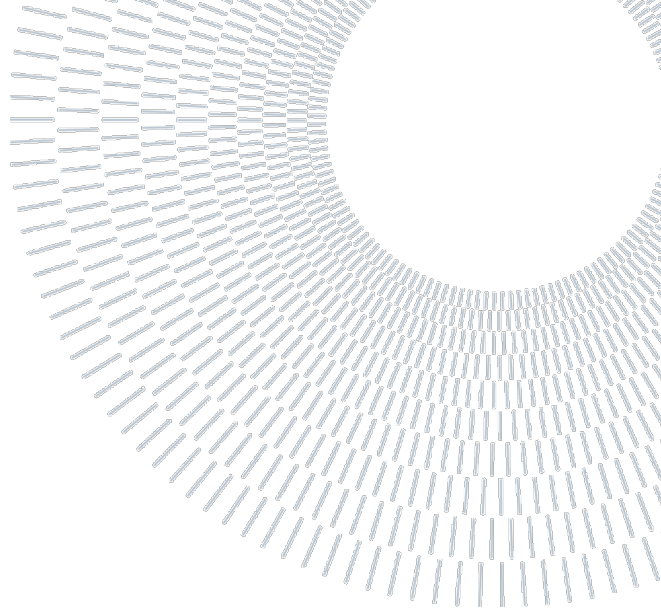




**POLITECNICO MILANO 1863**



# **Open Data in Public-Private collaborations: overcoming the impasse on limited data re-use**

TESI DI LAUREA MAGISTRALE IN  
MANAGEMENT ENGINEERING  
INGEGNERIA GESTIONALE

**Authors: Elena Di Girolamo, Errico Cascone**

**Student ID: 10798131, 10555326**

**Advisor: Giuliano Noci**

**Co-advisor: Irene Vanini**

**Academic Year: 2021-2022**

## **Acknowledgements**

We would like to thank our thesis supervisor Giuliano Noci, for allowing us to carry out this thesis work of great interest and relevance.

We would also like to thank Irene Vanini, our co-supervisor, for guiding us through the thesis research and writing process. Her support was essential, and her assistance was flawless.

Finally, we would like to thank everyone who took part in the GIOCOnda project, in particular Osservatorio Agenda Digitale of Politecnico di Milano.

# **Open Data in Public-Private collaborations: overcoming the impasse on limited data re-use**

Submitted to the Department of Management Engineering on November 24, 2022, in partial fulfillment of the requirements for the degree of Master of Science in Supply Chain Management

## **ABSTRACT**

This thesis work is part of the GIOCOOnDa project. The main objective of this project is to improve spatial governance through the use and sharing of Open Data between Public Administrations (PA) and private companies. This dissertation concerns two main topics. The first entails an examination of information availability and needs in the Italian-Swiss cross-border area. The second consists in developing an integration model to enable standardization and optimization of the data exchange process to integrate and maximize the re-use of the data. To achieve these goals, a literature analysis was done, followed by interviews with various representatives from the public and private sectors. The output of the thesis work is an integration model, designed based on the results obtained from these analyses. Then, the ultimate objective expected is to increase the database shared within the GIOCOOnDa platform and stimulate the use of Open Data through the model developed. The contribution of this study lies, therefore, in unifying best practices from different case studies to make the sharing process standardized and as efficient and effective as possible. It is expected, then, that there will be greater participation and inclination towards the Open Data ecosystem thanks to the use of the model developed, which demonstrates the method to undertake this path in the best possible way and the related benefits for the participants and the whole civil society.

**Key Words:** Open Data; Public-Private Partnership; Data Re-use; Integration; Collaboration.

## ABSTRACT IN ITALIANO

Questo lavoro di tesi fa parte del progetto GIOCOOnDa. L'obiettivo principale di questo progetto è quello di migliorare la governance territoriale attraverso l'utilizzo e la condivisione di Open Data tra Pubbliche Amministrazioni (PA) e aziende private. La affronta principalmente due argomenti. Il primo riguarda l'analisi della disponibilità e del fabbisogno informativo nell'area interregionale italo-svizzera. Il secondo consiste nello sviluppo di un modello di integrazione che consenta la standardizzazione e l'ottimizzazione del processo di condivisione dei dati per integrare e massimizzare il riutilizzo di questi ultimi. Per raggiungere questi obiettivi, è stata effettuata un'analisi della letteratura, seguita da interviste a diversi rappresentanti del settore pubblico e privato. L'output del lavoro di tesi è un modello di integrazione, progettato sulla base dei risultati ottenuti da queste analisi. L'obiettivo finale atteso è, quindi, quello di incrementare la base dati condivisa all'interno della piattaforma GIOCOOnDa e stimolare l'utilizzo degli Open Data attraverso il modello sviluppato. Il contributo di questo studio consiste, infine, nell'unificare le migliori pratiche provenienti da diversi casi di studio per rendere il processo di condivisione standardizzato e il più efficiente ed efficace possibile. Ci si aspetta, perciò, una maggiore partecipazione e propensione all'ecosistema degli Open Data grazie all'utilizzo del modello sviluppato, che dimostra il metodo per intraprendere questo percorso nel migliore dei modi e i relativi benefici per i partecipanti e l'intera società civile.

**Parole chiave:** Open Data; Partnership Pubblico-Privato; Data Re-use; Integrazione; Collaborazione.

## Table of Contents

EXECUTIVE SUMMARY OF THE THESIS .....	7
<b>1. INTRODUCTION.....</b>	<b>14</b>
1.1. TOPIC AND CONTEXT.....	16
1.2. THESIS' FOCUS AND SCOPE .....	18
1.3. STRUCTURE.....	19
<b>2. METHODOLOGY OF THE WORK .....</b>	<b>21</b>
2.1. LITERATURE METHODOLOGY .....	21
2.1.1. <i>Systematic Literature Review</i> .....	21
2.1.2. <i>Narrative and Grey Literature Review</i> .....	30
2.1.3. <i>Final Considerations</i> .....	33
2.2. EMPIRIC METHODOLOGY .....	35
2.2.1. <i>Qualitative research</i> .....	35
2.2.2. <i>Exploratory research model</i> .....	36
2.2.3. <i>Multiple case study approach</i> .....	36
2.2.4. <i>Semi-Structured Interviews</i> .....	37
2.3. MODEL DEVELOPMENT METHODOLOGY .....	41
<b>3. LITERATURE REVIEW .....</b>	<b>43</b>
3.1. INTRODUCTION .....	43
3.2. MAIN DEFINITIONS .....	44
3.3. EUROPEAN AND ITALIAN LANDSCAPE .....	46
3.4. SECTOR ANALYSIS.....	50
3.4.1. <i>Environment</i> .....	50
3.4.2. <i>Health care</i> .....	56
3.4.3. <i>Smart Cities</i> .....	64
3.4.4. <i>Transport Sector</i> .....	67
3.4.5. <i>Tourism Sector</i> .....	71
3.4.6. <i>Culture Sector</i> .....	73
3.4.7. <i>Government sector</i> .....	76
3.5. FINAL CONSIDERATIONS.....	80
<b>4. EMPIRICAL ANALYSIS .....</b>	<b>83</b>
4.1. DATA .....	83

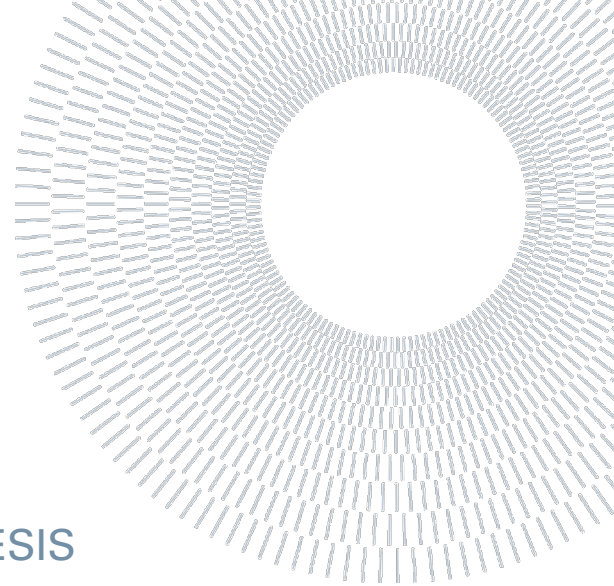
4.1.1.	<i>Data available</i> .....	84
4.1.2.	<i>Data enhancement</i> .....	84
4.1.3.	<i>Information need</i> .....	85
4.1.4.	<i>Critical factors</i> .....	87
4.1.5.	<i>Enablers factors</i> .....	89
4.2.	RELATIONSHIP WITH THIRD PARTY ORGANIZATIONS.....	89
4.2.1.	<i>Subjects from whom acquire/integrate data</i> .....	90
4.2.2.	<i>Limitations and difficulties</i> .....	91
4.3.	OPEN DATA .....	92
4.3.1.	<i>Open Data Culture</i> .....	92
4.3.2.	<i>Publication Venue</i> .....	93
4.3.3.	<i>Critical factors for sharing Open Data</i> .....	94
4.3.4.	<i>Enabling factors for Open Data sharing</i> .....	96
4.4.	POSSIBLE SCENARIOS.....	98
4.4.1.	<i>Current and/or future projects</i> .....	98
4.4.2.	<i>Pandemic impact</i> .....	99
4.5.	FINAL CONSIDERATIONS.....	99
<b>5.</b>	<b>LESSON LEARNED</b> .....	<b>105</b>
5.1.	SLR AND NLT .....	106
5.2.	EMPIRICAL ANALYSIS .....	107
5.3.	MAIN LESSON LEARNED .....	107
<b>6.</b>	<b>MODEL DEVELOPMENT</b> .....	<b>110</b>
6.1.	OPERATIONAL MODEL .....	111
6.1.1.	<i>Initiation</i> .....	113
6.1.2.	<i>Main activities</i> .....	115
6.1.3.	<i>Publish data for maximum reuse</i> .....	119
6.2.	CRITICALITIES AND RELATED SOLUTIONS .....	123
6.3.	BENEFITS AND COSTS .....	127
6.4.	FINAL CONSIDERATIONS.....	133
<b>7.</b>	<b>CONCLUSION AND FURTHER RESEARCH</b> .....	<b>136</b>
	<b>BIBLIOGRAPHY</b> .....	<b>137</b>
	<b>LIST OF FIGURES</b> .....	<b>144</b>
	<b>LIST OF TABLE</b> .....	<b>145</b>

<b>LIST OF ABBREVIATION .....</b>	<b>146</b>
<b>ANNEX .....</b>	<b>148</b>
<b>QUESTIONNAIRE.....</b>	<b>148</b>



**POLITECNICO  
MILANO 1863**

**SCUOLA DI INGEGNERIA INDUSTRIALE  
E DELL'INFORMAZIONE**



## EXECUTIVE SUMMARY OF THE THESIS

# Open Data in Public-Private collaborations: overcoming the impasse on limited data re-use

MASTER THESIS IN MANAGEMENT ENGINEERING – INGEGNERIA GESTIONALE

**AUTHOR: ELENA DI GIROLAMO; ERICO CASCONI**

**ADVISOR: GIULIANO NOCI**

**ACADEMIC YEAR: 2021-2022**

---

### Introduction

Digitization is a growing challenge, and this process is undoubtedly complicated. However, it is undeniable because it is a global trend, and Central Governments are working to hasten the digital transformation of their Countries through financial aid and education.

Focusing on Italy, an example is the “Piano Triennale per l’informatica nella Pubblica Amministrazione”. It is an essential tool for promoting the country’s digital transformation. One of its guiding principles consists of enhancing public data by making them available to businesses and citizen in an open and interoperable form.

In this context, Politecnico di Milano started a collaboration with SUPSI (Scuola Universitaria Professionale della Svizzera Italiana) on a project on Open Data: GIOCOOnDa. The main objective was to improve the governance of the cross-border area through the use and sharing of Open Data -

data that can be accessed, modified, re-used, and freely distributed (Bauer, Kaltenböck 2011) - between Public Administrations (PA) and private companies.

### *Aim and contribution of the research*

This thesis work is part of the GIOCOOnDa initiative with the final aim of developing two specific points that the project itself encompassed. The first concerned a thorough analysis of the information availability and needs of the areas of interest. The second consisted in developing an integration model to enable standardization and optimization of the data exchange process to integrate the data found to be necessary from the previous phase.

We currently operated in a context where a single standard process does not exist to support a company that wants to integrate and make its databases open. The contribution of this study lies, therefore, in unifying best practices from different case studies to make the sharing process standardized and as efficient and effective as possible.



The vision that guided our work was to generate greater participation and inclination toward the Open Data ecosystem through the use of the model developed, which demonstrates the method to undertake this path in the best possible way and the related benefits for the participants and the whole civil society.

The developed research is structured as follows: methodology, literature review, empirical analysis, results obtained, and model development. The executive summary follows the same format in order to preserve consistency.

## Literature and empiric methodology

We carried out the literature review using the Systematic Literature Review (SLR) method, an approach that guarantees the most objective and fair analysis possible Kitchenham (2006).

Since this study is synergetic to GIOCOOnDa project, the research question that guided us was given as it coincides with the objective of the project itself:

*Is it possible to develop a model for Open Data sharing across public and private organizations?*

Starting from it, we defined four search strings:

- *(ABS (integration AND public AND private AND data) AND ABS (transport OR health OR tourism OR culture OR environment OR government)).*
- *(ABS (data) AND ABS (reusable) AND ABS(LOD)).*
- *(ABS (LOD) AND ABS (public) AND ABS (private)).*
- *(ABS ("linked open data") AND ABS (USA)).*

After using these strings on Scopus and Google Scholar, only 21 articles were selected, which was insufficient for our objective. For this reason, having already a minimal knowledge of the topic and some essential papers to start from (e.g., Shoja et al. 2020), we were also able to carry out a Narrative Literature Review (NLR), which allowed us to obtain a more adequate number of articles. In addition, the review benefitted from the consultation of gray literature we also used Gray Literature when the topics allowed it.

As for the methodology of the empirical work, we decided opted for a qualitative (Bouchrika 2021)

analysis because it allows obtaining detailed information in a turbulent context.

We decided to adopt an exploratory (Naoum 2006) method, since it allows us to investigate a topic when there is limited information available.

Then, we decided to use the multiple case study approach to collect all the necessary information for model development. As a result, many scenarios were examined to gain a thorough understanding of all the individuals involved in the sharing process and their respective realities.

More in detail, we selected the Semi-structured interviews for this investigation (SSI) as they allow to ask open-ended questions and learn about the autonomous opinions of the respondent. The objective was to have on the one side specific information from all interviewees thanks to a set of predefined questions, but on the other side to ensure flexibility towards the different interviewees who in our case belonged to different sectors (Public Administration, private companies, Public Holdings and association of private companies). The questionnaire developed was designed with four sections: introduction; data availability; data necessity; criticalities, benefits, and future opportunities. The interview results were organized in a data collection model, developed by us, in order to be best evaluated.

The goal of the analysis consisted in gaining different information needed for the development of the sharing process. In particular, what and how much data is currently available, as well as what and how much data is required to start a win-win data-sharing process for the stakeholders has been researched. But also, the willingness of the interviewed to collaborate to open their databases has been the subject of study.

Finally, we made a link between the findings of these two investigations, on which we based the proposed model.

## Literature Review

The analysis was carried out in a sectoral manner with the goal to understand where sectors stand on the issue of data sharing. Some case studies were analyzed to have a more concrete approach to studying attempts to develop integration models. The main lesson learned from the literature were:

1) There is a gap in the literature, that is the lack of a widely used data-sharing model between Public Administration (PA) and private companies. Sharing models are not present in all sectors, and if so, they are inconsistent with one another.

2) Open Data is currently being expanded, and EU and various States are pushing for its deployment, through different directives and guidelines. Their goal is to make data more accessible and open to European citizens, and to increase transparency within local governments.

3) Several benefits emerged and, more specifically in our context, using Open Data in Public-Private Partnerships enables high-quality data by leveraging private technical skills and the resources of the PA.

4) The most significant issue: lack of interoperability, which requires time and money to be solved. These, however, are relatively precious resources for the private sector, which prefers not to invest because it does not perceive long-term benefits. This is due to a lack of culture and little readiness for change toward the use of Open Data.

## Empirical Analysis

We analyzed 12 different case studies. Each of them drew relevant best practices and practices to avoid embarking on a path of sharing. The information from the interviews was gathered using a data collection model. This section examines the various macro-areas of the model to determine which results are most relevant.

**Data.** Despite the large volume of data, both private individuals and the public expressed a desire for more information, which might be met through sharing. However, the biggest barrier is a lack of interoperability, which takes time and money to address.

**Relationship with third-party organizations.** The limitations in engaging a relationship with third parties lie, for example, in the presence of pseudo-silos that indicates that databases are not compatible with one another within the same organization. Moreover, the resources needed to keep the databases updated and in line with the regulations to avoid privacy issues are consistent.

There is also a strong jealousy of the information from privates hoping to profit from the information in their possession and the PA, concerned about the consequences of excessive transparency. On the contrary, the biggest benefit that emerged is related to the possibility of increasing the number of data available for the parties involved in the partnership.

**Open Data.** Interviewees are aware of what Open Data are; however, only 33% of them post Open Data. This is because the actors possibly involved are interested in Open Data and would like to embark on a path to that end, but the context is not ready. There is, in fact, little knowledge in general about Open Data, particularly on the part of civil society, which does not understand their utility. Another problem that slows down Open Data diffusion is the confidentiality issue because private companies are concerned about how competitors might use their data once they are public. The quality issue is another element to take into consideration because data providers must guarantee a certain level of data quality and it has a cost. Privacy is also a problem and to be sure to avoid it, the different actors must align their datasets to the GDPR policies, which requires money and time. Furthermore, there are no clear and defined processes for returning data. On the contrary, the most important benefits lies in the fact that sharing allows an improvement in business analysis and consequently, in the services and products developed.

**Possible scenarios.** Both public and private respondents have a strong desire to start the process of opening up the data. But, actually, of the 50% of the previously mentioned respondents who would have an interest in initiating an integration and openness project, only 25% have already concretely defined at least one. The main reason these projects have not taken off is due to cultural barriers, including authorities' lack of readiness and their perception that the environment is not yet ready to support this kind of business.

## Result

Comparing the outcomes of the literature review and the empirical analysis we established a relationship between them. From this link different best practices and practices to avoid were defined. Based on this information it was possible to conclude on the partnership features that need to be enhanced and avoided in order to expand the number of participants in PPPs and provide an effective database integration and openness process.

## Operational Model

The main parts of the operational model (Figure 3) are three: the initiation phase, the main process activities and the publishing data for the maximum reuse phase.

**Initiation.** The process starts when a data provider (private companies, private company associations, Public Administrations, and Public Holdings) expresses the willingness to integrate and open its database. The data provider who initiates a sharing project become the **Initiator**. Firstly, the Initiator should identify a so-called **Facilitator**, an expert in the data field, that serves as support and collaborate throughout all the process phase and collaborate.

**Main activities.** The goal of the main activities of the process is to prepare the data to be opened and to engage the needed data providers within the process.

The data mapping activity is the first to be carried out and has two objectives. The first is to understand how much and what data the initiator has and its data needs. The second is to understand how to obtain the needed data.

The second activity is focused on the engagement of the data providers. This activity is carried out differently depending on who initiates the project. Moreover, this phase is dependent on the previous one. Based on what emerged from the data mapping, in fact, the data providers who possess the identified information will be selected. Another parallel phase needs to run in parallel: the use case study phase. This stage is crucial because it allows for the clarification of the project's objectives, the

roles of the actors involved and, most importantly, the benefits that may emerge. It allows for the creation a solid foundation that can lead to increased awareness on the part of those promoting it.

At this point all actors are involved in the process and, since they use different standards for datasets development, a harmonization activity is required to make the databases interoperable. Finally, an additional process of checking and preparing data is required in order to ensure that the database is consistent with the guidelines for opening the data, which include an important anonymization step.

**Publish data for maximum reuse.** At this point, the process can take two roads: the data can be integrated and published on an open data network, or they can be published on an interoperable platform. The open data network is the set of open data portals that are developed by the countries' Public Administration. The interoperable repository, instead, is the set of open data portals that are developed by private organizations. Once the data is published on a platform it enters a self-feeding virtuous cycle. This means that the data will be used and provided continuously by the data providers who participated in the project to benefit from it, but it will also be used by final users who want to create additional value for the community that will passively benefit from it.

In the model, a block called High Value-Added data is also represented. The owners of these data of high value prefer to make profit from them. They are unwilling to integrate or make them accessible unless there is an immediate financial gain. Regarding the usefulness of these databases, what emerged is that the central government intervention would be necessary for funding the integration of these databases.

### *Criticalities and related solutions*

Four main criticalities created a well-defined need to find a solution: impasse, culture, non-interoperability, and costs.

**Impasse.** The facilitator will support the whole process, limiting the uncertainty that characterizes this context. Moreover, use cases are applied to demonstrate even before the project begins what its consequences will be. This will give a great deal

of structure to the project and will keep potential data providers from backing down because they do not know what they are up against and so they will help in overcoming the impasse.

**Culture.** The main problem is that since the ultimate goal is to create value for civil society, if there is no knowledge and then acceptance in the first place on their part, it is difficult to achieve success. Although to a lesser extent, a cultural problem exists within the organizations themselves. Private individuals and PAs have long been accustomed to viewing data as a valuable asset to be sold and it is difficult to separate the common imagination from this point of view. This is why efforts should be made internally and externally to raise awareness of Open Data and collaboration. As a result, a portion of the project budget will be allocated to training.

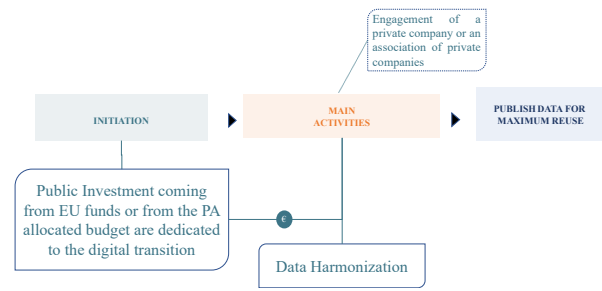
**Non-interoperability.** The facilitator will play an important role in speeding up the process of adhering to the existing standards that have the purpose of providing methodologies for homogenizing data in order to make them interoperable.

**Costs.** Private companies are concerned about the prospect of a return on investment. To understand the benefits of such an action, it is necessary to work on a cultural level and educate data providers in order to demonstrate what they can gain in the long run. For the PA the ultimate goal is to satisfy the citizen, and funds are allocated to various activities with this end goal in mind. If only citizens valued this activity and recognized the significance of PA digitization, they would be more empowered to allocate a majority share to openness, sharing, and Open Government. The intervention of the PA could then be used to push toward open data especially at this early stage when trust is still low.

### *Benefits and Costs*

The economic flow differs based on who initiated the project.

**Initiator: Public Administration or Public Holding.** In the case where the process is initiated by the Public Administration or Public Holdings, the flow is shown in figure 1 below.



*Figure 1: Economic flow when the Initiator is the PA.*

Because these authorities are public, they use public funds. As a result, they are unable to purchase independently and a tender is launched during the engagement phase to accomplish this need in order to guarantee the competitive dynamics without altering them. When it is the public that decides to undertake this path, it is the PA itself that will finance it. This financial support can come from European funds earmarked for digitization or from a portion of the budget allocated specifically for digitization by the PA that initiated the project. The only activity that will be funded by the private party joining the project is data harmonization. This decision is made because harmonization is an internal activity, which has to be done by the data provider of interest. The benefits that justify these investments both for the PA and the private are different.

Starting from the PA, by combining their data with the private companies' ones, they would have access to a wealth of information with which they can do market, feasibility or any kind of analysis in a much more realistic and reliable way. The direct consequence of advancement in information analysis is an improvement in the services they make available to society. The great amount of information will be useful also for budgeting and all the other activities that involve data. Another significant result is an increase in the PA's transparency. Moreover, improved management and a decrease in corruption follow greater transparency.

Similar to PA, the private sector will be able to develop high-quality analyses due to the increased amount of data obtained through integration. They will be able to save a significant amount of money as a result of these extremely reliable analyses. The volume of data available also enables in-depth analysis of company performance and the

products or services they provide, which allows it to continuously improve its offerings and increase the end customer's satisfaction. Another significant benefit is that the company becomes much more visible by sharing its data and opening it up to the public.

In conclusion, the data consumers are not true consumers. They are largely a passive beneficiary and are represented by civil society. This is because the average person is more interested in passively profiting from Open Data than actively consulting and using it.

**Initiator: private company or association of private companies.** In the case where the process is initiated by a private company or an association of private companies, the monetary flow is depicted in figure 2 below.

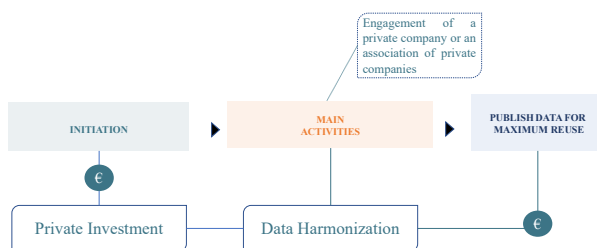


Figure 2: Economic flow when the Initiator is private.

In this instance, a single private person or group of individuals made all of the investment. The PA will be responsible for paying the costs of data homogenization when it joins the project during the engagement phase for the same reason as stated above. In this case, an interoperable platform needs to be created. Moreover, there will be not only standard users, the civil society, but also premium users that pay for higher service.

The advantages of this case are the same as the previous one, with just a few more for the private, that can have a return on the investment. This is because the premium users pay a small amount of money to benefit from the service offered, financing in this way the investment. In any case, the private party will be the owner of the interoperability platform, which could always be sold if the results are not satisfactory. Also, the premium users will have some benefits. They will access a source that is simple to get to and that

includes safe and reliable data at an affordable price. Last but not least, the premium user could use this data to benefit civil society as a whole rather than just themselves, feeding the virtuous cycle by, for instance, developing an application.

## Conclusion and further research

The contribution of this study lies, according to the objectives of the GIOCONDa project, in unifying best practices from different case studies to develop a sharing process standardized and as efficient and effective as possible. It is hoped, then, that there will be greater participation and inclination toward the Open Data ecosystem thanks to the use of the model, which demonstrates the method to undertake the “opening path” in the best possible way and the related benefits for the participants and the whole civil society.

This model can be viewed as a starting point that can subsequently be validated in order to widely disseminate the process of sharing and openness as a common practice.

## References

- Agid (may 2022), “Piano Triennale per l’Informatica | Agenzia per l’Italia Digitale.”
- Bauer, F., Kaltenböck, M. (2011). *Linked open data: The essentials. Edition Mono/Monochrom, Vienna, 710.*
- Kitchenham, B. (2006). *Evidence-based software engineering and systematic literature reviews Springer Berlin Heidelberg.*
- Shoja, M. M., et al. (2020). *A Guide to the Scientific Career : Virtues, Communication, Research and Academic Writing. [Chichester], Wiley Blackwell.*
- Bouchrika, I. (may 2021) “What Is Empirical Research? Definition, Types & Samples.” *Research.com, research.com/research/what-is-empirical-research.*
- Naoum, G. (2009) *Dissertation Research and Writing for Construction Students. Oxford, Elsevier/Butterworth-Heinemann*

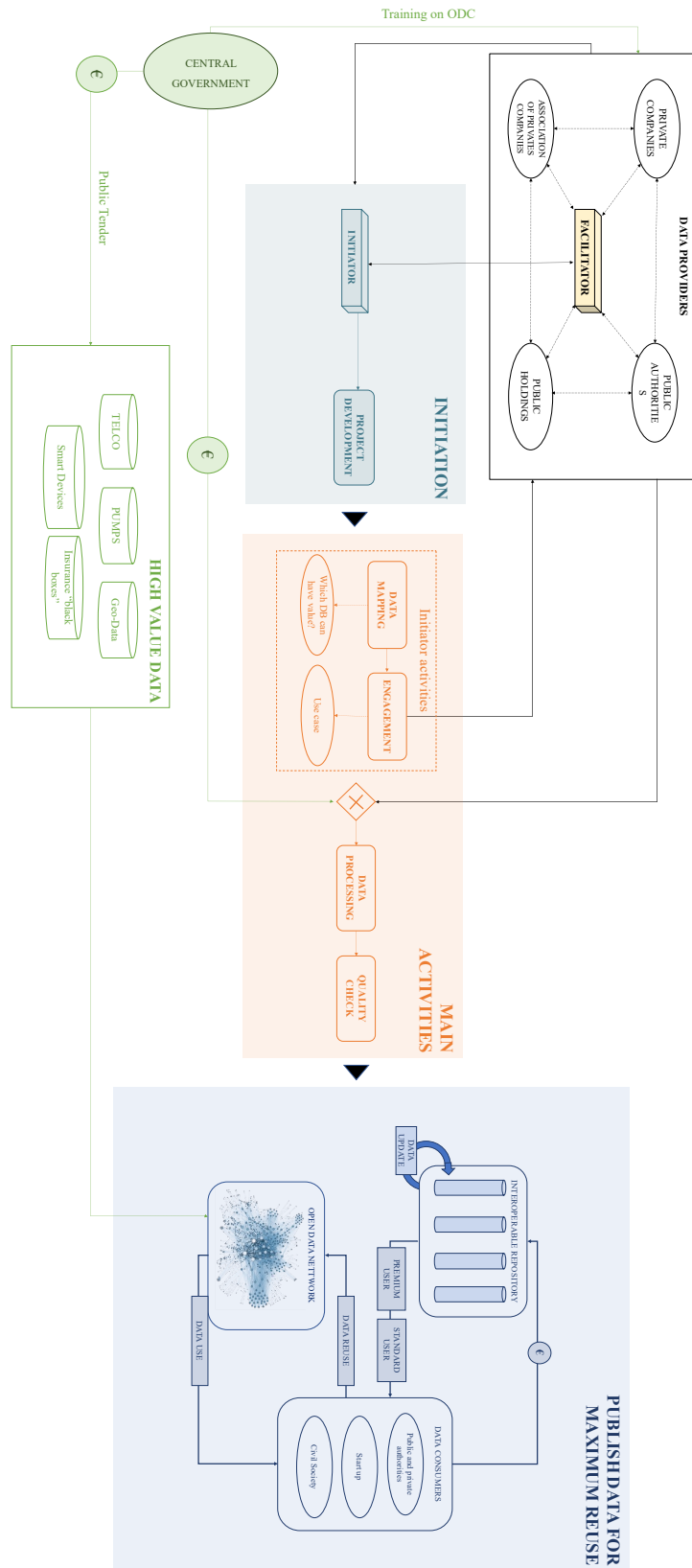


Figure 3: Operational Model

# 1.Introduction

The new trend of digitization implies the use of disruptive technology to improve social welfare, productivity, and value generation. Different studies delve into this concept. Strategic foresight studies, for example, have been prepared by several national governments, international organizations, and sector groups on the topic of digitization. Moreover, some other studies have shown that digitization has yearly growth and quick penetration. At the same time, however, some obstacles are slowing its spread, including inadequate or excessively heterogeneous company structures or cultures (Ebert et al. 2018).

One of the main objectives of digitization is to increase the satisfaction of civil society. So, to develop better user experiences for civil society, digitization entails a comprehensive redesign of rules, procedures, and services. Therefore, the final goal is to improve the efficiency and effectiveness of government services by completely revamping them to accommodate consumers' evolving needs. For this reason, the consumers of digital services are at the center of efforts to make this process more efficient. At the basis of these efforts there is the development of new increasingly powerful, sophisticated and dematerialized technologies. On the other hand, one of the most common errors is to just attribute the digitization to the idea of technology or to a fresh approach to marketing strategy. It is crucial to remember that IT and technology are only the tools through which these shifts take place. It is insufficient to transform merely the employed organizations and tools. In order to keep innovating in business, one must focus first and foremost on people.

Therefore, it is fair to assert that digitization is a multifaceted and complex process. Our thesis project falls under a specific branch of digitization: open data, data that can be accessed, modified, re-used, and freely distributed (Bauer, Kaltenböck 2011).

