource/authority/file-type/JSON_LD",
"license" : "http://publications.europa.eu/resource/authority/licence/CC_BY",
"title" : "Air Quality information for bilbao day 2021-05-01",
"accessURL" : "https://bilbao.urbanite.esilab.org/data/getTDataRange/airQualityObserved/bilbao?startDate=2021-05-01T00%3A00%3A00.000Z&endDate=2021-05-01T23%3A59%3A00.000Z"
}

The technology stack used to implement the component consists of three levels, as depicted in Figure 6.

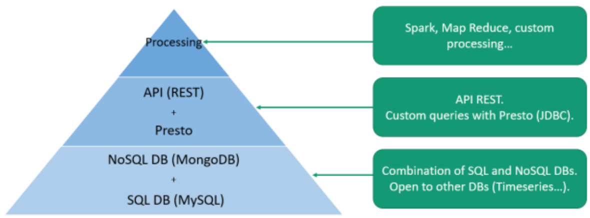

Figure 6. Data aggregation & storage technology stack

At the bottom level we have the storage repositories, being a combination of different types of databases: SQL databases, like MySQL and NoSQL databases, like MongoDB. The design is also open to the usage of other storage mechanisms that may be useful in the future, such as timeseries databases, files, semantic triple stores, etc.

The intermediate level offers the mechanisms for both storing and retrieving data. In turn it consists of two components: a REST API with predefined methods for inserting or accessing data and metadata, and a JDBC connection, through the Presto software, to perform custom queries (SQL statements) different to those offered by the API. All the interaction with the storage system is made through these two mechanisms, not allowing direct access to the data. This makes the choice of the specific database that stores the data transparent to the upper processing layer and can be modified without affecting the processes that make use of the component.

Finally, at the top level we have the processes that can be defined to feed the databases or make use of the data, e.g. data aggregation processes.

3.4 Data Catalogue

The Data Catalogue offers the functionalities to discover and access the datasets collected and managed by the components of the URBANITE ecosystem. Apart from the possibility to search over these datasets, the Data Catalogue also offers the possibility to search useful data across external "federated catalogues" (such as Open Data Portal) to increase the chance to find useful data.

The administrator is in charge of managing the federation of the catalogues, where a catalogue represents a data source. He/she can add new catalogues, delete or edit the existing ones. Moreover, the administrator can manage the platform configurations.

The end-user is then able to perform a federated metadata search among the harmonized DCAT-AP datasets provided by the federated catalogues. Moreover, the end-user can perform SPARQL queries over the federated RDFs provided by the federated catalogues, or he/she can access to statistics about the federated catalogues.

The Data Catalogue exposes APIs to access its functionalities; thus, an external system will be able to interact with the platform using such APIs.

The Data Catalogue is based on Idra [4]. Idra is a web application able to federate existing Open Data Management Systems (ODMS) based on different technologies providing a unique access point to search and discover open datasets coming from heterogeneous sources. Idra uniforms the representation of collected open datasets, thanks to the adoption of international standards (DCAT-AP) and provides a set of RESTful APIs to be used by third-party applications.

Figure 7 depicts the interaction among the Data Catalogue and the other URBANITE's components.

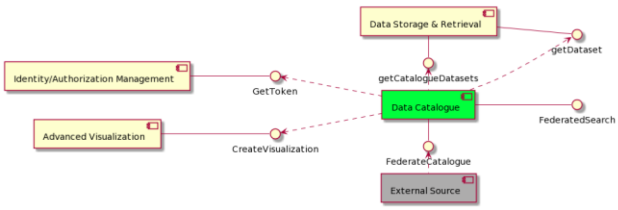

Figure 7. Data Catalogue - Component diagram

The Data Catalogue interacts with 1) Identity/Authorization Management component to allow administrators to access their specific functionalities retrieving the access token that will be further provided to the APIs, 2) Advanced Visualization to build visualization taking advantage of the DCAT-AP distributions it manages and 3) Data Storage & Retrieval to retrieve DCAT-AP datasets and distribution metadata. Finally, the Data Catalogue component is able to federate external sources such as Open Data portals or other sources providing DCAT-AP metadata.

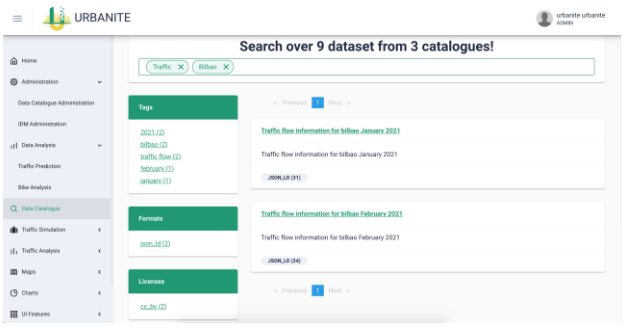

Figure 8. Data Catalogue – Dataset search (example)

Figure 9. Data Catalogue – Details of a dataset (example)

ACKNOWLEDGMENTS / ZAHVALA

This research is funded by the European Union's Horizon 2020 research and innovation program under grant agreement number 870338 (URBANITE: Supporting the decision-making in urban transformation with the use of disruptive technologies).

REFERENCES

[1] Kirstein F., Stefanidis K., Dittwald B., Dutkowski S., Urbanek S., Hauswirth M. (2020) Piveau: A Large-Scale Open Data Management Platform Based on Semantic Web Technologies. In: Harth A. et al. (eds) The Semantic Web. ESWC 2020. Lecture Notes in Computer Science, vol 12123. Springer, Cham. https://doi.org/10.1007/978-3-030-49461-2_38

[2] Piveau solution. https://github.com/piveau-data

[3] DCAT-AP 2.0.1. https://joinup.ec.europa.eu/rdf_entity/http_e_f_fdata_ceuropa_ceu_fw21_f32d70b6e-0d27-40d9-9230-017e4cd00bcc

[4] Idra - Open Data Federation Platform https://idra.readthedocs.io/en/latest/

Traffic Simulation for Mobility Policy Analysis

Maj Smerkol
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia
maj.smerkol@ijs.si

Miljana Sulajkovska
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia
miljana.sulajkovska@ijs.si

Erik Dovgan
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenija
erik.dovgan@ijs.si

Matjaž Gams
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenija
matjaz.gams@ijs.si

ABSTRACT

Recently urban mobility has been changing quickly due to the growth of cities and novel mobility methods' introduction. These changes are causing ever greater traffic congestion and urban pollution problems. To deal with the growing complexity of urban mobility and traffic systems we are developing a system to support the decision makers in the URBANITE H2020 project. An integral part of the system is a system for simulation of mobility policy proposals based on traffic simulation. The simulations are used for evaluation of policy proposals. We developed a system for automatic simulation creation and an algorithm for population synthesis based on open data available for multiple cities.

KEYWORDS

smart city, traffic simulation, mobility policy

1 INTRODUCTION

As cities are becoming more populous, traffic congestions, pollution and other problems are becoming harder to handle. Such complex and interconnected issues, also called wicked problems, are hard to deal with and any policy targeting these issues may be seen as undesirable from certain stakeholders' point of view or may have unforeseen side effects. An example of a wicked issues is moving the residents from using cars to driving bicycles or using public transit [6].

This paper presents a method for using traffic simulations among other analysis tools to analyse and evaluate mobility policies. The developed method aims to contribute to solving such problems in the urban mobility domain by simulation of the changes and calculation of key performance indicators (KPIs) [7], co-designed with the city stakeholders.

The method was developed with two goals in mind. The first is to empower the administration to easily run new simulations with less involvement of technical experts and thus shorten the feedback time from idea to the results of the simulations. The second goal is to enable the automatic creation of multiple simulations by variation of specific parameters within given constraints to algorithmically produce candidate solutions for specific problems.

The rest of the paper is organized as follows. Section 2 covers mobility policy simulations and overviews the selected traffic simulation tool. Section 3 describes the high-level view of the support system for mobility policy design and the role of the mobility policy simulation in the system. Section 4 overviews the process of partially automated simulation creation including the descriptions of data preparation processes, the design of the underlying relational database and the algorithms developed for each step of the simulation creation. The paper concludes with Section 5, which summarizes the paper and presents ideas for future work.

2 OVERVIEW OF THE SYSTEM FOR MOBILITY POLICY DESIGN

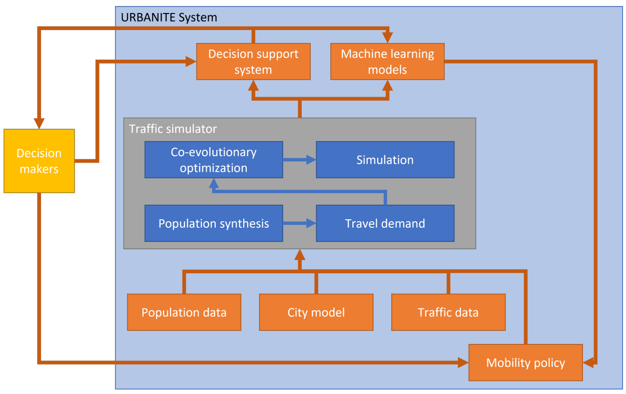

Figure 1: Architecture of the URBANITE system.

The URBANITE system consists of several components, including a data platform, AI-based tools including the mobility policy simulation, and tools for stakeholder engagement, including a forum and a social policy laboratory [11]. In this paper we focus on the architecture of the mobility policy simulation shown in Figure 1.

