ELABORATOR

Deliverable 1.5

Data domain protocol for transport research

CIVITAS 2030
Sustainable and smart mobility for all
THE CIVITAS INITIATIVE IS CO-FUNDED BY THE EUROPEAN UNION

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

Co-funded by the European Union

Document Control Page

Deliverable number	D1.5
Deliverable title	Data domain protocol for transport research
Deliverable version	1.0
Work Package number	Work Package 1
Work Package Title	Project management
Due date of delivery	31/05/2024
Actual date of delivery	04/10/2024
Dissemination level	Public
Type	Report
Editor(s)	Vangelis Angelakis, LiU
Contributor(s)	Monica Olyslagers, iRAP
Reviewer(s)	Hans Skov-Petersen KU; George Oikonomou, UBris.
Submitted by (Project Coordinator)	Dr. Angelos Amditis, ICCS
Project name	The European Living Lab On Designing Sustainable Urban Mobility Towards Climate Neutral Cities
Project Acronym	ELABORATOR
Project starting date	01/06/2023
Project duration	42 months
Rights	ELABORATOR consortium

Document History

Version	Date	Beneficiary	Description
0.1	11/09/2024	Vangelis Angelakis, LIU	Early Draft
0.5	27/09/2024	Vangelis Angelakis, LIU	Peer review
1.0	4/10/2024	Jason Sioutis, ICCS	Submission

Project Executive Summary

ELABORATOR stands for 'The European Living Lab on designing sustainable urban mobility towards climate neutral cities. The EU-funded project uses a holistic approach for planning, designing, implementing and deploying specific innovations and interventions towards safe, inclusive and sustainable urban mobility. These interventions consist of smart enforcement tools, space redesign and dynamic allocation, shared services, and integration of active and green modes of transportation.

They will be specifically co-designed and co-created with identified "vulnerable to exclusion" user groups, local authorities and relevant stakeholders. The interventions will be demonstrated in a number of cities across Europe, starting with six Lighthouse cities and six Follower cities with three principal aims:

I. to collect, assess and analyse user needs and requirements towards a safe and inclusive mobility and climate neutral cities;
II. to collect and share rich information sets made of real data, traces from dedicated toolkits, users' and stakeholders' opinions among the cities, so as to increase the take up of the innovations via a twinning approach;
III. to generate detailed guidelines, policies, future roadmap and built capacity for service providers, planning authorities and urban designers for the optimum integration of such inclusive and safe mobility interventions into Sustainable Urban Mobility Plans (SUMPs).

ELABORATOR Lighthouse cities

- Milan (Italy)
- Copenhagen (Denmark)
- Helsinki (Finland)
- Issy-les-Moulineaux (France)
- Zaragoza (Spain)
- Trikala (Greece)

ELABORATOR Follower cities

- Lund (Sweden)
- Liberec (Czech Republic)
- Velenje (Slovenia)
- Split (Croatia)
- Krusevac (Serbia)
- Ioannina (Greece)

Social Links:

 Twitter/X

 LinkedIn

For further information please visit www.elaborator-project.eu

Project Partners

Organisation	Country	Abbreviation
INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS	EL	**ICCS**
POLIS AISBL	EL	**POLIS**
EVROPSKI INSTITUT ZA OCENJEVANJE CEST	SI	**EURORAP**
MULTICRITERI-MCRIT AIE	ES	**MCRIT**
INSTITUT D'ARQUITECTURA AVANCADA DE CATALUNYA	ES	**IAAC**
COMUNE DI MILANO	IT	**CDM**
STEFANO BOERI ARCHITETTI SRL	IT	**SBA**
THINGS SRL	IT	**THIN**
AGENZIA MOBILITA' AMBIENTE E TERRITORIO SRL	AMAT	**AMAT**
KOBENHAVNS KOMMUNE	DK	**CPHK**
KOBENHAVNS UNIVERSITET	DK	**UCPH**
ANALYSE & TAL F.M.B.A	DK	**A&T**
FORUM VIRIUM HELSINKI OY	FI	**FVH**
TEKNOLOGIAN TUTKIMUSKESKUS VTT OY	FI	**VTT**
SOCIETE D'ECONOMIE MIXTE ISSY - MEDIA (SEM ISSY MEDIA)	FR	**ISSY**
COLAS	FR	**COLAS**
IFP ENERGIES NOUVELLES	FR	**IFPEN**
URBAN RADAR	FR	**URAD**
AYUNTAMIENTO DE ZARAGOZA	ES	**AYZG**
FUNDACION CIRCE CENTRO DE INVESTIGACION DE RECURSOS Y CONSUMOS ENERGETICOS	ES	**CIRCE**
JOC RENTAL S.L	ES	**MYR**
ANAPTYXIAKI ETAIREIA DIMOU TRIKKAION ANAPTYXIAKI ANONYMI ETAIREIA OTA	EL	**ETRIK**
URBANA	EL	**URB**

