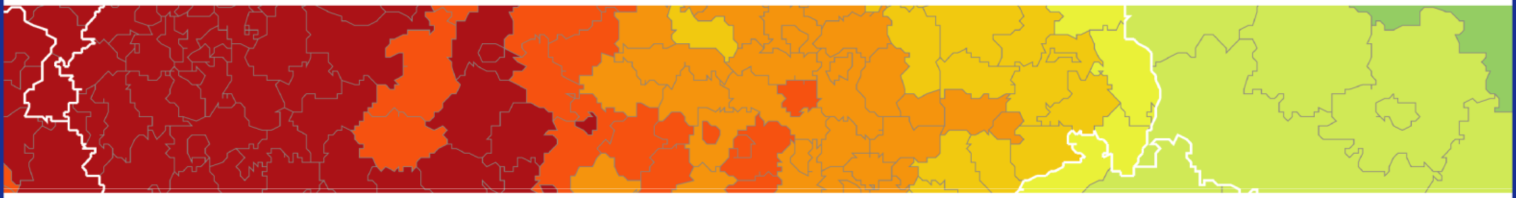
ESPON
EUROPEAN UNION
Co-financed by the European Regional Development Fund

Inspire policy making by territorial evidence



Potentials of big data for integrated territorial policy development in the European growth corridors (Big Data & EGC)

Targeted Analysis

Interim Report

Draft Version 08/10/2018

This Targeted analysis activity is conducted within the framework of the ESPON 2020 Cooperation Programme, partly financed by the European Regional Development Fund.

The ESPON EGTC is the Single Beneficiary of the ESPON 2020 Cooperation Programme. The Single Operation within the programme is implemented by the ESPON EGTC and co-financed by the European Regional Development Fund, the EU Member States and the Partner States, Iceland, Liechtenstein, Norway and Switzerland.

This delivery does not necessarily reflect the opinion of the members of the ESPON 2020 Monitoring Committee.

Authors
Helka Kalliomäki, Centre for Collaborative Research , University of Turku (Finland)
Ira Ahokas, Nicolas Balcom Raleigh, Finland Futures Research Centre, University of Turku (Finland)
Jukka Heikkonen, Pekko Lindblom, Paavo Nevalainen, The Department of Information Technology, University of Turku (Finland)
Siiri Silm, Anto Aasa, Mobility Lab, University of Tartu (Estonia)

Advisory Group
Nicolas Rossignol, ESPON EGTC (Luxembourg)
Antti Vasanen, Regional Council of Southwest Finland (Finland)
Dino Keljalic, Region Örebro (Sweden)
Liis Vahter, Ministry of Economic Affairs and Communications (Estonia)

Information on ESPON and its projects can be found on www.espon.eu.

The web site provides the possibility to download and examine the most recent documents produced by finalised and ongoing ESPON projects.

This delivery exists only in an electronic version.

Contact: info@espon.eu

Disclaimer:
This document is an Interim report.

The information contained herein is subject to change and does not commit the ESPON EGTC and the countries participating in the ESPON 2020 Cooperation Programme.

The final version of the report will be published as soon as approved.

Table of contents

List of Figures

List of Maps

List of Tables

Abbreviations

CNN	Convolutional Neural Network
EC	European Commission
ESIF	European Structural and Investment Funds
ERDF	European Regional Development Fund
ESPON	European Territorial Observatory Network
EU	European Union
FTA	Finnish Traffic Agency
GDP	Gross Domestic Product
LAM	Liikenteen Automaattinen Mittausasema (Finnish for Traffic Management System, see TMS)
MaaS	Mobility as a Service
NUTS	Nomenclature of Territorial Units for Statistics
NGZ	Northern Growth Zone
TEN-T	Trans-European Transport Network
TMS	Traffic Management System (see LAM)

1 Introduction

This interim report of the Big Data & EGC targeted analysis presents the second phase of the project with the focus on categorising and analysing new available data sources in the context of European growth corridors. In the previous inception report, the emphasis was on describing the conceptual and methodological framework of the study as well as presenting the public policy and stakeholder contexts and needs in the three countries of Finland, Estonia and Sweden. The aim of the Big Data & EGC project is to strengthen the knowledge-base for evidence-based and thus data-driven planning in the Northern Growth Zone (NGZ), which stretches from Oslo via Örebro to St Petersburg. Southwards is the North Sea-Baltic Corridor, also part of TEN-T core network and also covered by the analysis. The main objective of the project is to find and evaluate new available data sources for evidence-based policy making regarding the NGZ, and growth corridors more in general, and to research the potentials of big data and locations-based data mining to better inform comprehensive spatial policy in growth corridors. The especial interest is in finding new available datasets that can describe diverse flows and interactions along the corridors.

Big Data is essentially about generating insights from large datasets. In an early report about it by the consulting firm EMC, the concept was characterised in terms of the '3 Vs' -- Volume, Variety and Velocity (see Dietrich 2014). This list quickly grew to '7 Vs' with the additions of Veracity, Validity, Volatility, and Value (Kahn et al. 2014). In addition to referring to the size of the data, emphasis on variety, velocity and veracity means that data is collected faster and that there is more variation of data that can be tapped into. Veracity refers to the uncertainty of data. This has to do both with the quality of data, but also with the uncertainty of those dealing with the data of how accurate and complete this resource is (Giest 2017: 368). Irrespective of the many definitions of the term, big data describes broadly the volume and the complexity of the available data, as well as datasets that are too large for traditional processing systems and thus require new technologies (Provost & Fawcett 2013).

The perception of what is and is not big data varies by person and largely depends on the individual's institutional context. Countries with well-established national statistical systems might not consider their country's comprehensive population register as big data, whereas the term is used in this context in some other countries, as came out from the interviews. Big Data characteristics of the 7 Vs can be applied as a way to understand how a given study is or is not based on Big Data. For example, a study utilizing an already existing national population register would be more characterised as a big data study if it were combined with additional datasets, such as home ownership, registered vehicles, or electricity usage.

The data and methods used in the interim report are versatile. The methodology related to case studies is described more in detail in chapter 3.5. In analysing new available data sources in the study area, an extensive search was conducted in each country's key organisation's webpages to get an overall understanding about available datasets. Company-led (e.g. MaaS Global) and industry-led (e.g. ISO) practices to standardize data practices were also identified.

In addition, four interviews with six interviewees were made in diverse public organisations representing municipal, regional, and national perspectives to understand the various aspects related to utilising big data in the public sector. Furthermore, environmental scanning regarding data governance and big data driven futures was conducted.

The report is organised as follows. After the introduction, the second chapter presents a summary of data gaps and needs identified in the first phase of the targeted analysis to contextualise later chapters and their emphases. After that, typologies of new available data sources as well as brief case study descriptions are presented in the chapter 3. The fourth chapter presents a first overview of data governance and big data driven futures. Finally, conclusions are made about this second phase of the analysis.

2 Summary of data gaps and needs

In the interim report, an extensive overview was presented about the stakeholders' public policies and data related needs. Central themes were related to better governance and data driven innovations, and the different needs in the public and private sector. Especially the different processes and time-scales were discussed, as in private sector big data is typically used for short-term decision-making, whereas in public sector, democratic processes usually imply longer time horizons (e.g. Kim et al. 2014). However, interesting potentials concerning new forms of Big Data and public policy-making are partly related to the possibilities in utilising recent data to speed the production of provisional indicators for policy makers. Such potentials require increased monitoring capacities and access to real-time data to make short-term forecasts and revise long-term ones. The motivation for such efforts is to give decision-makers more up-to-date information about the present situation to enable more timely planning of development activities.

Big data has begun to derive insight in policy areas related, e.g., to transportation and infrastructure, smart mobility, economic growth, sustainable development, smart grid, energy efficiency, education and security (Kim et. al. 2014; Shi et al. 2017). In a recently published research roadmap for Europe (Cuquet & Fensel 2018), seven case studies were conducted in the key application areas - crisis informatics, culture, energy, environment, healthcare, maritime transportation and smart city - to understand the economic, legal, social, ethical and political aspects. Three large scale challenges were identified at the European level (2018: 75):

1. European policy may be unprepared for the positive and negative impacts of a rapid technological transition towards big data.
2. European policy may be poorly equipped for changes in the hegemony of big data.
3. European policy setting needs to be prepared to address both open, public data sources and closed, proprietary protections.

Central challenges in utilising big data in corridor development are related, e.g., to the missing spatial dimension from big data components (Shi et al. 2017) as well as to the fact that the majority of big data applications are designed for businesses and industries rather than for the government sector. Also, Santala (2016) addressed that one of the main challenges is that there is not enough big data suitable for territorial development or there is not enough knowledge on the available data. There are not many open data ecosystems where master data is available. There might be regional or national data available in many countries, but the data might not be similar enough to be combined for the purposes of territorial policy development. According to Giest (2017) one of the reasons for these challenges is that the data management structure is siloed and there is not institutional support and routine for sharing and collecting data. Companies have recently started to develop business models based on opening their data for the use of different networks.

In a workshop organised in the first phase of the project, stakeholders identified three themes as the most important policy dimensions related to corridor development that would benefit from big data: 1) infrastructure planning and transportation; 2) regional economic development, and; 3) land-use planning. The identified themes reflect the strategic objectives of major growth corridors in the study areas, as the key themes relate to transportation and smart mobility as well as developing the corridors as platforms for innovation. Even though big data utilisation was not recognised as such as being among the key strategic objectives, it was in the interviews said to be in-built in most corridor development practices as more efficient utilisation of data plays a central role in advancing smooth mobility and flows along the corridors as well as their development as uniform labour market areas and (digital) innovation platforms. Furthermore, the need for new sources of data describing the interactions and connectivity along the corridors was mentioned as evident in the practical attempts to improve the data-driven decision-making in corridor development.

The three themes identified in the needs analysis are taken as the basis for data categorisation concerning corridor development, however taking into account the different nature of the themes as well as their various interrelations. As the aim of the project is to identify datasets describing the interactions and flows along the growth corridors, the category of land-use presents a rather static framework for the flows and interactions happening in other identified categories. However, the land-use planning in the context of growth corridors essentially benefits from big data describing e.g. transportation flows and economic interactions hence highlighting its importance for the stakeholders and wider territorial development.

3 New available data sources in the study area

This chapter first presents a general picture of big data ecosystem and the different perspectives that should be taken into account in utilising big data. Next, developments related to open data are discussed in the study area, as well as the available data categories described in the context of corridor development. Finally, tentative frameworks of case studies are presented that go deeper in analysing the potentials of big data for corridor development.

3.1 Big Data Ecosystems

The organizational and network structures that facilitate the collection, maintenance, processing, exchange, and value-production related to data are described by some scholars and practitioners as data ecosystems. Like other systems, a data ecosystem features many elements in relation to each other. Many proposed definitions of a data ecosystem were carefully reviewed to create a meta-model of a generic data ecosystem. At a broad level, the parts of this metamodel includes actors, their roles, relationships, and resources. (Oliveira et al. 2018, 4.) Essentially, resources such as datasets, software, and analytical tools flow among the data ecosystem and are used by various individual and organizational actors in varying ways to achieve various goals. Taking a data ecosystem perspective is a key way for organizations to conceptualize how data of various types and forms flow within and outside of their organizations to produce value. It is also useful for exploring datasets viable in big data analysis.

