



TECHNISCHE  
UNIVERSITÄT  
WIEN

MASTERARBEIT

# **Integrating Geospatial Linked Open Data and Knowledge Networks into Location Business Intelligence**

Ausgeführt am Department für  
Geodäsie und Geoinformation  
der Technischen Universität Wien

unter der Anleitung von

**Dipl.-Ing. Dr. Markus Jobst, Bundesamt für Eich- und Vermessungswesen in Österreich**  
und

**Univ.Prof. Mag.rer.nat. Dr.rer.nat. Georg Gartner, TU Wien**  
**Dr.-Ing. Ekaterina Chuprikova, TU München**

durch

**Olha Ostrovna**

Matrikelnummer 11842020

Wien, 10.09.2020

---

Unterschrift (Student/in)

---

Unterschrift (Betreuer/in)



TECHNISCHE  
UNIVERSITÄT  
WIEN

MASTER'S THESIS

# Integrating Geospatial Linked Open Data and Knowledge Networks into Location Business Intelligence

Conducted at the Department of  
Geodesy and Geoinformation  
Technical University Vienna

Under the supervision of

**Dipl.-Ing. Dr. Markus Jobst, Federal Office of Metrology and Surveying in Austria**  
and

**Univ.Prof. Mag.rer.nat. Dr.rer.nat. Georg Gartner, TU Vienna**  
**Dr.-Ing. Ekaterina Chuprikova, TU Munich**

by

**Olha Ostrovna**

Matriculation number 11842020

Vienna, 10.09.2020

---

Signature (Student)

---

Signature (Supervisor)



**Cartography M.Sc.**

**Master thesis**

# **Integrating Geospatial Linked Open Data and Knowledge Networks into Location Business Intelligence**

Olha Ostrovna



2020

## ACKNOWLEDGEMENTS

The completion of this thesis involves the participation and assistance of people who made this journey to be special, unique and exciting. I would like to take a moment to express my sincere gratitude to everyone.

First, I would like to express my deepest gratitude and appreciation to my first supervisor and deputy head of the Information Management Department at the Austrian Federal Office for Metrology and Surveying, Dr. Markus Jobst. Thank you for always providing me with excellent supervision, properly-directed questions, up-to-date ideas, sharing valuable insights and giving useful comments.

I would like to thank a lot to Prof. Georg Gartner for his constant involvement and useful remarks and critiques at each stage of the research process, for support and encouragement, ensuring that I am moving in the right direction.

Our regular virtual discussions worth particular attention. Such meetings were essential in all the meanings during such uncommon conditions of 2020. Thanks a lot for always friendly and well-organized Jour Fixes, Wangshu Wang, and your immediate response and assistance.

I would like to show my sincere thanks to Dr.-Ing. Ekaterina Chuprikova for reviewing my research steps and ideas, assisting my work with good recommendations and paying attention to details.

Moreover, I am grateful to everybody within the Cartography MSc Consortium for this unique chance and experience.

Thanks a lot to everyone involved in this programme, from lectures to classmates, for fantastic time, unforgettable memories and exclusive experience expanding borders. I express my particular gratitude to the programme coordinator, Juliane Cron, who accompanied me through my Master's journey.

Furthermore, I want to show special appreciation to Yevgeniy Bartosh, my Geography teacher back from school times in Ukraine, for his inspiring approach that contributed to my achievements.

Finally, I express my hearty gratitude to people closest to me for being always by my side, even remotely. Thank you, dear Mykola Kozyr, for your support in all the meanings. Many thanks to my family for unconditional love.

## ABSTRACT

With the development of a breaking theory of Linked Open Data (LOD), many initiatives and projects appear and adopt the LOD concept to geospatial data publishing on the Web. The main idea behind is a provision of integrated access to geospatial data coming from heterogeneous sources. Such a concept aims to assure convenient reuse and interoperability of location data within various applications and domains. There are a significant number of domains where applying the Linked Data principles would be beneficial. Particularly, Location Business Intelligence (BI) comes across as one of the interested parties able to transform the potential of Geospatial LOD into valuable decisions and solutions.

However, the novelty and complexity of the geospatial direction in the LOD concept limit its integration into Location BI. The synthesis of Geospatial LOD and Location BI is mutually beneficial and would provide flexibility and uncover new opportunities for users. Hence, this paper identifies and describes the reasons and problems that cause this integration gap. By considering them, both Geospatial LOD providers and BI tools will improve the adaptation, performance, and scalability of this valuable amalgamation.

The research is based on the exploration and evaluation of geospatial data access and data transfer technologies behind the global BI market leaders. Moreover, it examines standards supported by the selected Linked Open Data providers. The approach applied in this paper provides valuable insights on overcoming the existing integration gap between the Geospatial Linked Open Data paradigm and Location Business Intelligence solutions.

*Keywords:* geospatial, Linked Open Data, Business Intelligence, Location Intelligence, interoperability and reusability

ACKNOWLEDGEMENTS	i
ABSTRACT	ii
LIST OF FIGURES	iv
LIST OF TABLES	iv
ABBREVIATIONS	v
1. Introduction	1
1.1 Motivation and problem statement	1
1.2 Research identification	2
1.2.1 Research goals and objectives	2
1.2.2 Research questions	2
2. Theoretical background and related work	3
2.1 Geospatial Linked Open Data and Knowledge Networks	3
2.2 Location Business Intelligence	5
3. Methodology	8
3.1 Thematic scenarios development	8
3.1.1 Scenario 1	8
3.1.2 Scenario 2	9
3.2 Overview of leaders of the global Location Business Intelligence market	9
3.3 Overview of Linked Open Data providers	10
4. Implementation	12
4.1 Data Access Interfaces in Linked Data providers	12
4.1.1 Eurostat	12
4.1.2 European Data Portal	13
4.1.3 Thematic data extraction	15
4.2 Geospatial Linked Data integration via visualization frameworks in Business Intelligence tools	16
4.2.1 Tableau	17
4.2.2 CARTO	18
4.2.3 OmniSci	20
4.2.4 ArcGIS Insights	22
5. Results and discussion	24
6. Conclusion	29
REFERENCES	30
ANNEX 1 – Scenario 1 & Scenario 2 visualizations   Tableau	36
ANNEX 2 – Scenario 1 & Scenario 2 visualizations   CARTO	37
ANNEX 3 – Scenario 1 & Scenario 2 visualizations   OmniSci	38
ANNEX 4 – Scenario 1 & Scenario 2 visualizations   ArcGIS Insights	39

## LIST OF FIGURES

Figure 1. The Five Star Open Data model (CC-0) (Berners-Lee 2006; Berners-Lee 5) applied to spatial data types (McKeague et al. 2020) .....	4
Figure 2. Data Types usage in the Datasets' Distributions .....	26
Figure 3. Data Types usage aggregated by the Five Star Open Data model (Berners-Lee 2006; Berners-Lee 5) applied to spatial data formats(McKeague et al. 2020) .....	26
Figure 4. Distribution of the Number of Supported Geospatial Formats over datasets.....	27

## LIST OF TABLES

Table 1. Spatial data formats available in Eurostat .....	13
Table 2. Geospatial data formats supported by the European Data Portal (EDP) .....	14
Table 3. Spatial data formats available in EDP .....	15
Table 4. Thematic data extracted from Eurostat.....	16
Table 5. Thematic data extracted from EDP .....	16
Table 6. Spatial data formats available in Tableau .....	18
Table 7. Linked Data integration via thematic scenario implementation in Tableau .....	18
Table 8. Spatial data formats available in CARTO.....	20
Table 9. Linked Data integration via thematic scenario implementation in CARTO .....	20
Table 10. Spatial data formats available in OmniSci .....	21
Table 11. Linked Data integration via thematic scenario implementation in OmniSci .....	22
Table 12. Spatial data formats available in ArcGIS Insights .....	22
Table 13. Linked Data integration via thematic scenario implementation in ArcGIS Insights ....	23

## ABBREVIATIONS

API	Application programming interface
BI	Business Intelligence
EFTA	European Free Trade Association
ETL	Extract, Transform, Load
EU	European Union
GIS	Geographic Information System
GISCO	Geographic Information System of the Commission
GPU	Graphics Processing Unit
GUI	Graphical User Interface
ILO	International Labour Organisation
INSPIRE	Infrastructure for Spatial Information in the European Community
LOD	Linked Open Data
NUTS	Nomenclature of Territorial Units for Statistics
OSM	OpenStreetMap
SaaS	Software as a Service
SPARQL	Query language for linked data (RDF)
SQL	Structured Query Language
W3S	World Wide Web Consortium



# 1. Introduction

This chapter presents the context in which this research is set. It explains the research motivation and states the research problem. Afterwards, the research objectives and research questions are specified.

## 1.1 Motivation and problem statement

Technological advancements and rapid development of open-source frameworks allowed to publish and access geospatial data on the Semantic Web much easier and by many more users (Mai et al. 2019). This enabled the world of the Web to be constantly enriched with a huge amount of data. Very often some data might be similar, correlate, or even overlap with other data that makes it more complex. Along with this, “data heterogeneity has surpassed the capacity of current data management technologies” (Maté et al. 2012). Even the biggest companies and the most advanced Business Intelligence (BI) systems try to avoid storing, processing, and managing all this data in local servers (Maté et al. 2012). This phenomenon became one of the driving forces for developing the breaking concept of *Linked Open Data* and its geospatial extension, *Geospatial Linked Open Data*, later.

Many public organizations, governmental institutions, and communities realized the importance of avoiding hidden and isolated data. Following this trend, some of them decided to publish and share on the Internet the data and information they manage as Linked Open Data (Maté et al. 2012). Most of the geospatial data available as Linked Open Data is provided and supported by governmental initiatives. The idea behind states the increase of transparency, and more importantly, data reuse and data interoperability.

Nevertheless, “all this work focuses merely on publishing geospatial data, converting shapefiles to RDF, fusing geometries from different sources (Giannopoulos et al. 2014), discovering links (Mai et al. 2016; Ngomo and Auer 2011), querying remote endpoints (Battle and Kolas 2012), or computing geospatial properties on-demand (Regalia et al. 2016) and getting geodata out of data silos” (Mai et al. 2019).

Centering attention on the mentioned processes, Linked Open Data providers miss out of sight availability of this data for outside use. The question of direct integration of Geospatial LOD with potential stakeholders remains largely unanswered (Mai et al. 2019). And Business Intelligence is one of those industries interested in LOD integration and might transform it into valuable decisions. This trend also applies to Location Business Intelligence. As stated by (Bolon et al. 2012), many leading enterprises and organizations tap location analytics to solve business problems and uncover hidden relationships across business data. To understand the significance of location information, businesses often refer to BI tools powered by GIS. And with an opportunity to accommodate easy access to Linked Open Data, Location BI would help users gain more value out of geospatial data.

According to (Folmer et al. 2019), when Geospatial Linked Open Data can be analyzed and visualized in Business Intelligence tools, “the best of both worlds can be combined. With Linked Data, it is possible to combine a large variety of data and query data at the source. Business Intelligence tools serve as an optimal GUI for the visualization of these data. The data would no longer need to be copied and extracted to data warehouses and could be analyzed and visualized directly from the source. For end-users who want to use the data in business Intelligence tools, the Linked Data technology will become much more accessible. Business Intelligence tools can serve as a ‘Killer App’ for Linked Data and give Linked Data technology a boost” (Folmer et al. 2019).

However, publishing geospatial Linked Open Data into the cloud does not ensure the required reusability and integration into BI solutions. Unfortunately, there is a data integration gap between Linked Data providers and Business Intelligence. This problem exists due to several reasons: on the one hand, Linked Data platforms adopt strict LOD technologies and provide low-level interfaces for data access. Furthermore, sometimes they do not support diverse data transfer standards. On the other hand, BI tools focus on particular data transfer technologies, not always compatible with the Linked Open Data concept. For these reasons, the geospatial LOD and Location BI remain largely separated. All mentioned factors limit interoperability and reusability of Geospatial Linked Open Data within the decision-making domain.

## 1.2 Research identification

This section specifies the research goals and objectives of this study and the extended research questions regarding the objectives.

### 1.2.1 Research goals and objectives

Based on the motivation and problem statement, this thesis concentrates on the following goals and objectives:

- Analysis of current leading BI tools in terms of Geospatial Linked Open Data Integration
- Analysis of relevant Linked Open Data providers in terms of integration with BI tools
- Visualization framework for the integrated Linked Data within the selected BI tool
- Identified reasons limiting the integration of Geospatial LOD with BI software

### 1.2.2 Research questions

Based on the research motivation and goals, the research questions addressed to meet the thesis main objectives are:

- What data transfer technologies currently define the leading BI tools integrations' support?
- What specific data transfer technologies have to be in focus for the Linked Data platforms to provide successful integration use-cases?
- What are the key components of the Linked Data concept required to be integrated into BI tools to provide access to the global Geospatial Linked Data and Knowledge Network platforms?

## 2. Theoretical background and related work

This chapter introduces two main concepts, the integration of which is in the focus of this research paper. The first part focuses on Geospatial Linked Open Data and the diversity of approaches to classify and evaluate it. The identification of the Location Business Intelligence industry is described within the second subchapter.

### 2.1 Geospatial Linked Open Data and Knowledge Networks

“The Semantic Web is not just about putting data on the web. It is about making *links* so that a person or machine can explore the web of data” (Berners-Lee 2006) in an interoperable way. The more data is interlinked with other data, the higher its value and usefulness. *Linked Data* refers to the provision of integrated access to data dynamically aggregated from a wide range of distributed and heterogeneous data sources (Bizer et al. 2009), collectively forming a *network of knowledge* (Mangaladevi et al. 2017). Knowledge networks are expressed through “graphs of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities” (Hogan et al. 2020).