LUNDS KOMMUN	SE	**LUND**
LINKOPINGS UNIVERSITET	SE	**LIU**
SENSATIVE AB	SE	**SENS**
STATUTARNI MESTO LIBEREC	CZ	**LIBER**
CESKE VYSOKE UCENI TECHNICKE V PRAZE	CZ	**CVUT**
MESTNA OBCINA VELENJE	SI	**MOV**
AV LIVING LAB, D.O.O.	SI	**AVLL**
INTERNET INSTITUTE, COMMUNICATIONS SOLUTIONS AND CONSULTING LTD	SI	**ININ**
GRAD SPLIT	HR	**SPLIT**
SVEUCILISTE U ZAGREBU FAKULTET PROMETNIH ZNANOSTI	HR	**FPZ**
CITY ADMINISTRATION OF THE CITY OF KRUSEVAC	RS	**KRUS**
MUNICIPALITY OF IOANNINA	EL	**IOANN**
PLATOMO GMBH	DE	**PLAT**
INTERNATIONAL ROAD ASSESSMENT PROGRAMME	UK	**IRAP**
UNIVERSITY OF BRISTOL	UK	**UBRIS**

List of abbreviations and acronyms

Acronym	Meaning
BOAI	Budapest Open Access Initiative
BY	Attribution
CC	Creative Commons
EC	European Commission
EU	European Union
FAIR	Findable, Accessible, Interoperable, and Reusable
GDPR	General Data Protection Regulation
HE	Horizon Europe
KPI	Key Performance Indicator
LL	Living Lab
NC	Non-Commercial
ND	No Derivatives
PII	Personally Identifiable Information
SA	Share Alike
WP	Work Package

Contents

1 Introduction

The ELABORATOR project aims to promote inclusive, safe, and sustainable urban mobility in European cities. Good data management is essential to this aim by ensuring all research data is carefully handled, shared, and preserved according to EU standards.

The *Data Management Plan* (DMP) of the ELABORATOR project (D1.4) aims to ensure the proper handling of research data in line with FAIR principles, as well as compliance with EU requirements and data protection regulations such as the GDPR. The DMP provides detailed guidelines on data collection, processing, sharing, preservation, and reuse of research data generated during the project.

In the ELABORATOR project, there are regular updates of the DMP. The first version was submitted six months after the commencement of the project and additional versions are to be submitted on a yearly basis (D1.6 M18, D1.7 M30 and D1.9 M42) reflecting adaptations required to fit the project needs.

This deliverable, the *Data Domain Protocol for Transport Research* (D1.5) supports the DMP by establishing a standard template for data management specific to the transport research domain. In doing so, it supports the FAIR principles (Findable, Accessible, Interoperable, and Reusable) for the transport research data produced and sourced during the project.

1.1 Overview of the deliverable

This deliverable (D1.5) includes the following chapters:

1. Section 1 (this section) provides an introduction outlining the purpose, scope and linkages with other WPs,
2. Section 2 outlines a best-practice model for a Data Management Plan (DMP),
3. Section 3 describes the ELABORATOR domain model, and
4. Section 4 discusses Open and FAIR data principles at large
5. Section 5 Concludes the report.

1.2 Purpose and Scope

This report is designed to facilitate the development of the ELABORATOR DMP so as to be used by project partners and stakeholders to streamline the process of handling and publishing mobility-related data, fostering collaboration across multiple cities and research initiatives. It outlines best practices for publishing transport data in open-access repositories, ensuring compliance with ethical and legal standards such as the General Data Protection Regulation (GDPR), and emphasises securing sensitive and personal data while promoting data sharing for research purposes. The template proposed is adaptable and can be used across different transport research projects. It defines standards, policies, methods, and tools to ensure that the collected data meets EU requirements for open access and data sharing.

1.3 Links to the other Work Packages and Reports

The *Data Domain Protocol for Transport Research* is tightly coupled with multiple work packages, tasks and deliverables. These are summarised here and details for their relations are provided in the subsequent sections of the report.

1.3.1 Inputs from WP2

WP2 has been establishing the approaches and tools for evaluating urban mobility interventions across the pilot cities. It has been focusing on developing an inclusion plan (D2.1), the project Evaluation plan (D2.2) and gathering the project approach to LL cocreation in a co-creation playbook (D2.3). Finally digital toolkits for the project focusing on co-creation and co-evaluation (both to establish baselines and to perform impact assessment) have been delivered (D2.4) For the work herein and towards deriving a concrete DMP, the co-creation

methodologies (Co-Evaluation in D2.2) and tools (Co-Cretion Tools D2.3 & D2.4) developed in WP2 provide a clear pathway for identifying, entities and relationships between stakeholders, methods and tools, and their input/output information requirements.

1.3.2 Relationship with WP3

The ELABORATOR WP3 is tasked with delivering the interventions co-definition with the pilot sites relevant stakeholders. In this regard it performs the project Discovery phase (T3.1), and delivers the definition of interventions at the 6 project Lighthouse Pilots D3.4 / D3.5, while also delivering the twinning methodology (D3.2) enables the definition of interventions at follower cities (D3.6). LL governance forms are being developed to assist the smooth operation of the pilot demonstrations in T3.4. The overall types of stakeholders of the project (project internal – externals) their roles and ownerships of different assets (from access to physical resources such as pre-existing sensors at the intervention locations, to software relevant for the data processing) have been exposed in the work already performed in T3.1, and T3.3. Conversely the work herein, jointly with that of WP4 (T4.1) drives the discussions on evaluation data and KPIs liked with the definition of interventions (T3.3).