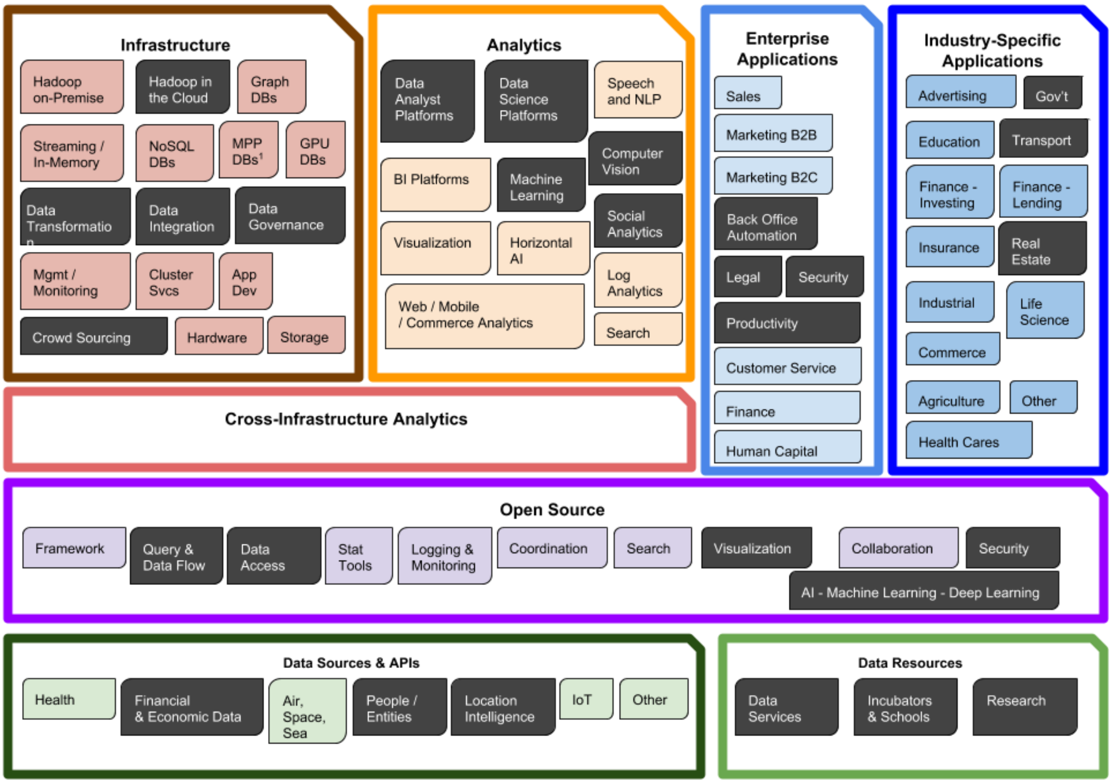


Figure 3.1. Categories and subcategories of big data actors and tools from the Big Data and Artificial Intelligence Landscape (Turck & Obayomi 2018) with the subcategories relevant to corridor development highlighted dark grey.

Another way to conceptualize a big data ecosystem is as a landscape of key players, companies, and toolsets which are dynamically interacting to develop big data capacities and applications for various use cases. The venture capital firm Firstmark has been tracking this big data landscape since 2012, producing the categorization of companies and big data tools for 2018 in Figure 1.

The landscape covers four main areas--infrastructure and analytics; applications and use-cases; open source communities; and sources and resources. Looking at the landscape from the perspective of growth corridor development some key areas of interest are cloud-based services, data integration, crowdsourcing; platforms for data science and analysis, machine learning, computer vision, and social analytics; internal applications for security, legal compliance, task automation, and productivity; big data applications designed for government, transportation, and real estate; open source tools for query & data flow, data access, data visualisation, and AI - Machine Learning - Deep Learning; data sources for Financial and Economic Data, People/Entities, and Location Intelligence; and resources such as data services, startup incubators and big data schools, and big data research centres. Additionally, the landscape indicates some of the new technology coming into the mix such as databases designed to run on a computer’s graphical processing unit and take advantage of parallel processing.

3.2 Approaches to new data sources

In the review of datasets which may be suitable as evidence for policy making, many datasets were encountered which were clearly outputs from some larger data systems (e.g. reports in the form of spreadsheets combining fields from multiple database tables). Rarely do organisations provide direct access to their primary *data systems.* In the ecosystem, this means most data that circulated is derived from somewhere else. In other words, within organizations that keep data, there is conventionally some ‘master set of databases’ and data products or flows emanating from and to them. For example, an organization’s IT department may run and closely guard the integrity of the consumer database while other departments in the organization are required to go through some gatekeeper system to update parts of this ‘master data’. While an increasing number of organizations are providing online access to their data via APIs, these APIs too are commonly designed to have limited read and write capacities and query operations. For example, Twitter throttles how many and what time-frame of tweets are available for querying via its API (Boyd & Crawford 2012).

Meanwhile, *data analysis*--which can include actions such as executing rudimentary SQL queries that produce reports to building machine learning models--is applied to data sets or source data to produce *data products*. There is a coupling between what systems for analysis emerge and what data is available, as the demands of the analysis shapes what data is considered valuable to keep or process. As organizations develop data science capacities, they demand more data from inside and outside their organizations to improve their tools for data analysis. For example, large datasets are often required to train neural networks--advanced data models produced using machine learning and deep learning (e.g. a neural network for identifying images)--and develop new systems. Furthermore, some analysis systems--especially the more advanced ones--themselves produce data in the form of simulation outputs which also can require storage and further processing.

These developments are tightly coupled with the rise in the deployment of networked “smart city” sensors and Internet of Things products. In addition to the adaptation of these sensor technologies, the development of automated and semi-automated driving systems are increasing the number of vehicles with advanced onboard sensor and computing assemblages. Geo-position tracking has also matured, with several classes of vehicles ranging from cars, city busses, trains, and commercial airplanes beaming their live locations to various computing systems. Combining these developments with trends in open data and data sharing, the present and near future availability of new, detailed, and real-time data is only expected to increase. In their report about IoT, NetGain Partnership concisely describe some of the challenges and opportunities arising from the widespread adaptation of networked sensors (Table 1):

Table 3.1. Challenges and Opportunities of the Internet of Things (Surman & Thorne 2016, 3)

Opportunities	**Challenges**
Savings and efficiency Improving public services Enabling citizens with data Democratizing product development Growing the movement for “open”	Erosion of privacy Surveillance on a global scale Inequity and reinforced social divides Threats to safety and security Centralization and monopolies

3.2.1 Forms of Big Data

The datasets reviewed can be categorized along many variables with ranges of attributes which describe general-level qualities. These ranges of qualities, summarized in Table 2, can help categorize any dataset, not just those ones most useful to corridor development.

One way to categorise datasets is by ***availability*** ranging from freely available *open data* (e.g. data published on a government-run portal) to *purchasable proprietary data* owned by a company to *unavailable proprietary data*. Another way to categorize datasets is by its **level of processing**--a spectrum of *rawness* to *highly processed.* Data streamed directly from a set of sensors is an example of *raw data* while data that has been evaluated and modified by data scientists in order to ensure its veracity and prepare it for use is an example of *highly processed data*. Similar to the level of processing is the **observational quality** of the data on a spectrum of *directly observed* to *synthetic data.* Synthetic Data is derived from some original aggregated data source to approximate finer grained details. Examples include taking average income data for a neighborhood and distributing it as average incomes for every building in the neighborhood or taking rough grain data about a vehicle’s travel path and velocity and approximating locations between recorded samples. (Grinberger & Felsenstein 2018, 109; Hwang et al. 2018, 135.)

Table 3.2. Typology of new data sources by variables and ranges of attributes

Data Categorization Variables	**Range of Attributes**
Availability	*Open Data* ←→ Exists but not easily Available ←→ *Purchasable Proprietary Data* ←→ *Unavailable Proprietary Data*
Level of Processing	*Raw (e.g. direct from sensors)←→ Pre-Processed Data ← → Processed Data← → Highly Processed Data*
Intended Audience	*Humans ←Programmers of Machines → Machines*
Observational Qualities	*Direct Observation ← → Synthetic*
Level-of Detail	*Fine-Grained (e.g. vehicle journeys) ← → Rolled-up (e.g. Indicators)*
Level of Structure	*Highly Structured ← → Semi-Structured ← → Unstructured*
Refresh Frequency	*Instant ←→ Milliseconds ←→ Daily ←→ Weekly ←→ Monthly ←→ Quarterly ←→ Annually ←→ Every few years*
Extraction Effort	*Requires much effort and resources to extract data ← → Requires little effort*
Clarity of Ownership	*Ownership is clear ← Perceptions of Ownership (e.g. "I own my Facebook data") → Ownership is unclear or shared by multiple actors*

Datasets can be categorized by their **intended audiences**, either *machine and human*. For example, data can depict an interaction between two machines (e.g. self-driving vehicle interacting with a Traffic System Management system), an interaction between a human and a machine (e.g. a set of search engine queries for a region), or an interaction between humans and humans (e.g. a social media post or email message). There are also classifications regarding the **level of structure** of a dataset, ranging from structured (meaning there are consistent fields and value types for every row of data) to semi-structured (meaning there inconsistencies between rows in how data is recorded) to completely unstructured (meaning data that is not even in tables). Structure can also refer to the type of database ranging from *relational* meaning the data is stored in a set of tables linked together by a primary key field (e.g. SQL) to non-relational (e.g. noSQL). Datasets can also be classified by the **refresh frequency**, ranging from *dynamic data* that is updated in nearly real time to *static data* that is updated less often. Some datasets are published in a form intended for human consumption (e.g. as tables in Excel files, or PDFs) and other datasets are published in a form intended for use by software developers (e.g. API endpoints or JSON value pairs). **Extraction Effort** refers to how much focused energy is required to extract the data from a given source (e.g. unstructured data in printed material). **Clarity of Ownership** refers to how obvious it is what entity owns the data. This seemingly straightforward concept becomes is blurred by user perceptions about their own data.

When looking for new datasets, additional broad categories can be helpful. Writing for *Forbes*, Kanellos (2016) describes five forms of data: *Big Data*, datasets that are gigantic due to the size and quantity of datasets; *Fast Data*, rapidly made forecasts from large datasets that produce 'good enough' projections and forecasts for the use-case at hand; *Dark Data*, data that exists but is not easy to access (e.g. security video) because it is un-networked, unstructured, or in unusual formats; *Lost Data*, this is data generated in industrial and commercial operations that is locked away in proprietary systems (e.g. an oil rig has 30,000 sensors but data from only 1% are ever utilized by data scientists); and *New Data,* the 'data we could get and want to get but aren't harvesting now'—data that will require new data products and services. This set of categories gives some insights into how to conceive of possibilities in data sets. *Emerging data*, data forms from new sources (e.g. sensors) still under development that may soon become mainstream, can be added to this list. An example of emerging data is LiDAR, sensors that use laser arrays to map 3D representations of objects in space.[1] In future, there may be several new and widespread types of sensors which are difficult to imagine today.

3.3 Review of Open Data in the Northern Growth Zone

Starting from the operational plan described in the inception report, a systematic search was conducted for open data portals of the nations that are included in the NGZ: Norway, Sweden, Finland, Estonia, and Russia. Additional open data portals for some of the major cities in the corridor were also identified. Additionally, the EU-level data portal ESSnet was reviewed.

The search produced following findings regarding the open data portals. All of these countries had at least one open data portal and the data portal was fairly mature and active. Many similarities were found among the portals. The portals generally have a large number of contributors sharing a variety of datasets. yet the veracity of these datasets is not always clear, and there is potential for duplication. Some interviewees also expressed that lack of standardisation and data management principles make the utilisation of open data challenging. It is typical that, e.g., all departments of the same organisation are not producing data in the same format or the data has some quality problems (for instance metadata missing). Furthermore, nearly all datasets encountered required some time dedicated to understanding its context and reviewing its metadata.

The online interfaces for accessing the data varies somewhat in their user experience designs and what dataset characteristics are forefronted for each data set (e.g. update frequency, data maintainer, kinds of data). The data portals supported search via a free-form search field, keyword tags, and data themes. The portals each used slightly different 'data themes' to categorize the data (Table 3). Sweden conveniently provides an easy-to-find link to download

[1] One identified smart city vendor sells a sensor product that integrates LiDAR object detection with a camera for use in Traffic Management Systems. This sensor can, for instance, determine how many people are standing at an intersection waiting to cross a street or the precise length of the vehicle that passes it on a highway.

all of the portal's metadata. Similarly, the Norwegian open data portal provides an API and RSS feed to access a list of all of its published datasets using the Data Catalog Interoperability Portal standard.[2] This practice, if it became more widespread across dataset portals, could facilitate more rapid comparisons among open data portals and efficient identification of viable datasets for policy questions.

