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

Technical University of Munich
TECHNISCHE UNIVERSITÄT WIEN Vienna University of Technology
TECHNISCHE UNIVERSITÄT DRESDEN
ITC UNIVERSITY OF TWENTE.

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 useful comments.

I would like to thank a lot to Prof. Georg Gartner for his involvement and useful remarks and critiques at each stage of the research process, 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 parents for their 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. 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 be able to improve 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 leaders of the global BI market. 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

LIST OF FIGURES

LIST OF TABLES

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 questions are specified. Finally, the outline of the thesis is given.

1.1 Motivation and problem statement

Given a continuously growing stack of open-source tools, it is now easier than ever to publish and consume geodata on the Semantic Web (Mai et al. 2019). This allows 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 what makes it more complex. Along with this, data heterogeneity has surpassed the capacity of current data management technologies. It is already not possible to process, store, and manage all this information in local servers even for the biggest companies and Business Intelligence (BI) systems (Maté et al. 2012). This phenomenon became one of the driving forces for the development of 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 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 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 geo-data 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 this plethora of data with potentially interested parties remains largely unanswered (Mai et al. 2019). And Business Intelligence is one of those industries that are 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 new opportunities. In order to understand the significance of location information, businesses often refer to BI tools, powered by GIS. And with an opportunity to accommodate access to Linked Open Data, Location BI would help users gain more value out of geospatial data.

According to (Folmer et al. 2019), when Linked 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.

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 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 still 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 will specify 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 Linked Data Integration
- Analysis of relevant Linked 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 in BI tools to provide access to the global 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 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. *Linked Data* refers to the provision of integrated access to data 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).

Rather than being stored in a traditional relational database, Linked Data is stored in a graph-based data model, typically indexed by a triple store, using standardized serialization formats like Turtle and RDF/XML (W3C 2016). The use of URIs/IRIs as identifiers in the data allows for the creation of links between datasets, providing context to the data, and thereby improving its understandability and usability (Bauer and Kaltenböck 2011). Furthermore, it allows for the discovery of new data by potential users. SPARQL is a standardized query language that can be used to answer complicated questions over one or more Linked Datasets (Folmer et al. 2019).

This is possible for Linked Data to be open. *Linked Open Data* is a powerful blend of Linked Data and Open Data: it is both linked and uses open licenses. One notable example of an LOD set is DBpedia – a crowd-sourced community effort to extract structured information from Wikipedia and make it available on the Web (Ontotext 2020).

The more things, events, people, locations, etc. are connected together, the more powerful the Web of Data is. However, in order to link, merge and integrate huge sets of data from disparate sources, some basic guidelines must be followed. The inventor of the World Wide Web and the creator and advocate of the Semantic Web and Linked Data, Sir Tim Berners-Lee (Ontotext 2020), laid down the 5-Star Linked Open Data model to indicate which criteria must be met by 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

Within this model, 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*, also due to advances in the Earth Observation domain (Koubarakis et al. 2017). Notably, the domain of geospatial data contains complex datasets (Eurostat 2015) and their linkage can assure convenient interoperability of location data. Moreover, as most datasets also have a geospatial dimension that is either explicit or implicit, it comes as no surprise that "location" is a convenient way for aligning and combining 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 advancement of the dataset containing *geospatial data* (see Figure 1).

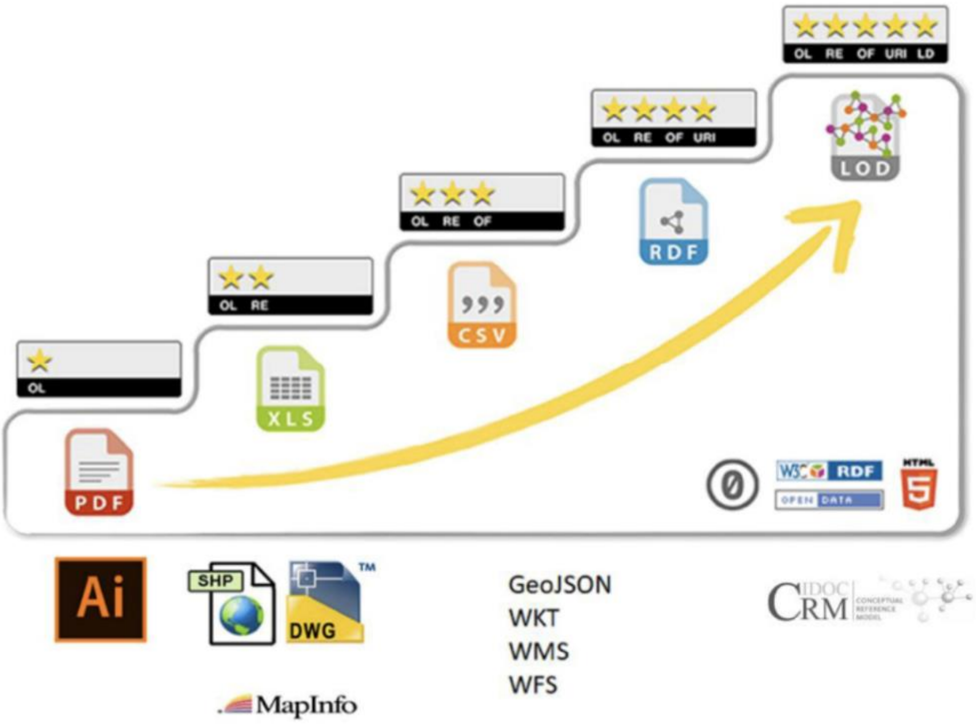


Figure 1. The Five Star Open Data model (CC-0) (Berners-Lee 2006; Berners-Lee 5) applied to spatial data types (McKeague et al. 2020)

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; Davies 2012).

Tim Berners-Lee proposed the 5 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 and are a real success in the Open Data ecology. Currently, however, there are no clear pointers or guidelines for *Open Data Portals* aiming at stimulating data reuse and improving data quality (Colpaert et al. 2013). Moreover, it concerns those platforms providing access to geospatial Linked Data. In order to fill this gap (Colpaert et al. 2013), in 2013 there was proposed a 5-Star evaluation model to guide Public Administrations in setting up an Open Data Portal. This 5-Star system for Open Data Portals is complementary to the star system of Open Data Engagement and Linked Open Data and, thus, helps Public Administrations to maximize their stars in these two other evaluation frameworks.

The stars' categorization starts with portals linking to various datasets and continues towards a meta-data portal for both the datasets and the re-use of the datasets. The fourth-star category of Open Data Portals takes care of the data publication itself. Finally, a data hub is set up where data becomes 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 particularly important was Michael S. Scott Morton's (1967) doctoral dissertation research. He built, implemented, and tested a system to support planning for laundry equipment.