The decision makers work with the system in an interactive mode by evaluating and improving policy proposals in an iterative fashion, i.e., by defining the mobility policy proposals, which are simulated and evaluated by the system. In each iteration, they use the insight gained to modify the policy proposals. However, the system is also able to search for mobility policy proposals within user provided constraints. These proposals are automatically simulated and evaluated, and the decision makers are presented with a selection of the best ones according to the selected KPIs.

The main inputs to the mobility policy simulation system are the proposed mobility policy and the data required for the simulation: the population data, the city model and the traffic data. Examples of mobility policies are described in Section 3.

The mobility policy simulation module executes the following steps to create and run the simulation:

- Population synthesis is the process of using the population data to create the artificial agents.
- Travel demand generation takes the generated agents and their activities, and generates the trips that the agent will take to arrive to the locations of their activities.
- Finally, before the simulation is run the trips are optimized to fit them to the known traffic data.
- The simulation run is performed and the simulation recorded for further analysis and visualization, and to provide data for the machine learning models.

The simulation results are used by the decision support system to calculate the KPIs, evaluate the mobility policy proposals and provide multi-attribute decision analysis, as well as by the machine learning models used for policy proposal.

3 MOBILITY POLICY SIMULATION

Historically, modeling of mobility started with analytical modelling, mostly based on economic models combined with numerical methods for traffic estimation [3]. At the same time, simulation models were also developed [12]. A significant improvement came with the introduction of agent-based simulations, which are used for simulation of complex self-organizing systems.

3.1 Traffic Simulations

The URBANITE mobility policy simulations are based on traffic simulations provided by the open source package MATSim [4], a Java framework for traffic simulations. The selected traffic simulation framework is a state-of-the-art microscopic multi-agent traffic simulation package that allows the creation of traffic simulations with features such as multi-modal trips, support for bicycles and micro-mobility, public transit support, and emissions estimation.

Some drawbacks of the MATSim framework are high computational complexity, demand for high quality input data[1] and highly complex process for creating high quality simulations. The selection of MATSim among the available microscopic traffic simulation software options is based mostly on its extensibility and flexibility.

3.2 Representation of Mobility Policies

Mobility policy is a very wide category and general policy representation is out of the scope of this work. Instead we focus on specific policies. Besides the policy representation, appropriate KPIs are also required. This section focuses of four distinct types of policies that are considered by the pilot cities within the URBANITE project and the KPIs selected to evaluate them.

3.2.1 Closing a Major Square for Private Car Traffic. Bilbao is a city near the northern shore of Spain and the largest city of the Basque Country with nearly 350,000 residents. The policy proposed for simulation is closure of the Moyua square in the city for all private car traffic. Main goals of the proposal are to improve the air quality at the square and to relieve traffic trough the square.

From the simulation point of view, this is a relatively simple change of the city's road network. The change is implemented by changing the properties of affected road segments in the network to disallow private car traffic. Public transport and emergency vehicles are not affected by this change as well as pedestrian and bicycle traffic.

KPIs are selected in accordance with the goals of the policy. To estimate the effects on the air quality in the area, the daily amounts of different air pollutants emitted at the square and in the nearby areas are recorded. We expect that there will be less pollutants emitted in the square and some increase in the surrounding areas as the traffic will be redirected to them.

3.2.2 Changing a Major Road to a Bicycle Highway. Amsterdam is the capital city of the Netherlands as well as its largest city with over 1.5 million residents in the urban area. It has a highly developed bicycling culture to the point where bicycle traffic jams often form at the peak traffic times. The policy proposed is to close one of the major roads into the center, the Oranje Lopper, for motorized traffic. While similar to policy proposed in Bilbao, the goals of Amsterdam are mainly to alleviate the bicycle traffic by introducing a new bike highway into the city center.

To represent the policy for the simulation, we change the properties of the road segments that make up the Oranje Lopper to disallow private car traffic and instead introduce a number of new bicycle lanes.

KPIs selected are the number of bicycles using the new bike highway and more importantly, average bicycle travel times between the city parts connected by the Oranje Lopper.

3.2.3 Alleviation of Ferry Traffic via Building of a new Tunnel. Helsinki is the capital city of Finland. The Helsinki port is also the busiest passenger port in the world, which causes regular traffic jams in the Jätkäsaari area where the traffic from the port to the mainland is forced to use a single road. The policy proposal in Helsinki includes building a tunnel connecting the port directly to the motorway with the goal of alleviating the traffic jams that form periodically when ferries arrive to the port.

This policy is represented by the addition of new links to the road network representing the tunnel. To test this proposal, the ferry arrivals are modelled as seafaring public transport with forced high loads of vehicles arriving as scheduled.

The main KPI for this policy is the traffic flow at the existing point of crossing to the mainland. Another significant factor for the evaluation of this policy is the amounts of air pollutants emitted in the Jätkäsaari area.

3.2.4 Addition of new Bus Lines to Under-Connected Areas. The last policy we consider is the addition of public transport lines to under-connected areas of the Messina municipality in Sicily, Italy. Generally the city is well connected by public transport, however as the city is caught between the sea shore and a mountain area, some of the more remote parts lack connectivity and are only accessible by private vehicles. These areas are also generally too remote to access the city by foot and too mountainous for everyone to use bicycles.

To simulate this policy, we add new bus lines by creating the GTFS data compatible with existing public traffic and including it in the simulation.

[1]In the context of URBANITE project, which this work is a part of, data gathering and quality assurance are parts of the project.

4 CREATING SIMULATIONS

To create a traffic simulation using MATSim, a set of input data needs to be defined, and the simulator's configuration must be specified. The simulation is run in two consecutive steps: first, the agents' actions are optimized using a co-evolutionary algorithm; second, the final run of the simulation is stored and analysed.

4.1 MATSim Input Data

In this section we describe the input data necessary to run the simulation, and the process for creating these files. For each of the files we also describe the data model and explain how the data is used.

4.1.1 Road Network. The road network represents the traffic infrastructure and its properties such as lane capacity and max speed. Currently, we rely on Open Street Maps [10], a crowd-sourced publicly accessible map database. It is a very valuable resource, however due to its nature it is not complete and there may be some inaccuracies in the data.

The resulting network is a collection of nodes and link. Each link represents a straight part of a single lane and connects two nodes. The link also contains other relevant information such road type, speed limit, road or street name, etc. The nodes represent the location via coordinates. This means that a single road or street is made of multiple links that may have different properties.

4.1.2 Facilities. To get the data about specific places in the city, such as hospitals, schools, parks and workplaces among others, we provide the simulator with a list of facilities. These are gathered from OSM along with the network itself and attributes are added from the city datasets. The attributes of interest include number of employees, average number of daily visitors, number of employee and visitor parking spots, etc. Before running the simulation, the facilities and the network are pre-processed together to match the coordinates and other attributes between the files.

4.1.3 Agent Plans. Agent plans are the daily plans of each of the agents that represent the population. These are described as a list of activities and a list of trips that allow the agent to engage in the activities. The agent plans are the results of the population synthesis and travel demand modelling, described in Section 4.4. There are multiple algorithms for population synthesis that are developed for working with different sets of available input data [2].

The trips for each agent are used for the final simulation run. The data contains a collection of trips, each made up of multiple trip legs that may use different transportation modes (e.g., an agent may take a bus to go to work but walk back home).

4.1.4 Vehicles. When interested in vehicle-related data, such as amounts of certain pollutants emitted, data about the vehicles in the city are necessary. Multiple types of vehicles can be defined with attributes such as vehicle type, engine technology, cylinder displacement and latest EURO emission standard it supports. A simulation using this data can be analysed for amounts of pollutants emitted per link for each vehicle by using the HBEFA [9] emission factors.

To use the defined vehicles, the agent population has to be split up into multiple subpopulations. Each subpopulation may link a specific vehicle fleet (a set of vehicles) that the agents can use.

4.2 Automating Simulation Creation

To automate the creation of mobility policy simulations, the following steps were taken:

- The entities representing input data are connected via appropriate relations and a relational database is designed. The input data is related to a simulation instance.
- Processes and algorithms for creating the input data are developed and implemented.
- Simulation results are stored and exported for visualization and further analysis.

We defined a database that allows the simulations to be automatically created and compared using a multi-attribute decision analysis methodology. To allow the user easy comparison of different simulation outcomes, the table `Scenario` includes multiple simulations. Each `Scenario` links to a `Decision Model`, that is designed and evaluated using the multi-attribute decision analysis tool DEXi [1]. This setup of entities `Scenario`, `Simulation` and `Decision Model` allows for a common and automated evaluation and comparison of different simulation results.

At the same time, the `Simulation` entity is linked to entities `Network`, `Agent Plans` and `Vehicles`. These include all the data that is needed to run the simulations.

4.3 Road Network Preparation

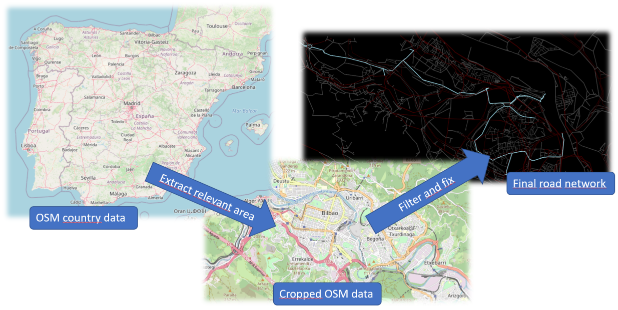

Figure 2: (1) The country wide map is retrieved from OSM. (2) The relevant area is extracted. (3) The map data is filtered and the minimal road network is stored.

OSM is limited by the area size it can export using the API. Instead, we download the binary map data for the entire country from the map catalogue. We use an open source tool Osmosis to extract the relevant area and filter out all the unneeded data in order to keep performance of the simulations to a minimum. Next, we remove any broken links, unconnected links and parts of network that are isolated from the greater connected network, usually artifacts of extracting the selected area.