## Introduction

The expression "Open Data" is being used more and more frequently, but it is intriguing to know where it came from. Robert King Merton, who is regarded as the father of the sociology of science, first theorized the concept of the common good and open access when he stated: "*Each researcher must contribute to the "common pot" and give up intellectual property rights to allow knowledge to advance*" (Merton et al. 1973). Hence, Open Data had its roots in scientific practice long before it became a technological issue or a political movement and the first professionals to recognize the advantages of transparency and data sharing were researchers. The interaction of this scientific notion with the principles of open source and free software resulted in the idea of Open Data as we know it today. Hence, the OD format was born in the scientific field as an open, interoperable and machine-readable format. In the 21st century, then, OD has quickly evolved from a specialized area of study to a component of mainstream world policy and global government-led Open Data projects have proliferated. On the contrary, technological or civil society experiments utilizing data to enhance governance have been spreading gradually. Subsequently, at an international level, the Open Government strategy concept was devised in 2009. The President of the United States Barack Obama, in fact, released a "*memorandum*" on Open Government and transparency to the executives of his administration on the day of his inauguration and as a first action. Due to its potential to empower civil society, and companies, and change how the government provides public services, OD are a valuable resource. Greater accountability, the provision of new, higher-quality services, lower operating costs, and the encouragement of open innovations in both government organizations and enterprises will all be made possible by the opening up of government data (Manyika et al., 2013). In this contest, many public and private entities have started to recognize the inherent benefit of OD, and an increasing number of projects have emerged with the following goals:

- To guarantee adherence to Open Data regulations and requirements.
- To promote and accelerate the usage of OD.
- To ensure that public access to government data and information.



## Introduction

- To provide a single point of access for OD in the public sector.

Finally, it is possible to assert that, in recent years, the value of sharing data in an open way has seen increasing success, and there has been an acceleration and increased attention following the pandemic. The three key pillars that led the process of establishing the OD culture were the ideas of data reuse, data accessibility, openness, and widespread engagement. According to the 2021 Capgemini report on the maturity level of the different countries on the use of Open Data has demonstrated that national policies for the management of public data, the impact of Open Data on the nation's economy, the resilience of Open Data portals, and the quality of the available data are all significantly higher than in previous years and in continuous improvement. Therefore, it is possible to state that there is a growing interest on the part of central governments in Open Data and large amounts of data are increasingly being released into the public domain by numerous governments, without fee or administrative costs. This thesis work is embedded in this context and aims to analyze the real interest both on the part of Public Administrations and private and attempts to map the interactions between the various data providers.

### 1.1. Topic and Context

The proposed research is a component of the larger GIOCOOnDa project. The GIOCOOnDA (*Gestione Integrata e Olistica del Ciclo di vita Degli Open Data*) project fits into the context described above and it officially starts in 2019. The project is managed by Politecnico di Milano and Scuola Universitaria Professionale della Svizzera Italiana (SUPSI) within the partnership that involves Provincia di Brescia, Provincia di Lecco, EasyGov, Varese Web and Fondazione Bruno Kessler. This project, which is funded by European Union Program, intends to add value by creating information products based on the reuse of Open Data that is freely available to the public. It specifically intends to contribute to the conception, development, and implementation of the infrastructure, technological tools, process architecture, and monitoring of the release of government data in an open format. In this regard, GIOCOOnDa objective is to enhance the Insubric Region's Italian and Swiss Public Administration's capacity for coordination and collaboration as well

## Introduction

as cross-border governance by creating a LOD platform. The first output of the GIOCOOnDa project has already been created: the GIOCOOnDa LOD platform, which was born with a high focus on the tourism industry following the project criteria. The earliest data available comprised information about historical sites, lodging options, and the environment and over time the Open Databases have been expanded. The platform is primarily targeted at domain and ontology specialists, who can authenticate in order to use the platform to generate and alter datasets. The datasets generated by the LOD platform are accessible through a public portal as well. New datasets can be submitted by Public Administrations as the platform also functions as a simplifier of the process for LOD conversion.

The GIOCOOnDA project, which has almost come to an end, has achieved additional results as the outcome of the platform's release in 2019 and the continued addition of datasets to it. In addition, an impact assessment model was created and is still widely used in Public Administration, and awareness-raising activities as well as technical and management training sessions for PAs were held.

The project finished in 2020, but due to the COVID pandemic's drive toward digitization, it was decided to construct an additional module that will end in 2023. The thesis inserts into this additional module (WP3 MAC) that aims to improve the analysis and data representation tool created within the GIOCOOnDa project by making it accessible to a broad and diverse audience through the use of a model that makes it easier to identify priority areas of intervention. The audience may be represented by local administrators, managers of the public sector, businesses and traders, but also citizens. More specifically, this work strategically fits into two of three tasks (Figure 1) that form the added module and they are (1) conducting the interviews to understand the actual information needed and the propensity of the organizations to share their database. Then, based on the existent literature and the collected information from the interviews, developing an interaction model (2) among various data providers within an Open Data ecosystem.

### 1. Analysis of new needs

- Sentiment and emotion analysis
- **Interviews**
- Focus Group

### 2. Enrichment of the LOD platform

- Publication of new datasets
- **Study of the interaction models between public and private authorities**

### 3. Improvement of LOD access tools

- Dashboard for data visualization
- Development of the user interface

*Figure 1: Tasks of the WP3 MAC module*

## 1.2. Thesis' focus and scope

Even though the environment in which the GIOCONDA project fits appears to be developing and receives a lot of interest from central governments, there are still a relatively small number of data providers who are actively engaged in the process of opening up their databases and keeping them updated. Moreover, the number of entities collaborating with each other by sharing and integrating databases to create value appears to be small anyway. Therefore, this thesis aims to study in depth the interaction patterns between private companies and public entities. In particular, it focuses on the timing and technical-organizational ways of collecting, uploading and updating data in terms of agreement with suppliers. The objective is to comprehend what best practices and success factors have resulted in the most effective utilization of the sharing process based on the information acquired. Critical elements that contributed to difficulties or shared failure are also examined at the same time. Based on this information the goal is to develop a sharing model

## Introduction

in order to take advantage of best practices and avoid running into the critical factors to increase the probability of success of the project and entice different actors to take part in it. The vision that guided our work was to generate greater participation and propensity towards the Open Data ecosystem through the use of the model developed in this thesis project. The model aims to demonstrate the best method for undertaking the path integration and opening of databases in the best possible way and to show the relative benefits for the participants and for the whole of civil society.

For the reasons previously mentioned and in agreement with the GIOCONDa initiative, the following research question will, from now on, serve as the study's primary focus:

*Is it possible to develop a model for Open Data sharing across public and private organizations?*

### 1.3. Structure

Before entering in the core of the study, a quick overview of the way in which it is organized is given so that readers won't get lost in the analysis. An executive summary, and different sections that analyze the research's findings, and a summary of the study results make up the thesis.

At order to have a better understanding of the processes and methods employed in each stage of the work, SECTION 2 provides further detail on the methodology used to perform the whole study. Then, the state of the art is covered in the sections that follow. The results of the current literature are summarized in detail in SECTION 3 and are organized by sector. The main commonalities in terms of the advantages and critical issues raised by the research are then provided in the last paragraph of the same section.

The material gathered during the interviews is all presented in a structured model in the SECTION 4, in accordance with the emerging analytical dimensions.

Subsequently the SECTION 5 is strongly related to the previous two; the goal of this section is to tie together all the pertinent data that was derived from the literature

## Introduction

and the interviews in order to establish the groundwork for the public-private interaction model's design phase.

Then, the text moves towards the main output of the thesis work. The public-private interaction model that was created to support public and private entities in the process of cooperation, integration, and database openness is presented in SECTION 6.

Finally, in SECTION 7 the conclusions, the contribution of the study presented here, and possible further research are presented.

## 2. Methodology Of the Work

This section focuses on deeply describing the methodology adopted in writing this dissertation. In particular, three different paragraphs are presented. The first concentrates on the methodology used to develop the Literature Review, the second on the methodology used to conduct the empirical analysis, and the last on the methodology used to conduct the development of the model, the main objective of the research.

### 2.1. Literature Methodology

The Systematic Literature Review (SLR) process entails recognizing, analyzing, and interpreting the most relevant research on a specific topic based on predefined criteria (Kitchenham 2006). It was decided to use this systematic approach to conduct the literature review because it allows for the most objective and fair analysis possible. However, it was later realized that systematic research would not be able to produce all of the desired results, so a Narrative Literature Review (NLR) was also developed. The methodology used to conduct these literature reviews is discussed in detail in the paragraphs that follow.

#### 2.1.1. Systematic Literature Review

According to Kitchenham (2006), the systematic literature review consists of three main stages, followed by sub-stages. The three major phases are planning the review, executing the review, and reporting the review.

##### **Planning the review**

The main goal of this phase is to understand why a literature review is required and define a plan with the essential procedures that have been used to conduct the review (Kitchenham 2006). Concerning the first scope of the planning phase, in general, the need for the literature review is fundamental because it allows to fully comprehend the actual state of what has been done in a specific context (Kitchenham 2006). Regarding this dissertation, the literature was necessary to

## Methodology Of the Work

understand the surrounding of Open Data and of the Public-Private Partnership, but more in detail, it was essential to apprehend the level of utilization of OD in the PPP and the perspectives of different sectors on these topics, with the main aim of verifying the existence of an integration model. The development of a sharing model is the main goal of the research work. The following parts explain in detail what it is and why it needs to be developed.

As a result, this plan was crucial to the formulation of the research question because it will serve as the foundation for all SLR techniques (Kitchenham 2006). Before entering into this topic is important that a premise be made. This work was introduced in an Interreg Italy-Switzerland project called GIOCOOnDa, funded by the European regional development fund for the years from 2014 to 2020. The main objective of the project was to improve spatial governance through the use and sharing of Open Data between Public Administrations (PA) and private companies. At the end of the project, a portal for Linked Open Data, called Gioconda LOD was made available, whose exposed datasets were already reachable from the Open Data portal of the Lombardy Region. More in detail, the aim was to involve the territories through information and training, as well as to integrate the extensive and relevant databases of private individuals with those of the Public Administration. Due to the pandemic, it was decided to create a new project module that was divided into three major phases and activities: analysis of new needs, enrichment of the GIOCOOnDa platform, and enhancement of tools for accessing Linked Open Data (LOD). This work was supportive in the development of the first two phases. In particular, in the first part, the goal of conducting the interviews to understand the need for data and the data availability was our responsibility, as it is fully explained in the methodology section on the empirical part. And, in the second part, the final objective of developing a model of integration between PA and private companies was also our responsibility. Having said that, the research question for this study was already known since it coincides with the project objective. Thus, the requirement to develop this work has been simply confirmed through the literature review. So, the research question is:

*Is it possible to develop a model for Open Data sharing across public and private organizations?*

## Methodology Of the Work

Concerning the second scope of the planning phase, following the definition of the research question, it is possible to proceed by developing the review protocol. It defines the method by which the review has been conducted, allowing for replicability and reducing the possibility of bias. The protocol specifically defines how the study search has been conducted and which research has been included or excluded from the analysis (Kitchenham 2006). As the first step in protocol definition, eligibility criteria were defined, followed by keyword extraction for the creation of a search string, always starting from the search question. The eligibility criteria are detailed below.

1) All articles written prior to 2006 were excluded.

The decision to limit articles to 2006 was made because the concept of Linked Open Data, on which the project is based, was not recognized prior to that year. Tim Berners-Lee, in fact, invented the term, as we know it today, in 2006 (European Union 2019).

2) All articles not written in Italian or English were excluded.

Simply to allow for the best understanding of the articles, only those in languages that can be fully understood were retained.

3) All articles not in the final stage were excluded.

Exclusively articles in the final stages of the approval journey were picked to ensure greater reliability of the information. This guaranteed that only scientific papers were being analyzed; gray literature analysis, on the other hand, was done as needed, at a later time.

These criteria were used in the various research, that has been done and described in the following step.

### **Executing the review**

After defining the eligibility criteria, the keywords for the search string were determined at this point. The following keywords have been identified: integration, public, private, partnership, data, and LOD. Moreover, it has been decided to limit the research to some sectors of particular importance: transport, health care, environment, culture, tourism and government. These sectors were selected primarily for different reasons: the GIOCOOnDa project has already begun in the



## Methodology Of the Work

fields of tourism, culture, transportation, and the environment; hence, it is obvious that these areas, which are closely tied to the project itself, were chosen to be included in the study. Then, the government sector was therefore included since it is one of the sectors where Open Data is utilized the most<sup>1</sup>. Moreover, it is one of the businesses where the idea of Open Data was born under the concept of Open Government Data and, for this reason, it is also an industry from which a number of considerations and best practices may be drawn. Additionally, this industry is ideal for integration because it consistently collaborates with both private citizens and private companies. Health care was then added since, by definition, it is a sector with a large number of public and private authorities that would greatly benefit from a sharing mechanism. Furthermore, although Open Data is not especially common in this industry, it was believed that interesting applications could be discovered to illustrate its value.

So, based on a combination of Boolean ANDs and ORs operators and based on these considerations, the first search strings were developed:

*(ABS (integration AND public AND private AND data) AND ABS (transport OR health OR tourism OR culture OR environment OR government)).*

It was decided to conduct research on the Scopus database using both abstracts and keywords and titles at first, but it was later discovered that doing so was leading to an unsatisfactory result, so it was decided to limit the search to abstracts only to make it more relevant. Once the previously defined criteria were applied to the search, the result was 421 articles. The next step was to read the abstracts and do a preliminary skimming. By this point, we realized that the number of interesting articles had been reduced to a quite low number of 90, which became only 15 after reading the full article (as can be seen in Figure 2). The primary criteria used for deciding whether or not to keep an article were principally three:

---

<sup>1</sup> [Open Data Use Across Sectors \(opendataimpactmap.org\)](http://opendataimpactmap.org)

## Methodology Of the Work

- The presence of a collaboration between private companies and Public Administration;
- The presence of information sharing between the two parties;
- The presence of a case study.

Mainly applied articles, i.e., case studies, were chosen because this research is precisely about the development of a model that will be applied to reality, and moreover in a rapidly evolving context. As a result, rather than theoretical arguments, it is important to reflect on what has been done.

Anyway, despite a good starting point, the number of papers chosen using these criteria was remarkably low.

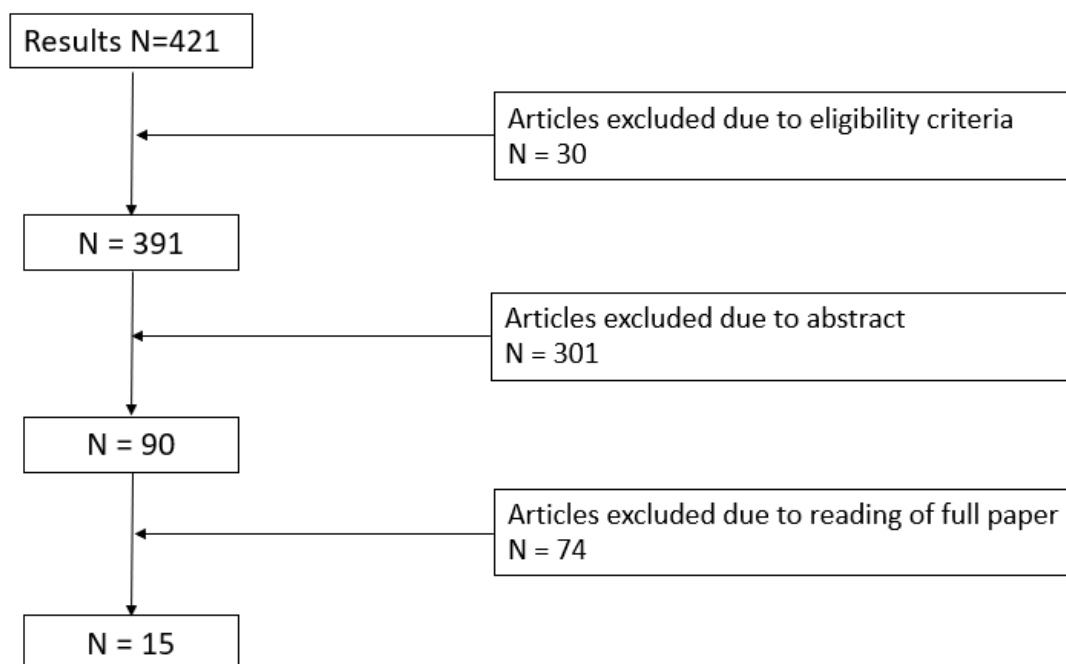


Figure 2: First search string result on Scopus

Not satisfied with this result, it was decided to try using the same string on another article search database, Google Scholar. For this research, the same criteria as previously listed were used, and the result was only 27 articles, which were reduced to 5 after reading the abstracts and 1 after reading the full papers (as can be seen in Figure 3).

## Methodology Of the Work

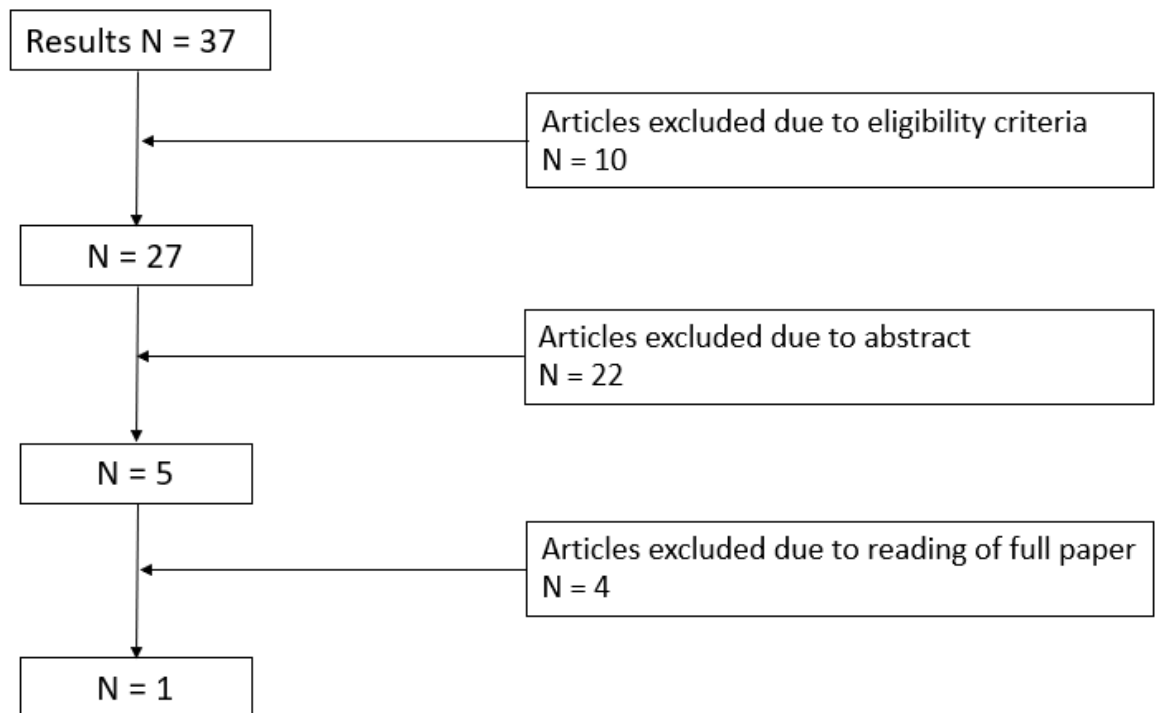


Figure 3: Second search string on Google Scholar

Clearly, even this research did not produce the desired results. As a consequence, it was decided to conduct more focused research on data use, its reusability and a particular emphasis on the USA due to its advantageous stance on these matters (Mastroianni 2018). The three strings that were created were as follows:

- 1) *(ABS (data) AND ABS (reusable) AND ABS(LOD)).*
- 2) *(ABS (LOD) AND ABS (public) AND ABS (private)).*
- 3) *(ABS ("linked Open Data") AND ABS (USA)).*

These strings were also used on the Scopus scientific article search database, and the same principles as previously listed were applied to the results obtained. The total result was 52 articles, which became 17 after reading the abstracts and only 5 after reading the full papers (as can be seen in Figure 4).

## Methodology Of the Work

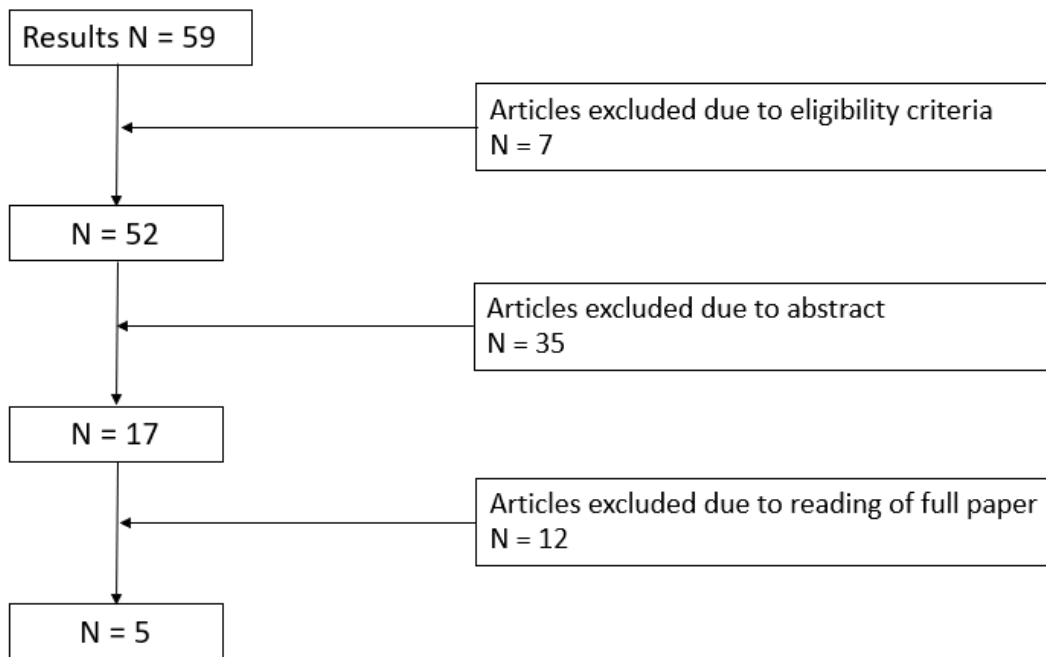


Figure 4: Third search string on Scopus

These selected articles were then extracted from their respective research databases to Excel in order to then be organized, sorted, and kept up to date. At the end of the search, these documents were uploaded on Refworks, a tool for the management of the references, to have the ability to handle them easily and safely.

### Reporting the review

As a result, the number of papers discovered was only 21. This result can be explained by the fact that the concept of Linked Open Data, particularly in Public-Private Partnerships, is still in its early stages, making it difficult to find papers that combine these two concepts.

Year of publication	Number of articles
2009	1
2011	1
2012	1
2013	1
2014	1
2015	4
2016	3
2019	2
2020	1
2021	6

*Table 1: Publication years SLR*

In fact, as we can see from Table 1, most of the papers have been published since 2015. It is important to specify two things. First of all, the years not listed in the table were excluded because no papers published in those periods were selected. Then, the research was done at the beginning of 2022, for this reason no papers were founded. All this could be a sign that this topic is becoming increasingly important and, consequently, our research could be useful in speeding the process of diffusion of Linked Open Data in the Partnership Public Private.

In terms of the sectors involved, the situation is depicted in Table 2 below.

## Methodology Of the Work

Sector	Number of articles
Environment	3
Health Care	5
Smart City	3
Transport	6
Tourism	1
Culture	1
Government	2

*Table 2: Publication sectors SLR*

The number of papers found by segment was slightly different than expected; in fact, the number of papers related to the healthcare industry was higher than those related to the government one, which is unusual given the government sector's higher prevalence of Open Data. This outcome could be explained by the fact that the healthcare-related publications are more recent - two from 2021, one from 2020, one from 2016, and one from 2019 - than those from the government sector, which are from 2019 and 2015. This may be justified by the fact that, unlike the government, which has stabilized somewhat further in this context and thus the related research has decreased, the benefits of health care have emerged more recently and therefore there are more studies on it.

Also, an analysis of the authors and citations of the articles was not performed because, as previously stated, the majority of the articles are case studies rather than theoretical in nature. As a result, we thought it was more important to focus solely on years of publication and sectors of application rather than their position in the scientific debate.

### 2.1.2. Narrative and Grey Literature Review

Since the number of papers selected was limited, it was decided to conduct a Narrative Literature Review (NLR). This type of literature is used when there is already knowledge of the topic and when there are fundamental papers from which to start (Shoja et al. 2020). Based on this, it is possible to proceed in three ways: using keywords to search the various databases unsystematically; starting with the fundamental papers and analyzing the paper cited in the latter; and looking for recent publications that cite those papers. It was feasible to use this approach because we already had a basic understanding of the topic and the key papers that addressed it. So, all three different search methods were used in order to gain a significant and satisfying number of papers. Indeed, after this work, the number of new papers selected was 33.

## Methodology Of the Work

Year of publication	Number of articles
2001	1
2007	1
2010	2
2011	3
2012	1
2014	2
2016	2
2017	3
2018	3
2019	6
2020	5
2021	3
2022	1

*Table 3: Publication Years NLR*

Doing the final analysis related to years of publication is possible to see how the number of papers increased in recent years. These findings support what the Systematic Literature Review found, namely the growing importance of these topics. Furthermore, as previously stated, the research was developed in early 2022, which is why there are not many papers published in that year.



Sector	Number of articles
Environment	3
Health Care	6
Smart City	5
Transport	5
Tourism	4
Culture	4
Government	6

*Table 4: Publication sectors NLR*

As for the sectoral analysis, however, the results obtained were more in line with what was expected than with what was achieved with the SLR. In fact, the government sector has the highest concentration, followed by smart cities and transportation, and then tourism and culture, as highlighted by the Open Data Impact Map<sup>2</sup>, with the exception of the environment sector, which has a significantly lower number, which is slightly smaller than expected. Furthermore, in the health care sector, a large number of papers were discovered, which, like SLR, could be attributed to recent growth given the increased presence of articles in recent years. Finally, what can be noted is that in any case, there is a good number of studies, also quite balanced, in the different fields. This may lead to think of a good diffusion of the use of Open Data in recent years and a slight clearing of these concepts to such an extent that they spread a little more homogeneously in the various fields.

---

<sup>2</sup> [Open Data Use Across Sectors \(opendataimpactmap.org\)](https://opendataimpactmap.org)

## Methodology Of the Work

In conclusion, during the study of literature, the moment something was not too clear or wanted to be more in-depth and detailed, the Gray Literatures were consulted. It refers to all sources that have not been reviewed by a commercial publisher (Schöpfel, Joachim 2010).

### 2.1.3. Final Considerations

The total number of selected papers at the end of the various searches was 54. This is a small number, but it was more than adequate and satisfactory for the Literature Review analysis. A summary of the origin of the articles can be seen in Figure 5.

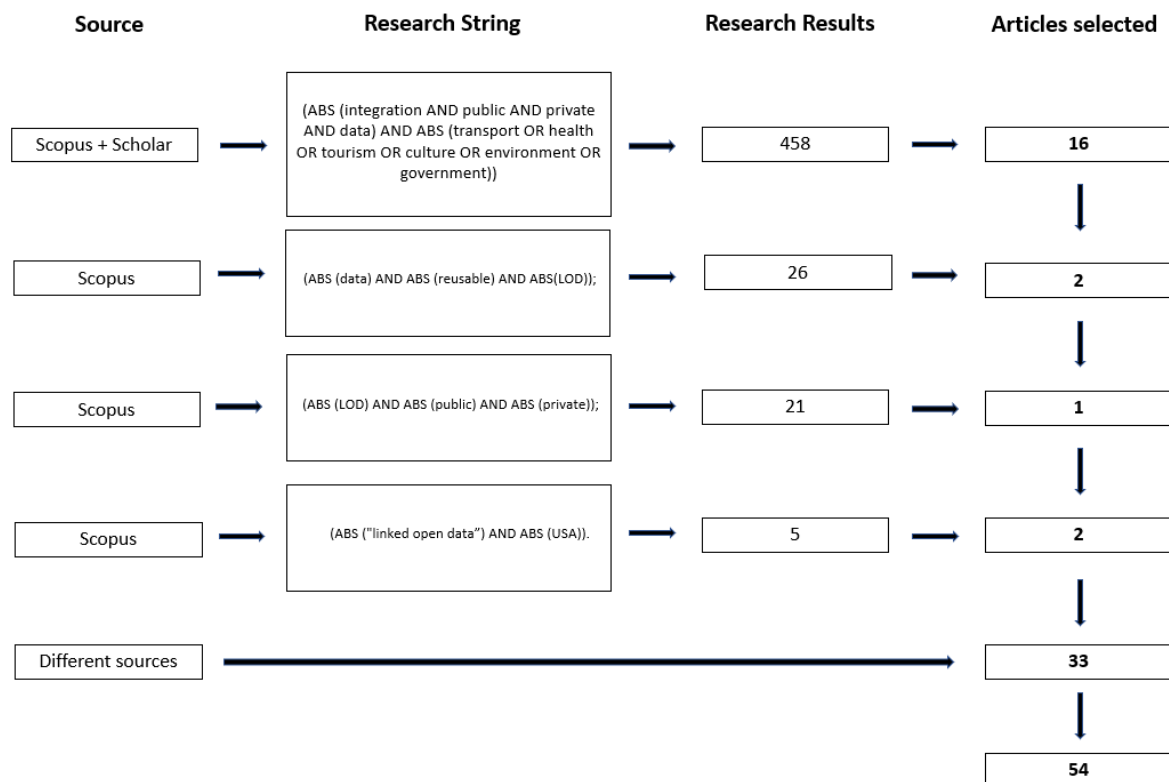


Figure 5: Summary of the articles selected

To summarize, as shown in Figure 6 below, an overall analysis confirms the increase in articles in recent years due to the significant growth of these topics. Always making an exception for 2022, because the research was conducted at the beginning of the year. Moreover, the peak was precisely reached in 2021, most likely as a result of the needs that arose following the covid period, which may have accelerated the Open Data dissemination.

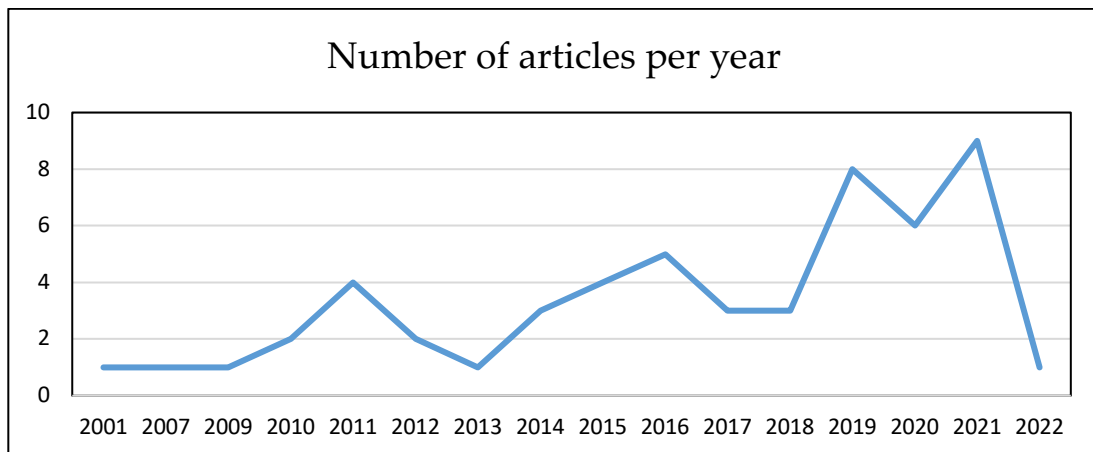


Figure 6: Number of articles per year

Finally, as shown in Figure 7, the articles by sector are fairly evenly distributed, with the exception of tourism and culture, which are still in development, as confirmed by the Open Data Impact Map. Similarly, as previously stated, what was not quite in line with what was expected is related to health care. In fact, the number of items discovered is extremely high, comparable to the transportation sector. In the preceding paragraphs, it was explained why this might be the case.