Later, he published an influential Sloan Management Review article (Gorry and Scott Morton 1971) and book (1971) that helped spread decision support concepts and provided a name for these analytical applications — *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. Hugh J. Watson (Watson 2009) mentions some of the directions in his article, in particular, executive information systems (EIS), group decision support systems (GDSS), and *geographic information systems (GIS)*. Transformations within the DSS domain, together with technological advances, led to the development of the *Business Intelligence* term. Gradually, *model-centric* DSS, those of focus of academicians with quantitative backgrounds, were getting less of interest. By way of contrast, *data-centric DSS* became more valued in meeting business needs and forming the BI industry.

The more use-cases and applications appeared the more diverse data entering the Business Intelligence world was of interest. Most businesses are immersed in a world of readily accessible data that could significantly improve their companies' efficiency, effectiveness, and profitability (Bowes 2007). In 2012 Zeljko Panian (Panian 2012) emphasized that only a minority of firms takes advantage of this rich mine of data – or, apparently, are even aware that it exists. Here he mentions data that pertains to *location* – the spatial environment in which a given organization operates, interacts with its customers, and transacts business. 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 that will benefit all aspects of most enterprises. 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 industry. It is a means of combining location data with business data/processes in order to draw meaningful insights, discover relationships between various factors and identify trends. It has a wide range of applications ranging from rendering basic features such as map-based visualization to enabling complex analytics for scenario analysis. The value chain of Location Intelligence begins with creating software or platforms, geocoding engines with analytics features, added with map rendering capabilities and simulation algorithms. This is followed by capturing map and location data, which includes getting the map data (base maps, satellite imageries, map layers, etc.) and geographic coordinates of objects and Points-of-Interest (PoI). These are integrated with business data and processes to create a *user and industry-specific solutions*.

GeoBuiz (GeoBuiz 2019) emphasizes that *geospatial* becomes the 'by-default' technology in various domains, industries and simply in daily life. Many BI tools accommodate GIS software technology as a valuable extension that provides a framework to manage, visualize, analyze, and ultimately understand the significance of location information.

3. Methodology

This chapter explains 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

Across different fields and practices, the term 'scenarios' does not have only one generally accepted meaning, serve the same purpose, or involve one form of production (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 problem regarded is influenced by many factors, that these factors are interlinked in a manifold way, and that there are a number of possible answers to the problem (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 implementation of the scenarios 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 are having a strong impact in nearly all European labor markets (International Labour Organisation 2020; COVID & Empl Working Group). On the basis of 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

In this section, BI tools that will be used within this thesis project are introduced. 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 for operations on geospatial data.

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 (Magic Quadrant for Analytics and Busi...; Magic Quadrant Research Methodology).

It offers a visual-based exploration experience that enables business users to access, prepare, analyze, and present findings in their data. Tableau allows users to ingest data rapidly from a broad range of data sources, blend them, and visualize results using best practices in visual perception. Data can easily be manipulated during visualization, such as when creating groups, bins, and hierarchies (*Magic Quadrant for Analytics and Business Intelligence Platforms*).

Moreover, Tableau Maps supports Location Intelligence operations. Tableau is designed to make the most of geographical data so that users can get to the "where" as well as the "why." With instant geocoding, Tableau automatically turns the location data and information users already have into rich, interactive maps with 16 levels of zoom - or use custom geocodes to map what matters to your business (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 tools. The company is positioned as a Location Intelligence platform due to tools with an aptitude for data analysis and visualization that do not require previous GIS or development experience (CARTO).

Data Scientists, Developers, and Analysts solve spatial problems using CARTO's data and analysis to understand where and why things happen, optimize business processes, and predict future outcomes through the power of Spatial Data Science (CARTO).

OmniSci

OmniSci (former MapD) is an open platform that unites analytics, data science, and location intelligence workflows (*OmniSci | Accelerated Analytics Platform*). Gartner recognizes OmniSci as a platform that provides database management and visual analytics software optimized for graphics processing unit (GPU) accelerators. OmniSci offers a built-in data visualization option that can leverage the rendering capabilities of the server-side GPUs, and also offers support for third-party visualization tools via ODBC or Apache Thrift (Feinberg *et al.*).

The platform is used in business and government to find insights in data beyond the limits of mainstream analytics tools. OmniSci removes the technological limitations preventing frequent, high-speed decisions on huge streams of operational data.

OmniSci is at the forefront of location intelligence for big data, combining unprecedented volumes of geospatial and BI data into one interactive experience (OmniSci for Geospatial Analysis and V...).

ArcGIS Insights

Part of the Esri Geospatial Cloud, ArcGIS Insights, is an analysis software that fuses location analytics with open data science and business intelligence workflows (Esri).

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

Insights for ArcGIS is available as a part of ArcGIS Enterprise and Online and provides a new user experience with an analysis that focuses on simpler ways to work with data. This includes visualizations, drag and drop analytics, and on-the-fly filtering and aggregations, making interactive and exploratory analysis fast and intuitive. By blending business intelligence (BI)-like capabilities with the best in location analytics, with 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

In this section, Linked Open Data providers that will be used within this thesis project are presented. 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

With over 300 million statistical data, Eurostat is a mine of statistical information and covers all areas of European society. Eurostat's primary role is to process and publish comparable statistical information at the European level (*What we do - Eurostat*).

Eurostat does not collect data. This is done in the Member States by their statistical authorities. They verify and analyze national data and send them to Eurostat. Eurostat's role is to consolidate the data and ensure they are comparable, using a harmonized methodology. *Eurostat* is *actually the only provider of statistics at the European level,* and the data we issue are harmonized as far as possible (*What we do - Eurostat*). Furthermore, Eurostat offers access to spatial Linked Open Data.

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

European Data Portal (EDP)

The European Data Portal (EDP) is an initiative 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-use by everyone ('European Data Portal'). It offers more than 1 million datasets across all areas of European Member States.

The European Data Portal harvests the metadata (data about the data) of Public Sector Information available on public data and geospatial portals across European countries. Portals can be national, regional, local or domain-specific. They cover the EU Member States, EFTA countries and countries involved in the EU's neighborhood policy (European Data Portal).

The European Data Portal is a central access point for metadata of Open Data published by public authorities in Europe and acquires data from more than 70 national data providers. The platform is a starting point in adopting the Linked Data specification DCAT-AP, aiming to increase interoperability and accessibility of Open Data (Kirstein et al. 2019).

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

4. Implementation

4.1 Data Access Interfaces in Linked Data providers

4.1.1 Eurostat

Eurostat provides detailed statistics for the entire European Union as well as additional statistics for major non-European countries (Halb et al.). Eurostat data covers a range of thematic direction/topics that can be used by governments, businesses, the education sector, journalists and the public for fulfilling various tasks and purposes.

