

Emilio Insfran · Fernando González ·
Silvia Abrahão · Marta Fernández ·
Chris Barry · Michael Lang · Henry Linger ·
Christoph Schneider *Editors*

Advances in Information Systems Development

Crossing Boundaries Between
Development and Operations in
Information Systems

Lecture Notes in Information Systems and Organisation

Volume 55

Series Editors

Paolo Spagnoletti, Rome, Italy

Marco De Marco, Rome, Italy

Nancy Pouloudi , Athens, Greece

Dov Te'eni, Tel Aviv, Israel

Jan vom Brocke, Vaduz, Liechtenstein

Robert Winter, St. Gallen, Switzerland

Richard Baskerville, Atlanta, USA

Lecture Notes in Information Systems and Organization—LNISO—is a series of scientific books that explore the current scenario of information systems, in particular IS and organization. The focus on the relationship between IT, IS and organization is the common thread of this collection, which aspires to provide scholars across the world with a point of reference and comparison in the study and research of information systems and organization. LNISO is the publication forum for the community of scholars investigating behavioral and design aspects of IS and organization. The series offers an integrated publication platform for high-quality conferences, symposia and workshops in this field. Materials are published upon a strictly controlled double blind peer review evaluation made by selected reviewers.

LNISO is abstracted/indexed in Scopus

More information about this series at <https://link.springer.com/bookseries/11237>

Emilio Insfran · Fernando González ·
Silvia Abrahão · Marta Fernández · Chris Barry ·
Michael Lang · Henry Linger · Christoph Schneider
Editors

Advances in Information Systems Development

Crossing Boundaries Between Development
and Operations in Information Systems

 Springer

Editors

Emilio Insfran 
Department of Computer Systems
and Computation
Universitat Politècnica de València
Valencia, Spain

Silvia Abrahão 
Department of Computer Systems
and Computation
Universitat Politècnica de València
Valencia, Spain

Chris Barry 
Cairnes School of Business & Economics
National University of Ireland Galway
Galway, Ireland

Henry Linger 
Faculty of Information Technology
Monash University
Melbourne, VIC, Australia

Fernando González 
Department of Business and Economics
Universitat Politècnica de València
Valencia, Spain

Marta Fernández 
Department of Business and Economics
Universitat Politècnica de València
Valencia, Spain

Michael Lang 
Cairnes School of Business & Economics
National University of Ireland Galway
Galway, Ireland

Christoph Schneider 
IESE Business School
University of Navarra
Barcelona, Spain

ISSN 2195-4968

ISSN 2195-4976 (electronic)

Lecture Notes in Information Systems and Organisation

ISBN 978-3-030-95353-9

ISBN 978-3-030-95354-6 (eBook)

<https://doi.org/10.1007/978-3-030-95354-6>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Conference Organization

Conference Chairs

Emilio Insfran, IUMTI—DSIC, Universitat Politècnica de València, Spain
Fernando González, DOE, Universitat Politècnica de València, Spain

Program Chairs

Silvia Abrahão, IUMTI—DSIC, Universitat Politècnica de València, Spain
Marta Fernández, DOE, Universitat Politècnica de València, Spain

Virtualization Chair

Abel Gómez, Universitat Oberta de Catalunya, Spain

International Steering Committee

Chris Barry, National University of Ireland Galway, Ireland
Michael Lang, National University of Ireland Galway, Ireland
Henry Linger, Monash University, Australia
Christoph Schneider, IESE Business School, Spain

Track Chairs

Managing IS Development and Operations

Ana Moreira, Universidade Nova de Lisboa, Portugal
João Araújo, Universidade Nova de Lisboa, Portugal

IS Methodologies and Education

Alfred Zimmermann, Reutlingen University, Germany
Giuseppe Scanniello, University of Basilicata, Italy

Security, Privacy and Trust in IS

Alberto Rodrigues da Silva, University of Lisbon, Portugal
Daniela Cruzes, Norwegian University of Science and Technology, Norway

Incorporating Human-Computer Interaction into IS

Jean Vanderdonckt, Université Catholique de Louvain, Belgium
Mikko Rajanen, University of Oulu, Finland

Green and Sustainable IS

Coral Calero, Universidad Castilla-La Mancha, Spain
Dorina Rajanen, University of Oulu, Finland

Current Topics in IS Development

Malgorzata Pankowska, University of Economics Katowice, Poland
Vaclav Repa, University of Economics in Prague, Czech Republic

Program Committee Members

Track: Managing IS Development and Operations

Dominique Blouin
Markus Borg
Dominik Bork
Nelly Condori Fernández
Maya Daneva
Jose Luis de la Vara
Ilias Gerostathopoulos
Abel Gómez
Jane Hayes
Eric Knauss
Seok-Won Lee
Grischa Liebel
Mariana Maia Peixoto
Beatriz Marin
Miguel Mira-da-Silva
Barbara Paech
Claudia Pons
Paulo Rupino-Da-Cunha
Mehrdad Sabetzadeh
Kurt Schneider
Christopher Vendome
Colin Venters

Track: IS Methodologies & Education

João Barata
Danilo Caivano
Oscar Diaz
Jānis Grabis
Carmine Gravino
Björn Johansson
Birger Lantow
Yoshimasa Masuda
Elena Navarro
Ovidiu Noran
Malgorzata Pankowska
Iris Reinhartz-Berger

Kurt Sandkuhl
Julio Sandobalín
Marco Santorum
Rainer Schmidt
Monique Snoeck
Karen Stendal
Janis Stirna
Ann Svensson
Torben Tambo

