Ref. Ares(2023)6231821 - 14/09/2023

Open Source Packages and APIS

Deliverable number: (D.4.9)
Author(s): Dimitrios Rizopoulos, Ioanna Fergadiotou, Gerasimos Kouloumbis
Author'(s') affiliation (Partner short name): INLE

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 861598. LEAD is a project under the CIVITAS Initiative. Read more - civitas.eu

THE CIVITAS INITIATIVE IS CO-FUNDED BY THE EUROPEAN UNION

Deliverable No.		**4.9.**	
Workpackage No.	**4**	Workpackage Title	**LEAD Adoption and Growth Roadmap**
Task No.	**4.6**	Task Title	**Exploitation Plan – Open Source Packages and APIs**
Date of preparation of this version:		**8/9/2023**	
Authors:		**Dimitrios Rizopoulos (INLE)** **Ioanna Fergadiotou (INLE)** **Gerasimos Kouloumbis (INLE)** **Anastasios Kakouris (INLE)** **Thanos Karydis (INLE)** **Beatriz Royo (ZLC)** **Rodrigo Tapia (TUD)** **Sebastian Horl (IRTX)** **Abdelhadi Belfadel (IRTX)** **Carla De Oliveira Leite Nascimento (MOLDE)** **Aletta Büki (BKK)** **Paulo Reis (SONAE)**	
Status (F: final; D: draft; RD: revised draft):		**F**	
File Name:		**D4.9 Open Source Packages and APIs v1.docx**	
Version:		**1.1**	
Task start date and duration		**01/09/2021 – 30/9/2023**	

This document is issued within the frame and for the purpose of the LEAD project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 861598.

Revision History

Version No.	Date	Details
0.1	07/09/2022	1st draft version, including some initial open outputs.
0.2	9/9/2022	2nd draft version, including 15 open outputs as communicated with partners, missing links to GitHub and licensing schemes for re-use.
0.3	27/9/2022	A more detailed explanation of models and smaller refinements after exploitation management team internal meeting.
0.4	17/3/2022	After communications with LEAD partners (i.e., open outcome owners), actions were taken to enhance visibility and searchability and add licensing options. Relative discussion is included in the deliverable for each open output.
0.5	30/6/2023	Open outputs update after confirming open outputs for LEAD LL4, LL5, and LL6, compared to the previous version.
1	1/8/2023	Formatting revisions and open outputs revision based on partner feedback.
1.1	8/9/2023	Final revisions after LEAD project internal review.

Reviewers List

Name	Company	Date	Signature
Bas Groothedde	ARG	29/08/2023	
Sebastian Horl	ISX	31/08/2023	

Executive Summary

LEAD's D4.9 Open Source Packages and APIs catalogues the LEAD project's open outcomes that have been developed during the LEAD project and have the potential to be re-utilized by external stakeholders of the LEAD project. The open outputs of the LEAD project are classified into three main categories, open digital models, open datasets, and open APIs & software packages. All open outputs, as included in this report, have the potential to be re-used for several reasons. Some examples are the following:

- *i)* academic and applied research purposes,
- *ii)* software development (Internet-Communication Technology (ICT) Platforms, data pipelines, data processing infrastructure, simulation environments, tracking and monitoring systems),
- *iii)* Transport & Logistics system modelling and development of decision support tools,
- *iv)* Consulting services,
- *v)* Data science and analytics applications,
- *vi)* Extraction of domain-specific insights through analysing open demand and survey results in datasets.

Based on these main re-use examples, this deliverable includes the summary of open results of LEAD, all under the umbrella of a single report, including references to the open platforms that they are hosted, respective documentation, and deliverables in which the open output is described in detail. The work done under Subtask 4.6.4 and this report D4.9 adds to the work done by open output owners by providing them with specific suggestions on improving the searchability, licensing, and visibility of the outputs on the online platforms. The report also includes a reference to how the collection of open outputs took place, guaranteeing a proportionality of open outputs (i.e., fifteen in number, seven open digital models, seven open datasets, one software API package) of LEAD to the whole body of work done in the project, as well as standards that the publishing framework (i.e., GDPR, FAIR, etc.) abides by. To supplement the re-usability and adoption of LEAD open outputs by stakeholders, and to highlight their reuse value in the field of last-mile logistics, use cases and re-use examples are given for each type of output.

Contents

List of Figures

List of tables

Acronyms

Acronyms	Explanation
API	Application Programming Interface
B2B	Business to business
B2C	Business to consumer
CLOGIN	CityLogin (Project partner)
COPERT	EU standard vehicle emissions calculator
DT	Digital Twin
EC	European Commission
EMT	Empresa Municipal de Transportes de Madrid (Project Partner)
EU	European Union
ICT	Internet-Communication Technologies
INLE	Inlecom (Project partner)
IRTX	IRT SystemX (Project partner)
KER	Key Exploitable Result
KPIs	Key Performance Indicators
LL	Living Lab
LMT	Last Mile Team (Project partner)
LSP	Logistic Service Provider
STAR	Sustainability Assessment of Mobility Projects
T&L	Transportation and Logistics
UCC	Urban consolidation centre
UPM	Universidad Politécnica de Madrid – Madrid Polytechnic University (Project partner)
WP	Work Package
ZLC	Zaragoza Logistic Centre (Project Partner)

1 Introduction

1.1 LEAD project overview

LEAD has created Digital Twins of urban logistics networks in six cities to support experimentation and decision-making with on-demand logistics operations in a public-private urban setting. Innovative solutions for city logistics have been represented by a set of value case scenarios that address the requirements of the on-demand economy while aligning competing interests and creating value for all different stakeholders. Each value case has combined several measures (LEAD strategies): a) innovative business models, b) agile urban freight storage and last-mile distribution schemes, c) low-emissions, automated, electric, or hybrid delivery vehicles, and d) smart logistics solutions.

Cost, environmental and operational efficiencies for value cases have been measured in 6 Living Labs (LL). Evidence-proven value cases and associated logistics solutions will be delivered in the form of exploitable Digital Twins, incorporating the models that support adaptation to different contexts and that provide incentives for PPPs.

The LEAD consortium comprises 22 partners, all of whom are involved in the Living Labs, supported by five international partners for knowledge transfer. This structure has incentivized the co-creation of solutions by city authorities, logistics industry leaders, start-ups, and research experts in freight modelling, complex simulation, and logistics optimization.

1.2 Deliverable purpose and scope

This report is prepared in the context of LEAD's WP4 – LEAD Adoption and Growth Roadmap that aims to discuss and promote the adoption and exploitation of LEAD project outcomes. The purpose of this document D4.9 – Open Source Packages and APIs is to specifically present the main open outcomes (i.e., outputs, terms used interchangeably) as they have been collected by the exploitation manager, according to LEAD project partners' input. This report is public and will include a series of open tools that have been produced or extended under the LEAD project. The goal of making these tools available on online open platforms is that they are expected to be re-used by researchers, software developers, data scientists, and other practitioners and organizations within the Transportation & Logistics (T&L) sector (and especially within last-mile logistics). The three main categories of open outputs are open software packages & APIs, open digital models, and open datasets. The re-use options for software engineers, researchers, SMEs, and other interested types of organizations are also described in this deliverable through representative examples.

1.3 Target audience

The main target audience is the international research community, as well as practitioners in the field of T&L, with a particular focus on the last-mile logistic area. The (mainly) technological tools that are described in this report can be re-used for various reasons, from scientific method development to software development or the derivation of domain-specific insights for organizations in the sector, such as Logistics Service Providers (LSP, LSPs), municipalities and public authorities.

1.4 Matching between GA requirements and deliverable content

This report is conducted in the context of WP4 and the Exploitation management activities of LEAD project that are described in the following task and respective deliverables:

- T4.6 Exploitation Plan and Open Source Packages and APIs, and deliverables:
    - D4.8 Exploitation plan – Due M31 (Dec 2022)
    - D4.9 Open Source Packages and APIs – Due M36 (September 2023)

While Task 4.6 includes four overall subtasks this report D4.9 is especially focused on the work conducted under subtask 4.6.4 Open Source Packages and APIs. The other three subtasks and the work conducted there has been described in report D4.8 Exploitation plan (market assessment, technology monitoring, SWOT analysis, LEAD's Outcomes Registry and Business Planning of Key Exploitable Results (KERs)), which has been submitted in Month 31 of the project and is currently under review. In more detail, in Table 1 below, one can notice the matching between the GA requirements under subtasks 4.6.4 and the contents of the current report.