Table 3.3. A comparison of top-level data categories used on five national open data portals with the categories used by all portals in bold and categories used by only one portal in italics

Norway	**Sweden**	**Finland**	**Estonia**	**Russia**
data.norge.no	oppnadata.se	opendata.fi	opendata.riik.ee	data.gov.ru
Top-Level Data Categories				

Agriculture, fisheries, forestry and food;
Economy and Finance;
Education, Culture and Sport;
Energy;
Environment;
Government and public sector
Health;
International themes;
Justice, justice system and general security;
Population and society;
Regions and cities;
Science and technology;
Transport
	Agriculture, fisheries, forestry and food;
	Download all metadata;
	Economy and Finance;
	Education, culture and sport;
	Energy;
	Environment;
	Government and the public sector;
	Health;
	International Affairs;
	Justice, justice and public security;
	Population and society;
	Regions and cities;
	Science and technology;
	Transport
		Agriculture, fisheries, forestry, food;
		Economy and finance;
		Education, culture and sport;
		Energy;
		Environment;
		Government, public sector;
		Health;
		International issues;
		Justice, legal system, public safety;
		Population and social conditions;
		Regions, cities
		Science and technology;
		Transport.
			Agriculture, fisheries, forestry;
			Culture;
			Economy and industry;
			Education and research;
			Employment;
			Environment;
			Finances and budgeting;
			Geography;
			Government services;
			Health;
			Open governance;
			Population and vital statistics;
			Social sphere;
			Transport
				Cartography;
				Construction;
				Culture;
				Economics;
				Education;
				Ecology;
				Electronics;
				Entertainment;
				Government;
				Health;
				Security;
				Sport;
				Tourism;
				Trade;

[2] For the Data Catalog Interoperability Standard, see http://spec.dataportals.org/.

				Transport *Weather*

Finland, Sweden, and Norway have nearly identical data themes. Estonia and Russia's portals vary from the nordic countries in their themes. For instance, Estonia group Education and Research together and give Culture its own category while the three Nordic countries group Education, Culture and Sport together. Russia offers a category for Cartography while the others do not. Russia is missing several of the categories found in the other portals, at least at its top level: Population; Agriculture, Fisheries, and Forestry; and Science and Technology, and Regions and Cities are missing. The Russian portal uses the following top-level themes which are not used in the others, but could be valuable for corridor-level analysis: Cartography, Construction, Electronics, Tourism, Trade, and Weather. All national open data portals in the five nations of the study area use the following top level data categories: Economy, Education, Government, Health, and Transport.

Furthermore, the found datasets had varying commitments to future updates. In some cases, a data set is essentially a tabular data report that is a year or two old. In other cases, a data set is an ICT developer-friendly API that is updated on some regular frequency ranging from milliseconds to monthly. If working on a project that requires a data model that is continually updated, these near-real-time APIs are likely more useful than the 'frozen in time' datasets. However, if working on a project with a specified time scope, the more traditional reports would be applicable, but still require some trust in future commitment if a policy decision is to be later verified. For most data scientists, having confidence there will be dependable and regular future updates to a data set may be a top requirement. However, in the open data portals, it can be difficult to determine how confident one can be that the dataset will be updated.

Several open data license schemes are supported by various advocates for those particular kinds.[3] All of the national open data portals have some indication about the open data license assigned to the datasets. There is a lot of variation, however, in which open data standards are used.

The datasets vary with many being very specific to a particular kind of data (e.g. number of pupils in a city's schools). In most cases, they appear to be 'reports' generated from various data systems, exported to Excel or .csv spreadsheets. Few datasets available in these portals are real-time or near real-time and accessible through an API endpoint or other developer-friendly technology. In other words, most of the datasets are outputs from data storage and

[3] A short list of 'Open Definition conformant licenses' is available at https://opendefinition.org/licenses/ (accessed 27 August 2018) and an inventory of open licenses at https://opensource.org/licenses/category (accessed 27 August 2018).

retrieval systems rather than datastreams from live data systems. This contributes to challenges in finding a dataset most appropriate to a specific policy question or context, which is why thorough metadata documentation and explanations of the purpose of the dataset are key to efficient data sharing.

An alternative to open data are non-open agreements among actors for **data interchange**. This form of data ecosystem takes the shape of private and secure networks for communicating among many independently managed databases. One example is TESTA, developed as part of the ISA2 program of the European commission.[4] It for example is used to run EUCARIS, an exchange of driver license and vehicle information among EU member states.

For example Statistics Finland and the Bank of Finland are experimenting with ‘nowcasts’, or forecasts based on more recent data than previously has been used to generate indicators. A nowcast is essentially the product of a data analysis process, yet as a forecast, it does not hold the veracity of rarified statistics.

3.4 Big Data categories in corridor development

The presented big data categorisation for corridor development is based on the themes identified in the stakeholders’ needs analysis, thus focusing the analysis of available datasets to the three themes relevant for growth corridors: 1) transportation and infrastructure planning; 2) regional economic development, and; 3) land-use planning (following major corridor development dimensions identified by Zonneveld and Trip 2003). The analysis focused especially on identifying new datasets that could describe the functionalities of corridors and provide new insights for data-driven policy-making.

The functionality of a growth corridor consists of spatial interactions, which, in the common practices of network analysis, “naturally form a network/graph, where each node is a location (or area) and each link is an interaction between two nodes (locations)” (Guo 2009: 1041). Whereas the term flow refers commonly to unidirectional flow from one location to another (e.g. a number of migrants), also called origin-destination flows (see, e.g., Patuelli & Arbia 2016; Spiekermann & Wegener 2007), the term interaction normally refers to two-way, reciprocal action between people or things (see, e.g., Spiekermann & Wegener 2007; Van Dijk 2012). In analysing spatial connectivities by using diverse datasets, flow data is also often used as a proxy for interaction thus assuming interaction based on connectivity. For the purposes of this targeted analysis, both flow and interaction data are used to describe functional growth corridors, although most interesting potential is seen to be related to analysing not just the

[4] See https://ec.europa.eu/isa2/solutions/testa_en (accessed 8 October 2018).

quantity but also the quality and nature - the content - of interaction. Therefore the interest is not only in describing the connectivities but also explaining them.

In the inception report for this project, a way to categorize datasets was proposed by the types of interactions they represent, followed by the relevant 'data objects' (e.g. actors, vehicles, buildings, infrastructure) involved in those interactions, followed by the kinds of data that is likely to exist for those objects. The types of interactions or flows specified were proposed by this project's key stakeholder and include physical, digital, administrative, organizational, intellectual, and social. To this table a column is now added for example datasets to help make this somewhat abstract approach more concrete (Table 4). The types of flows or interactions represent the different themes that were identified by the stakeholders in the early phase of the project, whereas the categorisation in figure 3.3 is based on the key policy categories arising from stakeholders' needs analysis.

Table 3.4. Categorisation of available data sources by types of interactions, data objects and data

Types of Flows or Interactions	**Relevant Data Objects**	**Data about Data Objects**	**Example Datasets**
Physical	Highway E18, Roads Intersecting E18, Land, Buildings, Parking Lots, Vehicles (cars, trucks, busses, etc), Freight/Cargo, Materials, Individuals	Transport Maps; Landuse Zoning Codes; Landuse Planning; Driver and Vehicle Registration; Port/trucking Manifests; Individual Level Temporal-Spatial Movement; Vehicle geo-spatial movement.	FTA Inspire Dataset (Finland transportation networks)[5] Trafi (Finland)[6] Call Detail Records (see Case 3) LAM Data (see Case 2) Eurostat Land Cover and Use
Digital Communication	Wired Infrastructure; Wireless Towers; Wireless Internet Access Points; Internet Connected Devices; Internet Connected Vehicles; Navigation Systems; Websites for companies, organizations, or governments in corridor.	Internet Peering Points; Internet Service Providers; Telecom Companies; Webmail Providers; Company and Organization IT Departments; Digital Infrastructure Details; Geolocation transponders; Geolocation satellites.	Google Trends FICIX Stats (live); NIX Statistics (live); Netnod IX Statistics; MSK-IX Traffic. Europe Media Monitor

[5] See https://www.avoindata.fi/data/fi/dataset/fta-inspire-transport-networks-theme-dataset

(Accessed 8 October 2018).

[6] Trafi data can be requested at https://asiointi.trafi.fi/en/web/asiointi/organisaatiot/tieliikenne/ajoneuvotietopalvelut (Accessed 8 October 2018).

Administrative	EU Administrative Entities, National Governments and Agencies, Provincial/Regional Governments, Municipalities, Park Services (national, regional, municipality), Neighbourhood Associations (if existing)	Open data from government and agencies; open data from civic-sector, housing-, environmental- (park), and economic services; news and blogs from all of the above.	Monthly Forecast for State Budget (Sweden) Quandl Alternative Data
Organizational	Large Businesses, SMEs, Non-profit Organizations, Societies, Public Sector	Business Registries (for business names and ownership); Real Estate Information (for office and facility locations); Business Websites; City planning.	PRH (Finland); Global Places by Factual

Intellectual	Universities; Research Institutions, Research Groups, Think tanks; Consortiums; Academic Writings (e.g. articles, essays, theses); Academic Networks (e.g. ResearchGate).	DBs of research funders; Bibliographic DBs; University websites; Academic Blogs.	EU Intereg Funding;
			EU Horizon Funding;
			Web of Science;
			Scopus; JSTOR;
			Conference Abstract Books.

Social	Homes; Inhabitants of Homes (e.g. families, couples, residents); Workplaces; Relationships among People (who interacts with who, and how).	Population Register (Nordics); Tax Office; Social Media (connections); Place-based Social Media.	Statistics Finland Microdata program; Geotagged Tweets (Twitter) Geotagged Instagram Photos Snapchat Groups

The relevance of any given dataset is strongly determined by the questions one is trying to answer with that data. In reviewing existing datasets it became clear that any policy-related question may have some relevant set of datasets ‘out there’. The first ideas of what questions to ask help surface some range of viable datasets, and after closer examination of those datasets, the question can be targeted to fit the kinds of insights the dataset is capable of answering. This aligns with the general data science research process proposed by Harvard professors Blitzstein and Pfister (2015) (Figure 3.2):