Unlike data stored in a traditional relational database, Linked Data is stored in a graph-based data model, typically within triple stores, using standardized exchange data formats like Turtle and RDF/XML (W3C 2016). Linked Data is identified and linked to other data on the Web via Uniform Resource Identifiers (URIs). Such identifiers establish links between datasets providing extra information, adding standardization, and removing ambiguity. Moreover, URIs, in conjunction with the HTTP protocol, allow further discovery of data available on the Web by potential users. SPARQL is a W3C-standardized query language that helps retrieve and manipulate data stored in RDF format (Folmer et al. 2019).

When Linked Data can be freely accessed, used, and distributed by anyone, it is *Linked Open Data (LOD)*. In other words, Linked Open Data is “a powerful blend of Linked Data and Open Data”: it is both linked and released under open licenses (Ontotext 2020).

The more things, events, people, locations, etc. are connected on the Web, the higher is the need for some particular guidelines on linking, managing, and integrating all this data diversity. The inventor of the World Wide Web and the creator of the Linked Data concept, Sir Tim Berners-Lee (Ontotext 2020), suggested the 5-Star Linked Open Data model indicating criteria that define Linked Open Data (Folmer et al. 2019):

- ★ Available on the web under an open license
- ★★ Available as machine-readable structured data in proprietary formats
- ★★★ Available as machine-readable data, but in a non-proprietary format (e.g. CSV instead of XSLT)
- ★★★★ Using open web standards from W3C (IRIs for identifiers, RDF for data model, SPARQL for querying)

★★★★★ Linked to other Linked Open Data on the web providing content

Under this star scheme, the fourth star and the fifth star are assigned to the data published using the Linked Data rules (Berners-Lee 2006; Ronzhin et al. 2018).

The LOD concept provides a huge potential in various industries, like the *geospatial industry*. Due to constant advances in the Earth Observation field (Koubarakis et al. 2017), the geospatial domain is expressed by complex datasets (Eurostat 2015). Combining all those datasets semantically can assure convenient interoperability of location data. Moreover, most datasets, even non-geospatial, might contain a geospatial dimension that is either explicit or implicit. This means that “location” appears to be an important key for aligning and interlinking different datasets (Roosens et al. 2019).

The 5-Star Linked Open Data model also finds its application for *geospatial data*. Within the archaeological research paper published in 2020 (McKeague et al. 2020), this model appears to classify the technical level of dataset’s advancement containing *geospatial data* (see Figure 1).

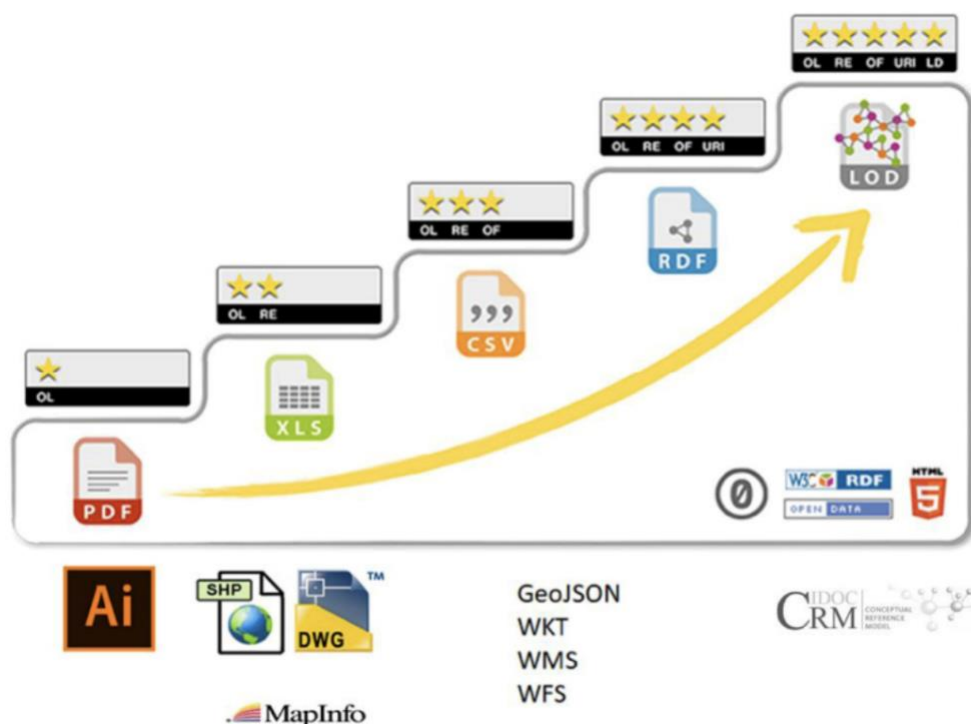


Figure 1. The Five Star Open Data model (CC-0) (Berners-Lee 2006; Berners-Lee 5) applied to spatial data types (McKeague et al. 2020)

Later, Tim Berners-Lee’s 5-Stars schema was complemented by the 5-Star Open Data Engagement model. It aims to highlight key steps that open data initiatives can take to engage with data users. Engaging open data should:

- ★ Be demand-driven
- ★★ Put data in context
- ★★★ Support conversation around data

- ★★★★ Build capacity, skills and networks
- ★★★★★ Collaborate on data as a common resource

“Each star includes a set of questions to unpack what might be involved in taking that step towards” (Davies 2012).

Tim Berners-Lee laid down the Five Stars of Linked Open Data mostly from a technological perspective, while the 5 stars of Open Data Engagement refers to a social perspective. Both these “star systems” have been adopted widely within various applications and prototypes. However, there is a lack of some clear guidelines for *Open Data Portals*, which are responsible for the stimulation of data reuse and data quality improvement (Colpaert et al. 2013). Moreover, it concerns those platforms providing access to Geospatial Linked Open Data. In order to fill this gap, in 2013 (Colpaert et al. 2013), a Five Star evaluation model was proposed to guide Public Administrations in setting up an Open Data Portal. This *Five Star system for Open Data Portals* is complementary and densely related to the star system of Open Data Engagement and Linked Open Data and, in the end, aims for data to become a common resource (Colpaert et al. 2013).

- ★ A dataset registry: *a list of links to Open Datasets*
- ★★ A meta-data provider: *maintain, structure and open up your meta-data*
- ★★★ A co-creation platform: *gather tools and stimulate conversations about re-use*
- ★★★★ A data publishing platform: *provide the data itself in common formats*
- ★★★★★ A common data hub: *open governance, provenance, trust and versioning*

All the discussed models imply primary concepts for Linked Open Data and are applicable for geospatial data.

## 2.2 Location Business Intelligence

“*Business intelligence (BI)* is a broad category of applications, technologies, and processes for gathering, storing, accessing, and analyzing data to help business users make better decisions” (Watson 2009). The concept of BI has a pretty dynamic history and involves various ideas, people, systems, and technologies (Power 2007). Its roots began in the 1960s with the first computer applications that later evolved into the first decision support applications. One of the first and most influential works in this field belongs to Michael S. Scott Morton’s. In 1967 in his doctoral dissertation research, the author developed, implemented, and tested a system to support laundry equipment planning. Later, he published a prominent Sloan Management Review article (Gorry and Scott Morton 1971) and a book (1971) that boosted the worldwide spread of decision-support concepts and applications and named them— *management decision systems*.

Over time, with the appearance of new problems, needs, requests, and technologies, the concept's name was transformed to *Decision Support Systems (DSS)*. The primary idea of DSS was advanced and expanded by new directions with their unique characteristics, in particular, executive information systems (EIS), group decision support systems (GDSS), and *geographic information systems (GIS)* (Watson 2009). Transformations within the DSS domain, together with technological advances, led to the development of the *Business Intelligence* term. Gradually, *model-centric* DSS were getting less interest. By contrast, *data-centric* DSS became more valued in meeting business needs and forming the BI industry.

The more use-cases and applications appear, the more diverse data enter the Business Intelligence world. Most businesses are running side-by-side with a world of ready-to-use data that could significantly improve their enterprises' efficiency, effectiveness, and profitability (Bowes 2007). In 2012 Zeljko Panian (Panian 2012) emphasized that all the companies operate, interact with their customers, and transact business within a *geospatial environment* rich in powerful data. However, only a minority of them is aware and takes advantage of location data. Unfortunately, the phenomenon indicated by Panian back in 2012 still exists. *Geospatial data* has huge potential to be a crucial addition to business intelligence applications across various industries. "For instance, absolute data like a city's population can serve as a rough gauge of a market's potential, but only relative data – a matching of store locations to local demographics, or the comparison of a company's own wireless network to those of its competitors – can yield true Location-based Intelligence (LBI)" (Panian 2012) that will offer value to business.

According to the latest report by GeoBuiz (GeoBuiz 2019), *Location Intelligence* solutions are one of the relatively new components of the geospatial market and business in general. *Location Business Intelligence (BI)* involves a broad category of applications, frameworks, and technologies of combining geospatial data with business data and business processes to create user- and industry-specific solutions and help make better decisions (GeoBuiz 2019). It has a big variety of applications ranging from rendering basic features such as map-based visualization to enabling complex analytical algorithms for particular tasking. The value chain of Location Intelligence engages some key processes. In particular, developing software or platforms, supplemented with geocoding engines with analytics functionality, added with map rendering capabilities and various algorithms. This is followed by capturing location data, which involves obtaining the map data (base maps, satellite imageries, map layers, etc.) and geographic coordinates corresponding to Points-of-Interest (PoI). These are integrated with business data and processes to deliver insights, identify trends and relations, and uncover new business opportunities (GeoBuiz 2019).

GeoBuiz (GeoBuiz 2019) emphasizes that *geospatial* becomes the "by-default" technology across industries and simply in daily life. GIS software technology became a valuable extension for BI tools providing a framework to manage, visualize, analyze, and ultimately understand the significance of location information.

Various investigations and potential solutions were suggested aiming to preserve data from the LOD world to the BI environment and GIS frameworks. However, not so much research was done particularly in the direction of straightforward integration of Geospatial LOD into BI solutions. Even though there are some notable works worth mentioning.

In 2019, for instance, Folmer et al. (Folmer et al. 2019), apart from discussing the value of cadastre data being accessible as Linked Open Data, emphasized the benefits of Linked Data integration into Business Intelligence. Moreover, this paper demonstrates the successful effort to visualize Linked Data in BI software, Tableau.

The same year, Mai et al. (Mai et al. 2019) delivered prominent and comprehensive work, where the authors attempted to implement a Linked Data connector framework prototypically. The suggested set of toolboxes aims at deriving geospatial data out of RDF triples directly to Esri's ArcGIS analytical environment.

## 3. Methodology

This chapter explains the methods used to investigate aimed results. It includes a review of the developed thematic scenarios. Furthermore, it introduces the selected Linked Data providers and Business Intelligence (BI) tools together with an explanation of the selection approach.

### 3.1 Thematic scenarios development

By its nature, 'scenario' can contain different meanings, serve a different purpose, and deliver different insights across fields and practices (Ramirez et al. 2015). In this paper, 'scenario' stands for a research technique that helps to understand, illustrate and answer (complex) situations, cases, or problems for a particular purpose, typically to provide inputs for further work and decision-making. *Complex* means that the regarded problem is dependent on multiple criteria and is influenced by many factors that are related and/or interlinked and, thus, deliver several possible answers to this problem rather than a single answer. (Ulrich and Probst 1991; Ulrich 1992).

Two primary thematic scenarios are developed within this paper. They will demonstrate the intended approach based on simple thematic examples with a particular topic, data requirements, and workflow. The output of the scenarios' implementation is expected to be visualization products such as maps, web-maps, charts, histograms, and others.

#### 3.1.1 Scenario 1

##### *Potential impact of COVID-2019 crisis on EU labor markets*

This scenario aims to discover the potential impact of the early 2020 COVID confinement measures on EU labor markets.

According to the latest reports by International Labour Organisation (ILO) and European Commission, COVID confinement measures have a strong impact in nearly all European labor markets (International Labour Organisation 2020; COVID & Empl Working Group). Based on an analysis of the restrictions on economic activity imposed in the EU Member States, the European Commission suggests an estimated level of the crisis impact on each economic sector. This level of influence is represented as an index for each economic sector depicted within the report in Table 2. "*A summary of the COVID sector lockdowns in three European countries as of early April 2020*" (COVID & Empl Working Group).

Eurostat database offers the data on the number of employees by different economic sectors (NACE Rev.2) in EU countries by region (NUTS-3; European Commission; European Commission). Based on this data and indexes developed by the European Commission, this is possible to roughly calculate the potential decrease of employees number within economic sectors in regions. Thus, the hardest-hit regions in EU countries in lockdown conditions will be identified.

For the implementation of this scenario, the following data is used:

- Business demography and high growth enterprise by NACE rev.2 and NUTS 3 regions
- Population in EU countries by regions
- Area data for NUTS-3 regions
- NUTS-3 regions geometry for EU countries



Used data sources: [European Data Portal](#), [Eurostat](#).

### 3.1.2 Scenario 2

*Closeness of bike-sharing services to the areas of Alpine ibex colonies concentration within Switzerland*

Ecotourism is one of the most popular vacation activities, and Switzerland suggests perfect conditions for this. This scenario aims to identify bike-sharing spots closest to the concentration areas of Alpine ibex in Switzerland. This can help tourists to know and choose bike-sharing services and ibex colonies according to their location.

For the implementation of this scenario, the following data is used:

- Bike-sharing and bicycle hire locations in Switzerland
- Distribution of ibex colonies in Switzerland

Used data sources: [European Data Portal](#), [Eurostat](#).

## 3.2 Overview of leaders of the global Location Business Intelligence market

This section introduces the Business Intelligence software that is used within this thesis project. Each BI tool was selected according to global market reports, works produced by comprehensive research and advisory companies, scientific papers, and companies' documentation. Each BI tool is supported with a basic description, together with the main references. All the selected BI tools are relevant for location intelligence applications since each of them offers the functionalities on geospatial data processing and analysis.