LEAD GA Requirements	Document section & Justification
ST4.6.4 Open Source Packages: This subtask will deliver open source packages along with open sample datasets and user guides for development teams, in a format ready to be used by external city stakeholders and businesses for further exploitation. This will also support SMEs and entrepreneurs to boost their business and convent research challenges into business opportunities. Sample datasets, models, APIs specification and documentation will be uploaded in GitHub and EC Open data portal in M30.	Section 3.3 of this report includes a listing of LEAD project open outputs with link to their open repositories. Section 3.3.1 includes open digital models of LEAD project, Section 3.3.2 includes open datasets and Section 3.3.3 includes open software tools (incl. APIs). Section 3.2 includes re-usability examples per output type included in the report (open digital models, open datasets, open software tools). For each open output included in the report, its re-use value is given alongside potential stakeholders that may be interested to use it after the lifetime of LEAD.
	Collection and publication methodology as well as supporting activities towards project partners (i.e., open outputs owners) are described in Section 2 of the report.

Table 1: Deliverable's adherence to LEAD objectives and Work Plan

1.5 Document structure

With respect to the sections above, the deliverable is organized into four main sections, starting with Section 1, the introduction, and Section 2, which presents the overall open outputs collection & publishing approach for LEAD open assets (i.e., open outputs). In Section 3, the main listing of the open outputs of the LEAD project is given, alongside their description and main expected re-use examples after the end of the LEAD project. Finally, in Section 4, the concluding remarks of this report are given.

2 LEAD Open Innovation Process - Open Assets

2.1 Communication steps & Collection approach

The goal of Subtask 4.6.4 and Deliverable report D4.9 is to collect and produce reusable and open assets within LEAD project. For this we have followed an open innovation process reassuring communication, collection and publishing of open assets (i.e., outputs). To achieve this goal several communication steps have been performed with partners to re-assure that all open outputs are effectively presented within this deliverable, that can in turn serve as a supporting tool to the re-use of assets produced within LEAD. As an integral step to this process, INLE's team has internally discussed and analysed the contents of this report. The main criteria have been the maximization of ease-of-reuse of the outputs as well as to provide a strong overall framework that communicates to external stakeholders the importance of these outcomes. Central to these first steps has been on the one side to collect partner input and provide guidance to those who need it to publish their open outputs.

To achieve those, three perspectives had to be included: the first is knowledge about the scientific methods used within the LEAD project, the second is technical knowledge about the software components that have been developed within LEAD, and the third is the overall requirements (i.e., searchability and findability, licensing, privacy rules, documentation) for the publication process. With those three perspectives in mind, the exploitation manager has decided upon the suitable templates (as included in Annex A) that provided the initial information for each of the open outputs. Four of these templates have been derived, and although they share several similar fields among them, they were made specific to the type of open output shared. The first template concerns open digital models, the second concerns open datasets, the third concerns open software modules, and the fourth concern open APIs. Although APIs are, in essence interfaces for software modules, it has been decided that a separate template focused specifically on this type of software is developed (in the case that partners want to publish APIs). For that reason, in the context of this report, outcomes collected with the APIs and software packages template are described in a common Section 3.3.3. Open digital models are described in Section 3.3.1 and open datasets are described in Section 3.3.2.

Based on these templates, as a next step, a request has been launched to the model owners of the project. Their role has been very central to both the digital models produced within LEAD as well as the Digital Twin platform. So is the case with this subtask and deliverable report as well. The model owners have been the starting point for the exploitation manager's search for open outcomes/outputs within the project. In more detail, the reason for that is that the digital models developed within LEAD are at the centre of the Digital Twin platform and the simulation and validation activities that have been performed through it. Thus, in our request for open outputs in the LEAD project, it has been considered reasonable to assume that most of the project outcomes that are of the three types described above have been developed to facilitate the efficient use of their models. In that regard, the templates included in Annex A have been individually sent to them starting on Apr. 13, 2022 (after a presentation to the regular model owners' meeting) and completed in August 2023, after the finalization of open outputs details and respective communications. In all this communication process, except for requesting information about the uploading process, the open output of basic information, guaranteeing searchability and findability, licensing, privacy, and security concerns, INLE has supported partners with an overall overseeing of the process and addressing their questions regarding the uploading of open outputs to several well-established platforms (GitHub, Zenodo, GitLab, etc.).

Finally, it is worth mentioning that the exploitation manager has communicated the need for including open outcomes in this report at several locations by addressing the issue at consortium meetings. More specifically presentations were held about Subtask 4.6.4 at LEAD's June 2022, December 2022, March 2023 and June 2023 Consortium meetings. Also, based on individual partners progress with uploading their open outputs, the exploitation manager has checked online repositories of each of the open outputs included in this report, and personalized support has been sent to each outcome owner during the first quarter of 2023, before the submission of this report.

2.2 Compliance with GDPR, FAIR, and the LEAD's Data management plan

Overall, in the process of collecting and publishing open outputs of LEAD projects, it has been very crucial to guarantee that EC standards about Ethics, Privacy, and Security have been guaranteed. For that reason, several feedback processes have taken place as a part of this Task and deliverable in order to guarantee that all partners have taken the proper steps in the publishing process. In the next few subsections, some of the major compliance frameworks that have been taken into consideration are discussed.

2.2.1 GDPR

The General Data Protection Regulation (GDPR) imposes several privacy and security laws, and although it has been drafted and passed by the European Union (EU), it imposes these restrictions on organizations anywhere, as long as the data that they collect and process concern EU citizens.

Several dimensions of the legislation have been guaranteed in the publication process of information in the context of deliverable and Task 4.6.4. Personal data and information that are connected to specific individual citizens have been fully anonymized, and any references that can indicate any specific individual have been left out. This anonymization does not only concern names and locations of any individual but several other aspects such as gender, ethnicity, web cookies, and political opinions. Although sensitive personal data have not been at the epicentre of any of the open outputs of the LEAD project, all open output owners have been asked to remove any references to specific individuals or data fields that may indicate any specific individual's preferences or behaviour.

2.2.2 FAIR

As also discussed in LEAD project Data Management Plan (DMP), LEAD project adheres to the FAIR data management principles that are focused on four pillars:

- Findable data: All data within the project will be documented and stored in such a manner that they are efficiently searchable. In the case of open outputs of the LEAD project, this translates into incorporating the modern, suitable nomenclature across output titles, descriptions, and guides, as they have been included in this deliverable and the platforms. As a very central step in the publication process, LEAD's open outputs have been published on well-established web platforms and are well-adopted by professionals globally. The main focus has been GitHub and Zenodo platforms, which are both very well received, especially by software engineers and researchers. Finally, as a last note, all open outputs owners have been instructed on how to include a keywords section on the platforms where the open output is uploaded. In this way, LEAD's output are expected to be search engine optimized and rank as high as possible in relevant internet search engines.
- Accessible data: All data generated by the project that will be made available as open outputs shall be available through the LEAD website. A link to this report, including all the links to open outputs will be included in the LEAD project, once this report is finalized, peer-reviewed and accepted.
- Interoperable data: Data generated and published by the LEAD project are formatted in such a way so that they can be re-used by modern computer systems and software libraries. Taking a step further, some of the LEAD datasets (where possible) include all the necessary information for the reproduction and extension of academic and applied research studies that have been developed by the project.
- Re-usable data: As mainly concerned, with Exploitation subtask 4.6.4 and the current deliverable D4.9, all data that have been either collected or produced and will be made openly available have been guaranteed to be made available only after the relevant data owners and Data Protection Officers (DPO, DPOs) have been informed.

2.2.3 Open source software and open outputs licensing

The importance of open-source software and open datasets has been well-known among communities of software developers for several decades [1], [2] but it is only in the past few years that its importance has been well-received by public administrations [3], [4]. Open outputs, as they are described in this deliverable, are either software (i.e., open digital models and open software packages) or open datasets. In accordance with European Commission's (EC) Open Source Software Strategy 2020-2023 [5] and EC's Vision for open-source software:

"The Commission leverages the transformative, innovative, and collaborative power of the open source, encouraging the sharing and re-use of software solutions, knowledge and expertise, to deliver better European services that enrich society and focus on lowering costs to that society."

The LEAD project has identified a licensing scheme for the re-use, adaptation, and overall exploitation of each outcome. In that context, all open output owners have been requested to provide a licensing scheme for the re-use of their outcomes that is not restrictive, allows re-use, and abides by EC's definition of open-source, which is:

"Open source software (or free software) combines copyright and a license to grant users the freedom to run the software, to study and modify it, and share the code and modifications with others. It facilitates collaboration, innovation, and agility."

Similarly, the LEAD project and the procedures followed within this deliverable abide by the rules regarding the licensing of open datasets and the overall strategic direction given by the EU to make openly available all research data that emerge from EC-funded research projects and are becoming more and more openly available in online portals [6].