Track: Security, Privacy & Trust in IS

Chris Barry
Karin Bernsmed
Achim Brucker
Danilo Caivano
Ana Cavalli
Miguel Correia
Christophe Feltus
Martin Gilje Jaatun
Tim Majchrzak
Andreas Poller
Sven Türpe

Track: Incorporating Human-Computer Interaction into IS

José Creissac Campos
Netta Iivari
Marianne Kinnula
Yann Laurillau
Sophie Lepreux
Pedro J. Molina
Norbert Pataki
Jorge Luis Pérez Medina
Costin Pribeanu
Philippe Renevier
Ahmed Seffah
Marco Winckler

Track: Green and Sustainable IS

Elina Annanperä
Ruzanna Chitchyan
Felix Garcia
Patricia Lago
David Langley
Irene Lizeth Manotas Gutiérrez
M^a Ángeles Moraga
Stefan Naumann

Track: Current Topics in ISD

Cheuk-Hang Au
Peter Bernus
Tomáš Bruckner
Alena Buchalcevova
Justin Filippou
Ahmad Ghazawneh
Mairéad Hogan
Henry Linger
Nearchos Paspallis
Dijana Plantak Vukovac
Marios Raspopoulos
Karel Richta
Artur Strzelecki
Oleg Svatoš

Additional Reviewers

Vita Barletta
Alejandro Catala
Shalini Chakraborty
Hyo Cheol Lee
Sangeeta Dey
Umberto Fiaccadori
Vinh Hoa La
Michael Lang
Rakshit Mittal
Lorraine Morgan

Huu Nghia Nguyen
Sihn Hye Park
Valentina Kirinic
Dijana Oreški
Joanna Palonka
Jaroslav Pokorný
Victoria Paulsson
Martin Potančok
Mariia Rizun
Paulina Rutecka
Irish Singh
Anna Soltysik-Piorunkiewicz
Zlatko Stapić
Filip Vencovsky
Mariusz Żytniewski

Preface

The **International Conference on Information Systems Development (ISD)** is an academic conference where researchers and practitioners share their knowledge and expertise in the field of information systems development. As an Affiliated Conference of the Association for Information Systems (AIS), the ISD conference complements the international network of general IS conferences (ICIS, ECIS, AMCIS, PACIS, HICSS). The ISD conference continues the tradition started with the first Polish-Scandinavian Seminar on Current Trends in Information Systems Development Methodologies, held in Gdansk, Poland in 1988. This seminar has evolved into the International Conference on Information Systems Development.

Throughout its history, the conference has focused on different aspects, ranging from methodological, infrastructural, and educational challenges in the ISD field to bridging the gaps between industry, academia, and society. Advancements in information systems foster technological developments. The deployment of the resulting technologies in all areas of society, including in the public and private sector, the community, and in the home is greatly beneficial. ISD has always promoted a close interaction between theory and practice, setting a human-centered agenda focused on advancing the methods, tools and management of IS.

This volume is a selection of papers from ISD2021, the 29th Information Systems Development conference hosted by Universitat Politècnica de València, and held in Valencia, Spain on September 8–10, 2021.

All accepted papers have been published in the AIS eLibrary, which is accessible at <https://aisel.aisnet.org/isd2014/proceedings2021>. This volume contains extended versions of the best papers, as selected by the ISD2021 Proceedings Editors.

The theme of the conference was *Crossing Boundaries between Development and Operations (DevOps) in Information Systems*. It focused on the latest developments in ISD and particularly on emerging concepts, novel approaches and ideas that are likely to shape information systems research and practice in the next years.

The ISD2021 conference attracted contributions in the general area of information systems development, management, and education, as well as in more specialized topics including Security, Privacy and Trust in IS, Green and Sustainable IS and Incorporating Human-Computer Interaction into IS. ISD2021 focused on these and

associated topics in order to promote research into theoretical and methodological issues and ways in which these advances enable better synergies between theory and practice.

We believe that the innovative papers assembled in these lecture notes will inform the reader of important contributions in this regard.

Valencia, Spain
Valencia, Spain
Valencia, Spain
Valencia, Spain
Melbourne, Australia
Galway, Ireland
Galway, Ireland
Barcelona, Spain

Emilio Insfran
Fernando González
Silvia Abrahão
Marta Fernández
Henry Linger
Chris Barry
Michael Lang
Christoph Schneider

Contents

A Data-Driven Optimization Computational Tool Design for Bike-Sharing Station Distribution in Small to Medium-Sized Cities: A Case Study for Cuenca, Ecuador	1
Fernando Cueva, Pengcheng Shi, and Priscila Cedillo	
A Method to Build Serious Games for Older Adults ADE2: Evaluation of User Perceptions and Threats of Validity	19
Priscila Cedillo, Cristina Sánchez-Zhunio, Emily Arteaga, Jefferson Arias, and Marcos Orellana	
Accessibility of Digital Content in Europe: Users’ Perspective	37
Inés López Baldominos, Vera Pospelova, Luis Fernández Sanz, and Ana Castillo-Martínez	
Authentic Learning for Groupwork Skills: A Comparison Between a Physical and a Virtual Experience	53
Mairéad Hogan	
Business-IT Alignment Anti-patterns	67
Jean-Philippe Gouigoux and Dalila Tamzalit	
Fuzzy Personalization of Mobile Apps: A Case Study from mHealth Domain	91
Fabio Sartori and Lidia Lucrezia Tonelli	
Government IT Projects: Current Evidence of Cost and Schedule Overrun and Their Antecedents	109
Adam Alami, Christian Østergaard Madsen, and Oliver Krancher	
Heat Maps from Decision Flaps—What We See and What We Think ...	127
Chris Barry, Mairéad Hogan, and Michael Lang	
Modeling Boundary-Spanning Business Processes in Industry 4.0: Incorporating Risk-Based Design	143
Vítor Ribeiro, João Barata, and Paulo Rupino da Cunha	