1.3.3 Collaboration with WP4

WP4 is tasked with the sharing of mobility intervention indicators across the partners' LLs in order to monitor the impact of interventions and support knowledge transfer. Having a core objective to "Devise a framework for the harmonization, homogenization, and unification of mobility intervention data" the work in D1.5 acts as the stepping stone to establish the data models needed in WP4 (see Sections 3.1.2 and 3.1.4).

1.3.4 Contributions to WP5 and WP6 (Implementation Phase)

WP5 and WP6 are ones delivering all activities relevant to implementation rollout and demonstration of interventions at the six lighthouse and six follower pilot sites, respectively, implementing the twinning programme of the project. For this a D5.1 lays the foundation for the data management delivering the domain model on which the platform of WP4 will be built upon towards efficient and effective information exchange across the sites and stakeholders.

1.3.5 Collaboration with WP7 (Evaluation and Impact Assessment)

WP7 addresses the evaluation and assessment of all interventions, for which the KPIs of D2.2 will be revisited – to this end the definition and interaction of different components defined in the domain model pertaining to the relationship between the Evaluation Plan, KPIs and the Datasets has central importance allowing the future DMPs to facilitate the flow of information across the necessary APIs.

1.3.6 Collaboration with WP8 (Evaluation and Impact Assessment)

WP8 is tasked amongst the dissemination, outreach and upscale planning to deliver the handbook of the ELABORATOR FAIR data (D8.3) to this end the present deliverable together with D1.4 the initial DMP of ELABORATOR provide the basics on which D8.3 will deliver, in conjunction with WP4 and the updated DMP (D1.6) on M18.

2 A Model DMP Template facilitating Mobility Research in the ELABORATOR

2.1 DMP Components

Towards establishing a complete Data Management Plan (D1.6) for the ELABORATOR project, going beyond the initial sketch of D1.4 the following items should be included in a final DMP:

1. Mobility and Intervention Data Description

- **Types of Data**: Describe the nature and type of data (e.g., quantitative, qualitative, raw, processed) that will be collected, generated, or reused
- **Purpose of data**: Describe the need for each data item / dataset based on the information in the intervention definitions (D3.4), and the evaluation methodology (D2.2)
- **Format**: Specify the file formats (e.g., CSV, JSON, Excel, video, image formats) in which data will be stored.
- **Ownership**: Define who owns the dataset and how (conditions and methods) the data or what derivatives of the data can be shared

2. Data Collection Methods

- Outline how data will be collected, generated, or acquired subject to GDPR national laws constraints (See D1.4)
- Include tools, software, instruments, or procedures used in data collection.
- Document any data standards or protocols followed.

3. Data Documentation and Metadata enabling FAIR data management

- **Metadata Standards**: Describe the metadata (data about the data) standards that will be used for describing the dataset (e.g., Dublin Core, DDI, ISO standards).
- **Documentation**: Detail how data will be documented, such as using codebooks, data dictionaries, or README files, to ensure others can understand and use the data.

4. Responsibilities and Resources

- **Roles and Responsibilities**: Identify the individuals responsible for data management tasks (e.g., PI, data manager) per organization, and at project level as a whole.

5. Ethics and Legal Compliance vis a vis GDPR

- **Consent**: Outline how informed consent will be obtained if the data involves human participants, provide standardized consent forms.
- **Data Privacy and Confidentiality**: Describe how personally identifiable information (PII) will be protected and pseudo- or anonymized.
- **Intellectual Property Rights**: Address issues around data ownership and intellectual property, beyond the ones stated in the project Consortium Agreement.

5. Storage and Backup

- **Data Storage**: Specify where each dataset will be stored during the project (e.g., academic/ municipal servers, partner cloud services).
- **Backup Strategy**: Provide details on backup frequency, methods, and locations to prevent data loss.

- **Long-term Preservation**: Specify how data will be preserved after the project, and for how long (e.g., how may data be redacted after 5 or 10 years).
- **File Formats**: Use formats that are widely used and sustainable for long-term access (e.g., non-proprietary formats like CSV or TXT).

6. Security

- **Access Control**: Describe measures to secure sensitive or confidential data, such as encryption or restricted access.
- **Security Protocols**: Detail how unauthorized access will be prevented.

7. Sharing and (Open) Access

- **Data Sharing**: Specify how when and which data will be (i) shared across the project partners (ii) open for external parties.
- Ensure that all aspects of the research (e.g., code, algorithms) are reproducible and well-documented.
- **Repositories**: identify the repository or archive where open data will be deposited (e.g., Zenodo, institutional repository).
- **Licensing**: Indicate the license under which the data will be shared (e.g., Creative Commons, Open Data Commons).

3 The Data landscape in an urban mobility intervention-centred project

The initial version of the ELABORATOR DMP (D1.4) introduces the main principles of the GDPR and FAIR data management. This provides the basic framework for data collection and storage for the ELABORATOR project. Moreover, the DMP introduced the project's Data Landscape and established guidelines for the processing of personal data and their protection providing details accountability and governance.

This section introduces and initiates the Domain model of the ELABORATOR project focusing on the co-creation (co-design through co-evaluation) of urban mobility interventions in the context of interconnected Living Labs.