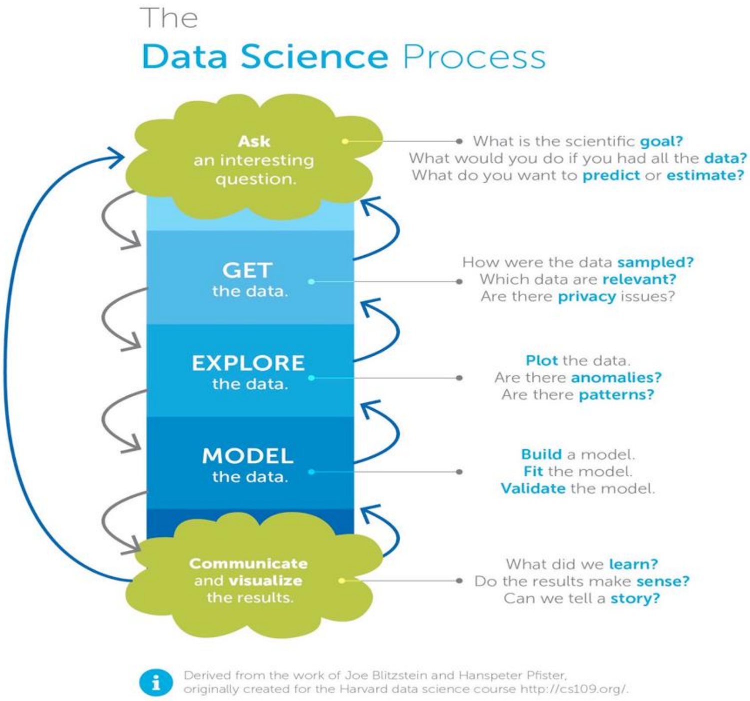
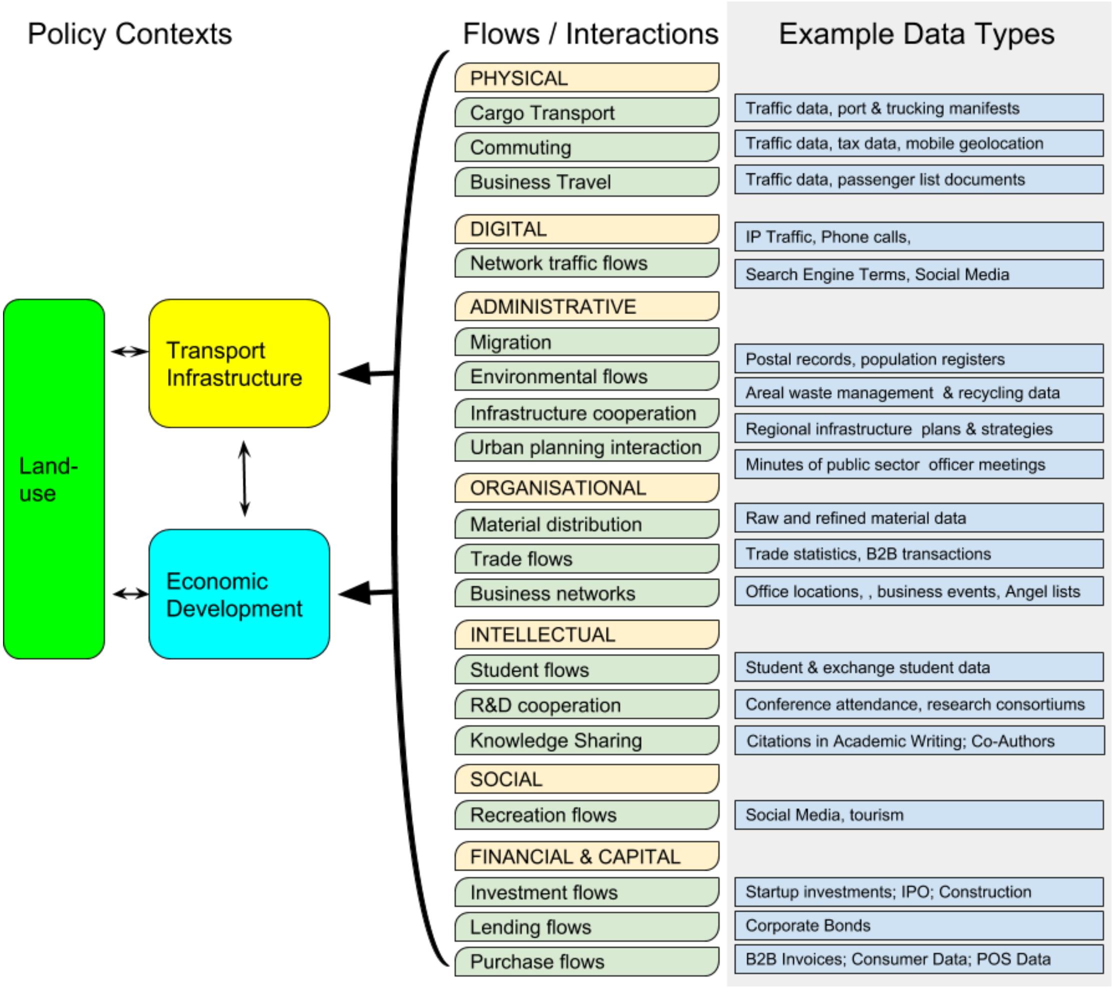

The
Data Science Process
Ask an interesting question.
What is the scientific goal?
What would you do if you had all the data?
What do you want to predict or estimate?
GET the data.
How were the data sampled?
Which data are relevant?
Are there privacy issues?
EXPLORE the data.
Plot the data.
Are there anomalies?
Are there patterns?
MODEL the data.
Build a model.
Fit the model.
Validate the model.
Communicate and visualize the results.
What did we learn?
Do the results make sense?
Can we tell a story?
Derived from the work of Joe Blitzstein and Hanspeter Pfister, originally created for the Harvard data science course http://cs109.org/.

Figure 3.2. Example of a Data Science Process, derived from the work of Harvard professors Joe Blitzstein and Hanspeter Pfister for their Harvard data science course (see http://cs109.org/).

Because the context of questioning is so strongly linked to what data is potentially useful to a given production of ‘evidence’ for policy making, it follows that data can be categorized based on the flows which are most significant to various policy contexts. How the three policy contexts identified by the stakeholders of this research are comprised by various flows which can be described by data is presented in the figure 2.

Figure 3.3: Data that represents key flows and interactions of various policy contexts.



Flows / Interactions	Example Data Types
PHYSICAL	
Cargo Transport	Traffic data, port & trucking manifests
Commuting	Traffic data, tax data, mobile geolocation
Business Travel	Traffic data, passenger list documents
DIGITAL	IP Traffic, Phone calls,
Network traffic flows	Search Engine Terms, Social Media
ADMINISTRATIVE	
Migration	Postal records, population registers
Environmental flows	Areal waste management & recycling data
Infrastructure cooperation	Regional infrastructure plans & strategies
Urban planning interaction	Minutes of public sector officer meetings
ORGANISATIONAL	
Material distribution	Raw and refined material data
Trade flows	Trade statistics, B2B transactions
Business networks	Office locations, , business events, Angel lists
INTELLECTUAL	
Student flows	Student & exchange student data
R&D cooperation	Conference attendance, research consortiums
Knowledge Sharing	Citations in Academic Writing; Co-Authors
SOCIAL	
Recreation flows	Social Media, tourism
FINANCIAL & CAPITAL	
Investment flows	Startup investments; IPO; Construction
Lending flows	Corporate Bonds
Purchase flows	B2B Invoices; Consumer Data; POS Data

In this model, land-use policy is inter-related with transport infrastructure policy and economic development polcy. While land use tends to describe physical, more static environment, planning activity, infrastructure and economic functions are all dependent on landuse policy. Landuse policy in turn requires economic data such as real estate estimates, income levels, and business demand forecasts. It is easier to comprehend the idea of flows and interactions as a function of transportation and economic activity. The built environment, instead, is planned in relation to the networks in which they are situated. However, for the purposes of this research, it may be clearer to look for interaction and flows in landuse policy context at the governance level—the administrators, planners, business stakeholders, and community interests. Landuse policy benefits greatly from taking a corridor-wide perspective as it places local activity in relation to a wider network. It is difficult to find data that sufficiently describes cooperation among municipalities. One could look, for example, in plans and workshop materials for lists of participants and begin to map these relationships. The extraction effort is quite high to retrieve such data.

Seaking a temporal dimension from unlikely datasets can also be one way to generate new insights. The population register is often treated as a static snapshot of people in place. But such a dataset can describe the internal migration of people as they move from home to home over their lifetimes. Postal records may also serve this purpose in nation states with no population registers. When seeking temporal insights from the data, it is able to describe flows of people as the sequences of places individuals have lived become evident.

3.5 Case studies examining new data sources

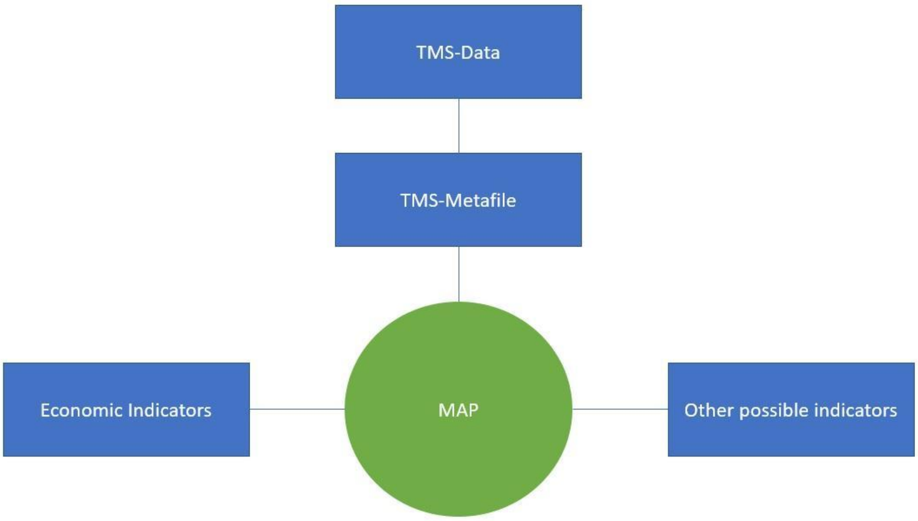
3.5.1 Case 1: Traffic measurement points

Case study 1 focuses on automated traffic measurement. Moreover, it elaborate the possibilities with selected data sets to observe the predictability and possible dependencies with econometrics.

E18 route traffic analysis

The dataset is gathered by the Finnish Traffic Agency (FTA), and the sample used in the analysis comprises E18 route (Turku-Helsinki-Lappeenranta) traffic data over the period 2010-2017. Vehicle speeds, vehicle categories (7 of them), vehicle speed and direction are recorded. There are vehicle observations in the database. Fig. 3.4 describes the overall system, and how the visualization connects the data to the map. TMS (Traffic Measurement System) data has a descriptive metafile attached to it, that has pinpoint location data of these measurement stations. Economic indicators are municipal datasets from Statistics Finland, which display indicators related to economic activity.

Figure 3.4: The overall system description.



The prediction problem uses the observed traffic waves (see figures 3.5 and 3.6). The traffic state *x* is observed every 6 minutes, and it is compared to the existing municipal economic data *y*. A prediction model is created using a Convolutional Neural Network (CNN), which is spatially restricted to the existing municipality areas and constrained to operate on the traffic flow waves described below.

Figure 3.5: Traffic along E18 at 10 AM.

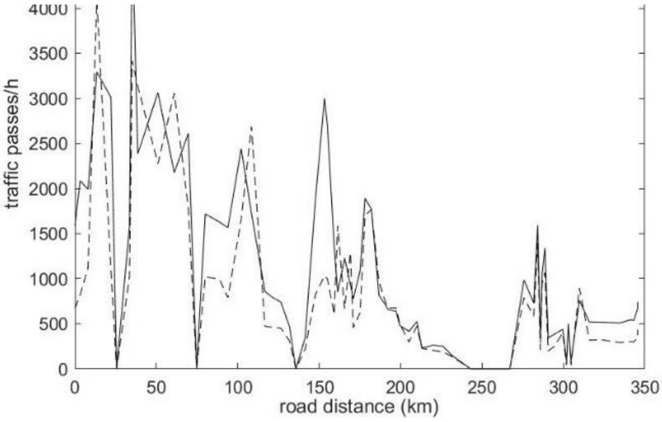
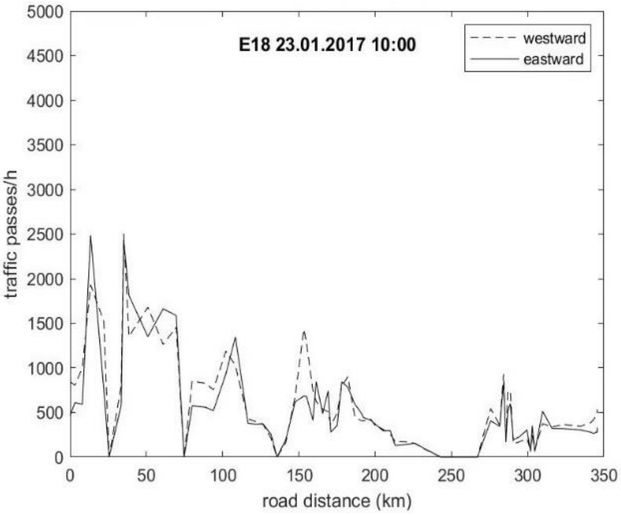


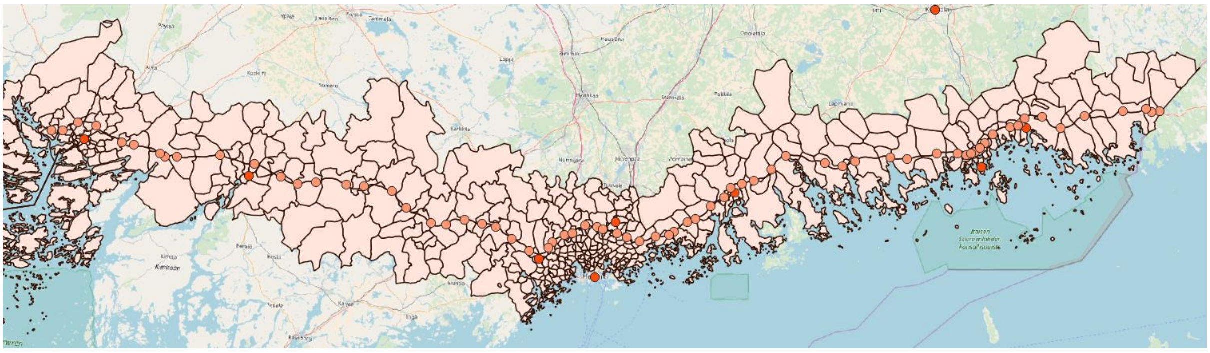
Figure 3.6: Traffic along E18 at 5 PM.