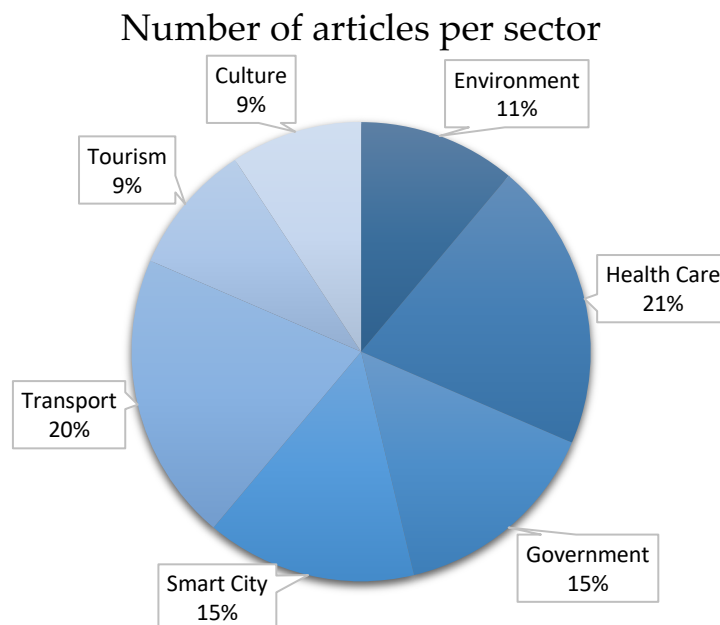


Figure 7: Number of articles per sector

To conclude, what can be seen from the analysis of the literature methodology is that the topic of OD in PPPs is a recently developed topic, but at the same time it is a relevant matter that is becoming increasingly widespread in different sectors and applications. For this reason, the proposed research push towards these issues and it can be important in order to boost the OD diffusion in a more coherent way.

## 2.2. Empiric Methodology

Empirical research is simply a study based on evidence and on actual facts that have been analyzed and studied. It can be done in either a qualitative or quantitative manner (Bouchrika 2021). For this study it was decided to conduct qualitative empirical research and this section is focused on the methodology used to do that.

### 2.2.1. Qualitative research

To have a more clear and detailed vision of the current situation it was decided to study in a qualitative way the concrete interaction models among enterprises and public authorities. The main objective of qualitative research is to emphasize and study the meaning, experiences, behaviors, and attitudes of different subjects involved (Catherine Dawson, 2009). This type of research allows obtaining detailed information about the analyzed specific context, but in a flexible and dynamic way. It was decided, therefore, to use this type of research rather than a quantitative one because the context studied was not yet well defined and only in this way was it possible to obtain a comprehensive approach towards the study of the topic. Furthermore, thanks to the direct interaction obtained through the interviews, which are analyzed later, this methodology allowed us to truly connect with the research participants and fully understand the interesting topics.

Going more in detail into the analysis, through qualitative research it is possible to gather information in two different ways: exploratory research and explanatory research (Naoum 2006). The former is used to analyze cases that have not been thoroughly studied, while the latter is used to understand why and how insights are required and should be made (Tegan 2021). In this context, it was decided to use the exploratory research model.

### 2.2.2. Exploratory research model

Exploratory research, as mentioned above, is a research method that investigates a particular topic when there is limited information available. It can help in gaining a better understanding of a subject, determining how or why a particular phenomenon occurs, forecasting future events, and suggesting how to improve the current situation. It was decided to use this method because the research topic was developed, but not yet well defined. For this reason, through the exploratory research methodology, was possible to analyze what was already being done for then comprehend the patterns, the critical and the enabling factors with the aim of formulating new hypotheses for the further development and evolution of the topic itself. Indeed, the goal of our research was to provide a push toward the use of Open Data in PPP contexts through the definition of a new model.

So, the objectives of the research proposed in this dissertation were as follows:

- Understanding the current data needs;
- Understanding the willingness of authorities to share their databases;
- Understanding the interaction models among public and private partners.

Then, the approach for conducting this kind of research can be carried out in different ways and this depends on the objective of the analysis and on the kind of data needed. For this study, it was decided to proceed with data collection according to a multiple case study approach.

### 2.2.3. Multiple case study approach

When you wish to learn more specifically, contextually, in-depth information on a particular real-world topic, a multiple case study approach is a suitable research design. This approach made it possible to investigate the main features, meanings and ramifications of the cases. All the information that emerged from this working methodology was then used during the design phase of the interaction model.

Typically, qualitative techniques are used in case study research designs (Naoum 2006) in order to gather information. In fact, according to this, the model used for data collection was the Semi-Structured Interview (SSI).

### 2.2.4. Semi-Structured Interviews

As just mentioned above, it was decided to use the SSI. This method allows to ask open-ended questions and to know the independent thoughts of the interviewees. It was done in order to exploit an evidence-based approach for the study and the interpretation of the actual status of art in LOD context. To this end, in collaboration with Scuola Universitaria Professionale della Svizzera Italia (SUPSI<sup>3</sup>), Brescia Mobilità<sup>4</sup> and Fondazione Bruno Kessler (FBK<sup>5</sup>), twelve interviews have been realized that have involved both public and private authorities of the Lombardy region and the Italy-Switzerland cross-border areas. Semi-Structured Interviews are distinguished by a flexible and interactive questionnaire, but with the same backbone in order to be adaptable and to facilitate the interview's data analysis and feedback collection phase. As a result, it was necessary to define the initial interview protocol, which was then tailored to the various interviewees. In fact, the choice of this method was necessary to have on the one hand very specific information from all interviewees thanks to a set of predefined questions, but on the other hand to ensure flexibility towards the different interviewees who in our case belonged to very different contexts (such as associations of individuals, private companies or representatives of Public Administration).

The method of Semi-Structured Interviews involves six steps (Tegan 2021): defining the objectives; designing the questionnaire; defining the participants; defining the conducting method; conducting; and analyzing.

---

<sup>3</sup> [SUPSI](#)

<sup>4</sup> [Brescia Mobilità](#)

<sup>5</sup> [FBK](#)

### 1. Defining the objective

It was first necessary to define the purpose of the interviews, that is, the information that would be obtained from them. It was to understand what and how much data is currently available, as well as what and how much data is required to start a win-win data sharing process for the stakeholders. It was also to understand the willingness of the interviewed to collaborate to open their own databases. These information, along with what was reported in the literature, served as the foundation for the model's development.

### 2. Designing the questionnaire

The following step was the definition of the design of the questionnaire. It was decided to be split into four main sections:

- 1) *Introduction*: to understand how data management and exploitation fit into the corporate business model.
- 2) *Data availability*: the focus of this part was on the available information assets of the enterprise (or Public Authorities) and its tendency to open and share datasets.
- 3) *Data necessity*: the main topic of this section was related to the understanding of the outside information needed thanks to an integration process in order to improve the actual situation.
- 4) *Criticalities, benefits and future opportunities*: the last part, was concentrated on the understanding of the difficulties, opportunities and future possibilities of the integration of different datasets from the point of view of the interviewees.

Regardless of how thoroughly the design structure was prepared (Annex), the questionnaire always needs to be considered a work in progress due to the type of interview chosen. Indeed, the questionnaire was always subject to changes that were made based on the interviewee, as previously mentioned, but also in the field as a result of the feedback quickly accumulated (Newcomer et al. 2015). As a result, the order of the questions was reconsidered several times, and the manner in which some questions were asked was revised.

So, the questions of the interviews have been designed to stimulate the interviewees to pay more attention to their own information assets and stimulate integration among data from different sectors.

### **3. Selecting the participants**

Participants were identified and involved at this point. Representatives from various industries as tourism, transportation, and culture, were chosen. The experience came from both the public and private sectors, as well as from the regions of Lombardy and Canton Ticino. The regions were determined solely by the territory involved in the GIOCOOnDa project.

Interested parties were contacted via email, and one-hour sessions with each one was scheduled. Following the research phase, twelve representatives decided to join, including three from the private sector, one from an association of private companies, two from the public sector, three from the private sector controlled by the public, and three from associations formed by both private individuals and public companies.

In this manner the study involved all the possible actors that can actually be introduced in a project for sharing data, allowing for an understanding of all their different points of view and of their position towards these topics.

### **4. Defining the conducting method**

It was decided to conduct the interviews in a direct manner through conversations rather than through paper questions. This is because direct conversation avoids the over-construction of written responses, which lose validity, and instead allows for more spontaneity. Moreover, to make calendar synchronization easier, interviews were conducted via video call on Microsoft Teams. Other than the aspect just highlighted, video calls had the very positive aspect of making the interviewee more comfortable, but on the contrary did not fully allow for attention to the nonverbal aspects as well.



### **5. Conducting the interviews**

As soon as the date and time of the interview were set, the protocol was adapted to make it as relevant as possible to the world from which the representative was coming. This was done, as also mentioned earlier, to allow for the most objective and fair data collection possible.

At the beginning of the interview itself, consents were asked to analyze and use what would emerge. In some cases, permission was also asked to record the video call so that the material could be accessed at any time. In contrast, permission was not asked to make explicit references to the names of the interviewees and the relevant company or entity and, for this reason, anonymity of them is maintained within the dissertation. After that, the GIOCOOnDa the thesis projects were presented in each interview in order to make the participant aware of the context and objectives of the work. Once all these rigorous introductions were made, the questions were submitted.

### **6. Analyzing the interviews**

Finally, all of the collected information had to be analyzed in order to extract all the relevant knowledge for the model development.

To this end, it was created a data collection model to arrange and store information within in (Figure 8) to summarize in brief the key feature of each interview. So, once all the interviews were conducted, they were transcript and then returned to the interviewee for confirmation of what was understood and, if necessary, the addition of information deemed important. This ensured that the model's subsequent development would be based on hard data. So, at this point the keys information were collected and arranged within the common data collection model to have all the information organized. The aim of the visualization framework was to find patterns, meeting points and similarities among the different authorities belonging to different sectors. Indeed, by putting together the different experiences of the interviewees it was possible to extrapolate common practice in term of implemented policies, benefits, and limitation of opening datasets.

## Methodology Of the Work

Figure 8 shows a summary of the data collection model. The first column represents the four areas of intervention that were considered fundamental for gaining all the information needed for the development of the model; instead, the cells summarize the key aspects of each topic that might have been useful for the purpose of the analysis. In this section, it is shown only the methodology to reach the desired information, while all the results and the meaning of each dimension and metrics are described in detail into section 4.

Dimensions	Metrics
DATA	Available Data
	Data Enhancement
	Information Need
	Critical Factors
	Enabling Factors
RELATION WITH THIRD PARTIES	Data Providers
	Data Integration
	Limitations and Difficulties
	Best Practice
OPEN DATA	Open Data Culture
	Publishing Venue
	Critical Factors
	Enabling Factors
POSSIBLE SCENARIOS	Actual and Future Projects
	Pandemic Impact

Figure 8: Data collection model

### 2.3. Model development methodology

The creation of the model was the main objective of this thesis project. The creation process began immediately with a diagnosis of the literature's findings in order to comprehend the state of the art and examine any existing interaction models. More in detail the model's design phase made use of the findings from the literature as well as the data gathered through interviews.

## Methodology Of the Work

All potential data providers have been divided into reference categories as a first component so that any actor wishing to take part in an integration process can identify within the model developed. New figures were then defined (initiator and facilitator) and added to the process based on the critical issues gathered during the analysis phase. Once the roles of all the actors involved in the process were clarified, the main activities to be carried out were defined in order to guarantee efficient and effective collaboration. Finally, the data flow that should exist once the LOD has been created and entered into the Open Data ecosystem was devised.

At the end the model was officially presented at a special Webinar hosted by the Digital Observatory of the Politecnico di Milano. By receiving the participants' approval, it was possible to obtain the model's first validation as a result of this presentation.

In conclusion, it is important to point out that this section has only analyzed the methodology of the work done. Indeed, in the next sections, the literature and then the empirical part are analyzed in detail. The analysis of the interview results and the development of the model are developed too.

## 3.Literature Review

### 3.1. Introduction

The literature review was conducted using a systematic method, and the research question, as explained in the methodology section, is consistent with the GIOCOOnDa project's goal of developing a new common data integration model. Therefore, the objective of this analysis is twofold. First, it seeks to highlight a gap in the literature, that is the lack of a widely used data sharing model between Public Administration (PA) and private companies; second, it seeks to understand the availability and demand for data in different sectors so as to better understand the possible features of a common data integration model. The analysis was carried out in a sectoral manner. The goal for each sector, environment, health care, tourism, culture, transport, Smart Cities and government, was to understand where they stand on the issue of data sharing. Sought information relates to what data involved parties require and what data they need, how the integration of data from both public and private sources can benefit them, and what problems have been encountered in data sharing. Some case studies were analyzed to have a more concrete approach in studying attempts to develop integration models. Consistent with our research goal, the study of the literature review highlights the fact that models are not be present in all sectors, and if they are, they are inconsistent with one another. As a result, this analysis is critical to the subsequent development of the proposed model, as it demonstrates its significance and necessity. To that end, three main different basics definitions are provided firstly in order to be able to better understand what is explained in the later paragraphs. Moreover, a brief overview of the European and Italian landscape is made, before delving into the individual sector analyses, to understand what different regulations exist on the topic of data sharing.

## 3.2. Main definitions

Three main definitions are required to fully comprehend the literature analysis. The first refers to Open Data (OD), the second to Linked Open Data (LOD) and the third to Public-Private Partnerships (PPP), which are named extensively afterward.

The term Open Data refers to data that can be accessed, modified, re-used, and freely distributed (Bauer, Kaltenböck 2011). To be considered as such they should be comprehensive, kept as basic as feasible to avoid aggregation, updated frequently, available to everyone, understandable and processable by machine, as well as non-proprietary and license-free. Hence the main advantages are the access and the free reuse of the data by any Data Providers.

To get the most out of Open Data, the more sophisticated concept of Linked Open Data (LOD) was developed. In fact, this term refers to Open Data that can be linked together to generate new knowledge, applications, and services. The use of data in the form of LOD allows for the maximum reuse of information. Sir Tim Berners-Lee was the one who introduced the concept of LOD, as explained in the methodology section. Furthermore, in 2010, he created the “5 Star Model” (Figure 9), which explains the costs and benefits of each step, starting with simple data publication and progressing to an Open Data format and linked open format (Janowicz et al. 2014).

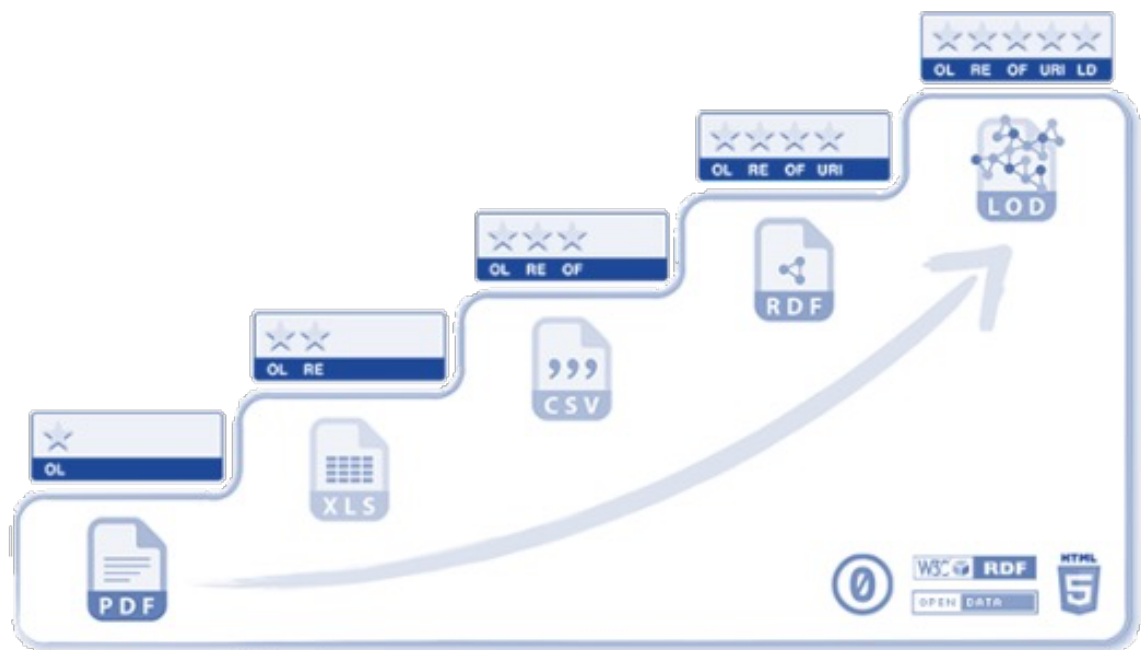


Figure 9: 5-star Open Data ([5-star Open Data \(5stardata.info\)](http://5stardata.info))

The first star represents the simple publication of data on the Web in an open format, which can be represented by a pdf, for example. The advantages of this stage of publication are that this information can be easily published, freely accessed, edited, and shared, but it is difficult to exploit and extrapolate this information outside of the document itself. The second star represents the publication of more complex and structured information, such as an Excel document, that can be consulted openly and understood by the machine. The advantages here, in addition to those of the previous stage, are primarily related to the fact that this information is easily publishable, can be used to perform calculations, and can also be transformed into other formats; however, the data itself cannot be exported from the document as in the previous case because it is dependent on proprietary software. The third, on the other hand, is data published in a non-proprietary format, such as CSV. The additional benefit of this step is that the data can be safely extracted and used without the need for any proprietary software package and thus without cost; however, a problem may arise for the publisher who may have to convert the data in the right format for being analyzed by the software. The fourth star is recognized if URI identification is used to enable web sharing; at this stage,

the RDF format is an example of a publication type. The benefits added at this stage are solely related to the fact that the URI allows for the linking of different sources; however, the understanding of the information may be more complex, resulting in a greater effort in terms of time. Simultaneously, more effort is required for publishing this data because it must be properly prepared, such as by assigning a URI for the purpose. Finally, the fifth star consists of the LOD. It is therefore possible to derive greater benefit from a network of data that are linked one to another, but the problem may be related to privacy issues that must be properly addressed.

Another concept arising from the literature is the necessity of close cooperation between public and private; for this purpose, it is necessary to introduce the concept of Public-Private Partnership (PPP). PPP refers to long-term collaborations between private companies and PA with the ultimate goal of creating a benefit from both parties (Garving, Bosso 2008) sharing risk, costs and resources (Van Ham, Koppenjan 2001). One of the fundamental activities of the PPPs consist in sharing information (Sehgal, Dubey 2019). These data sharing between companies and PA, thanks to the OD and LOD technology, as would possibly be seen in the later paragraphs, is a critical success factor to be able to reach the intended results and a digitization of the different countries.

### 3.3. European and Italian Landscape

On the basis of the use and diffusion of Open Data there is a broader and more radical idea of the so-called Open Government. According to this doctrine, the citizens should be able to have access to all the data used and produced by the government of the Country. An analysis of the current European context shows that there have been efforts by the EU in this direction. To this end, the European Union has defined directives on Open Data and on the re-use of public sector information. The purpose of this paragraph is to highlight some directives that can be considered as an input for the European Open Government and, more in general, for the use of the Open Data in a more comprehensive way.

The key directives introduced by the European Union to promote data reuse and increase territorial openness are summarized in Figure 10.

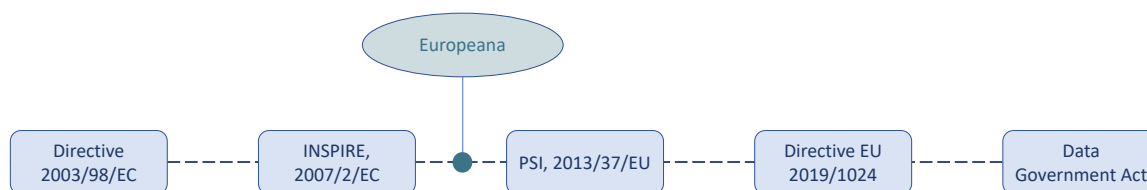


Figure 10: European directives

### Directive 2003/98/EC

The PSI Regulation (Directive 2003/98/EC) was the first directive in Europe to assist the reuse of public sector information. It defined a minimal set of guidelines and practical ways to make it easier for organizations in the EU's public sector to reuse their existing information.

### INSPIRE 2007/2/EC

Slightly shifting the focus on the use of common standards, in 2007, the Infrastructure for Spatial Information in the European Community directive promoted *“an EU initiative to establish an infrastructure for spatial information in Europe that will help to make spatial or geographical information more accessible and interoperable for a wide range of purposes supporting sustainable development”* (INSPIRE, 2007/2/EC). The INSPIRE Directive pushes the Member States to make a greater effort to standardize their databases. Starting from the work done by different countries, INSPIRE has also expanded to different sector, as the public transport network, the environment, the health and safety system and so many others. Nowadays there are rules for the interoperability on thirty-four different themes. Following the INSPIRE directive, the European Union and each of its members have begun to publish unique standard definition for the publication of data. However, the impression is that this great activity of writing standards has not had a strong response from local authorities in the practical application (Caroti, Piemonte 2010). In the Italian scenario, there was a significant lack of feedback from local authorities, that are the main users of the database. The consequence is that the European specifications remain only a steering document and they do not become a normative reference. The lack of adherence to the standards is influenced by several factors: the lack of sufficient resources for the migration of existing data archives to the new formats, the need for training courses for PA employees, but also the fact that often the



standards developed by central authorities do not take into account the specific needs of local authorities (Caroti, Piemonte 2010).

### **Europeana**

Based on this trend towards openness and sharing, a great success was the launch of the Europeana portal in 2008. It is a digital library that is available in thirty languages and has already been digitized by various organizations from the twenty-seven European Union member states. Books, movies, artwork, newspapers, sound archives, maps, manuscripts, and archives from libraries and historical repositories are among its equipment. By keeping the conversation on Open Data portals going by mentioning [data.europa.eu](http://data.europa.eu), the official European data portal that compiles and links Open Data from all of the member states. the European Open Data Portal<sup>6</sup> facilitates the digitization process of PA, and it represents an important step forward for open government process. The portal provides referenced datasets relating to Open Data from 34 European Countries and allows access to public data from all over Europe. It is structured into different categories with 13 levels of classification, such as health, education, transport, science, and justice. It allows all citizens and operators to search, access, and reuse data for any purpose. Hence it is a concrete contribution to the development of many projects in the Smart City of the IoT.

### **PSI, 2013/37/EU**

Following the debut of the European portals for both cultural and open government, a series of rules and guidelines aimed at enabling the release of data and supporting Open Data portals were published. Going in detail in 2013, the Public Sector Information was updated by establishing *“a minimum set of rules governing the re-use and the practical means of facilitating re-use of existing documents held by public sector bodies of the Member States”* (PSI, 2013/37/EU); the goal was to

---

<sup>6</sup> [The official portal for European data | data.europa.eu](http://data.europa.eu)

increase the commercial utilization of data, to promote economic growth, and specifically increase tourism.

### **Directive EU 2019//1024**

The interest from EU has grown over time, the legislation was updated by following the new digitization trend. The new Directive (Directive EU 2019/1024) expanded and improved on the PSI Directive (PSI, 2013/37/EU). The objective was to ensure a better harmonization of national rules and practices on the re-use of publicly funded information. Hence, to facilitate data's free circulation through the interchange of data in a structured and shared dimension, the European Union encourages public organizations to handle data in a format that permits portability. In particular, the 2019 Directive envisions the adoption of this concept through the provision of documents in any pre-existing language or format by public bodies and businesses, encouraging the accessibility of data from the design stage. People and companies will be able to find new uses for this information and develop ever-more-innovative goods and services if it is made available, preferably in a widely used electronic format.

### **Data Government Act**

Finally, it is important to mention the Data Governance Act (DGA) published by the European Commission on 25 November 2020 in response to the public consultation on the European Data Strategy. The consultation was used to gather input for numerous planned projects on data access and reuse as well as to analyze the opinions of stakeholders on the data strategy (including Open Data, data sharing, and data spaces). The consultation included a legal framework for shared European data spaces and an implementing law for a list of high-value datasets under the Open Data Directive. The directive that was developed is focused on two essential ideas in the European market: fair competition and transparency.

After examining the European legislative environment, it is feasible to categorize the development of the Open Data regulations into three major categories:

- Encourage the usage of Application Program Interfaces and the disclosure of dynamic data (API).

- Reduce the exclusions that now permit public organizations to charge more for data reuse than marginal distribution costs.
- Strengthen the transparency requirements for agreements involving public sector information between public authorities and private individuals in order to avoid exclusive agreements. Extend the scope of the directives to include data held by public enterprises, based on a specific set of rules and research data produced by public funding.

### 3.4. Sector Analysis

The findings of the literature review were categorized by reference sector, as was previously mentioned in the methodology section. The sectoral analysis is crucial since it serves to provide details about the internal dynamics of a sector's structure. In fact, dependent on the sector's level of competition, quite diverse behaviors have been found. This section provides a thorough study of the industries that use Open Data and partnerships and have the most potential for growth in the initiatives that come after them: environment, health care, smart cities, transport sector, tourism sector, culture sector.

#### 3.4.1. Environment

Environmental sciences must be thoroughly researched in order to prevent the development of a sustainable society from being held back. To that end, the collection of data from environmental sciences is critical for environmental protection and development (Song et al. 2018). These data, in particular, can be used to reduce environmental risks. The issue is that there is a large amount of data, but it is heterogeneous and diffuse, but it must communicate in order to be useful (Tomas et al. 2015). Two case studies were studied in order to understand how combining heterogeneous data from different sources can be beneficial to the environment, the first conducted by Tomas et al. (2015) on natural risk reduction in general and the second conducted by Magarey et al. (2009), in greater detail, on natural risk reduction related to biosecurity.

Natural hazards are processes or phenomena that can result in death, injury, or other negative health effects, property damage, loss of livelihoods and services,

socioeconomic disruption, or environmental damage. Data harmonization at the European level, in this environmental sector, is a goal of the previously mentioned Infrastructure for Spatial Information in the European Community (INSPIRE). Due to the large amount of data from various sources required for the assessment of environmental hazards, a complex situation has arisen with data that are not easily accessible, comparable, or integrable. In order to improve the accessibility and integrability of these data, an Interoperability Framework (IF) was created using the INSPIRE conceptual model. Its ultimate goal is to create a geo-portal with a single point of access to multiple sources (Tomas et al. 2015).

The IF is made up of four parts: the adoption of a common terminology; the development of a common data model; the development of a common model for evaluation; and the adoption of a common classification (Tomas et al. 2015).

For the adoption of common terminology, efforts were made to use terminologies that were adaptable to all the different types of natural hazards and that were able to represent a clear, unambiguous and consistent reference for the development of the framework. The problem is that different environmental hazards refer to different sciences, such as hydrology or geology, causing terminology confusion (Tomas et al. 2015). In fact, about ten different meanings have been found for the term risk alone (Aven, Renn 2010), and finding a way to unite the interpretation each term acquires is not an easy job.

Then, one of the most significant barriers to achieving interoperability is the use of different data models. As a result, the FI's primary goal is to create a common data model, following these steps: defining a use case; recognizing customer's needs; analyzing results; developing specifications; gap analysis; implementation; testing and validation (Tomas et al. 2015).

Moreover, assessing a natural hazard entail comprehending its magnitude, likelihood of occurrence, and level of vulnerability of those who come into contact with it. There are obviously several ways to interpret these factors, and the goal of IF is to find a common one. This common model (Figure 11) provides methods for assessing natural hazards both qualitatively and quantitatively (Tomas et al. 2015).

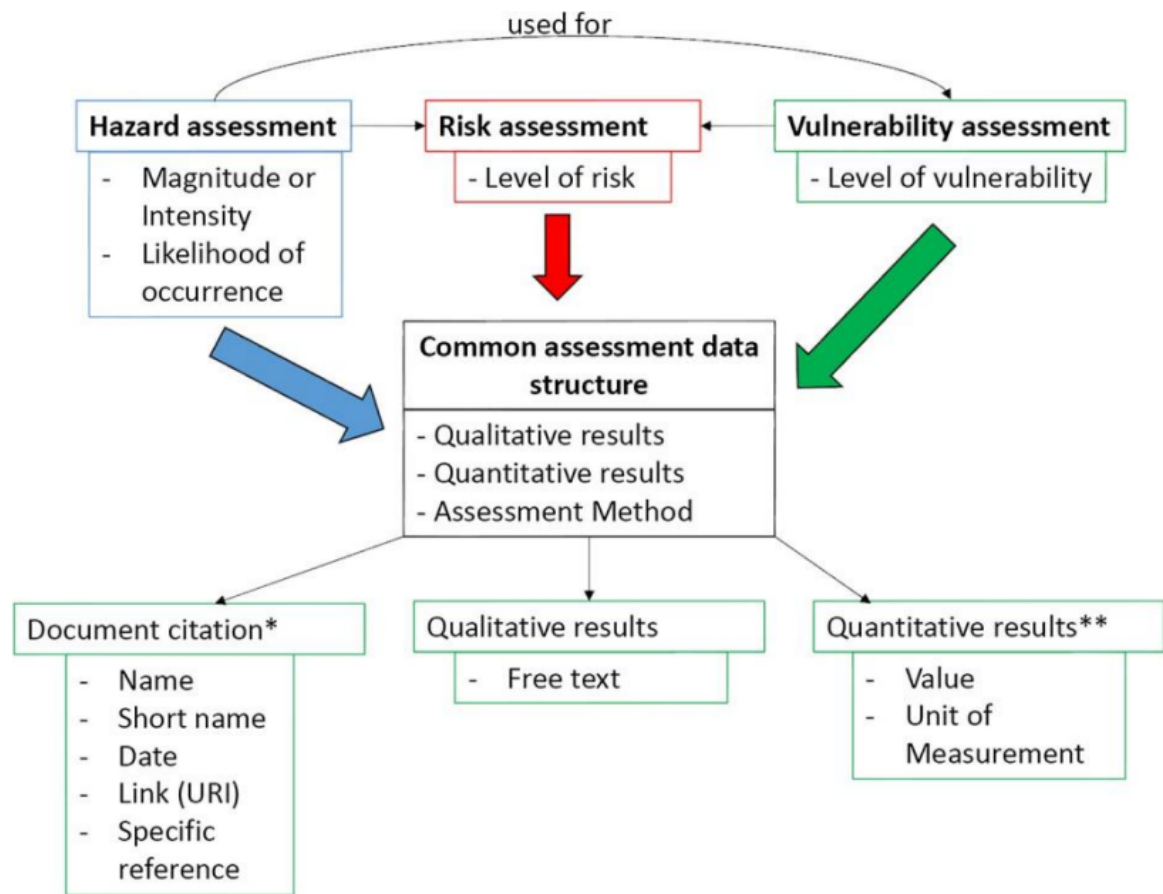


Figure 11: Common natural hazard assessment model

Natural hazards are classified differently depending on how a particular phenomenon is viewed and studied. With the goal of improving data cataloging and interoperability, the proposed model for a common classification employs the origin or type of hazard approach. The issue arises, however, because many hazards are similar to one another, and the terminology associated with them may differ depending on the environmental context in which they occur. A mudslide, for example, may be classified as a landslide or a flood. To avoid these misunderstandings, a two-level more detailed classification is proposed (Tomas et al. 2015).

The European Forest Fire Information System (EFFIS) was used as a use case to test the applicability of IF. EFFIS collects a variety of data and information, such as meteorological data and satellite images, in order to create hazard and fire maps, map burned land, and calculate various numerical indices. The IF enables EFFIS

## Literature Review

data provided via WMS (Web Mapping Services) to be interoperable with data provided by other suppliers. The greater the number of providers who can adapt to IF standards, the easier it will be to achieve efficient and complete integration (Tomas et al. 2015).

To fully utilize the knowledge and data on natural hazards, the definition of the FI is consequently crucial. Thanks to the FI these data will be able to be comprehensively compared, integrated and analyzed in order to reduce the possibility of disasters occurring (Tomas et al. 2015).