### [Tableau](#)

According to the latest market research on [Analytics and Business Intelligence Platforms](#) reported by [Gartner Magic Quadrant](#) Tableau is one of the BI market leaders as for January 2020 (Gartner Analytics Magic Quadrant; Magic Quadrant Research Methodology ).

Tableau enables users to access data rapidly from a broad range of data sources. Moreover, data can be processed, analyzed, manipulated, and visualized using best practices in visual perception. Tableau accommodates those necessary functionalities that help to deliver insights and identify patterns in users' data (Magic Quadrant for Analytics and Business Intelligence Platforms).

With "Tableau Maps", users can immerse in the Location Intelligence environment and get the most out of their geospatial data. With a geocoding toolbox, Tableau automatically turns the location data into meaningful, informative, and interactive maps that can serve as direct solutions across problems and use-cases (Tableau Maps).

### [CARTO](#)

CARTO is one of the world's leading Location Intelligence platforms and key market players, according to (GeoBuiz 2019). CARTO is a Software as a Service (SaaS) cloud computing platform that provides GIS, web mapping, and spatial data science capabilities. The CARTO company is positioned as Location Intelligence platform, offering geospatial data analysis and visualization frameworks with no GIS or developer pre-requirements (CARTO).

Individuals, scientists, governments, and enterprises solve location-related problems using geospatial data within CARTO to understand where and why things happen, optimize

internal and external processes and performance, and predict possible outcomes and results (CARTO).

#### [OmniSci](#)

OmniSci (former MapD) is an open platform that blends analytics, data science, and location intelligence workflows (OmniSci). [Gartner](#) recognizes OmniSci as a platform that provides powerful database management and comprehensive visual analytics software optimized for graphics processing unit (GPU) accelerators. OmniSci offers interactive querying over billions of records, followed by instant results and built-in data visualization functionality that can leverage the rendering capabilities of the server-side GPUs (Feinberg et al.).

The platform is used among various user audiences, particularly, among big enterprises having huge streams of operational data and demanding speedy data-driven decisions (OmniSci).

OmniSci removes technological limitations and ensures high-speed results and instant decisions based on a comprehensive combination of location intelligence with big data (OmniSci).

#### [ArcGIS Insights](#)

ArcGIS Insights is a powerful component of the Esri Geospatial Cloud, ArcGIS Insights that offers location data analytics within business intelligence workflows (Esri).

According to the latest Geobuiz report ('Geobuiz Report 2019'), ArcGIS Insights is one of the world's largest players within the Location BI industry.

Users can access ArcGIS Insights as a part of ArcGIS Enterprise and ArcGIS Online. This analytical benchmark aims to maximize individuals' and companies' productivity by offering simpler data analytical experience added with interactive and explanatory visualization frameworks like maps, charts, and tables (Esri). By blending business intelligence capabilities with the best in location analytics, ArcGIS Insights users can quickly and easily explore both spatial and nonspatial data from one application ('Location Analytics- The Key to More Powerful Analysis ebook').

### 3.3 Overview of Linked Open Data providers

This section introduces Linked Open Data providers that will be used within this thesis project. Each of the selected Linked Open Data providers offers access to geospatial data. Both are recognized as ones of the most used data sources within the European Union.

#### [Eurostat](#)

Covering all European society areas with more than three hundred million statistical data, *Eurostat (European Statistical Office)* is a remarkably rich data source (What we do - Eurostat). Eurostat's primary purpose is to process and publish statistical data and information at the European level (What we do - Eurostat).

Eurostat does not collect data by itself. This is a task for the Member States and their local statistical authorities. They collect, validate, and analyze national data and forward them to Eurostat for further processing. In turn, Eurostat ensures the quality of data using a harmonized methodology, integrates and makes them available for potential users. Since Eurostat's primary mission is a provision of harmonized statistical data, it is considered to be *the only provider of statistics at the European level* (What we do - Eurostat). Furthermore, Eurostat offers access to geospatial data that can be queried as Linked Open Data.

Google Scholar query results can confirm a high use frequency and significance of Eurostat in Linked Data research - more than 11 500 references since 2019 and referenced about 79 000 times in total (Google Scholar).

#### [European Data Portal \(EDP\)](#)

The European Data Portal (EDP) is a central point of access to public data initiated by the Publications Office of the European Union and by the European Commission that aims to increase the impact of open data by making it easy-to-find and re-used by everyone (European Data Portal). It offers over one million datasets across all areas of European Member States.

The European Data Portal collects the metadata (data about the data) on data and information managed and maintained by EU institutions, public agencies, and bodies across European countries at national, regional, local, or domain-specific levels. They cover the EU Member States, EFTA (European Free Trade Association) countries, and countries involved in the EU's neighborhood policy (European Data Portal).

The EDP acquires statistical, geospatial and non-geospatial data from more than 70 national data providers. Aiming to increase interoperability and accessibility of Open Data, the European Data Portal adopted Linked Data technology, including DCAT-AP specification DCAT-AP (Kirstein et al. 2019).

According to Google Scholar search results, EDP appears to be one of the most used data sources referenced concerning to Linked Data technology research - 19 000 references since 2019 and about 144 000 references in total (Google Scholar).

## 4. Implementation

### 4.1 Data Access Interfaces in Linked Data providers

#### 4.1.1 Eurostat

Eurostat delivers detailed, well-structured, and up-to-date statistical data as well as geospatial data for the entire European Union, and additionally, for some major non-European countries (Halb et al.). Eurostat data covers a range of thematic directions/topics of interest for governments, enterprises, education sector, journalists and individuals to fulfill various tasks and purposes (Eurostat).

The *data section* is the main access point to Eurostat's *database*, where all the datasets are stored and structured by theme. Since within Eurostat datasets are updated frequently, twice a day (at 11:00 and 23:00), Eurostat's database always contains the latest version of the datasets (Eurostat). Users can access and request datasets from Eurostat in several ways:

**Database.** The majority of data available within Eurostat can be accessed from the *data navigation tree*, which is structured according to statistical themes. Here datasets can be extracted directly from the database via the *Data Explorer interface*. Within this interface, data can be downloaded in various formats, such as XLS, CSV, HTML, PC AXIS, SPSS, TSV, and PDF (Statistics Explained - Eurostat).

**Bulk Download.** This facility allows downloading individual datasets or the complete database in two formats TSV and SDMX-ML, compressed in ZIP files (Eurostat BulkDownload Guidelines 2019).

**Web Services.** Eurostat's datasets can be accessed through:

- [SDMX Web Services](#), as well as
- [JSON and Unicode Web Services](#).

Both types of services are programmatic access to Eurostat data, with the possibility of customizing requests for data according to the user's requirements.

**Homepage.** Here the most recent and popular datasets can be quickly accessed.

**Statistical Atlas.** This tool is a map viewer where users can explore interactive maps for various topics at the NUTS level.

**SPARQL endpoint.** Furthermore, Eurostat provides its data as *Linked Open Data* via the SPARQL endpoint at the [EU Open Data Portal](#). SPARQL queries allow searching for the metadata stored in the EU Open Data Portal triple store and access data as RDFs.

Most of the statistical data collected and provided by Eurostat are related to a *location* at the country, region, or administrative unit level. Within Eurostat, [GISCO](#) manages a database of geographical data and information at three primary levels: the European Union, EU member states, and their regions. Users can access core geospatial data covering the whole of Europe, such as administrative boundaries and statistical units, ports and airports, Digital Elevation Model (DEM), population distribution, Land Cover/Land Use information. In addition, GISCO allows creating thematic maps based on the fusion of statistical data with geospatial data. Users can find most of the Eurostat's data is open for the general public and non-commercial use. However, some data is supposed to be used only for Commission purposes or under specific permissions and licenses (GISCO Eurostat 2020; Geodata Eurostat 2020).

Within Eurostat's GISCO, *geospatial data* is available in such formats as Esri Shapefile, TopoJSON, GeoJSON, GDB (Geodatabase), and SVG.

Additionally, users can obtain desired datasets containing data in different projections via the [GISCO data distribution REST API](#) (NUTS GISCO 2020). For instance, such geographical data as [NUTS](#) (Nomenclature of territorial units for statistics classification) can be accessed as Linked Open Data via SPARQL access endpoint, multiple geo-formats, GISCO data distribution API and RDF.

Table 1 below summarizes data standards according to The Five Star Open Data model (Berners-Lee 2006; Berners-Lee 5) applied to spatial data formats (McKeague et al. 2020) available within Eurostat.

Stars	Spatial Open Data formats available
★	PDF
★★	XLSX, Esri Shapefile, GDB, SPSS*
★★★	CSV, GeoJSON, TopoJSON, TSV, CSV*
★★★★	RDF
★★★★★	Yes

*Table 1. Spatial data formats available in Eurostat*

#### 4.1.2 European Data Portal

The European Data Portal is a notable component of European public data infrastructure, which enabled data retrieval under Linked Data principles. The technology behind is based on the Resource Description Framework (RDF) and Data Catalogue Vocabulary (DCAT). The EDP framework is designed to increase interoperability and data reuse and allow users to search and harvest various Open Data across Europe's portals and agencies (Kirstein et al. 2019).

As of June 2020, the EDP lists more than 1 million datasets harvested from 85 data providers covering 14 thematic categories (European Data Portal).

The *datasets section* appears to be the main access point for browsing, filtering, and searching the datasets (European Data Portal 2019). Here data can be accessed and acquired from the various source portals in several ways:

**CKAN-API.** API defines the core interaction functionalities available within the portal, in particular those related to the user web interface and data harvesting operation. Therefore, the harvested data and information can be used by external code that calls the portal API (European Data Portal).

**SPARQL.** The SPARQL Manager provides a graphical user interface (GUI) that allows users to define queries and send them to the Virtuoso SPARQL query engine. This SPARQL technology involves SPARQL Protocol and RDF Query Language as primary components responsible for querying metadata as Linked Data (European Data Portal 2019). Here three main modes are available: SPARQL - Search, SPARQL - Assistance, SPARQL - Queries. The EDP's SPARQL allows users to retrieve every dataset in different RDF serialization formats.

**Datasets Feed** is available as RSS Feed as well as ATOM Feed - small text files that provide information about the content on the EDP datasets webpage.

**File Download.** Datasets distributions are not stored directly within EDP, but they are only linked to the original providers (European Data Portal 2019). Thereby, the “Download” function directs the user to the initial data provider, where a dataset can be accessed and retrieved.

Some dataset distributions contain geospatial component and can be accessed not only in a tabular format, but also expressed via visualization frameworks. **“Open Geo-Visualisation”** option for open geospatial data offers previewing functionality. Also, it allows users to check the selected dataset meets particular needs in terms of geospatial and thematic parameters and coverage. Among supported geospatial formats are OGC Web Map Service (WMS) and GeoJSON (European Data Portal 2019).

Datasets can contain distributions in different formats. Thus, the same dataset can be downloaded in several formats according to the user’s requirements. Geospatial data is available within EDP and it is stored in particular geospatial formats. According to the EDP’s datasets section statistics, the following *geospatial formats* are supported (Table 2):

Format	Datasets amount
WFS	96656
WMS	92218
CSV*	90670
ZIP*	23210
JSON*	21521
KML	7979
Esri Shape	3550
GeoJSON	1988
Service*	1300
GML	1118
ArcGIS Map Service	771
ArcGIS Map Preview	637

*Table 2. Geospatial data formats supported by the European Data Portal (EDP)*

\* Although CSV, ZIP, JSON and Service are not geospatial formats by their nature, very often they are used for storage of location data. Hence, they appear in the table above.

Table 3 below is a summary of data standards according to The Five Star Open Data model (Berners-Lee 2006; Berners-Lee 5) applied to spatial data formats (McKeague et al. 2020) available within EDP.

Stars	Spatial Open Data formats available
★	PDF
★★	XLSX, Esri Shapefile, GDB
★★★	CSV, GeoJSON, TSV
★★★★	RDF
★★★★★	Yes

*Table 3. Spatial data formats available in EDP*

#### 4.1.3 Thematic data extraction

This section offers Table 4 and Table 5 that introduce detailed information on the datasets obtained from the selected Linked Data providers, integrated into BI tools and used for further implementation of Scenario 1 and Scenario 2. All data were derived from Eurostat and European Data Portal according to the technical capabilities of each platform.

##### Scenario 1

Used Data	Description	Available interfaces to access data	Available formats
<a href="#">Business demography and high growth enterprise by NACE rev.2 and NUTS 3 regions</a> (bd_hgnace2_r3) 2017	The annual Business demography data collection covers variables that explain the characteristics and demography of the business population ( <a href="#">Eurostat 2008</a> ).	File download via <a href="#">Database</a> , <a href="#">Bulk Download</a> , <a href="#">SPARQL endpoint</a>	XLS, CSV, HTML, PC AXIS, SPSS, TSV, PDF, RDF
<a href="#">Population on 1 January by age group, sex and NUTS 3 region</a> (demo_r_pjangrp3) 2017	"Each year Eurostat collects demographic data at the regional level from 37 countries as part of the Unified Demography (Unidemo) project. UNIDEMO is Eurostat's main annual demographic data collection and aims to gather information on demography and migration" ( <a href="#">Eurostat 2018</a> ).	File download via <a href="#">Database</a> , <a href="#">Bulk Download</a> , <a href="#">SPARQL endpoint</a>	XLS, CSV, HTML, PC AXIS, SPSS, TSV, PDF, RDF

<a href="#">Area by NUTS 3 region</a> (reg_area3) 2016	“Total Surface Area (TSA) – Total Surface Area is defined as the area of any given statistical area and includes land area and inland waters (lakes, rivers, etc.)” (Eurostat 2019). Statistical Regions form the Nomenclature of territorial units for statistics (Eurostat 2019).	File download via <a href="#">Database</a> , <a href="#">Bulk Download</a> , <a href="#">SPARQL endpoint</a>	XLS, CSV, HTML, PC AXIS, SPSS, TSV, PDF, RDF
<a href="#">NUTS-3 regions geometry for EU countries</a> 2020	This dataset contains location data of regions within EU countries for level 3 of the Nomenclature of Territorial Units for Statistics (NUTS).	<a href="#">GISCO Rest API</a> , <a href="#">Statistical Atlas</a> , <a href="#">SPARQL endpoint</a>	SHP, JSON, TopoJSON, GeoJSON, GDB, SVC, RDF

Table 4. Thematic data extracted from Eurostat

### Scenario 2

<a href="#">Bike-sharing and bicycle hire</a> 2017	This dataset offers data on bike-sharing and bicycle hire stations across Switzerland belonging to NextBike, PubliBike, Rent a Bike, Schweiz rollt and Velospot.	File download, <a href="#">SPARQL Search</a>	SHP, WMS, RDF, XML, JSON
<a href="#">Distribution of ibex colonies</a> 2015	This dataset contains data on the distribution of Alpine ibex colonies across Switzerland.	File download, <a href="#">SPARQL Search</a>	SHP, WMS, WMTS, RDF, XML, JSON

Table 5. Thematic data extracted from EDP

## 4.2 Geospatial Linked Open Data integration via visualization frameworks in Business Intelligence tools

This section offers a detailed description of explored semantic data standards for spatial data, data transfer, and integration technologies supported by each of the selected BI tools. This research is based on technical documentation and capabilities. Moreover, it involves exploration through the initial visualization framework functionality in BI tools.