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

Database. The majority of Eurostat's statistics may 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*. Data can be downloaded from the Data explorer in various formats (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 a range of different 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.

Statistical data collected and disseminated by Eurostat can also be geo-referenced: in fact, statistics are always associated to at least a country and sometimes further down to a region, a smaller administrative unit, or even to a point. Within Eurostat, GISCO is responsible for meeting the European Commission's *geographical information* needs at three levels: the European Union, its member countries, and its regions. In addition to creating statistical and other thematic maps, GISCO manages a database of geographical information and provides related services to the Commission. Its database contains core geographical data covering the whole of Europe, such as administrative boundaries and statistical units, ports and airports, Digital Elevation Model, population distribution, Land Cover/Land Use information. Some data are available for download by the general public and may be used for non-commercial purposes (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.

Access to the datasets is additionally provided via the **GISCO data distribution REST API**, which includes data in different projections (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 endpoint access, multiple geo-formats, GISCO data distribution API and RDF.

Table 1 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 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 Europe's Linked Data-enabled one-stop-shop for open public sector information. It is not limited to a metadata registry but forms an entire ecosystem for fostering the manifestation, reuse and quality improvement of Open Data. The profile is based on Linked Data principles and the Resource Description Framework (RDF) vocabulary Data Catalogue Vocabulary (DCAT). It is designed to increase interoperability and allows the user to search for Open Data across multiple portals (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* is 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. Under API are available the portal core functionalities. This encompasses all the user can do with the web interface, including data harvesting operation. The information retrieved can thus be used by external code that calls the portal API.

SPARQL. The SPARQL Manager provides a graphical user interface (GUI) for sending user-defined queries to the Virtuoso SPARQL query engine. The powerful SPARQL Protocol and RDF Query Language are primarily aimed at professionals 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. Besides special formats distributions are not displayed within the platform but only linked (European Data Portal 2019). Thus, the “Download” function directs the user to the initial data provider where a dataset can be accessed and retrieved.

If a dataset distribution is supported by the geo-spatial visualization, an additional “**Open Geo-Visualisation**” option for data access is available. The visualization of geo-spatial data within the European Data Portal provides previewing functionality for spatial open data. The aim is to allow the user to assess whether a dataset meets specific requirements in terms of spatial and thematic coverage. Supported 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 EDP (European Data Portal)

* 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. Thematic data extracted from Eurostat and EDP

4.1.3 Thematic data extraction

This section offers the Table "Thematic data extracted from Eurostat and EDP" that introduces 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
Business demography and high growth enterprise by NACE rev.2 and NUTS 3 regions (bd_hgnace2_r3) 2017	The annual Business demography data collection covers variables that explain the characteristics and demography of the business population (Eurostat 2008).	File download via Database, Bulk Download, SPARQL endpoint	XLS, CSV, HTML, PC AXIS, SPSS, TSV, PDF, RDF
Population on 1 January by age group, sex and NUTS 3 region (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	File download via Database, Bulk Download, SPARQL endpoint	XLS, CSV, HTML, PC AXIS, SPSS, TSV, PDF, RDF

	demography and migration (Eurostat 2018).		
Area by NUTS 3 region (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.). Statistical Regions form the Nomenclature of territorial units for statistics (Eurostat 2019).	File download via Database, Bulk Download, SPARQL endpoint	XLS, CSV, HTML, PC AXIS, SPSS, TSV, PDF, RDF
NUTS-3 regions geometry for EU countries 2020	This dataset contains location data of regions within EU countries for level 3 of the Nomenclature of Territorial Units for Statistics (NUTS) .	GISCO Rest API, Statistical Atlas, SPARQL endpoint	SHP, JSON, TopoJSON, GeoJSON, GDB, SVC, RDF

Table 4. Thematic data extracted from Eurostat

Scenario 2

Bike-sharing and bicycle hire 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, SPARQL Search	SHP, WMS, RDF, XML, JSON
Distribution of ibex colonies 2015	This dataset contains data on the distribution of Alpine ibex colonies across Switzerland.	File download, SPARQL Search	SHP, WMS, WMTS, RDF, XML, JSON

Table 5. Thematic data extracted from EDP

4.2 Geopatial Linked 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 functionality of the BI tool. 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.

Before a user can build a view and analyze data, they must first connect Tableau to their data. Tableau supports connecting to a wide variety of data, stored in a variety of places. For example:

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 in a user's enterprise. Tableau interface offers a big list of sources available for connection, such as PostgreSQL, Spark SQLOracle, and others. If the user's file or database type is not listed, they might have the option of creating their own connection using **Other Databases (JDBC)**, **Other Databases (ODBC)**, a **Web Data Connector**, or a **Connector Plugin** built using the Tableau Connector SDK.

Connect to published data sources. Additionally, one might connect to public domain data available on the web such as U.S. Census Bureau information, or to a cloud database source, such as Google Analytics, Amazon Redshift, or Salesforce.

In *Tableau Desktop*, users can connect to the following spatial file types:

- 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)

With a Creator license in *Tableau Online* or *Tableau Server*, users can upload spatial file formats that only require one file (KML, GeoJSON, TopoJSON, Esri shapefiles packaged in a ".zip", and Esri File Geodatabases with the extension ".gdb", ".zip").

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 illustrated in a way of describing main processes 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, nether 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.

4.2.2 CARTO

CARTO fits the needs of organizations looking to design and deliver effective web maps without having to invest 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 (with the possibility to create 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 provides the user with more flexibility and more options to define the way in which the desired data should be uploaded.

Database Connectors

In addition to 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 very useful, but due to the fact that it has the file size and rows limits for files that can be imported, it might force users to split their files into smaller ones before importing. CARTO SQL API's COPY command allows importing big CSV files without their splitting much faster than the default import process.

CARTO Libraries

Users could use CARTO Python SDK or CARTOframes to upload data to the CARTO account. These two libraries provide methods that use 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 is a summary 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 scenarios is illustrated in the following 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

4.2.3 OmniSci

In this research, OmniSci Immerse was used. OmniSci Immerse is a browser-based, interactive data visualization client that works seamlessly with the OmniSci server-side technologies, OmniSciDB, and Render. Immerse uses an instantaneous, cross-filtering method that creates a sense of being at "one with the data". Exceptionally easy to use, Immerse generates SQL queries to the OmniSci backend on the fly, with clicks, not code. Advanced users can also generate stunning visual analytics from running hand-written SQL queries with Immerse if they desire (OmniSci Immerse | Interactive Visual ...).

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 scenarios is illustrated in the following table 11.