The traffic flow model over time *t* describes the traffic waves. These are similar to standing waves in physics, and their parameterization allows building a predictive model for economical descriptors, see Figure 3.6. Some subtleties are included because of the spatial dependencies of traffic measurement points and the municipality areas. Each measurement point has to be associated to a small set of nearby municipal or postal areas, where the economic data can be differentiated.

Map 3.1 presents a preliminary visualization of the E-18 Route with the ZIP-code areas in selected scope. E-18 represents the growth corridor in Southern-Finland, and is depicted here from the Finnish-Russian border to the western coast Turku/Naantali harbour. The visualization connects the TMS-points from the Digitraffic-database and spatial location data. The traffic dynamics (x((t)) may prove to be viable predicting economic descriptors (y) in the future analysis.

Map 3.1: Finnish section of the Route E-18 with selected ZIP-code areas and TMS-points.



3.5.2 Case 2: Project networks

The purpose of the case study 2 is to analyse the opportunities and challenges of open data on joint EU project funding, combined with the regional economy. As the aim of this overall project is to explore various data sources and their uses, it should be mentioned that Case 2 does not aim to create a product or service that is ready to use, but to map potential direct or indirect usage possibilities of big data, and report the first outputs of the selected tests. This is done via interaction matrix, which is currently being constructed. This methodology may shed more light to both existing project partnerships and their implications to regional economy, but also hidden possibilities in the European perspective as well to these matters. The interaction-oriented approach to economic analysis will hopefully prove beneficial not only for regional developers, but also for regional policy- and decision-makers.

European focus

The framework of our project provides an opportunity to look at the regional economic factors from fresh network-oriented aspect. This is especially important in the context of big data, because traditional methodologies which are based on statistical documentation remain relatively far from the current world of open, rapidly updated or even near real-time data sources. Moreover, big data analysis is considered as a very likely source to generate also metadata- or other indirect new information, e.g. when looking at the connections of different datasets.

Despite the fact that the European Union is closely monitoring the effects of its various financial instruments as the ERDF:s and ESIF:s in this case, the indirect and procedural nature of the economic effects in particular, poses significant challenges for a comprehensive assessment of project efficiency and performance. In addition to the lacking criteria for quantitative monitoring and evaluation, more qualitative indicators should be developed. Authorities that are responsible for project execution underline that qualitative evaluation criteria and indicators would often take into account the nature of the project activity more coherently. A vast variety of reporting methods and data sources has been mapped in order to produce even more accurate understanding towards qualitative indicators used, alongside the quantitative economic regional data.

Process of the case study and methodology

The data-acquisition phase of this case is in its later stages, and the available data has been extensively mapped. The network is constructed from three EU-programmes, from the 2014-2020 programming period, including Interreg Cross-border, Interreg Transnational & Interreg Networking. In the first stage a key goal is to connect three member states, Finland (FI), Estonia (EE), and Sweden (SE). These member-states are partners in over 300 projects, and the availability of additional economical datasets and their similarity will be explored to apply to our model in the later stages. The aim in the first stage of the model is to generate a simple interaction matrix, and visualize it. Second stage is to weight the interactions based on the amount of funding for each project.

The project data is very useful, as it can be analysed for interactions on multiple levels, starting from country (NUTS) to municipal. At present the case study work focuses on visualizing interactions between municipalities. After the visualization the interaction matrix will be utilized to connect into local data sets, starting from Finnish statistical data gathered on the municipal level. This connection supports the study of the dataset in a more statistical manner.

Municipality areas and the econometric indicators are used to connect the traffic data to economic activity around the traffic areas. Municipal data is available for:

- employment (2012, 2013, 2015),
- education levels (2012, 2013, 2014 & 2016),
- tax-income (uncertain),
- disposable income (2012 – 2016),
- zoning and construction (2013 - 2016) and
- GDP of the municipalities (2000-2015).

This data is ordered either by year or quarterly, making the comparison more of a periodical outlook of the trends in traffic and economic activity. This data is compiled by statistics Finland, but unfortunately especially the Gross Domestic Product (GDP) and taxation data has been difficult to acquire for this moment. For the GDP there is a comprehensive data set but it covers larger administrational areas called regional municipalities. In practice, the interaction matrix has zn amount of partners, the dimensions of the matrix are zn * zn, and every cell inside the matrix has the number of interactions the two municipalities have in the data. The number of interactions that can be pointed towards a pair of municipalities can then be used to connect the two with programmable tool sets available for python and R software, for instance. The same data can also be converted to interactions per municipality, when the municipalities and the number of their interaction to economic factors and their development can be compared.

Mapping of the data sources

Some examples of the evaluated ERDF and ESIF open data sources:

- **Keep.eu**
- Available data and visualizations on Interreg, Interreg IPA CBC and ENI CBC.
- Provides partner data, currently used in the case study to map the interaction matrix.
- **European Structural and Investment Funds**: ESIF Open Data Platform.
- Database by the European Commission, categorisation by country, theme and funding instrument.
- Available programming period 2014-2020.
- **Socrata Open Data API.**
- Open interface targeted for developers' further use.
- Operated by a private Socrata organisation.

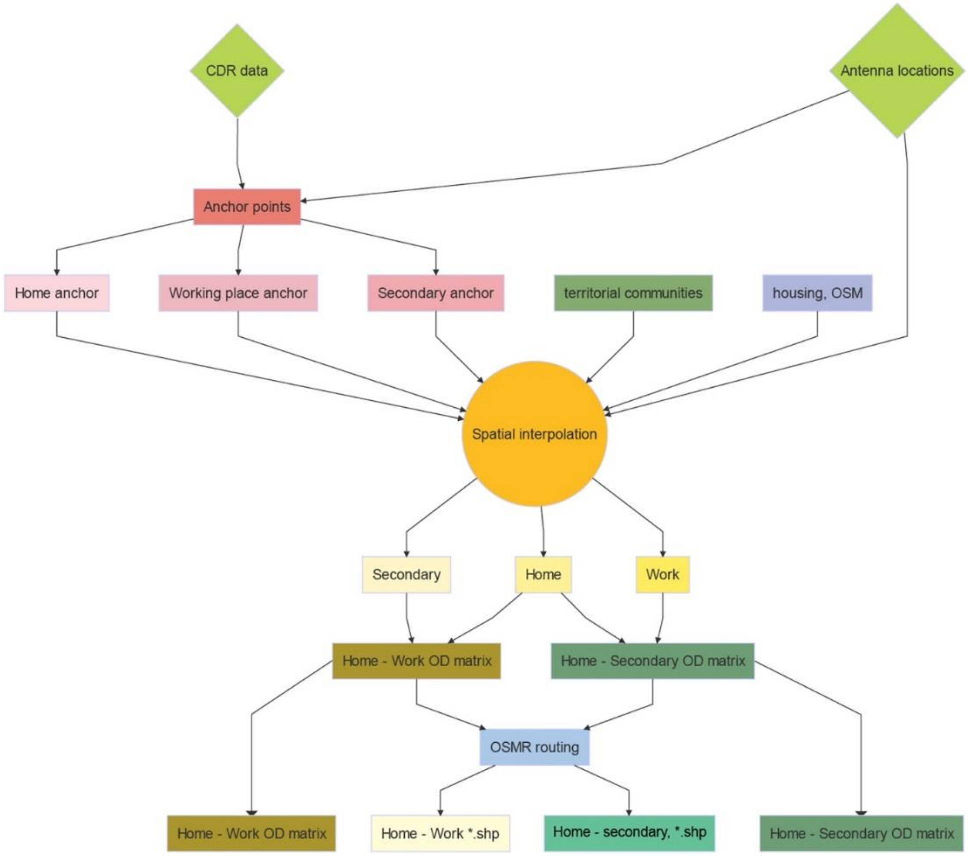
Examples of other accessible big data sources

- *Estonia:* Statistics Estonia and PX WEB interface.
- *Sweden:* Open Statistical database operated by the SCB and the Swedish Transport Administration.
- *Finland:* Statistics Finland's Paavo-database open data by postal code area, and Finnish Transport Agency's Digitraffic-service.

3.5.3 Case 3: Mobile positioning data

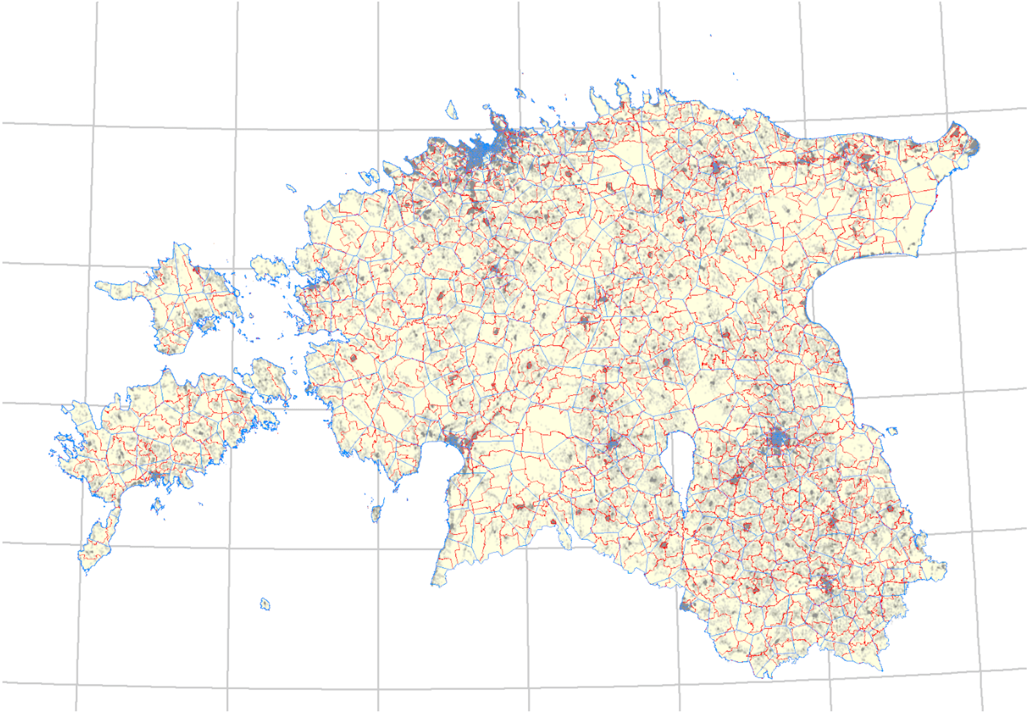
In the case study 3 an everyday mobility database is developed in collaboration with the Estonian stakeholder. Study period extends from January 2017 to April 2018. Dataset will be prepared for every separate month. Based on earlier experience and preliminary analysis, the spatial granularity of the database is planned to remain on level of territorial communities (Figure 3.7). As the mobile network is very dense in Tallinn the spatial accuracy can probably be higher. Completed datasets will be in machine readable format (*.csv) and available directly from Mobility Lab web service.

Figure 3.7: Workflow of case 3.



The main data source for the current mobility database is passive mobile positioning. Passive mobile positioning data is automatically stored in the memory files of mobile operators for call activities or movements of handsets in the network. The current study uses the database of the locations of call activities in network cells: the location, time and random unique ID. Passive mobile positioning data is normally collected with the precision of network cells. For the collection of passive mobile positioning data, mobile operators can aggregate anonymous geographical data from log files, ultimately not violating personal identity and privacy, and researchers can use it in surveys for scientific purposes. Passive mobile positioning data has been used in many transport and urban studies. The information on the crowdedness of network cells is used for research, planning or traffic management. Due to privacy issues, the database is anonymous and does not contain any backtraceable personal information about the user of the phone.

Map 3.2: Study area and spatial accuracy (grey – housing; red – territorial communities; blue – theoretical coverage areas of mobile antennas).