Within this paper, the Spatial Linked Data integration process is not limited only to importing data to each BI tool. The integration process is considered complete only when the imported data can be successfully manipulated and operated within the BI tool's functionality. Thus, the integration goes through a visualization framework in the selected BI tools. This allowed identifying integration inconveniences that cannot be revealed at the early data import stage.



### 4.2.1 Tableau

Tableau is an interactive data visualization software and platform that supports solutions in a wide range of domains. In particular, Tableau offers powerful functionality for manipulations on location data.

In order to process, analyze, and visualize data in Tableau, users have to connect Tableau to their data, and there is several possible ways for doing that, in particular:

**Connect to a file.** Data stored on one's computer in different data formats can be connected to Tableau.

**Connect to a server.** This is also possible to connect data "stored in a big data, relational, or cube (multidimensional) database on a server" (Tableau Maps) in a user's enterprise. Tableau interface offers a big list of sources available for connection, such as PostgreSQL, Spark SQL, Oracle, etc. Besides, users are granted an opportunity to create a custom connection using **Other Databases (JDBC)**, **Other Databases (ODBC)**, a **Web Data Connector**, or a **Connector Plugin** built using the Tableau Connector SDK (Tableau Maps).

**Connect to published data sources.** Additionally, one can connect to open data distributed on the Web by public agencies and authorities, or to cloud databases, like Google Analytics, Amazon Redshift, or Salesforce.

In *Tableau Desktop*, users can connect to the following geospatial file formats:

- CSV (Comma-separated values) files (.csv)
- Esri Shapefiles (.zip)
- Excel files (.xlsx)
- JSON (.json)
- Geodatabases files (.gdb)
- GeoJSON files
- TopoJSON files
- KML (Keyhole Markup Language) files (.kml)
- MapInfo tables
- Portable Document Format (PDF) files (.pdf)
- Text files (.txt)

The "Creator license" in *Tableau Online* or *Tableau Server* allows users to upload geospatial file formats as a single file (KML, GeoJSON, TopoJSON, Esri shapefiles packaged in a ".zip", and Esri File Geodatabases with the extension ".gdb", ".zip") (Tableau Maps).

Table 6 below is a summary of data standards according to The Five Star Open Data model applied to spatial data formats available within Tableau.

Stars	Spatial Open Data formats available
★	PDF
★★	XLSX, Esri Shapefile, MapInfo, Geodatabase
★★★	CSV, TXT, GeoJSON, TopoJSON, KML, KMZ
★★★★	SPARQL



Table 6. Spatial data formats available in Tableau

The Linked Data integration implementation via thematic scenarios is expressed through the main processes described in Table 7.

Process Name		Status	Comments
Import		Success	Datasets obtained from Eurostat and EDP were imported to Tableau without inconveniences.
Join		Success	Join operation between spatial tables (Esri Shapefile, GeoJSON) and non-spatial tables (CSV) was implemented successfully via Data Source Connections interface.
Processing	Scenario 1	Success	Simple calculation operations on data worked out without inconveniences using Tableau Analysis.
	Scenario 2	Fail	Calculation of distance was not implemented. Current Tableau Analysis functionality does not offer Nearest Neighbour analysis, neither creation of a buffer for polygon features. A distance calculation, as well as creation of a buffer, is applicable only for point features.
Visualization		Success	Visualization framework resulted in created maps and histograms which demonstrate Scenario 1 and Scenario 2 (only data visualization, no analysis).

Table 7. Linked Data integration via thematic scenario implementation in Tableau

#### 4.2.2 CARTO

CARTO fits the needs of organizations looking to design and deliver effective web maps without investing in significant GIS software infrastructure. CARTO is a self-contained mapping platform that enables users to upload, manage, and style content to then deliver through customizable interfaces (CARTO | NGIS).

Data can be uploaded to CARTO in various ways:

[CARTO Dashboard](#). Within Carto Dashboard, one can upload data in two different ways:

1. *Drag and drop file*. Data can be imported by the “drag and drop” function at the CARTO Datasets dashboard.
2. *Import data from the “Add datasets” menu*

This menu allows importing local files, URLs (might be used for creation of Sync tables) or import data from other sources such as Google Drive and ArcGIS Server.

[CARTO Import API](#).

Data can also be imported programmatically. CARTO Import API gives the user more flexibility and more options to define how the desired data should be uploaded.

### Database Connectors

Besides uploading files to CARTO, one can also connect the CARTO account to the database. That allows importing database tables to the CARTO account.

### CARTO SQL API: COPY command

The CARTO Import API is handy, but it has the file size and rows limits for imported files. Splitting files into smaller ones might be inconvenient for users, so alternatively, CARTO SQL API's COPY command allows importing big CSV files without their splitting much faster than the default import process.

### CARTO Libraries

Users can also take advantage of [CARTO Python SDK](#) or [CARTOfames](#) to upload data to the CARTO account. These two libraries deliver methods that are based on the Import API, the SQL API, and other CARTO APIs behind-the-scenes and help to define the desired workflow for uploading data to the CARTO account.

CARTO supports a large number of [data types and file formats](#), including:

- Esri Shapefile (.zip or .gz)
- CSV (Comma-separated values) files (.csv)
- Geopackage files (.gpkg)
- Geodatabase files (.gdb)
- GeoJSON (.geojson)
- GPX (GPS Exchange Format) files (.gpx)
- KML (Keyhole Markup Language) files (.kml)
- KMZ (.kmz)
- MapInfo files (.DAT, .ID, .MAP, .TAB) as a single compressed file in (.zip) or (.gz)
- OSM (OpenStreetMap) files (.osm)
- Spreadsheets (.xlsx or .odt)
- WMS (Web Map Service)
- WMTS (Web Map Tile Service)

Table 8 below summarizes of data standards according to The Five Star Open Data model applied to spatial data formats available within CARTO.

Stars	Spatial Open Data formats available
★	
★★	XLSX, Esri Shapefile, MapInfo, Geodatabase
★★★	CSV, GeoJSON, KML, KMZ, OSM, Geopackage, GPX, WMS, WMTS
★★★★	
★★★★★	

Table 8. Spatial data formats available in CARTO

The implementation of thematic scenarios is illustrated in Table 9.

Process Name		Status	Comments
Import		Success	Datasets obtained from Eurostat and EDP were imported to CARTO without inconveniences.
Join		Success	Join operation between spatial tables (Esri Shapefile, GeoJSON) and non-spatial tables (CSV) was implemented successfully via the CARTO SQL interface.
Processing	Scenario 1	Success	Simple calculation operations on data worked out without inconveniences using CARTO SQL.
	Scenario 2	Success	Calculation of distance was made using the CARTO SQL interface without inconveniences.
Visualization		Success	Visualization framework resulted in created maps and histogram which demonstrate Scenario 1 and Scenario 2.

Table 9. Linked Data integration via thematic scenario implementation in CARTO

#### 4.2.3 OmniSci

In this research, OmniSci Immerse was used. “OmniSci Immerse is a browser-based, interactive data visualization” (OmniSci Immerse) tool that offers powerful and speedy functionality on huge data volumes. Immerse interacts continuously with the OmniSci server-side technologies, OmniSciDB, and Render. OmniSci Immerse allows SQL querying to the OmniSci backend on the fly, with no coding pre-requirements providing a user-friendly experience even for users with no developer’s background. For advanced users, Immerse also offers custom SQL queries allowing to generate even more sophisticated visual analytics (OmniSci Immerse).

OmniSci supports such sources for data upload as:

**Kafka.** Apache Kafka is a distributed streaming platform. It allows creating publishers, which create data streams, and consumers, which subscribe to and ingest the data streams produced by publishers (Kafka). Users may operate OmniSciDB [KafkaImporter](#) C++ as a programmatic way of uploading data.

**OmniSci Immerse Data Manager.** This is possible to upload data from a *local file*, a delimited file from [Amazon S3](#), or import data from *Catalogue*.

OmniSci Immerse supports file upload for:

- Esri Shapefile (.shp) together with its types and extensions
- CSV (Comma-separated values) files(.csv)
- GeoJSON (.geojson)
- JSON (.json)

- KML (Keyhole Markup Language) files (.kml)
- KMZ (.kmz)
- Tab-separated values (.tsv)
- Text files (.txt)

Immerse also supports the upload of *compressed delimited files* in TAR, ZIP, 7-ZIP, RAR, GZIP, BZIP2, or TGZ format.

**Loading Data with SQL.** Geodata can be loaded to an OmniSci distributed cluster using a “COPY FROM” statement within the SQL interface.

Table 10 below is a summary of data standards according to The Five Star Open Data model applied to spatial data formats available within OmniSci.

Stars	Spatial Open Data formats available
★	
★★	Esri Shapefile
★★★	CSV, TSV, TXT, GeoJSON, JSON, KML, KMZ
★★★★★	
★★★★★	

*Table 10. Spatial data formats available in OmniSci*

The implementation of thematic scenarios is illustrated in Table 11.

Process Name		Status	Comments
Import		<b>Success</b>	Datasets obtained from Eurostat and EDP were imported to OmniSci without inconveniences.
Join		<b>Success</b>	Join operation between spatial tables (Esri Shapefile, GeoJSON) and non-spatial tables (CSV) was implemented successfully via the OmniSci SQL interface.
Processing	Scenario 1	<b>Success</b>	Simple calculation operations on data worked out without inconveniences using OmniSci SQL.
	Scenario 2	<b>Success</b>	Calculation of distance was made using the OmniSci SQL interface without inconveniences.
Visualization		<b>Success</b>	Visualization framework resulted in created maps, histograms, and charts which demonstrate Scenario 1 and Scenario 2.

Table 11. Linked Data integration via thematic scenario implementation in OmniSci

#### 4.2.4 ArcGIS Insights

ArcGIS Insights is a data analytics software that provides spatial and non-spatial analysis capabilities powered by business intelligence workflows. Insights can be used in three deployment options:

- ArcGIS Insights in ArcGIS Online
- ArcGIS Insights in ArcGIS Enterprise
- ArcGIS Insights Desktop

When a user enters the Insights environment, they might create feature layers first via:

[Importing a dataset](#) in Insights. Users can import their data from the following sources in Insights:

- Hosted or registered feature layers created by other users, groups, or organizations within ArcGIS Pro or ArcMap can be copied to other users' the ArcGIS organization.
- Esri Shapefiles (.zip)
- Excel files (.xlsx)
- Comma-separated value files (.csv)
- Feature layers from ArcGIS Living Atlas of the World
- GeoJSON files (.json, .geojson, or GeoJSON files in a .zip file)
- Map image layers
- Connections to Microsoft SQL Server, SAP HANA, PostgreSQL, and Oracle databases (available in Insights in ArcGIS Enterprise and Insights desktop). The following data types are supported:
  - Spatial tables
  - Non-spatial tables
  - Relational database views
- Connections to enterprise geodatabases (Supported data—Insights Get Started).

Table 12 below summarizes of data standards according to The Five Star Open Data model applied to spatial data formats available within ArcGIS Insights.

Stars	Spatial Open Data formats available
★	
★★	XLS, Esri Shapefile, Map image layers
★★★	CSV, GeoJSON
★★★★	
★★★★★	

Table 12. Spatial data formats available in ArcGIS Insights

The implementation of thematic scenarios is illustrated in Table 13.

Process Name		Status	Comments
Import		<b>Success</b>	Datasets obtained from Eurostat and EDP were imported to ArcGIS Insights without inconveniences.
Join		<b>Fail</b>	Function "Create Relationships" between data tables worked properly for joins between CSV tables. Hence, it did not work for joins on spatial tables (Esri Shapefile, GeoJSON) with non-spatial tables (CSV).  Possible reason - Insights version. Even within the online deployment ArcGIS Insights is available in three versions: 3.3, 3.4 and 2020.1. Within this research, only one version was used.
Processing	Scenario 1	<b>Success</b>	"Calculate" function for operations on data tables worked without inconveniences.
	Scenario 2	<b>Success</b>	"Spatial analysis" operations, in particular, "Find Nearest" worked without inconveniences.
Visualization		<b>Success</b>	Visualization framework resulted in created maps and histogram which demonstrate Scenario 1 and Scenario 2.

*Table 13. Linked Data integration via thematic scenario implementation in ArcGIS Insights*

## 5. Results and discussion

The presented work focuses on identifying conceptual challenges arising within the idea of integration of *geospatial Linked Open Data* with *Business Intelligence tools*. Based on the motivation and problem statement, the research goals and objectives were intended. Considering them, the research questions were proposed for this thesis, which are stated here again:

- What data transfer technologies currently define the leading BI tools integrations' support?
- What specific data transfer technologies have to be in focus for the Linked Data platforms to provide successful integration use-cases?
- What are the key components of the Linked Data concept required to be integrated into BI tools to provide access to the global Linked Data and Knowledge Network platforms?