3.1 The ELABORATOR Domain Model

3.1.1 The Need for a Domain Model in the context of the ELABORATOR

A **Domain Model** is a conceptual framework that represents the **key entities, their relationships, and the interactions** within a specific domain or system. It serves as a high-level abstraction, **organising** and **defining** how different components of a system or environment relate to each other and operate together to achieve specific goals.

A domain model is not tied to a particular implementation or technology; nor it is a detailed description of individual entities. Rather, the domain model is intended to provide a clear, structured view of the **core concepts** within the domain.

Key Features of a Domain Model:

1. **Entities** are the primary objects, components, or participants that are involved in the system. These could be physical objects, digital components, people, or abstract concepts. For example, in the context of the ELABORATOR: 'Stakeholders', 'Devices', and 'Interventions',
2. **Relationships** define how entities are connected or associated with each other. This could describe ownership, dependency, or interaction between entities. For example, a Stakeholder owns a (monitoring) Device in the Area of an Intervention,
3. **Interactions** describe the behaviours and processes that occur between entities. This includes the exchange of information, communication, or actions taken by one entity in response to another. E.g. A Stakeholder reviews the evaluation outcome of an Intervention.

The role therefore of a Domain model is to delimit and define the foundational structure that will guide the definition of the architecture of concrete solutions in the digital thread of a large multi-stakeholder project, serving as a guide providing a common language for the derivation of the architectural blueprint.

Purpose and Benefits of a Domain Model:

- **Clarification**: A domain model helps clarify the structure of a system by providing a **visual representation** of its core elements and their relationships.
- **Communication**: It facilitates better communication among stakeholders, developers, and business analysts by providing a **shared understanding** of the system's core concepts.
- **Design Guidance**: It helps guide the system design by offering a blueprint for implementing processes, interactions, and data management.

- **Problem Understanding**: By focusing on the key entities and their relationships, a domain model aids in understanding the **core problem** that the system is designed to address[1].

3.1.2 From Domain to Data Model

The **Domain Model** and the **Information/Data Model** of a system are closely related but serve distinct roles in system design and architecture. While both models focus on the structure of a system, they operate at different levels of abstraction and serve different purposes. The **Domain Model** focuses on conceptual entities and relationships, whereas the **Data Model** deals with the precise structure of data in the system, guiding how data is stored, retrieved, and managed.

Role of the Domain Model in Relation to the Information/Data Model:

- The **Domain Model** is a **conceptual** representation that defines the key **entities**, their **relationships**, and **interactions** within the specific domain. It helps to identify the main components and behaviours without worrying about technical details like storage or database structures.
- The **Data Model** (which includes both **information models** and **physical (deployment) models**) is a more **concrete representation** that describes how the domain's entities are mapped into **data structures**, such as databases, tables, fields, and schemas. It specifies how data will be **stored, organised, and managed** in the system. The report of D4.1 will contain the Information model of the ELABORATOR, while D4.2 ought to contain the physical / deployment aspects, completing the Data Model.

3.1.3 The Domain Model for ELABORATOR as a multi-stakeholder intervention-centred project

In Figure 1 we depict the domain model for ELABORATOR, which in fact can be generalised, as is, to describe the domain *for any similar multi-stakeholder intervention-based project*. The overall concept is centred on the **Intervention** entity, which takes place in one of many **Areas** (location) of the project.

To describe the Intervention, we introduce the concept of an **Intervention Plan** which in turn is composed of multiple **Tasks**. One such Task is that of the instantiation of the **Co-evaluation Method** which has been selected for gauging the Intervention impact. To deliver that, a set of **KPI**s are needed, each of which is represented by a **Dataset** associated with the Intervention.

In turn, a Dataset is produced by a **Data Processor**, which represents all manipulations (e.g. prediction fusion, redaction, pseudo- or anonymisation, etc.) that need to take place to data generated by the **Devices** that have been used to collect them for the purposes of the Intervention. The aim of the Data processor is to deliver data for FAIR storage in the ELABORATOR such that they:

(i) respect the policies of the DMP (ensuring the data is FAIR, OPEN -where appropriate, and GDPR compliant), and

(ii) format them such that they are directly associated to the relevant KPIs, or other usages in the project.

Such a different usage, for example in ELABORATOR, can be considered the use by the Community of Practice in project Knowledge Hub.

Project data is generated by Devices. There is no technological or methodological restriction on the Device entity. For example, a Device could be:

(i) a sensor deployed for pre/post- evaluation of an Intervention,

(ii) a questionnaire used for the same purpose, or

(iii) the materials (miro boards / menti / physical post-it notes on a board etc.) that have been used, by different stakeholders, in a workshop to co-design an Intervention.

[1] In the ELABORATOR this would be how to identify and structure the mobility data under a FAIR an Open considerations to co-design and co-evaluate interventions with a broad array of stakeholders, such that solutions' twinning will be enabled.

(iv) external data sources (e.g. geodata)

It is straightforward to see that the need of a Data Processor for the latter two is far more complex as data from such Devices can be not just qualitative but even analogue in nature.