Paving the Way for Smart Community Composting 163
José Monteiro, João Barata, Hugo Gomes, and Emanuel Castro

**Reducing Cognitive Load in Learning to Model UML Sequence
Diagrams** 179
Sohail Alhazmi, Charles Thevathayan, and Margaret Hamilton

**Studying the Influence of Empathy Maps on Brainstorming
for Requirements Elicitation: A Quasi-Experiment** 199
Ezequiel Kahan, Emilio Insfran, Marcela Genero, and Alejandro Oliveros

**Supporting a Bottom-Up Evolution of Microservice Compositions
Based on the Choreography of BPMN Fragments** 219
Jesús Ortiz, Victoria Torres, and Pedro Valderas

Team Performance in Large-Scale Agile Software Development 237
Tomas Gustavsson

Towards Conceptual and Logical Modelling of NoSQL Databases 255
Jaroslav Pokorný and Karel Richta

**Towards the Development of a Contingent Use of Systems
Development Methodologies Model** 273
Benson Moyo, Magda Huisman, and Lynette Drevin

A Data-Driven Optimization Computational Tool Design for Bike-Sharing Station Distribution in Small to Medium-Sized Cities: A Case Study for Cuenca, Ecuador



Fernando Cueva, Pengcheng Shi, and Priscila Cedillo

Abstract Faced with heavy vehicular traffic at present, the strategic implementation of Bike Sharing Systems (BSSs) in cities as an alternative means of transport for users is increasingly being adopted. These solutions reduce the environmental burden posed by other means of transportation, decrease costs for citizens, improves people's health due to physical activity, among other advantages. However, aspects such as the definition of bike stations' locations represent a challenge when these solutions are being implemented. Therefore, this paper presents a software tool design that supports a method that defines the location and number of stations within a BSS. Also, the tool uses a data-driven optimization model to establish the location of stations. Finally, a case study carried out in Cuenca—Ecuador, demonstrates the proposal's feasibility, showing a significant concordance with the consulting firms-consortia results (70–90% of coincidence) at a lower cost.

Keywords Bike sharing system · Public bike system · Optimization · Pollutants · Data-driven

A prior version of this paper has been published in the ISD2021 Proceedings (<http://aisel.aisnet.org/isd2014/proceedings2021>).

F. Cueva (✉) · P. Shi
Rochester Institute of Technology, Rochester, USA
e-mail: wfc9474@rit.edu

P. Shi
e-mail: pengcheng.shi@rit.edu

F. Cueva · P. Cedillo
Universidad de Cuenca, Cuenca, Ecuador
e-mail: priscila.cedillo@ucuenca.edu.ec

1 Introduction

Population growth and city urbanization lead to a greater demand for different means of transport [1]. It is increasing the number of vehicles [2]. It drives congestion with direct implications on travel times and air pollution, increasing mortality and morbidity [3].

There are several alternatives to reduce the problem of traffic congestion. Among them, public transport, vehicle access limitation by ordinance [4], shared vehicle [2], alternatives to the traditional vehicle (small ones with electric batteries) [5], motorcycles, and the implementation of a Bike Sharing System (BSS) [6]. However, the first one is limited by not having access to all city points, while the second one presents logistical and political problems [7]. The shared vehicle alternative reduces traffic congestion and provides users lower costs than private vehicles [8]; however, it is not environmentally friendly. The fourth alternative is not always accessible to all users due to the cost [5]. In the case of motorcycles, it allows users' connection with their places of destination and takes up little space, but pollution is more significant than bikes.

A BSS is a public transport alternative that allows users to take a bike at one point and return it at another, either in exchange for a fee or free. These systems enable traveling through a network of stations with multimodal connections that can be point-to-point or round-trip in an urban environment. BSS is an alternative to motorized transport that reduces travel time, generating economic savings - compared to other transport systems [9, 10], energy savings, optimizing the use of space on roads, parking, and mobilization. In addition, it can be integrated with the mass transport system, friendly to the environment, and promotes livable and healthier cities [10, 11].

However, the BSS implementation's success depends on the correct location and number of stations in the urban environment that satisfy the demand and guarantee an adequate service throughout the city. Therefore, this study aims to propose a tool that supports a method that estimates the number of stations and their optimal location. The presented informatic tool design provides an affordable solution to help designers during the stations' distribution establishment without using too many resources (e.g., economic, human). Medium and small cities could start their BSSs planning from a standardized platform supported by a data-driven software tool to define stations. This tool will consider aspects related to historical information and other public means of transport. Therefore, this study can lead to the necessary support to diminish the environmental impact produced by pollutants from cars, motorcycles, and traditional means of transportation using BSS systems with reachable stations that serve citizens around cities.

Although several companies provide consulting services for locating the optimal places for implementing the BSS's stations, those solutions are expensive. Besides,

most of the time, small and medium cities could not afford their prices; being necessary to have software that uses available information sourced by other public organizations (e.g., population and demographic distribution, public transportation companies, weather historical data sources) allow making decisions automatically. On the other hand, the reviewed literature indicates that data available for each city and its characteristics have defined the stations' location differently and often without a scientifically defined, replicable, and repeatable method [12, 13].