4.4 Population Synthesis Algorithm

The selection of the algorithm is limited by the data available in the four pilot cities. Another important goal of the algorithm selection is to use mainly open data. Often the studies in this field are not reproducible due to use of proprietary data or data that is not publicly available, as well as the use of proprietary software. The common population model used for population synthesis is shown in Figure 3. The city is split into existing statistical districts and each district is modeled separately. Each district has a population model and a corresponding vehicle fleet used for estimation of air pollutants emitted.

We adapted an algorithm developed for population synthesis using publicly available data in Paris [5], in order to process the data available in the pilot cities. It consists of the following steps:

(1) Sample the marginal distributions of the socio-economic data. Each household is assigned a home location and agents are generated for the household by sampling the marginal distributions of the population attributes.
(2) Iterative Proportional Fitting [8] is used to improve the matching of the agents' attributes by fitting to a small sample of the census data.
(3) Households are assigned income levels sampled from the income level marginal distribution.
(4) Activities are generated and activity chains are assigned to each agent. First, the primary (work and education) activities are considered, then secondary (shopping and leisure) activities are added, based on the travel surveys and facility data.

The lists of households, persons, activities and trips generated need to be optimized to match the traffic data before the final simulation. The initial version of the algorithm was already developed and is based on the open-source implementation of the algorithm described in [5], while the final version is under development.

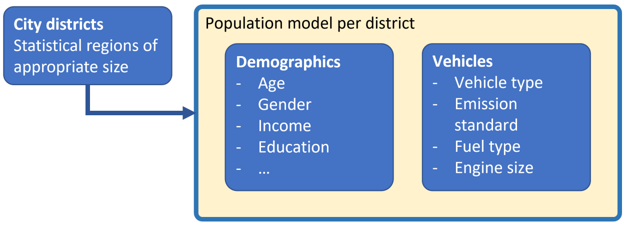

Figure 3: Each district has a separate population model and vehicle fleet.

5 CONCLUSION

We are developing a mobility policy simulation module as a part of the URBANITE system for mobility policy design support. The designed module consists of an open source multi-agent traffic simulation system, population synthesis algorithm including travel demand modelling and the co-evolutionary optimization algorithm for fitting the simulations to existing traffic data.

The system enables mobility policy simulation by implementing the processes for creating the simulations using open data and with no proprietary software required. Using open data allows the users to develop algorithms applicable to multiple cities and ensures the reproducibility of results.

The algorithm for population synthesis and travel demand modelling was selected and adapted to the data available and pilot cities' needs, and preliminary simulation were developed.

The research on this topic is far from concluded. Some of the future work include development of the co-evolutionary simulation fitting algorithm and the final implementation of the population synthesis algorithm.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870338.

REFERENCES

[1] Marko Bohanec. 2008. Dexi: program for multi-attribute decision making user's manual. *Ljubljana, Slovenia: Institut Jozef Stefan.*

[2] Abdoul-Ahad Choupani and Amir Reza Mamdoohi. 2016. Population synthesis using iterative proportional fitting (ipf): a review and future research. *Transportation Research Procedia*, 17, 223–233. International Conference on Transportation Planning and Implementation Methodologies for Developing Countries (12th TPMDC) Selected Proceedings, IIT Bombay, Mumbai, India, 10-12 December 2014. ISSN: 2352-1465. DOI: https://doi.org/10.1016/j.trpro.2016.11.078. https://www.sciencedirect.com/science/article/pii/S2352146516306925.

[3] Juan de Dios Ortúzar and Luis G Willumsen. 2011. *Modelling transport*. John wiley & sons.

[4] Andreas Horni, Kai Nagel, and Kay Axhausen, editors. 2016. *Multi-Agent Transport Simulation MATSim*. Ubiquity Press, London, 618. ISBN: 978-1-909188-75-4, 978-1-909188-76-1, 978-1-909188-77-8, 978-1-909188-78-5. DOI: 10.5334/baw.

[5] Sebastian Hörl and Milos Balac. 2021. Synthetic population and travel demand for paris and île-de-france based on open and publicly available data. *Transportation Research Part C: Emerging Technologies*, 130, 103291. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2021.103291. https://www.sciencedirect.com/science/article/pii/S0968090X21003016.

[6] Louise Kold-Taylor and Donald W de Guerre. 2020. From cars to bicycles: an ecosystem view of montreal traffic as a wicked problem. *Systemic Practice and Action Research*, 33, 1, 55–75.

[7] Puji Adiatna Nadi and AbdulKader Murad. 2017. Review of methods and indicators in sustainable urban transport studies overview from 2000 to 2016. *Communications in Science and Technology*, 2, 2.

[8] Paul Norman. 1999. Putting iterative proportional fitting on the researcher's desk.

[9] Benedikt Notter, Mario Keller, Hans-Jörg Althaus, Brian Cox, Wolfram Knörr, Christoph Heidt, Kirsten Biemann, Dominik Räder, and Marie Jamet. [n. d.] Hbefa 4.1.

[10] OpenStreetMap contributors. 2017. Planet dump retrieved from https://planet.osm.org. https://www.openstreetmap.org. (2017).

[11] Anne Fleur van Veenstra and Bas Kotterink. 2017. Data-driven policy making: the policy lab approach. In *International conference on electronic participation*. Springer, 100–111.

[12] Michal Čertický, Jan Drchal, Marek Cuchý, and Michal Jakob. 2015. Fully agent-based simulation model of multimodal mobility in european cities. In *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 229–236. DOI: 10.1109/MTITS.2015.7223261.

Machine Learning-Based Approach for Estimating the Quality of Mobility Policies

Miljana Sulajkovska
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia
miljana.sulajkovska@ijs.si

Maj Smerkol
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia
maj.smerkol@ijs.si

Erik Dovgan
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenija
erik.dovgan@ijs.si

Matjaž Gams
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenija
matjaz.gams@ijs.si

ABSTRACT

Cities are increasingly turning towards specialized technologies to address issues related to their significantly increased transport demand. Municipalities and transport authorities try to face these problems in order to achieve their objectives by taking various actions in the domain of public transport, air and noise pollution, road accidents, etc. The primary objective of this research is to explore the role of machine learning (ML) in mobility policy quality estimation using microscopic traffic simulations. The main idea is to use one simulation run as one training example. The features are represented by several group of parameters that are related to the input and output of the simulation, while the target variables are represented using key performance indicators (KPIs). The city of Bilbao is chosen as a use case. We have analyzed how closing the Moyua square in the city center and changing the number of cyclists there can affect the air pollution by estimating the CO2 emissions. Several machine learning algorithms are tested and the results show that by closing the main square in the city center and increasing the number of cyclists the CO2 emissions reduce.

KEYWORDS

machine learning, smart cities, mobility policy

1 INTRODUCTION

According to United Nations population estimation, the total population is exponentially increasing and by 2050 will reach 9 billion, i.e. it will increase for 2 billion from now [11]. This demographic growth will greatly impact on the transportation system in metropolitan areas since most population will be located there. As a result far more attention must go towards serving the needs and aspirations of the people with the aim to maintain the environmental, social, and economic costs at the same time [12].

In this context different mobility policies are tested and evaluated in order to achieve the desired city goals. Since implementing different scenarios in real life is an expensive process microscopic traffic simulations are widely used as a valuable support tool for evaluating transportation facilities or systems. Using the simulations we can see how some actions may impact the dimensions that we are interested in without making those changes in real life.

Currently, most of the mobility policy evaluation techniques rely on experts in urban/spatial planning using on simulation results [10]. Since the simulations create large amount of data, including data from optimization steps, various data analysis can be applied. In this context machine learning techniques can be applied to automate the evaluation of mobility policies and address the objectives of the cities.

As part of the URBANITE project we are developing a machine learning module using data from microscopic transport simulator that will help decision makers in the what-if analysis. More precisely, we propose a system to estimate the quality of previously simulated mobility policies using machine learning methods.

The rest of the paper is structures as follows. Section 2 explains the URBANITE approach and the relevant modules for this research. In Section 3 the data collection process is explained. Then, Section 4 presents the results of the machine learning module. Finally, Section 5 concludes the paper with ideas for future work.

2 OVERVIEW OF THE URBANITE APPROACH

The main objective of URBANITE approach is to build an intelligent platform that can use data from heterogeneous sources in order to help the city managers in the decision-making process. To achieve this aim, several modules are developed. In this section we will give an overview of only the relevant ones shown in Figure 1

The traffic simulator is used to simulate various mobility policies during the the policy evaluation process. The input files to this module are related to the network map, travel demand, public transit data etc. Based on the simulation output, target variables e.g. air pollution levels for the machine learning algorithms are calculated, based on which the models are build. This approach is able to process large amount of data in order to find the best mobility policy.

Besides automatic selection of mobility policy URBANITE also supports policy selection by the experts with the use of the decision support system. In addition to processing the simulation data this system also relies on expert knowledge in order to build the hierarchical decision models and satisfy the user preferences.

Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia

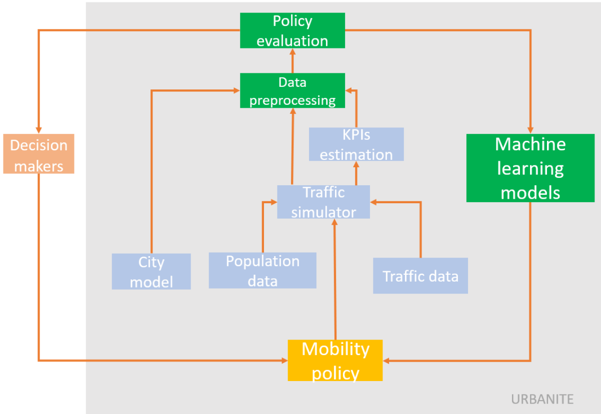

Figure 1: Modules of URBANITE approach.

