GEOProcessing 2024

The Sixteenth International Conference on Advanced Geographic Information Systems, Applications, and Services

ISBN: 978-1-68558-168-8

May 26 - 30, 2024

Barcelona, Spain

GEOProcessing 2024 Editors

Oren Gal, University of Haifa, Israel

Roger Tilley, Sandia National Laboratories, USA

GEOProcessing 2024

Forward

The Sixteenth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2024), held between May 26th and May 30th, 2024, in Barcelona, Spain, continued a series of events addressing the aspects of managing geographical information and web services.

The goal of the GEOProcessing 2024 conference was to bring together researchers from the academia and practitioners from the industry in order to address fundamentals of advances in geographic information systems and the new applications related to them using the Web Services. Such systems can be used for assessment, modeling, and prognosis of emergencies.

GEOProcessing 2024 provided a forum where researchers were able to present recent research results and new research problems and directions related to them. The topics covered aspects from fundamentals to more specialized topics such as 2D & 3D information visualization, web services and geospatial systems, geoinformation processing, and spatial data infrastructure.

We take here the opportunity to warmly thank all the members of the GEOProcessing 2024 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to GEOProcessing 2024. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the GEOProcessing 2024 organizing committee for their help in handling the logistics of this event.

We hope that GEOProcessing 2024 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field of geographic information systems, applications, and services.

GEOProcessing 2024 Chairs

GEOProcessing 2024 Steering Committee Chair
Claus-Peter Rückemann, Universität Münster / DIMF / Leibniz Universität Hannover, Germany

GEOProcessing 2024 Steering Committee
Thomas Ritz, FH Aachen, Germany
Yerach Doytsher, Technion - Israel Institute of Technology, Haifa, Israel
Roger Tilley, Sandia National Laboratories, USA
Alexey Cheptsov, Information Service Center of the University of Stuttgart (TIK), Germany
Jui-Hsin (Larry) Lai, Ping-An Technology - Research Lab, USA

GEOProcessing 2024 Publicity Chairs
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain
Laura Garcia, Universidad Politécnica de Cartagena, Spain

GEOProcessing 2024
Committee

GEOProcessing 2024 Steering Committee Chair

Claus-Peter Rückemann, Universität Münster / DIMF / Leibniz Universität Hannover, Germany

GEOProcessing 2024 Steering Committee

Thomas Ritz, FH Aachen, Germany
Yerach Doytsher, Technion - Israel Institute of Technology, Haifa, Israel
Roger Tilley, Sandia National Laboratories, USA
Alexey Cheptsov, Information Service Center of the University of Stuttgart (TIK), Germany
Jui-Hsin (Larry) Lai, Ping-An Technology - Research Lab, USA

GEOProcessing 2024 Publicity Chairs

Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain
Laura Garcia, Universidad Politécnica de Cartagena, Spain

GEOProcessing 2024 Technical Program Committee

Mahdi Abdelguerfi, University of New Orleans, USA
Alia I. Abdelmoty, Cardiff University, Wales, UK
Danial Aghajarian, Georgia State University, USA
Nuhcan Akçit, Middle East Technical University, Turkey
Zaher Al Aghbari, University of Sharjah, UAE
Bharath H. Aithal, IIT Kharagpur, India
Md Mahbub Alam, Dalhousie University, Canada
Heba Aly, University of Maryland, College Park, USA
Francisco Javier Ariza López, Escuela Politécnica de Jaén - Universidad de Jaén, Spain
Thierry Badard, Centre de Recherche en Géomatique (CRG) | Université Laval, Canada
Abderrazak Bannari, Arabian Gulf University, Bahrain
Fabian Barbato, UDELAR-School of Engineering, Uruguay
Melih Basaraner, Yildiz Technical University, Turkey
Peter Baumann, rasdaman GmbH Bremen / Jacobs University Bremen, Germany
Mikael Brunila, McGill University, Canada
Lorenzo Carnevale, University of Messina, Italy
Mete Celik, Erciyes University, Turkey
Aizaz Chaudhry, Carleton University, Canada
Wei Chen, University of Birmingham, UK
Dickson K.W. Chiu, The University of Hong Kong, Hong Kong
Alexey Cheptsov, Information Service Center of the University of Stuttgart (TIK), Germany
Daniel Enrique Constantino Recillas, TESE, Mexico
Giuliano Cornacchia, University of Pisa / Institute of Information Science and Technologies, National Research Council of Italy (ISTI-CNR), Italy
Alexandre Corrêa da Silva, HEX Geospatial Technologies, Brazil

Ying Lu, DiDi Research America, Mountain View, USA
Giovanni Ludeno, Institute for the Electromagnetic Sensing of the Environment | National Research Council of Italy, Napoli, Italy
Dandan Ma, Northwestern Polytechnic University, China
Ali Mansourian, Lund University, Sweden
Jesús Martí Gavilá, Research Institute for Integrated Management of Coastal Areas (IGIC) | Universitat Politècnica de València, Spain
Giovanni Mauro, University of Pisa / IMT School for Advanced Studies, Lucca / Institute of Information Science and Technologies, National Research Council of Italy (ISTI-CNR), Italy
Sara Migliorini, University of Verona, Italy
Sobhan Moosavi, The Ohio State University, USA
Hanan Muhajab, Cardiff University, UK / Jazan University, Saudi Arabia
Tathagata Mukherjee, The University of Alabama in Huntsville, USA
Beniamino Murgante, University of Basilicata, Italy
Ahmed Mustafa, The New School University, New York, USA
Purevtseren Myagmartseren, National University of Mongolia, Mongolia
Aldo Napoli, MINES ParisTech - CRC, France
Maurizio Napolitano, Fondazione Bruno Kessler, Trento, Italy
Rayan Nas, Yüksel Proje Inc., Ankara, Turkey
Rouhollah Nasirzadehdizaji, Istanbul University - Cerrahpaşa, Turkey
Alexey Noskov, Philipps University of Marburg, Germany
Lidia Ortega Alvarado, University of Jaén, Spain
Xiao Pan, Shijiazhuang Tiedao University, China
Shray Pathak, School of Geographic Sciences | East China Normal University, Shanghai, China
Akshay Patil, AiDASH.inc, India
Kostas Patroumpas, Athena Research Center, Greece
Davod Poreh, Università degli Studi di Napoli "Federico II", Italy
Viktor Prasanna, University of Southern California, USA
Kuldeep Purohit, Michigan State University, USA
Honggang Qi, University of Chinese Academy of Sciences, China
Thomas Ritz, FH Aachen, Germany
Ricardo Rodrigues Ciferri, Federal University of São Carlos (UFSCar), Brazil
Armanda Rodrigues, Universidade NOVA de Lisboa | NOVA LINCS, Portugal
Claus-Peter Rückemann, Universität Münster / DIMF / Leibniz Universität Hannover, Germany, Germany, Germany
André Sabino, Universidade Europeia, Portugal
Arpan Sainju, Middle Tennessee State University, USA
Markus Schneider, University of Florida, USA
Raja Sengupta, McGill University, Montreal, Canada
Elif Sertel, Istanbul Technical University, Turkey
Shih-Lung Shaw, University of Tennessee, Knoxville, USA
Yosio E. Shimabukuro, Brazilian Institute for Space Research - INPE, Brazil
Rajat Shinde, Indian Institute of Technology Bombay, India
Dericks Praise Shukla, IIT Mandi, India
Spiros Skiadopoulos, University of the Peloponnese, Greece
Dimitris Skoutas, Athena Research Center, Greece
Srushti Rashmi Shirish, Woven planet holdings, Tokyo, Japan
Francesco Soldovieri, Istituto per il Rilevamento Elettromagnetico dell'Ambiente - Consiglio Nazionale

Copyright Information

Table of Contents

Improving the Digital Cartographic Reference Data of the Walloon Region, Belgium (PICC) : A Comprehensive Methodology for Documenting Updating and Quality Control Processes

Sophie Petit[1], Benjamin Beaumont[1,2], Éric Hallot[1]
[1]Cellule Télédétections et Géodonnées
Institut Scientifique de Service Public
Liège, Belgium
e-mail : {s.petit, e.hallot}@issep.be

Florence Jonard[2], Jean-Claude Jasselette[2]
[2]Production géomatique et traitement de la donnée
Service Public de Wallonie
Namur, Belgium
e-mail : {florence.jonard, jeanclaude.jasselette, benjamin.beaumont}@spw.wallonie.be

Abstract— **The Service Public de Wallonie aims to set up an INSPIRE-compliant Georepository, using the "Projet Informatique de Cartographie Continue" (PICC) as a baseline. This dynamic geodatabase, initiated in 1992, covers the entire Walloon Region, in Belgium, with precision below 25 cm in x, y, and z coordinates, serving various sectors and undergoing continuous updates. As a major cartographic reference, it includes features, such as buildings, roads, and addresses, providing a comprehensive database for the entire region. Implementing the Business Process Model and Notation (BPMN) methodology, the study models the PICC update workflows and quality controls, resulting in 7 processes divided into 46 diagrams. This approach enhances the PICC management, streamlines processes, strengthens quality controls, and optimizes data architecture, ensuring its relevance and usefulness in a dynamic geospatial context.**

Keywords-geodata; process workflow; quality controls; BPMN.

I. Introduction

The Public Service of Wallonia (SPW) aims to set up a Georepository in accordance with the INSPIRE directive [1], guaranteeing the quality of geodata. The "Projet informatique de cartographie continue" (PICC) has been selected as a basis for its development. The PICC plays a key role as the tree-dimensional cartographic reference covering the entire Walloon Region in Belgium. It includes all the identifiable elements of the Walloon landscape, with a precision below 25 cm in their x, y, and z coordinates. Initiated in 1992, the PICC is a dynamic geodatabase that is continuously updated to reflect the constant evolution of the Walloon territory. Freely accessible through a Web Service via WalOnMap [2], it serves a wide range of sectors, including network operators and geometers, or is used for example as a basis for spatial analyses in combination with remote sensing technologies.

A first study [3] proposed a quality control methodology for three geodata features: buildings, road axes, and point addresses. The latter led to the definition of a theoretical basis for the validation of geodata quality. As a complement to this study, it was necessary to model the flows of the PICC update process, together with the existing quality analysis processes, to ensure their compliance with current standards, consolidate and improve them where necessary. To this end, the international standard methodology for business process modeling, "Business Process Model and Notation" (BPMN) [4], has been implemented.

The paper is organized into three sections: the BPMN, the BPMN application to the PICC and a conclusion.

II. BPMN

The BPMN approach provides a clear and intuitive graphical notation of activities that can be understood by all, while also being capable of representing complex processes. Several categories of elements in BPMN exist, the main ones being shown in Figure 1.

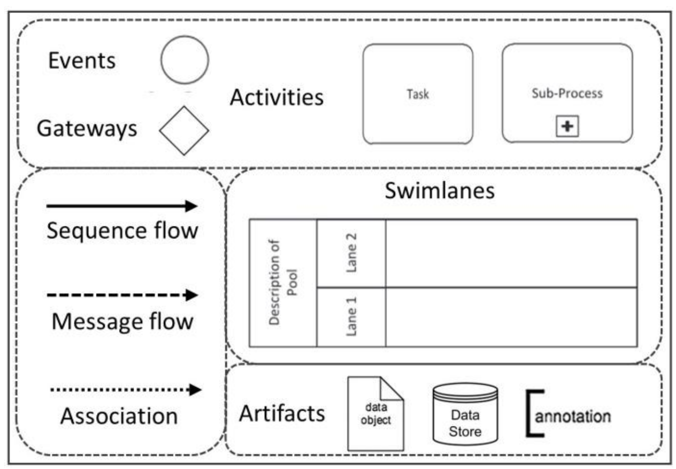

Figure 1. Basic elements in BPMN.

The methodology offers a number of advantages, particularly through the implementation of a collaborative approach in which diagrams are co-constructed by leveraging the experience of experts.

III. BPMN Application to the PICC

Our approach consists in modeling the update workflows and associated quality controls of the PICC geographic data.

Figure 2 shows the PICC update workflows consisting of two separate branches, one per database. This arrangement stems from the coexistence of the PICC database with an official database specially designed for address management. The two branches are handled together at the final stage, for data distribution.

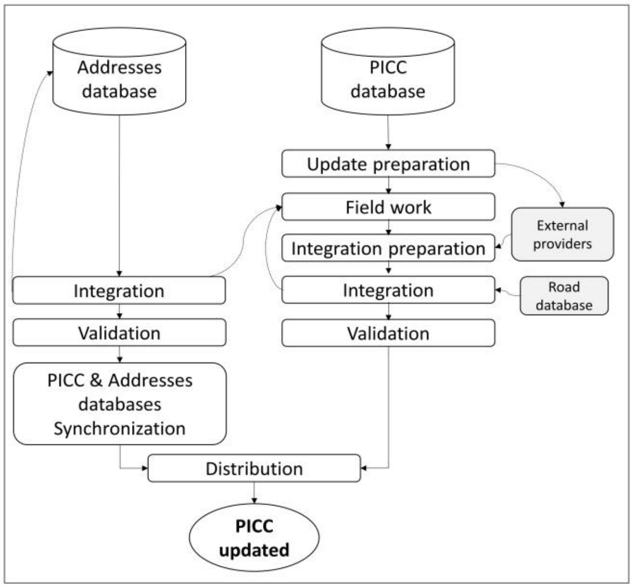

Figure 2. The PICC update workflow.