During the main method that is supported for the system of this contribution, four stages were considered, starting with (i) defining the factors for a correct distribution of the stations; (ii) preparing and preprocessing data; (iii) determining the suitability of the stations. Finally, the operational research model defines the locations, and the number of stations considering the previous models' algorithms and outputs to design a BSS from strategic planning and its respective evaluation case study.

Consequently, having a data-driven based software tool helps in a scientific and repeatable way to optimize the station distribution to design a BSS.

The solution has been used when the main method was applied to a specific case study in the city of Cuenca-Ecuador to measure precision when comparing the method's results versus the stations' empirical location [14]. Cuenca became an interesting case study because it is a city with traffic problems, declared a World Heritage Site. Furthermore, it is a medium-size city due to its number of inhabitants (i.e., half a million approximately) and its size (72 km²) [15].

The study is organized into sections. Section 2 presents the background and related work that identifies the factors considered in implementing a BSS in other cities. Section 3 describes the main contribution of this research. Sections 4 and 5 present the main results, conclusions, and further work.

2 Background and Related Work

According to international guidelines [8, 16], the factors that influence and must be considered for an optimal distribution of the stations are (i) demographic and socio-economic factors such as population and job density; (ii) geographic and topographic factors; (iii) climate; (iv) places of cultural and recreational attraction, public spaces and parks; and (v) infrastructure and mass transportation. Therefore, these factors should be considered general and a starting point for the design of BSS locations.

Additionally, the specific needs, objectives, and data sources must be considered, among other specific case study aspects. In this context, BSS experiences in other countries are analyzed to define the factors that could be used in determining the location of the stations [7, 14, 17, 18].

The literature available on BSS is varied. Studies related to BSS design have focused on stations, prices, bikes, balance, and service stations. A key component

to the success of the BSS design is the location of the bike stations. However, from the perspective of strategic planning, very few research studies have been conducted in this regard [19], so this component is the research's starting point. Also, as far as we know, there are a lack of software tools that support methods and allow making decisions related to the location of BSS stations.

The distribution of stations must guarantee the optimum level of service for users when they require a bike. It implies covering the total demand of the area with an optimal number of stations that ensure the service's quality [12–15], allowing to minimize the costs derived from the distance traveled and considering a penalty for unmet demand [20]. Several studies have used optimization methods for this purpose [21].

Appendix summarizes the objective, methods, and factors used by a set of studies focused on bike inventory management, station capacity, and location. Despite the contributions in using models and inventory management, works such as [20, 22] do not validate the proposals with a case study. On the other hand, other studies do not consider other factors or ignore the integration with the transport system to define the stations' location, in the case of Garcia-Gutierrez et al. [16] and Romero et al. [21], respectively. In Guo et al. [23], the approach used was robust to improve the optimization model. As it can be concluded, there are not highly flexible solutions that employing a software tool can join different data sources to establish the location of the BSS stations considering real information originated in the city in which the bike system will operate.

3 Data-Driven Strategies to Determine BSS Station Distribution

This section explains the model, factors analysis, data sources, data processing, and the proposed software solution.

3.1 Models

Previous steps allow us to establish a method which is shown in Fig. 1. Also, the method needs to be supported by a tool. Both are useful for making decisions in the station distribution's strategic planning based on different data sources. As shown in Fig. 1, four main activities were stated as part of the main method. In order to be harmonic with the main method, the software tool design presented in this paper supports each of the activities, integrating them in a solution that provides as output an optimal model with the final BSS station's locations.

This approach shows the multi data-driven philosophy with the different steps carried out to deal with the stations' distribution based on the information analysis.

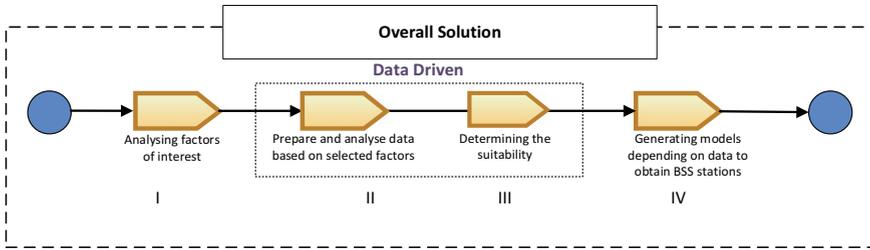


Fig. 1 Overall solution with data-driven philosophy

The different data used in the study (see Fig. 1-I) are summarized in Table 2. After handling and processing (see Fig. 1-II), these data allow obtaining the suitability areas, the cost of the pollutants (i.e., noise and NO_2) per city block (i.e., an area of the city surrounded by streets), distance matrix, and demand matrix. Finally, with the processed data, the data model is built whose objective function is to minimize the stations' costs with their penalties, distributing the stations with a number that covers the city. The final model is achieved using econometric models (i.e., hedonic prices), optimization models (i.e., set covering problem), data analytics, and machine learning algorithms (i.e., clusters) that have been proposed for data management.

This contribution constitutes a good starting point for places without resources, expertise, and a high need for a BSS, saving them time and money.

3.2 Factor Analysis, Data Sources, and Data Processing

This subsection explains the data-driven approach considering the impacting factors analyzed in previous studies, the data sources used in the case study, and the correlation with the general data used in other implementations. The data-driven approach allows preparing, analyzing, and determining the model's different inputs to define the station distribution.