In order to reach the goals of this thesis and answer the research questions, a review of the related works and findings was carried out first. It allowed getting familiar with the state-of-the-art and realizing the value of this thesis research among existing contributions.

The literature review showed that the majority of the related research heavily relies on the development of prototypes of adapters and connector frameworks between Linked Open Data providers and Business Intelligence tools. They are supposed to serve as add-ons to BI and GIS software. However, this trend cannot serve as a forever way-out. It is a temporal solution valid only until any changes are applied to standards either on the side of Linked Open Data or Business Intelligence. Moreover, these solutions do not always ensure automated updates for data (in case of changes) integrated into BI software. For this reason, there should be constant consideration and support by the interested parties running such connectors. This kind of initiative often requires financial support, which means a chance for connectors being not available for free. Together, all the mentioned factors emphasize the limitation of the connector frameworks. Furthermore, they demonstrate a need for a deeper review and improvement of existing technologies supported by LOD providers and BI tools.

The development of two primary thematic scenarios was in the next step aiming to demonstrate the intended approach. Both scenarios are based on simple thematic examples with a particular topic, data requirements, and workflow. The thematic data for the scenarios comes from *Eurostat* and *European Data Portal* as Linked Open Data representatives selected for this research. This paper provides an overview of both platforms (Section 3.3) extended with the description of their data access capabilities in detail (Section 4.1.1, Section 4.1.2). The conducted analysis showed three major ways to retrieve geospatial data and use it within BI software. The traditional and still most common approach is to obtain data from a data source and upload it into BI. Alternatively, users can stream data directly into BI software by connecting to an API or OGC Web Services, such as a Web Feature Service. This option guarantees that data remain up-to-date and will be automatically updated in case of any changes on the data source side (Mai et al. 2019). Finally, Eurostat and EDP provide SPARQL endpoints allow querying metadata stored as triples and accessing data as RDFs. What is more about Eurostat, its advantage is that geospatial data can be connected with non-geospatial data if needed. All available datasets always contain a key field of unique values that correspond to NUTS units providing extra flexibility for data retrieval.

In terms of supported geospatial data formats, both providers match the standards determined by (McKeague et al. 2020) and (Berners-Lee 2006; Berners-Lee 5).

There was additional research conducted on European Data Portal. EDP provides very high-level information about the data transfer technologies used. It offers highly-aggregated data



from all the datasets and distributions associated. This generic view is not enough to classify the technical level of advancement of the dataset mentioned in 2.1. It also includes interoperability exploration. In addition to data aggregation, the following issues were noticed:

- Data Formats are not correctly classified (i.e., "ESRI Shape" and "ESRI Shapefile" are different formats)
- Only 649434 distributions out of 1076124 datasets included information about data format. Others were under the "Unknown" tag. Considering that most datasets have more than one distribution, the partition of the "Unknown" data format is significant.

Two specific questions were designed to cover the mentioned points:

- What is the real distribution of geospatial data formats across the European Data Portal?
- How many different geospatial data formats are available for a single dataset?

Based on the EDP API, custom script collected the following data sample:

- 153951 Unique Datasets
- 390757 Unique Distributions
  - 139259 Distributions (35,6%) with no Geospatial formats

The additional script allowed to reduce the partition of the "Unknown" data format to 10,49% across this sample distributions.

The following figure 2 represents the frequency of Geospatial formats:

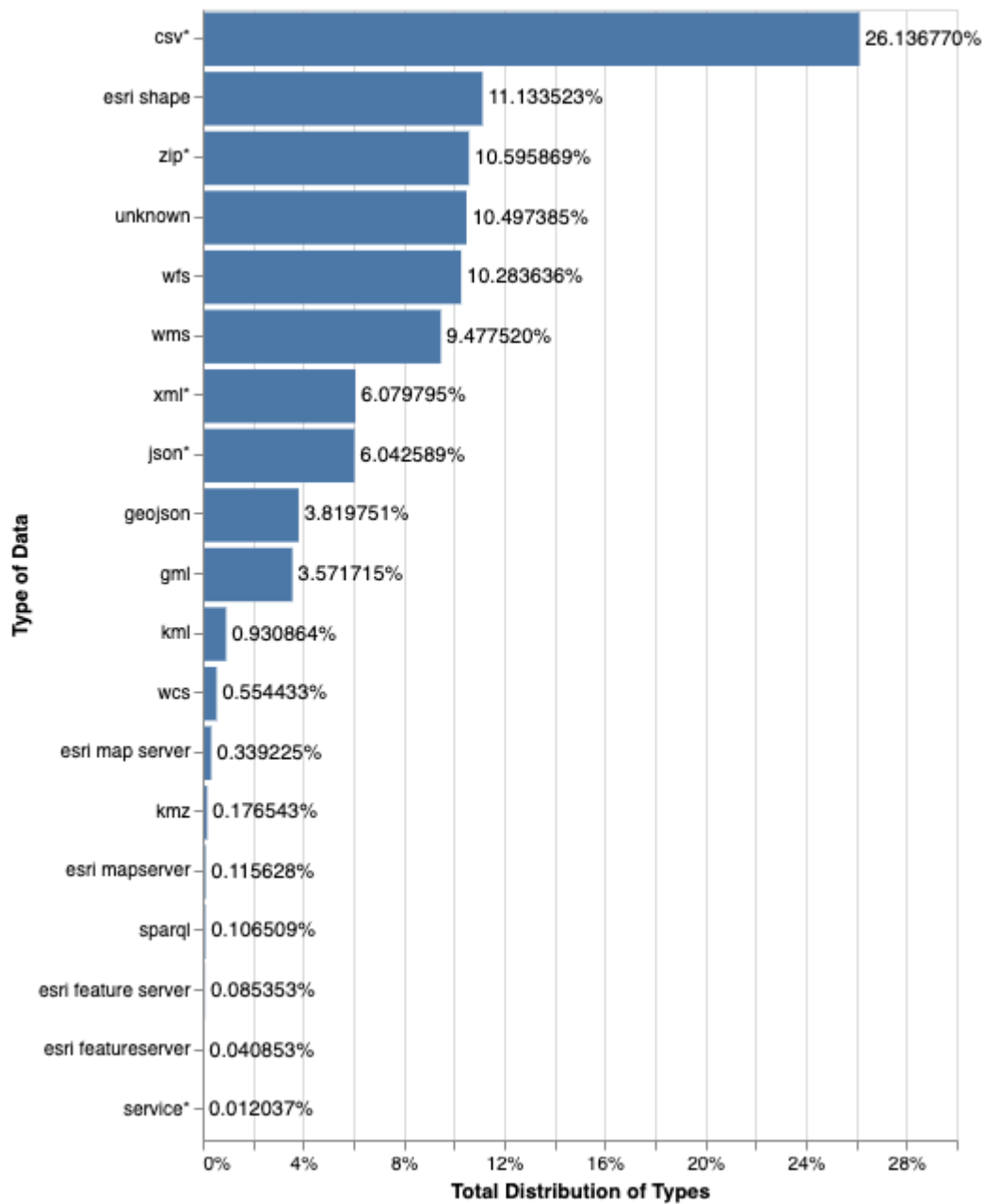


Figure 2. Data Types usage in the Datasets' Distributions

Based on this representable sample, the overall conclusion is the following: *most of data at the European Data Portal distributed on the third level of advancement* (figure 3). The RDF format is available over all the datasets representing the metadata of it.

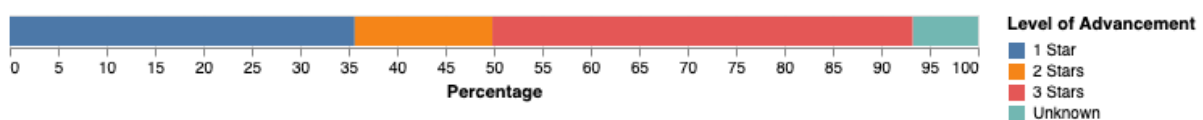


Figure 3. Data Types usage aggregated by the Five Star Open Data model (Berners-Lee 2006; Berners-Lee 5) applied to spatial data formats (McKeague et al. 2020)

Another conclusion worth mentioning is that the existing (Geospatial) Open Data Models must be reviewed, updated, and advanced. Keeping on the same stage file formats like GeoJSON and services like WMS is inefficient since they require a different level of further integration. The

knowledge of geospatial formats, their standards, and the processing chain is of high value for map production.

Pivoting the table with a focus on the number of different data formats produced the following visualization (figure 4):

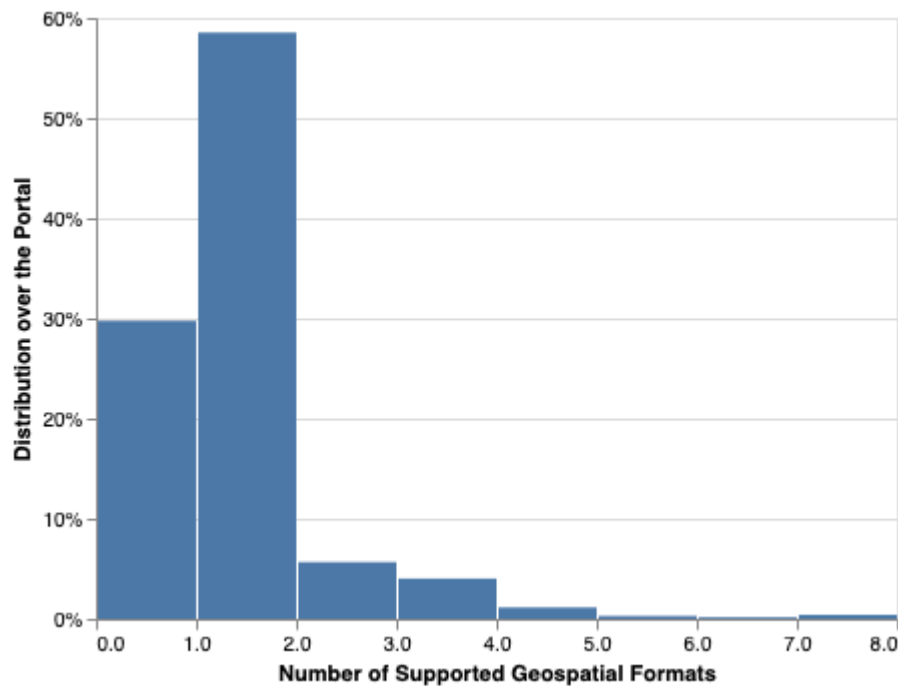


Figure 4. Distribution of the Number of Supported Geospatial Formats over datasets

It supports the understanding of options variety for further data integration. *The significant majority of data sources provide only two or one types of data distribution.* This poor result should be the topic for the following discussion within the INSPIRE initiative to support data variation towards its interoperability.

At first glance, the availability of data for being queried within interlinked RDFs completely satisfies The Five Star Open Data model's requirements. Moreover, Eurostat and EDP are positioning themselves as LOD providers. But in fact, while trying to integrate geospatial LOD into BI tools, most users are following a typical Extract, Transform, Load (ETL) process "in which Linked Data becomes just another data source" (Mai et al. 2019). This happens because the Linked Open Data paradigm is still largely isolated from the needs of its potential users. New Linked Open Data providers appear without previous identification of potential applicability, and their target users are out of focus. Therefore, some questions of The Five Star Open Data Engagement model (Davies 2012) remain unanswered by geospatial Linked Open Data providers. In particular, those that define the fourth and the fifth stars: "Do you provide or link to tools for people to work with your datasets?", "Do you broker or provide support to people to build and sustain useful tools and services that work with your data?" (Davies 2012).

The next stage involved selection, overview, and exploration of leaders of the current location BI market in terms of the supported geospatial data transfer technologies. This research went through four BI tools such as *Tableau*, *CARTO*, *OmniSci*, and *ArcGIS Insights*. The examination revealed that, on average, the second and the third star-levels in relation to The Five Star Open Data model (Berners-Lee 2006; Berners-Lee 5) applied to spatial data formats (McKeague et al. 2020) are supported by the selected BI tools. And Tableau is one that allows running SPARQL queries on data that corresponds to the fifth star-level. Although, this

functionality available with the assistance of the third-party XML driver or tool (in some cases Virtuoso instance is used). The analysis conducted on Business Intelligence software showed that all the selected tools allow integrating data from outside via Download & Import technology, public data domains available on the web, API, and SQL queries on databases. Although mentioned data transfer technologies do not match the initial technology behind the Linked Open Data paradigm, they are still more common for the significant majority of geospatial data users. Almost all BI and Location BI software developers try to accommodate data transfer functionality that users are well familiar with. In this way, such tools impressively reduce users' effort and time, guarantee flexibility, and favor users' needs. Hence, it would be great if the Linked Open Data community could focus on making LOD more affordable, usable, and BI-friendly, taking into account already widely-used geospatial data standards and data integration technologies.

The thematic scenarios developed in the previous steps were implemented within the visualization framework of each BI tool. Linked Data extracted from the research LOD providers was not only integrated into the software but also processed and transformed into demonstrative use-cases. Such an approach helped to get deeper insights into the characteristics of the integration process. As expected earlier, the scenarios' performance resulted in such visualization products as maps, web-maps, histograms, charts and others. The visualization outputs demonstrating both scenarios for each BI tool are illustrated within the ANNEXES 1-4.

## 6. Conclusion

Geospatial data appear as a fundamental component of the global Linked Data as a *location is a primary key* defining and densely interlinking events, objects, people and relationships.

Linked Open Data has the potential to be one of the convenient ways to *democratize access to geospatial data*. The LOD paradigm puts forward a unique and different perspective on structuring, publishing, discovering, accessing, reusing, and integrating geospatial data, hence addressing various challenges for its potential users, particularly within Location Business Intelligence.