In this context, BPMN provides an opportunity to have an exhaustive vision, both globally and in-depth, in order to make further improvements where required. BPMN has seldom been used in the field of geographic information science. Reference [5] applies this methodology to formalize the conceptualization of spatio-temporal data and the management of business processes.

The outcome of the BPMN application to the PICC is a set of 7 processes divided in 46 diagrams which, when combined, offer a complete and detailed vision of the PICC update processes and sub-processes. Figure 3 presents an example of one of these diagrams, where external data is used as input for the integration of changes in the PICC. Specifically, it presents the integration of the Road database into the "Integration" process of the PICC database branch from the PICC update workflow shown in Figure 2.

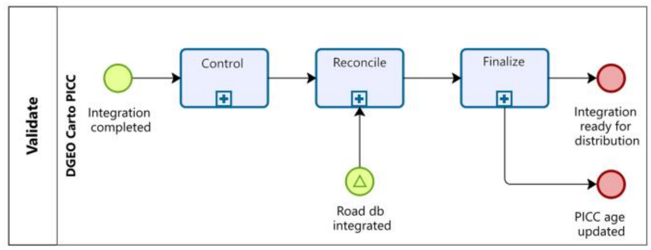

Figure 4. Diagram example: the PICC update "Validation" process.

Figure 4 shows a diagram of the "validation" process which is divided into three sub-processes. It illustrates the connections between the different diagrams, with the starting event, occurring during the second sub-process, corresponding to the end of the diagram in Figure 3.

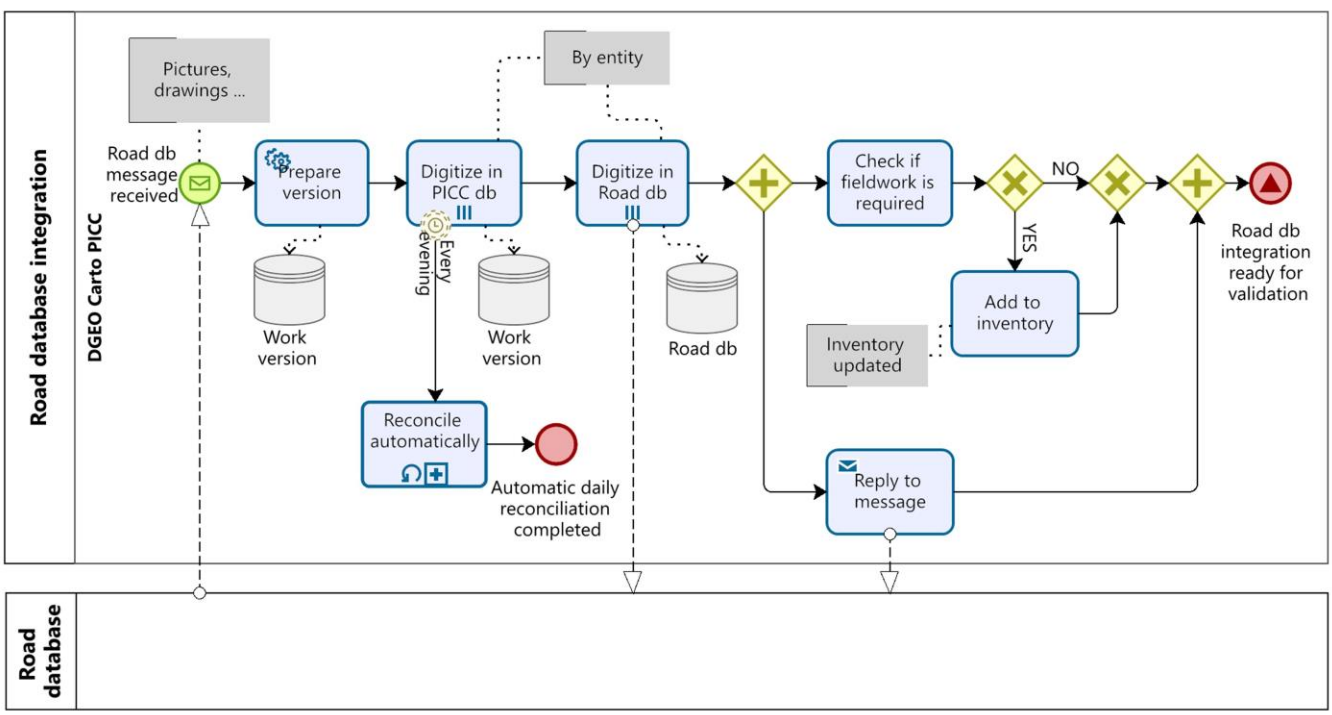

Figure 3. Diagram example: sub-processes of the PICC update "Integration" process.

 ISBN: 978-1-68558-168-8

IV. CONCLUSION

The use of BPMN methodology in the context of geographic data has seldom been used. Its application to the PICC has shown its relevance by providing an exhaustive vision, both globally and in-depth, of updating processes and associated quality controls. This approach has contributed to the overall improvement of the PICC management, leading to an ongoing restructuring of processes and a strengthening of quality controls. In addition, an in-depth analysis of data architecture has been undertaken to simplify existing structures, while ensuring that the final quality of the PICC is enhanced. This methodological initiative has enabled the overall optimization of the PICC, while reinforcing its relevance and usefulness in a dynamic geospatial context.

REFERENCES

[1] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). (2007). Official Journal of the European Union (EurLEX). Available from: https://eur-lex.europa.eu/eli/dir/2007/2/2019-06-26 (accessed on 07 february 2024)

[2] Service Public de Wallonie. *WalOnMap* [Online]. Available from: https://geoportail.wallonie.be/catalogue/b795de68-726c-4bdf-a62a-a42686aa5b6f.html (accessed on 07 february 2024)

[3] G.-A. Nys, C. Dubois, C. Goffin, P. Hallot, J.-P. Kasprzyk, M. Treffer and R. Billen, "Geodata Quality Assessment and Operationalisation of the Inspire Directive: Feedback" "Évaluation de la qualité des géodonnées et opérationalisation de la directive INSPIRE : retour d'expérience", Bulletin de la Société Géographique de Liège , 78, 2022, pp. 179-188, doi: 10.25518/0770-7576.6698

[4] Object Management Group, Inc. (OMG), "Business Process Model and Notation (BPMN)", Version 2.0, 2010

[5] C. Zaki, C. Claramunt, A. Nasser and S. Bahmad, "Merging Spatio-Temporal Objects and Business Processes: Land Reform Process Case Study", Appl. Sci. 2023, 13, 12372. https://doi.org/10.3390/app132212372

Advanced User Interface for a Geospatial Data Integration Platform

Lassi Lehto, Jaakko Kähkönen, Panu Muhli and Juha Oksanen
Department of Geoinformatics and Cartography
Finnish Geospatial Research Institute
Espoo, Finland
e-mail: lassi.lehto@nls.fi, jaakko.kahkonen@nls.fi, panu.muhli@nls.fi, juha.oksanen@nls.fi

Abstract—**Various solutions for dynamic, service-level integration of geospatial data have been developed in the EU-funded Geospatially Enabled Ecosystem for Europe (GeoE3) project. The GeoE3 geodata integration platform is a cloud-based service providing single point access to the thematic content offered by the project. In particular, the project has focused on on-the-fly cross-border integration of datasets originating from legacy country-level services. Access to the platform is based on the Open Geospatial Consortium's (OGC) latest interface standards OGC API Features and OGC API Coverages (API: Application Programming Interface). A separate layer of functionality has been added to the platform to combine the available content themes into a single service end point and to seamlessly integrate the country-level datasets across national borders. As the final step, an enhanced Web user interface for accessing the GeoE3 integration platform has been designed and developed. The new user interface facilitates visualization, interaction with and download of geospatial content across national borders and thematic domains.**

Keywords-user interface; data integration; cross-border; OGC API Features; OGC API Coverages.

I. INTRODUCTION

The Geospatially Enabled Ecosystem for Europe (GeoE3) project has developed advanced service-level solutions for cross-border and cross-domain geodata integration to support renewable energy-related applications [1][2]. By focusing to the needs of a well-defined use case, the project could focus on developing data integration solutions for a limited number of data sets. Energy-related use cases focused on solar energy necessitate the availability of 3D buildings, Digital Terrain Models and Digital Surface Models. A considerable effort was done to make 3D buildings available on the platform. 3D buildings are available from all participating countries, either as genuine Level of Detail 2 (LoD2) models or as experimental LoD1 models, dynamically generated by the service [3].

Based on the selected use cases, integration of climate attributes was also deemed important. The data portals of national meteorological agencies were accessed to retrieve climate information and the resulting point observation values were transformed to an interpolated coverage. Further on, the coverage is queried with a point location to derive a climate value for instance for an individual building [4].

An enhanced Web user interface has been developed to efficiently visualize, interact with and access the datasets available on the service platform of the project. The user interface is divided into two sections: one supporting easy access for the general public and the decision-maker level users and the other aimed at professional software developers.

The rest of the paper is organized as follows. Section II introduces the GeoE3 integration platform. Section III discusses the solutions used for integrating the separate datasets together. Section IV describes the new enhanced two-part user interface. The paper ends with final conclusions in Section V.

II. INTEGRATION PLATFORM

The GeoE3 integration platform is a cloud-based service providing single point access to all thematic content offered by the project [5]. The platform is predominantly based on pygeoapi, a Python implementation of the OGC API family of service interface standards [6]. On the GeoE3 platform, pygeoapi has been adapted to run on top of the web service development framework Django [7]. The capabilities of pygeoapi have been extended in the project, for example to facilitate the use of multiresolution raster data storage.

The platform provides harmonized access to the following themes: 2D/3D buildings, 2D/3D roads, Digital Terrain Model (DTM) and Digital Surface Model (DSM). In addition, the platform also provides climate-related parameters, like temperature, windspeed and sunshine to support the selected renewable energy-related use cases. At the moment the platform contains altogether 40 datasets/services (2D buildings: 6, 3D buildings: 6, 2D roads 4, 3D roads: 3, DTM: 5, DSM: 5, temperature: 5, windspeed: 4, sunshine: 2). Most of the datasets are accessed from the country-level services in query time.

Specialized content provider modules have been developed within the pygeoapi software to access national service end points providing data in national schemas. The integration platform takes care of all required harmonization procedures to make the content offering consistent across national borders.

Participating countries are treated as individual data collection datasets inside the single theme-specific OGC API service instance. For example, the theme 'Buildings 2D' contains currently five data collection datasets, one for each participating country. This arrangement enables natural cross-border data integration inside a single service end point.

 ISBN: 978-1-68558-168-8

III. Integration Approaches

The project's work in implementing OGC API Features and OGC API Coverages standards on the access interface of the platform deviate in some significant aspects from the traditional approach [8][9]. For instance, the GeoE3 services aim at maintaining the map zoom-in metaphor in content access. Instead of relying on tabular, paged feature browsing, the GeoE3 user interface allows for map zoom-in and panning while locating the user's area of interest - which might cross national borders. Data retrieval is commenced only when the zoom level of the map in the user interface has reached the pre-defined minimum value. In the case of the raster datasets and the OGC API Coverages access interface, the multiresolution nature of the source datasets makes it possible to enable visualization and content retrieval across the whole scale range. For this to work, support for the processing of 'scale-size' parameter was added to the base software pygeoapi [10].

To facilitate the map-browsing based content retrieval, support for a background map was added relying on map services provided by national mapping agencies. In the integrating visualization view, Explore, OpenStreetMap is used as the background map.

The handling of Coordinate Reference Systems (CRSs) is a challenging task in the applications integrating content across borders and domains. In the case of the GeoE3 Integration Portal, several CRSs have to be supported. In the OGC API Features services the output CRS is required by the standard to be the longitude, latitude WGS84 CRS. As the national level services are mostly based on Web Feature Service (WFS) standard, the request can be made in the desired CRS, using WGS84 in this case, and no coordinate transformations are needed on the Integration Platform. However, the order of the CRS axes still requires careful consideration in the computing processes on the Platform.

In the case of the raster data and OGC API Coverages, the issue is a bit more complicated. In the country level services, the raster data content is stored in the national CRSs. In the GeoE3 Portal, the access interface accepts queries only in the national CRS. In the map application of the user interface, the query bounding box is transformed to the national CRS and the returned dataset is again transformed to the projection of the map application. This transformation causes the raster image to warp. To ensure that the resulting image will fully cover the map viewport, the request bounding box has to be somewhat extended by the service.