2.2.4 LEAD's data management plan

The publication of open datasets of the LEAD project has also been considered and abided by the rules established in the LEAD D5.4 Data Management Plan (DMP). The DMP provides the general guidelines for the handling and utilization of datasets within the LEAD project, as well as procedures for the open sharing of data. Some of the steps for compliance with the DMP include the request to dataset owners to make sure and confirm that datasets published online are at their final versions, are up to date, that proper metadata exists and that an expiration date exists for the dataset. Another action has been to communicate with each dataset owner and make sure that their Data Protection Officer (DPO) has to review and confirm open data publication actions.

On the data security side, it has been made sure that all LEAD open datasets that have been made publicly available have done so by being uploaded to secure and trustworthy web platforms, in which credibility can be given to LEAD and output owners. Although the datasets are open, and thus anyone can have access to them and download them, it has to be made sure that the websites/web platforms hosting the files will not allow 3rd parties to modify the metadata without that fact being communicated to the LEAD partner who is a data owner. Finally, as stated in LEAD's DMP, all open datasets have been examined for their coverage of ethical aspects and partners have been suggested to consider this aspect. In cases that any information was included in the dataset where sensitive personal information may be exposed, LEAD partners, who are open dataset owners, have been requested to remove this information from the overall dataset.

2.3 Guarantying the proportionality of shared outputs

One of the main concerns of the exploitation managers of the LEAD has been to achieve proportionality of the open outputs of the project in comparison to the work conducted within the project. In other terms, although some developments and deliverables of the project had to be confidential and have limited access to specific partners, or only for the consortium, some others have been made publicly available.

For that reason, INLE has made sure to contact all partners regarding potential open outputs that they want to make available online to be re-used by researchers, practitioners, and professionals outside LEAD project. In that way, the greatest possible coverage could be achieved regarding the collection and presentation of open outputs of LEAD project.

2.4 Connection to other tasks of LEAD project

This report is connected to the work done generally in Task 4.6, which is LEAD project's "Exploitation Plan - Open Source Packages and APIs". The exploitation manager, as described in report D4.8: Exploitation plan, has applied an exploitation management methodology enabling the accurate cataloguing of all outcomes of the LEAD project as well as their exploitation path planning. Based on this work in D4.8, many of the open outputs included in this report also exist (as-is) within D4.8 or are parts of the wider outcomes as described in D4.8. For example, the 2-echelon digital model included in this report in open digital models is also an outcome within the Outcomes Registry as presented in D4.8 report. Another example is the Lyon open parcel demand dataset that is included in this report, which in D4.8 falls under the umbrella of the "Advances in last-mile network modelling approach in the Lyon LL" outcome.

Also, this report is connected to Task 2.2 and deliverable report D2.2 "Digital Twins Model Library" which includes all the models developed, extended and utilized in LEAD. Most of the open outputs in this report are also part of the LEAD Models Library.

3 LEAD's open outputs

3.1 Types of open outputs

In the context of Subtask 4.6.4 of LEAD project, three types of open outputs have been considered as the basis for the collection and publication of outcomes of the project. The three categories have been:

- Open digital models,
- Open datasets,
- Open software APIs.

Based on these three types, project outputs that can be potentially packaged and uploaded on the platform with the ultimate goal of being re-used have been collected and classified. Specific information to each type of outcome has been collected as well, which can be of benefit to external LEAD stakeholders who aim to access the open outputs and would require an overall guiding text on how to re-use the output.

3.2 Re-usability examples

Re-usability examples can be considered for each of the three types of open outputs, as they are presented in Section 3.1 above. These re-usability examples aim at showcasing the main use cases per open output type so that their overall utility is effectively communicated. The actual listing and description of open outputs start in Section 3.3 below.

3.2.1 Open digital models types and applications areas

Digital models are, in essence, the implementation of scientific modelling methods into a computer program that can “automate” and “apply” the scientific model to real-world situations based on input data. According to LEAD's Deliverables D1.2 and D2.2, the digital models can be categorized as:

- Optimization models that deal with minimization or maximization of a cost-related function,
- Network models that provide routing and allocation to road networks,
- Impact assessment tools to analyse and measure the impact of different policies (including externalities assessment,
- Agent-Based Models (ABMs) that simulate individual travellers (i.e., agents) behaviour and,
- Demand models that focus on obtaining and analysing preferences of agents (including demand synthesis).

Digital models can derive decision support functionality when they are part of the Digital Twin platform or can work as standalone tools. In the context of LEAD project, several models have been utilized in the LLs to validate and test real-world planning scenarios and strategies.

Regarding the first category of models, optimization models are the ones that usually include the maximization or minimization of an objective function over a set of constraints. These types of models are suitable for finding a real-world solution to transportation problems that regard the minimization of travel time or travel distance, as well as problems that include scheduling and resource allocation. For example, the Vehicle Routing Problem (VRP), the Travelling Salesman Problem, or the location/reallocation problem of facilities that can help support a transport network (delivery depots, consolidation centres, cross-docking facilities, charging stations). In the LEAD project, several approaches included routing and optimization modelling such as the case of Lyon LL (using jsprit) and in Porto LL (estimating Electric Delivery Vehicle fleet size based on the VRP).

Regarding network models, these regard models that, simulate the transportation demand on road networks or find optimal routes on transport networks. Such an example is the 2-echelon model, which estimates aggregate demand for last-mile logistics services according to several parameters. The availability of data is a key factor for the reproduction/re-use of this model as well that can be used by Logistics Service Providers (LSP, LSPs) to make decisions they need to take regarding network configuration, fleet size, and other performance metrics.

Impact assessment tools are another case of digital models developed within LEAD that can be re-used after the end of LEAD project. Such models are used to evaluate last-mile logistics after they have been produced or decided upon prior to the application of the assessment tool. For example, UPM's Electric Vehicle (EV) emissions calculator can be used to calculate emissions from the usage of an Electric Delivery Vehicle (EDV) fleet after the careful planning of deliveries and extractions usage metrics (kilometric distances covered, etc.)

Agent-based models are another example of a digital model that can be re-used to extract useful insights for the utilization of transportation networks. At the very centre of the agent-based simulation tool, there is a computational model that simulates the actions and interactions of autonomous agents on the micro-level and derives insights for the system level. Several agent-based models focused on last-mile logistics are introduced in the context of LEAD project. A re-use example is the Hague LL's Parcel Tour Formation/Parcel scheduling model, which takes as input data about the origin of parcel shipment and their destination and then generates a sequence of stops to fulfil them. This model enables a different type of flexibility for modelers that can utilize it as a sub-model for bigger modelling and research tasks. An example of such a wider task would be the solution of a Vehicle Routing Problem that would need to be solved based on clusters that are created based on the consolidation of the parcel in hubs or according to different types of vehicles.

Finally, there are demand synthesis models that can combine some sources of data and produce estimates for transport and delivery demand according to some criteria. Several of those have been introduced in the LEAD project. They can be further re-used if one combines data from different spatial contexts and produce what can be called synthetic parcel deliveries demand for other locations. These types of models are very important because they enable the study of different attributes of last-mile logistics networks in cases where no real datasets are available. An example of LEAD's open outputs can be the utilization of Lyon LL's synthetic parcel delivery demand generator pipeline (as described in Section 3.3.1.8), which can utilize census data, Origin-Destination data, populations data, income tax data, among a few more other datasets to derive a synthetic parcel delivery demand dataset for any other region.

3.2.2 Open dataset re-use example

Open datasets are, in essence, digital files, including multiple tables that include respective information on the dataset subject. The data values have been gathered or produced from various sources, either actual delivery data, surveys, software programs, or scientific methods used to produce synthetic datasets representative of some type of transport or delivery demand. Datasets can usually be used by some software or scientific method to be analysed or provide decision support functionality. Their utilization goes hand in hand with the digital model, or in a broader sense, with analytics applications. The re-use examples will be analysed from the perspective of analytics and the following categories:

1. Descriptive analysis,
2. Diagnostic analysis,
3. Predictive analysis,
4. Prescriptive analysis.

Re-using LEAD's produced data for descriptive analysis can lead to producing metrics and insights from the data that can be re-used for understanding several phenomena that have happened in the past. With this analysis, several measures of frequency, central tendency, and dispersion can be derived, which can mean something for the future steps to take and the overall planning of an organization that is active in last-mile logistics. For example, in the case of LEAD project and the crowd shipping data, they can be studied by organizations that are interested in expanding their services in last-mile logistics towards crowd shipping services to compare the metrics that emerge from the crowd shipping survey data and compare them to the metrics that emerge for their own customer or bringers surveys.