Linked Data concept has experienced a number of investigations in terms of theoretical guidelines determination (star models) as well as practical prototypes development. Many initiatives appear to adopt LOD technology to publish geospatial data on the Web. The main aim behind is a provision of integrated access to geospatial data coming from heterogeneous sources. However, this idea's novelty and complexity do not ensure its successful integration across applications and domains, especially with *Location Business Intelligence*.

Geospatial data integration in BI is a fundamental step for the creation of map products. It is inherited in the first Model of Cartographic Communication (Crampton 2001). Therefore, the result of successful Geospatial LOD integration is the source for potential visualization products, which lead to prospective use-cases, solutions and decisions across industries. Unfortunately, Geospatial Linked Open Data experiences relatively slow adoption speed. Therefore, the geospatial data is still not ideally preserved from the LOD world to the BI solutions, and the gap remains uncovered. The connector frameworks and adapters for BI and GIS software proposed by various initiatives are mainly temporal solutions with significant performance and scalability limits.

In most cases, Linked Data still serves as just a *theoretical precondition* to actual practical applications. The conducted research highlights that the integration gap between Geospatial Linked Open Data providers and Business Intelligence tools has to be a *core direction* for further profound investigation and development. Engineers who initiate technological advancements in Geospatial Linked Open Data ecology have to focus not on the over-engineering but rather *target successful user experience*. Moreover, they have to encourage (government) data owners to publish their data according to already existing basic LOD principles and constantly control this process's quality.

To avoid isolation, LOD providers have to grab decision-makers and stakeholders' attention by publishing Geospatial Linked Open Data in a clear and easy-accessible manner using common interfaces for the users.

Thereby, the identified and analyzed integration challenges emphasize the need for a more in-depth review and improvement of present technologies supported by LOD platforms and BI tools. In particular, the existing (Geospatial) Open Data Models must be reviewed, updated and advanced to provide not only theoretical insights but rather real practical applications. Successful deep integration of both would allow Business Intelligence tools to *directly transform the potential of Geospatial Linked Open Data* into valuable decisions and solutions across domains.

## REFERENCES

- ArcGIS Maps for Microsoft Power BI. Available at: <https://www.esri.com/en-us/arcgis/products/arcgis-maps-for-power-bi/overview> (Accessed: 27 May 2020).
- Battle, R. and Kolas, D. (2012) 'Enabling the geospatial Semantic Web with Parliament and GeoSPARQL', *Semantic Web*. IOS Press, 3(4), pp. 355–370. doi: 10.3233/SW-2012-0065.
- Bauer, F. and Kaltenböck, M. (2011) *Linked Open Data: The Essentials*, *Semantic Web*. Available at: <https://semantic-web.com/project/lod2-the-linked-data-technology-stack-for-enterprises/> (Accessed: 9 August 2020).
- Berners-Lee, T. (5AD) 'Star Open Data', 5 Star Data. Available at: <https://5stardata.info/en/> (Accessed: 10 August 2020).
- Berners-Lee, T. (2006) *Linked Data*. Available at: <https://www.w3.org/DesignIssues/LinkedData.html> (Accessed: 9 August 2020).
- Bizer, C., Berners-Lee, T. and Heath, T. (2009) 'Linked Data - The Story So Far'. Available at: <https://www.semanticscholar.org/paper/Linked-Data-The-Story-So-Far-Bizer-Heath/9f54a0057d0694bc7d1dcf69d186e313ca92775c> (Accessed: 9 August 2020).
- Bolon, B. R., Bull, J. F. and Ward, M. L. (2012) 'Location intelligence management system', US Patent. Available at: <https://patentimages.storage.googleapis.com/27/42/a1/32454b2c5a2833/US8224348.pdf> (Accessed: 8 September 2020).
- Bowes, P. (2007) 'Location Intelligence: The New Geography of Business', *BUSINESS WEEK RESEARCH SERVICES*. Available at: [http://media.govtech.net/RC\\_PITNEYBOWES/BusinessWeek.pdf](http://media.govtech.net/RC_PITNEYBOWES/BusinessWeek.pdf) (Accessed: 9 September 2020).
- Crampton, J.W. (2001) 'Maps as Social Constructions: Power, Communication, and Visualisation'. *Progress in Human Geography*. 25, 235-252.
- CARTO. Available at: <https://carto.com/> (Accessed: 3 August 2020).
- CARTO Platform. Available at: <https://carto.com/platform/> (Accessed: 14 May 2020).
- CARTO | NGIS. Available at: <https://ngis.com.au/Our-Technology/Carto> (Accessed: 11 July 2020).
- Chaturvedi, A. (2019) *Advancements in location to alter human-technology interaction*, *Geospatial World*. Available at: <https://www.geospatialworld.net/article/advancements-in-location-to-alter-human-technology-interaction/> (Accessed: 9 September 2020).
- Colpaert, P. et al. (2013) 'The 5 Stars of Open Data Portals', in. 7th international conference on methodologies, technologies and tools enabling e-Government (MeTTeG), pp. 61–67. Available at:

[https://www.researchgate.net/publication/265140938\\_The\\_5\\_stars\\_of\\_open\\_data\\_portals](https://www.researchgate.net/publication/265140938_The_5_stars_of_open_data_portals)  
(Accessed: 26 May 2020).

COVID & Empl Working Group. The COVID confinement measures and EU labour markets. Available at:  
[https://publications.jrc.ec.europa.eu/repository/bitstream/JRC120578/jrc120578\\_report\\_covid\\_confinement\\_measures\\_final\\_updated\\_good.pdf](https://publications.jrc.ec.europa.eu/repository/bitstream/JRC120578/jrc120578_report_covid_confinement_measures_final_updated_good.pdf) (Accessed: 27 May 2020).

Davies, T. (2012) 5-Stars of Open Data Engagement?, Open Data Engagement. Available at: <http://www.timdavies.org.uk/2012/01/21/5-stars-of-open-data-engagement/> (Accessed: 10 August 2020).

Davies, T. (2012) Open Data Engagement. Available at:  
<http://www.opendataimpacts.net/engagement/> (Accessed: 10 August 2020).

Esri ArcGIS Insights. Available at: <https://insights.arcgis.com/> (Accessed: 26 May 2020).

European Commission. NACE Rev. 2. Available at:  
<https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF>  
(Accessed: 27 May 2020).

European Commission. NUTS-3. Available at:  
<https://ec.europa.eu/eurostat/web/nuts/background> (Accessed: 27 May 2020).

European Data Portal (2019) Portal Version 4.3 - User Manual. Available at:  
[https://www.europeandataportal.eu/sites/default/files/edp\\_s1\\_man\\_portal-version\\_4.3-user-manual\\_v1.0.pdf](https://www.europeandataportal.eu/sites/default/files/edp_s1_man_portal-version_4.3-user-manual_v1.0.pdf) (Accessed: 4 June 2020).

European Data Portal. The European Data Portal: Opening up Europe's public data. Available at:  
[https://www.europeandataportal.eu/sites/default/files/edp\\_factsheet\\_what\\_is\\_edp\\_project\\_online.pdf](https://www.europeandataportal.eu/sites/default/files/edp_factsheet_what_is_edp_project_online.pdf) (Accessed: 3 June 2020).

Eurostat (2015) Eurostat - European Commission, Regions in the European Union. Nomenclature of territorial units for statistics.

Eurostat (2017) Eurostat BD Metadata. Available at:  
[https://ec.europa.eu/eurostat/cache/metadata/en/bd\\_esms.htm](https://ec.europa.eu/eurostat/cache/metadata/en/bd_esms.htm) (Accessed: 5 August 2020).

Eurostat (2018) Eurostat Demo Metadata. Available at:  
[https://ec.europa.eu/eurostat/cache/metadata/en/demo\\_r\\_gind3\\_esms.htm](https://ec.europa.eu/eurostat/cache/metadata/en/demo_r_gind3_esms.htm) (Accessed: 5 August 2020).

Eurostat (2019) Eurostat Reg Metadata. Available at:  
[https://ec.europa.eu/eurostat/cache/metadata/en/reg\\_area3\\_esms.htm](https://ec.europa.eu/eurostat/cache/metadata/en/reg_area3_esms.htm) (Accessed: 5 August 2020).

Eurostat Database. Available at: <https://ec.europa.eu/eurostat/data/database>  
(Accessed: 26 May 2020).

Eurostat Bulk Download Guidelines (2019). Available at: <https://ec.europa.eu/eurostat/data/bulkdownload> (Accessed: 11 June 2020).

Feinberg, D. et al. (2016) 'Cool Vendors in DBMS, 2016'. Available at: <https://www.gartner.com/en/documents/3288923/cool-vendors-in-dbms-2016> (Accessed: 10 September 2020).

Folmer, E. et al. (2019) 'Enhancing the Usefulness of Open Governmental Data with Linked Data Viewing Techniques', Proceedings of the 52nd Hawaii International Conference on System Sciences. doi: 10.24251/hicss.2019.352.

'Geobuiz Report 2019' (2019). Available at: <https://geobuiz.com/geobuiz-report-2019/> (Accessed: 10 September 2020).

Geodata Eurostat (2020). Available at: <http://ec.europa.eu/eurostat/web/gisco/geodata> (Accessed: 10 June 2020).

Giannopoulos, G. et al. (2014) 'FAGI: A Framework for Fusing Geospatial RDF Data', in On the Move to Meaningful Internet Systems: OTM 2014 Conferences. Springer Berlin Heidelberg, pp. 553–561. doi: 10.1007/978-3-662-45563-0\_33.

GISCO Eurostat (2020). Available at: <https://ec.europa.eu/eurostat/web/gisco/overview> (Accessed: 10 June 2020).

Google Scholar (2020). Emerald Group Publishing Limited. Available at: [https://scholar.google.com/scholar?q=Eurostat++linked+data&hl=en&as\\_sdt=0,5](https://scholar.google.com/scholar?q=Eurostat++linked+data&hl=en&as_sdt=0,5) (Accessed: 27 May 2020).

Google Scholar (2020). Available at: [https://scholar.google.com/scholar?hl=en&as\\_sdt=0%2C5&q=European+Data+portal+linked+data&btnG=](https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=European+Data+portal+linked+data&btnG=) (Accessed: 27 May 2020).

Google Scholar since 2019 (2020). Available at: [https://scholar.google.com/scholar?hl=en&as\\_sdt=0%2C5&as\\_ylo=2019&q=Eurostat++linked+data&btnG=](https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&as_ylo=2019&q=Eurostat++linked+data&btnG=) (Accessed: 27 May 2020).

Halb, W., Raimond, Y. and Hausenblas, M. 'Building Linked Data For Both Humans and Machines\*'. Available at: <http://ra.ethz.ch/CDstore/www2008/events.linkeddata.org/ldow2008/papers/06-halb-raimond-building-linked-data.pdf> (Accessed: 9 June 2020).

Hogan, A. et al. (2020) 'Knowledge Graphs'. Available at: <http://arxiv.org/abs/2003.02320> (Accessed: 9 August 2020).

International Labour Organisation (2020) ILO Monitor: COVID-19 and the world of work. Third edition. Available at: [https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/briefingnote/wcms\\_743146.pdf](https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/briefingnote/wcms_743146.pdf) (Accessed: 27 May 2020).

Kafka (2020) OmniSci. Available at: <https://docs.omnisci.com/loading-and-exporting-data/supported-data-sources/kafka> (Accessed: 22 July 2020).



Kirstein, F. et al. (2019) 'Linked Data in the European Data Portal: A Comprehensive Platform for Applying DCAT-AP', in *Electronic Government*. Springer International Publishing, pp. 192–204. doi: 10.1007/978-3-030-27325-5\_15.

Koubarakis, M. et al. (2017) 'Big, Linked Geospatial Data and Its Applications in Earth Observation', *IEEE Internet Computing*, 21(4), pp. 87–91. doi: 10.1109/MIC.2017.2911438.

'Location Analytics- The Key to More Powerful Analysis ebook'. Available at: <https://www.esri.com/en-us/arcgis/products/arcgis-insights/free-ebook> (Accessed: 18 June 2020).

Magic Quadrant for Analytics and Business Intelligence Platforms. Available at: [https://www.gartner.com/doc/reprints?id=1-3TXXSLV&ct=170221&st=sb&ocid=mkto\\_eml\\_EM597235A1LA1](https://www.gartner.com/doc/reprints?id=1-3TXXSLV&ct=170221&st=sb&ocid=mkto_eml_EM597235A1LA1) (Accessed: 14 May 2020).

Magic Quadrant Research Methodology Gartner. Available at: <https://www.gartner.com/en/research/methodologies/magic-quadrants-research> (Accessed: 27 May 2020).

Mai, G. et al. (2016) 'A Linked Data Driven Visual Interface for the Multi-perspective Exploration of Data Across Repositories', in *VOILA@ ISWC. The Fourth International Workshop Visualization and Interaction for Ontologies and Linked Data*, pp. 93–101. Available at: <http://dx.doi.org/> (Accessed: 7 September 2020).

Mai, G. et al. (2019) 'Deeply integrating Linked Data with Geographic Information Systems', *Transactions in GIS*, 23(3), pp. 579–600. doi: 10.1111/tgis.12538.

Mangaladevi, K., Beek, W. and Kuhn, T. (2017) 'Understanding Knowledge Networks', *CEUR workshop proceedings*. CEUR Workshop Proceedings, 1946, pp. 38–49. Available at: <https://research.vu.nl/en/publications/understanding-knowledge-networks> (Accessed: 9 August 2020).

Maté, A., Llorens, H. and de Gregorio, E. (2012) 'An Integrated Multidimensional Modeling Approach to Access Big Data in Business Intelligence Platforms', *Lecture Notes in Computer Science*, pp. 111–120. doi: 10.1007/978-3-642-33999-8\_14.