As we have seen, a natural hazard can cause environmental damage. Going into more detail, environmental damage can be caused, for example, by invasions of pests and alien species. According to the literature, biosecurity, which deals with strategies for managing these pests, alien species, modified organisms, and infectious diseases (Meyerson, Reaser 2002), is becoming an increasingly important topic. This is due to the fact that the environmental damage caused by these organisms, as well as the costs associated with it, have been steadily increasing over the years (Magarey, Colunga-Garcia, Fieselmann 2009). In fact, a study was conducted in the United States in 2009 to develop a biosecurity-related information management infrastructure that is analyzed more in depth. The objective of this infrastructure is to facilitate the collection, integration, and analysis of data from various sources and stakeholders in order to reduce the risks associated with these phenomena. Furthermore, it would allow businesses, governments, and academic institutions to work together to significantly improve biosecurity (Magarey, Colunga-Garcia, Fieselmann 2009).

Biosecurity strategies are carried out by phytosanitary agencies that have several tasks, including the detection and management of pests outside the national territory, entry controls in the territory, and control, identification and management of pests that creep into the territory. Biosecurity agencies are responsible for a variety of tasks, including pest detection and management outside of the national territory, entry controls within the territory, and pest control, identification, and management within the territory. In the United States, the main agencies in charge of these tasks are the US Department of Homeland Security's Customs and Border Protection (CBP) and the US Department of Agriculture's Animal and Plant Health

## Literature Review

Inspection Service's Plant Protection and Quarantine (APHIS PPQ) (USDA) (Magarey, Colunga-Garcia, Fieselmann 2009).

Understanding the plant biosecurity procedure used to control and reduce the risk associated with the entry of invasive species is essential to determining the true need for this infrastructure. This is necessary to comprehend what and how much information is needed by phytosanitary agencies, as well as how integrating with other data sources might improve their effectiveness. The process involves several activities on four levels: off-shore, port-of-entry, detection and pest program (Magarey, Colunga-Garcia, Fieselmann 2009).

Among the offshore activities is international capacity building, which means training and technology transfer to developing Countries to improve their phytosanitary agencies to prevent invasion. Another activity is to conduct inspections for approval to import products and to generate increasingly effective methods to prevent pests from being introduced into the territory along with these products. Phytosanitary standards are also set at this stage by the International Plant Protection Convention (IPPC), an international biosafety treaty. Information to consider at this stage is related to foreign pests and plants, which can be provided by the North America Plant Protection Organization, the European Plant Protection Organization, or private companies (Magarey, Colunga-Garcia, Fieselmann 2009).

The main activity of the port-of-entry is to conduct inspections at the entry of products into the territory to intercept potential pests and prevent their spread. Some treatments and quarantine periods may be necessary if pests are discovered on shipments in order to get rid of them. The most important information at this stage of the process relates to the shipments that required quarantine and treatments, the types of treatments per pest type, the description of the contents of the shipment, the route that this shipment had to take, and the regulatory action. The majority of these data are included in the PPQ and are then analyzed by the Agricultural Quarantine Inspection Monitoring (AQIM) system to obtain information on the various risks at diverse entry points and the likelihood of pests entering the Country (Magarey, Colunga-Garcia, Fieselmann 2009).

## Literature Review

Post-border detection activities are primarily carried out by the APHIS PPQ, with assistance from the Cooperative Agriculture Pest Survey (CAPS), whose main task is to monitor risk areas and then conduct surveys on the detected pests. The information gathered by CAPS during these activities is fed into the National Agricultural Pest Information System (Magarey, Colunga-Garcia, Fieselmann 2009).

Finally, if all of the previously mentioned activities fail to keep the pest from entering the State and it manages to establish itself on the territory, additional measures will be required. The New Pest Advisory Group (NPAG) is in charge of conducting assessments to understand the risk and potential for the pest to spread, and then establishing guidelines to follow in an effort to exterminate it. Collaboration among the various agencies is critical in particularly dangerous situations. The main issue is that pest-control activities are very costly and politically intricate due to the consequences on the state (Magarey, Colunga-Garcia, Fieselmann 2009), both economically and socially. This level of data collection involves very in-depth information about the pest that has established itself, its phenology, how it spreads, and potential eradication methods (Magarey, Colunga-Garcia, Fieselmann 2009).

As can be seen from the analysis of this process, there are several pieces of information involved, and the cost to manage them increases with each level. A structure is required to collect, manage and integrate all of this knowledge from governments, industries, and universities in order to get a comprehensive picture of the situation and improve the processes of risk analysis and the biosecurity in the territory. The challenge in developing this infrastructure is related to the difficulty in developing sharing mechanisms between heterogeneous datasets. This difficulty is exacerbated by the fact that the information involved is frequently from neighboring but distinct States. Incorrect information sharing may result in serious State security issues. To avoid these issues, the legislation defining the purpose, roles, and responsibilities of data processing would be required, but there is currently none. Because of these issues, farmers and governments are frequently forced to hold back information in order to avoid crop losses and subsequent profits. Actually, farmer cooperation would be essential since their knowledge may



result in better research, assist phytosanitary decisions, and enhance pest tracing (Magarey, Colunga-Garcia, Fiesemann 2009).

The examination of these studies shows that coordination, integration, and collaboration between Public and Private Administrations are crucial for protecting the environment. Without joint efforts, it is difficult for risks related to natural hazard to be reduced. It is clear that this is a difficult assignment that requires several efforts. But we can infer from what we just witnessed that some integration-related measures are being made. So, based on what we just learned, the heterogeneity of the data and the delicacy involved in exchanging sensitive data make this challenging as they do in almost all follow situations.

### 3.4.2. Health care

The field of health care in this context has been considered in its broadest sense. Indeed, methods of data management in medical, biomedical, and pharmaceutical research, as well as nutrition, is addressed. In the following paragraphs, through different studies, these fields are analyzed in order to fully understand how the use of data sharing and how the collaboration between different institutions can bring benefits to the health care system.

Health services research studies several factors, such as “social determinants, organizational structures and processes, technologies, financing and reimbursement, individual choices and behaviors on the access and quality of health care delivery and overall health and well-being of individuals” (Ozaydin, Zengul, Oner, Feldman 2020).

Health service researchers use data from various sources to conduct research aimed at improving health care. The lack of an aligned dataset, however, prevents them from quickly consulting and analyzing the data; instead, they are forced to work on numerous datasets, wasting a lot of time and not fully utilizing their potential. For this reason, in order to fully benefit from these heterogeneous databases, an infrastructure for their integration must be built (Ozaydin, Zengul, Oner, Feldman 2020).

## Literature Review

The data used in the health care sector can be classified into four levels (Figure 12). The first level includes patient data associated with clinical, administrative, research, and systems that manage medical devices used by patients. Clinical systems, in particular, include everything that is part of an Electronic Health Record (HER). At the second level, patient data is aggregated in an institutional enterprise data warehouse and based on this aggregated data, third and fourth-level data are generated through data analytics & business intelligence and reporting & visualization systems (Ozaydin, Zengul, Oner, Feldman 2020).

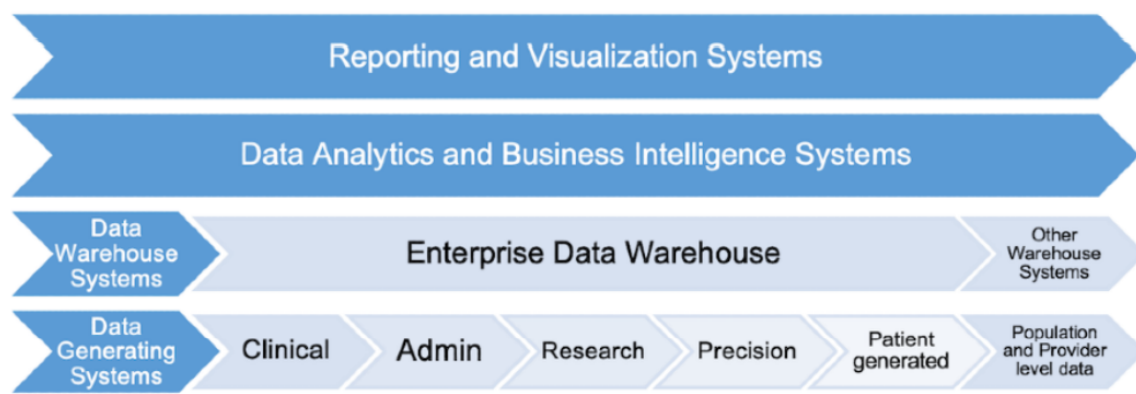


Figure 12: Four levels of data (Ozaydin, Zengul, Oner, Feldman 2020)

In more detail, the data contained in the EHR are derived from the combination of different data from various EMR (Electronic Medical Record) systems of hospitals or private centers. In 2016, Venezuela conducted one of the first studies on merging EMR systems in order to create an HER ecosystem (Clotet, Hernandez, Huerta 2016). All information related to the reason for admission or visit, the outcome of the visit, medical history, and everything about the patient is included in EMR systems. In order for the various EMRs to be integrated, the data must be anonymized in order to avoid privacy concerns while also being able to be used for research purposes. To avoid data dissemination and privacy issues, access to the entire EMR must be granted only to those who have the right to do so through the use of roles. It is also critical to develop standards for the integration of various heterogeneous EMRs in various facilities, as well as to reduce the size of large data sets. Even with standards, synchronization of heterogeneous databases in order to build a HER ecosystem is a difficult task (Clotet, Hernandez, Huerta 2016).

## Literature Review

At a higher level, Ozaydin et al., (2020) sought to develop a single platform that combined all four previously defined data layers from governmental or nongovernmental administrations. This platform, known as the Healthcare Research and Analytics Data Infrastructure Solution (HRADIS), is thus aimed at integrating healthcare data in a dependable, safe, and consistent way so that it can be envisioned, analyzed, and exploited with the help of BI (Business Intelligence) systems. For the full exploitation of this platform, there are a few challenging difficulties connected to architectural and socio-technical issues that still need to be solved. Concerning architectural issues, these mostly revolve around data interpretation, updating, and security. The socio-technical ones, on the other side, focus on how technology interacts with its users, some of whom may not be ready to adopt new technologies (Ozaydin, Zengul, Oner, Feldman 2020).

For the development of this platform, the Jonas (2007) iterative model was used (Figure 13). The analysis of this model is very useful in understanding how interoperability can be achieved, which is why it is discussed in greater depth. It involves four phases: analysis, projection, synthesis and communication. Each of these phases is, in turn, made up of four steps: research (data collection), analysis (study of the data collected), synthesis (expression of the problems that emerged) and realization, which allow for the transition from research (theoretical) to realization (practical) (Jonas 2007).

		Steps of the iterative <b>micro-process</b> of design			
		Research	Analysis	Synthesis	Realization
Phases of the iterative <b>macro-process</b> of design	ANALYSIS – “the true” (how it is today)	Gathering data on the current situation	Understanding the current situation	The current situation relative to the whole situation	Usable presentation of the current situation
	PROJECTION – “the ideal” (how it could be)	Gathering data on future changes	Understanding future needs	Scenarios incorporating possible future situations	Usable presentation of goals and future evolutions
	SYNTHESIS – “the real” (how it is tomorrow)	Gathering data for immediate use	Understanding the requirements	Design solutions	Usable solutions for immediate use
	COMMUNICATION – “the driver”	Understanding the process for moving it forward			

Figure 13: Iterative model (Ozaydin, Zengul, Oner, Feldman 2020)

The first phase of analysis seeks to comprehend the current situation and data availability. After downloading all currently offered information from various databases, it was linked together, but without the creation of a single database (Ozaydin, Zengul, Oner, Feldman 2020). This phase’s problem statement was discovered in the management of large amounts of data, the definition of responsibilities for data access, and, most importantly, the lack of standards among different datasets. Based on what was learned in the analysis part, the projection phase attempts to identify the ideal future State. This phase’s problem statement includes several objectives, including developing the platform with end users’ needs in mind. To reach these conclusions, use cases were used to gain a better understanding of the user-technology relationship and how to optimize it. The third synthesis phase focuses on the real, that is, how to make what was highlighted in the previous two phases possible. The final product is a blueprint for the HRADIS platform’s initial implementation. Communication, the fourth and final phase, is present at all times during other stages and is critical to the process’s success. When working in large groups, communication becomes even more important and must be more formal in order to raise awareness (Ozaydin, Zengul, Oner, Feldman 2020).

## Literature Review

The use of this model has enabled the development of long-term win-win solutions as well as the creation of a structure capable of integrating various data sources in order to exploit and analyze them (Ozaydin, Zengul, Oner, Feldman 2020). The use of this platform can be aimed not only at improving health services but also at improving pharmaceutical or biomedical research.

In particular, in the biomedical and pharmaceutical fields, research serves as the foundation for innovation. Public-Private Partnerships (PPPs) have been created in these areas to enable, through open information exchange, more in-depth studies to produce new scientific discoveries to enhance the healthcare system (Davis et al. 2021). Indeed, also as stated by De Souza et al. (2014), databases of information on the structure and activity of substances from numerous sources enable the creation of novel findings to improve scientific investigations.

One of the most important PPPs in the pharmaceutical sector is the Innovative Medicine Initiatives (IMI), whose main goal is to develop new drugs, vaccines and antibiotics through collaboration between companies, institutions and government (Davis et al. 2021). It focuses on four different areas: those with no incentives for private companies; those with very complex diseases that require pre-competitive consortia; those with data sharing platforms that can advance research and development and those with gaps in the process of discovering and developing new drugs (Lavery, Meulien 2019).

Because there are multiple players involved in PPPs, integrating the data that each provides is critical (Davis et al. 2021). In this regard, IMI has launched the IMI Open PHACTS (Open Pharmacological Concept Triple Store) project, with the goal of delivering and maintaining an 'Open Pharmacological Space' (OPS). OPS is a platform that enables research activities to be carried out in a more efficient way by using interoperable language and common standards (Williams, Harland 2012).

In response to the needs as well as the new necessity for machine-readable data, FAIR (Findable, Accessible, Interoperable, Reusable) principles were developed (Davis et al. 2021). Given the complexity and rapid spread of data volume, the FAIR principles are intended to support data reusability (Shanahan, Hoebelheinrich, Whyte 2021). They allow computational systems to find, access, interoperate, and

reuse data with little or no human intervention<sup>7</sup>. The FAIR Data Principles, for example, provide a data processing framework to assist researchers in managing their data. Furthermore, by sharing FAIR data, researchers facilitate knowledge discovery and increase the possibility of collaboration. IMI has, also on this occasion, launched a project called IMI2 FAIRplus with the aim of applying FAIR principles to data generated according to IMI principles (Davis et al. 2021).

In the pharmaceutical and biomedical context, then, there is also a very large Open Dataset called PubChem containing information on substances, components, and bioassays integrated from different sources (Kim et al. 2021). This is primarily used to conduct studies for new drug development, but it is also used to better understand the positive and negative effects of existing drugs (Kim et al. 2021). It is thus a great source for the advancement of research, which has, however, had significant development difficulties due to the very many heterogeneous sources that have been integrated.

Biobanks are another context in which PPPs are becoming increasingly important. There are currently three major ones underway: the International Society for Biological and Environmental Repositories (ISBER), the Biobanking and BioMolecular Resources Research Infrastructure-European Research Infrastructure Consortium (BBMRI-ERIC), and the European, Middle Eastern and African Society for Biobanking (ESBB) (Davis et al. 2021). Biobanks are collections of raw material used in both public and private sectors to fuel scientific, medical and biomedical research (Harris et al. 2012). By having access to this vast collection of human biological markers and related clinical and research data, it is possible to make the most of emerging technologies that can improve patient care and result in significant advancements in the development of new treatments. Biobanks are internationally acknowledged as sharing platforms and, in order to be properly used, they must be made interoperable, flexible and sustainable. For this reason, it is essential that there is participation and, above all, the collaboration of different actors, such as research facilities, but also funders and policymakers. To this end,

---

<sup>7</sup> [What are the FAIR Data Principles? | Augustus C. Long Health Sciences Library \(columbia.edu\)](#)

## Literature Review

national biobanks participating in the BBMRI are working to develop information collection and sharing standards that will allow different sources to be harmonized, increasing the quantity and quality of usable data. In addition to cooperation between many parties and information harmonization, other crucial factors to concentrate on for the better development of biobanks are connected to funding, culture, education, and legislation (Harris et al. 2012). With regard to funding, there is a European program called ESFRI (European Strategy Forum on Research Infrastructures) that specifically aims to build a system that enables biobanks to be founded internationally (Hancock et al. 2010). Concerning culture and education, instead, there is currently a lack of widespread knowledge regarding these topics, so it is crucial to train future professionals in such a way as to make these data related issues accepted and recognized and consequently make them more open and trusting about sharing (Harris et al. 2012). Finally, because many actors are involved in these sharing processes, conflicts of interest can arise, making it critical to collaborate with policymakers (Harris et al. 2012) to establish a legal framework based on legislation. Legislation is also necessary to address the issue of privacy during information sharing, but the regulations currently in place are outdated and ineffective. Biobanks should not only be integrated with one another but also with national health information systems, to increase the chances of obtaining funding, to speed up the cultural process of normalizing biobanks, and to maximize the benefit of having even more information available (Harris et al. 2012).

These PPPs have accelerated research and science, but for their full potential to be realized and for them to be able to integrate the growing amount of available data, funding and investment must increase (Davis et al. 2021).

Another important aspect related to health care is nutrition. Both for the health of the individual, but also for the health of the environment, the topic of nutrition needs to be addressed more carefully. Climate change, indeed, is one of the largest issues we face today, and irresponsible food management is one of its main causes. In fact, the UN created 17 Sustainable Development Goals (SDGs) to be accomplished by 2030 in an effort to solve these issues, several of which include better food management (Conrad et al. 2021). Specifically, to increase access to safe and nutritious food, more sustainable agricultural systems, and better management

## Literature Review

of natural resources is required. The biggest obstacle to attaining these objectives is that also this field's research is constrained by a lack of comprehensive data, which prevents it from providing insight at the economic, social, and environmental levels. This was the subject of a study conducted in the United States. Data on food consumption, food prices, environmental resources, and production-related emissions were examined in depth. Several data sources were found, such as the National Health and Nutrition Survey (NHANS), which collects information on dietary data and health behaviors supplemented with additional data to have more information on food groups, individual foods, ingredients and nutrients as well (Conrad et al. 2021). The other sources provide datasets on the nutrient content of different foods, but also information on food waste, for example. The integration of the many sources with NHANS has the objective of providing information on recipes, on conversion factors to estimate food groups, on the number of ingredients in food, on prices, on the changes in consumer prices and on health behavior. The study's goal is to determine how these databases can be integrated in order to conduct more in-depth studies and improve sustainability. The findings, indeed, indicate that by linking the different datasets it is possible to assess food loss and waste in great detail and is also possible to assess the environmental impact of individual diets. Finally, the cost of purchased, wasted, inedible, and consumed food can be estimated. There are several constraints that make integrating these datasets challenging. First and foremost, these are disaggregated datasets, and linking them in order to analyze them is difficult (Conrad et al. 2021). Furthermore, there is an update problem because, for example, the NHANES data are updated on a regular basis, but when linked with the data about prices, if these prices are not updated as well, it results in incorrect datasets. Another source of complexity is data incompatibility. For example, the data about prices only include home food consumed, but if estimates are to be made for out of home food consumed, assumptions will be required, making the results less reliable (Conrad et al. 2021).

In conclusion, the integration of public and private data in the broadly defined health sector can bring significant benefits in terms of development and innovation, though there are challenges. From the analysis of these various abovementioned studies and different attempts at integration, the major difficulties observed can be summarized as follows: the presence of different standards makes data integration



and querying complex; the relationship between people and technology is underestimated, increasing the chances of project failure; data updating is not always consistent across datasets, resulting in unrealistic information; the concepts of privacy and regulation are essential, coming into contact with sensitive patient data, but it is not easy to ensure that only those who have the right to access the data are granted access; funding is not sufficient; the current culture does not recognize the issue of sharing as “normal,” making it difficult to accept such initiatives.

### 3.4.3. Smart Cities

In recent years local governments pushed in the direction of smart cities because it led to an improvement in urban economic performance (Yigitcanlar et al. 2018). The purpose of this paragraph is to understand the role of private partners in the efficient development of the Smart Environment and the advantages and risks related to Open Data in Smart Cities.

For city services to function what they need is access to clean reliable Open Data. These data are used to build apps and services that represent a value-added for citizens. The objective is to create basic applications which make Smart Cities closer to reality. Smart city realization requires huge investments and the implementation of efficient and cutting-edge technologies; these barriers can be overcome by partnering among public, private, and nonprofit organizations for the rapid development of Smart City projects (Liu et al. 2020). In the Smart City landscape, PPP becomes of relevant importance. This cooperation is often considered the basis of Smart City creation, as it enables the financing and management of digital infrastructure and related services. Considering this structure, the benefits involve both sides of the partnership. The private sector provides its managerial, commercial, and innovative capabilities during the design, financing, construction, and operation phases and achieves an economic return. Therefore, to continuously attract private sector investment and leverage their technical and management expertise, it is important to ensure the economic sustainability of the project through specific financing. One factor pushing the private sector for improvement is related to the payment terms; the private sector is paid according to the quality of the service it provides, and this pushes partners to use their expertise to improve the

## Literature Review

service level (Liu et al. 2020). The public sector gets economic and financial benefits from the presence of the private sector because, with the same number of public capitals committed, it is possible to optimize the use of available resources capable of generating a virtuous circle between public spending and the provision of public services.

Efficient public-private collaboration requires a model that clearly defines the collaboration between the two actors. The most commonly used models in Smart City projects are Build-Operate-Transfer (BOT) and Design-Build-Finance-Operate (DBFO). Based on these two models, the private sector manages and designs the public project under a long-term contract (Liu et al. 2020). PPPs are becoming increasingly common as large companies such as IBM, Cisco, and Siemens use Smart Cities as a testbed for new technologies (Liu et al. 2020).

Considering the Smart City context, it is also necessary to analyze the data problems. The high number of smart devices and external sources of information generates an increase in the volume of the data flow among public authorities and all the actors involved in the project. Lee et al. (2019) pointed out the lack of interoperability and openness of data among the various sectors of Smart Cities reduces the reusability of city data and limits the availability of various data services. Therefore, it is of paramount importance to integrate and standardize data from different sources so that the various players in the project have visible and available all the data they need in real-time (Stone et al. 2018).

The SmartSantander Augmented Reality app, for example, was developed in Barcelona in 2010 and represents an excellent integration initiative. The app, in fact, combines data from various sources on a single platform and allows anyone to access it to consult it, as well as analyze and reuse the various information on it. SmartSantander proposes a digital infrastructure in support of typical applications and services for a smart city. This facility is sufficiently large, open, and flexible to enable horizontal and vertical collaboration with other players. The project envisions the deployment of 20,000 sensors in Belgrade, Guildford, Lübeck, and Santander, exploiting a large variety of technologies (Barcelona Ciutat Digital 2018). It leverages publicly available municipal data to give users access to traffic and

public transportation information and weather information in a variety of locations, including beaches, parks, gardens, and tourist attractions (Stone et al. 2018).

Another central point that it is necessary to take into consideration for the survival of Smart Cities is the presence of a large number of sensors and smart devices that inevitably produce a huge amount of data that will necessarily have to be collected, analyzed, and managed optimally to ensure excellent services. Hence, it is important to use smart city platforms that, acting as operations centers, are able to process and make the best possible use of big data from the urban environment. Integrated management of a large amount of data should be handled by centralized platforms that ensure linking and mixing databases that have different domains and structures. A concrete example is the ALMANAC project, funded by the European Commission (Bonino et al. 2015). The goal of this initiative is to create a platform that delivers services, and corresponding technological solutions, by integrating IoT networks with metropolitan networks. The platform ensures semantic interoperability between private and public networks; thus, the platform's native data (i.e., data from smart devices) are aggregated and analyzed with datasets from both public and private external sources. I-LOD is another example of a platform that has the objective of overcoming the lack of cooperation among different datasets within a Smart City (Lee et al. 2019). This system supports the industrial problem-solving occurring in urban areas by creating a semantic integration of industrial real-time data. It is possible to state that, before the development of a Smart City, there is a strong need for a platform that enables communication among completely different devices, but the presence of these platforms could also push both public and private actors to share their datasets.

In conclusion, the evolution of Smart Cities represents both opportunities and risk. For this reason, it is also important to underline the possible barriers of the current context. The most important one is the involvement of the private sector entails certain risks to the privacy of citizens. This is because the implementation of smart city projects requires the collection and profiling of big data; this amount of data could be used by private organizations to profile customers for profits. Thus, one of the main looming problems, when public sector databases are opened and shared with the private sector, is the privacy of citizens' personal data. To avoid those

private partners use personal data for commercial purposes, it is fundamental that during the design phase of the PPP the rights to acquire, analyze, share, and use data must be clearly defined (Liu et al. 2020).

### 3.4.4. Transport Sector

Among the many dimensions of a Smart City, transport is one of the critical components of today's urban challenges (Mor et al. 2021). The development of Information Technology Services and Smart Cities suggests that it will be necessary that the existing information system will be extended in order to embody newly available information, as well as to integrate information from all the stages of road design, construction, and maintenance. The main obstacle is that all these information come from different data sources (Gristina et al. 2016). The focus of this paragraph is the "Open Transport Data" which refers to the goal of maximizing the reuse of transport datasets (Colpaert 2017); the goal is to use open standards for transport and mobility, and it allows to share of transport account data (tickets/purchases, usage/journeys, and discounts/concessions) between different providers and platforms. Within the last decade, the impact of information and communication technologies on improving the efficiency of city operations had incredibly grown. This section of the literature shows some of the potential applications arising from the opening of data sets.

According to the actual European trend, the impact of public and private transportation on the environment is very important. Hence, traffic congestion and traffic-related air pollution should be more properly managed (Gristina et al. 2016). These elements previously described require integration and interoperability among different information systems so that the roads are uniquely described, indexed, and identified. A standard used to manage all stages of a road's life cycle is the Geographic Information Systems (GIS) for registering and managing roads; the generic nature of GIS, organizing information by location, is interdisciplinary and not specific to any application. Gristina et al. (2016) carried out a study to investigate the integration and interoperability of the Italian Road Inventor, the road cadastre, with the goal of developing a 3D GIS model for road management. Furthermore, through interviews and workshops, it was demonstrated the validity

of the model and also usefulness, for the various organizations involved, to have access to a 3D model integrated with road inventory and cadastral data.

One of the main operational challenges that it is necessary to overcome in order to create successful public-private integration is the alignment of heterogeneous datasets. This limitation can only be effectively addressed with the help of ontologies. Having the ability to offer a conceptual schema for any data set, independent of format, structure, or size, ontologies play a significant role in the integration and standardization of static and dynamic entities. The system must be capable of evolving while remaining backward compatible and should also have a low entry barrier, as it will be used by both data users and data publishers (Colpaert 2017). Once a standard framework is defined, it can be used by administration and city stakeholders for traffic planning design and policy-making purposes to provide services to commuters through specific applications. A successful example of ontology standardization is the Smart City Umbrella Ontology (SCUO) developed in the context of Indian surface transportation (Mor et al. 2021). It is a multilayer ontology that provides a comprehensive foundation as well as a skeleton for designing and improving existing infrastructure to provide administrative services. The designed framework serves as the bases for future innovative platforms aimed at offering new interconnected transportation services based on linked data. What has been done in the Indian context demonstrates how at the basis of any process of integration among databases there must be an effort on the part of an entity to create an ontology that goes beyond the point of view of the single actor, and it is the starting point for future integration projects.

Another issue raised from the literature is the inconsistency among public transport companies' Data Base; in a perspective of efficient data management, the datasets of PA should be published "once only" and be aligned with all other institutions with the aim of maximum reuse (Colpaert 2017). The identifiers of different datasets must be interoperable. An Open Data interoperability framework extremely streamlines the process of integration because all the available data could communicate with each other. A particular use case is intermodal transportation where route planning is an integration, alignment, and interoperability data problem (Colpaert 2017). In order to create a reliable and solid digital infrastructure,

## Literature Review

it is necessary that data providers, it is important that they give to the platforms or open good quality data; data must be valid, consistent, accurate and complete. Data quality is the measure of the alignment between the real world and its representation in an information system; the only way to improve the data quality is to increase the use of Data (Orr et al. 1998). Good data quality is crucial for taking advantage of opportunities in the public transport market and finding optimal solutions for citizens. The trend toward Open Data is stimulating data providers to increase the quality of their own data in order to reach more benefits from the process of opening databases.

An obstacle in the integration process is the presence of high modifiability datasets of linked data; the parties involved should keep high-quality data up-to-date data in real time. This process is quite difficult and also very big projects, such as DBMedia which extrapolate information from Wikipedia and publish them on the web as LOD, have low modifiability data (Iijima et al. 2011). Hence, regarding Linked Open Transport Data it is also important that the involved part, not only open the databases but also actively participate in the maintenance and updating of data on the infrastructure.

A challenge that slows down the integration process among different party is the initial absence of resources needed to clean and align the datasets. Usually, in particular the databases of Pas, cannot be opened directly. Actions aimed at cleaning the data and actions aimed at protecting the privacy of citizens are necessary. From an internal point of view, the effort of manually fixing the heterogeneity of datasets is still too high in comparison to the company's benefits in the short term (Walravens et al. 2016).

Some papers (Iijima et al. 2011; Mor et al. 2021; Colpaert 2017) show that the practices used for data linkage are similar to each other; the recurring steps are as follows: Collecting Data and Creating Master Data; Converting Data to RDF Format; Creating Links; Storing the Data in the Database and Creating a SPARQL Endpoint. The underlying problem is another: the propensity of the various entities to share and integrate their databases. It is, therefore, necessary to go beyond technicalities in order to involve more closely public and private entities in the integration process. Usually, the opening process is not spontaneous, but the various actors are

## Literature Review

involved in projects with the aim of creating ontologies, frameworks, and software applications to integrate data management.

Evidence proves that a useful method to involve more public and private institutions is the use of workshops to inform and collect feedback from institutions (Colpaert 2017). The aim is to create ad hoc solutions and share possible future benefits with the involved parties. For instance, before starting an integration project it is quite important to highlight the virtuous circle that is created as it increases the number of possible institutions that are interested in the project.

In conclusion, a common theme resulting from many transportation-related papers is the goal of integrating heterogeneous databases from the public and private sectors in order to create a data source readable by humans and machines and to develop new useful applications. Furthermore, the alignment and the integration of different data sources lead to a high understanding of the existing infrastructure usage in terms of traffic roads, road congestion, and CO<sub>2</sub> emissions (Fredericque et al. 2011). Once the infrastructure data are consolidated and harmonized it is also possible to go into further integration of geospatial data with heterogeneous documents such as legal documents, city plans, technical reports, images, and existing scanned plans, allowing models to be created to serve as an index for other data.

Once the parties decide to share the collected information it is important that the data can communicate among them, which means that the same standard formats must be used. At the end of a set of workshops, Colpaert et al. (2014) identified four main variables to warrant interoperability among different entities: technical, syntactic, semantic and organizational interoperability. The most effective solution, to cope with the possible issues, is to use Linked Open Transportation Data. LODs guarantee complete interoperability between different repositories both from a formal point of view, as they are coded in a common format, and a conceptual point of view, as defined using homogeneous classes. Hence, the goal is to push both public and private institutions to use the same format in order to allow interoperability among different datasets.

### 3.4.5. Tourism Sector

The Open Data movement is perfectly reversed in the tourism domain as it makes it possible to communicate and aggregate data belonging to various intermediaries in the tourism sector, both public and private; this creates value added both for tourists and data providers. There is an increase in service level for the tourists and greater visibility of the tourists for the data providers. According to Longhi et al. (2014) the Open Data movement and other disruptive technologies have led to a deep redefinition of the whole attributes of the tourism industry. Moreover, new Web applications (TripAdvisor, Facebook) have profoundly altered the tourism sector, shifting the focus from mass travel tourism to individual behavior analysis (Longhi et al. 2014). The new technologies allow to share more and more information and the data consumer can become a data provider. However, open access to information could become a two-edged sword. For instance, the available information on some open sites or applications can increase interest in specific locations, but there is also the risk of diffusion of distorted information (Longhi et al. 2014). The new topic of LOD fits in a strategic way within the tourism domain. The integration among different Open Data sets can be used to enrich tourism available data and content in order to meet the tourists' requests (Yochum et al. 2020). People are focused on the quality of their tourist experience; hence it is necessary that data providers give information in an aggregate and linked way. By doing that, more information availability leads to an increase in number of the tourists that are interested in a specific location. Furthermore, Open Data can be a vector through which public authorities and private partners collaborate in order to promote local territory (Longhi et al. 2014).