While doing integrated visualization in the user interface, the map application is using the OpenStreetMap background map with Pseudo Mercator CRS. The requested raster dataset, potentially crossing national borders, must be projected to the Pseudo Mercator CRS and integrated into a same image by the Integration Portal.

IV. User Interface

The new Web user interface of the GeoE3 integration platform has been enhanced significantly to support easy access to all provided geospatial content. The user interface has been published in the dedicated Web domain locationeurope.eu and aims at promoting the use of the already existing content and also encouraging new organizations to join the platform as data providers.

The development of the Web user interface has been guided by a professional Web designer and is based on the use of the Bootstrap framework [11]. Map-related functions of the user interface are based on the widely-used Web map library OpenLayers [12]. The user interface is branded according to the Finnish Location Innovation Hub styling and is being further developed and marketed in the context of this initiative.

The most significant new idea in the user interface design is the division of the functionality into two separate parts, aimed at two distinct user groups: decision-makers and general public on one hand and professional software developers on the other. The part designed for the latter group follows closely the API structure of the OGC API family of standards and is similar with the pygeoapi default user interface and many other implementations of the OGC API standards. The part aimed at decision-makers and general public is more innovative and provides some brand-new functionalities.

At core of the decision-maker's user interface is the Explore map view that provides an integrated view to all content available on the integration platform. In the Explore view, all the content themes are listed and can be visualized on top of the OpenStreetMap-based background map. The user interface thus aggregates the output of several individual OGC API Features and OGC API Coverages services into a single map view. Behind the service end point, the platform combines all country level services of each theme as a single integrated output. On the Explore view, the country polygons are shown to indicate, in which areas all the currently active themes are available, see Figure 1.

The new cross-theme integration creates a new conceptual computing layer on top of the theme-specific service instances that takes care of the content integration tasks, see Figure 2. In this arrangement, the integration layer treats the theme-specific services as its data collections and the country-level services, that formerly were seen as data collections, are now behind the scenes automatically combined as a single seamless theme.

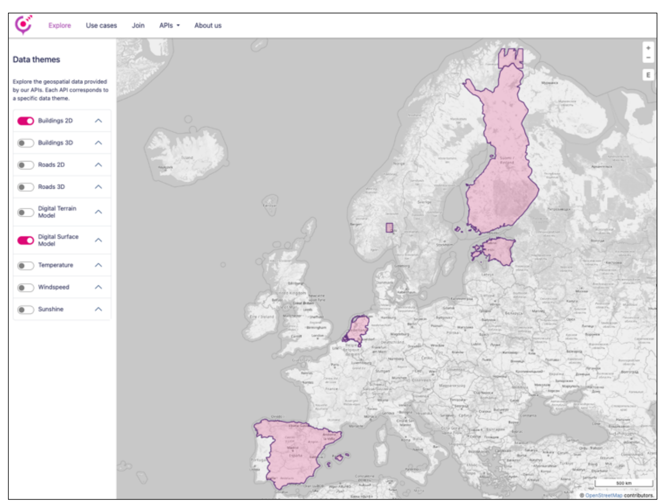

Figure 1. The Explore view with two themes selected and the areas, from which those themes are available, shown as coverage polygons.

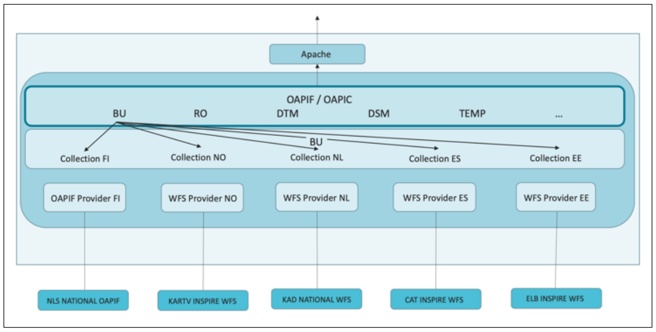

Figure 2. The new computing layer on the integration platform, responsible for combining the themes together and performing automated cross-border integration (Buildings (BU) as an example).

During the query processing, the integrating layer has to overlap the query bounding box with the country polygons to determine, which background services have to be included into the process. The resulting individual country-level responses have then to be combined into a single dataset. The processing involves several CRS transformations, carried out both on query parameters and on the response data.

The Explore view also allows the user to switch individual countries' datasets on and off in the map view, if desired. This functionality necessitates information about the visible countries to be passed to the server side. For this purpose, a new query parameter 'collections' has been introduced to the service request. This represents an extension to the OGC API Features and OGC API Coverages service interface specifications and has to be regarded as an experimental function. However, this functionality closely resembles the concept 'cross-collection query' that was originally introduced in the OGC API Features Part 3, but finally removed from it in the published standard document [13].

In the integrated Explore view, several vector themes, like buildings and roads, can be visualized together. From the available raster themes, only one can be visible at a time. This enables the user to visualize for instance a Digital Terrain Model or a Digital Surface Model together with buildings and roads, see Figure 3. The opacity of each content layer can be adjusted individually.

Buildings and roads can also be viewed as three dimensional objects. A 3D visualization tool called ThreeJsViewer, developed by the Delft University of Technology's 3D Geoinformation group [14], is used for visualizing of and interacting with the 3D models in the Web browser, see Figure 4. 3D content is requested from the integration platform via the OGC API Features interface and encoded in the CityJSON format [15]. The content integration capabilities of the integration platform enable cross-border visualization of 3D models, too, see Figure 5.

In some cases, the LoD1 3D building models are created dynamically by the integration platform using building 2D footprints and heights derived from DTM and DSM. In the case of Spanish buildings, the 3D model is generated by utilizing knowledge about the number of floors in each part of the building, see Figure 6.

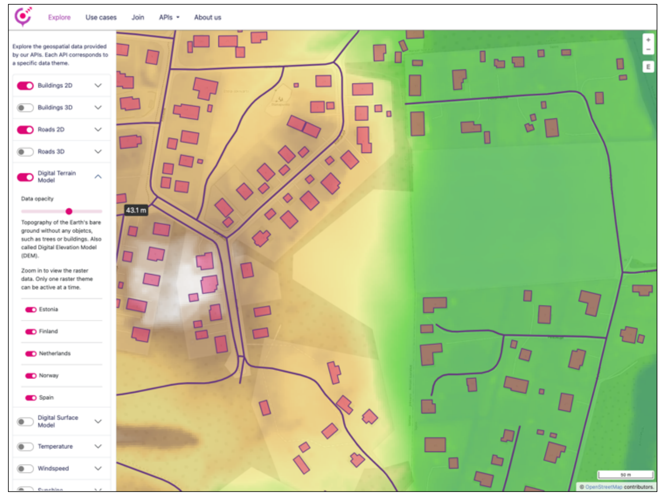

Figure 3. The Explore view showing Buildings and Roads, requested from an OGC API Features service interface, together with Digital Terrain Model, requested from an OGC API Coverages interface.

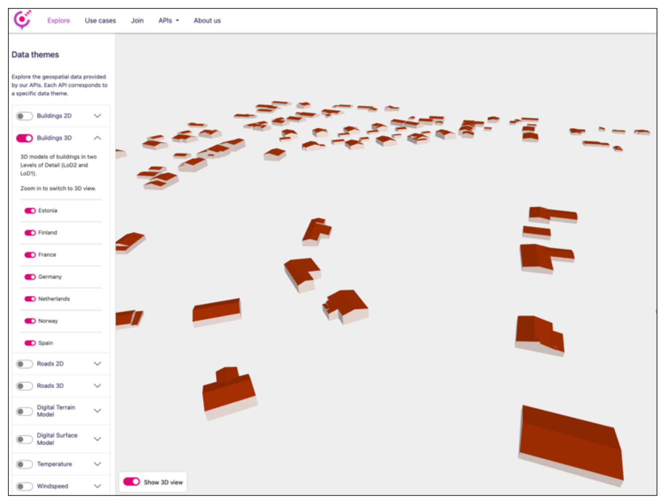

Figure 4. The Explore view showing 3D Buildings, visualized with the ThreeJsViewer component.

The starting page of the user interface represents at the same time the landing page of the integrated cross-theme service end point. All the available themes are introduced already on this page to smooth the browsing experience of an uninitiated user. The landing page thus also includes information usually available on the 'collections' level of the API. The themes are presented as info cards containing an example image of the theme visualization and direct links to the corresponding API landing page and to the Explore view, see Figure 7.

 ISBN: 978-1-68558-168-8

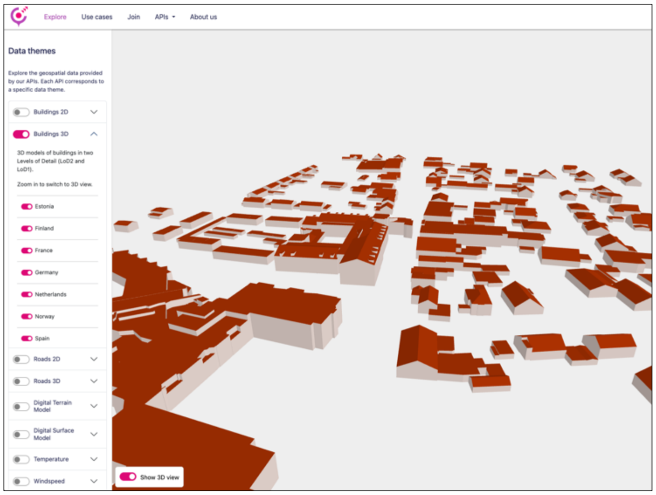

Figure 5. The Explore view showing 3D Buildings in a cross-border area (Dinxperlo, The Netherlands, on the left and Suderwick, Germany, on the right).

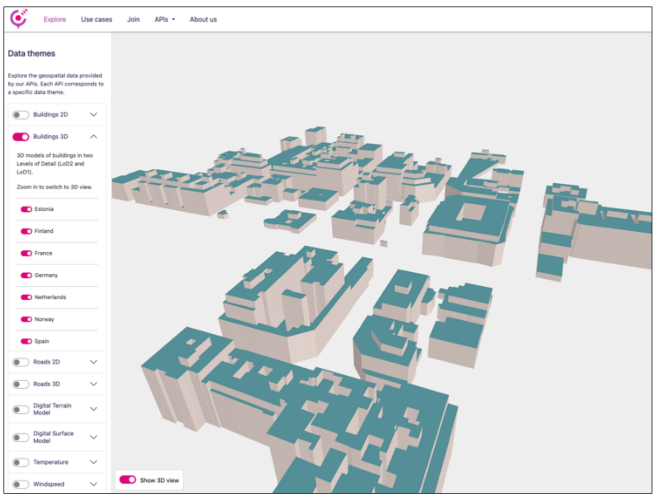

Figure 6. Dynamically created LoD1 3D models, based on buildings' 2D footprint and information on the number of floors in each building part (Barcelona).

Other functionalities provided by the user interface for decision-makers' category include pages for introducing the current members of the data provider network and a page for joining the network as a new content provider.

The user interface designed for the developer community strictly follow the API structure providing separate Web page for the API landing page, collections list, description of a collection, collection items or collection coverage, and the collection item for visualization of a single vector feature together with its attributes, see Figure 8.

Additional user interface pages for the developer include a page for requesting a free API key required for direct calls to the APIs from the user's own application, and a page providing detailed instructions on the formulation of the queries.

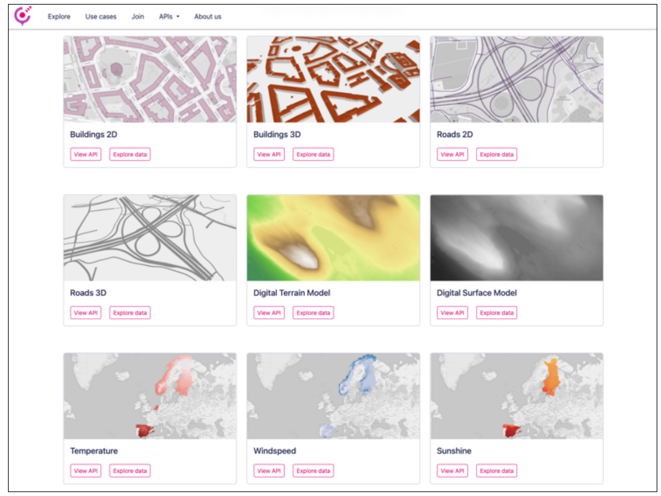

Figure 7. The info cards on the landing page introducing the available content themes (Describe Collections in OGC API).

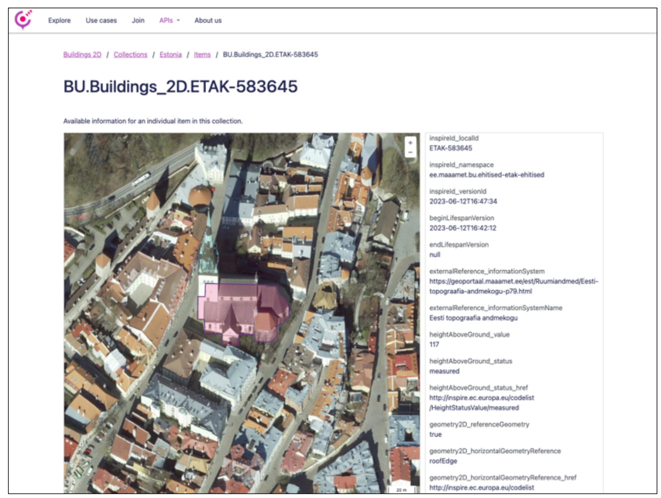

Figure 8. A single vector item, a Building, in the API view (Get Item), (Oleviste Church, Tallinn).