In contrast to descriptive analysis, which is concerned with extracting metrics and insights based on past events, diagnostic analysis can also be used on the data to go a step further into the analysis and discover correlations between different phenomena. This type of analysis brings in the qualitative perspective of things and can combine different (external) data sources that can test or reject hypotheses about correlation and causation. For example, based on LEAD's open outputs and the survey results dataset on the importance of last-mile logistics sustainability criteria, according to experts, a research team can further work on comparing these survey results with the results of their own surveys. In the bigger picture and the global context, this dataset can be the basis for comparing the importance of sustainability criteria across European regions, countries, or continents.

Moving on to predictive analysis, here datasets can be re-used to produce predictive metrics about the future. Time series analysis, data mining, and regression models are only some of the tools that can be applied here in order to extract meaningful quantitative results that, in one way or another, predict the future behaviour of people and systems or approximate demand. An example within the LEAD project is the utilization of the real-world services logistics services from the Madrid LL, which can be leveraged for estimating future demand, especially if one considers external data sources for the economic performance of markets, and the weather, among others.

Finally, LEAD datasets can also be utilized across a series of prescriptive analytics applications and for subsequent analysis. In this case, the dataset is utilized not for understanding a past phenomenon or predicting a future phenomenon, but in this type of analysis, the dataset can be used to describe the best or optimal course of action. In this case, the dataset serves as an input to a Decision Support System or an optimization module, which combines inputs to derive a set of decisions for the optimal planning of a network or of an LSPs operator. An example derived from LEAD's open outputs is the utilization of parcel delivery demand data (as they emerge from several LLs and are independent of whether they are real-world or synthetic) in order to develop a VRP algorithm or a routing algorithm that derives the respective solutions for each real-world problem based on the needs of algorithms' user, either that is an LSP, an individual driver or a team of researchers and consultants. In Figure 1, the open datasets re-use cases are summarized.

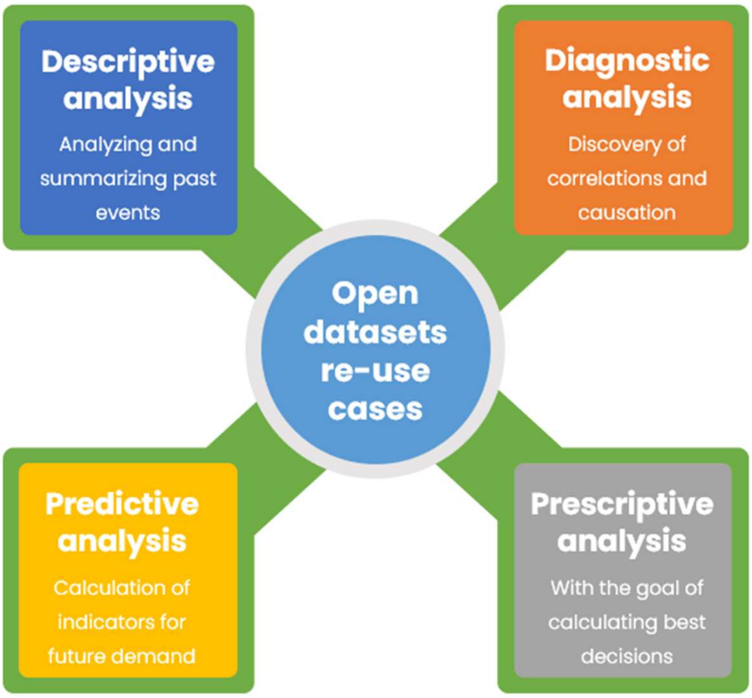

Figure 1: Open datasets re-use cases

3.2.3 Open-source software package and re-use examples

Open-Source software is one most important pillars of innovation in our days, and that is also depicted by the EC's Open Source Software Strategy 2020-2023 [5]. Within LEAD the focus has been on APIs that enable the exchange of data between different software modules to enable complex modelling and computation processes.

Due to the intensive integration efforts conducted to realize the LL's model execution workflows, there have been several software interfaces developed between models. The interfaces/APIs usability within LEAD has been to collect parameters, variables, and results from the execution of one model or platform and provide them to another model or platform in a standardized and automated way. An example of an open output from the LEAD project is the interface between COPERT emissions software, which runs in Window as an executable. The open API made available by LEAD project is expected to enable users of COPERT to make use of the simulation results from other ABM, optimization and network models and, through the open API provide them as input to COPERT. This Open API is expected to further facilitate Digital Twin applications, visualization platforms, or other systems that require the combination of several digital models under common scenarios.

3.2.4 An overview of re-use examples

LEAD's open outputs are independent of each other, meaning that they can be used as standalone tools. Still, they can also be part of new or existing modelling approaches and ICT systems across pilots and Living Labs in Europe. However, these same tools, as included in this report, can be combined to reproduce outcomes (decision support) derived within LEAD for the same or new areas.

Two main examples can be discussed here, with the first being the application of MASS-GT LEAD version by utilizing its four submodules and the Hague LL Parcel Delivery demand dataset, as well as the API to COPERT, to reproduce the agent-based simulations for Hague and estimate effects on the last-mile network as well as CO2 emissions. Based on this reproduction of the simulations for Hague, an individual research, developer or any other organization can adapt MASS-GT to work with other input datasets for other regions under study. Similarly, the Charging Station Facility Location Problem model can be used with the open Porto LL Parcel Delivery Demand dataset to reproduce insights for the Porto area, and then based on this, any dataset can be used by the respective organisation to apply the Charging Station Facility Location Problem model for their area of study. Several other LEAD open outputs can be combined, as included in this report, such as the 2-echelon model with the Madrid LL Parcel delivery demand dataset or the eqasim demand synthesis model with the Lyon LL Parcel delivery demand dataset. A very generic overview of potential re-use opportunities are given in Figure 2.

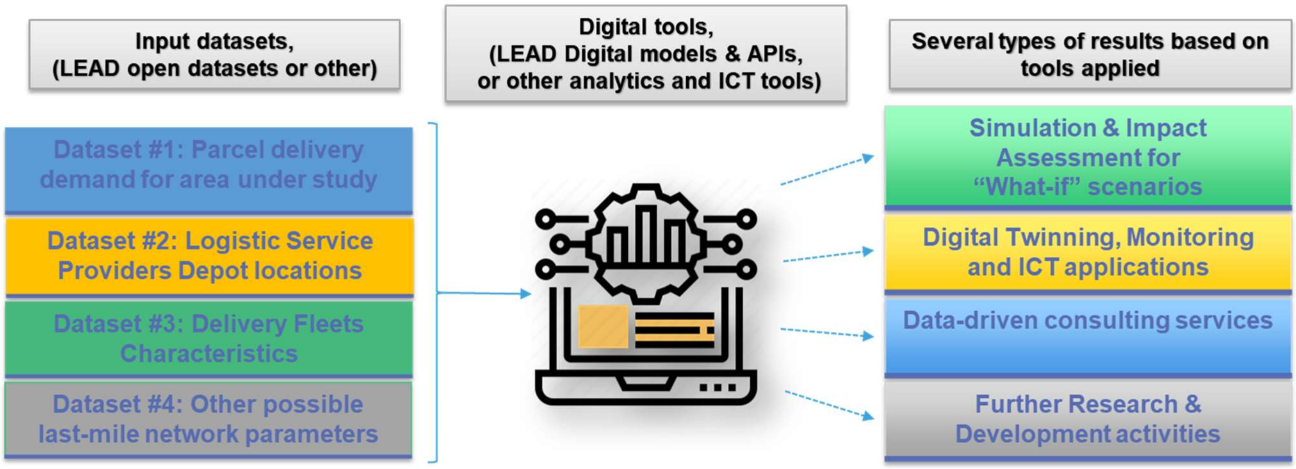

Figure 2: Generic overview of re-use opportunities for LEAD open outputs

The reuse examples, as discussed in Section 3.2, are just a few of the possible examples that are feasible. Given the potential progress in modern R&D and technology sectors, the re-utilization possibilities for LEAD outcomes depend on the last-mile logistics industry needs and how European R&D teams will attempt to support the sector. However, it is important to keep in mind that the re-use of the open outputs of this report requires personnel with high expertise and multidisciplinary teams of scientists that will be required to spend several person months to analyse the outcomes and re-apply them.

3.3 LEAD's Open outputs

In this section, the catalogue of open outputs of the LEAD project is presented. This section is organized based on three sub-sections, each one for each type of open output within LEAD.

All of the listed open outputs are results of the R&D activity within LEAD, and have either been developed (or extended) and used to satisfy project requirements. While their utilization and development within LEAD are described in several other deliverables of the LEAD project, this section focuses exactly on their re-use opportunities by researchers, businesses, and other authorities outside the LEAD project. The goal is not to present again the same details about each outcome/output, but rather showcase how it can be re-used, by whom, and with what goal. For that reason, for each open output listed in the current section, a general description is given, alongside a re-use value explanation. In Figure 3, one can notice the categories of open outputs in LEAD project.