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

Visualization	**Success**	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

4.2.4 ArcGIS Insights

ArcGIS Insights is a data analytics workbench that offers spatial and nonspatial analysis capabilities to explore data and deliver powerful results. Insights is available in three deployment options:

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

One of the data sources that can be used in Insights is feature layer items. Feature layers can be created by:

Importing a dataset in Insights. Users can import their data from the following sources in Insights:

- Available hosted or registered feature layers from the user's content, groups, or organization. Users can publish a feature layer from ArcGIS Pro or ArcMap to be hosted in the ArcGIS organization. When publishing a hosted feature layer, data is copied from the user's data source to your 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 is a summary 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 scenarios is illustrated in the following table 13.

Process Name		**Status**	**Comments**
Import		**Success**	Datasets obtained from Eurostat and EDP were imported to ArcGIS Insights without inconveniences.
Join		**Fail**	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	**Success**	“Calculate” function for operations on data tables worked without inconveniences.
	Scenario 2	**Success**	“Spatial analysis” operations, in particular, “Find Nearest” worked without inconveniences.
Visualization		**Success**	Visualization framework resulted in created maps and histogram which demonstrate Scenario 1 and Scenario 2.

Table 13. Linked Data integration via thematic scenario implementation

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 connectors 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 not always ensure automated updates for data (in case of changes) that was integrated into BI software. For this reason, there should be a constant consideration and support by the interested parties running such connectors. Very often this kind of initiatives requires financial support, what 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 already 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 there are three major ways to retrieve geospatial data and use it within BI software. The traditional and still most common approach is to download or otherwise obtain data and load it into BI. Alternatively, one can connect to an API or OGC Web Services, such as a Web Feature Service, to stream data into a BI tool and thereby ensure the data remain up to date (Mai et al. 2019). Finally, Eurostat and EDP provide SPARQL endpoint that allows 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 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:

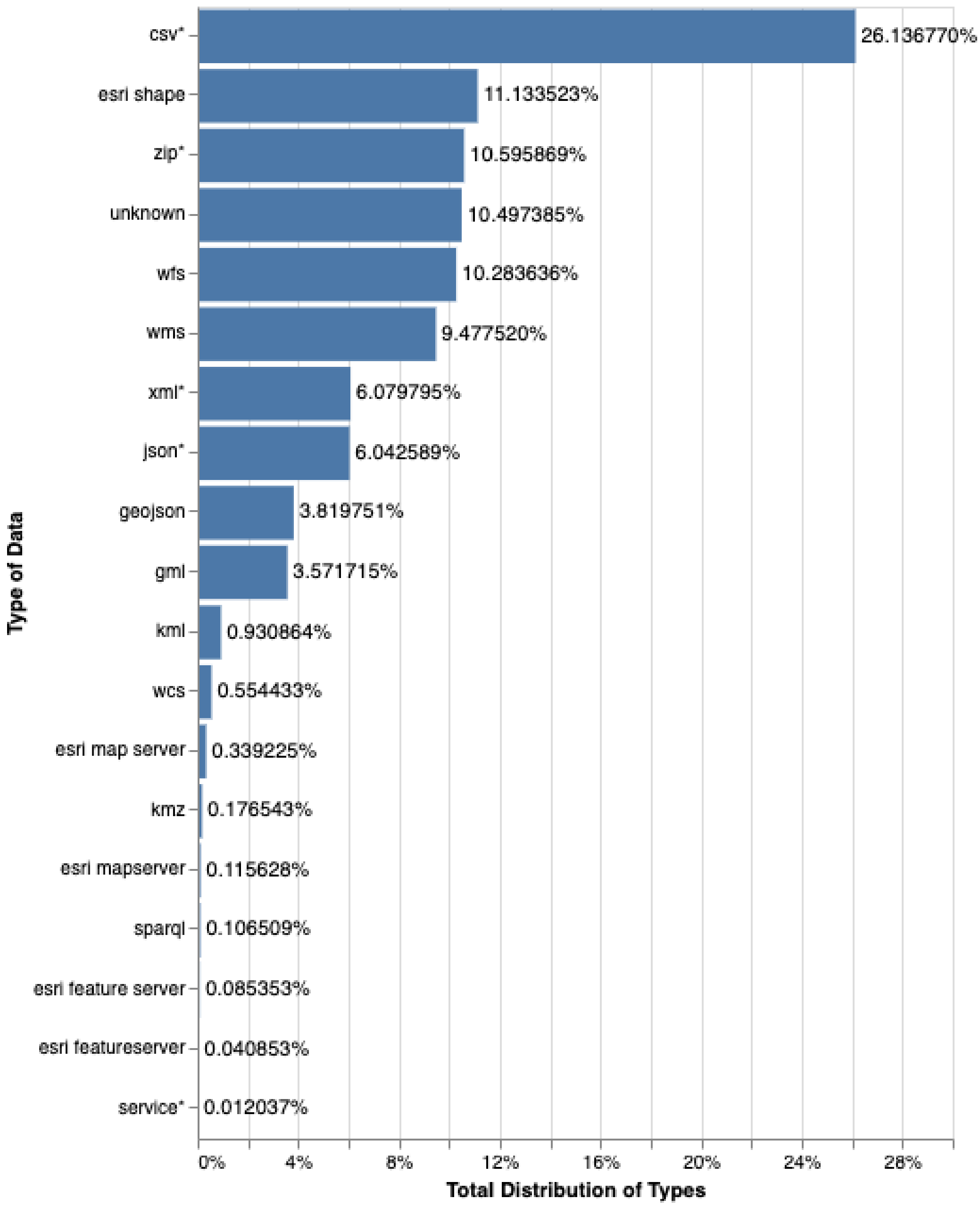


Figure 2. Data Types usage in the Datasets' Distributions

Based on this representable sample, the overall conclusion is the following: *the majority 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.

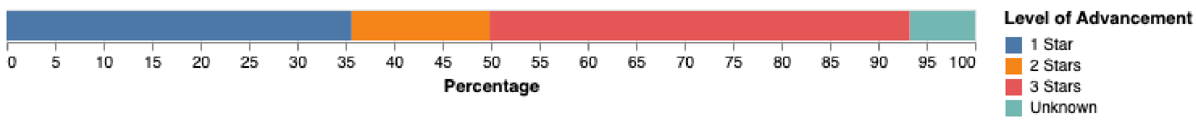


Figure 3. Data Types usage aggregated by the Five Star Open Data model (Berners-Lee 2006; Berners-Lee 5) applied to spatial data formats(McKeague et al. 2020)

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):