Analyzing the factors

According to the literature review, multiple factors have been classified as low, medium, and high impact related to the level of influence on the objective (to establish the shared bike stations' location) (see Table 1). On the other hand, data on pollutants produced by automobiles, such as noise and Nitrogen Dioxide (NO_2), are used to reduce their impact on the city and its inhabitant's health, proposing an alternative transportation system. These pollutant data are analyzed with the software tool and represent an input towards searching the optimization model.

It is recommended to use at least medium and high impact factors for a good result. The factors were selected according to the application case (Cuenca city), the literature review, and the data availability in this research. High impact factors were selected: (i) demographics, (ii) bike lanes, (iii) tourist attraction, (iv) bus stops, (v)

Table 1 Factors used in the case study

Impact	Factors	Previous studies
Low impact	Stations nearby automobile parking, origin–destination matrix	[6, 17, 18, 24, 25]
Middle impact	Residential locations, recreational parks, university locations, and topography	[11, 26–33]
High impact	Job density, public transit location (bus stops), demographics (e.g., age, education, income, equity), tourist attraction, bike lanes, and land use	[6, 10, 18, 24, 34–39]
New proposal	Pollutants and safety at both the road and city levels	

land use, and (vi) job density. Medium impact factors: (vii) the topography of the city, (viii) residential locations, (ix) parks, and (x) university locations. Low impact factors: (xi) Stations nearby automobile parking.

Additionally, new factors are proposed, such as using pollutants as penalties, traffic patterns, and safety at both road and urban levels. These factors were not present in previous studies. All these factors were used with a data-driven approach. Data were analyzed using methods that eliminate duplications, such as Principal Component Analysis (PCA), a dimensionality-reduction method used to reduce large datasets without losing representativeness. Once they were refined, they served as input for the software that allowed obtaining maps at the city block level for the city's stations' location.

Data sources

The city data are stored in different databases (public and private institutions) (see Table 2). In some cases, the information collected does not directly provide the required information but instead infers the information necessary to plan a BSS from specific data.

According to the studies analyzed in the previous sections, the data used in this research were collected to establish, on the one hand, the demand based on the factors analyzed (e.g., demographic, land use) and, on the other, the optimization of the number of stations and its distribution in the city to meet the user's needs. These data are produced in the daily activity of the city and collected by different organizations.

The data were analyzed and crossed between different sources to obtain the information used by other authors. Then other data that had not been used in previous studies were included; this allowed to obtain a better result of the station's distribution in a BSS. The data-driven approach allows establishing systematic and repeatable steps to establish the distribution of stations in countries where resources are limited.

Data processing and analysis

In this process, the data sources used and the methodologies applied are detailed according to the different factors considered for the study. Therefore, information processing was carried out to debug the data (outliers, exceptions, inconsistencies,

Table 2 Data sources

Source	Information
Commission for Environmental Management (CGA)	Noise pollution
Municipal Mobility, Transit and Transportation Company (EMOV-EP)	Nitrogen dioxide pollution
ANT (National Transit Agency)	Vehicle registration database with segmented information and information on vehicular traffic
EMOV EP	Bus database (Includes: 1. Household survey on the socio-economic conditions of the users, 2. Conditions of the bus service, and 3. Register of counts of getting on and off people)
	Tram Database: Tram stops, bus stops, and public parking lots are considered last-mile points for private vehicles
	Route of existing bike lanes
	Vehicle mobility patterns at certain times in the city
Citizen Security Council (CSC)	Accident and robbery database
A survey, database built within the doctoral thesis	Demand database. Mobility survey applied to people between 16 and 65 years old in the city of Cuenca. Reference year: 2018. Geographic coverage: City of Cuenca. The mobility survey collects sociodemographic and behavioral information on transport to characterize the user who uses a bike as a means of mobility
Internal Revenue Service (SRI)	Database of taxpayers who have economic activity in the municipality of Cuenca
ANT	Database of vehicles that pay for vehicle registration in Cuenca
Public Telecommunications, Water and Sewerage Company (ETAPA)	The database contains the telephone numbers of users throughout the Cuenca canton with a location at the property level (a)
Institute of Statistics and Censuses (INEC)	Population and housing data at the block level for the entire study area
The decentralized autonomous government of the municipality of Local government of Cuenca	The topography of the city in raster format allows obtaining the slope levels of the city
	Database of properties and their characteristics. Map of the city at the level of blocks and land use with the main activities categorized as commercial, educational, and recreational

Note (a) These data are cross-referenced with the SRI information to obtain business or business locations to establish jobs' density

or other problems) and prepare it for the following phases. After debugging, that data were passed down to a block-level and then exported to an array. The array's dimensions are made up of the number of blocks in the city that were 4172 (i.e., rows), 990 bus stops, 27 tram stops, and five public parking lots defined as last-mile points (i.e., columns). In these arrays, the distances between those elements were stored (e.g., distance matrix between blocks and bus stops 4172×990). Another array was also created containing the components resulting from the dimensionality reduction for each city block. This process is critical as it is time-consuming, and the model's accuracy depends on this stage.

Analyzing the data

Factor analysis is carried out in order to reduce variables without loss of relevant information. The extraction method that gave the best results was the Principal Component Analysis (PCA). All the factors explain 55.28% of the variation, and a value of 0.783 was obtained in the Kaiser, Meyer, and Olkin test and Bartlett's specificity, indicating that the analysis is adequate. Intrinsic factors were grouped into three components (i) traffic factors that addressed characteristics such as bus weight, topography, parking, rush hours (traffic patterns), and accidents. (ii) Demographic factors made up of potential users, grouped into clusters and population density. Finally, (iii) component 3, composed of land use and job density. Some factors were grouped into a single layer of information called land use (i.e., university locations, tourist attractions, parks). On the other hand, pollutants were used for the hedonic price model to determine the penalty cost in the cost function.