Both approaches are used in the policy evaluation process with the difference that machine learning module relies on algorithms that automatically select the best mobility policy. In the next sections the simulation and machine learning process are described in detail.

3 DATA COLLECTION

3.1 Simulation

In order to collect the data, a microscopic traffic simulation tool was used. Several state-of-the-art solutions were tested and MATSim was chosen as the most suitable one. MATSim [6] is an open-source tool implemented in Java. It is used for microscopic modeling that enables us to simulate and analyze components on the network such as traffic flow, congestion, public transport, behavior of cyclists, etc. One of the core concepts is the co-evolutionary optimization where the individuals' plans are evolving in the presence of all other persons doing the same.

To run the simulator several input files need to be provided that are related to the city model, traffic and population data. For the creation of the transport demand e.g. persons with their daily plans and mode of transport real data from census and other travel surveys is required. Since there is no complete dataset containing the socio-demographic characteristics of individuals at a small geographic scale because of privacy concerns a transport demand was generated based on known random variables.

After providing all the required input we can run the simulation which is optimized by configurable number of iterations (see Figure 2). Each individual agent learns by maintaining multiple plans which are scored by executing them in the mobsim, selected according to the score and when needed, modified. The iterative process consists of the following steps:

- Mobsim simulation
- Scoring
- Replanning

Every iteration starts with an initial demand simulated by the mobility simulation and then evaluated by the scoring module as a central element of the simulator [9].

The MATSim scoring module evaluates the performance of a plan in a synthetic reality and determines the choice of person's plan in the next iteration. Next, only plans with higher scores are selected by the agent - others are deleted in the replanning step. The scores are computed using scoring function taking into

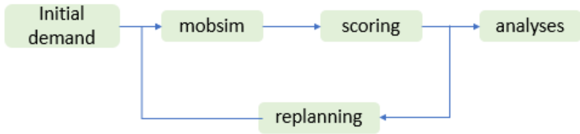

Figure 2: MATSim cycle.

account the performance of activities and travel time. A typical score is calculated as follows:

$$S_{\text{plan}} = \sum_{q=0}^{N-1} S_{\text{act},q} + \sum_{q=0}^{N-1} S_{\text{trav,mode}(q)} \quad (1)$$

Here utility functions are used to represent a basic scoring function or in other words the utility of a plan S_{plan} is computed as sum of all activity utilities $S_{\text{act},q}$, plus the sum of all travel utilities $S_{\text{trav,mode}(q)}$. N represents the number of activities. For scoring, the last activity is merged with the first activity to produce an equal number of trips and activities. Positive scores are obtained for desired events and negative for unwanted ones.

Finally, the optimization step takes part where four dimensions are considered: departure time, route, mode, and destination. Each of the agents has a memory of M plans that have been observed in the past and which have obtained a score. In the first step the replanning process checks whether the agent's memory exceeds the limit. If so one of the existing plans is removed according to previously computed scores. If the plan is removed that was currently selected for execution, a random one among the remaining ones is selected. After a certain number of iterations an equilibrium state on the network is reached improving the initial scores.

Several files are produced as output of the simulation that are related to specific iteration or they summarize a complete run, e.g. events file that contains every action taken on the network, and travel distance statistics showing the distance traveled per mode. These results are used to compute the features for the machine learning module and to define the target variable. More precisely the input features consists of simulation input and output data which can be directly influenced by the user. On the other hand the target variables depend on the simulation results and cannot be directly set by the user which makes them relevant for the decision-making process of particular mobility policy. These variables are summarized in Table 1.

Additional MATSim package was used to calculate the CO_2 emissions which are used as a target class in the prediction process. The tool calculates warm and cold-start exhaust emissions by linking MATSim simulation output to detailed emission factors for road transport.

3.2 Scenarios

In order to gather the dataset, 14 simulations were executed by applying different policies in the city of Bilbao, Spain. The main objective is to see the impact of closing the Moyua square in the city center for private traffic.

Two scenarios are implemented: the baseline scenario of the current network situation and the modified scenario representing the closure of Moyua square as one possible policy. All other

Table 1: Input data and target variable of the machine learning module

ML input		Target variable
Sim input	Sim output	
Surface of a road	Number of cars	CO_2 emission
Capacity of road	Number of cyclists	
Number of lanes	Number of public transport vehicles	
Type of district		
Number of bus stops	Average travel time	

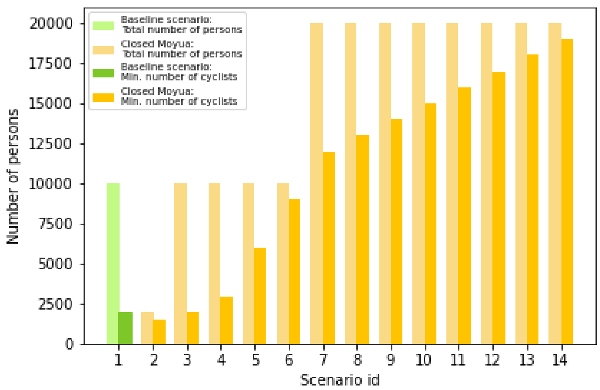

Figure 3: Number of persons and cyclists per scenario instance.

instances are variations of the second scenario where the number of cyclists varies from 1500 to 19000 while changing the number of inhabitants from 2.000 to 20.000, respectively. This change in number of cyclists does not represent a specific policy but shows what would happen if in conjunction with the applied policy also the number of cyclists changes. Figure 3 depicts these variations where with green is marked the baseline scenario and with orange the other scenario and all its variations. The first instance represents the baseline scenario with 10000 inhabitants and 2000 cyclists, while the remaining instances represent the second scenario with all variations. The second instance contains 2000 inhabitants with at least 1500 cyclists. The next four instances are representing 10000 population with up to 9000 cyclists while reducing the private transport. The rest of them represents 20000 inhabitants with up to 19000 cyclists. The number of public transport vehicles stays the same in all variations.

Figure 4 shows the results of the applied policy. More precisely it shows the relationship between number of cyclists and the level of CO_2 emissions. The x-axis represents the number of cyclists nearby the square and and y-axis represents number of cyclists in the center. The different colors denote the amount of CO_2 emissions as a target variable where with orange circle is marked the baseline scenario. This figure shows that by closing the main square for private traffic and reducing the number of private vehicles nearby it, the level of CO_2 emission is decreasing.

4 MACHINE LEARNING

4.1 Methods

Several machine learning models were applied using Orange [3]: k-Nearest Neighbors, Decision Tree, Support Vector Machines,

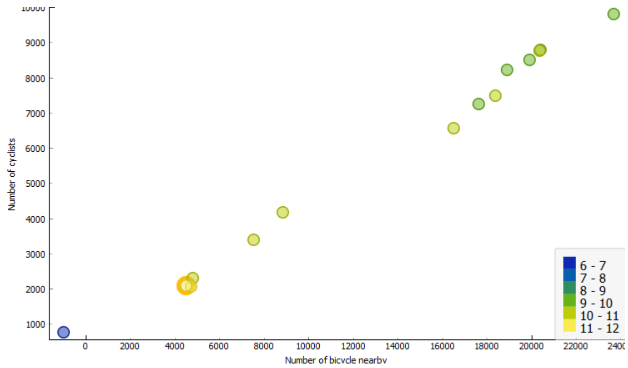

Figure 4: CO_2 **emissions in the Moyua square. The baseline scenario is marked with orange circle.**

Random Forest, Linear Regression, Gradient Boosting, and Neural Network.

The k-nearest neighbor (kNN) is a semi-supervised learning algorithm that requires training data and a predefined k value to find the k nearest data based on distance computation. If k data have different classes the algorithm predicts class of the unknown data to be the same as the majority class [1].

Tree splits the data into nodes by class purity. The top-most node is called root, the bottom ones leaves, and all other nodes are internal nodes connected to each other with edges. Each edge represents satisfaction of the node condition, and each leaf node determines the class assigned to the instances that met the conditions of the internal nodes on the path from the root node to the leaf node [5].

Support Vector Machines (SVM) is a two-grouped classifier where input vectors are non-linearly mapped to a high-dimension feature space. In this feature space a linear decision surface is constructed. Special properties of the decision surface ensures high generalization ability of the learning machine [2].

Random Forest consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes model's prediction. A large number of relatively uncorrelated models (trees) operating as a committee outperform any of the individual constituent models.

Linear Regression is commonly used in mathematical research methods, where it is possible to measure the predicted effects and model them against multiple input variables. It is a method of data evaluation and modeling that establishes linear relationships between variables that are dependent and independent [8].

Gradient Boosting tries to convert weak learners into strong ones by training many models in a gradual, additive and sequential manner where the gradient of the loss function is being minimized, with respect to the model values at each training data point evaluated at the current step [4].

Neural Network model simulates a large number of interconnected processing units that resemble abstract versions of neurons where the processing units are arranged into layers. The units are connected with varying connection strengths (or weights). Input data are presented to the first layer, and values are propagated from each neuron to every neuron in the next layer. Eventually, a result is delivered from the output layer. The network learns by examining individual records, generating a prediction for each record, and making adjustments to the weights whenever it makes an incorrect prediction. This process is repeated many times, and the network continues to improve its predictions until one or more of the stopping criteria have been met [7].