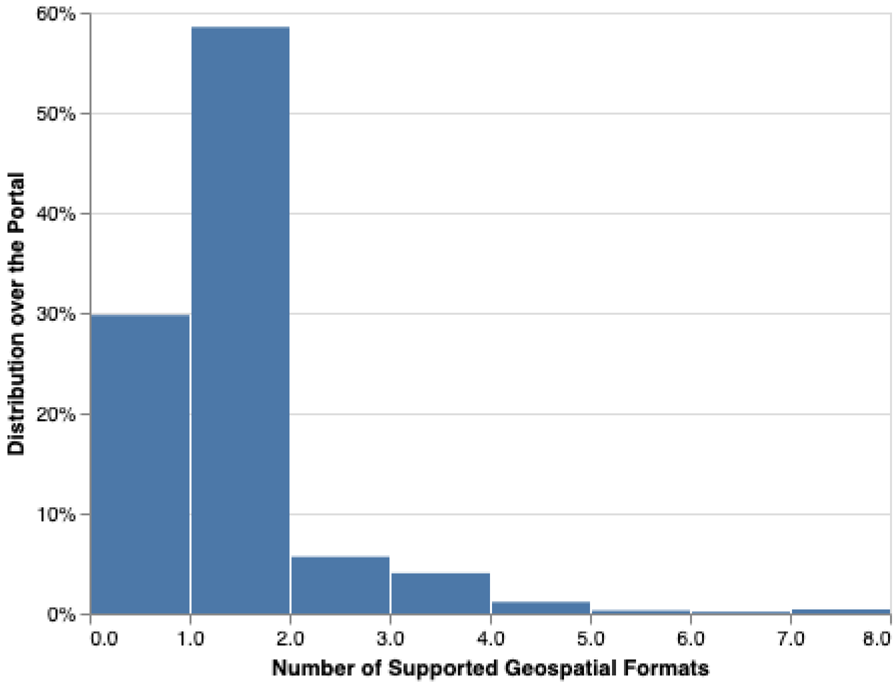


Figure 4. Distribution of the Number of Supported Geospatial Formats over datasets

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 requirements of The Five Star Open Data model. Moreover, Eurostat and EDP are positioning themselves as LOD providers. But in fact, while trying to integrate geospatial LOD into BI tools, a majority of 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 due to the fact that 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 5 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?”

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 (such as U.S. Census Bureau information), API and SQL and PostgreSQL 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 are in favor of 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 performance of the scenarios 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

Linked Data, and more generally, the idea of making semantically annotated raw data available on the Web, has taken information technologies by storm (Mai et al. 2019). A few conceptual models were developed to determine the main requirements for data to serve as LOD. The LOD concept has spread to the geospatial industry, boosting more profound research in this direction. What is more, 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 the Location Business Intelligence industry. Unfortunately, the geospatial data is still not ideally preserved from the LOD world to the BI solutions, and the gap remains uncovered. The connectors and adapters for BI and GIS software proposed by various initiatives are mainly temporal solutions with significant limits.

The conducted research emphasizes that the integration gap between geospatial Linked Data providers and Business Intelligence tools has to be a topic for further discussion and analysis. 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 maps, which lead to prospective use-cases, solutions and decisions across industries.

To avoid isolation, LOD providers have to focus on their potential users' needs and experiences continually. Maps are created automatically within web visualization frameworks in BI, often by non-cartographic experts. Due to this fact, it is highly reasonable to support this new developing domain of data scientists with a cartographic methodology of the LOD integration. In turn, the identified and analyzed challenges for integration make clear the need for a more in-depth review and improvement of already existing 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 various 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 (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 (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 (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 (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 (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 (Accessed: 5 August 2020).

Eurostat (2018) Eurostat Demo Metadata. Available at: 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 (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 (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= (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= (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 (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 (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 (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://ubikcan.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.

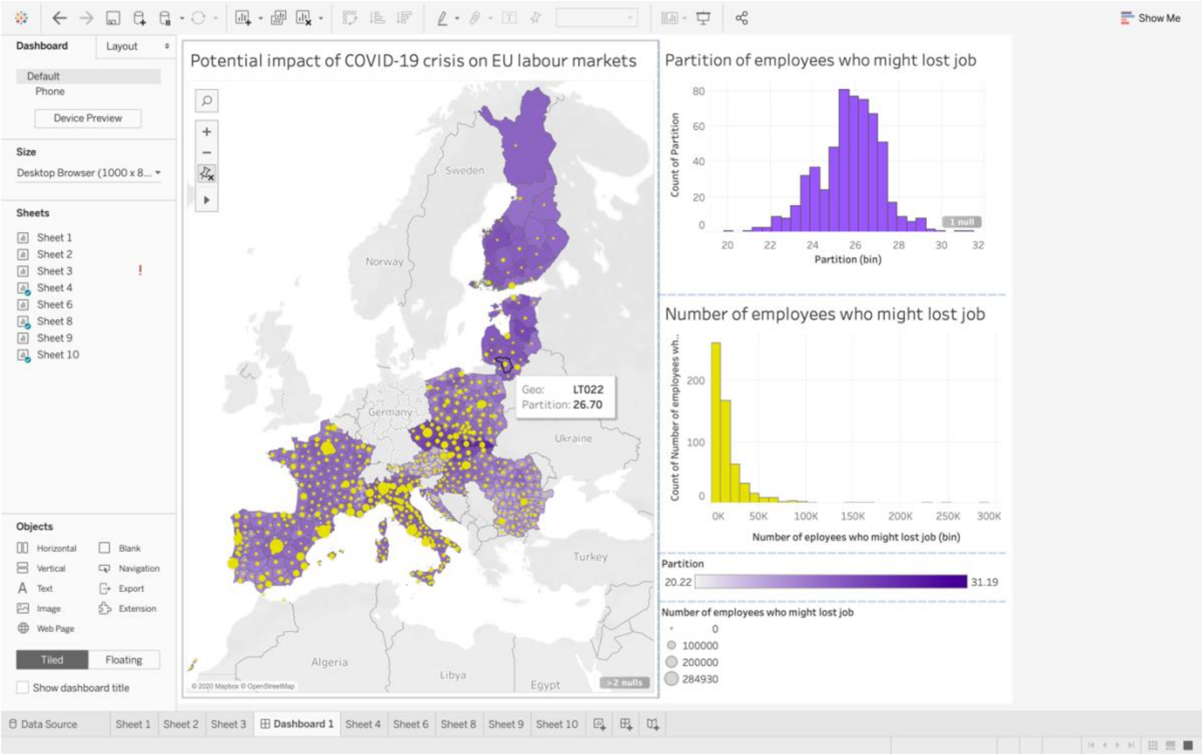
Ulrich, H. and Probst, G. J. B. (1991) 'Anleitung zum ganzheitlichen Denken und Handeln: ein Brevier für Führungskräfte'. Haupt.