To provide a solution to high-pollution areas through implementing a BSS, the proposed model includes a monetary penalty for the level of air pollution (NO_2) and noise. For this, the hedonic price model was used (Eq. 1) [40] to establish the value per square meter that people would pay to avoid contamination. Then, this value is multiplied for each block, obtaining the penalty per block. Thus, blocks with more contamination result in a higher value.

$$P_h = f(S_h, N_h, G_h, e_h) \quad (1)$$

Where the market price P_h of a single-family detached house with its structural (S_h) and neighborhood (N_h) characteristics, as well as environmental variables (G_h), and the error term (e_h), are the standard hedonic framework. The variables used in the hedonic price model are:

1. Environmental variables of pollution: levels of nitrogen dioxide (NO_2) and noise pollution at any time of the day or night, during the day and night, measured in decibels.
2. Property features such as age, land area, length of the front, number of floors, number of bathrooms, type of floor, state of construction, the material of columns, beams, mezzanines, wall, deck, floors, doors, windows, plastered and lying.

3. External or neighborhood factors such as traffic accidents, robberies of people and houses, distance to the historic center. Other factors are if it is in front of a river if the property is a building, municipal records (RMO) for business, if the house is located in an industrial area, the services available 250 m around, type of road, and if it is located in the historic center.

The results show that the Cuenca market penalizes the property value by \$21.2 year/property per unit of NO₂ emitted in the area. The penalty values for noise pollution are lower than for NO₂ emissions. The price of the property decreases by \$3.4 on average annually for every decibel. These values make up the pollution cost penalty.

Determining origin/destination and demand arrays

In this phase, distance matrices are established between the main travel generators or origins and attractors or destinations [41], which are necessary to establish the stations' distribution. The distance matrix was calculated through the Geographic Information System (GIS). The matrix's dimensions are made up of the number of blocks in the city that were 4172 (i.e., rows), 990 bus stops, 27 tram stops, and five public parking lots defined as last-mile points (i.e., columns). It was necessary to use the distances of the central points of the blocks with each of the bus stops (4172 × 990 order matrix) and trams (4172 × 27 order matrix), parking lots (4172 × 5 order matrix), and between the blocks (square matrix of order 4172).

Determining the demand

Determining the characteristics of the potential demand for the BSS is essential to infer the leading market segments that can use the bike share system (potential users). Therefore, based on the demand database (see Table 2), a cluster analysis was first performed to find sociodemographic data patterns (e.g., age, occupation, sex, among others) of the potential demand. As a result, two clusters were obtained with adequate and almost reasonable quality (Silhouette coefficient). Cluster 1 includes private employees and people with their businesses between 20 and 55 years old, whose occupation is destined to the work environment; they are mobilized for labor issues. Cluster 2 is made up of students between the ages of 18 and 30.

Determining the suitability

In this phase, it is required to determine the areas where a station's establishment is necessary. Suitability means which blocks have the best characteristics for locating a bike station. The suitability blocks matrix makes up an input within the model. The weights obtained through the principal components' method (Analyze the data section) are taken as input and multiplied by the stations' cost to establish suitability. With this, the most suitable blocks with a lower cost were identified.

Building the model

With the previous stages, it is possible to obtain the optimal number of stations covering the city and meet the objective and the proposed restrictions. The proposed

optimization model aims to maximize BSS coverage between bus stops, tram stops, last-mile points, and the city's main origins and destinations. The model is represented by an objective function, where the decision variable is the number of stations that satisfy the highest possible demand.

Here, the system provides the optimization model based on each activity of the methodology. The used metrics are part of the system and calculate the final value. The city of Cuenca is divided into 4,172 blocks. It has 990 bus stops, 27 tram stops, and 5 points in the last mile (proposed); those stations must cover within an established radius (e.g., 300, 500 m). The written model's formulation is presented using the objective function (Eq. 2) subject to the restrictions (Eqs. 3–7). The problem formulation is represented in Table 3.

Table 3 Subscripts and sets

Subscripts and sets	Input parameters	Decision variable
	$D_{bb}, J_{bg}, K_{bt}, L_{bm}$: the distance between candidate stations and blocks, between bus stops and blocks, between tram stops and blocks, between last-mile points and blocks, respectively	
	P, j, k, n : the maximum distance to cover demand in the block, the bus stops, tram stops, and last-mile points, respectively	
$b \in B$: the set of potential bike stations locations	$d_{ba}, u_{bg}, w_{bt}, r_{bm}$: equals 1 if the area, bus stops, tram stops, the last-mile points, respectively, can be covered by candidate station b in p, j, k, n metres, respectively, and 0 otherwise	Y_b : equals 1 if demand in area b is not covered within p distance and 0 otherwise
$g \in G$: the set of bus stops	S_b : defines overall suitability of an area for candidate stations where $S_b > 0$	X_b : equals 1 if station location b is allocated and 0 otherwise
$t \in T$: the set of tram stops	h_b : is the fixed cost of setting up a station at candidate site b	
$m \in M$: the set of last-mile points	δ : is the unit penalty cost for uncovered demand	
$o \in O$: the set of last-mile points	θ_{bo} : is the amount of pollutant I on site b	
	v_o : cost per pollutant	
	λ : is the population to be served in area b	

$$\text{Min.} \sum_{b \in B} h_b(X_b + X_b S_b) + \delta \sum_{b \in B} \lambda_b Y_b + \sum_{o \in O} \sum_{b \in B} X_b \theta_{bo} v_o \quad (2)$$