4.2 Evaluation Results

We have evaluated the machine-learning algorithms described in Section 4.1. The data for evaluation of the algorithms is split randomly 10 times and the average results are computed. We have compared four evaluation metrics:

- Mean squared error (MSE) measures the average of the squares of the errors or deviations (the difference between the true and estimated values).
- Root mean squared error (RMSE) is the square root of the arithmetic mean of the squares of a set of numbers (a measure of imperfection of the fit of the estimator to the data).
- Mean absolute error (MAE) used to measure how close forecasts or predictions are to eventual outcomes.
- R2 is interpreted as the proportion of the variance in the dependent variable that is predictable from the independent variables.

The results are shown in Table 2. According to MSE, RMSE, and R2 the best model is kNN, while according to MAE the best model is SVM.

Table 2: Evaluation results

Model	MSE	RMSE	MAE	R2
kNN	4.718	2.172	1.416	-0.372
Tree	5.387	2.321	1.431	-0.567
SVM	4.953	2.225	1.298	-0.441
Random Forest	5.093	2.257	1.404	-0.482
Neural Network	11.264	3.356	2.583	-2.277
Linear Regression	8.076	2.842	2.041	-1.349
Gradient Boosting	5.193	2.279	1.324	-0.511

5 CONCLUSION

In this paper we showed how machine learning can be used in mobility policy evaluation helping the urban development in cities. As large amount of data is produced from simulations, machine learning techniques can be applied to automatically choose the best policy.

We defined the mobility policy for Bilbao. Then, using microscopic traffic simulation the two scenarios were implemented: baseline scenario of the current network state and the modified scenario representing closure of Moyua square for private traffic. In order to gather more data, variations of the second scenario were produced by changing the proportion of cyclists and private car users. After gathering sufficient data, machine learning techniques were applied to evaluate the performance of the policy. Changing the number of cyclists in combination with the second scenario showed that the level of CO_2 emissions can be decreased or in other words, the proposed policy proved fairly good.

In future work, more policies will be tested and evaluated using the proposed approach. Then, advanced machine learning and deep learning techniques will be applied to improve the current results. Finally, data from simulation runs in the optimization step can be used to expand the current dataset.

ACKNOWLEDGMENTS

This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870338. The authors also acknowledge the financial support from the Slovenian Research Agency (research core funding No. P2-0209).

REFERENCES

[1] Kittipong Chomboon, Pasapitch Chujai, Pongsakorn Teerarassamee, Kittisak Kerdprasop, and Nittaya Kerdprasop. 2015. An empirical study of distance metrics for k-nearest neighbor algorithm. In *Proceedings of the 3rd international conference on industrial application engineering*, 280–285.

[2] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning*, 20, 3, 273–297. DOI: 10.1007/bf00994018. http://dx.doi.org/10.1007/BF00994018.

[3] Janez Demšar, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, Miha Štajdohar, Lan Umek, Lan Žagar, Jure Žbontar, Marinka Žitnik, and Blaž Zupan. 2013. Orange: data mining toolbox in python. *Journal of Machine Learning Research*, 14, 2349–2353. http://jmlr.org/papers/v14/demsar13a.html.

[4] Jerome H. Friedman. 2002. Stochastic gradient boosting. *Computational Statistics and Data Analysis*, 38, 4, 367–378. DOI: 10.1016/s0167-9473(01)00065-2. http://dx.doi.org/10.1016/S0167-9473(01)00065-2.

[5] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2001. *The Elements of Statistical Learning*. *Springer Series in Statistics*. Springer New York Inc., New York, NY, USA.

[6] 2016. *Introducing matsim. The Multi-Agent Transport Simulation MATSim*. Ubiquity Press, 3–8.

[7] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature*, 521, 7553, 436–444.

[8] Dastan Maulud and Adnan M Abdulazeez. 2020. A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, 1, 4, 140–147.

[9] 2016. *A closer look at scoring. The Multi-Agent Transport Simulation MATSim*. Ubiquity Press, 23–34. DOI: 10.5334/baw.3. http://dx.doi.org/10.5334/baw.3.

[10] Soledad Nogués, Esther González-González, and Rubén Cordera. 2020. New urban planning challenges under emerging autonomous mobility: evaluating backcasting scenarios and policies through an expert survey. *Land Use Policy*, 95, 104652. DOI: 10.1016/j.landusepol.2020.104652. http://dx.doi.org/10.1016/j.landusepol.2020.104652.

[11] Gilles Pison. 2019. How many humans tomorrow? the united nations revises its projections. *The Conversation*, 1–6.

[12] Eleni I Vlahogianni, John C Golias, and Matthew G Karlaftis. 2004. Short-term traffic forecasting: overview of objectives and methods. *Transport reviews*, 24, 5, 533–557.

Visualizations for Mobility Policy Design

Maj Smerkol
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia
maj.smerkol@ijs.si

Miljana Shulajkovska
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia
miljana.sulajkovska@ijs.si

Erik Dovgan
Jožef Stefan Institute
Ljubljana, Slovenija
erik.dovgan@ijs.si

Matjaž Gams
Jožef Stefan Institue
Ljubljana, Slovenija
matjaz.gams@ijs.si

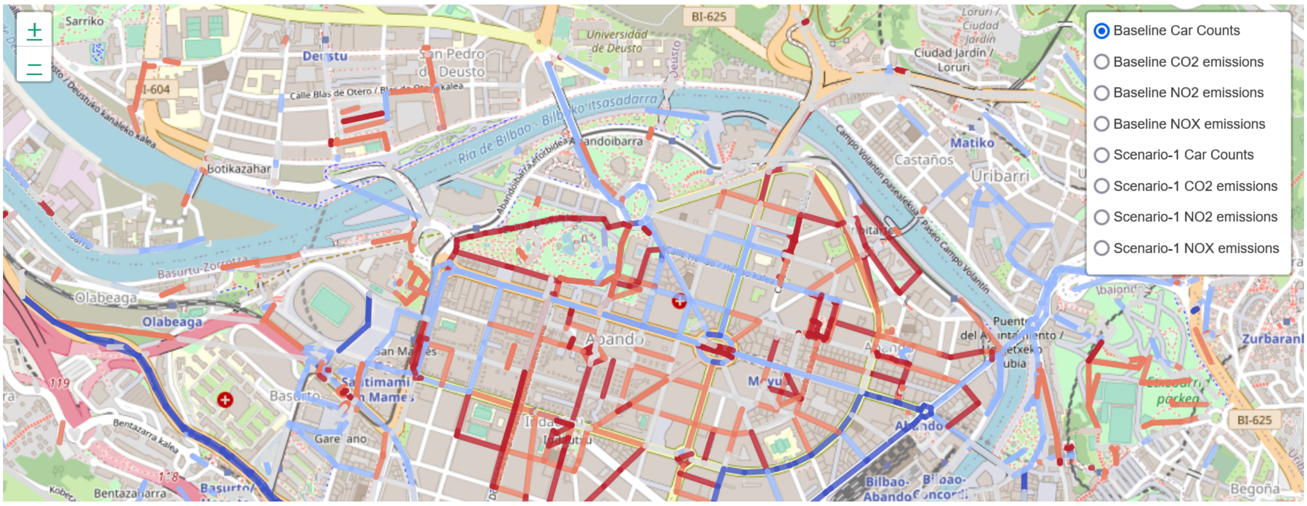

Figure 1: Visualization of simulated traffic flow intensity in center of Bilbao.

ABSTRACT

Cities around the world are rapidly gaining population as more people are moving to urban areas from the periphery. At the same time, novel urban mobility solutions are emerging such as e-cars and micro-mobility. Thus, urban traffic is getting heavier and more complex. To deal with these problems the H2020 URBANITE project is developing tools for city administrations including a data platform, mobility policy validation via traffic simulation, decision support for multi-attribute decision analysis and a visualizations module, described in this paper. We consider different types of visualizations in the domain of urban traffic and select most appropriate, implementing a module used for data visualizations for the system.

KEYWORDS

smart city, urban mobility, visualization

1 INTRODUCTION

Cities' mobility landscapes are rapidly changing due to the raising populations as well as new and disruptive mobility modes that are emerging. As the populations of cities are growing, so are occurrences of traffic congestions, air pollution and traffic noise in urban areas. At the same time, introduction of novel mobility modes both in the sector of micro-mobility (e-scooters, bike sharing etc) and sharing services in other sectors cause dynamics of the urban traffic to change [3]. This makes it increasingly difficult to model the growing complexity of the urban mobility as well as predict the effects of specific policies. This makes the prediction of possible effects of mobility policies more difficult but also more important. The importance of policy effects has been shown on the example of e-scooters [1]. They offer a clean and sustainable way of travelling the first and last kilometer of the trip but can also be very dangerous when not properly regulated.

We are developing a new AI assisted tool set for supporting the development and evaluation of urban mobility policies as part of the URBANITE H2020 project [7]. The project includes a mobility policy simulation, a recommendation system for policy design support, advanced visualization suite and a machine learning component for quick evaluation of expected policy results. This paper focuses on the visualizations of the traffic data, simulation results and decision analysis.

The end users of the tool set are city administrations (generally not technical personnel), urban mobility planners (experts on mobility) and interested citizens (laymen). These groups differ greatly with regards to their knowledge as well as the intent of interacting with the system. Citizens are interested in understanding the administration's actions but do not have a direct

Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia

possibility to work with the system. Urban mobility planners are mostly already using traffic simulation tools and data analysis tools and require a highly detailed view of the data and analysis results. They are not limited by their knowledge and are used to working with complex tool sets. The city administrators however are similar to the citizens in that they are generally not experts on urban traffic and mobility, but are expected to make decisions about policies. The visualizations developed are primarily aimed at the city administrators with the goal to help them understands and interact with the data available, compare the results of different policy proposals and help with the interaction of the experts with the administrators. The second goal of the visualizations is to inform the public and democratize mobility planning. To achieve that the visualizations should be self explainable.