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

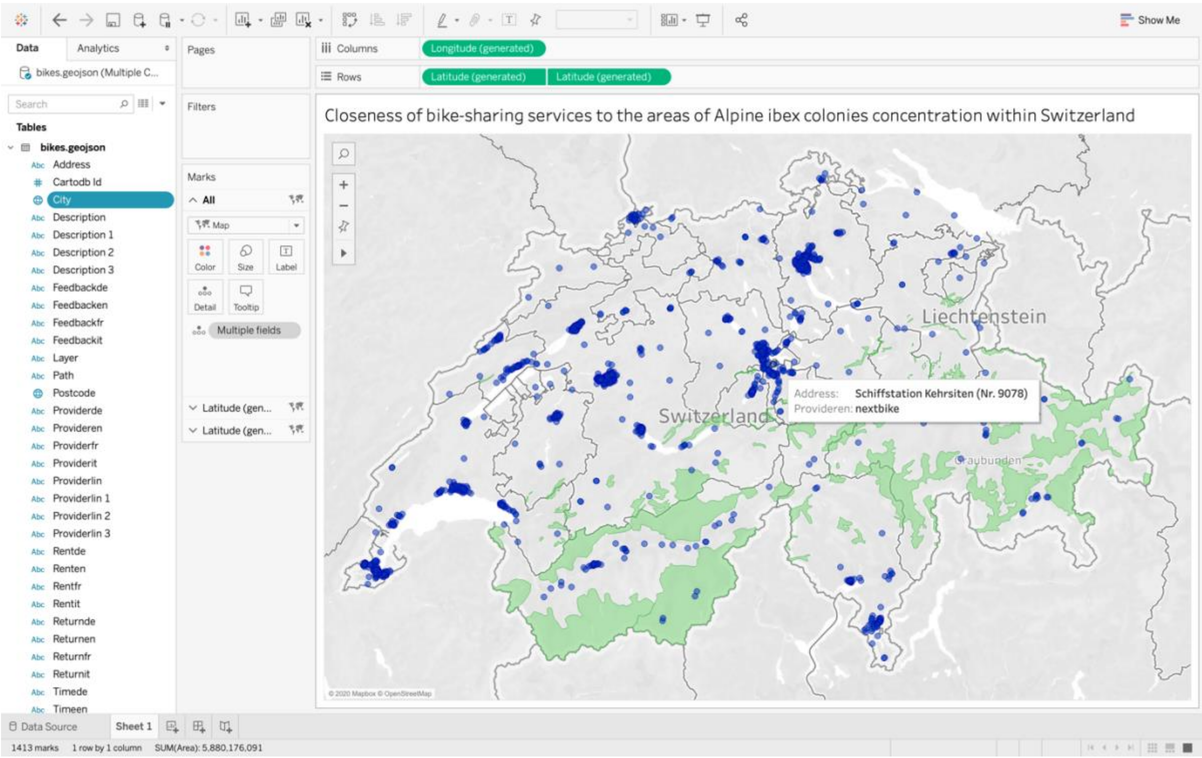
Watson, H. J. (2009) 'Tutorial: Business Intelligence – Past, Present, and Future', Communications of the Association for Information Systems. doi: 10.17705/1CAIS.02539.