Open outputs in LEAD Project		
Open Digital Models	**Open datasets**	**Open API Software**
2-echelon model	Madrid LL Parcel delivery demand dataset	API to COPERT model
Mass-GT LEAD version	Lyon LL Parcel delivery demand dataset	
parcel demand synthesis pipeline	Hague LL Parcel delivery demand dataset	
eqasim demand synthesis model	Budapest LL Parcel delivery demand dataset	
Charging Station Facility Location Problem	Porto LL Parcel delivery demand dataset	
Oslo LL's CED Model	Madrid LL's Survey results on the sustain. criteria importance	
Oslo LL's EDM Model	Hague LL's Crowd shipping Survey results	

Figure 3: Overall listing of the 15 open outputs described in this deliverable per outputs type (7 open digital models, 7 open datasets, 1 software APIs)

3.3.1 Open digital models

Several open digital models are presented in this section that have been developed within LEAD to facilitate the Digital Twinning process in the several LLs of the project. On a general note, digital T&L models are implementations of scientific methods (modelling, solution approach and analysis) into software, that enables the application of the method by satisfying its computational requirements. These digital models are essential parts of the decision support functionality, and each one focuses on different aspects of the LLs planning and operation. The current section focuses on the latter case of Digital Models that can be re-used as standalone tools with different types of inputs and outputs to serve the needs of stakeholders in last-mile logistics. The open digital models of the LEAD project will be placed when fully developed at the LEAD GitHub Repository at the following link:

- https://github.com/Horizon-LEAD

As discussed in Section 2.3.2 of this report, GitHub has been chosen as the platform for uploading most of the open digital models, and generally open outputs, under the LEAD project since it is the most well-adopted platform for open-source software in the world. Based on this, it is a platform that both allows searching for open outputs that are hosted on the platform and, at the same time, it is very highly ranked on internet search engines (e.g., Google, Bing, Yahoo), which overall leads to the greater exposure possible for LEAD's outputs. GitHub also allows the licensing of its repositories in a standardized way[1], which enabled LEAD Partners to take the necessary licensing measures for their open outputs.

In each of the GitHub repositories, documentation for digital model re-use has been placed based on the re-exploitation requirements for each digital asset. The documentation is either provided as part of the markdown file in the GitHub repository (i.e., readme.md), or as a Microsoft Word document (.docx) included in the repository. An example of the documentation in the readme file is in the case of the COPERT webservice:

- https://github.com/Horizon-LEAD/copert-webservice

and an example of the documentation in the Microsoft Word:

- https://github.com/Horizon-LEAD/2echelon_v2

3.3.1.1 2-echelon digital model

This open tool is an R model for estimating the resources for 1-echelon and 2-echelon distribution last-mile networks. The objective of this model is to provide decision support to actors in last-mile logistics environments and help them compare the impacts of different decisions they need to take regarding network configuration, fleet size, and other performance metrics. Based on the results (number of vehicles, distances, and time), the user can calculate the environmental, operational, and financial impact and examine several "what if" scenarios.

The re-use value of the 2-echelon model is based on the fact that it is a decision support model that can be re-used within various spatial contexts and different urban environments. In essence, organizations that are interested in improving their last-mile logistics networks can utilize this model as is and provide it with their own data to broadly estimate the impact of moving from a 1-echelon network to a 2-echelon network logistics network. In contrast to other models that provide similar evaluations, the 2-echelon model does not require the exact delivery points (as they are described in the parcel deliveries demand dataset) but can base its calculations on Daganzo's formula for calculating approximate distances within the delivery areas.

This model is useful for public authorities that are examining the shift from a 1-echelon model to a 2-echelon model to derive initial estimates for the changes that they would need regarding several parameters of the network, such as the vehicle fleet and distances to the customer and expected outputs such as environmental and financial impact. This model can be especially useful for LSPs that would need to examine these alternative scenarios without collecting detailed data or employing specialized personnel for the development and maintenance of scientific models and other related software. Nonetheless, consultancies can utilize this digital model as a tool of a wider analysis suite that can be utilized for the provision of consulting services to stakeholders in last-mile logistics. Except for its re-use, several possibilities exist for further development of the 2-echelon model, with the main

[1] *Licensing a repository, GitHub, https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/licensing-a-repository*

direction being the examination of going from multiple echelon models to other multiple echelon configurations.

- https://github.com/Horizon-LEAD/2echelon_v2

This digital model is available under MIT license. The searchable keywords that have been added on the GitHub repository are: “transportation-planning”, “network-modelling”, “last-mile”, “2-echelon”. The model is described in greater detail in LEAD’s deliverable D3.3 Madrid Value Case (v1) and D3.4 Madrid Value case (final demonstrator).

3.3.1.2 MASS-GT LEAD version

MASS-GT is a multi-agent simulation model of the logistics decisions in the urban freight context that models urban freight demand on a long-term and short-term tactical level. MASS-GT LEAD version is an adaptation of MASS-GT available at:

- https://github.com/mass-gt

MASS-GT LEAD version consists of 5 sub-models from the model library that are interlinked:

1. Parcel Generation: https://github.com/Horizon-LEAD/Parcel_Generation
2. Parcel Market: https://github.com/Horizon-LEAD/Parcel_Market
3. Parcel Tour formation: https://github.com/Horizon-LEAD/ParcelTourFormation
4. Shipment model: https://github.com/Horizon-LEAD/Shipment_Module
5. Network model: https://github.com/Horizon-LEAD/Network-Assignment

All models are licensed released under GNU General Public License v2.0. The MASS-GT LEAD version is an event-based simulation used for the following logistic processes: 1) sourcing, 2) distribution channel choices, and 3) shipment size & vehicle type. The software script is developed in Python programming language and makes use of several spatial data, mainly in shapefile format.
Specifically, the parcel generation model makes synthetic parcel demand (Origin-Destination pairs) for the study region. The Parcel Market model simulates the fulfilment of the demand: whether it goes through the courier network, parcel locker network or crowd shipping. The parcel tour formation generates tour for each individual logistic agent’s parcels. The shipment model generates demand and tours for all other segments of urban logistics that are not e-commerce. Finally, the network model assigns the tours to the road network.

The main benefit that comes from the re-use software script is that it enables researchers in the Hague area and the Netherlands to use some standardized data sources about the region and produce a synthetic dataset that can be used to realistically study the last-mile logistics network in the area. This software script can be further used by researchers/software developers and respective SMEs and teams in other locations other than Hague, Netherlands. However, in that case, the entity to re-use will have to extend the software and adapt it to read and process different data sources that concern the area of focus. For this reason, and because of the heterogeneity of multiple stakeholders considered, MASS-GT enables the examination of multiple logistics network configurations and setup scenarios.

Its re-use value is focused on the fact that it is a decision support model that accounts for high data granularity and actors heterogeneity that can be adapted to different networks and spatial contexts. As in the case of other digital models within LEAD, MASS-GT can be used by public authorities, LSPs, and other stakeholders in last-mile logistics in order to assess last-mile logistics solutions and compare them to understand trade-offs and advantages/disadvantages between them.

All the models above are openly available under GNU General Public License v2.0. The models developed for the Hague LL are also described in detail in LEAD's Deliverable D3.5, D3.6 and D2.2. The searchable keywords (i.e., GitHub topics) per MASS-GT sub-repository are:

1. Parcel Generation: parcel-synthesis; freight-simulation
2. Parcel Market: parcel-fulfilment; freight-simulation
3. Parcel Tour formation: freight-tour-simulation; freight-simulation
4. Shipment model: shipment-synthesis; freight-simulation
5. Network model: network-assignment; freight-simulation

3.3.1.3 Oslo LL's Discrete Choice Model (DCM)

Oslo LL's DCM is an econometric model that has been implemented in R language and, in essence, is an analytical tool to model agents' preferences, assess adoption/acceptance/reaction, optimize stimuli selection, and identify compromise solutions. This model can be used to investigate stakeholders' preference heterogeneity to forecast their choice behaviour related to some scenario, policymaking, or decision.

The model takes as input the result of the Stated Preferences survey about consumers' decisions towards specific attributes of the last-mile logistics networks and utilizes random utility theory in order to derive quantitative metrics about the systematic utility of some of these attributes. In the context of LEAD and the Oslo LL, the DCM has been used in order to model the preferences of consumers towards alternative delivery configurations based on socioeconomic and environmental aspects.