On average, approximately 420,000 active respondents per month were noted whose home anchor points could be defined using the anchor point model (with a varying maximum of approximately ±10%). Within this study, the meaningful places for a respondent originate from the use of the anchor point model, which was developed by the Mobility Lab at the University of Tartu and Positium LBS. The model helps to assign locations that are meaningful to mobile phone users for every calendar month (Table 3.5), including the most likely home and work locations based on the respondent's calling activities over time. This study uses home and multifunctional anchor points. The former is defined as an everyday anchor point that is the probable location of a respondent's home, and the latter is defined as an everyday anchor point where the home and work-time locations are positioned at the same base station and therefore cannot be separately identified (Ahas et al. 2010). These anchor points allow us to investigate meaningful locations and people's daily activity spaces as well as more permanent moves such as changes of residence.

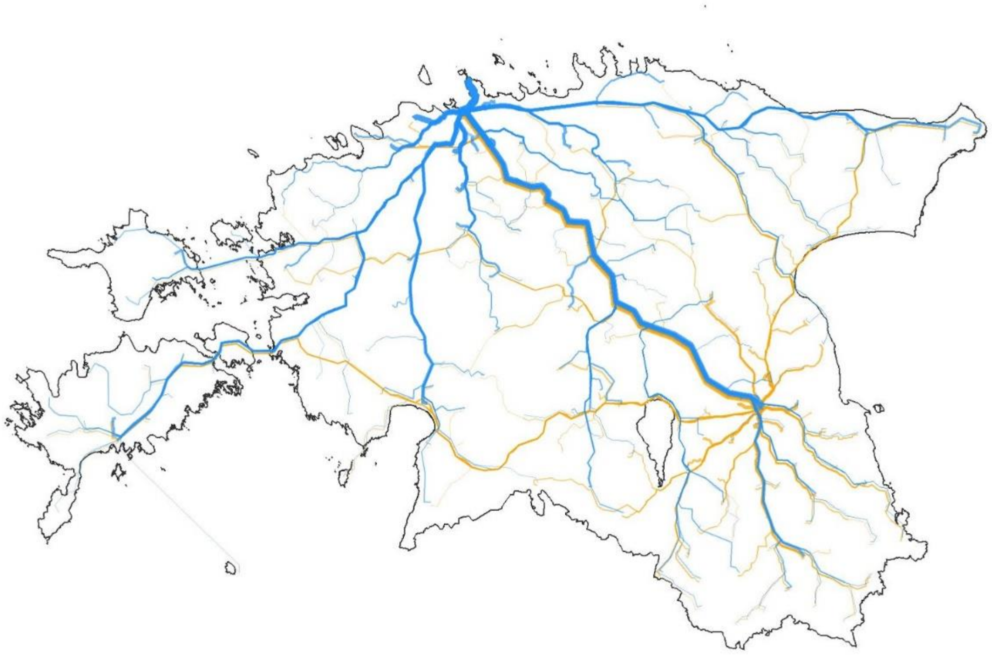
Table 3.5: Structure of mobile positioning data used in the analysis.

Field	**Description**
Respondent ID	Unique numerical pseudonym of a respondent
Site ID	ID of an anchor point at the mobile site level
Timestamp	Month level
Location information	Longitude and latitude coordinates for a mobile site
Type of anchor	Possible values: home, work-time, multifunctional, secondary

Anchor ranking	Importance of anchor (according to days spent in specific location)

All the locations the respondent visits in at least 5 days during a month will be added to the OD-matrix of everyday mobility (Fig. xxxx). For every user ID one home and one work location is defined per month. The number of secondary anchors may be higher. Movement to work and secondary anchors will be distinguished. Everyday mobility is described as movement between regularly visited places (home, work, free time). As the current dataset does not allow to construct real routes of movements, all the movements are constructed as routes that are most likely between anchor points. As a simplification, all movements start from home, meaning that the real travel chains stay hidden.

Table 3.6: Example of OD-matrix modelled on road network (Tallinn related mobility - blue; Tartu related mobility - orange).



3.5.4 Case 4: Hackathon about big data utilization potentials

In addition to the detailed case studies in the stakeholders' territories, a data hackathon is organised in close collaboration with the stakeholders to promote the innovative utilisation of new data sources. Hackathons are events where a semi-random group of people will meet in a facilitated and predetermined time to co-create solutions to the challenge presented to them. Process utilises the concept of Business innovation camp developed by Centre for Collaborative Research (CCR) at the University of Turku, however this time directing it more towards hackathon type of working due to the specific aim to harness the potentials related to new data sources. The hackathon is organised in cooperation with another regional EU-funded project called Open Data as a Service, which aims at promoting the innovative utilisation of open data. The teams will consist mainly of students from the University of Turku and Turku

University of Applied Sciences (possibility for international collaboration with e.g. University of Tartu is explored).

The hackathon is scheduled for the beginning of the next academic semester 2019. The intense execution period will take place on 17-18 January 2019. Open call for the event will be opened in November-December 2018. The launch will be executed as a briefing session, most likely on Thu 10.1.2019, immediately after the semester starts at week 2 of the year 2019.

4 Data governance and big data driven future – first overview

In order to create understanding about the efficient utilization of big data and possible development directions, an environmental scanning was utilized to seek information relating to the data governance of big data and potential, challenges and obstacles for big data driven future. The analysis included technical aspects, as well as non-technical aspects such as economic aspects and practices linked with organizational culture. The state-of-the art also includes European analysis on data collection, processing and analysis challenges.

4.1 State-of-the-art in data governance

4.1.1 Technological aspects of data governance

Venture capitalist Matt Turck has been categorizing the key players and tools in the big data field starting in 2012. In 2018, he added Artificial Intelligence to the name of his annual Big Data Landscape infographic to acknowledge how AI and Big Data developments are today deeply interconnected. The categories and subcategories used in the landscape paint a picture of the variety of activity in this growing industry sector (see Figure 3.1, Chapter 3.1). Contextualizing his 2018 infographic, Turck (2018) observes the following key trends for the year:

- The 'late majority' of large and small firms are now acquiring data infrastructure and analytics capacities through cloud services. These services in turn are adding more machine learning tools to their service offerings.
- Startups are actively developing new products for streaming, data governance, data fabrics/virtualizations, AI chips, GPU databases, AI devops tools, and platforms for distributing data science and machine learning capacities throughout an enterprise.
- AI researchers have not made breakthroughs in developing Artificial General Intelligence, but have developed new ways to apply deep learning. There is worry of "overhyping" AI could lead to a loss in market faith if promises never materialize.
- While the infrastructure vendors have reached late majority markets, enterprise and so called 'vertical AI' applications are still in the early majority.
- Data privacy concerns related to ownership and security are widespread and AI could make things worse. Centralization of data ownership also contributes to these issues.

As these tools become available, organizations such as multi-state actors (e.g. EU), national ministries, regional administrations, municipalities governance will increasingly need internal data governance policies that evolve with technological capacities and emerging legal contexts. Already, a data governance policy from 2016 for any such organizations dealing with data from EU residents is already outdated because of the GDPR. Meanwhile, the future use-cases foreseen in GDPR may not prove adequate to govern future practices. For instance, in a Netgain Partnership report, the authors posit that justice systems may have a difficult time keeping up with how advanced analytics are used in practice. For instance new concepts such as "algorithmic accountability" will require definition (Robinson & Bogen 2018, 50). Big technology players such as IBM are already initiating sophisticated open source projects like AIF360 to help companies address legal risk from biased AI algorithms. These biases can hold real-world consequences, such as death from misdiagnosed cancer or poor credit scores based on your social grouping. (Feldman 2018b.)

With new technologies, mistakes are bound to happen, yet some mistakes are bigger than others. In 2017, a large credit bureau with records about nearly all adults living in the United States, forgot to patch a web server. As a result over 100 million records were stolen by hackers. Since that data breach, little has changed in the company's practices and the company has faced few penalties (Whittaker 2018). That was a case of a clear mistake. But what happens when a well-intended and 'legal' data process produces an outcome that is unethical or breaks the law?

As early as 2006, two cases had emerged regarding data sharing and privacy. Two large Internet-based companies AOL and Netflix released large sets of anonymized data to researchers. AOL released 20 million search queries made by 657,000 of its users across two months. The company was careful to remove personal information and IP addresses from the record set. Yet newspaper reporters were able to identify of one of the users based a constellation of search terms associated with one of the anonymized users and was soon knocking on her door to write a story about how they did it. Netflix, meanwhile, launched a contest to improve its recommendation system by 10 percent.[7] They gave research teams 100 million anonymized rental records for 500 thousand users. A group of researchers demonstrated how one could combine the data with other sources and identify the users identity with 99 percent accuracy based on their ratings of at least 6 obscure movies and the dates those ratings were made. (Mayer-Schönberger & Cukier 2013, 155.) Potential for **deanonymization** continues to be a potential threat today. Despite legally compliant efforts to protect the privacy of individuals observed in a dataset, an individual's identity could in some cases be revealed with enough effort and enough additional relevant information from other datasets. For example, knowing the detailed movement of an individual or vehicle from point to point can indicate where a person lives and works. When combined with other data, this information could potentially be deanonymized.

The technological capacities regarding big data are rapidly evolving in connection to the development of mainstream and increasingly available artificial intelligence, machine learning, and deep learning tools. Some of these tools are being developed as additions to existing services offered by big tech firms while others are being developed or applied by agile startups. The next phase of this project seek practical information about these new analytical techniques and tools. The goal of this exploration is to support national, regional, and municipal governments in strategically developing their big data capacities for policymaking in corridor development.

[7] At the time, Netflix had not yet launched its movie streaming service and only distributed media via mailed DVDs.

4.1.2 Data governance and economic aspects

Big data has great economic potential. It should be bear in mind though that there is a risk of overinvesting in data governance. Data governance practices should maintain a balance between value creation and risk exposure in the new organizational imperative. Organisations should balance between risk and overinvesting. The risk is that organisations are liable to make mistakes. This could lead to technical, economic or reputational risk if organisations underinvest in storage technology for highly valuable data (e.g. storing clinical trial data for a next blockbuster drug). On the other hand, organisations could waste resources by overinvesting in storage technology when the value of their data is low, e.g. needlessly replicating data. Data can only create value when it is used, so the intent is to motivate use within a framework of safeguards that seek to protect that value at all times. Policies that strike a balance between acceptable and affordable risks are the key to an effective data governance regime (Tallon 2013).

The underlying economic benefit of big data is that it can transform economies and deliver a new wave of production growth. Business sector has recognized that big data will help them increasing operational efficiency, informing strategic direction, develop better customer services, identifying and developing new products and services and identifying new customers and markets (Philip Chen & Zhan 2014). In business, organisations are trying to define emerging industries and find innovative ways to differentiate themselves from competitors by becoming more collaborative, virtual, accurate, synchronous, adaptive and agile. They are trying rapidly responding to market needs and changes. Organisations have noticed that the data they own and how they use it can make them different from others. Data and information are becoming primary assets for many organisations (Demirkan & Delen 2013; Nathan & Rosso 2018). According to Santala (2017) biggest potential related to big data is in creating new innovations and businesses. This can be beneficiary especially to SME´s as big, global companies have better resources to collect data from different sources

Data-driven improvements and innovations are not limited to the private sector, but it can also increase the productivity in public sector as public sector planners can make decisions about how to apportion and optimize public resource. Public sector has published big data and open data strategies, as they see that data will possibly create new businesses and new enterprises creating positive societal growth, such as new jobs and more tax revenue. Big data is also linked with public sector productivity development. Big data can increase public sector efficiency as big data can support decision making and increase exchange of information between different departments (Santala 2017). Governments around the world are facing adverse conditions to improve their productivity. Namely, they are required to be more effective in public administration. Particularly in the recent global recession, many governments had to provide a higher level of public services with significant budgetary constraints. Therefore, they should take big data as a potential budget resource and develop tools to get alternative solutions to decrease big budget deficits and reduce national debt levels. Big data functionalities, such as reserving informative patterns and knowledge, provide the public sector a chance to improve productivity and higher levels of efficiency and effectiveness. European public sector could potentially reduce expenditure of administrative activities by 15-20 percent (Philip Chen & Zhan 2014). On the other hand, there are also economical obstacles hindering the development of utilising big data in public sector. The biggest economical factor hindering the development is the lack of resources in public sector, such as lack of resources hiring new personnel that can advance the data governance development and the use of big data (Santala 2017).