Subject to:

$$\sum_{b \in B} d_{ba} X_b \geq 1 \forall a \in A \quad (3)$$

$$\sum_{b \in B} u_{bg} X_b \geq 1 \forall g \in G \quad (4)$$

$$\sum_{b \in B} w_{bt} X_b \geq 1 \forall t \in T \quad (5)$$

$$\sum_{b \in B} r_{bm} X_b \geq 1 \forall m \in M \quad (6)$$

$$X_b, Y_a = \{0, 1\}, \forall a \in A, b \in B \quad (7)$$

where

Equation (2) minimizes the sum of the bike station's cost, pollution penalty costs, and unmet demand. The set of restrictions are represented by Eqs. 3–7 and ensure that residents of each defined area can access at least one station in less than p meters (3), that users can access at least one station in less of j , k , and n meters from the bus stops (4), from the tram stops (5), and the last mile points (6), respectively. Finally, the integrality constraint ensures that the decision variables take 0 or 1, Eq. (7).

3.3 Designing a Software Solution for Station Distribution of a BSS

Software tools play an important role when integrating a series of components to collect, store, and process data. In order to present support for the station's location in a BSS, it has been presented the architecture of a tool that shows each component necessary to support designers towards a method that allows the location of stations in small and medium cities.

Therefore, this research has included a tool that provides designers with information about station distribution in the city. Data collected and analyzed will serve as input for the provided solution to obtain the optimization model (see Fig. 2-1). The pollutants are a data source to be considered in order to obtain a hedonic price model (see Fig. 2-4), which through a proposed information system, will report the price of the pollutants (see Fig. 2-3) used, and that will be the value that people are willing to pay in order to maintain the lowest environmental impact in their living area. On the other hand, optimization models will be used based on available data to provide a data-driven approach; with these data, the city's ideal areas obtained from analyzing the different available information will be defined. The result of these models will

BSS SYSTEM FOR DETERMINING STATION LOCATION – DATA DRIVEN APPROACH

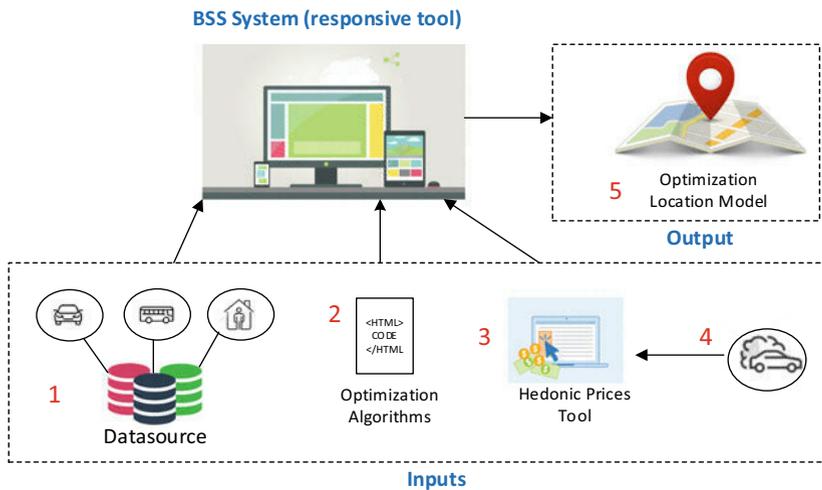


Fig. 2 BSS system for determining station’s location—data-driven approach

serve as an input for the proposed system in such a way that after analyzing them, it allows the visualization on a map. This map contains the stations’ distribution covering the city’s different points based on citizens’ demands and needs, as shown in Fig. 2-5. This software tool should be friendly and easy to use based on Web technologies. This proposed architecture is illustrated in Fig. 2.

The software has been designed to be implemented by using as backend node.js and for the front-end Angular.js. Also, several web services will be integrated into a Software Oriented Architecture (SOA), which implements a business layer that can be easily updated with new services that new data sources can feed. Also, other considerations related to event-driven patterns and asynchronous calls will be implemented. The orchestration of those services allows them to work together to implement and reach the objectives of the method.

Although the software is in an initial phase (i.e., requirements elicitation and design), a simulation of each step has been manually performed in which step by step has been followed in order to prove the feasibility of the activities when they will be implemented into the system.

4 Results

This article addresses the optimal distribution of stations (location and number) within a BSS as an alternative means of mobilization in small and medium-sized cities through an established methodological process. After applying the proposed



Fig. 3 **a** Location of bike stations and bus stops at 300 m. **b** Location of bike stations and tram stops at 300 m

optimization model, the results show different stations depending on the radius used (300 or 500 m). The ideal radius is 300 m, and the maximum limit is a radius of 500 m [42]. Therefore, the ideal radius for the analysis was considered.

The model obtained by the method and supported by the system estimated 345 and 146 stations for a distance radius between 300 and 500 m, respectively. These values provide coverage to the entire city applying the restrictions established in the model. Furthermore, the maps presented reveal that the stations' location covers the bus stops (see Fig. 3a), and the tram stops see Fig. 3b, demonstrating that a BSS can be integrated with the public transport systems available in the city (i.e., buses, trams) and private cars (last mile points).