The model can be-reused by researchers who are interested in analysing similar phenomena in other regions. This modelling approach can be re-used to i) derive the same insights about other cities and the corresponding stakeholders, or ii) it can be used to directly compare results between regions. LSPs, within Norway, can benefit from the results of the application of this DCM and the Stated Preferences survey by analysing the importance and significance of the attributes within the survey for their customers. Finally, LSPs from other locations can reproduce the approach in order to model and estimate the significance of several attributes (environmental footprint, socioeconomic attributes, cost, etc.) of their delivery services and the perception of value by their customers. This model is available at the LEAD project public GitHub repository:

- https://github.com/Horizon-LEAD/DCM_Molde

This module can be re-used under the terms of the MIT license. The model is described in greater detail in LEAD's deliverable D3.11 and 3.12. The output is searchable on GitHub by the following keywords “r”, “econometrics-model” and “choice-model”.

3.3.1.4 Oslo LL's Experimental Design Model

A choice experiment is a survey approach designed to draw consumer preferences based on a hypothetical scenario. This choice is expected by the researcher to occur by the trading of the individual attributes of the different goods available and choosing the good (or alternative) that provides the most utility. Therefore, it can allow to model an individual choice and behaviour and distribution of behaviour patterns of the population. This analytical tool, created using R Studio, will supply the building blocks to design the choice experiments. The results will help in forecasting decisions, suggesting to respondents' questions about their possible choices in hypothetical situations given a specific set of conditions created thanks to experimental design.

The model can be used in the beginning phase of data collection process, when a Stated preference (SP) questionnaire is being used, once the variables and the range of the variables are stipulated. Hence, the model can be re-used with the objective to create a new or re-design a SP questionnaire. This model is available at LEAD project public GitHub repository:

- https://github.com/Horizon-LEAD/EDM_Molde

This module can be re-used under the terms of the MIT license. The model is described in greater detail in LEAD's deliverable D3.11 and 3.12. The output is searchable on GitHub by the following keywords “r”, “experimental-design”, “stated-preference” and “questionnaire”.

3.3.1.5 CS-FLP: Charging Station Facility Location Problem

A digital model under the title of “Charging Station Facility Location Problem (CS-FLP)” has been developed specifically for the Porto LL to address research questions related to the living labs’ first use case. The CS-FLP is built upon a weighted-p-median model and its Mixed Integer-Linear Problem (MILP) formulation. By leveraging the OR-Tools library, the CS-FLP model can determine optimal placements for Charging Stations across an LSP's network, effectively identifying the most suitable locations. This newly implemented digital model can be customized to cater to the requirements of various LSPs operating in Europe. The CS-FLP can be formally described as follows: Given a set of locations, denoted as N, and considering that the LSP intends to install Charging Stations at Y out of the N locations, the CS-FLP provides a solution that determines the specific N locations where the Charging Stations should be placed. This solution minimizes operational costs by taking into account factors such as distances and demand. The what-if scenario that has been addressed by this model is:

- Based on past delivery demand and real-world distances in between the LSP’s depot locations in the PORTO area, **where** are the most suitable locations to add charging stations so that all depots are served with the minimum operational cost?

CS-FLP can be mainly re-used by researchers, but with higher maturity levels it can be re-used by LSPs and technology providers in last-mile logistics. Researchers can leverage CS-FLP as a foundation for further investigation and advancements in the field of charging station facility location optimization. They can build upon the existing model, modify its components, and explore new features or algorithms to enhance performance. Researchers can contribute by refining the model's objective function, considering additional constraints, or incorporating novel optimization techniques. Additionally, LSPs can utilize CS-FLP to optimize their charging station deployment strategies. By customizing the model according to their specific requirements and network characteristics, logistics companies can identify the most suitable locations for placing charging stations. Finally, technology providers focused on last-mile logistics can integrate CS-FLP into their software solutions or services. By incorporating the CS-FLP model into their platforms, technology providers can offer advanced features related to charging station optimization. This integration allows them to assist logistics companies in determining optimal charging station locations based on their specific networks and operational requirements. Possible future extensions are the incorporation of heterogenous fleets of EDVs and different types of charging stations, including conventional charging and ultrafast charging [7].

The digital model is described in further detail in LEAD's deliverable report D3.14 with 'Porto Value Case (final demonstrator)' and has been made available on GitHub under the following link:

- https://github.com/Horizon-LEAD/CS-FLP

This digital model is openly available under Apache-2.0 license and is searchable under the keywords: “routing”, “vrp”, “logistics”, “last-mile” and “rescheduling”.

3.3.1.6 Eqasim: A household data generator pipeline

LEAD project and the Lyon LL have contributed to the development of eqasim, a software package that is available online and currently being further developed by the global scientific community on GitHub. The goal of eqasim is to produce synthetic household activity data based on publicly available data sources. Household activity data describe the mobility patterns of populations in specific areas and can then be used in simulations or other types of descriptive analysis. Eqasim mainly requires the following data sources:

1) Census data,
2) Origin-destination data,
3) Population totals,
4) Income tax data,
5) Service and facility census,
6) National household travel survey data,
7) IRIS zoning system data and zoning registry data,
8) Enterprise census data,
9) Addresses data,

After successfully consolidating the data and running the eqasim pipeline, the users of the model can derive the following synthetic data:

- Person and household data, including persons and households of the region under study, including all of their sociodemographic attributes,
- Activities and trips data, which include transport demand based on the activity-based framework, including the purpose of travel and transport modes,

This data pipeline is a very important tool that produces synthetic datasets that are very close to reality (based on state-of-the-art modelling) when real-world household and transport demand datasets are not available (which is the most common for most cities in the world). The main benefit that stems from this synthetic household demand pipeline is that it is a standardized way to produce synthetic populations that are used by researchers and transport planners in agent-based simulations and other research and planning efforts. One example is the analysis of the synthetic transport demand data produced by eqasim for the evaluation of changes in a city's transportation infrastructure. Another example is the analysis of changes in transport vehicles of the population of the city, which can be examined by analyzing the transport demand data and alternating the transport modes in the

data before feeding this data to simulation tools. These types of data have also been used to successfully study the spreading of diseases and the placement of services and facilities.
The software is released under the following license at the following links:

- https://github.com/eqasim-org/ile-de-france (GPL license)
- https://github.com/Horizon-LEAD/irtx-synpop (MIT license)

The utilization of eqasim within the LEAD project is further explained in LEAD's deliverable D3.8. The searchable keywords for the repositories are “simulation”, “open-data”, “transport”, “population-synthesis”, “agent-based”, “France”, “matsim”.

3.3.1.7 Lyon LL's Parcel deliveries demand synthesis pipeline

Eqasim pipeline, as described above in section 3.3.1.7, produces household and transport demand synthetic datasets. However, in order to study last-mile and logistics networks to a greater extent, parcel delivery datasets are required. This is the gap that is filled by Lyon LL's open deliveries demand synthesis pipeline, which utilizes the results of eqasim to provide parcel demand datasets in the cases where the last-mile logistics network needs to be studied, but there is no real-world parcel dataset.

The parcel synthesis pipeline is an extension to the eqasim software, which is utilized to generate parcel demand for a region based on synthetic household data. A household dataset usually represents sociodemographic attributes of residents of an area along with mobility patterns. In that sense, the parcel synthesis pipeline is a standardized way in which researchers can go from mobility pattern data used for typical transport science studies to datasets for logistics-focused studies. Logistics-focused studies can include Vehicle Routing Problem solutions and simulations.

The main re-use value of the dataset is based on the fact that real-world delivery data are hard to be found by researchers and by R&D projects that need to study and plan last-mile logistics networks. Often LSPs do not share their delivery data being afraid that they may disclose personal information or information that may give advantages to competitors. In these cases, researchers and other practitioners in the field can utilize Lyon LL's parcel demand synthesis pipeline to derive synthetic data and gain insights into how the last-mile logistics network may benefit from measures and interventions. The pipeline is based on eqasim results but can be utilized with other data sources (such as real household survey data, etc.).

The software is released under MIT license at the following link:

- https://github.com/Horizon-LEAD/irtx-parcels

And is described in greater detail in LEAD's D3.8 along with eqasim. The searchable keywords for this repository are “synthetic”, “demand”, “parcels”, “logistics”.

3.3.2 Open datasets

Several open datasets have been utilized in the context of LEAD project in order to facilitate the application of the DT Platform in the LLs. These datasets have usually been the input or the output to a digital model, which in some way processes the data.