In an Open Data domain, the main objective is to integrate different data sources in order to create new services for tourists (Yochum et al. 2020). The cooperation among various public and private partners brings benefits to both sides. The data consumers, the tourists, have available more and more information related to the quality and the services offered from a possible selected host structure, such as the various points of interest nearby available. According to Yochum, et al. (2020), from the point of view of data providers, the main benefits are:



## Literature Review

- An improvement in sightseeing as there are more available information to organize guided tours.
- Better management of transportation services as it is possible to increase the accuracy of forecast on tourist traffic.
- Improve marketing actions as it should be possible to use a unique channel to share information about relevant points of interest.

One of the factors that creates competitive advantage is given by tourism recommendations systems, which assist tourists in identification of their ideal holiday (Yochum et al. 2020). The presence of LOD related to tourism can be used to improve, first the performances of the recommendation systems by exploiting semantically interconnected data; on the other hand, it could improve the veracity of recommendations by exploiting the huge amount of integrated data (Yochum et al. 2020). Furthermore, along these lines, the widespread use of mobile applications allows users to share their visited locations and other information, which generate user check-in records. The integration among all these data and private data of local entities, such as geospatial information and available places left, can be used to improve the accuracy of recommendation systems (Yochum et al. 2020).

Tourism is strongly linked to the availability of services and local mobility efficiency. Many Regions are trying to develop Mobility as a Service applications to increase the service offered to citizens and tourists. MaaS applications require data to be available in machine-readable formats, which is often not the case, and is still a major limitation. The Velopark project fits into this context, it is an example of data integration between public authorities and public transport agents (Rojas et al. 2020). The goal of the Velopark project is to promote the massive use of bicycle mobility. According to Rojas et al. (2020) the data, such as the parking spaces available for bicycles, are present on the websites of each operator but are not presented in an aggregate and unique format. For this reason it was decided to create a digital infrastructure that allows cyclists to know where and how to access and use physical cycling infrastructure. This successful example can be seen as a base case for future applications in many other sectors, such as culture and tourism. Machine-readable data is critical for building applications that provide useful and reliable information (Rojas et al. 2020). For example, let inspiration by Velopark,

which gives information on the real-time occupation of a specific car park, it would also be useful to create applications that monitor tourist traffic in real-time, in order to share aggregate information on the number of people who are planning to visit a museum.

Summing up, the core benefits related to massive usage of LOD in the tourism sector are an increase in transparency, discoverability, accessibility, usability, and interoperability of tourism data (Yochum et al. 2020). Cooperation between local authorities and PPPs are necessary to increase the competitive advantage of a region in the tourism industry, as it is necessary to have as much real-time information as possible in order to deeply understand tourists' needs and to adapt the local resources to them (Longhi et al. 2014). According to Kim et al. (2017) the process of "production-processing-utilization", done by tourists, small and medium-sized enterprises and public institutions, creates a virtuous cycle of benefits for all the players involved in the network. It means that a deep understanding of Open Data practices gives benefits both to data providers and data consumers. Tourists can get information from a unique platform without searching for information on different websites (Rojas et al. 2020; Kim et al. 2017). Small and medium-sized businesses can create new information services by combining private data they hold with public data, and they can easily acquire trustworthy tourism information to use for offered services (Kim et al. 2017). It appears from the various papers that the first step in pushing public and private entities to share and align their datasets is to find mutual interest within the industry. In addition, another common aspect is to have public funding before the start of an integration project.

### 3.4.6. Culture Sector

LOD methodology is influencing the ways that institutions organize and share their data on the web. LOD projects are contributing to the creation of an open and uniform information environment. They can be used to solve some problems that have always affected cultural institutions, such as interoperability and the integration of data and the enhancement of information in a broader context of individual websites or application portals.

Regarding the culture sector, the first entities that feel the urgency to identify new forms of integrated and shared communication and mediation are the archivists. The goal of cultural institutions is the birth of thematic portals aimed at disseminating on the web the local cultural heritage out of the restricted circles of scholars. Cultural institutions that open their databases acquire a whole series of advantages: usability, searchability, discoverability and interoperability of their information and reuse by the community (Guernaccini et al. 2019). Despite all the benefits deriving from the use of LODs, most of the participants of these projects only parts of the future benefits, limiting themselves to the benefits deriving from immediate visibility on the platform (Guernaccini et al. 2019). The main consequence is that few organizations are still involved in a wider process of production and use of Linked Data.

As previously highlighted in the transport sector, ontologies play a central role as they disambiguate the semantics of data and uniquely identify data objects. In this way, there is a minimization of the arbitrariness of the data, and it is possible to create new knowledge. The analysis of the literature shows how there are various ontologies for the representation of archival heritage<sup>8</sup>. We are therefore faced with a reality that is undoubtedly rich in content, but each system has developed its own modules and tools for interrogation, research and consultation, making any interoperability impossible. A good attempt of interoperability has been made in the Italian scenario in which there have been a set of projects aimed at the development of cultural heritage. These projects<sup>9</sup> have in common the use of the same ontology and integration with resources connected to the LOD Cloud (Guernaccini et al. 2019).

Regarding the cultural sector, another area where LOD play a central role is the library management. LOD are important for libraries because they offer the possibility of integrating catalogs which, by their nature, can be very heterogeneous.

---

<sup>8</sup> Europeana Data Model; CIDOC CRM; SAN; RiC-O

<sup>9</sup> Archivio storico della Camera dei Deputati; Complessi archivistici degli Istituti culturali emiliano-romagnoli; Archivio storico della Presidenza della Repubblica; Archivi della scienza

The traditional “walled garden approach<sup>10</sup>” has limited the access and use of library resources. Instead, LOD, based on the principles of openness, offers the possibility of displaying own data to every data consumer. This is a key factor to favor and improve the collaboration between cultural institutions, also including different fields.

Cultural Heritage is the result of physical artifacts and intangible attributes of a Country collected over time. LODs play a crucial role in the management of cultural heritage because the high quality and the richness of linked data create great value for local societies (Carriero et al. 2019). In Italy, MiBAC (Ministero per I Beni e le Attività Culturali) is in charge of the management of the Italian cultural heritage. Within MiBAC, instead, ICCD (Istituto Centrale per il Catalogo e la Documentazione) is in charge of listing and updating Italian cultural heritage; in this regard, it was realized the ArCo platform that is the Knowledge Graph of the Italian Cultural Heritage (Carriero et al. 2019). The ArCo project has led to interoperability between national or regional institutional bodies, public or private, which manage cultural heritage within the Italian territory. It is important to highlight how this platform extremely facilitates the interoperability process by minimizing the effort required by public and private entities to open their data set as ArCo also includes software to automatically convert ICCD compliant catalog records into ArCo-compliant LOD (Carriero et al. 2019).

In conclusion, open and machine-readable data that are published so that they are connectable to others of the same type represent one of the most important resources for enriching the cultural sector. Considering the European landscape, the promotion of open and interlinked data is one of the “best practices” adopted by the main platforms for the management of digital cultural heritage. Hence, the Italian and European landscape is digitally ready for a change in this direction, but still there are problems related to the inaction by public institutions; furthermore, large quantities of raw data are not well structured and they cannot be opened.

---

<sup>10</sup> “A Walled Garden is an enclosed environment that controls the end user’s access to certain websites and services”

Currently, the simplest way to transform data into LOD is to join [CulturaItalia](#)<sup>11</sup>. The [CulturaItalia](#) system is designed to transform metadata into LODs and to automatically send them to [Europeana](#)<sup>12</sup>, the European digital library that collects digital objects of European public and private institutions. At the end of this analysis, it is possible to assert that at the bases of cooperation and integration among different public institutions it is necessary to find an agreement about the ontologies to use to represent the data. It is also necessary to avoid that local entities develop ad hoc models customized to their cultural heritage.

### 3.4.7. Government sector

The term Open Government Partnership (OGP) refers to a 2011 initiative to increase the openness of governments in order to make them more accountable and responsive (Janev 2020). The first activity required for governments to become more open is data sharing, specifically the development of “e-government” (Machado, Parente de Oliveira 2011). It entails using information systems to quickly and easily interact and share information with citizens, departments, and other governments to improve government efficiency (Charalabidis, Lachana 2020). In recent years, there has been an increase in the sharing of these data (Ding et al. 2011) that have been published in an open format in order to make this exchange even more easy and more valuable. Linked Open Government Data (LOGD) is a type of Open Data that is published in a raw format so that it can be freely accessed and reused to create new services and improve existing ones (Machado, Parente de Oliveira 2011). To date, this concept is represented by the term Government 3.0, which refers to governments opening up data in order to improve public services, transparency, and interaction with citizens, as well as more broadly with other governments (Janev 2020). It is critical that the data disseminated remains of high quality for Government 3.0 to be truly effective. As a result, in this context, governments must

---

<sup>11</sup> [Dati Cultura Italia](#)

<sup>12</sup> [Europeana](#)

be flexible, inclusive, and willing to share data and information also with nongovernmental actors such as private organizations, NGOs, businesses, and academic partners through public-private partnerships (Espey, Badiee, Dahmm, Appel, Noe 2019). In the following paragraphs, the benefits and problems associated with the development of Government 3.0 is discussed in order to better understand why it is critical to fully deploy this system and what actions must be taken in order to reap the greatest number of benefits.

One of the major advantages of using Open Data is transparency (Ding et al. 2011; Na Lamphun 2016; Janev 2020). Greater transparency simplifies data access for citizens and other stakeholders. Data consumers will be more aware of the government situation, they will be more participatory, and corruption will be reduced as a result (Na Lamphun 2016). This is due to the ability of LODG to connect datasets, which increases their significance and the possibility of comparison. Another advantage is the reduction in costs associated with publishing data in open format (Ding et al. 2011). Previously, in order to gain access to data, requests had to be made, after which the data was processed and delivered. Instead, they are published directly, which reduces the processing, time, and costs associated with them. The other main benefits are, as mentioned above, related to the increased efficiency of public services through enhanced proximity and collaboration with citizens and other organizations (Moon 2020).

But, to fully benefit from Government 3.0, some changes in governance as well as legal and political structures are required (Espey, Badiee, Dahmm, Appel, Noe 2019).

In terms of governance, different changes could be made. First, a new figure, the Chef Data Officer (CDO), would be appointed to manage all data issues between governments and partnerships. The CDO's contribution would be critical not only in managing economic funds and finding new ones, but also in encouraging public-private partnerships by taking care of data integration, making it easier, more effective and innovative. The CDO figure was first introduced in France in 2014, and it has since been adopted in Estonia, New Zealand, and the United States, with great success in all cases. In these situations, the CDO has assisted in simplifying information sharing between governments at both the technical and strategic levels,

as well as supporting the spread of Open Data and closing data gaps (Espey, Badiee, Dahmm, Appel, Noe 2019). Second, in order for a CDO to be possible, there must also be a supportive policy framework. That is, we require regulations that effectively enable the CDO to form relationships with various stakeholders in order to securely share their data (Espey, Badiee, Dahmm, Appel, Noe 2019). This is a significant constraint that currently prevents many governments from opening up, and it should be addressed urgently. The third point on which governance changes are required is in national development plans. Each Country will need to create its own route and comprehensive plan, addressing and emphasizing the importance of data. Local governments must also collaborate for the success of a Government 3.0, because the majority of the population lives in urban areas and the first contact with the institution occurs at the municipal level. In Patiala, India, for example, a project was developed to monitor data on waste management, pollution, parking, stray animals, and road safety, which was led by various city stakeholders who were able to align national goals with local goals through a neighborly approach, yielding positive results (Espey, Badiee, Dahmm, Appel, Noe 2019). Furthermore, in order for this systematic shift to a more open government to be properly implemented, it will be necessary to hire personnel with new skills, knowledge, and abilities in data processing, in the use of artificial intelligence and new devices, as well as in negotiating with various stakeholders (Espey, Badiee, Dahmm, Appel, Noe 2019). Finally, improvements have been made at the international governance level; in fact, two World Data Forums were held in 2017 and 2019 to create a space where governments, companies, and organizations can express their stance on the topic of data development and collaborate to innovate. The problem is that these are informal actions that fail to establish universally applicable standards. What studies should focus on to achieve real progress is standardization through legislation that all nations can adopt in order to create an official system of interoperability and development. The United Nations will play a significant role in this task, and only through their massive intervention will these radical changes be possible (Espey, Badiee, Dahmm, Appel, Noe 2019).

As previously stated, one of the most significant obstacles to the development of a Government 3.0 is a lack of standards, and several actions on the political and legal structures must be taken to find a common plan. First, at the international level,

different states should agree on a common terminology so that data from various sources can communicate easily. Of course, this is a difficult task because each term has different meanings, especially from Country to Country, and reaching an agreement can be challenging. At the same time, even if a hypothetical standard is identified, adapting to it necessitate significant effort on the part of data providers who will have to modify their data structures (Espey, Badiee, Dahmm, Appel, Noe 2019). However, once these obstacles are overcome, the benefits of using standardized data will be substantial. Another fundamental legal change that must be made is to define the boundaries for Open Data policy. That is that must be established the extent to which data can be open in order not to run into privacy problems. Countries have been increasingly becoming “open by default” in recent years. In 2016, 112 Countries declared that they would follow this principle, which means that government data would be made open by default in favor of open access and greater transparency. The issue with this principle is that it must be determined what its limit is and how far data can be opened up. This is why we would need proper regulations to specify what can and cannot be open. Because there is no clear information about data sharing and openness, when companies and governments form partnerships, they must enter into very specific agreements to avoid legal issues concerning privacy or data ownership. In this regard, in 2018, an initiative called Contracts for Data Collaboration (C4DC)<sup>13</sup> was launched with the goal of simplifying the legal processes for entering into contracts for public-private partnerships in order to incentivize data sharing. It essentially consists of an online library of past agreements that can be freely accessed to provide support for those who need to enter into a new contract. Contracts are time-consuming and costly to enter into; however, by using this library, these can be reduced. In fact, the system will quiz interested parties in order to better understand what they require before offering comparable contracts from which they may learn how to make more informed choices. This is only one example of how things may be made simpler, but

---

<sup>13</sup> [Contracts for Data Collaboration](#)



the UN's involvement in creating legal guidelines is essential if we want to make the kind of real change we need (Espey, Badiee, Dahmm, Appel, Noe 2019).

In conclusion, it is clear from the literature analysis of the government sector that there are numerous advantages, but that numerous systemic adjustments in terms of governance and legality are also required in order to fully capitalize on them. Through the Interoperability Framework (IF) and the PSI directive, which were previously discussed, Europe has already taken some steps to try to start this change and unify the usage of data, but there are still many difficulties related to them (Janev 2020) and there is still a lot of work to be done.

### 3.5. Final considerations

This study provides a thorough review of linked Open Data from many domains. Per each domain, the objective has been to analyze the current state of the art and try to understand the risks and benefits and some experimental results.

Generally, the EU pushes in the Open Government direction to pursue initiatives that make data more accessible and open for European citizens. The aim was to create greater transparency within local governments and facilitate the flow of data of interest between States Member. For this reason, a series of directives have been created to facilitate the transition of PA towards Open Data. In addition, general guidelines have been created to lead Countries toward the digitization process. To take stock in Europe, the study proposed by Capgemini<sup>14</sup> concerning the maturity of European countries relating to the transition to Open Data is interesting. It is based on 4 analysis dimensions had split the 34 participating European countries into 4 groups trendsetters, fast trackers, followers, and beginners. Italy has a good positioning as fast trackers characterized by a maturity score between 89% and 92%.

More in detail, the literature has highlighted several benefits of LOD application in different domains. For instance, some proposed cases studied in the environment

---

<sup>14</sup> [Open Data Maturity Report](#)

sector show as several problems regarding biosecurity and management of natural risks can be efficiently handled by data sharing and integration. Furthermore, an integrated infrastructure facilitates research and creates benefits for citizens. The analysis of the papers has revealed a theme that unites the various sectors proposed is the need for PPPs for the management of projects based on Open Data. They provide better infrastructure solutions than an initiative that is wholly public or wholly private. Each participant does what he does best; in particular, the private partners provide their technical competence to manage the infrastructure. As reported in the health and in smart cities sector, greater interest from local entities is needed in order to increase the funding for these structures. Existing ones, such as IMI, have already brought satisfactory results and higher funding is also a factor to involve both small and large entities.

The analyzed articles taken into consideration brought out that the main limitation for the development of PPPs is the lack of culture from the various authorities in sharing and integrating their databases. They look only at the immediate economic benefits with little focus on the long term. Another problem that is a barrier and slows down the spread, as highlighted in the transport and government sector, is the risk to citizens' privacy; on the one hand private partners can use data for commercial purposes and on the other hand the public partner can have too detailed profiles of citizens. As highlighted in the government sector, another constraint is the bureaucratic apparatus that surrounds the PPP which slows down its development. The evidence shows that the case studies that have collected positive results are characterized by a unique ontology behind the digital infrastructure. The latter is fundamental to overcoming the issues related to data heterogeneity. Furthermore, a predefined ontology that standardizes the language to represent a specific domain could be used in the future as the backbone for interoperable applications.

Existing literature focuses on the problems, benefits, and interoperability models of a specific relationship or specific projects. This turns out to be an important limitation as local authorities are focused on the creation of platforms aimed at managing only their own reality. In conclusion, it can be said that the foundations for the explosion of the Open Data movement already exist but there is a need for a

## Literature Review

great and widespread cultural change. The literature does not present a univocal supra-sectoral model involving, facilitating, and supporting public and private authorities in the opening and sharing process. The suggestion is the creation of a sector-neutral approach that can be used by any organization that aims to develop an integrated platform and open its datasets. The definition of a protocol to create an open partnership would help to overcome all the cultural problems and it should be also possible to define best practices to address the challenges that arise during projects aimed at data integration and at opening data.

## 4. Empirical Analysis

As previously explained in the methodology section, an empirical analysis has been done after the analysis of the literature. This section shows the main results and information collected from empirical evidence within a data collection model. The main objective of this investigation is to then formulate the output of this study, the sharing model.

The data collection scheme, as already mentioned in the empirical methodology section, has four main dimensions of analysis: Data, Relationship with third Parties, Open Data and Possible Scenarios. Each dimension is evaluated basing on different metrics. The model does not report the specific results of each interview, but the most relevant results found in both the Public Administration and private sectors. The goal is to make the evidence collected more structured and to use the results extrapolated from this pattern and the results from the literature to create a standard model of interaction between PA and private companies. The model can be used to define a set of recommendations and best practices for improving the Open Data ecosystem within Public-Private Partnership.

In the following sections, therefore, the four dimensions are reported and analyzed in detail.

### 4.1. Data

This dimension aims to analyze the propensity, possibilities, and abilities from data providers to make databases available, integrated and subsequently opened. The driven question from the interview protocol was: *“what is the propensity and willingness from Public Administration and private authorities to make their databases available to third party?”*

This concept resulted to be of fundamental importance even before the concept of Open Data. Indeed, having authorities willing to collaborate with third parties to create new value-added services is the first step toward opening up databases. It is pointless to discuss Open Data without this awareness and willingness. As a result,

## Empirical Analysis

the actors involved must recognize that there is value creation. The PA generate value for the civil society in agreement with its final purpose; the private sector, as it is possible to see, gains benefits in long term.

This section consists of five sub-dimensions: data available, data enhancement, information need, critical factors and enabler factors; that are detailed below.

### 4.1.1. Data available

This indicator analyzes the availability of data from respondents. Moreover, the data presented into this section belong to those public and private authorities that has shown the willingness to open their databases-

Considering the gathered information, it emerged that PAs have a lot of data with great intrinsic value that are not yet processed. It should be also highlighted that the "data publication" is mandatory and regulated for PAs, but this does not imply having available integrated and Open Databases. From the interviews it was clear that the purpose of publishing data only for administrative transparency purposes has been progressively replaced over time in favor of a reuse of open PA data.

Below are reported three examples of high value fragmented data among PA, Public Holding and private sector.

- 1) Data related to mobility management: car parks, parking meters, bike sharing, limited traffic zone, traffic light and traffic control system, sustainable mobility services.
- 2) Meteorological and environmental data whose economic value depends on the immediate availability and on the constant updating.
- 3) Data related to the cultural heritage: archival, literary and paintings heritage.

### 4.1.2. Data enhancement

This section presents how the specific player enhances his available information assets. The aim is therefore to understand the reasons why the authority invests in its information assets to create value. For clarity in this analysis, the term "data enhancement" refers to that process aimed at reworking the available data and

## Empirical Analysis

developing services that add value both inside and outside the confines of the company landscape.

During the studies four areas of high interest for data enhancement were identified: policy making, economic return from data, reporting on the effectiveness and efficiency of the exercise and creating data enhancement systems. Each of this area is detailed in the following section.

One of the main common elements between Public Administration and private authorities regarding data enhancement is the policy making; therefore, the objective is to define a set of practices and processes that help to ensure the optimal use of resources within the company. Data enhancement is also used as a support tool for long-term decisions at both the operational and strategic levels.

An emerging element mainly in the private sector is data enhancement with economic purposes. Large investments are done for commercial use of the data, such as for customer profiling and market analysis. In addition, the data is also internally processed to resold information to third parties. The economic benefits related to data enhancement represent a limiting factor for the purposes of the carried-out analysis. This is because the private individual is not motivated in undertaking a sharing process for which he does not see a benefit, when on the other hand he has at his disposal the possibility of selling the data and generating profit.

An element that is strictly aligned with thesis project is reworking of data in order to create systems for data enhancement. It consist in the provision of platforms that align the needs of data providers and data consumers with the aim of creating value for the community. In the case of the private companies these need to develop the platform from scratch, while the public authorities usually use already existing portals to offer this service.

### 4.1.3. Information need

Being in the era of digitization, all the players involved have shown a growing need for information. In this discussion, this need regard all those data that the interviewee does not currently have available, or they are struggling to obtain.

## Empirical Analysis

These data, however, could be integrated with the internal ones to create new added value both for the company itself and for the community.

Different information need raised from the interviews and specific requirements are not dealt with the discussion below, but the commonalities among the various actors involved are. The work done was to collect and cluster the most widespread of them into the following categories:

- *The geolocation of movements.* From the evidence it was very interesting how these databases could be used to create origin-destination matrices to map movement of people. Beyond privacy problems related to their specific location, these databases allow to create value both for private and public authorities. The PA could better manage public transports and services; the private partners could have higher visibility on tourism domain. As previously said, this study requires a high data processing investment to be aligned to General Data Protection Regulation (GDPR)<sup>15</sup>.
- *The database of major mobility players.* These databases resulted to be very important because they facilitate the creation of Mobility as a Service (MaaS) applications by third parties. For instance, the starting point could be the information sharing and cooperative ventures between organizations that oversee public and private national transportation and those who offer services for alternate mobility. This collaboration would create benefits both for data providers and data consumers. A truly functioning ecosystem could generate benefits for all those actors involved in the mobility ecosystem. For citizens, who could access new digital services for every travel need, making more informed mobility choices. For MaaS mobility operators, thanks to a system of clear rules that lowers barriers to entry and encourages the development of the free market. For the Public Administration, which would

---

<sup>15</sup> GDPR is the toughest privacy and security law in the world. Though it was drafted and passed by the European Union (EU), it imposes obligations onto organizations anywhere, so long as they target or collect data related to people in the EU. The regulation was put into effect on May 25, 2018. The GDPR will levy harsh fines against those who violate its privacy and security standards, with penalties reaching into the tens of millions of euros. ([GDPR.EU](https://gdpr.eu))

be able to govern the mobility of its territory, collect data and optimize the use of resources. The interviewees commonly highlighted which entities should be involved: Azienda Trasporti Milanesi, Trenitalia, Swiss Federal Railways (SFR) and Private Transport companies (Italo, Trenord). These organizations own thousands of data gathered from the different covered stations and they can be used to create new value-added applications for the civil society.

- *Public and private data.* The large amount of data collected from the private sector was also considered of interest. The standardization and integration of databases of various sectors such as food, agriculture, tourism, automotive, could be used for the creation of statistics to study social society attitudes and improve product and services consequently. The goal of the public sector is to make analysis on public behaviors; the private sector can use these analyses to study possible end users and market viability before even entering it.
- *High value-added data:* a common aspect from the analysis is the high interest on high value data, such as mobile phone datasets. It is a set of datasets already processed, cleaned and ready to be integrated. The factor that limits the access to these data is the great intrinsic economic value of these databases. Data Providers process them in order to resell them, but the PA and the private companies not always have the economic availability to pay for them. Due to the high interest resulting from the interlocutors, these datasets are later be dealt with in the discussion on paragraph 6.1.3. Hence it is possible to assert that over the next years, high-value datasets will gain in popularity, and the Digital Governance Act and the Open Data Directive (described in the “European Legislative Landscape”) offer a preliminary framework for their introduction and adoption.

### 4.1.4. Critical factors

This section focuses on the limiting variables regarding data integration. Only later the focus shifts to the critical factors concerning the relationship with third parties and to Open Data.



## Empirical Analysis

The first aspect to consider is the lack of interoperability among different databases. Some Italian and Swiss actors have highlighted the failure of some integration projects due to the onerous process to align different datasets; furthermore, it was detected that within the same sector are used different standards and vocabularies to store the data. Referring to the literature, this process could be facilitated by the definition of a univocal sectoral standard that makes the DBs interoperable. An example that validates this thesis is that some DB integration processes among different partners were successful only within the project itself. These are projects in which a standard was defined to be used by all the partners involved into the project.

An element that has limited the data integration process is the investment needed to harmonize the different databases. Particularly, regarding PA, it is difficult to obtain databases already processed ready to be integrated with third parties. Data harmonization is a key process in the development of data infrastructures. The goal of the process is to transform datasets of different nature in such a way as to make them homogeneous, both with respect to geometry and semantics. At the end of the process, the aim is to have a unified view of the data from different sources, where conflicts and inconsistencies between the data will be removed. This process requires specific resources and investments, it is therefore necessary to be able to justify them. Starting from this concept, the evidence has shown a lack of resources, in terms of time and budget, from the PA to harmonize their internal DBs. Anyway, this problem is addressed in the next section, when the benefits that justify the investment are explained.

Furthermore, it is also necessary to emphasize that integration and publication projects do not always translate into Open Data. This happens because the various partners use standards that are not aligned with the open format. At this point to shift from public to open it is necessary the intervention (time and money resources) of another player who acts as a converter towards Open Data. In contrast, all the benefits gained from the integration project are bound within the project itself without the possibility of interoperability with external portals.

An example of limit generated by the sharing standards was what is done in the Trentino Alto Adige area where a long process of mapping the territory was carried

out by the PA relating to the use of public and private transports. Once the data of this project was published, the biggest limitation was the use of a vocabulary that was not easily interoperable with third parties. In detail, it was used a Germanic standard (VDV), technically superior to others and very advanced, but not directly interoperable with Maps which uses the General Transit Feed Specification (GTFS).

### 4.1.5. Enablers factors

In this section the focus is shifted on the positive factors that have pushed public and private actors to work together.

The analysis revealed that when data from private partners is transmitted from PA to citizens, the private sector becomes more interested in PPP. It means that it is necessary that the PA, within PPP, should become the touchpoint between the privates and the users in order to enlarge the user base of the private partners involved. Furthermore, the interaction allows the creation of new ways of accessing services and new paths that could stimulate other ideas and cross different profiles.

An example relates to the potential for improving Automatic Vehicle Monitoring (AVM) system effectiveness that is an essential instrument for the integrated management of urban transportation. The AVM makes it possible to handle the information flows required for service modulation by automatically acquiring the TPL's (Local Public Transport) key operational data. It was noted that the effectiveness of AVM systems would increase if efforts were made to integrate the suggested databases by both commercial parties (transportation firms and alternative transportation services like bike sharing) and the public authorities. As a consequence, cross-border movement, which is presently handled by several agencies, would benefit. This would also have positive effects on the tourist and cultural sectors. The alignment of the Italy-Switzerland data has been emphasized several times as a way to boost coastal, mountain, and cultural tourism.

## 4.2. Relationship with third party organizations

The goal of this dimension of analysis is to investigate how the various subjects interact with each other, whether they are public or private entities. This section

then presents the information gathered regarding public-private interactions as well as the limitations discovered for shared information asset management.

This section consists of two sub-dimensions: subjects from whom acquire/integrate data and limitation and difficulties; that are detailed below.

### 4.2.1. Subjects from whom acquire/integrate data

The first section of the analysis focuses on "how" and "how many" respondents use third-party data acquisition and integration solutions.

The results showed that the simplest way to obtain data in the quickest manner is to join associations (which can be translated into partnerships) or start projects aimed at integrating data. This is because working through projects with well-defined scopes facilitates and motivates the actors involved.

An example is what happens among industrial associations, such as Confindustria, where the various partners freely share part of their databases in order to exploit the combined information to study the market and create new value for all the partners. In this case the collaboration results optimal in terms of effort and investment; the drawback of this scenario is that the results of the integration are aimed at the project itself and not to be made available to third parties.

Going in detail into specific contexts:

- Public Holdings. Two cases that were analyzed during the interviews has shown that once the data are integrated, processed and new data is created, the information remains within the project itself and shared through internal channels among PAs. These data can be accessed upon request; the problem is that the number of demands received is small.
- Private company. The ones who are able to better manage and rework the data are those companies whose core business is the data enhancement. From the interviews it appears that they are oriented to Open Databases and they have the capabilities to produce LOD. The main problem is that they work on information of data providers and the responsibility in this case bears on their owner. It sometimes happens that those who provide these services

prepare the data in linked format (ready to be opened), but the decision to open them belongs to the data providers who, if private, do not open them because they do not find an economic return; if Public Administration, they exchange them through channels inside the PA and make them available to everyone, but only upon the request from the data consumer, therefore not in OD format.

### 4.2.2. Limitations and difficulties

Considering the existing partnerships, the greatest difficulty encountered during the integration was related to the presence of pseudo-silos which indicates that databases that were not compatible with one another within the same organization were discovered. In order to create value among different entities it is necessary to align different datasets and enabling osmosis requires time and resources that must be justified within the annual budget; this led to the completion of some integration projects in the bud.

Some of the interviewees were further along in their business. These, in fact, were already used to combining data from various sources and publishing it, in some cases also in Open format. Shifting the focus to these individuals, they stated that it was difficult to find someone willing to collaborate during the data provider search phase. Instead, a strong jealousy of the information has been detected, both from individual private companies hoping to profit from the information in their possession and from the PA concerned about the consequences of excessive transparency. A practical example, concerning the PA, could be the presence of a hole or an intersection with a high accident rate; non-timely intervention by the municipality could lead to complains from the citizens. Going into detail, this can be classified as a cultural problem. Citizens should appreciate the increase in know-how without understanding that it is not possible to do all the necessary interventions immediately. It should be highlighted to them that it is better to bring a problem to light and wait for it to be resolved than to be unaware of it. The “jealousy”, just mentioned, has also been found in the private sector, which is sometimes shortsighted regarding the benefits of possible integration in the short term and fears the alteration of commercial dynamics.

*Another aspect arising from the study is a methodological issue: how the data is collected?*

If data is collected differently, with different standards and vocabularies, putting them together becomes complex. This is one of the problems that, as is possible to see below, the proposed model tries to overcome.

Another limiting aspect for the processing and integration of databases is the initial investment for data collection and its maintenance. An effort at regional level would be needed to arrive at concrete results and more fund should be necessary to finance Public-Private Partnership because the resources available to public authorities, in terms of time and managerial capacity, for the digitization of heritage are limited for now.

### 4.3. OPEN DATA

This section is the center of the research that has been carried out. The maturity level of the cross-border area in the use of Open Data is first presented below and then the critical and enabling factors are analyzed.

This section consists of four sub-dimensions: Open Data culture, Platforms of publication, Critical factor and Enabling factors for Open Data sharing; that are detailed below.

#### 4.3.1. Open Data Culture

The interviewees are aware of the concept and the possibility of Open Data, however only 33% of the interviewees post LOD, according to the sample that was evaluated. Despite the low rate of compliance with Open Data policies, almost 50% of the interviewees saw the possibility of taking part in initiatives to integrate and perhaps open their own databases. These authorities do not actually have definite plans, but are considering starting one.