The current locations were established by the consortium formed by a Spanish company and the public mobilization company in Cuenca, named BiciCuenca [14]. The model shows high precision (70%) in comparison with the consortium. If the results are analyzed considering the contiguous blocks, the precision reaches 90% since only two existing stations' locations do not coincide with the model. The algorithms, models, and SI used in the present study provide a data-driven approach to establish the city's station distribution. Using a data-driven approach helps to design the strategic planning of a BSS used by decision-makers. This method constitutes a good starting point for places without resources, expertise, and a high need for a BSS, saving them time and money.

This research contributes to the design, improvement, or restructuring of a BSS using a method that defines the stations' distribution in a parameterizable, repeatable, and organized way, based on accepted standards.

5 Discussion, Conclusions, and Future Research

The proposed software solution design incorporates various techniques such as econometric models, data analysis, machine learning (i.e., clusters), and operations

research, developing a mathematical optimization model. Several sources of information will be used, both primary (via a survey) and secondary (multiple public institutions). The model used to establish the stations' location considers a penalty for air and noise pollution, unsatisfied demand, and suitability.

Currently, the city of Cuenca has a bike-sharing system. The bike stations were located according to expert considerations (without scientific studies). The model had a precision of 70% (for exact blocks) using all the method parameters. Accuracy was 90% when contiguous blocks were included compared to existing stations in Cuenca city (prepared with expert experience).

The proposed software is computationally feasible and will work with different data types. Therefore, it will allow the station distribution's strategy step by step, which has not been explicitly done in previous studies. The implemented strategy allows a systematic flow that makes it repeatable, reliable, effective, efficient, and replicable in other similar situations in small and medium-sized cities. Furthermore, the proposal is less expensive than other studies requiring extensive or expensive human experts' participation; consequently, it saves economic, human, and time resources. In this context, the study contributes considerably to the strategic planning of a BSS related to the stations' distribution. Also, it contributes to the empirical literature on the use of pollution penalties within the model to increase the precision of station locations since no studies have been found that determine the location based on pollution data (NO_2 and noise). On the other hand, the list of factors used allows standardizing the data management approach.

Furthermore, in this article, only two pollutants were used, such as NO_2 and noise. Other pollutants such as fine particles (PM), sulfur dioxide (SO_2), and carbon monoxide (CO) could be interesting to analyze with the model in further investigations. On the other hand, other components such as bike lanes and the initial bike fleet could be considered to improve the proposed optimization model. A sensitivity analysis of the stations' optimal location can be examined in the next steps by adjusting the different penalty parameters' values and using other pollutants to identify the effect of model changes. These elements could be resolved in future research.

As a next step towards implementing the proposed solution, the structure of the data (format, type) and how data will be collected and inserted into the database will be defined. Another important aspect is the design of an optimal architecture in which the main components and their interconnections or relations will be clearly defined. In order to reach this objective, on the one hand, it will be included all the mechanisms used to clean data according to the enabled data sources by using a method (e.g., Cross-Industry Standard Process for Data Mining (CRISP)) or any other technique that allows the storing of optimal information.

Finally, the case study shows several aspects to be considered when a general method is defined. Therefore, as a first step, it can be concluded that implementing a solution that contributes to technicians in designing the best locations of bike stations can represent a powerful and cheap solution when leading the design of a BSS. This solution, therefore, represents the first approach of a tool that will support a method, which will be the next step in our research. Thus, there is expected to implement the software with the tools and languages explained in Sect. 3.

Acknowledgements This research was supported by the *Secretaría de Educación Superior, Ciencia, Tecnología e Innovación* (SENESCYT), Government of the Republic of Ecuador [grant number 704-2012]. We thank the Department of Computing and Information Sciences Ph.D. of RIT and the *Universidad de Cuenca* for the support provided to the researchers. We also thank GAD of Cuenca through their company EMOV-EP, CGA, ETAPA, SRI, INEC, CSC, who provided the necessary information for this research.

Appendices

Literature Review

Research	Objective and method
[22]	<ul style="list-style-type: none"> – Minimize bike stock cost, service level measured by coverage, and available bikes – Greedy heuristic for stations (open/close station until reduced cost), Dijkstra’s algorithm to determine the lanes, and Palm’s theorem for inventory management
[20]	<ul style="list-style-type: none"> – Proposed a deterministic centralized approach based on an integer non-linear program model – It manages an optimal inventory level to restock the bikes inventory where the replenishment costs and the units are minimal, and the stations do not remain unserved in the system
[43]	<ul style="list-style-type: none"> – Maximize the revenue for stations, bikes, and relocation. The problem was solved in several days; it means computationally intensive and with synthetic data. Based on simulations of trips by the hour, the authors defined the fleet, station capacity, and location. They proposed a mix of strategic and operational planning
[21]	Minimize the cost of private cars and public bikes with a bilevel optimization problem. The authors use the population data to determine station locations, and they integrate two means of transportation
[23]	Minimize the travel distance between stations and the overlapping between predefined clusters defining station locations at scenic spots. The optimization cluster by cluster improves computing time; it was executed in a city’s specific zone
[16]	Used a bilevel optimization model to determine station location, bike number, and parking lots for bikes. They minimize the total cost of the whole transportation system (i.e., car, buses, cycling) with a study case in Mexico
[44]	Minimize travel time, consider initial bikes and parking lockers. The authors determine the dynamic demand from O/D in periods. They used an illustrative example and a mixed strategy with operational planning