The paper is organized as follows: Section 2 overviews the common visualizations and their advantages and disadvantages, Section 3 covers the selection of visualization methods for first release and Section 4 describes the implementation of specific visualization methods.

2 VISUALIZATIONS IN THE TRAFFIC AND MOBILITY DOMAINS

This section overviews the common data types in the domains of traffic and mobility and the visualization methods commonly used to represent the data. The data types overviewed are traffic flows, air pollution and specific pollutant emissions, general tabular data, geo-spatial data and geo-temporal data.

2.1 Spatio-Temporal Data

In the domains of urban mobility and traffic spatial data is very common, since most of the data is related to specific roads or locations within the city. This category contains all such data, including the road networks (city maps), traffic flows on roads and streets, trips made and modal splits of traffic in specific locations, as well as population properties in different statistical districts of the cities, locations of important facilities such as hospitals and schools, parking lots and public parking garages and public transport lines. Some of these types of data are further discussed bellow in the section Traffic Data.

These data are best represented as interactive layers on a map. Each layer must be visually distinct from the underlaying map to ensure the visibility of the data. Often it is less important that the map itself is easily readable as it serves mostly as a spatial anchor that allows the easy recognition of general locations. To ensure the understandability of the geo-spatial visualizations data layers can be interatively selected so only the relevant data is shown at the same time.

To keep the data minimally cluttered and therefore more understandable we do not show details of the data on the map unless the user hovers the mouse over some part. In this case, a popup with the details of the selected locations are shown. This can be seen on the Figure **??**, where the demographic data is shown for each district. Generally one of the attributes is shown using a color scale on the map and other attributes are only shown when user hovers over a specific city district.

The most intuitive way to show time-dependant data is by animating the visualiziations. To simplify the interaction and thus reduce the mental overhead we show a timeline below the geo-spatial view. The user may select the time they are interested in or play the animation at different speeds. An example is not shown due to the limitations of printed media.

2.2 Traffic Data

Traffic data includes traffic counts, trips, and Origin-Destination (OD) matrices. There are multiple ways of visualizing these data. Following is a brief overview of common visualization methods used on these types of data.

2.2.1 Traffic counts. Traffic counts at a specific location are commonly shown via line charts where the horizontal axis represents time (usually one day) and the vertical axis represents number of vehicles passing the location. To compare data from several specific locations we can show multiple lines on the same chart using different colors.

To visualize the traffic counts all over the city simultaneously geo spatial map based visualizations are commonly used, either as point-based or line-based map layers. Point-based visualizations are best suited when the traffic counts are measured using existing sensing devices such as induction loops or smart cameras. Such sensing devices are usually not available on every road segment. In this case the points locate the sensors while the values measured are commonly color coded.

2.2.2 Traffic flows. Traffic flow is the amount of vehicles that pass a certain point on the road in a time slot. Traffic flows are commonly visualized using line-based map layers, where the traffic flow is represented either by line thickness or color. To specifically show the modal split of the traffic flows we can show them separately or at the same time. In the latter case it is best to use color codes to represent different types of vehicles and line thickness to represent the traffic flow.

2.2.3 OD matrices. OD matrices hold the information about number of people moving from parts of the city to other parts. Commonly the spatial resolution of the OD matrices matches statistical regions. OD matrices are commonly obtained using travel surveys, estimated using GPS traces of trips and public transit data.

Common method for visualizing the OD matrix data is to show the matrix as a heat map with rows and columns labeled with the name of the district. Such visualizations are hard to understand and under certain conditions can be cluttered, decreasing their readability.

Alternatively OD matrices can be shown on a map via connections between districts. The intensity of travel between two parts is commonly represented via the connection thickness, while color is typically used to distinct different connections. This method is easier to understand, but readability depends on the geographical positions of the districts.

2.3 Air pollution

Air pollution levels are usually visualized using heat map layers on top of the city map, and are therefore counted among the geo-spatial visualizations. Generally the most common method for visualizing air pollution is an air quality index heat map. Some of the advantages of visualizing air pollution as a heat map are high understandability and very low visual cluttering. A negative aspect of heat maps in this use case is that air pollution often does not spread equally in all directions due to air movements and buildings blocking the pollutants' paths. This is however not very important as the users are mostly interested in general pollution levels and in the case of the URBANITE project the levels of specific pollutant emissions.

3 SELECTION OF METHODS

The selection of methods to be implemented was based on the pilot city requirements, data availability and the available simulation outputs. In order to support comparing the measured data and the simulation results we are limited to using visualizations that are appropriate to both.

This section covers the visualization methods we have selected and is split by the type of data into traffic, air pollution, and other data visualizations and concludes with a brief discussion of the color maps chosen to represent the values.

3.1 Traffic data

The category of traffic data contains multiple different data types that have to visually represented using different methods. Some of the data that is shown using the methods for geo-spatial data visualization are:

- Traffic counts, shown either geo-spatially by aggregating the counts per day or geo-temporally by aggregation of the counts per specific time slot, commonly hours. Traffic counts at a specific location depends on time.
- Traffic flows, measures in vehicles per hour passing a road. The traffic flow at a specific location depends on time. The specific flows for different modes, such as public transport, heavy duty vehicles, bicycles and pedestrians are currently visualized separately.
- Congested roads. Simplest way of identifying problematic roads or junctions is to show the locations of congested traffic. We can detect congested traffic and traffic jams by searching for road segments with high traffic density and travel speed below the free-flow speed [8].

Some of the more detailed traffic data are better visualized using simpler charts. Traffic flows at specific locations over time are shown using a line chart. Traffic flow predictions at specific locations are shown using a line chart with the confidence interval included to inform the user that these are not exact. Modal splits of traffic at specific locations as well as city-wide aggregations of modal splits are visualized using area charts or stacked area charts.

3.2 Air pollution data

Due to the data available in pilot cities as well as the results of the traffic simulations we are not able to map the air quality index. Instead of air quality index, the data available includes levels of specific pollutants at existing measuring stations and the simulated levels of the same pollutants.

Therefore we show the available data: measurements at existing air quality monitoring stations are shown as a sparse heat map layer showing the levels of selected pollutant while the simulated pollutant emissions are shown as a layer over each road segment that shows the selected pollutant level.

4 IMPLEMENTATION

The visualization were implemented as part of the URBANITE user interface (UUI). We focus on the UUI modules used to analyse the simulation results and the comparison of two simulations.

The UUI is implemented using the Angular framework [4] mostly in TypeScript with some JavaScript parts. For ease of integration and to be able to package the UUI module as an Angular module we opted to use JavaScript libraries Leaflet.js [2] and echarts [5]. Leaflet.js is used for all geo-spatial and geo-temporal visualizations. It provides an interactive map and the functionality to add custom layers to the map. We use the library Echarts to implement any line charts and spider charts.

4.1 Map based visualizations

Multiple visualizations were developed to visualize certain geo-spatial data:

- Visualization of the traffic flows is shown on Figure 1. Each street is overlaid with a line, colored according to the traffic flow intensity. Less intensive flows are shown with blue hue and more intensive flows are shown in red. The color scale consists of a five color ramp selected for best visibility on the base map.
- Visualization of emissions of specific pollutants. Each street is overlaid with a line, colored according to the amount of selected pollutant. Streets with less emitted pollutants are shown in blue and streets with more are shown in red.

These visualizations are implemented using JavaScript and based on maps provided by the library Leaflet.js[2]. The overlays are generated from the simulation results by aggregating road network links by street name and summing the selected attribute for the day or per hour, thus enabling animated visualization of changes throughout the day or a static daily attribute visualization.

4.2 Color maps

We use one color map for all the visualizations that are based on the city map. The color map selected must be diverging in order to highlight best and worst values according to their desirability. The chosen map a diverging color map using red colors for undesired values and blue for desired values. The color map should also be appropriate for color blind users to avoid potential misunderstandings. With the requirements of the color map defined, we selected a color map named cold-warm [6] that fits our needs. The color map is shown on Figure 2 and is a a diverging color map that is colorblind safe. We have opted to use a five step color map instead of the full gradient to make the extreme values stand out more.

Figure 2: Color map used for overlays. Blue color is used for desired values and red color is used for undesired values.

4.3 Interactive charts

We use interactive charts implemented using the echarts library to implement line charts, histograms and spider charts. Line charts are used to analyse the modal splits on specific streets and the level of selected emitted pollutant (CO_2) using an overlaid area chart. These were made interactive to allow the user to zoom in and move the viewport around. Hovering over any of the lines shows the number of trips of the appropriate modes as well as the mode the line represents.

The same visualizations were also implemented as 3D line charts as shown on Figure 3. Lines are replaced with strips and instead of adding an area map to show the emitted pollutant levels they are color coded using the strip color. These allow more interactions such as panning and rotation.

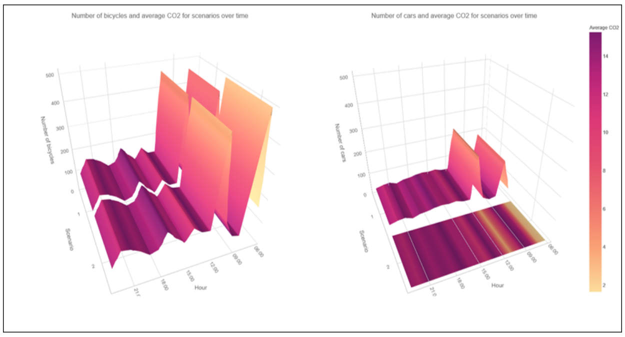

Figure 3: Visualization of the modal split between car and bicycle trips and the amount of the CO_2 emitted on the selected street The CO_2 levels are color coded.