The Intervention is characterised by an **Intervention Plan** brought about by a **Co-Design Method** (see as a point of departure **D2.3**), which jointly with the **Co-Evaluation** call for a host of **LL Actors** to be involved. These actors come from a pool of **Stakeholders**, which can be either **Project Partners** (e.g. the team of ELABORATOR beneficiaries from the local ecosystem of an Intervention Area) or **External** (i.e. non-beneficiaries) engaged in the Living Lab.

The Stakeholders can be both engaged in the co-creative processes (**Co-Design, Co-Evaluation)** of the Intervention, or merely own a certain Device or Data Set pertinent to the Intervention. Stakeholders may be (ELABORATOR or external) researchers or publishers who deliver publications. Allowing for external Stakeholders to deliver **Publications** based on the project is an intentional push towards Open Access to Datasets which we cover in later sections of this report.

In Figure 1, the notation of relationships between the entities utilises UML notation[2] (Fowler, 2003). In a nutshell the notation comprises boxes to indicate entities, arrows of different styles connecting the entities and numbers where connections meet the entities.

[2] Fowler, M. (2003). UML Distilled: A Brief Guide to the Standard Object Modeling Language. Addison-Wesley Professional.

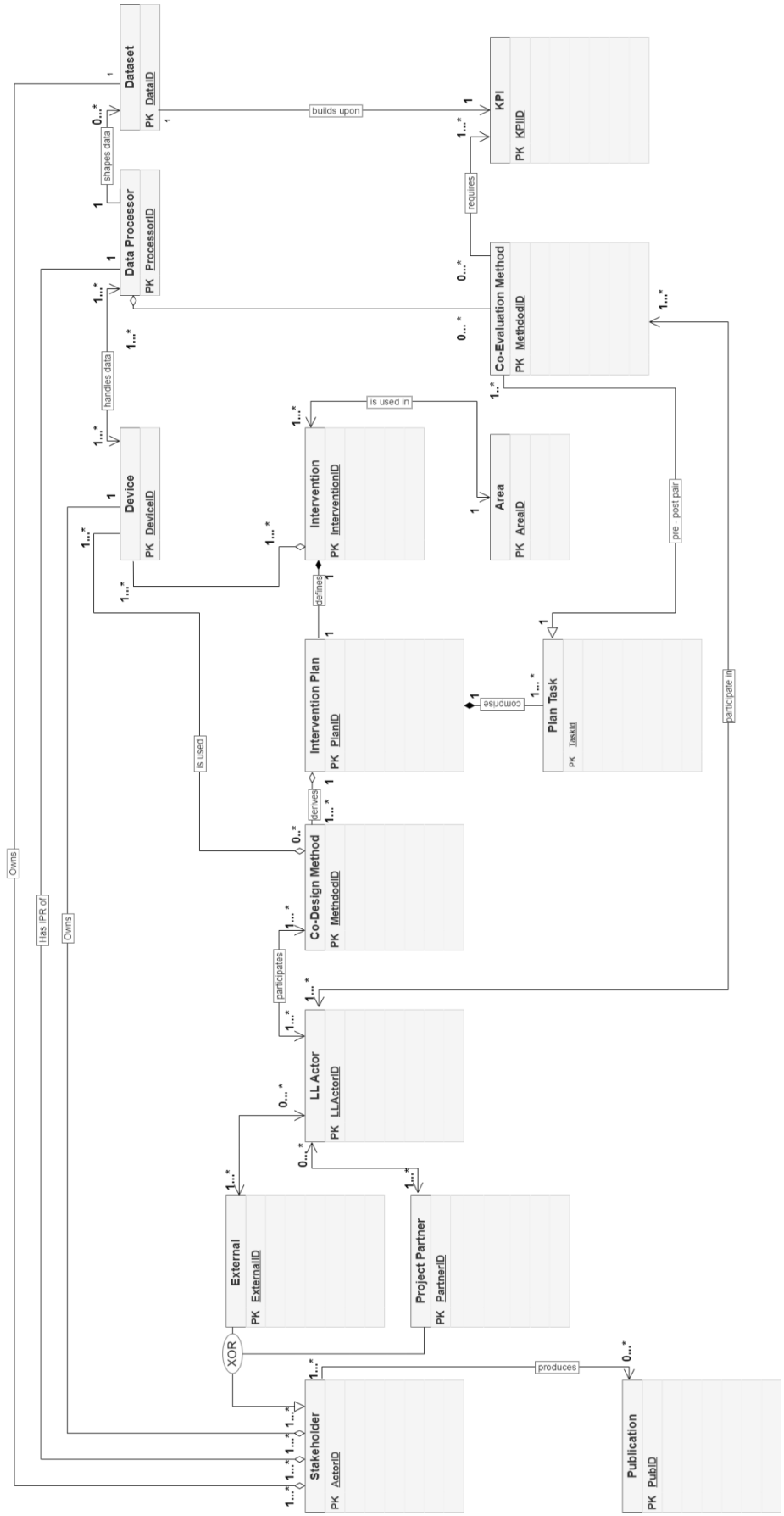

Figure 1: *The Domain Model for the ELABORATOR (explanations of signatures are found in the following text).*

The shape of the arrows bears the following significance:

- **(empty) Triangles – Generalisation,** is used to depict an *is-a* relationship, i.e. the arrow points to a "parent" entity, all of whose attributes are inherited by the "child" entity from which the arrow originated. The child, which can be seen, reversely as a specialisation of the parent entity, can have unique additional attributes. One such example are the External and Project Partner entities which in the context of the LL are considered to be children of the Stakeholder.
- **empty Diamonds – Aggregation,** is used to depict a *has*-a relationship, i.e. the diamond points to an entity that comprises others whose existence is may not be depended on its lifetime. A key example here is the Device entity: Any Intervention and Co-Design entities aggregate (multiple) Devices, which however, may have been deployed regardless of the Intervention.
- **filled Diamonds – Composition,** is used to depict an *is-part*-of relationship, i.e. the diamond points to an entity that comprises others whose existence is conditioned on it. A key example here is the Intervention Plan and Plan Task relationship, with the Plan Tasks not existing as standalone entities outside a complete Intervention Plan.
- **Arrows – one/two way associations,** single direction arrows are used to describe relationships where the arrow arriving entity does not need to be aware of the entity that the arrow originates from (e.g. the KPI – Co-Evaluation methodology – the KPI is a standalone measure, which does not need to be informed of where it shall be utilised). On the other hand: an LL Actor needs the information of the Co-Evaluation methodology they participate in, and conversely the Co-Evaluation Methodology is shaped by the information of the participating LL Actors.

The numbers indicated on the relationships indicate the association Cardinalities across the entities, with "*x…y"* or * indicating respectively "from x" with x being 0 to note "may be none" or x being 1 to indicate "there has to be at least one" "to y" with y being 1 (at most) or * to indicate *any* number.

3.1.4 Pathway to Information model in D4.1 and information handling in the following DMP

The Domain Model plays a key role in guiding the development of the Data Model (for T4.1) by providing a high-level understanding of the necessary to capture entities, relationships, and behaviors within the system's domain. It thus sets the foundation by identifying the core concepts that need to be reflected in data, ensuring that the Data Model maps the real-world objects and interactions the system must support. In the Domain Model, entities are defined conceptually, representing abstract ideas like "Stakeholder" or "Device." These conceptual entities are later translated into concrete structures in the Data Model, such as tables or objects, where more technical details—like data types, indexes, and constraints—are specified.

The Domain Model focuses on high-level business logic, identifying which entities are important to the system and how they relate to one another. For example, in a system where a **Stakeholder** *owns* a **Device**, this relationship is captured by an association in the Domain Model. Moving to the Data Model, this relationship is implemented through technical constructs like foreign keys, associations, or join tables in a relational database, or through references in a NoSQL database. Thus, the Domain Model provides the conceptual framework, while the Data Model turns this framework into a concrete blueprint that dictates how data will be stored and managed at a technical level. In this way, the Domain Model directly informs what data structures need to be created and how they will be connected.

Moreover, while the Domain Model may outline the entities their key attributes or properties of each (for instance, a **Stakeholder** having a *name*, *affiliation*, and *email*) are captured and elaborated in the Data Model with **technical specifications**. For example, in the Data Model, the Stakeholder's "name" would be defined as a string data type, perhaps with length constraints or validation rules. The Data Model focuses not only on defining data attributes but also on specifying how these data elements will be **validated, stored, and queried within a database** or data market (T4.2). Thus, while the Domain Model provides a conceptual view of entities and their relationships, the Data Model offers a detailed technical representation, ensuring the system can handle specific data operations efficiently.

The Domain Model also plays a critical role in defining the types of interactions between entities, such as behaviors and workflows. For instance, in the Domain Model, interactions like a **Stakeholder** *owns* a **Device** or an **Area** has-an **Intervention** may be described in a way that outlines the system's behavior at a high level. In the

Data Model, these interactions are translated into database-level **operations**, such as **join operations, transactions, or even triggers that ensure data consistency and integrity**. By conceptualizing behaviours through associations in the Domain Model, the Data Model is better equipped to support the necessary data operations and optimize for performance based on the real-world interactions that need to be reflected in the system.

Together, the Domain and Data Models form the foundation for a comprehensive Data Management Plan. As the Domain Model defines the entities and relationships conceptually, and the Data Model provides the technical specifications and implementation details, this structured understanding directly informs the Data Management Plan's core components. Thus the derived DMP should detail how the data will be stored, secured, and managed throughout the project lifecycle. Specifically, the insights from both the Domain and Data Models help optimize key areas such as data storage, access control, data sharing policies, and long-term preservation strategies. By accurately reflecting real-world entities and their relationships, the Data Management Plan ensures the data remains accurate, secure, and efficiently handled at every stage of the project.

Summarizing: the Domain Model's conceptual clarity and the Data Model's technical precision work in hand in hand towards ensuring that the Data Management Plan is not only aligned with the system's needs but also robust enough to support complex data operations, secure data handling, and sustainable data management practices over time. This synergy between models enables a more thoughtful, detailed, and effective approach to managing the project's data.

4 Open and FAIR Data

In the context of an Innovation Action, as the ELABORATOR project, open access refers to the practice of making research outputs (such as scientific publications and data) freely available online, without cost or access restrictions, in order to ensure transparency, foster innovation and facilitate collaboration and upscale.