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

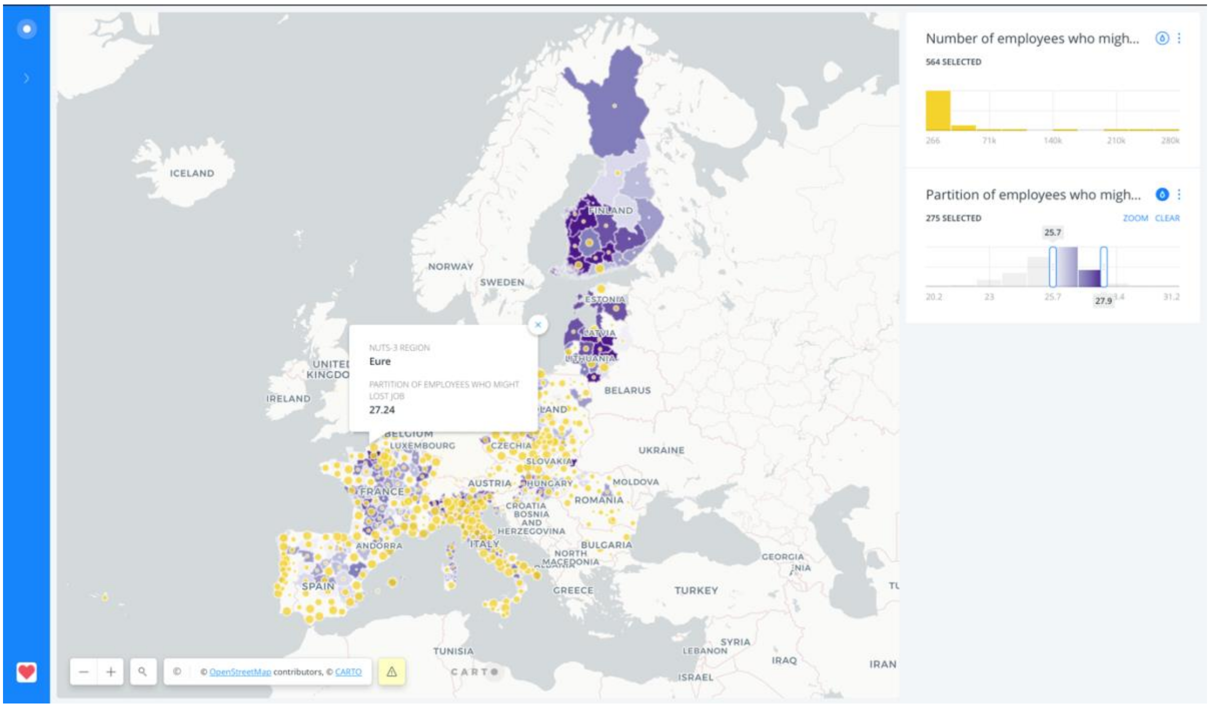


Potential impact of COVID-2019 crisis on EU labor markets | Tableau *

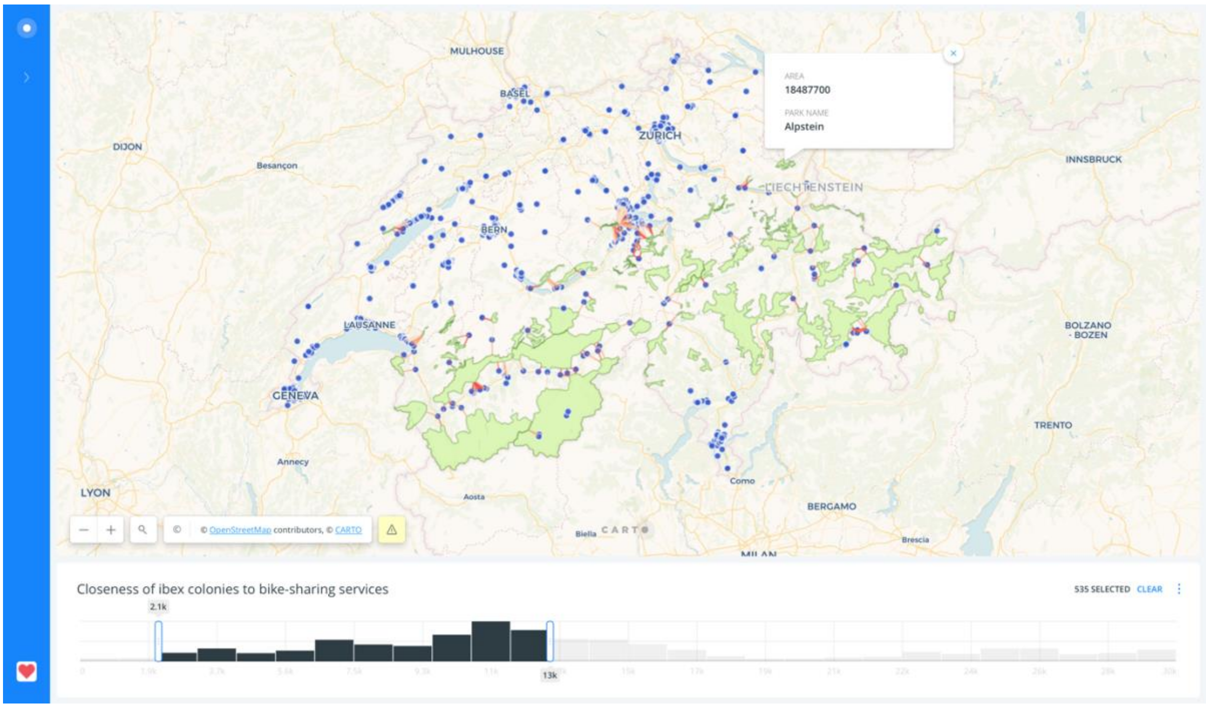


Closeness of bike-sharing services to the areas of Alpine ibex colonies concentration within Switzerland | Tableau *

ANNEX 2 – Scenario 1 & Scenario 2 visualizations | CARTO

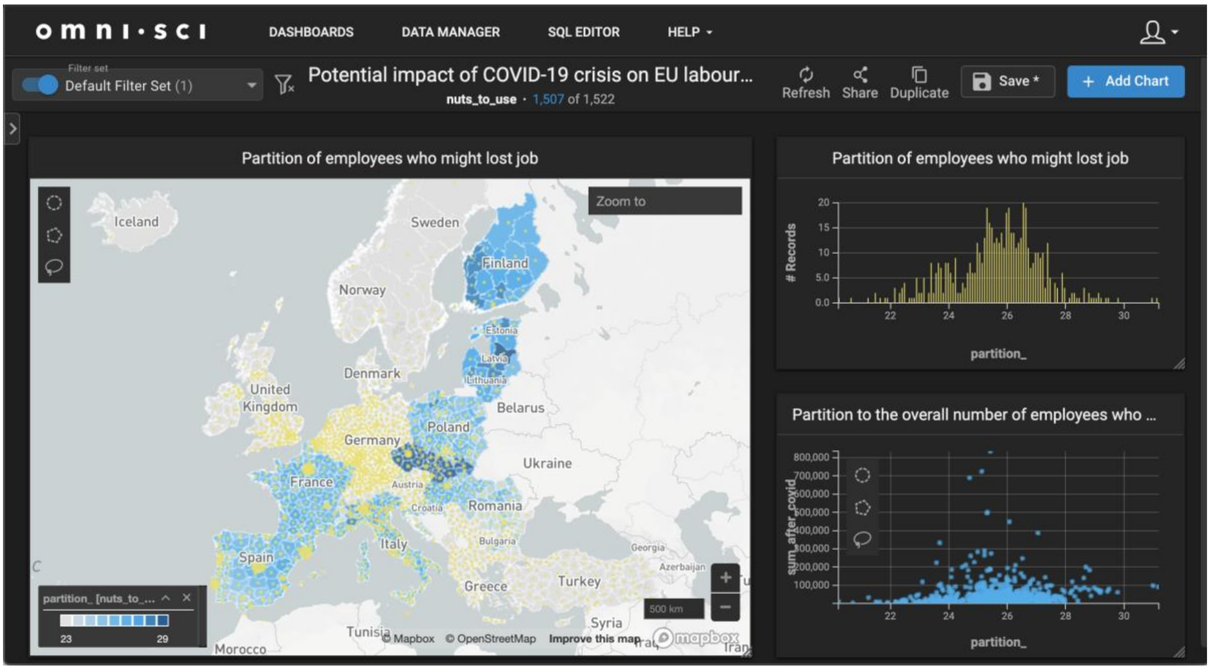


Potential impact of COVID-2019 crisis on EU labor markets | CARTO

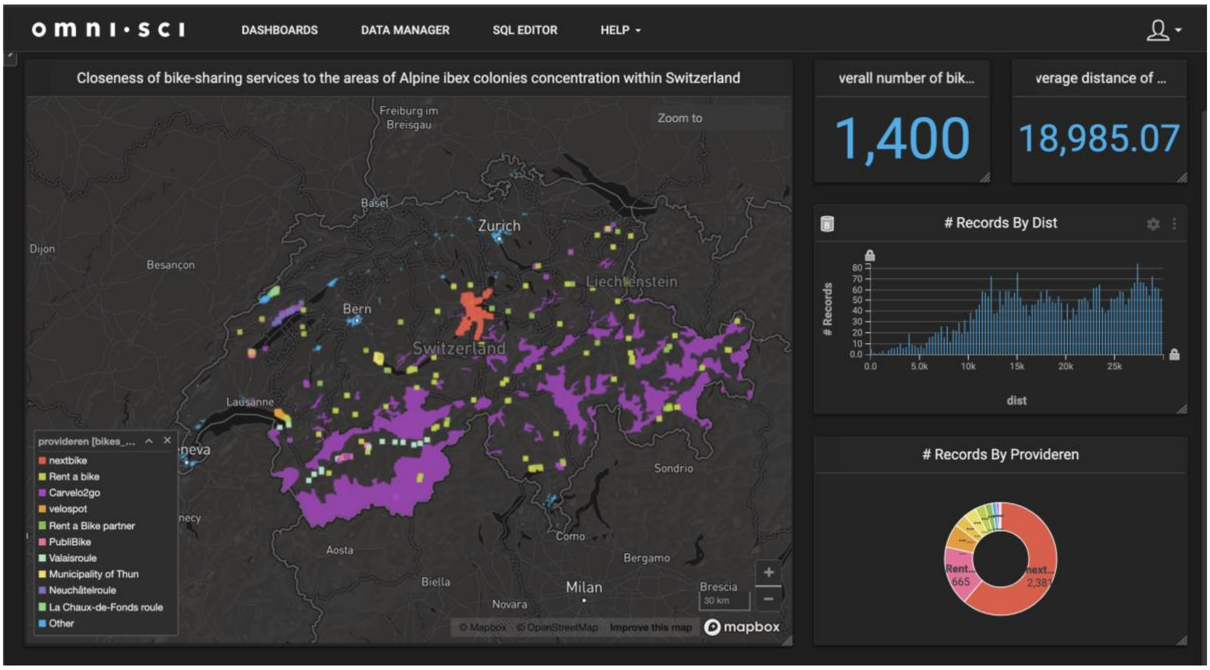


Closeness of bike-sharing services to the areas of Alpine ibex colonies concentration within Switzerland | CARTO