A spider chart was implemented using the echarts library for multi-attribute comparison of different simulation results. This allows the user to recognize the dominant solution at a glance based on the size of the area that represents the simulations. On the other hand they allow us to show the detailed values of multiple, potentially competing attributes. Thus the user is able to understand the data at a glance on some level while also providing the details when the user hovers the mouse over axes of the spider chart or the line representing a single simulation result. An example of the spider chart can be seen on Figure 4.

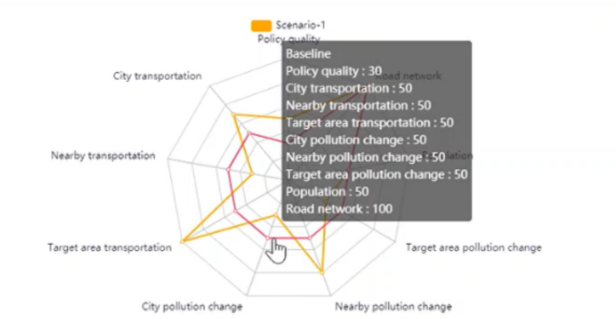

Figure 4: The spider chart shows the comparison of two different proposed mobility policies based on the simulation results. We can see the popup that shows values of all attributes of one of the simulations.

5 FUTURE WORK

There is a lot of room for improvement and the work is ongoing. Several visualization techniques covered are not yet finished, such as the geo-temporal visualizations and the heat maps. The next step is to finish the implementations of all the selected visualization types and to improve the visual appeal of the visualizations.

In order to compare the air quality measured with the results of simulations which provide estimations of the levels of emitted specific pollutants the data must first be transformed to an estimation of the air quality index. An alternative to this approach should it prove infeasible is to use the simulation results to estimate the measurements at the location of the measuring stations.

6 CONCLUSIONS

1 we overviewed the common methods of visualization of common traffic data

We have overviewed the mobility related open data-sets available in four major European cities and identified the most important for dealing with urban mobility policy. Several different sorts of data were analysed and appropriate visualizations were selected. Some of the visualizations are implemented, specifically traffic count, daily trips, and emitted air pollutant visualizations, among with some of visualizations of policy comparison.

The module implementing the visualizations supports the needs of the urban mobility analysis tool-set that we are developing. The visualization selection and implementation fits the needs of different users and will be further improved as we gather feedback from the pilots.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870338. The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P2-0209).

REFERENCES

[1] Gijs Alberts. 2021. Standing e-scooters, what to expect: micro-mobility with micro effects?: explorative research into the expected effects and policy implications of the introduction of e-scooters in the dutch traffic system.
[2] Paul Crickard III. 2014. *Leaflet. js essentials*. Packt Publishing Ltd.
[3] Mohamed El-Agroudy. 2020. Mobility-as-a-service: assessing performance and sustainability effects of an integrated multi-modal simulated transportation network.
[4] Nilesh Jain, Ashok Bhansali, and Deepak Mehta. 2014. Angularjs: a modern mvc framework in javascript. *Journal of Global Research in Computer Science*, 5, 12, 17–23.
[5] Deqing Li, Honghui Mei, Yi Shen, Shuang Su, Wenli Zhang, Junting Wang, Ming Zu, and Wei Chen. 2018. Echarts: a declarative framework for rapid construction of web-based visualization. *Visual Informatics*, 2, 2, 136–146.
[6] Kenneth Moreland. [n. d.] Diverging color maps for scientific visualization (expanded). *Proceedings in ISVC*, 9, 1–20.
[7] The URBANITE Project. 2021. https://urbanite-project.eu/.
[8] Martin Treiber and Arne Kesting. 2013. Traffic flow breakdown and traffic-state recognition. In *Traffic Flow Dynamics*. Springer, 355–366.

URBANITE Ecosystem: Integration and DevOps

María José López[†]
ICT Division
TECNALIA, Basque Research and Technology Alliance (BRTA)
Spain
mariajose.lopez@tecnalia.com

Iñaki Etxaniz
ICT Division
TECNALIA, Basque Research and Technology Alliance (BRTA)
Spain
inaki.etxaniz@tecnalia.com

Giuseppe Ciulla
Research & Development Laboratory
Engineering Ingegneria Informatica
Palermo, Italy
giuseppe.ciulla@eng.it

ABSTRACT

URBANITE is a collaborative research and innovation project whose outcomes are mainly software based. These outcomes will be implemented in a collaborative manner by different development teams from different partners. In order to manage the development environments, and the integration of the different software components in on time releases, the proper DevOps strategy and processes has been defined and set up.

This paper describes the URBANITE integrated architecture at month 12, with a theoretical vision of the URBANITE system that will cover all the functional and non-functional initial requirements set by the technical work-packages considering the social perspective and the input of the use cases.

The definition of the interactions among components is shown through the specification of the interfaces, considering the dataflows envisioned for meeting the needs of the different stakeholders. Different tools, environments and strategies envisioned for the management of the development, integration and validation stages of the software components to be implemented during the life cycle of the project are described as part of the integration strategy.

KEYWORDS

DevOps, Integration, Ecosystem, Requirements, Architecture, Prototype.

1 REQUIREMENTS

The process for setting up the URBANITE Ecosystem receives inputs from the rest of technical components, related to the data management and simulation processes, regarding to:

- Technical (software) requirements, expressing both functionality needs and non-functionality aspects.
- Architectural structure and configuration of the components implemented in different work packages.
- And about how to integrate them into the overall URBANITE UI.

Information Society 2021: 24th international multiconference, *4–8 October 2021, Ljubljana, Slovenia*

Also the Use Cases propose functionalities for the URBANITE ecosystem, that will be considered for being part of the URBANITE ecosystem and prioritize for their implementation.

This platform will comprise the URBANITE components implemented, Key Results or KR from now on, (Key Result KR1-Virtual Social Policy Lab, KR3-Data Management Platform, KR4- Algorithms) and their integration in the URBANITE ecosystem (KR5).

The elicitation of the first version of the functional and non-functional requirements for the URBANITE ecosystem and the related components is described as an iterative process where both the technology providers and use case providers' have participated.

For the functional requirements a combined approach has been followed: 1) a top down approach led by the technology provider partners, who have defined the first set of functional requirements and 2) a bottom up approach where the needs of the Use Cases have been monitored and UC initial requirements have been extracted. For the Non-Functional Requirements, these have been detailed per component, including relevant aspects, such as performance, usability or resources needs for deployment.

All these requirements will serve for the continuous development and improvement of the URBANITE ecosystem, through the different releases, validation processes and reviews of the requirements.

The URBANITE ecosystem will include all the components for data management, analysis and support to the decision making that are going to be created/developed/implemented in the context of the URBANITE project. The first version of the requirements will be updated in further reports and analysis.

Several sources will be used to elucidate the requirements for the URBANITE ecosystem:

- **Requirements coming from the URBANITE action specification**: These requirements cover the functionalities described in the URBANITE action specification. The first version of these requirements has been described by the Technology providers partners (Fraunhofer, Tecnalia, JSI and Engineering Ingegneria Informatica) based on the URBANITE approach and high-level architecture description included in the URBANITE action.

- **Requirements coming from the Use Cases**: The Use Cases proposed functionalities for the URBANITE ecosystem, so that the features offered can cover their needs.
- **Requirements coming from the co-creation sessions (SoPoLabs)**: It is expected that some requirements may be derived from the SoPolabs that will be conducted in the context of WP2. If relevant these requirements will be considered for the URBANITE ecosystem and prioritize for their implementation.

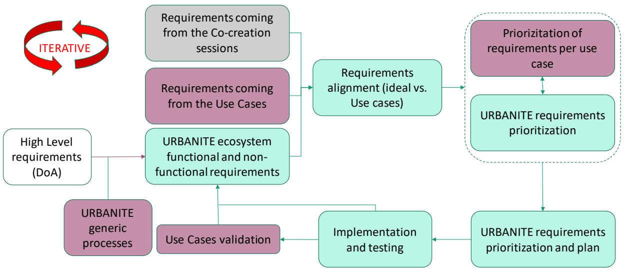

Figure 1: URBANITE process for requirements gathering and prioritization

The different users of the URBANITE platform to perform any of the previously described are:

- PA (Public Administration): This actor is the user from the public administration, usually the civil servants.
- Citizens: This actor is the citizen that is using any of the components in the platform.
- Platform administrator: This is the administrator of the platform who can install components, check the status of the included components, etc.

2 ARCHITECTURE

The detailed description of the entire global architecture of the URBANITE ecosystem as a general representation of it, is in its first version and can evolve following the needs of the project. Structural and behavioural analysis of each component of the architecture was performed identifying interactions and dependencies among them.

Three layers of components can be observed and identified by colours:

- Yellow components are those that manage the data, and implemented withing the WP3
- The purple ones are dedicated to the simulation and analysis of the data ingested to the system for the yellow ones.
- And the grey components are those related to the UI, as the entry point to the platform and for user management.
- There is also a green component considered as a repository of the datasets stored by the data management layer.