4.1.3 Data governance practices and organization culture

Data governance practices are organizational policies or procedures that describe how data should be managed throughout its useful economic life cycle. These compose three separate categories: structural, operational, and relational. *Structural practices* refer to setting policies and standards for protecting and using data. *Operational practices* are means by which organisations execute data governance. *Relational practices* describe the formalized links among the personnel in the organization e.g. the CEO, business managers and data users in terms of knowledge sharing, value analysis, education, training and strategic IT planning. In its best, relational procedures can support practices related to good data governance. On the other hand poor management can hinder data related knowledge sharing, education and training (Tallon 2013). Big data can also assist in rebuilding organisational procedures and policies, as well as increase transparency and information sharing in the organisations.

Tallon (2013) has identified aspects that seem to be data management enablers. Tallon noticed that organisations that have a focused business strategy tend to be more prepared to enact a standard set of data governance policies; those that are more diverse in their strategic orientation might struggle to find policies that are equally relevant to each department. On the other had a decentralized organization can find it difficult to create a common set of data governance policies or standards that will satisfy all users equally. Where organisations have bowed to internal pressure to allow divisions to opt out of established firm-wide policies in favor of personalized policies, the result has been a chaotic mix of complex and in many cases contradictory policies. Users learn not to game the system when they feel that policies are inconsistent. On the other hand, data governance enablers are: culture of promoting strategic use of IT, regulations, centralized organization structure, aligned IT and business strategy. Some interviewees emphasised especially the role of IT strategy being critical enabler. IT strategy can even hinder the use of big data if the strategy is not aligned with the organisation´s big data strategy.

Sivarajah et al. (2017) have identified challenges linked with big data. Extant studies surrounding big data challenges have paid attention to the difficulties of understanding the notion of big data, decision-making of what data are generated and collected, issues of privacy and ethical considerations relevant to mining such data. These challenges are related to management challenges, that tackle with governance and lack of skills related to understanding and analyzing data, as well as lack of skills tackling with data and information sharing, cost expenditure, data ownership, data governance, security and privacy. (Hargittai 2015; Crawford 2013; Lazer et al. 2009; Boyd & Crawford 2012).

Boyd and Crawford (2012) are raising skills as critical questions for big data and data governance. Large data sets from Internet sources are often unreliable and these errors are magnified when multiple data sets are used together. This requires skills to understand the properties and limits of a data set, regardless of its size. To make statistical claims about a data set, it is essential to know where data is coming from; it is similarly important to know and account for the weaknesses in that data. Also Manyika et al. (2011) point out that there is a shortage of data scientists in the world, and more to the point, there are critical gaps in the supply of individuals who have the skills and knowledge to use and process big data, and to interpret, apply and contextualize the results of the analysis.

Skills are key question regarding big data governance challenges. Analyzing big swathes of data is a skill set generally restricted to those with a computational background. Manovich (2011) writes of three classes of people in the realm of big data: those who create data, those who have the means to collect it, and those who have expertise to analyze it. The last group is the smallest and they often determine the rules about how big data will be used. This can create institutional inequalities. Also Santala (2017) sees that factors like lack of data skills and experience can hinder good data governance and the increase of using and opening data. In addition, some interviewed experts brought up that ability to visualize big data analysis results is also an important skill as policy-makers need to have results in a clear format (e.g. Infographics and maps) in order to be able to make decisions based on the data. Even though governments and different public sector organisations have published big data strategies, lack of skilled workforce in the operational level is problematic hindering the desirable development. In addition, other aspects related to organisational culture can hinder the use of big data in public sector. Poor level of managerial skills can lead to bad attitudes towards utilising big data. On the other hand, better leadership can encourage the increase of organisational cooperation and interaction that is needed to follow different aspects of data governance (Santala 2017). With better leadership, it is possible to create new operational models in the organisations to support both data governance and the use of big data.

4.2 European analysis on data collection, processing and analysis challenges

A recently published research roadmap for Europe (Cuquet & Fensel 2018) presents a comprehensive picture about the issues that should be taken into account regarding big data collection, processing and analysis (Table x). The key issues in the roadmap will be further explored in the next phase of the study. In addition, key aspects from the European data economy and single market strategies will be summarised regarding data collection, processing and analysis. Furthermore, the integration of big data to statistical systems is discussed according to the overall aims of the ESS Big Data Action Plan and Roadmap 1.0. Altogether, this chapter will be developed based on a literature review on recent research materials presenting the EU level discussion around the topic.

Table 4.1: Key aspects identified in the research roadmap (Cuquet & Fensel 2018).

Data management	○ handling unstructured and semi-structured data ○ semantic interoperability ○ measuring and assuring data quality ○ data lifecycle ○ data provenance, control and IPR ○ data-as-a-service model and paradigm ○ open data practices
Data processing	○ techniques and tools for processing real-time heterogeneous data ○ scalable algorithms and techniques for real-time analytics ○ decentralized and distributed architectures ○ efficient mechanisms for storage and processing
Data analytics	○ improved models and simulations ○ semantic analysis ○ event and pattern discovery ○ multimedia (unstructured) data mining ○ machine learning techniques, especially deep learning for business intelligence, predictive and prescriptive analytics
Data protection	○ complete data protection framework ○ privacy-preserving mining algorithms ○ robust anonymization algorithms ○ protection against reversibility
Data visualization	○ end user centric visualization and analytics ○ dynamic clustering of information ○ new visualization for geospatial data
Non-technical priorities	○ establish and increase trust ○ privacy-by-design ○ ethical issues ○ develop new business models ○ citizen research ○ discrimination discovery and prevention

4.3 Future potentials and challenges of big data for growth corridor spatial policies

The analysis of potentials, challenges and obstacles for big data driven future shows that developing tools for big data analysis holds central attention in the computing industry. One example of this is extreme data analysis. The term *extreme data analytics* is beginning to appear in grey literature starting in 2015 and is a subset of a research field called High Performance Computing. *Extreme data* is described as a next form of big data. Extreme data is purported to be bigger, faster, and of higher variation than Big Data. For example, Boman et al. (2015) call for new forms of hardware--such as systems with enough RAM to hold a large dataset and run analytic software--and IT systems capable of more rapidly delivering outputs. Additionally, the EU has recently announced it will invest 1.4B euros in developing new next-generation supercomputers using European technology (Feldman 2018). Already in data science practices with Big Data, some datasets are so large and so rapidly produced it is impossible to analyze all of it--which leads to a practice of running analysis on sample populations from the dataset.

Also the digital divide is an aspect that influence big data development. The implications of living with big data and algorithms is unclear. The digital divide in regards to forms and speeds of available Internet, particular between urban and rural area, is a key variable in the present that will likely have large influence on what regions and municipalities will have access to essential types of Big Data as well cloud-based analysis tools for Big Data. These disparities and variations across regions not only impact today's situation, but also could have big implications for future developments—especially in cases where past data would be required to develop highly tuned algorithms and deep learning networks. (Schintler 2018.)

There are also obstacles that can hinder the use of radical technologies in the future development of data driven spatial policies at the growth corridors. General Data Protection Regulation (GDPR) could serve as a huge bottleneck or constraint on the use of big data and achieving the benefits it can use. Privacy is a critical factor affecting the ability of governments and other organisations to use big data (Stough & McBride 2018). In the publication of the Committee for the Future of the Finnish Parliament (Kuusi & Linturi 2018) MyData and GDPR were estimated to be one of the key aspects affecting on the potential of radical technologies. Kuusi & Linturi (2018) suggested that. the relation between AI and GDPR should be clarified. GDPR should be interpreted so that the law does not hamper progress with AI and robotics, or otherwise regulation can hinder AI and big data development in Europe. MyData, on the other hand, can ease the challenge with GDPR as individuals can accept and control their personal data promoting innovations (Kuusi & Linturi 2018).

4.3.1 Big data and complex systems

According to Gillard et al. (2016) the increasing pressure of climate change and global urban population growth inspire two normative agendas: socio-technical transitions and socio-ecological resilience, both sharing a complex systems epistemology. They note that deploying sensors and big data is not enough when planning for resilient regions that are complex systems. Enormous volumes of data are generated, but only some are relevant and most of which describe static and stable conditions. Instead, it should be directed towards the ability to withstand shocks, such as extreme weather events or unexpected economic shocks and other stresses that might influence in the viability of both human and natural system (McPherson 2014). Big data driven policy-making could provide a capability to react more rapidly to these unpredictable events, and capability to anticipate systemic transitions. In businesses, big data is typically used for short-term decision-making processes. In public sector, decision-making usually takes longer time and long time horizon is important in visionary development and policy-making. The temporal insights gained from big data can provide significant value in business and the public sector. An interviewee noted that some datasets at the surface show a static snapshot in time, but when analysed for temporal information, can reveal insightful temporal patterns. Day and night locational studies are one example (EC). Another is to analyse a population register for changes in home locations over 10 to 30 years. The action oriented approach to using big data in business can help the public sector as well. Shortening the time between documenting observations and producing indicators could help policymakers be more prepared for emerging shocks, stresses, or opportunities across corridors.

Monitoring of both social and environmental systems is needed in order to study current and future states and improve the ability to adapt to potential state changes and develop methods for governing these systems in inclusive ways. In the absence of an integrated data-driven policymaking paradigm, communication between data scientists and policymakers only happen sporadically (Boyd & Crawford 2012). However, cities and regions are complex socio-economical systems consisting of many subsystems, including social, economic, technical and ecological. Establishing any intervention is not possible without a systematic holistic approach to how cities and regions function, which technologists are not trained for (Klosterman 2013). In the future a vision for how these pieces might work with the human and institutional pieces would be beneficial (Estiri & Afzalan 2018).

4.3.2 New data sources in policy-making

The interviews exposed an interesting observation that even though new data sources have a lot of potential in policy-making, many public sector data analysts and policy-makers take negative attitude to using these data sources such as social media data due to the unclear status of data in policy-making processes. For instance, geotagged Tweets and Instagram pictures have a lot of potential bringing new information of people´s spatial interactions. But there are obstacles why many officers still prefer using more reliable data sources, such as traditional statistics databases. e.g. public sector officers understand that social media data is usually biased as all age and social groups are not active social media users. Yet, there is interest in using untraditional data sources where doing so can speed up production of indicators and forecasts. There are also new companies offering analysis services utilising new data sources, like data of Internet trends. But if companies are not willing to reveal the model they use in the analysis, the lack of transparency can cause scepticism towards these service products. Therefore public sector officers are cautious of using these new services and data sources to support policy-making.

5 Conclusions

The Interim report presents the results of the second phase of the Big Data & EGC project, which will continue by further analysing new datasets in the case studies as well as by going deeper into the landscape of data governance and big data driven futures. Overall, big data and rapidly emerging analysis tools offer great potentials for business and policy making. Yet, using big data requires attention to the limitations of various analytical methods and asking 'why' various analytical results occur. This will require sensitivity to the cultural and daily contexts of the topic under study and a willingness to combine data science with other forms of research.

There are several challenges in utilising big data in corridor development, such as the missing spatial dimension from big data components as well as the fact that the majority of big data applications are designed for businesses and industries rather than for the government sector. In addition, there is not enough big data suitable for territorial development or there is not enough knowledge on the available data. Data structure is also siloed and there is not institutional setup and routine of sharing and collecting data. In a workshop organised in the first phase of the project, stakeholders identified three themes as the most important policy dimensions related to corridor development that would benefit from big data: 1) infrastructure planning and transportation; 2) regional economic development, and; 3) land-use planning. Even though big data utilisation was not recognised as such as being among the key strategic objectives, it was in the interviews said to be in-built in most corridor development practices as more efficient utilisation of data plays a central role in advancing smooth mobility and flows along the corridors as well as their development as uniform labour market areas and (digital) innovation platforms. Furthermore, the need for new sources of data describing the interactions and connectivity along the corridors was mentioned as evident in the practical attempts to improve the data-driven decision-making in corridor development.