Open Access and **FAIR Data** are both essential components of **Open Science**, working together to ensure that research is freely available, transparent, and reusable. While Open Access ensures that scientific publications are accessible to everyone, FAIR principles ensure that the research data underpinning those publications is well-organized, reusable, and shared under clear conditions, facilitating a more effective and open research ecosystem.

In this report we consider FAIR as a set of principles aimed at improving the management and sharing of research data so that it can be easily located, accessed, combined with other datasets, and reused by both humans and machines. To this end we contrast the interplay of Open and FAIR data.

Common Principles:

- **Accessibility**: Both Open Access and FAIR data promote the idea of increasing the accessibility of research. OA focuses on removing barriers to reading publications, while FAIR principles emphasize making data accessible (though not necessarily openly available) in a standardized, well-described, and retrievable manner.
- **Reusability**: Open Access allows unrestricted reading and reusing of publications (especially with open licenses like Creative Commons). FAIR data emphasizes that data should be reusable through the use of clear licensing, appropriate formats, and sufficient metadata for others to understand and apply it.
- **Transparency**: Both frameworks encourage transparency in research, enabling validation, replication, and secondary use of the research outputs.

Complementary Goals:

- Open Access often applies to **publications**, ensuring that research results can be freely read, while FAIR principles mainly apply to **research data**, ensuring that data can be easily found, understood, and reused.
- Together, Open Access and FAIR data support the larger vision of **Open Science**, where both scholarly publications and the underlying research data are freely available and reusable, promoting collaboration, reproducibility, and innovation in research.

Implementation Differences:

- **Open Access** typically requires removing paywalls or other restrictions for accessing research outputs (publications), sometimes through repositories or open-access journals.
- **FAIR Data** does not necessarily require data to be open but instead focuses on proper **metadata, standards, and repositories** that allow data to be findable and reusable, even if access is restricted due to privacy, security, or legal concerns. In that regard access/cost schemes to FAIR data becomes significant in view of the data market platform of T4.2 of the ELABORATOR project.

Licensing:

- In **Open Access**, clear and open licensing (see discussion on CC above) is key to enabling the free use of research papers.
- In **FAIR Data**, data can be openly licensed (if appropriate), but FAIR principles accommodate the possibility that some data might be restricted or controlled due to sensitive or proprietary reasons. What's crucial is that **metadata is open** and the **conditions for accessing and reusing the data are clearly defined**.

Mutual Reinforcement:

- **Open Access** to publications ensures that research findings are widely disseminated, and when combined with **FAIR Data**, it allows other researchers to access not just the publications but also the supporting data in a usable form.

- The integration of Open Access and FAIR Data creates a holistic environment where both the **outputs (publications)** and the **inputs (data)** of research are available for scrutiny, reuse, and advancement of science.

For a project handling GDPR-collected or sensitive data, open access should be carefully applied. Any research data that is collected under GDPR must be anonymized or redacted to protect personal information before it is made openly accessible. Only data that complies with ethical and legal standards can be shared, while ensuring that the **metadata** remains open to provide visibility and **findability**.

With regards to research papers or other academic publications (e.g. Theses), these must be made freely available in open access repositories immediately upon publication, with proper licensing (e.g., CC BY— see below). The overall project information, such as public reports or deliverables, should also be published on the project web page, ensuring that stakeholders can access and review the results of the project, while respecting confidentiality agreements and legal constraints on certain proprietary data.

The European Commission mandates, via the Grant Agreement, **strict Open Access** policies for research outputs from its funded **Horizon Europe** projects. Core requirements are:

1. **Immediate Open Access to Publications:**
   - **All peer-reviewed scientific publications** arising from Horizon Europe funding must be made **immediately available** without embargo in an open-access repository at the time of publication.
   - **No embargo periods** are allowed under Horizon Europe, which marks a stricter requirement compared to previous programs like Horizon 2020.
   - Research can be published in either **Gold Open Access** journals (which provide immediate open access upon publication) or opt for **Green Open Access** by depositing the author's accepted manuscript in a trusted repository (e.g., Zenodo, arXiv, or institutional repositories e.g. Diva for Swedish academia).
2. **Licensing Requirements:**
   - Publications must be released under a license that allows free access and reuse, typically a **Creative Commons Attribution (CC BY)** license, allowing others to copy, distribute, and build upon the work as long as they give appropriate credit (see below).
3. **Open Access to Research Data:**
   - Research data must follow the principle of **open by default**. All research data produced in Horizon Europe projects must be made open, unless there are legitimate reasons to restrict access (e.g., privacy, security, intellectual property rights).
   - If data cannot be fully open, the project must ensure the **metadata** is accessible, findable, and clearly explains how the data can be accessed or why it is restricted.
   - Researchers are required to create a **Data Management Plan (DMP)** outlining how they will ensure that data complies with the **FAIR principles** (Findable, Accessible, Interoperable, and Reusable).
4. **Repository Use:** Researchers must deposit both publications and data in **trusted repositories**. Publications should be linked to datasets, ensuring full transparency and reproducibility.
5. **Acknowledgment of EU Funding:** All open access publications must explicitly acknowledge the funding received from Horizon Europe and include a reference to the specific grant agreement.