In the case that the dataset is used as an input to a digital model, then it is usually associated with one of the parameters of the model, which can be the road network, transport, or parcel demand. On the other side, when the dataset is an output of the model, then it can be some form of optimized decision on an operational, tactical, or strategical planning level. Datasets in this section can also be the results of surveys that depict the perceptions of professionals in the field and have been used as input for the determination of decision criteria in scientific methods. Despite the initial use of each dataset within LEAD project, different use cases exist for each dataset.

3.3.2.1 Open parcel delivery demand datasets

3.3.2.1.1 Delivery services dataset from the Madrid area

This is an open dataset that contains information about logistics services provided by CityLogin from the Madrid LL, from December 2021 to July 2022, in monthly buckets, in comma separated values (.csv) format, that can be individually downloaded. At the time of writing this deliverable, it contains more than forty-five thousand geo-referenced services data that have been anonymized (not showing the actual real-world CLOGIN full waybill number). The services dataset will be updated monthly until the end of the project, and each service can be a drop-off service or a pick-up service.

This dataset can be used to represent the demand covered by an LSP in Madrid and can be utilized by decision support models that utilize analytics (descriptive, predictive, or prescriptive models). The main use case for this open output is that it can serve as a test dataset for researchers and software developers that aim to develop computational software for the optimization of deliveries (e.g., Vehicle Routing Problem solvers, facility location problems) or software platforms that help monitor and calculate different aspects of deliveries (such as emissions, kilometric distances, tracking services). The dataset is available at the following link on the GitHub platform:

- https://opendata.lastmile.team/

The dataset is available under the Attribution 4.0 International (CC BY 4.0) license, which enables users to Share (copy and redistribute the material in any medium or format) and Adapt (remix, transform, and build upon the material for any purpose, even commercially).

3.3.2.1.2 Parcel delivery demand dataset for the Lyon area

Similarly, to Madrid LL open logistics services dataset, this dataset represents the demand for LSPs in the Lyon LL. The main difference to the first dataset is that this one contains slightly different information (and thus columns) and is also synthetic (synthetically produced by software according to a scientific method).

As in the case with the Madrid LL logistics services open dataset, this dataset describes the potential demand for logistics services in the Lyon area. It can be used for the testing and development of several scientific methods, digital models, and software systems, but in comparison to the Madrid LL open dataset, this one is useful for the development of predictive and prescriptive models. Descriptive analysis of such a dataset may be of lower re-use value since this dataset is synthetic. The dataset is available at the following link:

- https://doi.org/10.5281/zenodo.7697117

The dataset is available under license CC-BY 4.0 International The searchable keywords for this repository on Zenodo are “synthetic”, “data”, “parcels”.

3.3.2.1.3 Parcel demand data for the Hague area

This dataset includes the results of the application of the parcel demand generator software (as a sub-module of the MASS-GT-LEAD Version, described in Section 3.3.1.2) for the Hague LL area. It is a synthetically generated parcel demand based on the number of households and business establishments in each geographical zones. Some of the information (table columns) that they include are parcel IDs, city zone of origin location, city zone of destination location, depot identification number, delivery service provider, and vehicle type. The dataset is available on the TU Delft portal, alongside the necessary guidelines, at the following link:

- https://github.com/Horizon-LEAD/Parcel_Generation/tree/main/Output

Similarly, to other parcel demand and delivery services datasets, this one can be re-used by external stakeholders (researchers, developers, SMEs, etc.) mainly to test the development of new software products and services or the underlying optimization and scheduling algorithms. The output owner has made the dataset available on a GitHub repository under license GPL-2.0. The repository is searchable by the keywords “freight-simulation” and “parcel-synthesis”.

3.3.2.1.4 Parcel demand dataset for Budapest LL

Another open dataset for parcel demand has been made available from BKK and the Budapest LL. The dataset will include information about parcels to be delivered over the territory for an average day. The files will be provided in the shapefile (.shp) of the delivery area, a table with the shop's list (their ID and coordinates) as well as the loading routes of the subnetwork area.

This dataset can ideally be re-used by researchers or other stakeholders that have access to PTV's VISUM simulation software. The dataset, alongside the necessary documentation, will be made available by the end of the project at the following link:

- https://zenodo.org/record/8144170

The dataset has been made available under Creative Commons Attribution 4.0 International and is searchable by the keywords “last mile logistics”, delivery demand dataset” and “last mile network”.

3.3.2.1.5 Parcel demand data for the Porto area

The Porto LL has decided to make the last-mile parcel delivery demand dataset [8] for SONAE's depot for the Porto region openly available to the research community. The dataset includes several features like order ID, delivery location (i.e., postal zone), time window of delivery, boxes per order. The dataset, similarly, to all of the datasets described in Section 3.3.2.1 can be re-used to study the last-mile logistics network of several regions in Europe. The dataset is available on Zenodo, alongside the necessary guidelines, at the following link:

- https://zenodo.org/record/8143163

At the same Zenodo repository, Porto postal zone information and SONAE's depot locations are available as accompanying datasets that can be used to facilitate research activities. Similarly, to other parcel demand and delivery services datasets, this one can be re-used by external stakeholders (researchers, developers, SMEs, etc.) mainly to test the development of new software products and services or the underlying optimization and scheduling algorithms. The dataset is searchable under the terms: "last mile", "parcel delivery demand", and "vrp", "routing", "optimization", and "digital twins".

3.3.2.2 *Survey results dataset*

3.3.2.2.1 Survey results dataset on the importance of last-mile logistics sustainability criteria according to experts: REMBRANDT scale

In the context of the development of the STAR Logistics model, UPM has surveyed the opinions of experts in last-mile logistics and produced a dataset that can be then used for the derivation of sustainability criteria that can aid decision support processes.

In more detail, the dataset provides the importance from the point of view of logistics, mobility, and sustainability experts (i.e., 53 responses included in the dataset) of each LEAD sustainability criterion in last-mile logistics obtained from a pairwise comparison method. Through a questionnaire, the experts are asked to compare the importance of the different sustainability criteria on a scale from -8 to +8, known as the REMBRANDT scale. The information and the overall dataset has been finalized, anonymized, during the last few months of LEAD project. The survey results will be anonymized and available by the end of the LEAD project at the following link:

- https://zenodo.org/record/7695814#.ZAW0RnbMK3C

The dataset is available under license Creative Commons Attribution 4.0 International and is findable under the following keywords: "Last Mile Logistics", "Multi-criteria Decision Analysis" and "Expert Weights".

3.3.2.2.2 Crowd shipping survey dataset based on the Hague region

This section action includes two datasets that have been collected under the crowd shipping survey in the Hague LL region. The first dataset includes the results of a Stated Preferences (SP) survey performed in the context of the Hague LL and aimed to collect the declared willingness of potential bringers (such as students) to carry some parcels. The dataset is available on the TU Delft portal, alongside the necessary guidelines, at the following link:

- https://data.4tu.nl/datasets/614eb537-ff5c-48e9-867d-0b7855a15951

The dataset has been made available under license CC BY-NC-SA 4.0 The searchable keywords for this open dataset are activity-based models, crowdshipping, discrete choice models, passenger trip generation, value of time. A second dataset is included related the preferences of consumers towards the choice of crowdshipping services or traditional couriers:

- https://data.4tu.nl/datasets/b1cd5e55-c3c2-4d5b-8ec7-58f071f53b6e

The dataset I also avavialble under license CC BY-NC-SA 4.0. The searchable keywords for this open dataset are Crowdshipping, Hybrid Choice Model, Stated Preference, Trust.

Each dataset provides insights for companies and researchers who are looking to plan and realize crowd shipping services in the Netherlands and elsewhere. Based on this, the crowd shipping service providers can decide on the pricing model and or other aspects of the crowd shipping model that will be adopted.

3.3.3 Open API software

In this section, an open API software that has been utilized within the LEAD project are presented. This set of software tools has been either developed or extended within the LEAD project with the ultimate goal of supporting the Digital Twinning process.

Usually, digital transport models utilize a series of data inputs, process them in a specific way and then produce an expected output, which, most of the time, can serve as an input for a decision-making process and logistics solutions planning. To facilitate the communication between models and always have the proper data formats LLs have developed software modules for the handling of data.

With that in mind, the open software packages, as included in this section, can not serve as a standalone replicable digital model but can support the application and reproduction of digital models in multiple regions of the world and for different stakeholders.

3.3.3.1 API to COPERT (EMISIA's environmental impact calculation model)

Environmental impact calculation of road transport is one of the biggest challenges in last-mile logistics but in another sector in the T&L industry. EMISIA's environmental impact calculation model called COPERT is the EU standard vehicle emissions calculator, which utilizes several parameters such as vehicle population, mileage, speed, and other data sources to derive emissions and energy consumption for a specific country or region. Its widespread acceptance by researchers globally [9] showcases its importance for estimating T&L traffic impact on the environment.