V. CONCLUSION

A flexible geodata integration mechanism has been established on the GeoE3 integration platform. Participating countries' datasets are treated as individual data collections inside an OGC API Features or an OGC API Coverages service instance. This setting enables easy cross-border integration of geospatial content. A new layer of functionality has been developed on top of these theme-specific OGC API service nodes. This layer combines all themes into a single service end point and carries out cross-border content integration automatically behind the scenes. As an extension to the OGC API standards, the service also supports a query parameter 'collections' enabling the client side to dictate, which countries' datasets are to be included in the response, if desired.

A new enhanced Web user interface has been developed to access the data offerings available on the GeoE3 integration platform. The user interface supports easy viewing of all the content on the platform by providing a cross-theme Explore map view, by which all content themes from all participating countries can be accessed. The Explore view also contains facilities for interactive visualization of 3D content – enabling even cross-border viewing of 3D objects. A separate section of the user interface supports the traditional developer-oriented approach, in which the structure of the Web pages directly reflects the structure of the API itself.

The cross-border and cross-domain content integration functionality offered by the GeoE3 integration platform facilitates applications requiring access to harmonized geospatial datasets throughout Europe, specifically supporting renewable energy-related use cases. In particular, it supports the efforts of companies working in this domain to expand their operations across national borders.

ACKNOWLEDGMENT

The GeoE3 project is co-financed by the Connecting Europe Facility (CEF) of the European Union with the grant agreement number INEA/CEF/ICT/A2019/2063390.

We made use of the GeoCubes service provided by the Open Geospatial Information Infrastructure for Research (Geoportti, urn:nbn:fi:research-infras-2016072513) funded by the Academy of Finland, CSC – IT Center for Science, and other Geoportti consortium members.

The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

REFERENCES

[1] GeoE3, Geospatially Enabled Ecosystem for Europe, GeoE3 Home Page. [Online]. Available from: https://geoe3.eu, 2024, [retrieved: Apr, 2024]

[2] L. Lehto, J. Kähkönen, J. Reini, T. Aarnio, and R. Tervo, “Cross-border and Cross-domain Integration of 3D Content in a European Geospatially Enabled Ecosystem”, *International Journal on Advances in Intelligent Systems*, ISSN 1942-2679, Vol. 15, no. 1 & 2, year 2022, pp. 1-11, http://www.iariajournals.org/intelligent_systems/

[3] L. Lehto and J. Kähkönen, “Cross-border Delivery and Web-based Visualization of 3D Buildings”, *GEOProcessing 2022, the Fourteenth International Conference on Advanced Geographic Information Systems, Applications, and Services*, Jun 26-30, 2022, Porto, Portugal, pp. 36-40. ISBN: 978-1-61208-983-6

[4] L. Lehto and J. Kähkönen, “Dynamic Integration of Climate Properties with Geospatial Data for Energy Applications”, *GEOProcessing 2023, the Fifteenth International Conference on Advanced Geographic Information Systems, Applications, and Services*, Apr 24-28, 2023, Venice, Italy, pp. 37-40. ISBN: 978-1-68558-079-7

[5] L. Lehto, J. Kähkönen, J. Oksanen and P. Muhli, 2023. Use Case Driven Dynamic Integration of Cross-Domain Content, *Abstr. Int. Cartogr. Assoc.*, 6, 136, https://doi.org/10.5194/ica-abs-6-136-2023, 2023

[6] T. Kralidis, pygeoapi Home Page. [Online]. Available from: https://pygeoapi.io, 2024, [retrieved: Apr, 2024]

[7] Django, Django Home Page. [Online]. Available from https://www.djangoproject.com, 2024, [retrieved: Apr, 2024]

[8] OGC, OGC API – Features Home Page. [Online]. Available from: https://ogcapi.ogc.org/features/, 2024, [retrieved: Apr, 2024]

[9] OGC, OGC API – Coverages Home Page. [Online]. Available from: https://ogcapi.ogc.org/coverages/, 2024, [retrieved: Apr, 2024]

[10] OGC, OGC API – Coverages – Part 1: Core. [Online]. Available from: https://docs.ogc.org/DRAFTS/19-087.html, 2023, [retrieved: Apr, 2024]

[11] Bootstrap, Bootstrap Home Page. [Online]. Available from: https://getbootstrap.com, 2024, [retrieved: Apr, 2024]

[12] OpenLayers, OpenLayers Home Page. [Online]. Available from: https://openlayers.org, 2024, [retrieved: Apr, 2024]

[13] OGC, “OGC API Features – Part 3: Filtering”. [Online]. Available from https://docs.ogc.org/DRAFTS/19-079r1.html, 2024, [retrieved: Apr, 2024]

[14] TUDelft, TU Delft 3D Geoinformation Home Page. [Online]. Available from: https://3d.bk.tudelft.nl, 2024, [retrieved: Apr, 2024]

[15] CityJSON, CityJSON Home Page. [Online]. Available from: https://www.cityjson.org, 2023, [retrieved: Apr, 2024]

Improved Prediction of Olive Crop Yield Using Satellite Imagery

A Case Study in Andalusia (Spain)

M. Isabel Ramos
Dept. Cartographic Engineering, Geodesy and Photogrammetry
University of Jaen
Jaen, Spain
miramos@ujaen.es

Lidia Ortega
Dept. Computer Science
University of Jaen
Jaen, Spain
lidia@ujaen.es

Angel Calle
Researcher hired for project
PREDIC I-GOPO-JA-20-0006
University of Jaen
Jaen, Spain
acalle@ujaen.es

Ruth M. Cordoba
Researcher hired for project
PREDIC I-GOPO-JA-20-0006
University of Jaen
Jaen, Spain
rcortega@ujaen.es

Juan J. Cubillas
Dept. Information and Communication Technologies Applied to Education
International University of La Rioja
Logrono, Spain
juanjose.cubillas@unir.net

***Abstract*—Agriculture is one of the strategic economic and social sectors in many countries. In Spain, agriculture has been and continues to be a fundamental sector on which the development of the rest of the sectors depends to a large extent. However, there are many factors that influence its performance, some are dependent on the farmer and others, such as the effects of climate change, do not depend on this. Therefore, crop yields are not always easy to forecast, especially not at an early stage, i.e. before any investment is made for the new cropping season. A case example is presented here in the olive grove in Andalusia, Spain. The objective is to analyze the most influential predictor variables on an early predictive model and the contribution, in this case, of data from satellite images.**

Keywords-olive crop; predictive modeling; multi-source data; satellite imagery.

I. Introduction

Agriculture plays an important role worldwide, but in specific areas, such as the Mediterranean basin, it has a special economic and social contribution. The climate of the countries in the Mediterranean area has conditioned the type of agriculture, mainly centered on vines, cereals and olives. However, there is currently growing concern about the lack of rainfall and abnormally high temperatures, which are determining factors in the fall in agricultural production, especially in the olive sector.

The productivity of olive crops depends on various factors such as soil fertility, weather conditions, the type of tillage carried out and also the amount of last year's crop [1]–[4]. Some of these factors are controllable by the farmer, but others, such as weather conditions, are not.

High olive harvest values are a key objective not only for growers but also for olive mills, distribution and insurance companies, as well as for the food industry. In this sense, it is reasonable to think that strategic decisions in this sector are made on the basis of the expected harvest. Thus, an early season of crop yield prediction, i.e. before planning the investment of resources, is essential in this sector. In the particular case of olive production, it makes it possible to establish market strategies, to know in advance the income of farmers and companies, the number of workers to be hired, or the machinery that will be needed. This information would make it possible to adjust production costs and contribute to environmental sustainability by optimizing the use of phytosanitary products or tillage.

In this sense, a few years ago, crop predictions relied heavily on the experience of farmers, but today much agricultural research focuses on improving crop forecasting. This shift has been complemented by other economic sectors using the predictive potential of Artificial Intelligence (AI) techniques, in particular Machine Learning (ML) and Deep Learning (DL). These techniques are now widely used for crop yield prediction [5]–[8]. However, crop yield prediction is one of the challenging problems in precision agriculture and, as indicated by Xu et al. [9] it is not a trivial task. There are several studies in olive crop yield prediction and most of them are based on the predictive value of pollen emission levels [10]–[14]. These studies take into account basic parameters of pollen levels, as well as other factors such as temperature, rainfall and relative humidity. In the latest study, a regression analysis is performed with results with an error of 0.96% in July. It should be borne in mind that in Spain the harvest is normally harvested between December and January. Therefore, in spite of the latter being a very reliable model the prediction is made after pollination, which occurs very late in the agricultural year, in April, May or June. It is therefore a very reliable model, but it provides a very late prediction, only 5 months before the olive harvest. These results provide useful information for the farmer, to help establish optimal harvesting periods. However, the purpose of our prediction is not that, but to help plan at the beginning of the crop year the investments and resources to be used for ploughing that year.

The importance of early crop prediction is a hot topic in the scientific literature [15], [16]. In the particular case of the olive

 ISBN: 978-1-68558-168-8

grove, most work focuses on the study of variables such as the amount of pollen in the air, vegetative indices that indicate the health of the plant or rainfall maps [11], [12], [17]. However, some of these values are collected in spring, between April and May. This forecasting period is too late to make strategic decisions of the type discussed above, as the investment of resources has already been made at that time. The challenge in this sector is an early forecast in the months of February, just when the previous season's olive harvest has finished.

The aim of this work is to generate an early predictive model of the harvest quantity of olives, our target. For this purpose, ML algorithms are applied using variables from different sources. Although not all the factors that will affect the next harvest are known at this time, it would be advisable to anticipate what the general trend will be if other negative factors do not intervene. In any case, this prediction model will be fed with different variables as the seasons progress in order to better adjust to the time of harvest. The current state of the work is in the analysis of the predictor variables used. At this phase the aim is to determine which are the most influential and to discard those that do not have a significant influence or introduce noise. In this sense, it is observed that the integration of variables from satellite imaginary improves the predictive error so far.

The document is structured as follows. Section 2 describes the methodology carried out and the data processing carried out using Google Earth Engine and QGIS together, as well as the ML algorithms used. Section 3 shows the first results with the different combinations of predictor variables, analyzing how the results improve with the incorporation of satellite image data. Section 4 describes the conclusions and future work to be developed.

II. METHODOLOGY

This section describes each of the phases comprising the workflow followed. The objective in this phase of the state of the research is to analyze how the predictor variables used influence the level of accuracy of the predictive model.

A. Research Area

The study area covers all the municipalities of the province of Jaen, located in Andalusia, southern Spain, between co-´ ordinates 38°N 3°W (WGS84). This province extends over a total area of 13,496 km², characterized by diverse topographies ranging from mountainous areas to wide valleys. In total, it has 97 municipalities (see Figure 1). The climate is Mediterranean, with partial variations depending on the configuration of the terrain and the proximity to the Guadalquivir River, which flows through the province. The average annual temperature varies between 15-17 °C, with marked differences in temperature between day and night, especially in areas close to the river. Rainfall, around 500 mm, is mainly concentrated in the coldest period of the year, although this pattern varies according to the geography of each area.

B. Dataset

At the core of any predictive analysis is reliable data to ensure an accurate prediction. Here, the data comes from official sources and entities, which supports its reliability. Examples of these data sources are: the Area of the Central Registry of Cartography of the National Geographic Institute of Spain which has provided the municipality boundary files, or the State Meteorological Agency (AEMET), dependent on the Spanish Government, meteorological data have been obtained. To achieve accurate results, some pre-processing is often necessary, such as outlier cleaning, data cleaning, categorization and normalization of information, among other steps.

In general terms, the yield of the olive grove depends on the climate, the quality of the soil, the presence of pests, tillage, olive tree varieties and planting density [1]–[4]. However, this study is conducted at municipal aggregation level, where several of these factors are inherently embedded in the harvest data or the municipality's geographical context, making it unnecessary to explicitly insert them into the dataset. For the research, influential data that do not conform to a recognizable yearly pattern are employed as training data. Therefore, environmental and meteorological variables are considered, as well as variables that indicated the state of the crop during the crucial seasons.

The olive tree follows a vegetative cycle composed of two growth stages, where proper care and feeding are crucial to ensure optimal yields after harvest. Olive cultivation is influenced by a variety of factors, some of which are specific to the local environment, such as environmental and weather conditions. Other factors are a direct result of the agricultural practices employed by the farmer on the farm. Within these stages, olive trees go through various phases, including bud break (February-April), flowering (April-May), fertilization and fruit set (May-June), fruit growth (June-September), fruit ripening (October-December) and dormancy (November-February). In this study, these stages are grouped into three seasons, coinciding with the seasons of the year in which the olive tree responds uniformly to the external factors affecting it. Table 1 shows the months covered by each season considered.