A data categorisation was produced based on the review of available datasets, which will be further complemented by utilising the results of the hackathon organised in January 2019. The diverse aspects of big data utilisation were first elaborated, after which a review about open data was conducted in the case area. All of the countries - Norway, Sweden, Finland, Estonia and Russia - had at least one fairly mature and active data portal, and there were many similarities among the portals, potentially allowing more comprehensive analysis along the corridor. However, the utilisation of portals has to start from a specific need and a policy question to be able to identify their more specific utilisation possibilities. Also the possible benefits of case studies have been discussed with stakeholders to match the analysis to the concrete policy needs. In building big data categorisation for corridor development, various datasets were recognised that could be utilised in deepening understanding about the corridor functionalities related to physical, social and digital connectivities along the corridors. In the next phase of the project, the case studies will provide new innovative ways to think about spatial connections as well as transferable examples for other European growth corridors.

The next phase of the Big Data & EGC project aims at deepening the analysis of new available datasets and analysis methods in the context of the three case studies in the study area, as well as identifying other possible data sources and utilisation potential by organising the data hackathon for diverse disciplinary students in Turku and possibly Tartu university. In addition, wider European level possibilities and challenges related to utilisation potential are further explored with an especial focus on territorial and corridor development. Futhermore, policy recommendations and suggestions to improve territorial governance and coordination as well as public-private-partnetships are presented.

References

Ahas, R., Silm, S., Järv, O., Saluveer E., Tiru, M. 2010. Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones , Journal of Urban Technology, 17(1): pp. 3-27.

Blitzstein and Pfister (2015) CS105 Data Science, course website. Available at: http://cs109.github.io/2015/ (Accessed 5 October 2018.)

Boman, E., Madduri, K., Rajamanickam S & Wolf, M. (2015) High-Performance Computing for Extreme-Scale Data Analytics. US Department of Energy's Office of Scientific and Technical Information. Available at: https://www.osti.gov/biblio/1253084 (Accessed 4 October 2018.)

Boyd, D. & Crawford, K. (2012) Critical questions for big data: Provocations for a cultural, technological and scholarly phenomenon. *Information, communication & society,* Vol. 15 No.5, 662-679.

Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018). Notes from the AI frontier. Insights from hundreds of use cases. Discussion Paper. McKinsey Global Institute. April 2018. Available at: https://www.mckinsey.com/mgi/ (Accessed 27 September 2018.)

Crawford, K. (2013). The hidden biases of big data. Harvard Business Review Blog, 1 April 2013. Available at: https://hbr.org/2013/04/the-hidden-biases-in-big-data (Accessed 14 September 2018.)

Cuquet, M. & Fensel, A. (2018) The societal impact of big data: A research roadmap for Europe. Technology in Society 54, 74-86.

Demirkan, H. & Delen, D. (2013) leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems,* 55 (2013), 412-421.

Dietrich, David (2014) Three Big Misconceptions about Big Data. 15.1.2014 *Dell EMC In Focus.* Available at: https://infocus.dellemc.com/david_dietrich/three-big-misconceptions-about-big-data/ (Accessed 2 October 2018.)

Estiri, H. & Afzalan, N. (2018) Towards data-driven cities. Incorporating big data into urban management.*In* Big data for regional science, *eds.* Schintler, L. A. & Chen, Z. Routledge, New York.

Feldman, Michael (2018) Europeans Budget 1.4 Billion Euros to Build Next-Generation Supercomputers. *Top 500 Supercomputer Sites,* 1 October 2018. Available at: https://www.top500.org/news/europeans-budget-14-billion-euros-to-build-next-generation-supercomputers/ (Accessed 4 October 2018.)

Feldman, Michael (2018b) IBM Introduces Software to Alleviate AI Bias. *Top 500 Supercomputer Sites,* 24 September 2018. Available at: https://www.top500.org/news/ibm-introduces-software-to-alleviate-ai-bias/ (Accessed 5 October 2018.)

Gillard, R.; Gouldson, A.; Paavola, J. & Van Alstine, J. (2016) Transformational responses to climate change: Beyond a systems perspective of social change in mitigation and adaptation: Transformational responses to climate change. *Wiley Interdisciplinary Reviews: Climate Change,* 7(2), 251-265.

Grinberger, A.Y. & Felsenstein, D. (2018) Using big (synthetic) data to identify local housing market attributes. In: Schintler, L. A., & Chen, Z. (eds.) *Big Data for Regional Science*. Oxon: Routledge.

Guo, D. (2009) Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data. IEEE Transactions on visualization and computer graphics, 15(6), 1041-1048.

Hargittai, E. (2015) Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science*, Vol. 659 No. 1, 63-76.

Hwang, S., Yalla, S. & Crews, R. (2018) Processing uncertain GPS trajectory data for assessing the locations of physical activity. In: Schintler, L. A., & Chen, Z. (eds.) *Big Data for Regional Science*. Oxon: Routledge.

Khan, M. A.; Uddin, M. F. & Gupta, N. (2014) Seven V's of Big Data: understanding Big Data to extract value. Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education. https://doi.org/10.1109/ASEEZone1.2014.6820689

Kim, G.-H., Trimi, S. & Chung, J.-H.(2014) Big-data applications in the government sector. *Communications of the ACM.* Vol. 57:No.3 March 2014.

Kanellos, Michael (2016) The Five Different Types of Big Data. *Forbes.com* 11 March 2016 (Accessed 10 September 2018) https://www.forbes.com/sites/michaelkanellos/2016/03/11/the-five-different-types-of-big-data/

Klosterman, R. E. (2013) Lessons learned about planning. *Journal of the American Planning Association,* 79(2), 161-169.

Kuusi, O. & Linturi, R. (2018) Suomen sata uutta mahdollisuutta 2018-2037. Yhteiskunnan toimintamallit uudistava radikaaliteknologia. *Eduskunnan tulevaisuusvaliokunnan julkaisu* 1/2018.

Lazer, D.; Pentland, A.; Adamic, L.; Aral, S.; Barabai´si, A.; Brewer, D., Christakis, N.; Contractor, N.; Fowler, J.; Gutmann, M.; Jebara, T.; King, G.; Macy, M.; Roy, D.; Van Alstyne, M. (2009) Life in the network: the coming age of computational social science. *Science,* Vol. 323 No. 5915, 721-723.

Manovich, L. (2011) Trending: the promises and the challenges of big social data. *In* Debates in the Digital Humanities, ed. M. K. Gold, The University of Minnesota Press, Minneapolis, 4/28/2011.

Manyika, J.; Chui, M.; Brown, B.; Bughin, J.; Dobbs, R.; Roxburgh, C. & Hung Byers, A. (2011) Big data: The next frontier for innovation, competition and productivity. McKinsey Global Institute.

Mayer-Schönberger, Viktor & Cukier, Kenneth (2013) *Big Data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt.

McPhearson, T.; Hamstead, Z. A. & Kremer, P. (2014) Urban ecosystem services for resilience planning and management in New York City. *AMBIO*, 43(4), 502-515.

Nathan, M. & Rosso, A. (2018) Exploring digital technology industry clusters using administrative and frontier data. *In* Big data for regional science, *eds.* Schintler, L. A. & Chen, Z. Routledge, New York.

Oxford Internet Institute (2018) In Conversation with Viktor Mayer-Schönberger. Interview. YouTube. https://www.youtube.com/watch?v=5d6QabsQ2V0 (Accessed 27 September 2018.)

Oliveira, M. I. S., Oliveira, L. E. R. A., Batista, M. G. R., & Lóscio, B. F. (2018) Towards a meta-model for data ecosystems. *Proceedings of the 19th Annual International Conference on Digital Government Research Governance in the Data Age - Dgo '18*, 1–10. https://doi.org/10.1145/3209281.3209333

Patuelli, R. & Arbia, G. (2016) *Spatial Econometric Interaction Modelling*. Springer.

Philip Chen, C.L. & Zhang C.-Y. (2014) Data-intensive applications, challenges, techniques and technologies: Survey on Big Data. *Information Sciences,* 275 (2014), 314-347.

Robinson, D. & Bogen, M. (2018) *Automation & the Quantified Society.* The Netgain Partnership. Available at: https://www.netgainpartnership.org/resources/2018/1/26/automation-and-the-quantified-society (Accessed 4 October 2018).

Santala, V. (2017) Avoimen datan kehitys pohjoisella kasvuvyöhykkeellä. Haasteet ja mahdollisuudet sekä merkitys kasvuvyöhykkeen kehittämiselle. (Translated: Development of open data in the North growth zone. Challenges and opportunities as well as the importance of developing the growth zone.) University of Turku Faculty of Mathematics and Natural Sciences, Department of Geography and Geology, 02-02-2017.

Schintler, L. A. (2018) The constantly shifting face of the digital divide: Implications for big data, urban informatics, and regional science. In: Schintler, L. A., & Chen, Z. (eds.) *Big Data for Regional Science*. Oxon: Routledge.

Shi, X.; Qian, Y. & Dong, C. (2017) Economic and Environmental Performance of Fashion Supply Chain: The Joint Effect of Power Structure and Sustainable Investment. *Sustainability* 2017(9): 961.

Sivarajah, U.; Kamal, M. M.; Irani, Z. & Weerakkody, V**.** (2017) Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70 (2017) 263-286.

Spiekermann & Wegener 2007. ESPON Project 1.4.4. Preparatory Study on Feasibility of Flows Analysis. Final Report.

Surman, M. & Thorne, M. (2016) We All Live in the Computer Now: A NetGain paper on society, philanthropy and the Internet of Things. The NetGain Partnership. 20 October 2016. Available at: https://www.netgainpartnership.org/resources/2018/1/26/internet-of-things (Accessed 3 October 2018.)

Turck, M. & Obayomi, D. (2018) Big Data & AI Landscape 2018. Infographic. Available at: http://mattturck.com/wpcontent/uploads/2018/07/Matt_Turck_FirstMark_Big_Data_Landscape_2018_Final.png (Accessed 8 October 2018.)

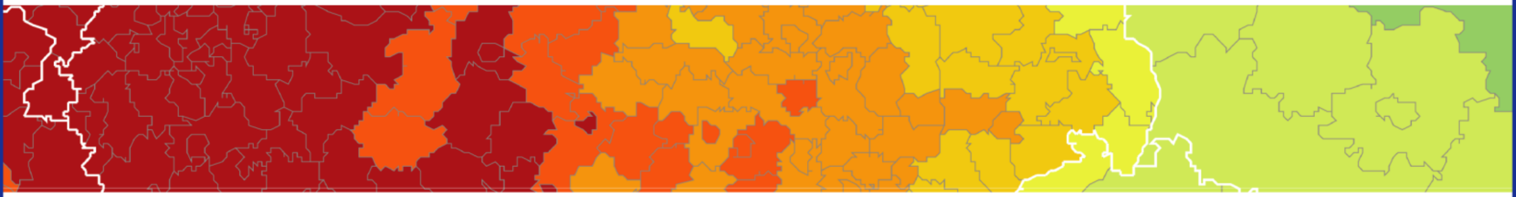
Tallon, P. P. 2013 Corporate Governance of Big Data: Perspectives on Value, Risk and Cost. *Computer,* Vol. 46. Iss. 6. June 2013.

Turck, Matt (2018) Great Power, Great Responsibility: The 2018 Big Data & AI Landscape. 26 June 2018 *Mattturck.com*. (Accessed 20 September 2018) http://mattturck.com/bigdata2018/

Van Dijk, J. (2012) *The network society.* 3rd ed. Sage, London.

Whittaker, Zack (2018) A year later, Equifax lost your data but faced little fallout. *TechCrunch* 8 September 2018. https://techcrunch.com/2018/09/08/equifax-one-year-later-unscathed/ (Accessed 8 October 2018.)

Zonneveld, W. & Trip, J.J. (Eds.) (2003) Megacorridors in North West Europe. Investigating a New Transnational Planning Concept. Delft, Delft University Press.



ESPON 2020 – More information

ESPON EGTC
4 rue Erasme, L-1468 Luxembourg - Grand Duchy of Luxembourg
Phone: +352 20 600 280
Email: **info@espon.eu**
www.espon.eu, **Twitter**, **LinkedIn**, **YouTube**

The ESPON EGTC is the Single Beneficiary of the ESPON 2020 Cooperation Programme. The Single Operation within the programme is implemented by the ESPON EGTC and co-financed by the European Regional Development Fund, the EU Member States and the Partner States, Iceland, Liechtenstein, Norway and Switzerland.