COPERT has been utilized within the LEAD project and several LL's model execution workflows to derive an estimate for the impact of last-mile logistics traffic on the urban environment. However, for the automated utilization of COPERT to be realized by the DT platform, an interface that triggers the execution of COPERT based on the parameters provided has been needed. This interface is now being openly available online to be re-used by researchers who are interested in examining the impact of transport traffic (through simulation) on the surrounding environment. The re-use value is especially focused on the fact that the standard version of COPERT (openly available online) is manually triggered through a command prompt interface within a single environment (e.g., Windows or Linux platform), while with this API, COPERT can be programmatically called from software programs/ICT platforms. In more detail, with the API, the COPERT model can be called multiple times (for multiple scenarios/simulations), while the input data can be retrieved from a database and outputs can be stored at a database. This open API/interface will be uploaded to GitHub by the end of the project at the following link:

- https://github.com/Horizon-LEAD/copert-webservice

The software package has been made available under license an Apache-2.0 license and is searchable according to the keywords “COPERT API”, “emissions calculation”, “interface”.

4 Conclusion

This overall report assembles the open outputs of LEAD project and provides a description of their potential re-use by several practitioners in the field of last-mile logistics or the wider T&L sector. In Section 2, the Open Innovation Process supporting the overall guiding principles for the collection and publication of open outputs are given. In Section 3, the listing of open outputs is given alongside illustrative examples of their potential re-use cases as they emerge from the LEAD project. For each output, as they are listed in subsection 3.3, a link to the platform that hosts the actual open output.

Overall, it is expected that several of these open outputs of LEAD project will be re-used since the LEAD project covered a lot of emerging research-intensive technical challenges to realizing the DT Platform for the analysis of last-mile logistics networks. LEAD's outputs can benefit both research institutions and industry practitioners, but most importantly, European citizens can expect long-term benefits from the re-use of open outputs, which can be both environmental and socioeconomic. However, it is important to state that the re-use examples are just indicative of what are some re-use cases for the open outputs, and it is expected that over the years, they will be used to address several emerging research questions. On the industry level, since LEAD's open outputs, as they are listed in this report, are digital tools, it can be expected that industry practitioners with strong IT and R&D departments will be able to extract the most utility out of this set of tools.

Based on Figure 3, one can notice that the LEAD project will make openly available seven digital models, which are focused on last-mile logistics networks and the examination of “what-if scenarios” regarding the several last-mile configurations in scope. They enable simulation on mainly two levels of network planning: *i)* the tactical level (e.g., MASS-GT) and *ii)* the strategic level (e.g., 2-echelon). Very important also are the synthetic parcel demand production pipelines and models, which enable researchers and practitioners in the field to derive realistic synthetic datasets in cases where real-world datasets are not available. This is one of the very critical research problems in modern T&L scientific literature with real-world value for both the public sector and authorities trying to guarantee the public welfare, as well as LSPs that are looking to have the most profitable configuration possible for their networks.

Additionally, the LEAD project makes available seven open datasets and one open-source software API. While the open datasets are very important for mainly addressing theoretical or applied research questions in T&L science, the open API software package made available as a central solution to the productionisation and re-usability of software produced by research labs and their deployment towards real-world applications. Especially in the case of open software APIs made available by the LEAD project, stakeholders will be able to connect and integrate into their modelling, the well-received COPERT digital model.

Finally, it is important to state the open outputs of the LEAD project are very important for European R&D and projects that will emerge over the next few years (e.g., the URBANE project under Horizon Europe). As underlined by the European Green Deal, which aims to reduce net emissions of Greenhouse Gases (GHG) to zero by 2050, it is very important for European cities to have access to the tools that will support the EC's vision and will improve the welfare of European citizens both from the socioeconomic and environmental perspective. LEAD project's open outputs are another important steppingstone in this direction, which will hopefully support upcoming European R&D projects in their missions.

References

[1] Edward M. Corrado, “The Importance of Open Access, Open Source, and Open Standards for Libraries,” *Issues in Science and Technology Librarianship*, 2005.

[2] Rishab A. Ghosh, Ruediger Glott, Bernhard Krieger, and Gregorio Robles, “Free/Libre and Open Source Software: Survey and Study”.

[3] A. van Loon and D. Toshkov, “Adopting open source software in public administration: The importance of boundary spanners and political commitment,” *Gov Inf Q*, vol. 32, no. 2, pp. 207–215, Apr. 2015, doi: 10.1016/j.giq.2015.01.004.

[4] F. Xiao, D. He, Y. Chi, W. Jeng, and C. Tomer, “Challenges and Supports for Accessing Open Government Datasets,” in *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, New York, NY, USA: ACM, Mar. 2019, pp. 313–317. doi: 10.1145/3295750.3298958.

[5] European Commission, “Open Source Software Strategy 2020-2023: Think Open,” Oct. 2020.

[6] European Commission, “Facts and figures for open research data.” https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor/facts-and-figures-open-research-data_en (accessed Sep. 08, 2022).

[7] S. Y. ten Have, K. Gkiotsalitis, and K. T. Geurs, “Investigating the Future of Ultrafast Charging: A Choice Experiment in the Netherlands,” *World Electric Vehicle Journal*, vol. 11, no. 4, p. 70, Nov. 2020, doi: 10.3390/wevj11040070.

[8] Paulo Reis, Dimitrios Rizopoulos, Andreas Alexopoulos, and Ioanna Fergadiotou, “Porto area Parcel Delivery Demand dataset,” *Zenodo*, 2023.

[9] Md. S. Alam, B. Hyde, P. Duffy, and A. McNabola, “Analysing the Co-Benefits of transport fleet and fuel policies in reducing PM2.5 and CO2 emissions,” *J Clean Prod*, vol. 172, pp. 623–634, Jan. 2018, doi: 10.1016/j.jclepro.2017.10.169.

Annex A

In this annex, we include the four templates that have been used to collect the required information from project partners about their open outputs. The templates, as described in Section 2: LEAD Open Innovation Process - Open Assets, are listed below, starting with open digital models request template:

Open digital model information		
Topic	**Questions**	**Answers**
Partner and model basic info	Partner	
	Please enter your name.	
	Please enter the name of your model.	
Digital Model detailed description	Please provide a description for your model: What is the model used for and what are the results that it can derive?	
	From whom is it expected to be re-used?	
	Is your model fully developed and ready to be used in its current version? Will it be ready by Month 30 (November 2022)?	
Ethics, Privacy, Security	Does your model include personal or sensitive information?	
Availability	Is your model already open-source on a platform? If yes, please indicate which one (GitHub, GitLab, etc.) If no, are you willing to make your model available to stakeholders outside of the project consortium through an online website?	
Publishing info	Are there any available documentation/guidelines?	
	Please describe any restrictions that are related to your model and its potential use by other interested organizations through an online platform or a publication.	

Table 2: Template for the request for open digital models in LEAD project

Next the template used for the open datasets is included. As one can notice it includes request sections that are specific to datasets.

Open dataset information		
Topics	**Questions**	**Answers**
Partner and dataset basic info	Partner	
	Please enter your name.	
	Please enter a title for your dataset	
Dataset detailed description	Please provide a description for your dataset: What is your dataset used for?	
	From whom is it expected to be re-used?	
Ethics, Privacy, Security	Does your dataset include personal or sensitive information?	
Availability	Is your dataset already openly available on a platform? If yes, please indicate which one (Zenodo, EC Open Data portal, etc.) If no, are you willing to make your dataset available to stakeholders outside of the project consortium through an online website?	
Publishing info	Is there any available documentation/guidelines?	
	Please describe any restrictions that are related to your dataset and its potential use by other interested organizations through an online platform or a publication.	

Table 3: Template for the request for open datasets in LEAD project

Finally, the last template used is portrayed below, that includes API software-specific fields.

Open API service information		
Topics	**Questions**	**Answers**
Partner and API service basic info	Partner	
	Please enter your name.	
	Please enter a title for your API service	
API service description	Please provide a description for your API service: What is your API service used for?	
	From whom is it expected to be re-used?	
	Can your API potentially serve as an open service?	
Ethics, Privacy, Security	Does your API service include personal or sensitive information?	
Availability	Is your API service already available at an endpoint? If yes, please indicate which one (address, etc.) If no, are you willing to make your API service available to stakeholders outside of the project consortium through an endpoint?	
Publishing info	Are there any available documentation/guidelines?	
	For how long can you keep the API service running after the end of the project? Is there any specific cost that is required to keep the API service running in the next few years? Is this cost covered or can be covered?	
	Please describe any restrictions that are related to your API service and its potential use by other interested organizations through an online platform or a publication.	

Table 4: Template for the request for open API software in LEAD project