The yield of the olive grove depends on the climate, the quality of the soil, the presence of pests, tillage, olive tree varieties and planting density. This study is conducted at municipal aggregation level, where several of these factors are inherently embedded in the harvest data or the municipality's geographical context, making it unnecessary to explicitly insert them into the dataset. It is obvious that meteorological factors are decisive for crop production. Taking into account the level of aggregation of the crop yield data used in this research, the data are grouped at municipal level of aggregation and the dates to which the data refer are the seasons indicated in Table 1. In short, the predictive variables used are:

• Olive crop yield data from each municipality/year. The training data contain a total of eight years. • Weather information from official weather stations. The Regional Government of Andalusia publishes annual olive crop yield data at different levels of aggregation. In this sense, from weather stations are used rainfall data.

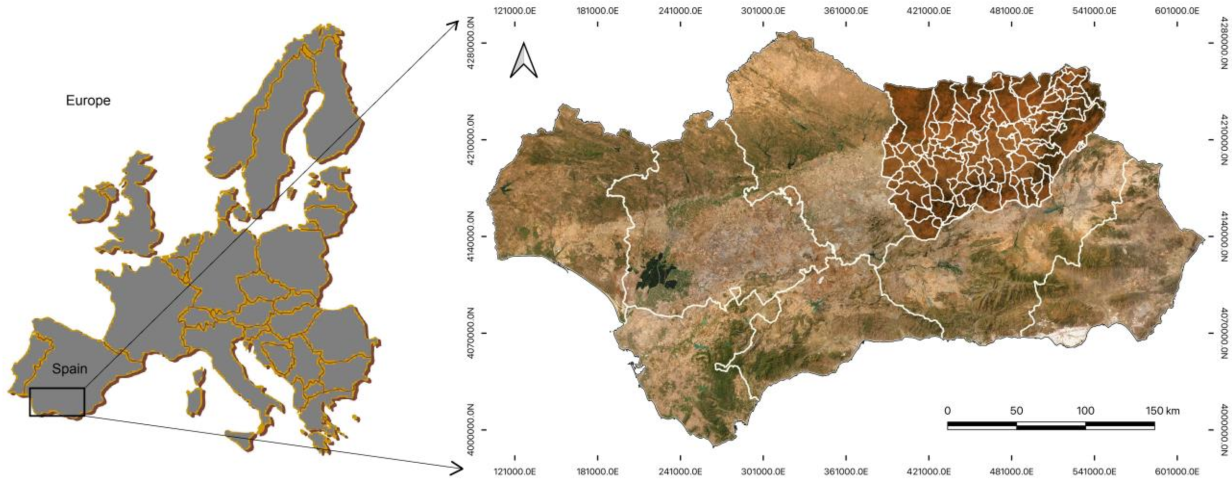

Figure 1. Research area. The left image represents the location of Andalusia in Europe. The right image is the map of Municipalities of Jaen inside Andalusia.´ The coordinates (m) are UTM, zone 30 referred to ETRS89.

TABLE I. SEASON DIVISION

Year	Dates
Winter	01-December to 29-February
Spring	01-March to 30 - June
Summer - Autumn	01-July to 30-November

• Satellite data. Taking into account the information to be obtained, the choice of satellite type becomes a critical decision. Consequently, different satellites are used to acquire data related to different fields, as it is temperature or precipitations.

C. Modeling process

The methodology followed consists of the following phases: (1) The understanding of the data, identify the information it provides, the format and the meaning of the values; (2) Its acquisition and preparation process, Once captured, the data requires prior preparation before being inserted into ML models, some require categorisation, others labelling, detection of outliers, null values, etc.; (3) The generation of the predictive models using different algorithms and, finally, (4) The validation or analysis of the accuracy obtained, in this sense the k-fold cross validation technique was used to evaluate results in statistical analyses. The output data of the model is the amount of olive harvest (the number of Kg of olives). This information is collected for each campaign and for each one of all the province municipalities.

Once the dataset is available, the following step is the detection of data anomalies. This is to identify unusual cases in apparently homogeneous data. A classification algorithm is used for this purpose because these anomalies can be considered as a particular case of classification. Specifically, in this phase, the algorithm used has been the Support Vector Machine (SVM) algorithm. The next phase consists of determining the level of influence of the variables used on the target attribute. In this case, the Minimum Description Length (MDL) algorithm [18]. Finally, regression analysis algorithms are used in the creation of the model. In this study, SVM with Gaussian kernel and Linear kernel, respectively, have been used. SVM regression supports two types of kernels: Gaussian kernel, which is used for non-linear regressions, and the linear kernel, which is suitable for linear regressions. SVMs work effectively on datasets containing many features, even if the number of cases to train the model is very small. Another algorithm used was Neural Networks (NN) which also offers good precision but not bettern than SVM which best fits the prediction in all cases.

III. RESULTS

The first results of this research are the analysis of the contribution of the use of satellite imagery to improve predictive models. We first analyze the results considering only the data from meteorological stations and then we analyze the improvement that the incorporation of satellite images entails.

A. Predictive models using only weather station data

The advantage of using weather stations is that they are daily data from stations maintained by official agencies. However, in this case, the particular disadvantage is the level of aggregation we are using and the distance between stations. Figure 2 shows the network of meteorological stations in Jaen.

Nevertheless, we do this analysis to see to what extent satellite imagery can overcome this drawback Municipalities, which have a station nearby, provide reliable rainfall data.

B. Predictive models using satellite imagery data

Temperature is monitored using MODIS constellation, given that it has spectral bands that allow the detection of surface temperature (see Figure 3). Finally, for rainfall information, we use ECMWF data, based on Copernicus satellite images and Google information. Specifically, ERAS5 (Earth Climate Reanalysis) project.

 ISBN: 978-1-68558-168-8

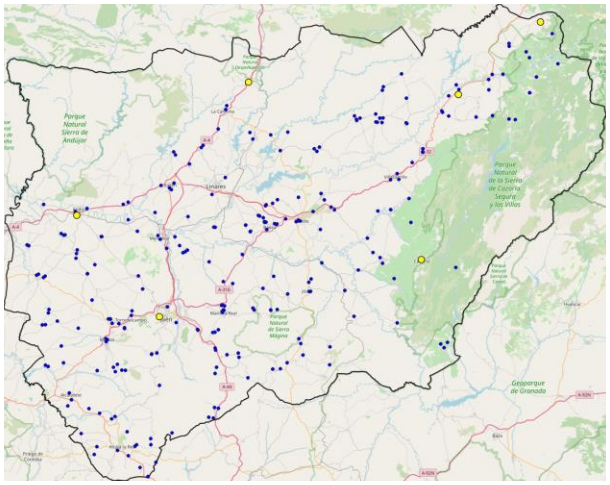

Figure 2. Location of the network of meteorological stations in the province of Jaen (yellow dots) and the oil cooperatives (blue dots).

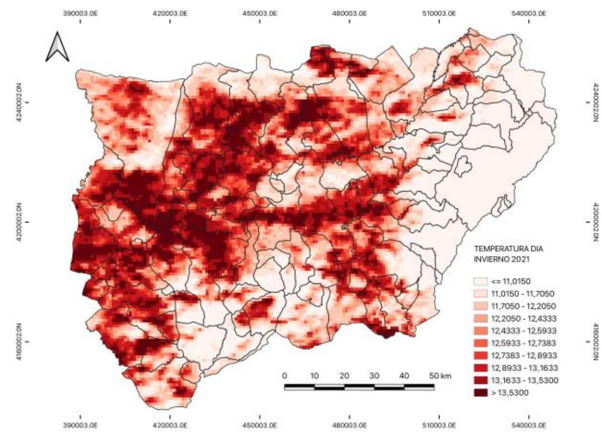

Figure 3. Map of temperatures in the province of Jaen calculated from ECMWF data.

TABLE II. PREDICTION MODELS

Predictor variables	Weather stations			Satellite imagery		
Algorithms	*SVM*		*NN*	*SVM*		*NN*
	Gaussian	*Lineal*		*Gaussian*	*Lineal*	
Only Rainfall	23.56%	24.29%	25.67%	24.62%	25.70%	27.07%
Rainfall + Temperature	23.63%	24.20%	25.42%	24.58%	25.48%	26.26%

Table II shows the mean absolute errors of the predictive models using different combinations of variables. These first tests confirm that, in general, data from weather stations can be substituted by satellite data in those areas far away from a station.

IV. CONCLUSIONS AND FUTURE WORKS

This study presents a workflow methodology describing the steps followed in the analysis of the predictive calculation of olive crop yield at an early stage. The novelty of the work is how early, within the agricultural year, it makes the crop prediction. Just before the investment of resources begins, when there is still no visible or measurable sign of pollen or the beginning of fruit. The correct selection of the predictor variables and the quality of these variables are fundamental. The integration of satellite imagery into the model improves crop yield prediction. This is due to the fact that a better diagnosis of the state of the satellite imagery can overcome this drawback weather next to the area studied contributes to a good early prediction of its production. In this sense, satellite images have been fundamental in order to have sufficient temporality covering all the municipalities in the province of Jaen. The methodology developed is applicable to future works at different spatial scales, even at local or farm detail level. The short-term future development of this work includes the generation of an intelligent system in which the variables obtained from the satellite images are extracted automatically as well as the downloading of the meteorological values from web services. Thus, a non-expert user will be able to use the system by simply inserting the harvest values of the farm or the area under study. The short-term future development of this work includes the generation of an intelligent system in which the variables obtained from the satellite images are extracted automatically as well as the downloading of the meteorological values from web services. Thus, a non-expert user will be able to use the system by simply inserting the harvest values of the farm or the area under study.

ACKNOWLEDGMENT

This research has been partially funded through the research projects PREDIC I-GOPO-JA-20-0006, cofinanced with the European Union FEDER, and the Junta de Andalucia funds, also the projects PID2021-126339OB-I00 and TED2021132120B-I00, funded by the Spanish Ministry of Science and Innovation.

REFERENCES

[1] C. Silveira, A. Almeida, and A. C. Ribeiro, “How Can a Changing Climate Influence the Productivity of Traditional Olive Orchards? Regression Analysis Applied to a Local Case Study in Portugal,” *Climate*, vol. 11, no. 6, p. 123, Jun. 2023, [Online]. Available: https://www.mdpi.com/2225-1154/11/6/123

[2] P. Deiana et al., “Effect of pedoclimatic variables on analytical and organoleptic characteristics in olive fruit and virgin olive oil,” *European Journal of Agronomy*, vol. 148, 2023.

[3] A. Awan and A. Rab, “Influence of agro-climatic conditions on fruit yield and oil content of olive cultivars,” *Pakistan Journal of Agricultural Sciences*, vol. 51, no. 3, pp. 625–632, 2014.

[4] S. Lavee and M. Wodner, “The effect of yield, harvest time and fruit size on the oil content in fruits of irrigated olive trees (*Olea europaea*), cvs. Barnea and Manzanillo,” *Scientia Horticulturae*, vol. 99, no. 3, pp. 267–277, Feb. 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304423803001006

[5] S. Archana and P. S. Kumar, “A Survey on Deep Learning Based Crop Yield Prediction,” *Nature environment and pollution technology*, vol. 22, no. 2, pp. 579–592, 2023.

[6] R. Beulah, “A Survey on Different Data Mining Techniques for Crop Yield Prediction,” *International Journal of Computer Sciences and Engineering*, vol. 7, no. 1, pp. 738–744, Jan. 2019.

[7] S. Iniyan, V. A. Varma, and C. T. Naidu, “Crop yield prediction using machine learning techniques,” *Advances in Engineering Software*, vol. 175, p. 103326, 2023.

[8] P. Saini and B. Nagpal, “Empirical Survey Analysis For Crop Yield Prediction & Identification Of Factors Affecting Yield Gaps,” *Journal of Pharmaceutical Negative Results*, vol. 13, pp. 1318–1329, 2022.

[9] X. Xu et al., “Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China,” *Ecological Indicators*, vol. 101, pp. 943–953, Jun. 2019.

[10] P. Filippi et al., “An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning,” *Precision Agriculture*, vol. 20, no. 5, pp. 1015–1029, Oct. 2019 [Online]. Available: https://doi.org/10.1007/s11119-018-09628-4

[11] C. Galán, L. Vázquez, H. García Mozo, and E. Domínguez, “Forecasting olive (Olea europaea) crop yield based on pollen emission,” *Field Crops Research*, vol. 86, no. 1, pp. 43–51, Feb. 2004.

[12] M. Fornaciari, L. Pieroni, F. Orlandi, and B. Romano, “A new approach to consider the pollen variable in forecasting yield models,” *Economic Botany*, vol. 56, no. 1, p. 66, Jan. 2002.