McKeague, P. et al. (2020) 'One Archaeology: A Manifesto for the Systematic and Effective Use of Mapped Data from Archaeological Fieldwork and Research', *Information. An International Interdisciplinary Journal*. Multidisciplinary Digital Publishing Institute, 11(4), p. 222. doi: 10.3390/info11040222.

Ngomo, A.-C. N. and Auer, S. (2011) 'LIMES -- A Time-Efficient Approach for Large-Scale Link Discovery on the Web of Data'. unknown, pp. 2312–2317. doi: 10.5591/978-1-57735-516-8/IJCAI11-385.

NUTS GISCO (2020). Available at: <https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts> (Accessed: 10 June 2020).

OmniSci | Accelerated Analytics Platform. Available at: <https://www.omnisci.com> (Accessed: 14 May 2020).

OmniSci for Geospatial Analysis and Visualization OmniSci. Available at: <https://www.omnisci.com/solutions/geospatial-analysis> (Accessed: 27 May 2020).

OmniSci Immerse | Interactive Visual Analytics OmniSci. Available at: <https://www.omnisci.com/platform/immerse> (Accessed: 22 July 2020).

Ontotext (2020) What are Linked Data and Linked Open Data?, Ontotext. Available at: <https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/> (Accessed: 9 August 2020).

Panian, Z. (2012) 'A New Dimension of Business Intelligence: Location-based Intelligence', *International Journal of Industrial and Systems Engineering* 6, pp. 338–343. Available at: <https://www.semanticscholar.org/paper/A-New-Dimension-of-Business-Intelligence%3A-Panian/91e48aa788ce88adbc2e8e0698d6071dedd7518a> (Accessed: 30 August 2020).

Power, D. J. (2007) 'A brief history of decision support systems', *DSSResources. com*, 3. Available at: <http://dssresources.com/history/dsshistoryv28.html> (Accessed: 6 September 2020).

Ramirez, R. et al. (2015) 'Scenarios as a scholarly methodology to produce "interesting research"', *Futures*, 71, pp. 70–87. doi: 10.1016/j.futures.2015.06.006.

Regalia, B., Janowicz, K. and Gao, S. (2016) 'VOLT: A Provenance-Producing, Transparent SPARQL Proxy for the On-Demand Computation of Linked Data and its Application to Spatiotemporally Dependent Data', in *The Semantic Web. Latest Advances and New Domains*. Springer International Publishing, pp. 523–538. doi: 10.1007/978-3-319-34129-3\_32.

Ronzhin, S., Folmer, E. and Lemmens, R. (2018) 'Technological Aspects of (Linked) Open Data', in van Loenen, B., Vancauwenberghe, G., and Crompvoets, J. (eds) *Open Data Exposed*. The Hague: T.M.C. Asser Press, pp. 173–193. doi: 10.1007/978-94-6265-261-3\_9.

Roosens, D., McGlinn, K. and Debruyne, C. (2019) 'Using Maps for Interlinking Geospatial Linked Data', in *On the Move to Meaningful Internet Systems: OTM 2019 Conferences*. Springer International Publishing, pp. 209–226. doi: 10.1007/978-3-030-33246-4\_14.

Statistics Explained - Eurostat. Available at: [https://ec.europa.eu/eurostat/statistics-explained/index.php/Accessing\\_European\\_statistics](https://ec.europa.eu/eurostat/statistics-explained/index.php/Accessing_European_statistics) (Accessed: 11 June 2020).

Supported data—Insights Get Started | Documentation. Available at: <https://doc.arcgis.com/en/insights/latest/get-started/supported-data.htm> (Accessed: 20 June 2020).

Tableau Maps. Tableau Software. Available at: <https://www.tableau.com/solutions/maps> (Accessed: 26 May 2020).

Ubickan blogspot (2006) 'The Map Communication Model and Critical Cartography'. Available at: <http://ubickan.blogspot.com/2006/08/map-communication-model-and-critical.html> (Accessed: 10 September 2020)

Ulrich, H. (1992) Anleitung zum ganzheitlichen Denken und Handeln. Stiftung zur Förderung systemorientierten Managementlehre.

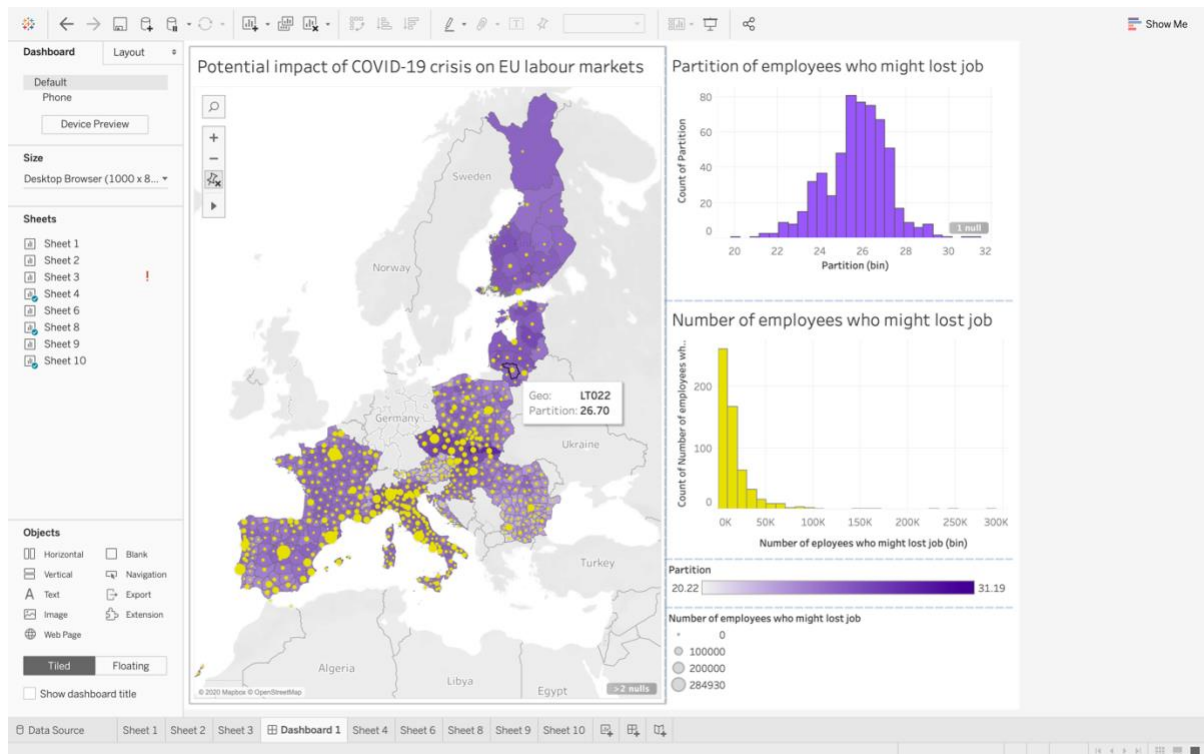
Ulrich, H. and Probst, G. J. B. (1991) 'Anleitung zum ganzheitlichen Denken und Handeln: ein Brevier für Führungskräfte'. Haupt.

W3C (2016) Linked Data, W3C. Available at: <https://www.w3.org/wiki/LinkedData> (Accessed: 9 August 2020).

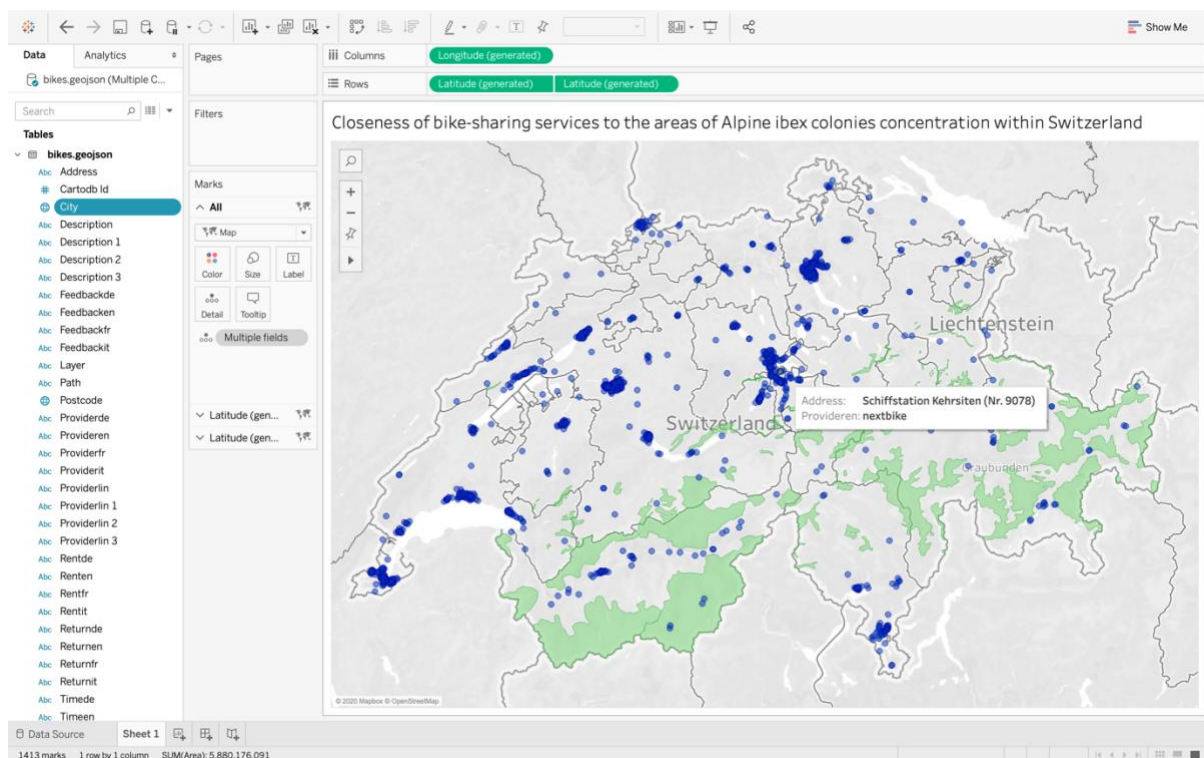
Watson, H. J. (2009) 'Tutorial: Business Intelligence – Past, Present, and Future', Communications of the Association for Information Systems. doi: 10.17705/1CAIS.02539.

What we do - Eurostat. Available at: <https://ec.europa.eu/eurostat/about/overview/what-we-do> (Accessed: 15 May 2020).