Focusing on Open Data publishers, it is important to consider other actors who, despite wanting to push in the direction of Open Data, have had problems with data providers due to the previously mentioned "jealousy of data". A case of particular interest dealing with this subject was examined. Thanks to a strong willingness to

move toward Open Data, the respondent from a private entity developed a software to organize the results of the projects with which he collaborates in a way that makes them openable. The company was able to provide a valuable service by using its in-house expertise to create LOD, except that the data remains the property of those who provided it and, in most cases, are not open to the public but only shared with members for commercial purposes. Concerning the PA instead, a lot makes data available upon request by data consumers or directly publishes it on the web, but not in LOD format. These examples show that many actors in Public Administration and the private sector, are actually very keen on Open Data, the problem is that there is strong resistance from the context that is not yet ready. This problem is cultural and severely limits the possibility of data openness and its potential.

Another example that demonstrates cultural limitations was collected by the interview of a Public Holding. Despite publishing high-quality data on their portal a firm in the Canton Ticino, that develop digital products and services, discovered that customers had little interest in their offerings. This highlights two different things. Firstly that the environment is still not ready for these innovations and secondly how much work is required from both data providers and data consumers to start the process toward a radical change in the conception and perception of Open Data.

Hence, basing on the experience of the responders it results that both in Italy and in Switzerland there is a low data culture for the end users that do not have a high usage rate of Open Data when they are available within portals or platforms.

### 4.3.2. Publication Venue

In this section the focus is on whose players who already publish Open Data. Considering the specifics cases of analyzed in the thesis project, it emerged that those who publish the majority Open Data are private companies that provide services for the maintenance and connection of data. In practice, 75% of private companies, among the respondents, public Open Data and obviously they receive an economic return for data processing. Their customers are for the most part public partners (libraries, municipalities), but also private individuals (art galleries and tourist organizations) who want to open their databases. These realities, having no

internal capabilities, turn to experts in the sector who open the data at the request of the data providers. In regard to the PA, data is made available with the intention of maximizing reuse, however no one publishes data in LOD format. The results show that the methods they use to open the data were mainly two:

- Open Data portals: portals financed by EU (like [europeana](https://data.europeana.eu)<sup>16</sup>) or by central government (like [opendata.swiss](https://opendata.swiss/de)<sup>17</sup> e [datiopen.it](http://www.datiopen.it)<sup>18</sup>). In this case the data provider should process data and open them. This concept is better defined in the section related to the model.
- Open Data infrastructures: thematic open portals created by private entities to be closer to the end users. Successful examples are Open Data Hub e Medialibrary. The international digital platform known as Open Data Hub aids start-ups, businesses, and research institutions in creating digital solutions based on actual data. It links data from many data suppliers and makes it simple for data consumers to access this data. Medialibrary Online is the digital lending platform that gathers and makes available eBooks from all Italian libraries and beyond for online usage. It is possible to examine and read open eBooks or legally borrow freshly published eBooks thanks to this digital network. Also, this way of open the data is detailed in the section related to the model.

### 4.3.3. Critical factors for sharing Open Data

During the interviews several limiting factors raised. This section analyzes the different issues that are important to be overcome to facilitate data opening:

- **Data availability.** When a data provider wants to start an opening project, he does not know what data are available and what is the potential of these data. This issue limits the number of actors that could be interested into an opening data project because there is a lack of visibility.

---

<sup>16</sup> <https://data.europeana.eu>

<sup>17</sup> <https://opendata.swiss/de>

<sup>18</sup> <http://www.datiopen.it>

- **Confidentiality issue.** This is especially true for private companies that are concerned about how competitors might use their data once it is public. The government should intervene in this matter to ensure competition without altering competitive dynamics. For instance, an opportunistic use of the available data can be used for customer profiling and for this reason, it is difficult to reach a full opening of the data, but it could be considered the opening of a filtered data.
- **Quality issue.** Data providers must guarantee a certain level of data quality when they open their datasets. Evidence suggests that organizations that have already integrated different databases and published them as LODs spent a significant amount of time and money improving the quality of the data of the various partners before they were opened. The low quality of available databases discourages third parties from participating in data opening processes. Regarding PA it is necessary a job by the administration to start a process of homogenization and harmonization of data to then obtain high quality in the documentation. For example, there was a cross-border project based on Italian and Swiss data, which was never carried out because it was not found a way to link the Italian data to the Swiss data, not being able to use the identifiers of one or of the other.
- **Privacy and economic return.** Before opening, especially sensitive data owned by Public Administration, it is necessary to clean and aggregate them in order to be aligned with GDPR policies and overcome privacy issue. This obviously requires time and money that the PA should invest into their databases. Regarding private sector, these jobs, are expensive and before investing it is essential to understand if the private company can have a return from its investment.
- **Structure of data sources.** Each source has its own structure in order to respond to a specific need. To make several sources integrable, work of harmonization has to be carried out. This work may be feasible in contexts where few actors are involved, but it becomes very complex when the number increases. In particular, when there are inconsistencies between many datasets, artificial intelligence technologists are used to fill the gaps of missing data and make proxies on this incomplete data to be able to obtain a final result. The problem is that these results can be misleading.



- **Processes.** There are no clear and defined processes for returning data. This way when requests are made, it takes a great deal of time to get a response. It does not happen because of incompetence or insensitivity of the works, but precisely because there is a lack of automatisms and widespread culture. During the interviews there were not found best practices to follow to support the efficiency and the effectiveness of a PPP in LOD ecosystem.
- **Motivational issue.** As regard the evidence collected from those private companies who do not Open Data, there is a big motivational issue. Companies realize more and more of data value and prefer to sell them to make a profit. A win-win solution should be found to persuade companies to make their data available, using it as a bargaining chip. Another critical factor is related to the purpose for which the data are used. The data are not relevant if they cannot be used for a clear, direct and defined purpose. There must be a motivation in the use of the data, which leads to the creation of value and a greater benefit than the cost and effort necessary to the collection and the opening of the data. So, basically if you are not able to understand the purpose and benefits from opening projects then it is not worth investing. "You have to prove its value, or you have to invest public money." It is possible to conclude that there is a strong "impasse" due to the fact that when data is shared, there is no certainty of how it will be interpreted. This causes an uneasiness of expectation and a lack of trust in sharing, due to a lack of context and a deconstruction.

### 4.3.4. Enabling factors for Open Data sharing

Starting from the information collected, the main elements that have led to the creation of successful Open Data projects are presented below. In particular, the variables that prompted the interviewees to integrate their own- and third-party databases in order to be opened.

From a general point of view data sharing has allowed an improvement in services. For instance, the developers of applications and systems for mobility have significantly increased the service provided thanks to greater data availability. Furthermore, greater availability of data by administrations, for example on the

## Empirical Analysis

characteristics of roads or on the level of accidents due to road maintenance, had help companies and businesses to study the potential market before starting a project to understand if it could have success in the market and to understand how to build new business model.

An example of relevance is what has been done with the international Image Interoperability Framework (IIIF<sup>19</sup>). It is a set of open standards for delivering high-quality digital objects at online scale. IIIF is supported by a consortium of important cultural institutions. This consortium is establishing itself for the exchange of images allowing those who join to publish their images and make them open to others. Particularly during the pandemic, many cultural institutes have digitized their image heritage using this standard which has enabled international interoperability of cultural assets.

From a "benefit to the community" point of view, it has been demonstrated how having LOD create value for the civil society. For instance, digital libraries, one of the first industries to use LOD technology, have been proven to have a significant impact on the national cultural heritage and to increase the number of readers and book lovers through the use of LOD platforms.

From the interviews it is clear that the private sector has benefited from the opening of their databases. Open Data has allowed greater dissemination and differentiation of the target audience. Furthermore, with new data available from the public sector, new ways of accessing and new pathways have been created that have stimulated new ideas. Along this line it is relevant to highlight the interest shown by the tourism sector (private); the possibility to share data within the same open platform allowed to create a "unique communication channel" between the tourist infrastructures and the end users and this has generated higher visibility and higher tourist flow. In this specific example, also the PA gained benefits in term of offered

---

<sup>19</sup> <https://iiif.io>

service to the community. In fact, a better manage of public transport means depending on the location of people crowd.

Another factor that has pushed both public and private authorities in the Open Data direction is an improvement in the quality of the data. Joining at Open Data project means to offer high-quality data to the platform; in this contest the most expert players had offered support and data cleaning techniques to entice the actors to participate and obtain a direct benefit. For instance, a provider of an Open Data infrastructure, that was analyzed, involved a great number of players within its projects by giving training about Open Data licenses and data cleaning techniques.

### 4.4. Possible scenarios

The final dimension addresses the future of Open Data to determine how and if stakeholders are willing to go down this road. Second, it investigates whether the pandemic has had an impact on the development of Open Data.

This section consists of two sub-dimensions: current and/or future projects and pandemic impact; that are detailed below.

#### 4.4.1. Current and/or future projects

Beyond those actors who already publish Open Data, what has been discovered is that both public and private respondents have a strong desire to start the process of opening up the data. But, actually, the one that have specific future project aimed at creating partnership from the integration of their databases is small. Against the 50% of the previously mentioned respondents who would have an interest in initiating an integration and openness projects, only 25% have actually already concretely defined at least one. The main reason these projects have not taken off is due to cultural barriers, including authorities' lack of readiness and, more importantly, their perception that the environment is not yet ready to support this kind of business.

### 4.4.2. Pandemic impact

According to the information gathered, the pandemic has filled the digital literacy gap; the number of users who have begun to interact with digital applications related to culture and mobility has increased significantly.

Regarding the PA he COVID-19 pandemic has reemphasized the importance of systematic data collection and public availability. Due to the need to respond to the emergency, many countries have begun publishing related data and developing initiatives and dashboards to make the data more understandable and insightful.

The interviewees highlighted that the pandemic-related digitization projects are still used today to get closer to the end customer. This could be a great place to start for future projects aimed at integrating and sharing databases.

## 4.5. Final considerations

In conclusion, it is critical to emphasize that these findings were highly inspirational and significant because they reflect the actual, ongoing situations. The participant summary Table 5 below provides the essential details so that is possible to get a specific, quantitative and schematic understanding.

## Empirical Analysis

Legal Form	Country	Data Integration with private companies	Data Integration with PA	Open Data	Future Projects on Open Data
Public Holding	Italy	No	Yes	No	No
Public Holding	Italy	No	Yes	No	Yes
Public Holding	Italy	No	No	No	No
Public Holding	Switzerland	Yes	Yes	Yes	No
Public Authority	Switzerland	No	Yes	No	No
Public Authority	Switzerland	No	Yes	No	No
Private	Italy	Yes	Yes	Yes	Yes
Private	Italy	Yes	Yes	Yes	Yes
Private	Italy	Yes	Yes	Yes	No
Private	Switzerland	Yes	Yes	No	Yes
Association of Private Companies	Italy	Yes	No	No	No
Association of Private Companies	Italy	Yes	No	No	Yes

Table 5: Summary of interviews

## Empirical Analysis

The analysis described in this section is based on the experience of the representatives of each organization in the world of data management and their relationship and knowledge of external actors for the integration and opening of data. Regarding their legal forms, the Swiss (33%) and Italian (67%) interlocutors were divided into the following four groups (Figure 16):

- 3 Public Holdings.
- 3 Public Authority.
- 4 Private companies.
- 2 Association of private companies.

Of this respondent, the majority were Public Holding and private companies in Italy and Public Administration in Switzerland.

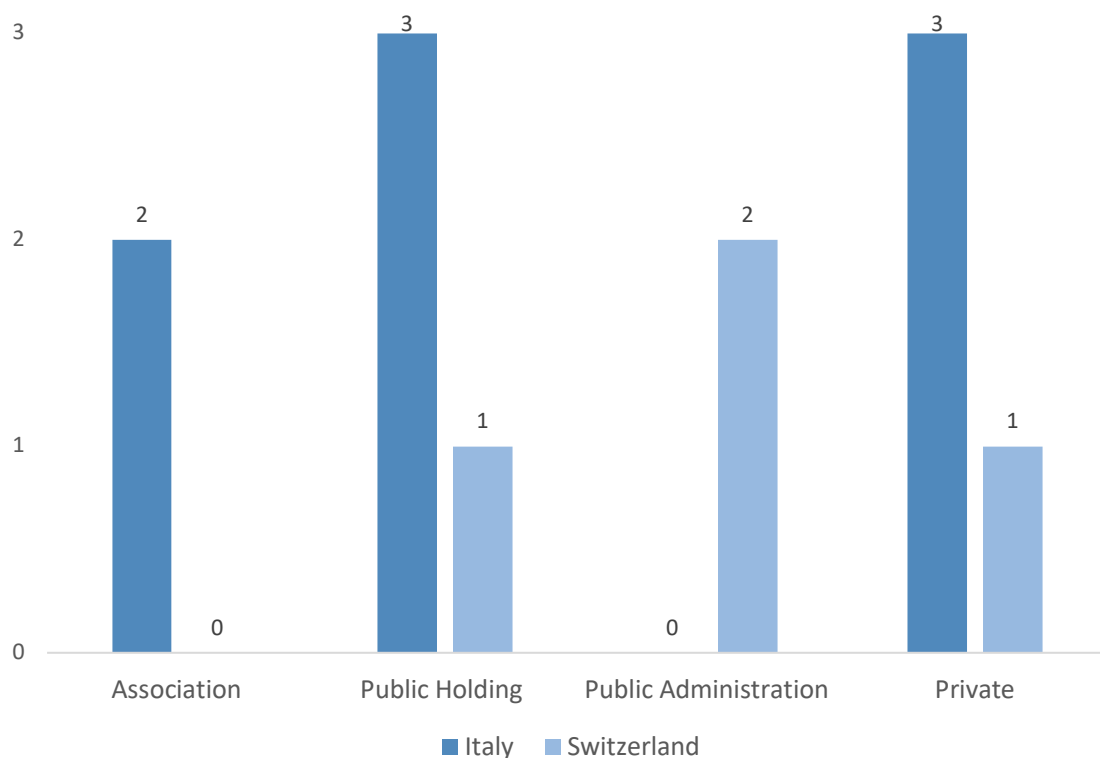


Figure 14: Type of respondent by country

Interestingly, as is possible to see in Figure 15, 42% of respondents are already in Public-Private Partnerships. This class is of particular importance since it was the reference group for the extrapolation of the technical difficulties related to

## Empirical Analysis

information sharing. Almost all of them have encountered barriers in the integration process related to the heterogeneity of the standards used and the presence and absence of internal harmonization of the internal DBs themselves. The respondents belonging to PPPs were 25% Italian entities, while 17% are Swiss and the same issue were found on both sides of the cross-border area.

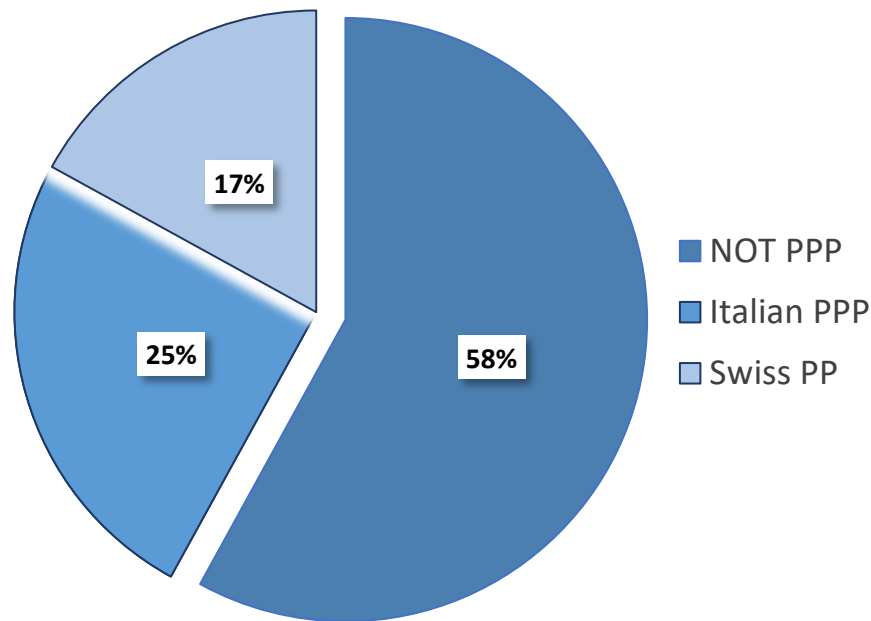


Figure 15: Respondents integrating data by country

Finally, with a focus on the four key aspects extrapolated from the interviews, it is possible to note some critical points from the Figure 16 and 17 below that respectively represent the positive and negative response towards these four main areas:

- Data Integration with private companies. This dimension showed that while conducting the analyses for the thesis project, there was a main interaction with entities that integrate data for the exclusive purpose of economic benefit. In fact, as previously demonstrated, the majority of respondents who are already in an integration situation are private companies. This occurs because the need for these companies to integrate arose in advance, with the ultimate goal of policy making and economic value creation.

## Empirical Analysis

- Data Integration with PA. The results of the integration dimension with PA were not as satisfactory as those with private individuals. This is primarily due to the fact that PA, unlike them, does not derive a purely economic return from these activities. For this reason, the process is taking longer to fully comprehend the benefits of even non-directly economic origin.
- Open Data. In terms of Open Data, the situation is very similar to the previous one on integration. In fact, the majority of respondents who have already published data in open format are private companies that have already integrated. This is an intriguing result that indicates that these companies were forerunners with respect to the context.
- Future Project on Open Data. As has already been shown, however, although the numbers are not yet very high, at least 50% of respondents are considering sharing and open their data. This is not yet happening simply because of a cultural problem, which, however, is being worked on. This last important element suggests that the environment is, anyway, beginning to move in this digital direction.



### Positive responses

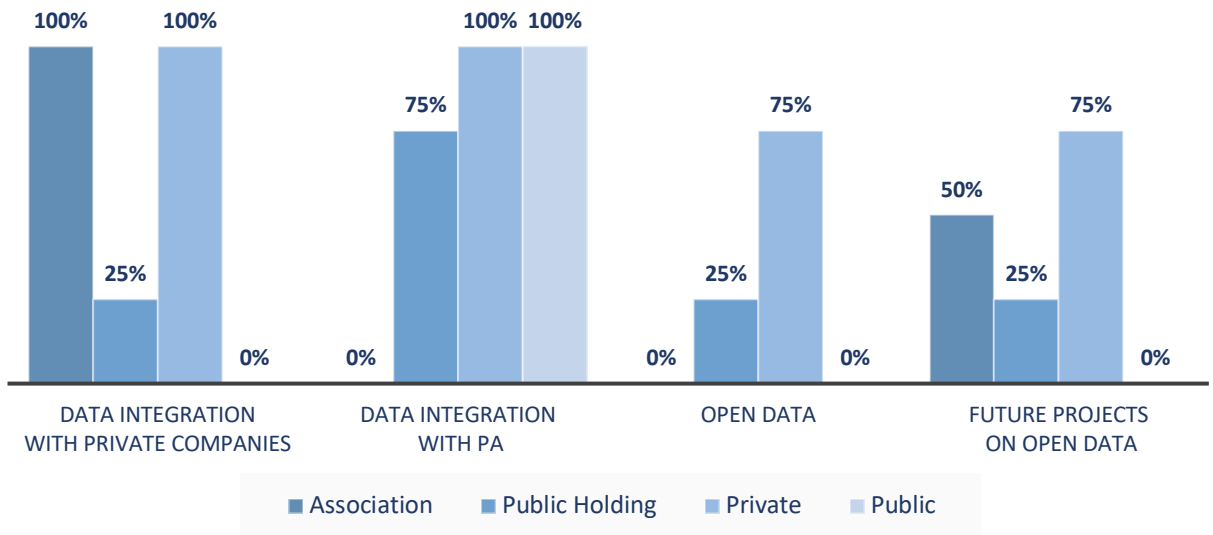


Figure 16: Positive response by type of interview

### Negative responses

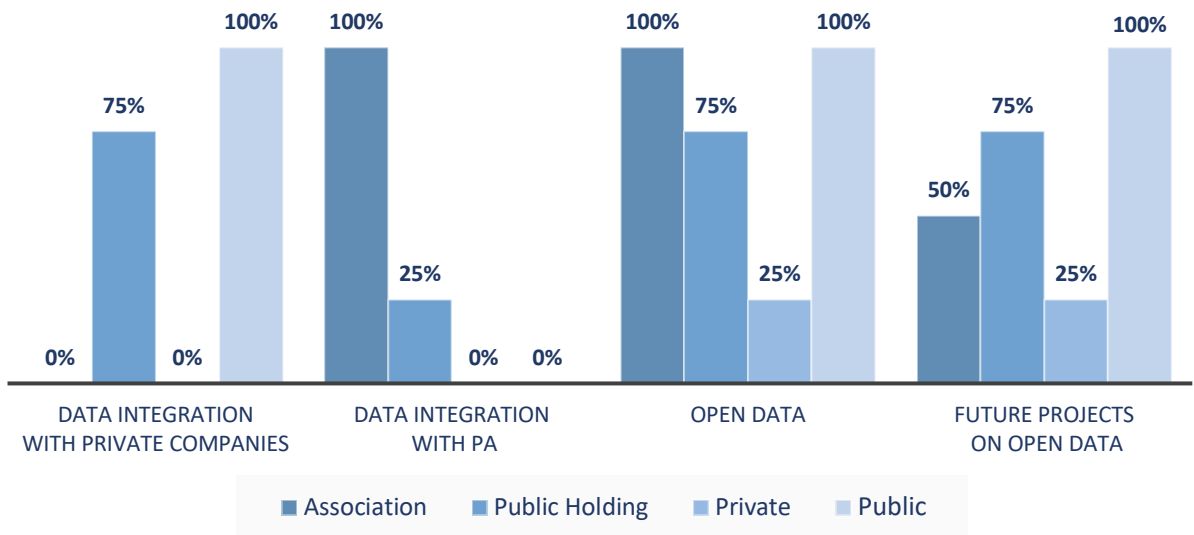


Figure 17: Negative response by type of interview

## 5. Lesson Learned

The aim of this paragraph is to show the commonalities and the similitudes resulting from the systematic literature review and the interviews. The main lesson learned from the collected information are used in the design phase of the model. From a broad perspective, it is reasonable to conclude that the European and the Italian-Swiss landscape is technically prepared to move toward Open Data, but there is still a dearth of partnerships that are dedicated to the integration, openness, and reuse of databases. Both the literature and the evidence have shown that it is necessary to join forces to efficiently manage the production and maintenance process of LODs.

From the perspective of national interest, many countries already have advanced Open Data policies in place, with improved functioning of specialized portals and national level coordination mechanisms amongst the various fields. It was able to clearly see the effort to add more functionality to the portals and comprehend the implications. The flip side of the coin was found both in literature and in interviews. The parallelism that has been discovered is in the end users' unwillingness to use publicly available data. Some organizations work hard to create, integrate, and then publish their own databases. In particular, some organizations, adhering to the principle of transparency, "publish" information not in an open format but in a format that is only available upon request and hence, companies are not encouraged to spend in opening their databases because to the limited number of requests from end users.

It is therefore finally possible to state that there is a trade-off:

- On the one hand, European governments have prioritized accessible data and demonstrated that they are aware of its importance in laying the foundation for the data economy.
- On the other hand, there are currently not many Public-Private Partnerships that integrate and Open Databases. There was no widespread culture or drive across all industries.

## 5.1. SLR and NLT

Section 3 present a sectoral analysis of the existent literature about Open Data and Public-Private Partnership in different sectors. Here, in the Figure 18 below, is possible to see a summary in order to have clear in mind the main lessons learned.

Dimensions	Results
DATA	<ul style="list-style-type: none"> <li>• Non-interoperable databases</li> <li>• Privacy concern</li> <li>• High investment for cleaning and preparing the database for the integration and opening</li> <li>• Possibility of using the same ontology in order to avoid or solve the non-interoperability limit</li> <li>• Low adherence to existing protocols</li> <li>• Relationship with people and technology is underestimated</li> </ul>
RELATION WITH THIRD PARTIES	<ul style="list-style-type: none"> <li>• Absence of an unambiguous and structured model for data integration</li> <li>• Data providers do not see the benefit in the short term in sharing their data for free</li> <li>• Collaboration for the purpose of openness leads to better infrastructure due to private technical expertise</li> <li>• Local authorities are used to share data with other authorities in the territory, but not to open it</li> </ul>
OPEN DATA	<ul style="list-style-type: none"> <li>• Absence of an unambiguous and structured model for Open Data</li> <li>• Higher involvement of the Central Government, especially in monetary term, would be helpful for speeding up the process of Open Data diffusion</li> <li>• Data providers do not see the benefit in the short term in opening their data for free</li> <li>• Cultural problem</li> <li>• Benefits for civil society that are passive beneficiary</li> <li>• Higher transparency for PA and less corruption</li> </ul>
POSSIBLE SCENARIOS	<ul style="list-style-type: none"> <li>• Italy's guidelines and European directives are pushing for the expansion of Open Data and reusability.</li> <li>• Introduction of a CDO (Chief Data Officer)</li> </ul>

Figure 18: Main results of SLR and NLR

## 5.2. Empirical Analysis

In Section 4 detailed analysis was made for each dimension of the empirical analysis. Figure 19 below summarizes the main results collected from it.

Dimensions	Results
DATA	<ul style="list-style-type: none"> <li>• Respondents are aware of the intrinsic worth of the several datasets they have available.</li> <li>• Enhancing internal corporate databases is typically done for financial gain or to keep track of what has been accomplished throughout the year.</li> <li>• Information was becoming more and more necessary, yet organizations were reluctant to share their information.</li> </ul>
RELATION WITH THIRD PARTIES	<ul style="list-style-type: none"> <li>• The respondents' interactions with third-party databases revealed gaps in their interoperability and the quality of the data of data providers.</li> <li>• Use of many different standards and vocabularies</li> <li>• The initial expenditure required for database integration is one that the public and private sectors struggle to justify.</li> <li>• There were no standard process to support the communication among parties within a PPP</li> </ul>
OPEN DATA	<ul style="list-style-type: none"> <li>• A sort of "jealousy of the data" was found</li> <li>• Poor open data culture</li> <li>• Critical Factors: data availability, confidentiality issue, quality issue, structure of data sources, look for short-term benefits, motivational issue</li> <li>• Successful examples can be used as a baseline to involve more and more actors in partnerships</li> </ul>
POSSIBLE SCENARIOS	<ul style="list-style-type: none"> <li>• There weren't many initiatives with the ultimate goal of merging and opening databases.</li> <li>• The pandemic has expedited the development of digital literacy.</li> </ul>

Figure 19: Main results of the empirical analysis

## 5.3. Main Lesson Learned

Comparing the results of the SLR and NLR it is possible to draw conclusions on the partnership features that need to be enhanced in order to expand the number of participants in PPPs and provide an effective database integration and openness process. The main takeaways to take into account are:

## Lesson Learned

- **No best practice.** A model that establishes a standard of procedure to assist public and private in a partnership that produces advantages for both sides and for civil society might be used to get around constraints described in the previous sections.
- **Project failure.** The results demonstrated that when publicly financed programs are finished, a lack of funding to maintain the infrastructure causes the platforms to shut down and the loss of the high-quality data that had been produced. Supporting organizations to work together to establish an infrastructure in order to provide the data acquired sustainability is a feature of the thesis study. The data gathered is preserved even when public financing stops, and using an open standard method makes it very simple to use data from other initiatives.
- **Win-win approaches.** Since both involved parties must be completely satisfied with the settlement or outcome, a win-win agreement must be the cornerstone of the partnership. It is an integrative process that can lead to a synergy of thoughts, feelings, and beliefs and produces the best outcome. By doing so, it would be possible on one hand to get over the economic constraints cited by private firms, on the other hand generate value-added services to civil society (main objective of the PA).
- **Cooperation.** Both private sector and public sector lack all necessary competencies to effectively implement projects involving the integration and opening of data sets. Therefore, it is essential to bring everyone's skills to the partnership in order to prevent the project from failing soon after the round of public funding.
- **Open data culture.** Taking this aspect of analysis into account, collaboration is required. The end user must become an "accomplice" in the entire process without complaining about the extreme transparency provided by the PA and some related problems, such as the intervention of public maintenance. The PA must start training courses for the best management of its databases. Private companies must concentrate more on the long-term returns resulting from the opening of databases.
- **Economic return.** It costs money to harmonize, clean, and normalize databases, and it has been discovered that certain high-value databases are

## Lesson Learned

difficult to access. The central government should take on the responsibility of boosting financed projects in order to support these developments.

- **Data updating.** It is crucial that the project does not fail once the databases have been opened, as in some cases found on the literature. The Open Databases need to be updated consistently and nearly in real time.

## 6. Model development

After the description of the results collected from the literature and from the interviews this section proposes a standardized method for making the sharing path as comprehensive as possible in order to encourage institutions to undertake it. The framework is a supportive tool for those organizations who want to collaborate, create value by integrating their databases and open their datasets. The aim is to standardize the interaction process among the various organizations and to define some suggestions and best practice to overcome the main obstacles that could affect relationship within PPP. It is also important to underline, as already mentioned, that the development of this model has been based on the evidences arisen from the evidence collected both from the empirical and literature analysis.

The first part of the section provides a detailed description of the realized model by going in detail into each section that makes it up. There are no technical-informatics insights because are outside the scope of the project, but the operational-management ones are undoubtedly defined.

In the second part, once the reader has a clear overview on the realized model, there is an explanation of how the model support the actors involved to overcome the main criticalities collected from the theoretical and practical studies.

Finally, the main costs and benefits gained from the whole ecosystem once it is widespread and fully operational is explained.

## 6.1. Operational model

As it is possible to see (Figure 20) from the picture, the model is characterized by three main building blocks:

- *Initiation*. It is the part of the process in which the first activities for the launch of the project are carried out.
- *Main Activities*. It represents all the required steps for launching the initial project and to create LOD
- *Publishing data for maximum reuse*. It represents the fully operation situation once the project generates the desired output (LOD) and the final users can benefit from them.

To get to the core of the model, each block is described in detail at this point.





### 6.1.1. Initiation

The initiation phase is the first stage of the process. The empirical analysis developed led to the decision to treat this process as a real project. In fact, the evidence suggested that using this type of approach would allow for better adaptation and greater dissemination of Open Data. This is true because the concept of project management is to clearly define downstream what the objectives of the work are. In an environment as uncertain as Open Data, the more clarity can be brought the better. Furthermore, the interviews revealed that "opening data to release them within a data lake is inefficient and involves only a few organizations in a possible partnership." As a result, this type of process is required to avoid databases being opened without a specific goal and resources being wasted without reaping benefits. Moreover, the presence of a project would allow all difficulties to be managed in a standardized and common manner. However, in any case, the willingness to start the project, is expressed by one of data provider. The term data providers, in this context, refers to all those actors that possess data that could potentially be shared. As shown in Figure 21, there are primarily four types of these actors: private companies, private company associations, Public Administrations, and Public Holdings. The main distinction between these four entities is that the first two operate in the free market in a dependent manner, whereas the latter, which uses public funds, must adhere to strict regulations. Section 6.3 shows as this affects the process, but for now, it is sufficient to be aware of who the various actors may be.