[13] J. Oteros et al., “Better prediction of Mediterranean olive production using´ pollen-based models,” *Agronomy for Sustainable Development*, vol. 34, no. 3, pp. 685–694, 2014.

[14] F. Aguilera and L. Ruiz-Valenzuela, “A new aerobiological indicator to optimize the prediction of the olive crop yield in intensive farming areas of southern Spain,” *Agricultural and Forest Meteorology*, vol. 271, pp. 207–213, Jun. 2019.

[15] L. Ortenzi et al., “Early Estimation of Olive Production from Light Drone Orthophoto, through Canopy Radius,” *DRONES*, vol. 5, no. 4, Dec. 2021.

[16] J. J. Cubillas, M. I. Ramos, J. M. Jurado, and F. R. Feito, “A Machine Learning Model for Early Prediction of Crop Yield, Nested in a Web Application in the Cloud: A Case Study in an Olive Grove in Southern Spain,” *Agriculture*, vol. 12, no. 9, p. 1345, Aug. 2022.

[17] L. Achmakh et al., “Airborne pollen of Olea europaea L. in Tetouan (NW Morocco): heat requirements and forecasts,” *Aerobiologia*, vol. 31, no. 2, pp. 191 – 199, 2015.

[18] P. D. Grunwald, J. I. Myung, and M. A. Pitt, “*Advances in Minimum Description Length: Theory and Applications*, ser. Neural Information Processing series, M. I. Jordan and T. G. Dietterich, Eds. Cambridge, MA, USA: A Bradford Book, Feb. 2005.

 ISBN: 978-1-68558-168-8

City Planning Dynamics: A Theoretical Framework of Urban Development Scenario of Belgium Provinces Using Logit Based Cellular Automata

Anasua Chakraborty
Department of ArGEnCo, University of Liege, Liege, Belgium
e-mail: a.chakraborty@uliege.be

Ahmed Mustafa
The New School University, Urban Systems Lab New York, US
e-mail: a.mustafa@newschool.edu

Jacques Teller
Department of ArGEnCo, University of Liege, Liege, Belgium
e-mail: jacques.teller@uliege.be

***Abstract*— The idea of compact city encourages a sustainable future in urban planning. Haphazard urban development affects the agricultural landscape of a country, degrades environmental qualities or leads to habitat fragmentation. Urban densification or development in already existing built up areas can help control such adversities. Addressing these challenges through the lens of compact city principles presents an intriguing avenue for sustainable urban development. Our study presents a theoretical framework as a solution to understand, analyze and implement urban densification in three provinces of Belgium: Brussels, Flemish Brabant, and Wallonia Brabant. Additionally, it talks about the future work based on regional planning scenarios for efficient land take. We use urban built up maps for three years 2000, 2010, and 2020 and then classified them into three density classes of 1 (low density built up), 2 (medium density built up), and 3 (high density built up). Using a genetic algorithm well known for its advanced single objective functionality, we optimized the neighborhood parameters for each density class built up. These are used in modelling urban densification from classes 1-2 and classes 2-3, using a multinomial logistic based cellular automata. Our findings show that while Brussels experiences high density with 80% of the region covered with residential built up, Wallonia still undergoes fragmented growth or urban expansion. Thus, our study can help the urban planners, researchers and stakeholders propagate an idea of smart solutions towards city planning.**

Keywords- urban densification; genetic algorithm; cellular automata; city planning; scenario modelling.

I. INTRODUCTION

One of the defining characteristics of the twenty-first century is the rapid urbanization that offers perspectives on sustainable development, as well as challenges [1], [2]. Urban growth frequently results in strained resources, degraded natural habitats, and encroachment into agricultural land, all of which calls for creative approaches to urban planning. Compact cities, which prioritize densification inside already existing urban areas, are becoming popular as a strategy to deal with these issues and to advance sustainable urban development [3].

Belgium, with its unique blend of urban and rural landscapes, serves as an ideal context for exploring the potential of urban densification strategies. A significant majority of Europeans reside in urban areas, with Belgium boasting an exceptionally high urbanization rate of 98% [4]. Particularly in the Wallonia region, located in the southern part of Belgium, urbanization proceeds at a rapid pace, expanding by approximately 17 km^2 annually [5]. This growth is anticipated to continue, fueled by demographic forecasts indicating a projected rise of 200,000 households from 2011 to 2026 [6]. Consequently, this trend is expected to lead to a further expansion of impermeable surfaces in the region.

The motivation behind this research stems from the pressing need to reconcile urban growth with environmental conservation, accessibility and socio-economic development [7], [8]. We aim to address the following key objectives:

- Mapping the spatio-temporal trend of demand for urban built up areas.
- Simulating futuristic urban densification scenarios.
- Analyzing the idea of densification as a pragmatic solution for 2050 no net land take.

Hence, we present a theoretical framework of approaching these objectives, and help planners and regional administration better understand the evolution of land use pattern and design policies for a smart and sustainable city planning. While our theoretical framework provides insights into urban densification, its real-world implementation may face challenges related to political, economical and social factors that influence decision making in urban planning. It is important to consider this as a limitation when formulating policies based on our findings.

Following the introduction, the rest of the paper proceeds to briefly explain the ideated methodology of the current ongoing research. In Section II, the area of interest is delineated, providing the rationale for our study area. Section III outlines the framework of the data and methodology used. Section IV includes the conclusion, summary of the main findings of the study and the highlights of its contributions to the field. Finally, Section V outlines avenues for future research, suggesting potential directions for further investigation, thus encapsulating the comprehensive narrative of the research endeavor.

II. AREA OF INTEREST

Belgium, as shown in Figure 1, is the 6th most urbanized country in Europe and the 22nd most densely populated country in the world, with 381 inhabitants per km^2 [9] The country is governed by three different federal bodies, making it a crucial area for studying the intricacies of different planning and policies.

On the other hand, there has been a growth of 17% in buildings for Flemish and Wallonia region. The growth in number of dwellings - to accommodate the growing population, has been alarming since 1995, at a rate of 31% in Flemish region and 29% in Wallonia. To address these challenges, policies and strategies aligned with the principles of "No Net Land Take" have gained prominence[10]Using this concept, the cross-border provincial region of Brabant can mitigate the adverse effects of urbanization by 2050. Furthermore, by identifying and strengthening centralities, urban planners and policymakers can foster compact, vibrant and well-connected urban environments.

Figure 1. Study Area

III. Methods

A. Data framework

Though, many studies used Land Use Land Cover (LULC) maps, the quality of data and geometric error can lead to erroneous processing and results. In our study, we use cadastral data provided by Belgium's land registry. This is a vector-based data which contains useful information like "construction year". Since vector data is highly computationally inefficient, we rasterize it for big data processing. Similarly, we prepare the factors impacting the built up development under geophysical, accessibility and socio-economic category, as shown in Table 1. All of these maps are rasterized at 100 × 100m or 1 hectare as it is a commonly used scale for LULC studies. The model was calibrated using 2000-2010 data and was then validated using 2020.

Hereto preparing the dataset, it is observed that our dataset comprises of static and dynamic factors. But, the most dynamic variable which needed to be calibrated for scenario modelling was "Neighborhood". This will be discussed in the below sub-section.

TABLE 1. LIST OF EXPLANATORY VARIABLES

Factors	Name
X1	Elevation (DEM)
X2	Slope
X3	Distance to highways
X4	Distance to primary roads
X5	Distance to secondary roads
X6	Distance to residential roads
X7	Distance to railway stations
X8	Distance to large-sized cities
X9	Distance to med-sized cities
X10	Employment rate
X11	Population
X12	Zoning

B. Modelling framework

We propose a hybrid modelling framework, where the independent variables are calibrated using MultiNomial Logistic (MNL) regression model [8], [11], [12]. On the other hand, the Cellular Automata (CA) model has been used to calibrate the neighborhood interaction based on a push-pull effect. This means the probability of a cell transitioning to a particular land use depends on its weightage and its distance from the cell of interest (where the weights are optimized using genetic algorithms). This technique was chosen because of its ability to incorporate both - the explicit and implicit nature of urban development. While similar models have been used to study land use land cover changes, there have been few applications that take into account different built up density. Additionally, the genetic algorithm has been proven to simulate futuristic scenarios more precisely than other commonly used models [13]. Our results are expected to produce three scenario conditions based on the planning policy of regional development authorities.

- In the first scenario, we will simulate a futuristic scenario until 2050 based on the historic trends of urban built up derived from calibration. The calibration was done based on years 2000 and 2010, to simulate the year 2020 and was validated with the observed built up map of 2020.
- In the second scenario, we will simulate a futuristic scenario until 2050 for development across different density class of low-density built up to medium

 ISBN: 978-1-68558-168-8

density built up, and medium density built up to high density built up. This is based on the idea of promoting gradual decrease in expansion to avoid overwhelming effect on land take.

- In the third scenario, we will simulate a futuristic scenario until 2050 using the centralities of a city as a promoting factor, while still a controlled part of the growth will be encouraged in the periphery of the city to maintain a balance between the economy and environment.

These studies are based on planning ideas of Belgium as a country and should be adapted contextually. However, it can aid to the foundation of "smart" planning especially for countries of global south, where unplanned growth is still prevalent.

IV. Results and Discussion

Urban development in Belgium presents complex challenges that require innovative approaches to city planning and management[6]. This research presents a theoretical framework for modelling urban growth scenarios by utilizing the power of MNL based CA modelling. Three scenarios present divergent ideas for the cross-border scenario of Brabant's future development: BAU, Stop au béton (STOP), and Centralities. BAU anticipates urban growth as a continuum of the past pattern of growth. In our results, BAU shows that urban expansion continues to grow rapidly, especially in Wallonia, which predominantly experienced the same over the years. On the other hand, STOP assumes that new urban development should be stopped beyond 2050, which means the growing population should accommodate the need for housing in existing built areas. Drawing inspiration from a centrality-focused approach, the third scenario aims to promote compact, integrated urban centers by 2040 and reduce a considerable amount of growth outside of these central zones. The implementation of these scenarios underscores the need for proactive and strategic land use policies tailored to the unique characteristics across borders. The STOP scenario highlights how critical it is to prioritize densification in order to stop urban sprawl and protect parks, green areas, and historical sites. On the other hand, in order to improve connection, accessibility, and livability for citizens and companies, the scenario of centralities emphasizes the possible advantages of supporting functioning urban areas and centralities.

In order to ensure that the model and its scenarios remain relevant and successful in guiding land use policy and urban planning activities in Belgium, a further study and collaboration are needed. The incorporation of both simulated and actual built up density maps into risk models facilitates the development of targeted policies, improved resource allocation, and a greater ability for policymakers and stakeholders to prioritize mitigation actions related to urban growth[11], [13], [14]. Stakeholders may collaborate to build prosperous, inclusive, and sustainable communities that improve everyone's quality of life by utilizing data-driven insights and proactive tactics.

V. Conclusion and Future Work

In this paper, we propose a framework for a modelling approach to understand and analyze the built up development type and its implications to the policy of no net land take. We have considered three scenarios for evaluating this. While BAU scenario shows that urban expansion still exists, STOP scenario assumes that this new development should be replaced by "rethinking" built up development in the form of densification, consequently slowly bringing it to zero land take. However, to address the complexity of planning and policy, the scenarios of centralities will still allow a 75-25 ratio of growth in central part of cities to its outskirts.

This study can lay a foundation for sustainable urban planning, although uncertainties are a part of futuristic simulation. This is why it is important to take into account the challenges of modelling perturbation, and to understand the difference between "What it is" and "What it should be". This will drive our futuristic research in a more realistic and accurate direction.

Acknowledgment

This research was funded by the INTER program and co-funded by the Fond National de la Recherche, Luxembourg (FNR) and the Fund for Scientific Research-FNRS, Belgium (F.R.S—FNRS), T.0233.20— 'Sustainable Residential Densification' project (SusDens, 2020–2023).

References

[1] B. C. Pijanowski, K. T. Alexandridis, and D. Müller, "Modelling urbanization patterns in two diverse regions of the world," *J Land Use Sci*, vol. 1, no. 2–4, pp. 83–108, Dec. 2006, doi: 10.1080/17474230601058310.

[2] B. Cohen, "Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability," *Technol Soc*, vol. 28, no. 1–2, pp. 63–80, Jan. 2006, doi: 10.1016/J.TECHSOC.2005.10.005.

[3] A. H. Allawi and H. M. J. Al-Jazaeri, "A new approach towards the sustainability of urban-rural integration: The development strategy for central villages in the Abbasiya District of Iraq using GIS techniques," *Regional Sustainability*, vol. 4, no. 1, pp. 28–43, Mar. 2023, doi: 10.1016/J.REGSUS.2023.02.004.

[4] "World Bank Open Data | Data." Accessed: Mar. 23, 2024. [Online]. Available: https://data.worldbank.org/

[5] N. Stephenne, L. Poelmans, E. Hallot, B. Beaumont, and I. Uljee, "From Regional LULC Model to Urban Population Density Simulation in Wallonia", doi: 10.2312/udmv.20161422.