# ANNEX 1 – Scenario 1 & Scenario 2 visualizations | Tableau

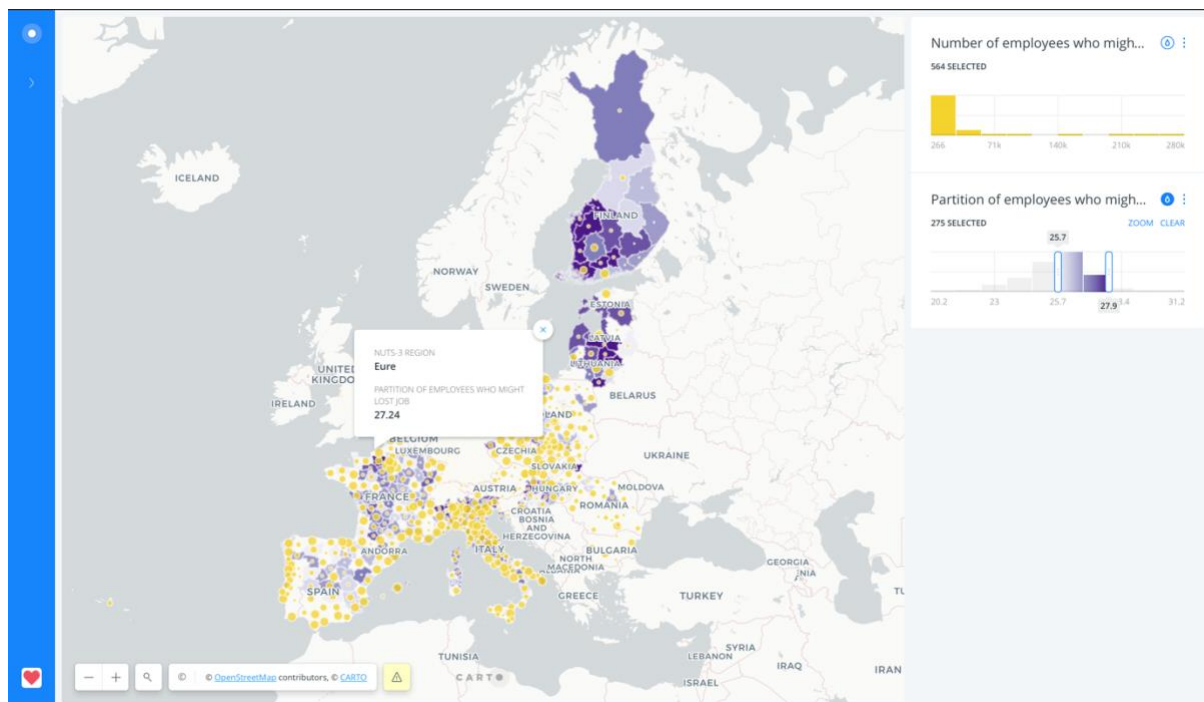


*Potential impact of COVID-2019 crisis on EU labor markets | Tableau \**

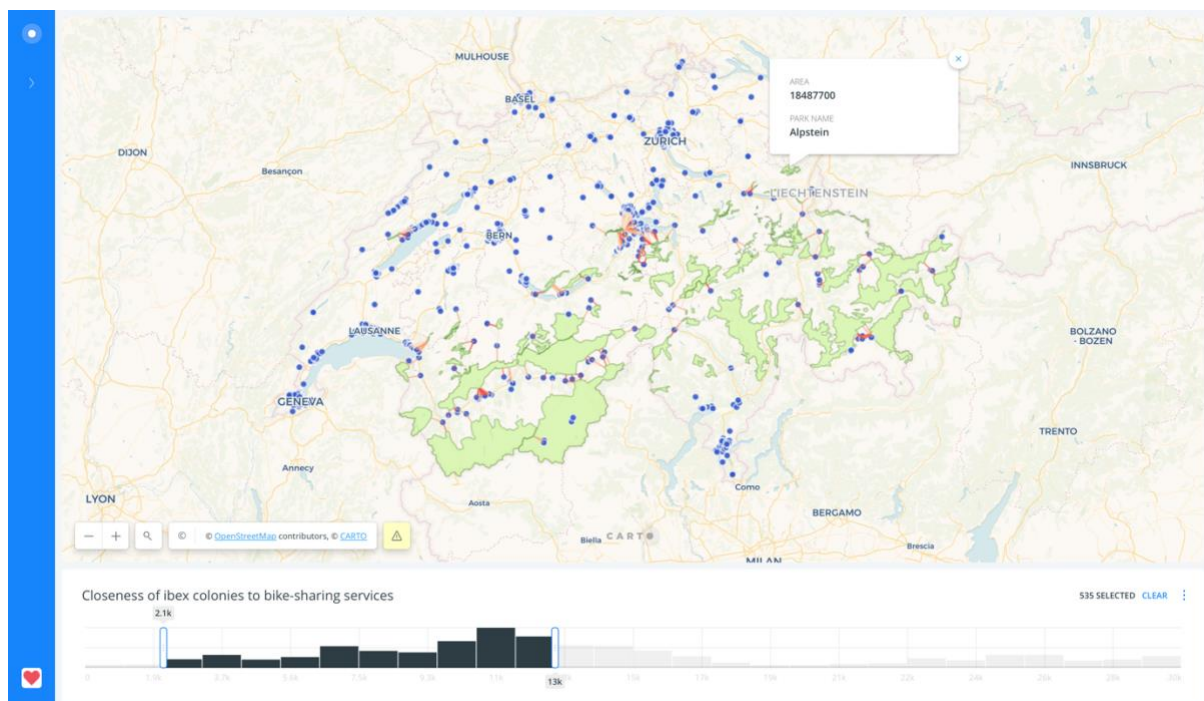


*Closeness of bike-sharing services to the areas of Alpine ibex colonies concentration within Switzerland | Tableau \**

## ANNEX 2 – Scenario 1 & Scenario 2 visualizations | CARTO



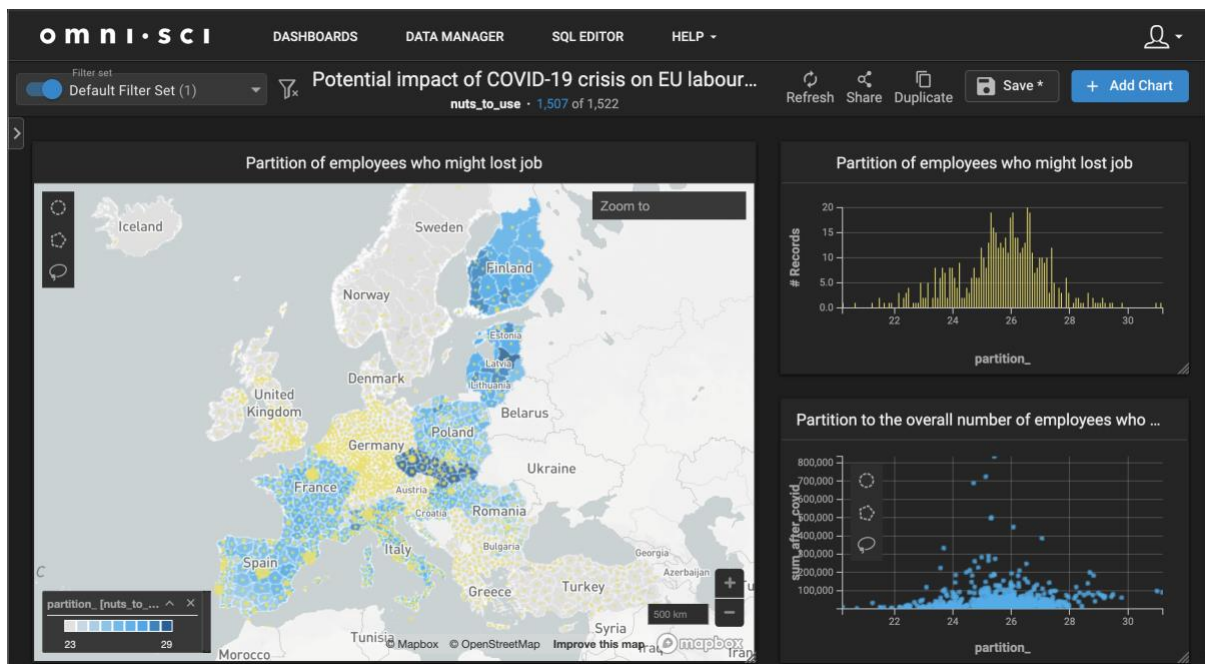
Potential impact of COVID-2019 crisis on EU labor markets | CARTO



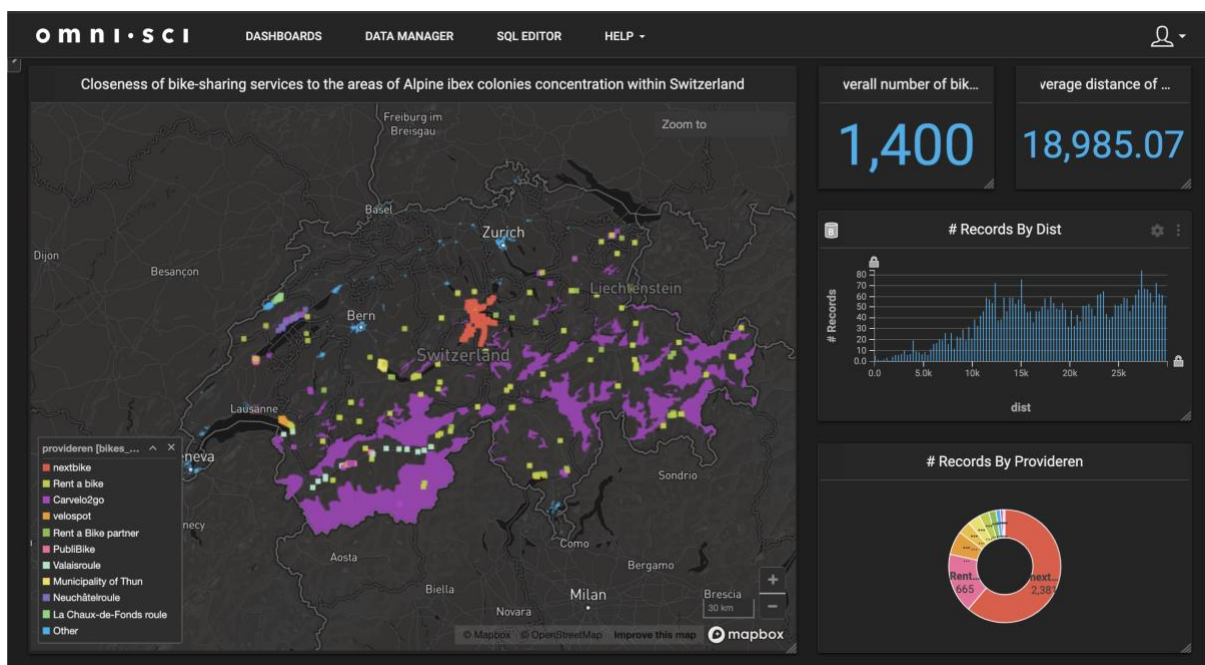
Closeness of bike-sharing services to the areas of Alpine ibex colonies concentration within Switzerland | CARTO



## ANNEX 3 – Scenario 1 & Scenario 2 visualizations | OmniSci

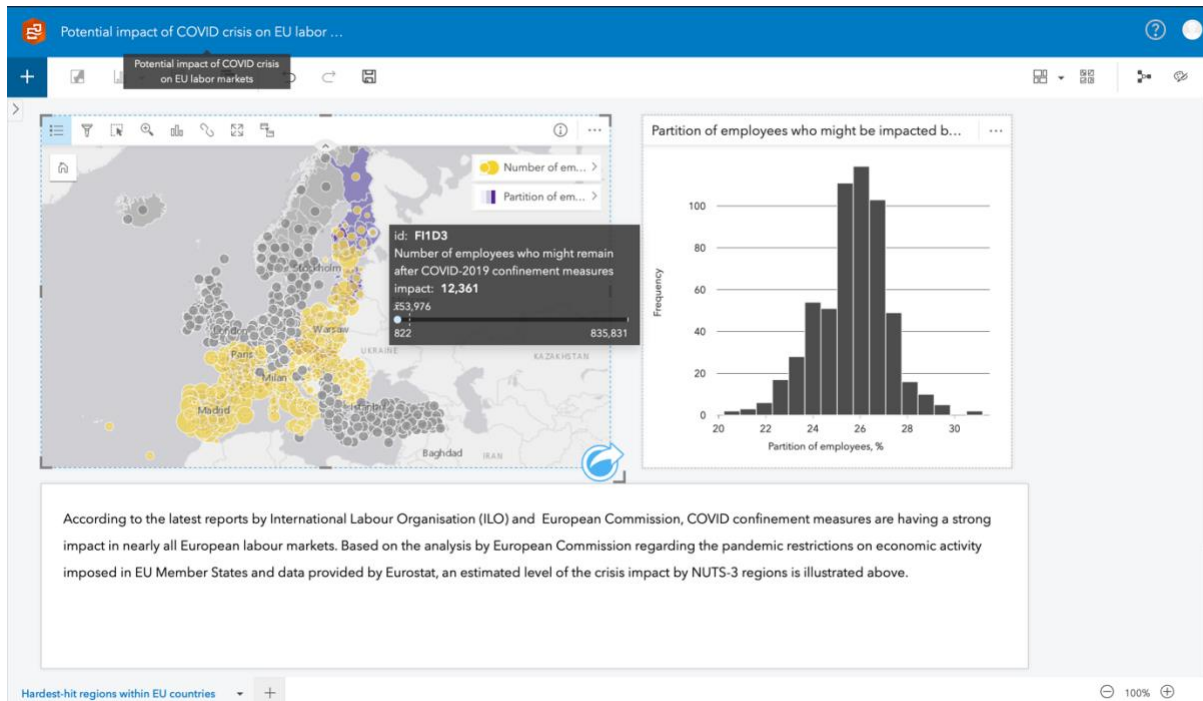


*Potential impact of COVID-2019 crisis on EU labor markets | OmniSci \**

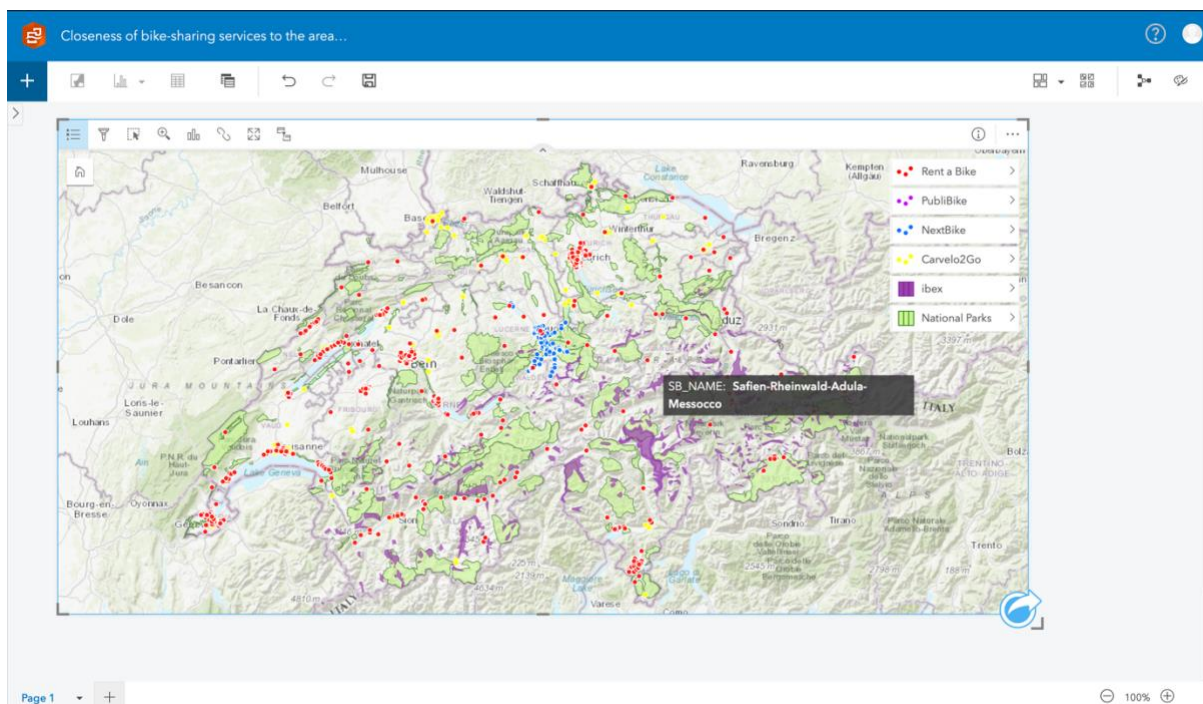


*Closeness of bike-sharing services to the areas of Alpine ibex colonies concentration within Switzerland | OmniSci \**

## ANNEX 4 – Scenario 1 & Scenario 2 visualizations | ArcGIS Insights



*Potential impact of COVID-2019 crisis on EU labor markets | ArcGIS Insights \**



*Closeness of bike-sharing services to the areas of Alpine ibex colonies concentration within Switzerland | ArcGIS Insights \**

\* Public sharing capabilities are limited