## Model development

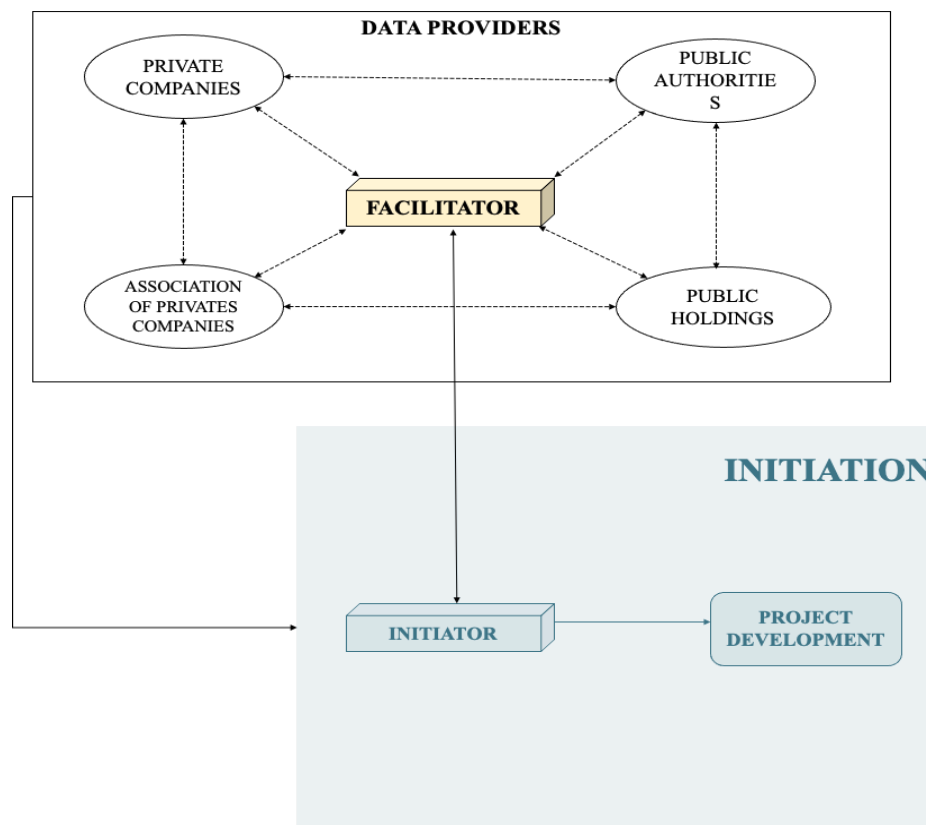


Figure 21: Data providers and Initiation

So, when one of the data providers, just mentioned, decides to start a sharing process, it becomes the Initiator. Each of these actors can, therefore, initiate a sharing project indifferently. Based on who initiates the process, there will be just minor differences in the various activities and some in the economic flow of the project, which, anyway, are explored in more detail later. The Initiator is characterized by a strong Open Data culture, and it fully understands the virtuous circle related to the opening of databases. The initiator recognizes that the partnership is the best way to launch this process effectively, share expenses and risks, and collect increasing amounts of data to combine with those already accessible.

For the time being, it is important to understand that the Initiator, once decides to embark on this path, should first identify a so-called Facilitator. The Facilitator is an extremely important figure. It serves as a support throughout all the process. It can be internal or external to the initiating entity, but the most important thing is that it is an expert in the field of data. Once included in the project, this figure will

## Model development

collaborate with the Initiator to structure all subsequent activities and, most importantly, will support them through his awareness. The facilitator fits into this model with the following functions:

- **Cultural intermediary:** The analysis has demonstrated that a lot of projects did not start or failed because some organizations does not fully understand the benefits from opening data sets. The facilitator, who is knowledgeable about the procedures, acts as a motivator for all potential partners. The objective is to include as many parties as possible while preventing them from walking away from the project.
- **Structural Function.** Many DB should be reorganized from scratch in order to be aligned with LOD policy. To start the database normalization process, the facilitator would offer assistance and training.
- **Communication.** It is fundamental to guarantee a continuous communication among the various partners of the PPP in order to be always aligned.

### 6.1.2. Main activities

The actual activities can begin after the project has been defined and the Facilitator has been chosen. These, as is possible to be seen in Figure 22, are mainly five, respectively data mapping, use case study, engagement, database processing and quality check. These phases of the project are critical because lay the groundwork for integrating and opening the databases. The goal of these activities is, then, to prepare the data to be integrated and opened in order to subsequently form partnerships that recognizes the value it can bring to itself and the community.

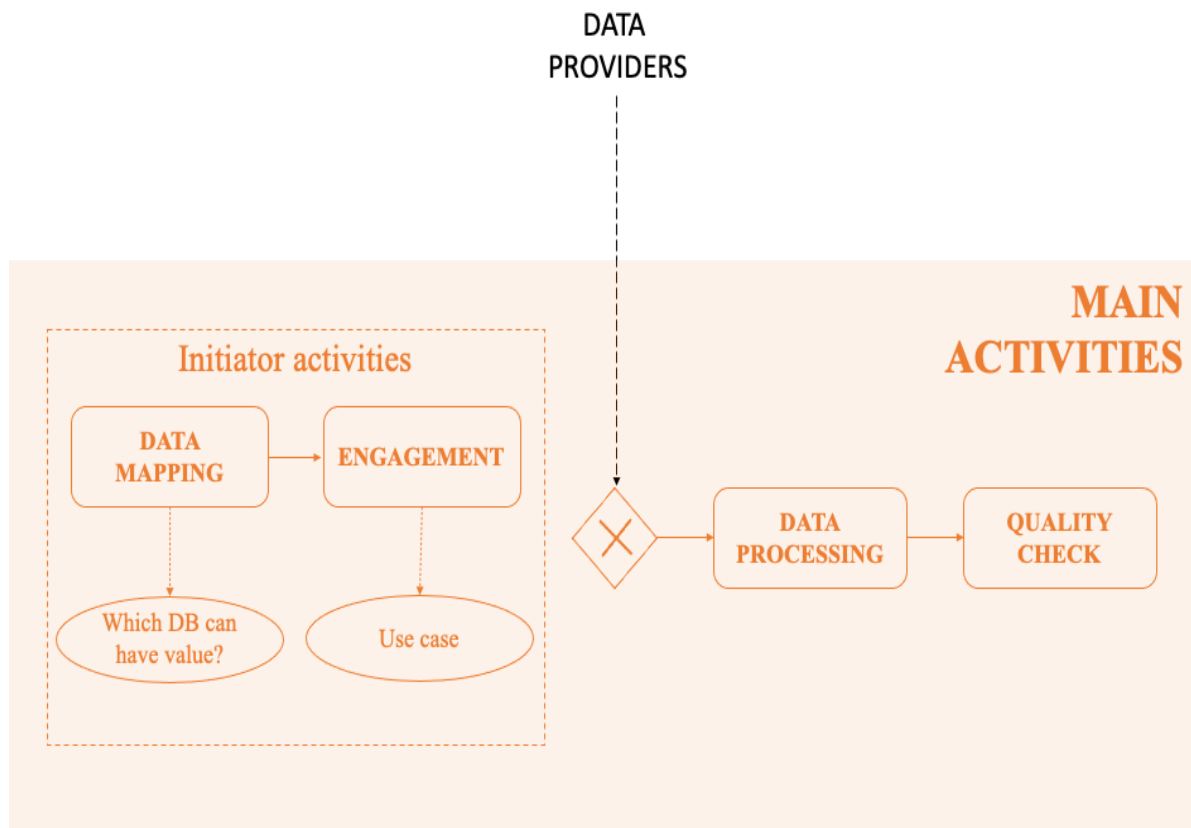


Figure 22: Main activities

## 1. Data mapping

In order to carry out a successful project, one of the most important things to do is data mapping. The term Data Mapping refers to an articulated process based on the union of strategy and technology that allows to select and combine the information required in a single scheme in order to achieve the desired project results. This activity has mainly two objectives. The first is to understand how much and what data the initiating entity has and needs. The second is to understand what data are available in the market and who has them in order to confirm the presence of those of potential interest. Of course, it is critical that the mapping be done clearly and precisely because all subsequent activities will be based on the results obtained. It is critical and fundamental to emphasize that this starting phase is carried out in close collaboration between the initiating actor and the facilitator, who will support and guide them throughout the process. Only later the other players enter the process.

## 2. Engagement

The engagement phase consists, as the name implies in engaging the data providers. This activity is carried out differently depending on who initiates the project:

- When it is the Public Administration or a Public Holding company that starts the project, having to follow well-defined practices, a tender will be necessary. Several private companies may participate in the tender, which will then be evaluated and chosen for the actual project participation. This mechanism is used because, as previously stated, the Public Administration must ensure the competitive dynamics without altering them.
- When it is the private party that initiates the project, the engagement phase is more open and based solely on the PA's interest in participating and in the compatibility of the data they have available.

Moreover, this phase is clearly dependent on the previous phase. On the basis of what emerged from the data mapping, in fact, the data providers who have the information that have been identified as necessary will be selected. But actually, it is also dependent on another phase that is run in parallel. This is the use case study phase. Based on the use case the actors will be presented with the project and motivated to participate.

The purpose of the use case study is to test the project, although in a small way. Based on the findings of the previous phase, then, a reference data provider with which to make this test is chosen. This stage is crucial because it allows for the clarification of the project's objectives, the roles of the actors involved, their responsibilities, and, most importantly, the benefits that may emerge. It allows for the creation of a well-defined and structured context for the project, as well as a solid foundation that can lead to increased awareness on the part of those promoting it.

Hence, the first step is to define a project that will be carried out; secondly the possible data providers of interest are identified and then the use case is made before actually involve the actor in the project. This "reverse approach" is the best

## Model development

way to overcome the skepticism of many Italian-Swiss entities regarding the opening of their databases by demonstrating them the project's potential benefits right away. In conclusion at the end of these phases is possible to engage the actors and proceed with the activities.

### **3. Database processing**

According to the evidence, one of the most important issues is data harmonization and homogenization. At this point all actors, public and private, are involved in the process. The problem is that they use various disparate standards for the databases development. These different standards severely restrict interactions between the public and private spheres. Indeed, most data providers are willing to collaborate, but they do not have databases that are interoperable. And, as we already know, non-interoperable databases cannot be integrated, shared and open. As a result, the harmonization phase is critical to the process. This entails making the databases that will be integrated compatible and interoperable. To accomplish this, the data providers involved, as well as the initiating entity, will need to work together to have their databases processed in accordance with the specifications. By adhering to these well-defined standards, it will be possible to make the databases communicate. Obviously, this phase will take time and is not without complexity, which is why the role of the facilitator will be critical in gaining the support of the participants in this work. The facilitator, in addition to being an expert in the subject matter, has a broad view of the project, which allows this figure to optimize and simplify the process even in these highly complex activities.

### **4. Quality check**

Subsequent to adhering to the interoperability standards to make the databases harmonized, an additional process of checking and preparing them is required. In fact, it is critical to ensure that the database is also consistent with the guidelines for opening the data, which include an important anonymization step. By following these specifications, it is possible not only to make the data formally correct to be opened, but more importantly it is possible to make the data secure and avoid privacy issues. As a result, adherence to standards and the resulting security must be ensured to guarantee high data quality.

### 6.1.3. Publish data for maximum reuse

At the conclusion of the main activities just described, databases will be ready, clean, and secure. They can then be integrated and made open to the public. Starting from the Figure 23 is possible to see how the flow of Open Data works when it is fully operational. In order to better understand this cycle, all the different part that form it will be defined.

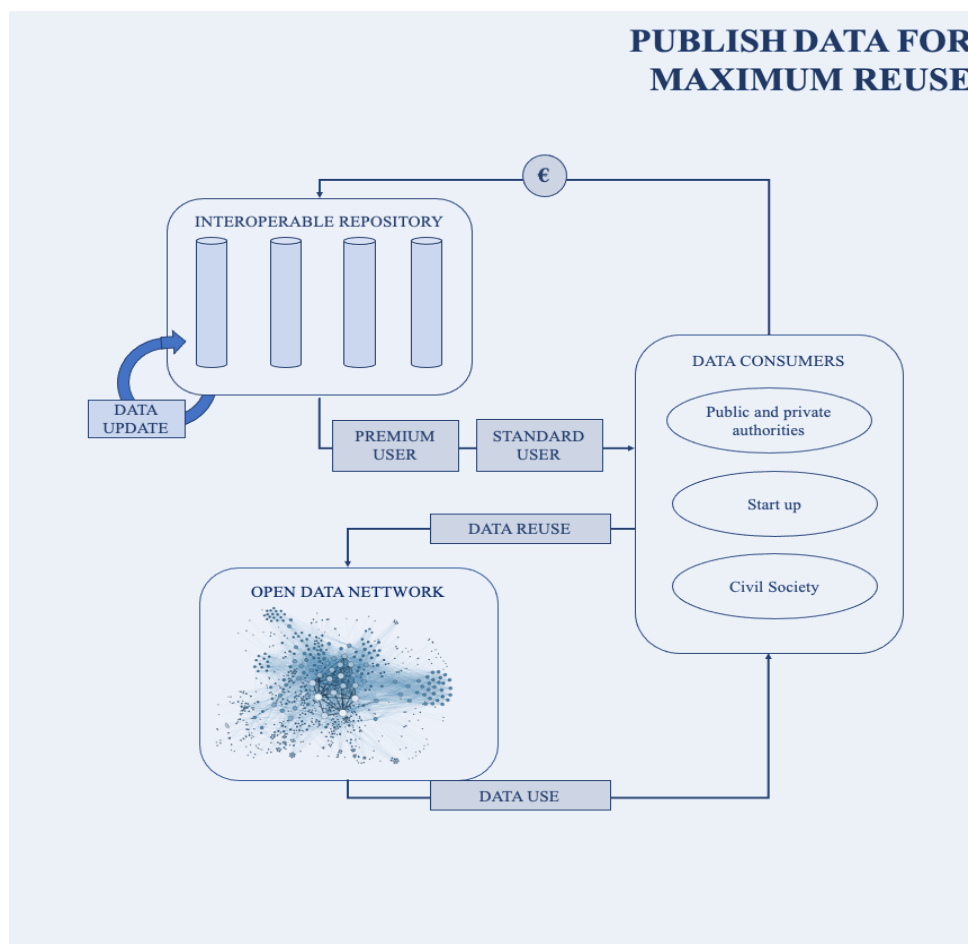


Figure 23: Publish data for maximum reuse

As just saw, after the development of the various preparatory activities, the data are ready to be shared and opened. At this point the process can be divided into two parts. In the first case, the data can be integrated and published on an Open Data network or, in the second case, they can be published on an interoperable platform. These platforms and portals are independent, and whether or not the data is



## Model development

published depends solely on design decisions. The interesting thing, however, is that at this point the platforms and networks even if are independent, are interoperable. That is, integrating the published data into the network or into the interoperable platform would be simple because the data was prepared using the same standards and vocabularies.

### **Open Data network**

The OD network is the set of Open Data portals that are developed by the countries' Public Administration, some of which are described in the discussion. Most of the information contained comes from publicly funded PPPs or directly from the PA itself. The main aim of these portals is to open public sector information with a double objective. Firstly, in order to increase the transparency of the Administration and, secondly, to maximize the reuse of data for the improvement of the services. The reuse of the information, in particular can be done by collecting and publishing different data catalogs to a wide audience for their further reuse. For example, private companies can access these sources in order to use them for develop new services or products or improve the already existing ones. It is important that the data is reused for creating value because, otherwise, it was opened for no reason. Reuse activity, as we have just seen can be done by the intervention of private companies, but individual citizens, start-ups or other Administrations can also participate. It is not important who uses the Open Data, but it is important that the use of this OD creates value. As previously demonstrated, value can be created by improving a service or developing a new application or product. In summary, with anything that has the potential to generate value. For this reason, is important to establish cooperation between data publishers, data consumers and final users (Figure 23). To this end, it is also important to intervene at the cultural level, as it is better explained later, to encourage the use of OD to fuel this virtuous cycle.

Finally, from the analysis emerged that Governments are moving towards e-government and the presence of open portals aims to facilitate these changes.

### **Interoperable repository**

The interoperable repository, instead, is the set of OD portals that are developed by the private organizations. It is done by those companies who want to start opening their own databases in order to gain benefit from them. The interoperable platform is an infrastructure for the management of Open Data, but differently from Open Data portals it is owned and developed by a private entity. It is a platform with two levels of management:

- The first tier, that standard users can access, is a repository where public and private data are made available as an output of a specific project. Data consumer finds them in machine-readable format and can use the data however they see fit. But, usually, the standard users do not directly access the data, they have a more indirect relationship with them.
- The second tier, instead, is accessible by premium users. These users will incur in an annual cost for the access to the information. This is done because in this manner the owner of the platform receives money with which is able to guarantee a high Service Level Agreement (SLA) as the data is processed and updated in real time. These are LODs of the highest quality whose reliability is guaranteed by the platform itself, but this concept is detailed later in the paragraph related to the costs. For instance, a premium user could be a startup that needs machine readable data to use as input for its applications. The ideal would also be to obtain “algorithms as a service” that would provide a reliable output to the premium community. In this way there would be a value created for the owner of the platform who receives positive cash flow, for the user who would obtain output of a certain algorithm instead of internally replicating the structure of the algorithm, civil society as in most cases applications based on Open Data provide services to citizens.

So, from the above description is possible to understand who will benefit from the process once the data is made available. Here there are actually three beneficiaries: the Public Administration, the private sector and civil society. The benefits of these parties are described in detail in a separate section, but it is important to point out that the main goal of the whole process is to create an improvement in service for

## Model development

the citizen. This happens because through this, private individuals and Public Administrations will be able to improve their operations and will be able to provide higher quality products and services, increasing the satisfaction of the end user, the citizen. It is a passive beneficiary because, precisely, the citizen does not actually enter into the merits of the various activities since he does not participate in them, but he will still be able to see and be positively affected by them. In fact, civilization can be divided into two categories. That of so-called standard users and that of so-called premium users. Standard users are those previously described, whereas premium users are the ordinary citizens who want to use the data directly. In fact, the latter benefit from the infrastructure that houses this Open Data by consulting it and using it for personal purposes, such as developing an application. These citizens are typically technology hobbyists or developers with little knowledge of the subject. In any case, while there is a win-win situation for developers, the process is designed to meet the needs of the end user.

As we have seen, once the data is published on the platforms described above, it enters a self-feeding virtuous cycle. This means that the data will be used and provided continuously by the data providers who have participated in the project to benefit from it, but it will also be used by premium users who will create additional value for the community that will passively benefit from it.

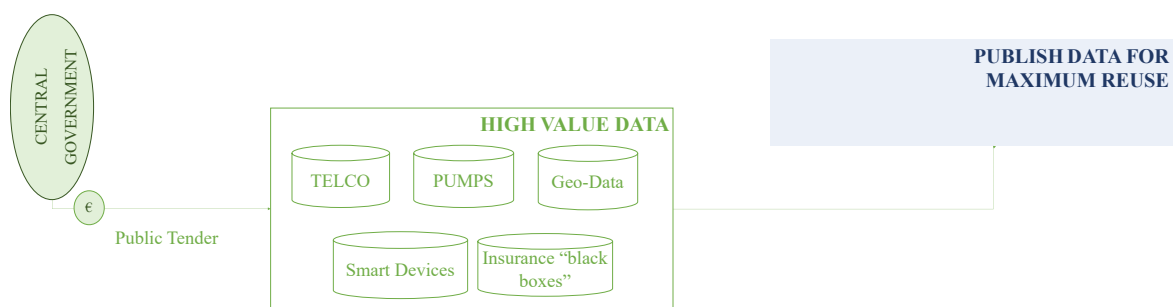


Figure 24: High value-added data

Moreover, in the model a block called High Value-Added data is represented. It was not presented before because it includes data that can be added and introduced in the process in order to give an even higher value, as the name suggest, but are not essential for the cooperation. For this reason, before going on with the cost and benefits paragraph this block is defined. When the results of the interviews were

## Model development

analyzed, the concept of High-Value Data became apparent (Section 4). The owners of these data of high value are characterized by strong jealousy. They are, in fact, unwilling to integrate or make them accessible unless with an immediate financial gain. This jealousy has arisen due to the high intrinsic value of these information, even if they are economic by nature. Precisely because of these characteristics, it would be optimum if these databases were open. This is because, due to their really high value, the information contained within them might be used by outside parties to develop new applications and value-added services for the community.

Figure 24 shows the most relevant databases that emerged from the empirical study:

- Telephone company data
- Insurance black box
- Data from urban mobility plans
- Data from smart devices

Currently, these databases are only accessed through a direct sale and only the buyer who has made a large investment can benefit from them. With reference to the usefulness of these databases, it was decided that the central government intervention would be necessary. In the model created (Figure 20), after the project has begun and its value has been demonstrated, the central government should step in by funding the investment for the use of this data with a community purpose. More specifically, since public money will be used, public tenders will be created according to the necessary information to allow the owners of these databases to participate. The winner of the tender will open in LOD format the information and will keep the dataset updated for a certain period, receiving an economic return and high visibility.

## 6.2. Criticalities and related solutions

In this section, having an overview of the developed model, the criticalities that created the need to include the different parts are described and motivated. The Table 6 shows the main criticalities that created a well define need to find a solution: impasse, culture, non-interoperability and costs.

## Model development

CRITICALITIES	SOLUTIONS
<i>Impasse</i> : uncertainty related to a lack of structure and contextualization	<b>Use Case; Facilitator</b>
<i>Culture</i> : lack of knowledge with the topic, existing laws and related benefits	<b>Training</b>
<i>Non – interoperability</i> : not following the guidelines to facilitate the datasets integration	<b>Facilitator</b>
<i>Economical</i> low motivation as a result of uncertainty about the Return On Investment (ROI)	<b>New central government policy</b>

Table 6: Criticalities and related solutions

### Impasse

The first issue is what is known as an impasse. This term refers to an uncertain and complex situation that results from a failure to structure and define the context. Indeed, in this case, the world of Open Data within PPPs is precisely uncertain and complex. This is because it is a relatively new concept, as we have seen, but mainly because it has been approached differently and independently according to individual needs. This has resulted in a great deal of confusion and a lack of structure in these matters. What is required is a unified approach to simplifying, structuring, contextualizing, and supporting these processes. To overcome these problems, two solutions have been introduced. The first is the figure of the facilitator, and the second is the use of use cases.

As previously stated, the facilitator serves as a professional support figure. This person can be chosen from within the initiating organization or from outside, such as a consulting firm. It is not so much where it comes from as it is knowledge of this figure. In fact, it must be well-versed in the subject of data in general, and Open

## Model development

Data in particular. The facilitator will be included in the project from the beginning, as has been previously shown, and will participate in the definition of the project itself and then support it in all its parts. Thanks to his great knowledge he will be able to bring and define a project context in a satisfactory way in order to make the whole process as easy as possible. Furthermore, he will serve as a liaison between the various participating actors in order to demonstrate them the benefits of the project; in order to maintain constant communication about the events and in order to provide support on all sides.

Applying use cases, on the other hand, makes it possible to verify even before the project begins what its consequences will be. That is, the actors involved, their roles, the necessary investments, but most importantly the benefits that derived from it and the way in which it is carried out are well defined later. In addition, a test will be done to verify its effectiveness. In this way everything is defined and proven. This will give a great deal of structure to the project and will keep potential data providers from back down because they do not know what they are up against.

## Culture

The empirical analysis showed that there is little awareness of Open Data. The lack of knowledge, on the part of organizations but also civilization, leads to being untrusting and reluctant towards these issues. This resistance then does not allow people to open up, accept and enter into these innovative projects, leaving room only for those who manage to go beyond the common imagination. This has resulted, therefore, in little adherence and acceptance.

The main problem is that, since the ultimate goal is to create value for the civil society, if there is no knowledge and then acceptance in the first place on their part, it is difficult achieve success. This was demonstrated by one interviewee who offered a very illustrative practical example. That is, if Open Data on the current state of roads were made available, a citizen may become aware of a pothole on a specific street. The citizen at this point is aware that the Public Administration is informed and expects timely action. However, as is possible to imagine, this is not feasible because interventions take time. The citizen may be annoyed by this lack of intervention and may complain. What this example shows is that the PA is hesitant

## Model development

to open up for fear of a negative reaction from citizens. The citizen, on the other hand, should recognize that by doing so, the PA will be made aware of the pothole and, albeit in its own time, will be able to repair it. In contrast, if that data had not been opened, the PA might not have been aware of it, or at least not in the short term, and the work to fix it would have taken even longer. All of this is to say that if citizens were better educated, it would be possible to help them understand and appreciate the potential benefits of openness and sharing. The knowledgeable and favorable end user motivates stakeholders to go down this road.

Although to a lesser extent, the cultural problem exists within the organizations themselves. Here, however, it is not just properly a problem of knowledge, which is scarce anyway, but more a problem of adaptation to change. Private individuals and PAs have long been accustomed to viewing data as a valuable asset to be sold. It is difficult to separate the common imagination from this point of view. This is why efforts should be made internally to raise awareness of Open Data and collaboration, as well as to demonstrate the potential and real benefits in order to persuade data providers to pursue alternative and innovative paths. As a result, it is critical that also the organizations themselves are made aware. Another problem inside the organization are the people. Unfortunately, if employees do not change their mind toward a more digital world, it is really difficult for the company to move in that direction. They must have a flexible mind, in order to adapt to changes. This do not happen usually, Because people that are used to work since different years in a specific manner, when are approached by a new technology are not ready. For these reasons, changing the cultural paradigm is fundamental.

As a result, a portion of the project budget will be allocated to training.

### **Non-interoperability**

In recent years, as much as the problem of poor contextualization and structuring is present, improvements have been made. Guidelines and standards for data sharing and Open Data have pushed the industry toward standardization. In fact, the purpose of these documents is to provide methodologies for homogenizing data in order to make them interoperable. The problem, however, is also related to the cultural one just seen. Because of a lack of knowledge, even these tools that could

## Model development

greatly simplify the work and speed up the dissemination of OD are not known. For this reason, the problem of non-interoperability has emerged as one of the most important ones, and here again training is crucial, in order to be more aware. At the same time, the role of the facilitator will play an important role in speeding up the process within the project itself. As previously stated, is an expert figure who is aware of these standards. As a result, he will share his knowledge and assist data providers in the path of data homologation to reference standards throughout the process. This figure is critical because he serves as a link between the various operational and structural functions. Only through his presence will be possible to simplify the overcoming of such a big problem as non-interoperability.

### Costs

The last but not least problem is economic. Data providers, particularly private companies, are concerned about the prospect of a return on investment. As previously stated, they have become accustomed to viewing data as valuable assets to be sold rather than shared for free. To understand the benefits of such an action, it is necessary to work on a cultural level. It is therefore critical to educate data providers in order to demonstrate what they can gain in the long run.

The public's economic situation is different. The ultimate goal of the PA is to satisfy the citizen, and funds are allocated to various activities with this end goal in mind. If only citizens valued this activity and recognized the significance of PA digitization, they would be more empowered to allocate a majority share to openness, sharing, and Open Government. The public's intervention could then be used to push towards OD especially at this early stage when trust is still low.

### 6.3. Benefits and costs

As previously stated, the economic issue is extremely important. It can actually be addressed in two ways, depending on who initiates the project. The economic flow, indeed, may differ based on that. This section depicts the flow in both scenarios, as well as the benefits that justify, motivate, and make the investment sustainable.



## Model development

### Initiator: Public Administration or Public Holding

In the case where the process is initiated by the Public Administration or a Public Holdings the flow is shown in Figure 25 below.

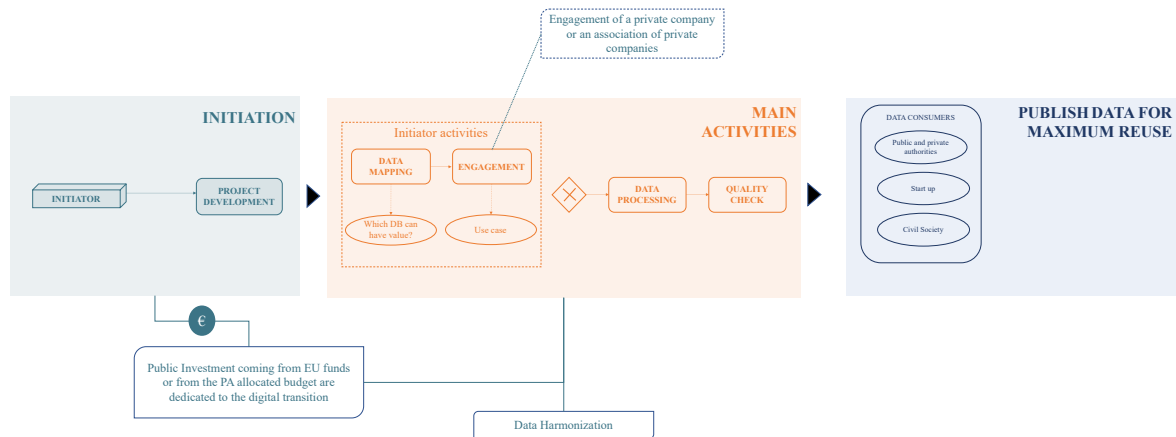


Figure 25: monetary flow for PAs and Public Holdings

Because these authorities are public, they use public funds. As a result, they are unable to purchase independently and must adhere to certain procedures. As shown in Figure 25, indeed, a tender is launched during the engagement phase to accomplish this need. The tender is required because, when private actors are involved, they cannot be chosen independently and must be able to compete in order to ensure a balanced challenging market. As a result, this activity is unique to this situation in which is the public that initiates the process. Another peculiarity of this situation is the economicity. When it is the public that decides to undertake this path, it is the PA itself that will finance it. This financial support can come from European funds earmarked for digitization or from a portion of the budget allocated specifically for digitization by the PA in question that initiated the project. The only activity that will be funded by the private party joining the project is data harmonization. This decision was made because harmonization is an internal activity, which has to be done by the data provider of interest; therefore, it will be self-funded. In any case, however, the facilitator figure will also be supportive to private during this phase, to make sure that there is always alignment and communication between the different parties.

## Model development

At this point, it is important to understand why the public and private sectors, albeit in a small way, should invest this money. The benefits that justify this investment are different and are all summarized in Figure 26 below. They are now analyzed in detail.

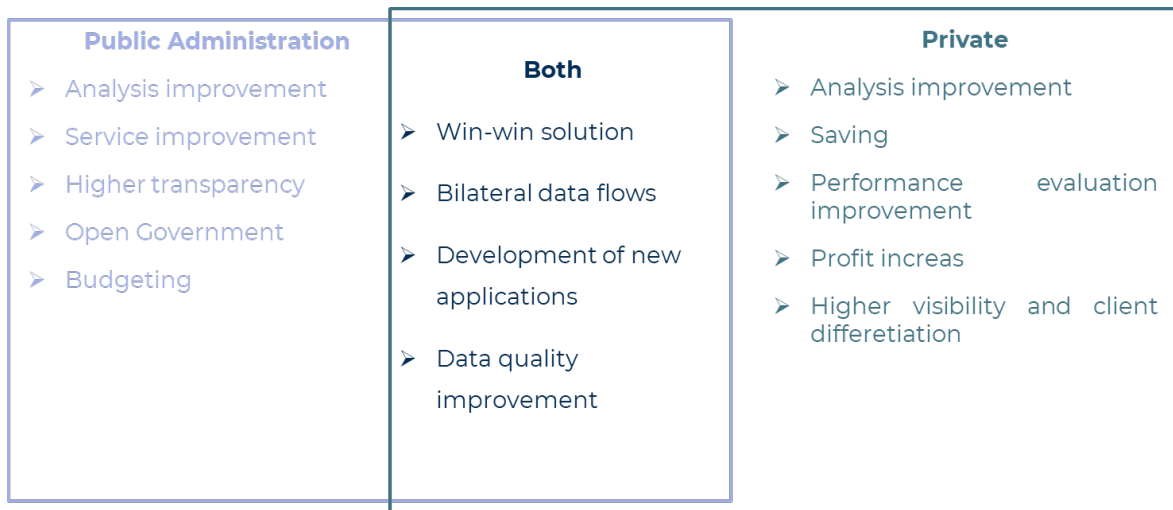


Figure 26: benefits for PA and private companies

Starting from the Public Administration, the initiator, the benefits depicted in Figure 26 are different. First of all, by combining PA data with that of private companies, the Public Administration would have access to a wealth of information. With this vastly greater amount of data than they would have had without private participation, they can do market, feasibility or any kind of analysis in a much more realistic and reliable way. The direct consequence of an advancement in information analysis is an improvement in the services they make available to society. Taking up the example shown just above about the pothole in the street, through the integration of this private information, the PA would be able to repair the problem and improve the service it provides. The great amount of information will be useful also for the budgeting and all the other activities that involve data. Another significant result of data integration and opening is an increase in the PA's transparency. According to the literature review, the PA has been attempting to get as close to the concept of transparency as possible in the last year. The principle of transparency requires the PA to publish specific information, as was already mentioned. Although published, these are not in an open format. Additionally, in Italy for example, according to two distinct laws, information has to be shared

## Model development

differently. In the first case, for the “Diritto d’accesso”, citizens are free to request access to documents that are not protected by state secrecy. They must, however, demonstrate their motivation in order to access them. It takes a lot of time and money for the PA to analyze the requests, find the information, and then prepare them for be shared. In the second case, for the “Accesso civico”, information must be granted without any verification of the reason for which the data are needed. In the latter scenario, the PA attempts to foresee them by directly publishing the information on its website rather than analyzing each individual request. However, transparency would benefit the PA greatly in both situations. First, by having pre-built databases, it can provide information to citizens without first having to prepare and clean them. Additionally, by making the information as accessible as possible, PA avoids dealing with numerous requests from citizens, saving both time and money for the analysis. So, this model allows for greater transparency since all the information will be open, but in a safe way. Moreover, improved management and a decrease in corruption follow greater transparency. Finally, this will allow PA also to digitize and align with the concept of Open Government.