[6] "A Common Land-Use Change Model for Both the Walloon and Flanders Regions in Belgium." Accessed: Mar. 23, 2024. [Online]. Available: https://www.researchgate.net/publication/357226412_A_Common_Land-

 ISBN: 978-1-68558-168-8

Use_Change_Model_for_Both_the_Walloon_and_Flanders_Regions_in_Belgium

[7] A. Chakraborty, A. Mustafa, H. Omrani, and J. Teller, "A Framework to Probe Uncertainties in Urban Cellular Automata Modelling Using a Novel Framework of Multilevel Density Approach: A Case Study for Wallonia Region, Belgium," *Urban Book Series*, vol. Part F270, pp. 325–341, 2023, doi: 10.1007/978-3-031-31746-0_17/FIGURES/4.

[8] A. Mustafa, I. Saadi, M. Cools, and J. Teller, "Understanding urban development types and drivers in Wallonia: A multi-density approach," *International Journal of Business Intelligence and Data Mining*, vol. 13, no. 1–3, pp. 309–330, 2018, doi: 10.1504/IJBIDM.2018.088434.

[9] "Statbel." Accessed: Dec. 20, 2023. [Online]. Available: https://statbel.fgov.be/en/about-statbel

[10] A. Chakraborty, H. Omrani, and J. Teller, "A Comparative Analysis of Drivers Impacting Urban Densification for Cross Regional Scenarios in Brussels Metropolitan Area," *Land 2022, Vol. 11, Page 2291*, vol. 11, no. 12, p. 2291, Dec. 2022, doi: 10.3390/LAND11122291.

[11] A. Mustafa, A. Van Rompaey, M. Cools, I. Saadi, and J. Teller, "Addressing the determinants of built-up expansion and densification processes at the regional scale," *https://doi.org/10.1177/0042098017749176*, vol. 55, no. 15, pp. 3279–3298, Feb. 2018, doi: 10.1177/0042098017749176.

[12] L. Poelmans and A. Van Rompaey, "Complexity and performance of urban expansion models," *Comput Environ Urban Syst*, vol. 34, no. 1, pp. 17–27, Jan. 2010, doi: 10.1016/J.COMPENVURBSYS.2009.06.001.

[13] A. Mustafa, A. Heppenstall, H. Omrani, I. Saadi, M. Cools, and J. Teller, "Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm," *Comput Environ Urban Syst*, vol. 67, pp. 147–156,Jan.2018, doi: 10.1016/J.COMPENVURBSYS.2017.09.009.

[14] L. Poelmans and A. Van Rompaey, "Detecting and modelling spatial patterns of urban sprawl in highly fragmented areas: A case study in the Flanders–Brussels region," *Landsc Urban Plan*, vol. 93, no. 1, pp. 10–19, Oct. 2009, doi: 10.1016/J.LANDURBPLAN.2009.05.018.

Mopsify: Gamified Spatial Crowdsourcing for Content Creation in Location-based Games

Nancy Fazal
School of Computing
University of Easter Finland
Joensuu, Finland
fazal@cs.uef.fi

Pasi Fränti
School of Computing
University of Eastern Finland
Joensuu, Finland
franti@cs.uef.fi

***Abstract*— Location-based games require players to step outdoors to interact with real-world objects. The gameplay depends on the player's current physical location, and it is crucial to have the content available everywhere. Global scalability of the content is a huge challenge, as many games find it difficult to scale worldwide. Even players of the successful games like Pokémon Go complain about the lack of locations in certain regions. Different approaches have been studied for content creation including Web crawling, Open Street Map data, and social media data. Crowdsourcing has been successfully used but only with established games with rich community support. In this paper, we present *Mopsify*, a location-based game prototype which combines the power of crowdsourcing, gamification, and Open Street Map data for content creation. The content generated as part of the *Mopsify* gameplay shows the potential of this approach. We tested it in 15 locations in three games in Joensuu City. Only 4 of the locations had issues with accessibility, as they happened to be inside other buildings, which were closed already by the time gameplay happened. The other 11 were classified as good for their place name, location from OSM, and photo captured by the player. The players gave positive feedback and were interested in enriching the content by using smaller, interesting objects in the gameplay instead of restricting them to only visible landmarks.**

Keywords: Location-based games; Spatial Crowdsourcing; Gamification; Open Street Map; Content Creation.

I. INTRODUCTION

Location-Based Games (LBGs) have been around since 2000 and are a subgenre of pervasive games. The common feature of these games is that they use the player's physical location to generate a gameplay. The first commercial game was *Botfighters*, in which players take control of a robot to destroy other robots [1]. *Shadow Cities* was another early game created by a Finnish company Grey Area. The game uses the player's physical GPS location on an in-game map to battle other players.

LBGs saw immense popularity when Niantic launched Pokémon Go. The game is free-to-play and has over 100 million downloads and about 14 million reviews on the Google Play Store. It received mixed reviews such as appreciation for the concept but criticism for the technical problems, causing public annoyance and contributing to accidents. Despite all the critics and challenges, Pokémon Go in-app purchase revenue surpassed 645.59 million dollars worldwide in 2022. The success of Pokémon Go made it evident that LBGs offer exciting commercial potential [2]. Since then, several companies have attempted to recreate the phenomena with their games and have achieved various levels of success [3].

One challenging factor for LBGs is content creation as it requires visiting outdoors, collecting game content, and attaching it to real-world locations [4]. Even a simple treasure hunt game would require many locations worldwide. As a result, many games hand over the content creation responsibility to the game administrators or the players themselves followed by a review mechanism.

Pokémon Go uses a collaboratively built database of locations spread worldwide. The players, however, still complain about the lack of locations in many places, despite community support and success. Other games such as *Hidden Lion* [5]; *See It* [6]; *Museum Scrabble* [7] and *O-Munaciedd* [8] have been built around a specific location and cannot be easily transferred to other places as they rely on specific stories and knowledge. Manual content creation around the globe is a hard problem, and as a result, many LBGs find it difficult to survive over long periods.

According to Laato et al. [11], the literature has shown that games using crowdsourcing procedures lead to geographic bias and so far, mainly web-based crowdsourcing has been used for content creation. With the advancement of mobile technologies, traditional web-based crowdsourcing such as *Amazon Mechanical Turks* (AMT), *Crowd4U*, and *Upwork* has now extended to mobile and spatial crowdsourcing such as *Uber*, *Waze*, *Open Street Map* (OSM), *GigWalk*, *Field Agent* and *TaskRabbit* [12].

The major contrast between web-based and spatial crowdsourcing is that the latter requires volunteers to move in the real world to solve tasks using their mobile devices [13]. Typical spatial crowdsourcing tasks include capturing photos, traffic checks, store audits, groceries, and location-aware surveys [14]. The major issue with crowdsourcing is to keep the interest of participants alive [15].

Martella et al. [16] present gamification as a practical solution by leveraging user's intrinsic motivations. Morschheuser et al. [17] investigate how gamification elements in crowdsourcing tasks can increase the motivation and participation of the crowdsources. Their empirical literature review revealed that gamification is an effective approach for improved crowdsourcing participation. Gamification boosts the intrinsic motivation of contributors to participate in an activity [12].

Bowser et al. [18] report the experiences in evaluating *Biotracker*, a gamified location-based mobile application for

 ISBN: 978-1-68558-168-8

citizen science. Their findings suggest that people who use *Biotracker* are more motivated by the gamification element in the form of "*competing with peers*" and "*earning badges*". Some users' feedback on *Biotracker* exhibited that the gamification makes it fun and less tedious to participate in citizen science.

Laato and Tregel [11] state that submissions in Niantic require the players to make decisions on what to submit, writing the proper name and description, placement in the real world, and capturing a proper photo for it. The submission is then followed by a review mechanism. During the review, players also have some influence when they rate the submissions and can mutually decide to either accept or reject it or even move it from its original location. The process increases the work of players and reviewers. It further allows only players who have reached a certain level in the gameplay to make submissions and reviews.

In this paper, we present *Mopsify*, a location-based game prototype that combines the power of gamified spatial crowdsourcing and OSM data for content creation in LBGs. The data from OSM is highly reliable for its location accuracy and place names, which would eventually lessen the burden of contributors and reviewers. Capturing photos as part of the gameplay reduces the chance of attaching copyrighted, promotional, or vandalized photos. *Mopsify* aims to evoke intrinsic motivations in players to step out on an adventure of randomly generated real-world locations and show off their photography skills.

The remainder of the paper is structured as follows: Section II presents the content creation approaches in LBGs. In Section III, we present Mopsify, a LBG prototype to accelerate and overcome the limitations of existing content creation processes. In Section IV, we discuss the games played, players experience and quality of the content generated. Finally, Section V presents the concluding remarks and future work.

II. CONTENT CREATION IN LOCATION-BASED GAMES

O-Mopsi [19] content creation has mostly been handled manually by game administrators and players which as a result limited its applicability within Finland and a few other European countries only. Various approaches such as Web Crawling [9], OSM [4], and Social Media Data [10] have been studied for content creation. Their results remained modest and only a fraction of the 6845 retrieved images from web crawling contained geotags. OSM provided a better approach, but the challenge was that the data itself lacked geotagged images. However, the data sometimes included external web links from which the desired content was found 21% of the time. OSM data, however, is greatly underrepresented in Asian countries and good results were limited to the urban or downtown areas only. The data from social media services has varying strengths for different aspects of the content. Flickr had the highest number of images, representative of the place. Yelp and Foursquare always had the correct location. The place names were accepted as it is except for Flickr where relevant place names had to be extracted using an external method called Tag-tag [20].

Pokémon Go relies on the web-based crowdsourcing alone and one must reach a certain level before they can make any contribution to the Niantic Wayfarer [2]. Pokémon GO requires the players to reach level 37+ and Ingress requires the level 10+. The nominations are later reviewed by mature players for their positional accuracy (using maps, street view, or human judgment), safety, photo representativeness, appropriate title, description, and possible duplicate. Niantic's acceptance criteria for submissions [23] [24] and O-Mopsi have similar design principles for the Point of Interest (POI) [21] classified as attractiveness, accessibility, location clarity, identifiability, and lifetime. In such a case, the content contributed by players is of varying quality and requires the proper inspection from game administrators or mature players.

The review mechanism is usually a time-consuming process and might take several weeks or months. Besides reviewing the photo quality, its title/description, category of a nomination, verifying its location accuracy is also extremely important. A reviewer can use satellite views and street views to help in determining its real-world location. In cases, where the nomination happens to be inside the park or under a tree, the reviewer must use their best judgment to decide where it could exist in the real world by looking at clues in the submitted photo background [24].

III. MOPSIFY

Mopsify is a single player location-based game prototype, which aims to combine the gamified spatial crowdsourcing and OSM data to support the content creation. Players can create new games in the area of their own choice containing a desired radius and the number of locations they want to visit. The randomly generated locations are extracted from the OSM database.

A. Motivation

Mopsify aims to simplify the content creation process by obtaining POIs automatically from OSM. Since OSM does not provide pictures itself, *Mopsify* players are encouraged to visit the selected locations and take photos as a part of its gameplay. OSM data is usually reliable for its positional accuracy and textual names, and therefore, reduces the work of both reviewers and contributors. OSM is a leading crowdsourced spatial database, published under the Open Database License. Due to its free and open availability, several games have utilized it without any legal restrictions [22]. Pokémon Go uses it for its in-game map.

Mopsify allows any player to act as a contributor. The location data from OSM removes the need for contributors to select POIs and write descriptions. Instead, they can fully focus on taking pictures and enjoy visiting random locations. Photos are very crucial elements in Pokémon Go as well as O-Mopsi and must therefore be checked if they represent the POI [23]. Mopsify prompts the players to take a photo of a place right away when they arrive at the POI, thus, reduces the possibility of uploading copyrighted edited or vandalized photos.

 ISBN: 978-1-68558-168-8

LBGs enforce a set of safety and security restrictions like omitting areas on a private property, company grounds, military areas as well as locations without pedestrian access like highways. Laato et al. [11] construct three tag groups for OSM data. The first group called *strict* corresponds to the Niantic's criteria. The second group called *relevant*, represents the locations with shops, relevant societal buildings, or leisure areas. The third group called *notable* includes tags for better coverage in suburban and rural areas. Mopsify, however, only uses the randomly generated POIs with the only exception of excluding *hospitals*, *weighbridges*, *atm*, *nightclubs*, *fire stations*, *banks*, *prisons*, *crematoriums*, *graveyards*, *waste disposals*, *recycling*, *toilets*, *wastebaskets* and *parking*.

B. Gameplay

The gameplay (see Fig. 1) begins once a player has successfully logged in using their existing account or entered as a guest. A player is then shown brief instructions before starting the game. Next, the player can specify the game name, the number of locations they want to visit, and the desired radius from their automatically identified location on the map. For the given player preferences, random locations are generated from the OSM database. A player can shuffle the locations as many times as needed and can then freely choose the location wanted to visit first. Once selected, the location is highlighted on the map. A player can now go to the location and his/her arrival is notified by playing a brief *ta-da* sound once he/she reaches within a 20m proximity to the selected location.