4.1 Open Access & Licensing core references

The *Budapest Open Access Initiative* (BOAI): One of the foundational declarations in the open access movement (2002). It defines open access as the free availability of scholarly research literature on the public internet. https://www.budapestopenaccessinitiative.org/

The *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*: A significant framework promoting open access by encouraging research institutions, libraries, and others to make scientific knowledge freely accessible (2003). https://openaccess.mpg.de/Berlin-Declaration

4.1.1 Creative Commons Licensing

Creative Commons (CC) licenses are a suite of open licensing models that allow content creators to define how others can access, share, and reuse their work. These licenses are designed to provide flexibility, making it easier for creators to legally share their work while retaining certain rights. Each CC license specifies different levels of permissions for users, balancing openness with the creator's control over their content.

The main types of **Creative Commons licenses** are:

1. CC BY (Attribution)

What it allows: This is the most open of the CC licenses. It allows anyone to **distribute, remix, adapt, and build upon the work** for any purpose, even commercially, as long as they give proper credit to the creator.

Use case: This license is ideal for research publications and data, where the creator wants to encourage widespread reuse and collaboration while ensuring they are credited.

This is the recommended license for publications under Horizon Europe, promoting maximum openness.

2. CC BY-SA (Attribution-ShareAlike)

What it allows: This license allows others to remix, adapt, and build upon the work, even for commercial purposes, as long as they give credit to the creator and **license their new creations under identical terms** (ShareAlike). It ensures that any derivative works will also be openly accessible.

Use case: Often used by creators who want their work to be freely used, but who also want to ensure that any derivative works are shared with the same openness. This license is common in open-source communities and collaborative projects.

3. CC BY-ND (Attribution-NoDerivatives)

What it allows: This license allows others to reuse the work for any purpose, including commercially, as long as it is passed along **unchanged** and in whole, with credit to the creator. Users cannot remix or build upon the work.

Use case: Suitable for authors who want to allow free distribution of their work but do not want others to make changes or adaptations, ensuring that the work is always shared in its original form.

4. CC BY-NC (Attribution-NonCommercial)

What it allows: Others can remix, adapt, and build upon the work **non-commercially**. They must credit the creator, but they do not have to license derivative works under the same terms.

Use case: This license is ideal for content creators who want to allow others to reuse their work, as long as it is not used for commercial purposes.

5. CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)

What it allows: Others can remix, adapt, and build upon the work **non-commercially**, as long as they credit the creator and license their new works under the **same terms**. This ensures that any derivative non-commercial works are also shared openly.

Use case: Often used in educational and non-profit sectors, this license allows free use and adaptation of content but ensures that it remains non-commercial and openly accessible.

6. CC BY-NC-ND (Attribution-NonCommercial-NoDerivatives)

What it allows: This is the most restrictive Creative Commons license. Others can **download and share** the work with proper credit, but they cannot change it in any way or use it commercially.

Use case: This license is suitable for creators who want to make their work freely available for viewing or sharing but do not want it to be used commercially or modified.

Key Features of Creative Commons Licenses:

- **Attribution** (BY): All licenses require that the original creator is credited.
- **ShareAlike** (SA): Derivatives must be licensed under the same terms as the original work.
- **NonCommercial** (NC): The work cannot be used for commercial purposes.
- **NoDerivatives** (ND): The work cannot be altered or adapted.

Choosing a License:

The creative commons web page provides a tool supporting the choice of license:

https://chooser-beta.creativecommons.org/

A general set of rules of thumb:

- **Maximum openness**: If the goal is to maximize reuse and dissemination, **CC BY** is the most appropriate license.
- **Collaboration and sharing**: For projects that aim to promote not only reuse but also collaborative building (with the same open licensing requirements), **CC BY-SA** is ideal.
- **Control over adaptations**: If the creator wishes to allow redistribution but retain control over adaptations, **CC BY-ND** or **CC BY-NC-ND** are more suitable.

4.2 FAIR Data

In D1.4 FAIR data has been explored for the ELABORATOR in terms of metadata to be included in research publications, namely the minimum fields of metadata that should come with a ELABORATOR project-generated scientific publication in a repository:

- The terms: “European Union (EU)”, “European Commission” (EC), “Horizon Europe” (HE)
- Identifier of the action (HORIZON-MISS-2022-CIT-01 -01)
- Acronym and grant number (ELABORATOR, 101103772)
- Publication date
- Length of embargo period if applicable
- Persistent identifier.

D1.4 noted that the DMP will not include the actual metadata to be produced in the ELABORATOR project, prior to the release of D4.1, and D3.1.

5 Conclusion

This report has been developed to (i) foster the evolution of the ELABORATOR DMP, and to support the development of the project data modelling facilitating the digital thread of the project. These will be used by project partners and stakeholders to streamline the process of handling and publishing mobility-related data, fostering collaboration across multiple cities and research initiatives. We have outlined best practices for publishing transport data in open-access repositories, ensuring compliance with ethical and legal standards such as the General Data Protection Regulation (GDPR), and emphasised the need for Open data for research purposes. The DMP template and the Domain model proposed are adaptable and can be used across different transport research projects. It defines standards, policies, methods, and tools to ensure that the collected data meets EU requirements for open access and data sharing.