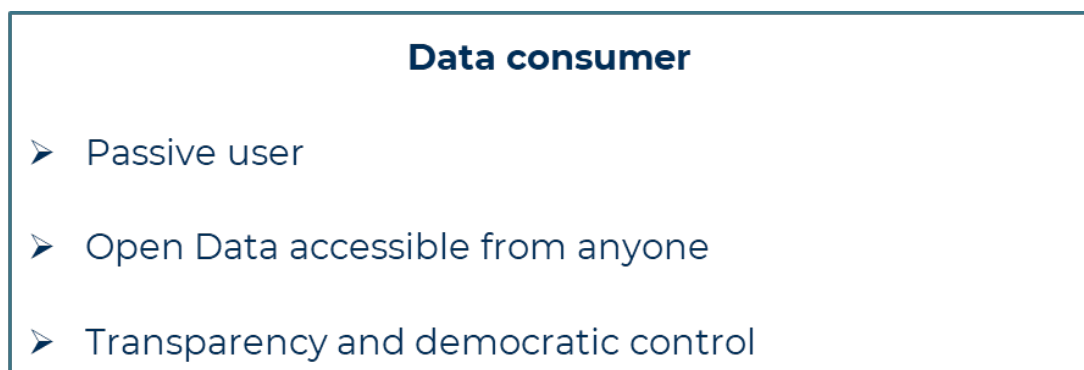
Moving on to private companies, the Figure 26 previously shown also their advantages are depicted. Similar to PA, the private sector will be able to develop high-quality analyses due to the increased amount of data obtained through integration. The private party will be able to save a significant amount of money as a result of these extremely reliable analyses. For example, consider a company that makes annual forecasts of the quantities of raw materials to be purchased for product development. Because of the large amount of high-quality information, these forecasts will be far more reliable than those conducted without the integration. This high quality allows for the purchase of the correct quantity of raw materials, avoiding unnecessary spending by purchasing more than is required and losing market share by purchasing less than is required. Another significant advantage is continuous improvement. The volume of data available also enables in-depth analysis of company performance and the products or services they provide. This allows the company to continuously improve its offerings, increasing the end customer's satisfaction. To top it all off, another significant benefit is that the company becomes much more visible by sharing its data and opening it up to the public. For example, if a company opens up its data on charging stations (where

## Model development

they are located and how many there are) for electric vehicles, which are still relatively scarce in the territory, it would be doing a form of silent publicity by demonstrating that there are indeed numerous charging stations owned.

As it possible to be seen in Figure 26 above, there are benefits that are shared between PA and private company. Although the PA invests public funds in the project in this case, it is still a win-win situation. Indeed, as we have seen, having a two-way data flow, for example, there are advantages for both parties. Furthermore, both the PA and private parties can benefit from the large amount of data by developing new applications. Finally, because they must adhere to strict regulations, the data they will possess will be of high quality, and continuing to pursue this model will allow the circulation of more and more clean, high-quality data.

Finally, in Figure 27 below the benefits of the data consumer are represented.



*Figure 27: Data consumers' benefits*

In this instance, the data consumer is not a true consumer. He is largely a passive beneficiary and is represented by civil society. This is due to the fact that the average person is more interested in passively profiting from Open Data than actively consulting and using it. That is, the moment the PA and private individuals succeed, through this process, in improving their products and services, society will be more satisfied and thus benefit. Anyway, there will still be an opportunity for them to openly consult all the information in case they are interested in it. Moreover, the opening also enables the PA to maintain complete transparency with the society. Transparency is highly regarded by citizens, as is well known. Therefore, the

## Model development

advantages that data consumers can derive from the process are extremely significant.

### Initiator: private company or association of private companies

In the case where the process is initiated by a private company or an association of private companies, the monetary flow is depicted in Figure 28 below.

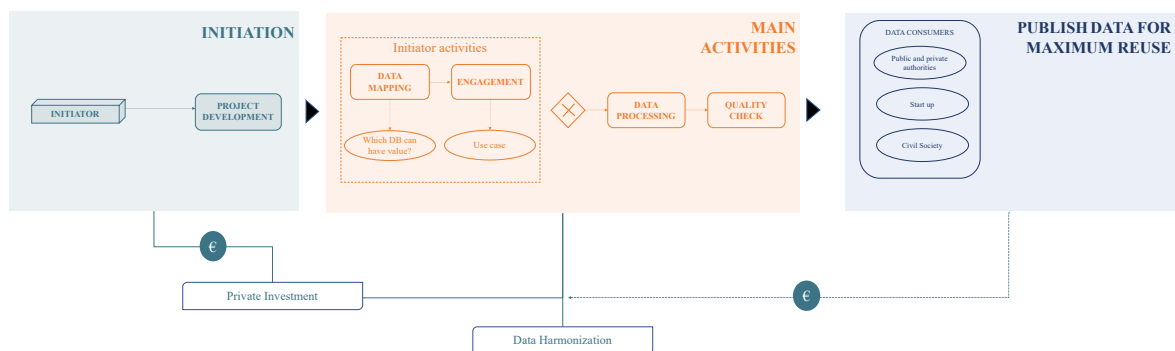


Figure 28: monetary flow for private companies and association of private companies

In this instance, a single private person or group of individuals made all of the investment. The PA will be responsible for paying the costs of data homogenization when it joins the project during the engagement phase for the same reason as stated above. This is the only part of the process that is auto financed by the public data provider.

During this process, two things primarily change. The first is that an interoperable platform needs to be created in this case, as opposed to the prior one where reference Open Data networks already existed. The private party will then need to make a financial investment for building it. The second distinction is in the data consumers. As was mentioned before, there will be not only standard users, the civil society, but also premium users. Premium users are unique users with deeper knowledge of data. Because of their knowledge, they can take advantage of OD to generate new information and giving it new value. However, occasionally they require more refined, precise, and processed data in order to develop an app, for example. They are willing to pay a relatively small sum of money to the private individual who started the project and owns the platform in order to use this service. In this manner, the premium user will continue to finance the platform and

## Model development

the company's investment through the payment of his subscription. As a result, the private company will directly benefit from their investment. Although this might not be sufficient in the short term to cover the whole investment, in the long run it is possible to expect significant gains from the opening, which will make it worthwhile.

The advantages that can be anticipated in this situation are the same as those previously observed. Only a few additional options are available for the private companies and the end user. In particular, for the private company, there will be the possibility of a return on investment, as we have just seen. Most importantly, the private party will be the owner of the interoperability platform, which could always be sold in the event that results are not satisfactory. Given the high caliber of the data on it, there is no doubt that this could generate a sizable return. As for the data consumer on the other hand, in this case the premium user is really a consumer. He will gain a number of advantages from using this information. The first has to do with the fact that this is a thorough source that is simple to get to. The second is that it includes safe and reliable data. The third is that it would not matter how much they pay; it will still be affordable. Last but not least, the premium user could use this data to benefit the civil society as a whole rather than just themselves, feeding the virtuous cycle by, for instance, developing an application.

### 6.4. Final considerations

What is important to highlight is that this model brings together best practices from empirical analysis and literature in order to improve the sharing process. The most crucial factors to take into account are the activities and the order in which tasks must be completed to prevent missing critical components in identifying the important elements that bring various parties together. This model provides, therefore, the starting point for data providers who want to embark on this path in a successful way. Finally, a summary of the model's main flow is presented. It is critical to provide an overview of the work completed for higher clarity.

## Model development

The entire process is started by the initiator, which can be public or private. The very first step is the involvement of a facilitator who will be supportive of the entire process from the very early stages of defining project objectives.

The initiator completes two initial phases independently, with assistance from the facilitator. The data mapping phase that is used to understand the project's information needs, and the engagement phase that is used to recruit organizations to join the partnership.

At this point, the partnership has been formed, and the opening process steps to normalize, align, and clean the databases to be opened begin.

Once preparation activities are completed, LODs have been created and can be published on both existing national and international portals or privately operated interoperable infrastructures.

At this point begins the virtuous cycle whereby these infrastructures are fed by the various participants, including end users.

Finally, from a critical point of view it is possible to say that the value of data always increases the more it is used, this has some fundamental implications. The first one is related to the impact on civil society: in line with the principles of transparency and digital transformation, there would be increasingly easy access to information. The second one is that the presence of a large amount of related machine-readable data would allow further investment by third-party entities to generate both a private economic return and additional and increasingly real-time services to citizens. Finally, after the first successful projects more and more public and private authorities would begin to enter into partnerships and this would generate a self-feeding circle leading the Italian-Swiss landscape to have more and more open, high-quality public and private databases that could be used for multiple purposes.

In conclusion, what is important to highlight is that this model brings together best practices from empirical analysis and literature in order to improve the sharing process. The most crucial factors to take into account are the activities and the order in which tasks must be completed to prevent missing critical components in identifying the important elements that bring various parties together. This model

## Model development

provides, therefore, the starting point for data providers who want to embark on this path in a successful way.



## 7. Conclusion and further research

This research aimed to understand the actual need for information and to investigate the interaction models for Open Data sharing across public and private organizations.

Based on this research theme it was analyzed the actual state of the art and it can be concluded that there is an increasing interest in this topic from both public and private sectors but only a few organizations really integrate and open their datasets with third parties. Best practices and governmental directives already exist, but they are ineffectively used. From our research, it can be stated that the lack of a standard process that directs organizations through the process of establishing databases is one of the key causes of these issues. As a result, there is a propensity for organizations to move toward open data, but at the same time, organizations are constrained by the multiple challenges they could encounter during a potential integration and opening process. Light of the findings of the study, a model was developed to aid any organization (public or private) wishing to begin the process of opening the databases in order to generate value for both the organization itself and civil society.

The model is expected to facilitate various organizations, from the creation of a public-private partnership to the opening of databases, by offering the assistance required to get over potential barriers that stop companies from working on projects of this nature. The desired outcome of the application in real world of the realized model is an enrichment of the LOD platforms, which is primarily driven by a greater understanding of the advantages associated with the opening process.

Finally, the model can be viewed as a starting point that can subsequently be validated in order to widely disseminate the process of sharing and openness as a common practice. Actually, the validation process has already been started; in fact, the model was presented by us at an official Webinar held by the Osservatorio Agenda Digitale of Politecnico di Milano and was circulated to all respondents to get initial feedback. This process is still ongoing and will hopefully continue in order to obtain the most satisfactory results possible.

# Bibliography

- AgID (may 2022), "Piano Triennale per l'Informatica | Agenzia per l'Italia Digitale."
- Andrés Rojas, J., Morlion, P., Tambuyzer, H., Baert, W., Colpaert, P., & Verborgh, R. (2020, June). Velopark: A linked Open Data platform for bicycle parkings. In International Conference on Web Engineering (pp. 53-64). Springer, Cham. Doi: 10.1007/978-3-030-65665-2\_6.
- Aven, T., Renn, O. (2010). Risk management and governance: concepts, guidelines and applications, vol 16. Risk, governance and society. Springer, Berlin. Doi:10.1007/978-3-642-13926-0\_1.
- Bauer, F., Kaltenböck, M. (2011). Linked Open Data: The essentials. Edition Mono/Monochrom, Vienna, 710.
- Bauer, F., Kaltenböck, M. (2011). Linked Open Data: The essentials. Edition Mono/Monochrom, Vienna, 710.
- Bonino, D., Alizo, M. T. D., Alapetite, A., Gilbert, T., Axling, M., Udsen, H., ... & Spirito, M. (2015, August). Almanac: Internet of things for smart cities. In 2015 3rd International Conference on Future Internet of Things and Cloud (pp. 309-316).
- Bouchrika, I. (may 2021) "What Is Empirical Research? Definition, Types & Samples." Research.com, [research.com/research/what-is-empirical-research](https://www.research.com/research/what-is-empirical-research).
- Bouchrika, I. (may 2021) "What Is Empirical Research? Definition, Types & Samples." Research.com, [research.com/research/what-is-empirical-research](https://www.research.com/research/what-is-empirical-research).
- C. Longhi, J.-B. Titz, and L. Viallis. (2014). Open data: Challenges and opportunities for the tourism industry," *Tourism Manage., Marketing, Develop.*, pp. 57–76. Doi: 10.1057/9781137354358\_4
- Caroti G., Piemonte A., 2010. Standard per la generazione del grafo stradale a cui appoggiare rilievi di catasto strade: il caso della Toscana e del Comune di Pisa. *Bollettino della Società Italiana di Fotogrammetria e Topografia* n.3, pp. 69-80.

- Carriero, V. A., Gangemi, A., Mancinelli, M. L., Marinucci, L., Nuzzolese, A. G., Presutti, V., & Veninata, C. (2019, October). ArCo: The Italian cultural heritage knowledge graph. In *International Semantic Web Conference* (pp. 36-52). Springer, Cham. Doi: 10.1007/978-3-030-30796-7\_3.
- Charalabidis, Y., Lachana, Z. (Jun 15, 2020). Towards a science base for digital governance. Paper presented at the pp. 383-389. Doi:10.1145/3396956.3400062.
- Clotet, R., Hernandez, E., & Huerta, M. (2016). EMR system synchronization. Paper presented at the pp. 1-6. Doi:10.1109/ETCM.2016.7750865.
- Colpaert, P. (2017). Publishing transport data for maximum reuse (Doctoral dissertation, Ghent University). Doi: 10.1049/iet-its.2016.0269.
- Colpaert, P., Van Compernelle, M., De Vocht, L., Dimou, A., Vander Sande, M., Verborgh, R., ... & Mannens, E. (2014). Quantifying the interoperability of open government datasets. *Computer*, 47(10), 50-56. Doi: 10.1109/MC.2014.296.
- Conrad, Z., Stern, A., Love, D. C., Salesses, M., Cyril, A., McDowell, A., & Blackstone, N. T. (2021). Data integration for diet sustainability analyses. *Sustainability (Switzerland)*, 13(14). Doi:10.3390/su13148082.
- Davis, A. M., Engkvist, O., Fairclough, R. J., Feierberg, I., Freeman, A., & Iyer, P. (2021). Public-private partnerships: Compound and data sharing in drug discovery and development. *SLAS Discovery*, 26(5), 604-619. Doi:10.1177/2472555220982268.
- De Souza, A., Bittker, J. A., Lahr, D. L., Brudz, S., Chatwin, S., et al. (2014). An overview of the challenges in designing, integrating, and delivering BARD: A public chemical-biology resource and query portal for multiple organizations, locations, and disciplines. *Journal of Biomolecular Screening*, 19(5), 614-627. Doi:10.1177/1087057113517139.
- Ding, L., Lebo, T., Erickson, J. S., DiFranzo, D., Williams, G. T., Li, X., et al. (2011). TWC LOGD: A portal for linked open government data ecosystems. *Journal of Web Semantics*, 9(3), 325-333. Doi:https://doi.org/10.1016/j.websem.2011.06.002.
- Ebert, C., & Duarte, C. H. C. (2018). Digital transformation. *IEEE Softw.*, 35(4), 16-21.
- Espey, J., Badiie, S., Dahmm, H., Appel, D., Noe, L. (2019). Counting on the World to Act: A Roadmap for Governments to Achieve Modern Data Systems for Sustainable Development. Sustainable Development Solutions Network.

## Bibliography

- European Union (2019). The origin of Linked Data. <https://data.europa.eu/en/news/origin-linked-data>.
- European Union (2022), *Indice di digitalizzazione dell'economia e della società (DESI) 2022 Italia*.
- Fredericque, B., Lapierre, A., Byrn, P., 2009. 3D City GIS for sustaining city infrastructure: the Bentley approach, In: ISPRS Archives, Lund, Sweden, Vol. XXXVIII-2/W11.
- Garvin, M. J., Bosso, D. (2008). Assessing the effectiveness of infrastructure public-private partnership programs and projects. *Public Works Management & Policy*, 13, 162-178.
- Gristina, S., Ellul, C., & Scianna, A. (2016, October). Developing a 3d road cadastral system: Comparing legal requirements and user needs. In *ISPRS Annals* (Vol. 4, pp. 223-231). International Society for Photogrammetry and Remote Sensing. Doi: 10.5194/isprs-annals-IV-2-W1-223-2016.
- Guernaccini, F., Mazzini, S., & Bruno, G. (2019). LOD publication in the archival domain: methods and practices. In *ODOCH@ CaiSE* (pp. 15-26).
- Hancock, J. M., Glauco Tocchini-Valentini, Aidinis, V., Schughart, K., Rosenthal, N., Smedley, D., et al. (2010). Sustaining the data and bioresource commons. *Science*, 330(6004), 592-593. Doi:10.1126/science.1191506.
- Harris, J., Burton, P., Knoppers, B., Lindpaintner, K., Bledsoe, M., et al. (2012). Toward a roadmap in global biobanking for health. *European journal of human genetics: EJHG*, 20, 1105-11. Doi: 10.1038/ejhg.2012.96.
- Iijima, C., Morita, T., Enomoto, Y., & Yamaguchi, T. (2011, September). A mobility service based on Japanese linked data. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 246-255). Springer, Berlin, Heidelberg. Doi: 10.1007/978-3-319-06826-8\_5.
- Janev, V. (2020). Open data: Challenges and opportunities for rowth.
- Janowicz, K., Hitzler, P., Adams, B., Kolas, D., & Vardeman II, C. (2014). Five stars of linked data vocabulary use. *Semantic Web*, 5(3), 173-176.
- Jonas, W. (2007), "Research through DESIGN through research: A cybernetic model of designing design foundations", *Kybernetes*, Vol. 36 No. 9/10, pp. 1362-1380. Doi: 10.1108/03684920710827355
- Kim, P. (2017). Linked Data Construction and Utilization of Public Data for Tourist Information Service. *International Journal on Recent and Innovation Trends in*

## Bibliography

- Computing and Communication, 5(2), 68–72.  
<https://doi.org/10.17762/ijritcc.v5i2.170>
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Research*, 49(D1), D1388-D1395. Doi:10.1093/nar/gkaa971.
- Kitchenham, B. (2006). *Evidence-based software engineering and systematic literature reviews* Springer Berlin Heidelberg. Doi:10.1007/11767718\_3.
- Kitchenham, B. (2006). *Evidence-based software engineering and systematic literature reviews* Springer Berlin Heidelberg. Doi:10.1007/11767718\_3.
- Lavery, H., Meulien, P. (2019). The Innovative Medicines Initiative – 10 Years of Public-Private Collaboration. *Frontiers in Medicine*, 6, 275. Doi: 10.3389/fmed.2019.00275.
- Lee, J., An, J., Jeong, S., & Song, J. (2019, November). I-LOD: Industrial Linked Open Data System for Semantic Integration of Industrial Real-time Data in Smart City. In 2019 IEEE International Conference on Industrial Internet (ICII) (pp. 297-298). IEEE.
- Liu, T., Mostafa, S., Mohamed, S., & Nguyen, T. S. (2020). Emerging themes of public-private partnership application in developing smart city projects: a conceptual framework. *Built Environment Project and Asset Management*.
- Machado, A. L., Parente de Oliveira, J. M., (Aug 2011). DIGO: An Open Data architecture for e-government. Paper presented at the 448-456. Doi:10.1109/EDOCW.2011.34.
- Magarey, R. D., Colunga-Garcia, M., Fiesemann, D. A. (2009). Plant biosecurity in the united states: Roles, responsibilities, and information needs. *Bioscience*, 59(10), 875-884. Doi:10.1525/bio.2009.59.10.9
- Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., & Doshi, E. A. (2013). *Open data: Unlocking innovation and performance with liquid information*. McKinsey Global Institute, 21, 116.
- Mastroianni, F. (2018). "Blog | Nel Mondo 2700 Portali Open Data. La Rimonta Dell'Italia." *Info Data*. [www.infodata.ilsole24ore.com/2018/09/20/nel-mondo-2700-portali-open-data-la-rimonta-dellitalia/?refresh\\_ce=1](http://www.infodata.ilsole24ore.com/2018/09/20/nel-mondo-2700-portali-open-data-la-rimonta-dellitalia/?refresh_ce=1). Accessed 12 Sept. 2022.
- Mastroianni, F. (2018). "Blog | Nel Mondo 2700 Portali Open Data. La Rimonta Dell'Italia." *Info Data*. [www.infodata.ilsole24ore.com/2018/09/20/nel-mondo-2700-portali-open-data-la-rimonta-dellitalia/?refresh\\_ce=1](http://www.infodata.ilsole24ore.com/2018/09/20/nel-mondo-2700-portali-open-data-la-rimonta-dellitalia/?refresh_ce=1).

## Bibliography

- 2700-portali-open-data-la-rimonta-dellitalia/?refresh\_ce=1. Accessed 12 Sept. 2022.
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago press.
- Meyerson, L. A., Reaser, J. K. (2002). A unified definition of biosecurity. *Science*, 295(5552), 44-44. Doi: 10.1126/science.295.5552.44.
- Moon, M. J. (2020). Shifting from old open government to new open government: Four critical dimensions and case illustrations. *Null*, 43(3), 535-559. Doi:10.1080/15309576.2019.1691024.
- Mor, A., Kumar, M., & Chaudhury, S. (2021, December). Smart City Umbrella Ontology: Context-Driven Framework For Traffic Planning. In *Forum for Information Retrieval Evaluation* (pp. 83-90). Doi: 10.1145/3503162.3503170.
- Na Lamphun, P. (2016). ENHANCED E-GOVERNMENT TRANSPARENCY THROUGH LINKED OPEN DATA. *Panyapiwat Journal*, 8(1), 225–236. Retrieved from <https://so05.tci-thaijo.org/index.php/pimjournal/article/view/55752>.
- Naoum, G. (2009) *Dissertation Research and Writing for Construction Students*. Oxford, Elsevier/Butterworth-Heinemann.
- Naoum, G. (2009) *Dissertation Research and Writing for Construction Students*. Oxford, Elsevier/Butterworth-Heinemann.
- Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2015). Conducting semi-structured interviews. *Handbook of practical program evaluation*, 492, 492.
- Orr, K. (1998). Data quality and systems theory. *Communications of the ACM* 41(2), 66–71. Doi: 10.1145/269012.269023.
- Ozaydin, B., Zengul, F., Oner, N., & Feldman, S. S. (2020). Healthcare research and analytics data infrastructure solution: A data warehouse for health services research. *Journal of Medical Internet Research*, 22(6), e18579. Doi:10.2196/18579.
- Schöpfel, J., Joachim S. (2010). *Grey literature and professional knowledge making* Springer International Publishing. Doi:10.1007/978-3-319-94177-6\_8.
- Sehgal, R., Dubey, A. M. (2019). Identification of critical success factors for public-private partnership projects. *Journal of Public Affairs*, 19(4), e1956.

## Bibliography

- Shanahan, H., Hoebelheinrich N., Whyte, A. (2021). Progress toward a comprehensive teaching approach to the FAIR data principles. *Patterns*, 2(10). Doi: 10.1016/j.patter.2021.100324.
- Shoja, M. M., et al. (2020). *A Guide to the Scientific Career : Virtues, Communication, Research and Academic Writing*. [Chichester], Wiley Blackwell.
- Shoja, M. M., et al. (2020). *A Guide to the Scientific Career : Virtues, Communication, Research and Academic Writing*. [Chichester], Wiley Blackwell.
- Song, ML., Fisher, R., Wang, JL. Et al. (2018) Environmental performance evaluation with big data: theories and methods. *Ann Oper Res* 270, 459–472. <https://doi.org/10.1007/s10479-016-2158-8>.
- Stone, M., Knapper, J., Evans, G. and Aravopoulou, E. (2018), “Information management in the smart city”, *Bottom Line*, Vol. 31, pp. 234-249.
- Tegan, G. (6 dec. 2021) “A Guide to Exploratory Research.” Scribbr, [www.scribbr.com/methodology/exploratory-research/](http://www.scribbr.com/methodology/exploratory-research/).
- Tegan, G. (6 dec. 2021) “A Guide to Exploratory Research.” Scribbr, [www.scribbr.com/methodology/exploratory-research/](http://www.scribbr.com/methodology/exploratory-research/).
- Tomas, R., Harrison, M., Barredo, J. I., Thomas, F., Llorente Isidro, M., Pfeiffer, M., & Čerba, O. (2015). Towards a cross-domain interoperable framework for natural hazards and disaster risk reduction information. *Natural Hazards*, 78(3), 1545-1563. Doi:10.1007/s11069-015-1786-7.
- Van Ham, H., Koppenjan, J. (2001). Building public-private partnerships: Assessing and managing risks in port development. *Public Management Review*, 4, 593-616.
- Walravens, N., Van Compernelle, M., Colpaert, P., Mechant, P., Ballon, P., Mannens, E. (2016). Open Government Data': based Business Models: a market consultation on the relationship with government in the case of mobility and route-planning applications. In proceedings of 13th International Joint Conference on e-Business and Telecommunications (pp. 64–71). Doi: 10.5220/0005948300640071.
- Williams, A. J., Harland, L., rowth, P., et al. (2012). Open PHACTS: Semantic Interoperability for Drug Discovery. *Drug Discovery Today*, 17, 1188–1198. Doi: 10.1016/j.drudis.2012.05.016.
- Yigitcanlar, T., Kamruzzaman, M., Buys, L., Ioppolo, G., Sabatini-Marques, J., da Costa, E.M. and Yun, J.J. (2018), “Understanding ‘smart cities’: intertwining

## Bibliography

development drivers with desired outcomes in a multidimensional framework", *Cities*, Vol. 81, pp. 145-160.

Yochum, P., Chang, L., Gu, T., & Zhu, M. (2020). Linked Open Data in location-based recommendation system on tourism domain: A survey. *IEEE Access*, 8, 16409-16439. Doi: 10.1109/ACCESS.2020.2967120.



# List of figures

Figure 1:Tasks of the WP3 MAC module.....	18
Figure 2: First search string result on Scopus.....	25
Figure 3: Second search string on Google Scholar.....	26
Figure 4: Third search string on Scopus.....	27
Figure 5: Summary of the articles selected .....	33
Figure 6: Number of articles per year.....	34
Figure 7: Number of articles per sector .....	34
Figure 8: Data collection model.....	41
Figure 9: 5-star Open Data (5-star Open Data (5stardata.info).....	45
Figure 10: European directives .....	47
Figure 11: Common natural hazard assessment model.....	52
Figure 12: Four levels of data (Ozaydin, Zengul, Oner, Feldman 2020) .....	57
Figure 13: Iterative model (Ozaydin, Zengul, Oner, Feldman 2020) .....	59
Figure 14: Type of respondent by country.....	101
Figure 15: Respondents integrating data by country .....	102
Figure 16: Positive response by type of interview .....	104
Figure 17: Negative response by type of interview .....	104
Figure 18: Main results of SLR and NLR .....	106
Figure 19: Main results of the empirical analysis .....	107
Figure 20: Operational Model.....	112
Figure 21: Data providers and Initiation.....	114
Figure 22: Main activities .....	116
Figure 23: Publish data for maximum reuse .....	119
Figure 24: High value-added data .....	122
Figure 25: monetary flow for PAs and Public Holdings.....	128
Figure 26:benefits for PA and private companies.....	129
Figure 27: Data consumers' benefits .....	131
Figure 28: monetary flow for private companies and association of private companies ..	132

# List of Table

Table 1: Publication years SLR .....	28
Table 2: Publication sectors SLR.....	29
Table 3: Publication Years NLR.....	31
Table 4: Publication sectors NLR.....	32
Table 5: Summary of interviews.....	100
Table 6: Criticalities and related solutions.....	124

## List of Abbreviation

<b>Abbreviation</b>	<b>Extended Version</b>
<i>API</i>	Application Program Interfaces
<i>AQIM</i>	Agricultural Quarantine Inspection Monitoring system
<i>AVM</i>	Automatic Vehicle Monitoring system
<i>BOT</i>	Build-Operate-Transfer
<i>C4DC</i>	Contracts for Data Collaboration
<i>CBP</i>	Customs and Border Protection
<i>CDO</i>	Chef Data Officer
<i>DBFO</i>	Design-Build-Finance-Operate
<i>DESI</i>	Digital Economy and Society Index
<i>DGA</i>	Data Governance Act
<i>EFFIS</i>	European Forest Fire Information System
<i>EMR</i>	Electronic Medical Record
<i>ESBB</i>	Middle Eastern and African Society for Biobanking
<i>ESFRI</i>	European Strategy Forum on Research Infrastructures
<i>GDPR</i>	General Data Protection Regulation
<i>GIS</i>	Geographic Information Systems
<i>GTFS</i>	General Transit Feed Specification
<i>HER</i>	Electronic Health Record
<i>HRADIS</i>	Healthcare Research and Analytics Data Infrastructure Solution
<i>ICCD</i>	Istituto Centrale per il Catalogo e la Documentazione
<i>IF</i>	Interoperability Framework
<i>IMI</i>	Innovative Medicine Initiatives
<i>INSPIRE</i>	Infrastructure for Spatial Information in the European Community
<i>IPPC</i>	International Plant Protection Convention
<i>ISBER</i>	International Society for Biological and Environmental Repositories
<i>LOD</i>	Linked Open Data
<i>LOGD</i>	Linked Open Government
<i>MaaS</i>	Mobility as a Service
<i>MiBAC</i>	Ministero per I Beni e le Attività Culturali
<i>NHANS</i>	National Health and Nutrition Survey
<i>NLR</i>	Narrative Literature Review
<i>NPAG</i>	New Pest Advisory Group

<b>Abbreviation</b>	<b>Extended Version</b>
<i>OD</i>	Open Data
<i>OGP</i>	Open Government Partnership
<i>Open PHACTS</i>	Open Pharmacological Concept Triple Store
<i>OPS</i>	Open Pharmacological Space
<i>PA</i>	Public Administration
<i>PPP</i>	Public-Private Partnership
<i>ROI</i>	Return On Investment
<i>SCUO</i>	Smart City Umbrella Ontology
<i>SDGs</i>	Sustainable Development Goals
<i>SFR</i>	Swiss Federal Railways
<i>SLA</i>	Service Level Agreement
<i>SLR</i>	Systematic Literature Review
<i>SSI</i>	Semi-Structured Interviews
<i>SUPSI</i>	Scuola Universitaria Professionale della Svizzera Italiana
<i>TPL</i>	Local Public Transport

# Annex

## Questionnaire

These questions represent the backbone of the questionnaire, but the structure of each question is flexible, and it was aligned with interview characteristics.

### **INTRODUCTION**

1. Do you use the data to further the goals of your business (products or services)? Do you value the data your services produce? (Strategic decision-making, new service development, and policy-making analysis and assistance.
  - a. If yes, when did you start doing it? Have you started any projects using this theme?

### **DATA ABOUT THE ENTITY**

1. What data do you have that you could share with others?
2. Do you own the data you have access and how you obtained it, or do you have agreements with third parties?
3. Do you use technology software to handle the data you manage? Is there a mechanism in place to control data quality?
4. Do you make data available in an Open Data format?

### **OTHER PARTIES DATA**

1. Do you combine your data with data from other users for the purposes of your activities?
2. What information (databases) would you need to start / improve the integration process (European, Swiss / Italian)?
3. Considering the pandemic experience, which data do you believe may be useful? Which data would have been helpful to you during the pandemic?

## **DIFFICULTY AND PROSPECTS FOR THE FUTURE**

1. What are the key challenges you've already faced or expect to face in the data integration process?
2. Do you have any future plans for this based on what you are now doing?