The player is now prompted to capture a photo of the location and submit it when ready. She can move around to find a suitable location and angle to take the best possible photo. Once the photo is submitted to the database, the visited location is removed from the map. A player then chooses another unvisited location and continues her tour unless has visited all the locations. Once the game is finished, the player is shown all the locations and images captured. She is then asked to give feedback on how well each of the visited locations meets the five design principles [21] by rating them on a scale from 1 (lowest) to 5 (highest). In addition, she can submit general comments on the gameplay if interested. Having such information from the player would eventually help to construct better OSM data retrieval queries as per the player's preferences.

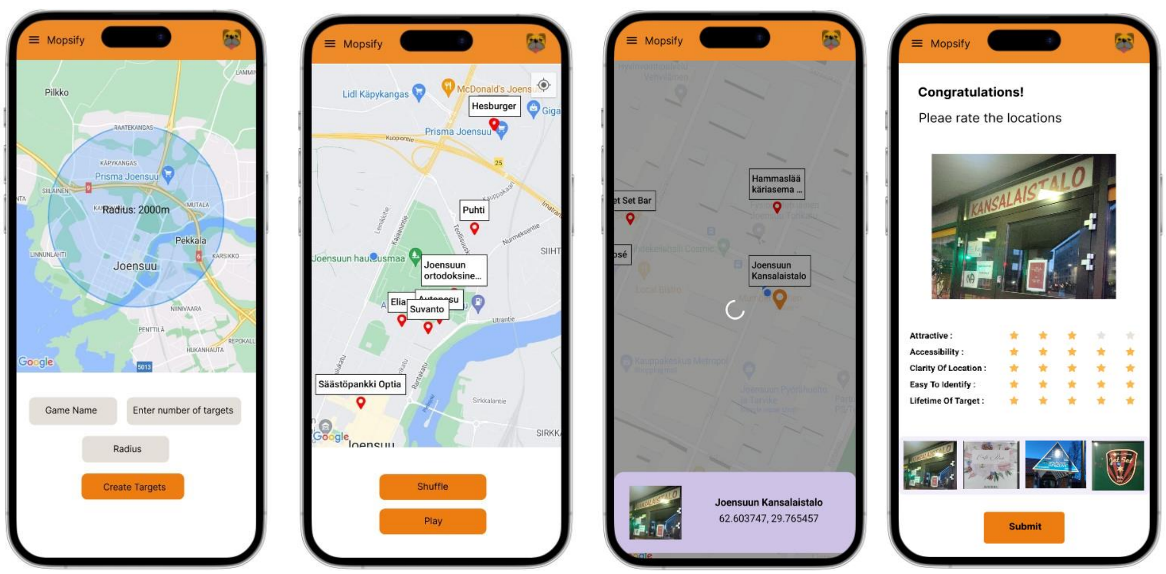

Figure 1. Mopsify gameplay.

IV. EXPERIMENTS

Three small games within the city of Joensuu, Finland, were played for data collection purposes. The games are named *Kanervala*, *Keskusta*, and *Kaislakatu*. The estimated game length varied from 2 to 3 km with an average number of 5 locations to visit. The content generated was examined for its quality by the authors, who are local experts. Besides the images captured, other location details such as POI type, original location, photo taken location, image EXIF data, device, network information, and accuracy are also recorded. Feedback from each location using the five design principles is additionally stored for analysis purposes.

Two female players who are residents of Joensuu played the games and provided the content shown in Tables I, II, and III. One player had good experience of playing LBGs, whereas the other had never played any such game before. Both players are software developers by profession. They gave good feedback on the gameplay, mentioning that they discovered places which they have often passed by but never noticed. One player talked about coming across a pizza shop called *Pizza Ella* which is about 53 meters from their workplace, and they were not aware of its existence.

The player showed excitement about this finding and ended up buying a pizza from there. The other player, however, complained about coming across places such as *schools* to take photos as they were CCTV monitored. Parents were picking up their kids by that time and the player hesitated about capturing a photo.

A player also reported that the game notified of her arrival at the location, she could not see herself. In such cases, the player just took a random photo to let the gameplay proceed. It happened with the restaurant *Amica* which the player could not find and ended up taking a photo of another restaurant nearby, *Kreeta*, to be able to continue the gameplay. A similar case was identified with *Punainen Risti – lajittelukeskus* and restaurant *Elisa* for which the player took a wrong photo of Hyvinvointi- ja kauneushoitola (a beauty salon). Outdated locations were also reported. The dentist shop *Hammaslääkäriasema Otso* is permanently closed even if still present in the OSM database. As a result, the player struggled to find it and ended up taking the wrong photo of Pihlajalinna (medical center) to let the game proceed. The content was later reviewed by the authors who concluded that such cases were mostly inside buildings that were closed at the time of gameplay. These cases are highlighted in red in Tables I, II and III.

In general, for places with good accessibility, both players provided positive comments on how they learned a bit more about their surroundings and taking photos was an enjoyable activity for them. Playing the game did not feel tedious but rather made their evening walks adventurous. Players also showed interest in having small objects in the gameplay such as *statues*, *benches*, *fountains*, *trees*, and *bird towers*.

TABLE I. CONTENT GENERATED FOR THE GAME "KANERVALA"

POI					
Name	Elias	Muistelupaikka	Heinosen Leipomo Oy	ABC Kanervala	Joensuun vapaaseurakunta
Type	restaurant	place of worship	cafe	fuel	place of worship
Design Principles Feedback (*Attractive*, *Accessibility*, *Clarity of Location*, *Easy to identify*, *Lifetime*)					
	(0,0,0,0,0)	(5,3,3,3,5)	(2,5,5,5,5)	(1,5,5,5,5)	(5,5,5,5,5)

TABLE II. CONTENT GENERATED FOR THE GAME "KESKUSTA"

POI					
Name	Joensuu Kansalaistalo	Amica	Hammaslääkäriasema Otso	Jet Set Bar	Café Rosé
Type	Community centre	restaurant	dentist	pub	cafe
Design Principles Feedback (*Attractive*, *Accessibility*, *Clarity of Location*, *Easy to identify*, *Lifetime*)					
	(3,5,5,5,5)	(0,0,0,0,0)	(0,0,0,0,0)	(4,5,5,5,5)	(5,5,5,5,5)

TABLE III. CONTENT GENERATED FOR THE GAME "KAISLAKATU"

POI					
Name	Joen Eläinlääkäriasema	Kanervalan koulu	Punainen Risti - lajittelukeskus	Joensuun Nuorisoverstas ry	Pizza Ella
Type	veterinary	school	social centre	social centre	fast food
Design Principles Feedback (*Attractive*, *Accessibility*, *Clarity of Location*, *Easy to identify*, *Lifetime*)					
	(3,5,5,5,5)	(1,1,5,5,5)	(0,0,0,0,0)	(1,5,5,5,5)	(5,5,5,5,5)

 ISBN: 978-1-68558-168-8

V. CONCLUSION

Mopsify is a location-based game prototype for content creation in location-based games. It combines gamified spatial crowdsourcing with Open Street Map data to ease the work of both contributors and reviewers. The OSM data is highly reliable for its positional accuracy and place names. We made a small-scale experiment of the app in Joensuu, Finland with three games. The results were promising and support the potential of this approach. The games length ranged from 2 to 3 km with 5 locations in each. In total, 15 locations were visited. Four of the locations were either not accessible or had been closed permanently. The content generated was reviewed by the authors and they found that 11 valid locations had correct place name, and location from OSM. Their images captured by players were also representative. Players mostly expressed positive feedback and showed interest in having *statues*, *benches*, and other interesting small objects in the gameplay. Places such as schools, fuel stations, community centers and veterinary were reportedly less attractive for the players.

In future, we aim to extend the gameplay as per players feedback, do extensive testing, and work on handling the accessibility issues. OSM data has been reportedly biased in rural areas, thus finding ways to overcome this constraint should also be addressed as part of the *Mopsify* gameplay. The data generated from *Mopsify* can be used in any location-based game and sight-seeing applications.

REFERENCES

[1] J. T. P. N Jacob and A. F. Coelho, Issues in the development of location-based games. *International Journal of Computer Games Technology,* 2011.
[2] S. Benford et al., Coping with uncertainty in a location-based game, IEEE *pervasive computing*, *2*(3), pp. 34-41, 2003.
[3] S. Laato, T. Pietarinen, S. Rauti, M. Paloheimo, N. Inaba, and E. Sutinen, A review of location-based games: Do they all support exercise, social interaction and cartographical training? *CSEDU (1)*, pp. 616-627, 2019.
[4] N. Fazal, R. Mariescu-Istodor, and P. Fränti, Using open street map for content creation in location-based games, *29th Conference of Open Innovations Association (FRUCT)*, pp. 109-117, Tampere, Finland, 2021.
[5] K. P. Chang et al., HIDDEN LION: A location-based app game of sword lion searching. ACM SIGCHI annual symposium on Computer-human interaction in play, pp. 323-326, October 2014.
[6] C. Neustaedter and T. K. Judge, See it: a scalable location-based game for promoting physical activity. ACM 2012 Conference on Computer Supported Cooperative Work Companion, pp. 235-238, 2012.
[7] C. Sintoris, A. Stoica, I. Papadimitriou, N. Yiannoutsou, V. Komis, and N. Avouris. MuseumScrabble: Design of a mobile game for children's interaction with a digitally augmented cultural space. Social and organizational impacts of emerging mobile devices: Evaluating use, pp. 124-142. IGI Global, 2012.
[8] D. La Guardia, M. Arrigo, and O. Di Giuseppe, A location-based serious game to learn about the culture. Int. Conf. on The Future of Education (2nd edition), pp. 508-512, 2012.
[9] N. Fazal, K. Q. Nguyen, and P. Fränti, "Efficiency of Web Crawling for Geotagged Image Retrieval", *Webology*, vol. 16, 2019.
[10] N. Fazal, P. Fränti, "Social media data for content creation in location-based games". (Submitted)
[11] S. Laato, and T. Tregel, Into the Unown: Improving location-based gamified crowdsourcing solutions for geo data gathering. *Entertainment Computing*, vol. *46*, 2023.
[12] Y. Tong, L. Chen, and C. Shahabi, Spatial crowdsourcing: Challenges, techniques, and applications. *Proceedings of the VLDB Endowment*, vol. 10, issue. 12, pp. 1988-1991, 2017.
[13] Y. Zhao, and Q. Han. Spatial crowdsourcing: current state and future directions. *IEEE Communications Magazine*, vol. 54, issue. 7, pp. 102-107, 2016.
[14] S. R. B. Gummidi, X. Xie, and T. B. Pedersen. A survey of spatial crowdsourcing. *ACM Trans. on Database Systems (TODS)*, vol. 44, issue. 2, pp. 1-46, 2019.
[15] F. Alt, A. S. Shirazi, A. Schmidt, U. Kramer, and Z. Nawaz, Location-based crowdsourcing: extending crowdsourcing to the real world. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pp. 13-22, October 2010.
[16] R. Martella, C. Kray, and E. Clementini, A gamification framework for volunteered geographic information. *AGILE: Geographic Information Science as an Enabler of Smarter Cities and Communities*, pp. 73-89, 2015.
[17] B. Morschheuser, J. Hamari, and J. Koivisto, Gamification in crowdsourcing: a review. In *2016 49th Hawaii international conference on system sciences (HICSS)*, pp. 4375-4384, IEEE, 2016.
[18] A. Bowser et al., Using gamification to inspire new citizen science volunteers. In *Proceedings of the first international conference on gameful design, research, and applications*, pp. 18-25, October 2013.
[19] P. Fränti, R. Mariescu-Istosdor, and L. Sengupta, O-Mopsi, Mobile orienteering game for sightseeing, exercising, and education. *ACM Trans. on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 13, issue. 4, pp. 1-25, 2017.
[20] N. Fazal, P. Fränti, Relevant Tag Extraction Based on Image Visual Content, In: Huang, DS., Premaratne, P., Yuan, C. (eds) Applied Intelligence. ICAI, 2023. *Communications in Computer and Information Science*, vol. 2015. Springer, Singapore, 2023.
[21] P. Fränti and N. Fazal, Design Principles for Content Creation in Location-based Games, *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 19, issue. 5, pp. 1-30, 2023.
[22] Games-OpenStreetMap Wiki [Online]. Available from: https://wiki.openstreetmap.org/wiki/Games#Games_using_OpenStreetMap_data 2024-05-20
[23] Photo Guidelines - Wayfarer Help Center [Online]. Available from: https://niantic.helpshift.com/hc/en/21-wayfarer/faq/2769-photo-guidelines/ 2024-05-20
[24] Contributing to the Pokémon GO Map — Pokémon GO Help Center [Online]. Available from: https://niantic.helpshift.com/hc/en/6-pokemon-go/faq/41-contributing-to-the-pokemon-go-map-1694119494/ 2024-05-20