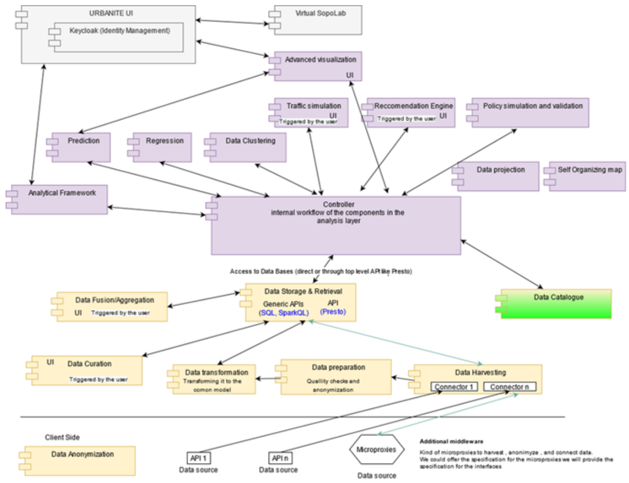

Figure 2: First version of URBANITE architecture

In the Month 15 of the project, the architecture is a reduced version of the overall URBANITE architecture and consists of the components that provide the functionalities designed covering the current version of the requirements.
The components are:

- URBANITE UI: The entry point to the URBANITE Ecosystem, that allows users to access the functionalities provided by the URBANITE platform at this point of the project.
- Identity Manager (Key Cloak): This component is in charge of securing the access to the other URBANITE's component, whenever security is needed. It is called by other components that interact with the user.
- City Bike Pattern Analysis: This module analyses GPS information related to the mobility of the bikes and transform it into more useful data.
- Traffic Prediction: It performs heuristic prediction for the vehicle flow at a location within the city by the processing of historical values measured by a fixed sensor and other information.
- Traffic Simulation: It offers the simulations of traffic under specified conditions, as proposed mobility policies, different weather conditions, changes to the traffic infrastructure, etc.
- Scheduler: It triggers a pipeline for the harvesting process, downloading data from a list of configured APIs within defined periods of time.
- Data Harvester and Transformation: It is responsible for fetching data from a given API, being the entry point of the data into the pipeline. Then a transformation is done into common models.
- Data Storage and Retrieval: This module stores and retrieves datasets metadata and related data in repositories DCAT-AP compliant metadata and transformed data.
- Data Catalogue: It allows to discover and access the datasets collected and managed by the components of URBANITE Ecosystem for data acquisition, aggregation, and storage.

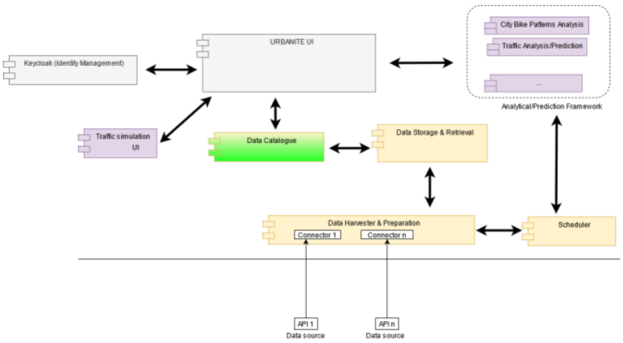

Figure 3: Month 15 version of URBANITE architecture

3 INTEGRATION AND DEVOPS STRATEGY

This section presents the infrastructure and tools planned to be used internally for the development and operation. The DevOps approach requires the set-up of a development and delivery pipeline, that consists in the stages an application goes through from development through production.

The URBANITE iterative and incremental approach mandates the adoption of a development and deployment process able to fully support it. That is why the project will adopt a DevOps approach in the development of all KRs. DevOps integrates development and operations into a single-minded entity with common goals: high-quality software, faster releases and improved users' satisfaction. DevOps also incorporates a number of agile principles, methods, and practices such as continuous delivery, continuous integration, and collaboration [1].

The different KRs, which are the outcomes of URBANITE [2], are composed of several software components that will be implemented by different partners following different technologies.

In URBANITE, the DevOps approach will be structured in three environments as depicted in the figure.

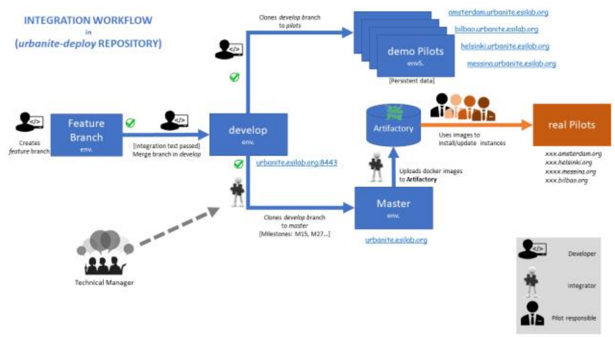

Figure 4: Continuous Integration and DevOps approach

The description of the environments that are part of the integration system:

FEATURE BRANCH: Temporary environment that is created each time a developer wants to integrate a new version of his component. It just checks that the new version of the urbanite platform builds without problems and is destroyed afterwards.

DEVELOP: Environment that contains the last version of the components running together. Dedicated to test new features, interfaces, and communications among components.

MASTER: Contains a specific version of the platform, frozen for specific Milestones.

DEMO PILOTS: Four environments, one for each city, where the integrated platform is replicated and adjusted to the characteristics of the use cases. It is a previous step for testing the platform before setting up in the infrastructure of the municipalities.

REAL PILOTS: the installation of the platform in each municipality's infrastructure. To be done after the integration phase once a stable version is achieved to test the use cases.

Apart from that, in order to support developers during the integration, we provide:

A Portainer instance that allows to access the logs and the console of every container in every environment.

An Artifactory instance to store the images of the containerized components. These images will be used to deploy the final version of the platform in the real Pilots.

4 URBANITE ECOSYSTEM

The main result of the URBANITE project is the URBANITE Ecosystem and aggregates all aspects of the project, namely the citizen participation, both social (citizen participation, attitude and trust in disruptive technologies, co-creation) and technical aspects (data management platform, algorithms and so on).

The URBANITE UI is an integration framework at User Interface level.

The integration strategy provides different approaches that can be followed:

1. **External component integration**
   - **Iframe:** the external application is included in the UI through an iframe
   - **External link:** the application is referenced in a dedicated section of the UI, and a specific link is provided to the user
2. **Template component integration:** the external application, that must provide a set of REST APIs for developing a specific component included in the UI.

The URBANITE UI is an Angular application built taking advantage of Nebular, ngx-admin frameworks and Eva Design System. With the addition of some of the most popular front-end libraries and packages.

The access to the Urbanite UI is provided through the Urbanite's Identity Manager component (an instance of KeyCloak whose theme is customized following Urbanite's colour palette).

The Urbanite UI provides Role-Based access to specific functionalities following the IDM returned role(s) of the user

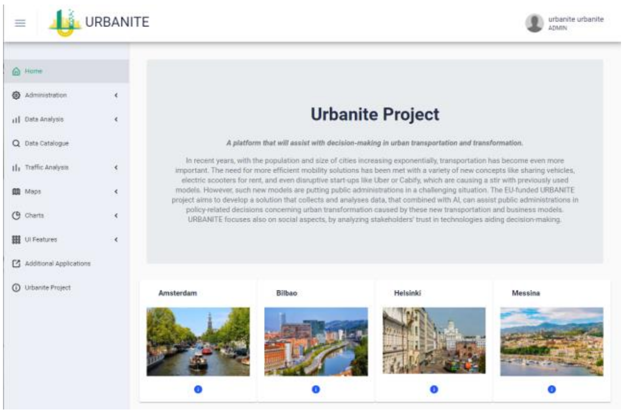

Figure 5: M15 Component's integration REST APIs based

The figure above shows is the URBANITE UI, integrating the components implemented for the M15 version of the prototype, covering the requirements and the functionalities provided by the implementation planned at this point.

The left part of the page includes the available options. Some of them are general utilities, and other are functionalities implemented within the different technical work packages.

- The Home page offers descriptions of the four municipalities and the basis of the URBANITE project. There is an additional information section for each description that allows to extend the details of the selected city.
- The Administration, Data Analysis, Data Catalogue and Traffic Analysis are specific sections that provide services related to the data of the different municipalities.
- Maps: where are two examples of how a developer can build and manage maps, using the libraries provided by this UI.
- Charts: this option displays three possible library alternatives provided by the Urbanite UI to build bars, pies and line charts.
- UI Features: about style examples as colors, icons, typography, and the grid system that should be used for implementing pages to provide responsiveness.
- Additional Applications: is a section where external links to other applications can be added through the URBANITE UI configuration file. For instance, the forum page is linked.
- And the Urbanite Project where included information about the objectives of the project.

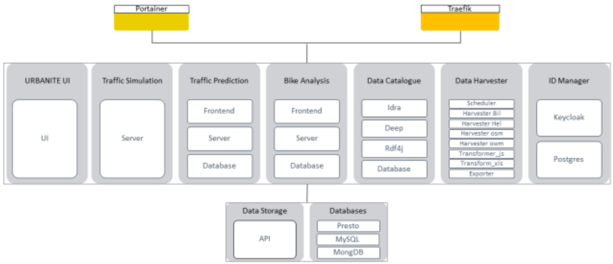

Figure 6: URBANIITE Ecosystem v1

The Figure 6 describes the schema that supports the before explained prototype.

REFERENCES

[1] New Relic, "“Navigating DevOps - What it is and why it matters to you and your business”," New Relic, 2014.

[2] URBANITE Consortium, "URBANITE Annex 1 - Description of Action," 2019.

Indeks avtorjev / Author index

Delavnica URBANITE 2021

URBANITE Workshop 2021

Sergio Campos, Shabnam Farahmand, Nathalie van Loon, Erik Dovgan