ANNEX 3 – Scenario 1 & Scenario 2 visualizations | OmniSci

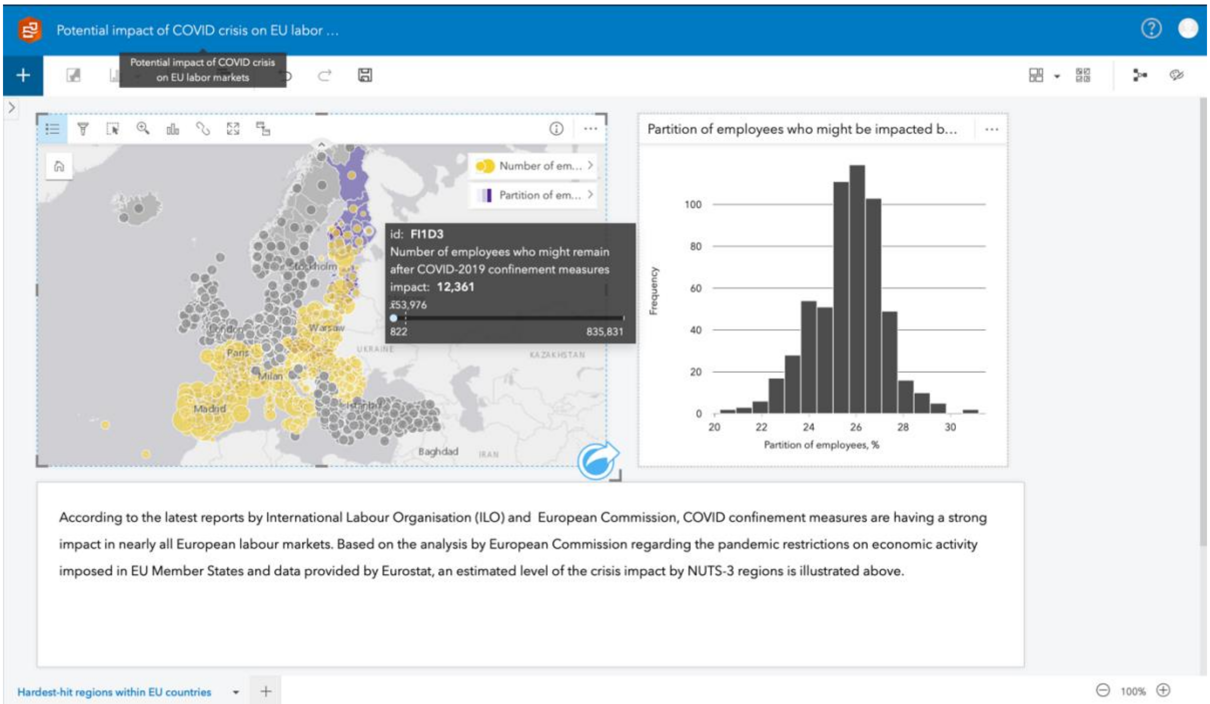


Potential impact of COVID-2019 crisis on EU labor markets | OmniSci *

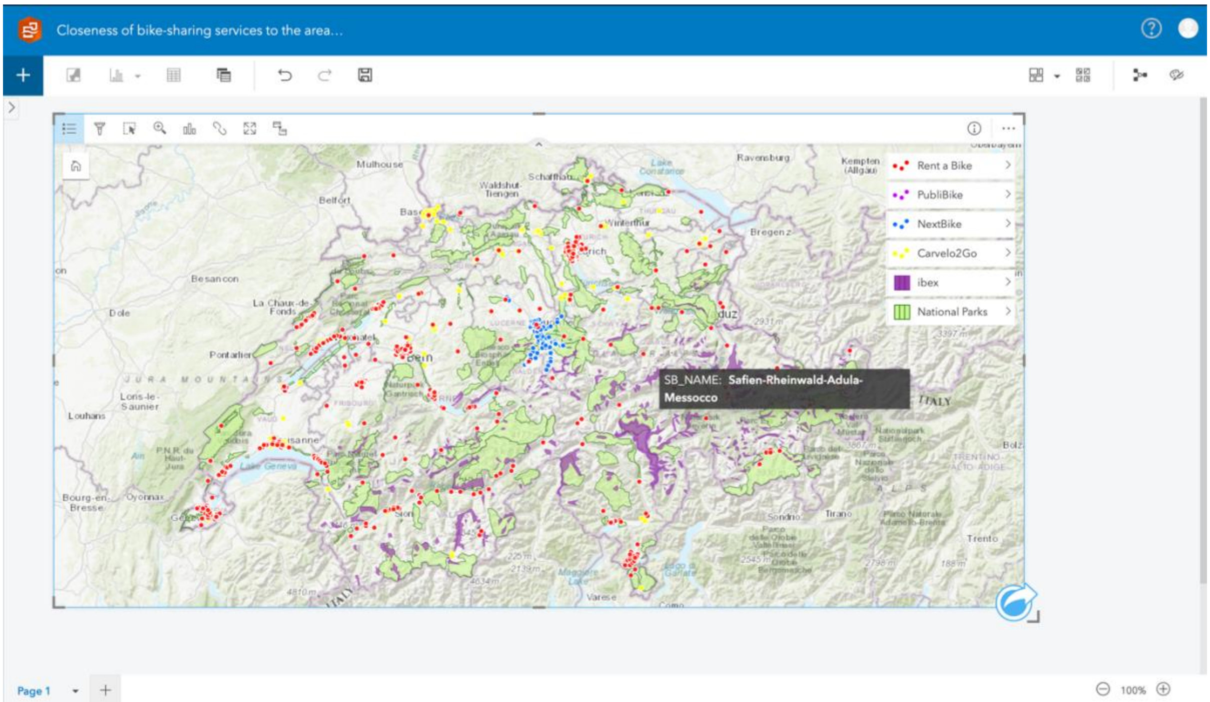


Closeness of bike-sharing services to the areas of Alpine ibex colonies concentration within Switzerland | OmniSci *

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