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

Targeted Analysis

Inception Report

Version 18/08/2018
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.

© ESPON, 2018

Printing, reproduction or quotation is authorised provided the source is acknowledged and a copy is forwarded to the ESPON EGTC in Luxembourg.

Contact: info@espon.eu

Disclaimer:
This document is an inception 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 content

1 Introduction...........................................................................................................................................i
2 Description of conceptual and methodological framework................................................................. 3
   2.1 Conceptual framework for addressing the research need ................................................................. 3
   2.2 Methodological framework ............................................................................................................ 5
3 Overview of stakeholders’ public policies and needs ............................................................................. 7
   3.1 Public policy considerations for big data .......................................................................................... 7
      3.1.1 Better governance ..................................................................................................................... 7
      3.1.2 Data driven innovations .......................................................................................................... 7
      3.1.3 Different needs in public and private sectors .......................................................................... 8
   3.2 Big data and corridor development .................................................................................................. 8
   3.3 Big data and public policy needs - cases Finland, Estonia and Sweden ....................................... 10
      3.3.1 Big data and public policy needs in Finland ........................................................................... 10
      3.3.2 Big data and public policy needs in Estonia .......................................................................... 12
      3.3.3 Big data and public policy needs in Sweden ........................................................................... 15
   3.4 Big data needs defined by the stakeholders .................................................................................... 16
      3.4.1 Infrastructure and connectivity planning ................................................................................... 16
      3.4.2 Economic development .......................................................................................................... 16
      3.4.3 Land-use planning .................................................................................................................. 17
      3.4.4 Data gaps defined by the stakeholders .................................................................................... 17
4 First categorisation of available data sources ....................................................................................... 18
4.1 Plan for conducting prospective analysis of new datasets at European level .................................. 20
5 Plan for identifying good practice examples ......................................................................................... 22
6 Conclusions ........................................................................................................................................... 24
References ............................................................................................................................................... 26

ANNEX 1 – PRESENTATION OF INCEPTION REPORT TO PROJECT STEERING COMMITTEE ......................... 29
ANNEX 2 – FRAMEWORK FOR CASE STUDIES 1 & 2 ............................................................................ 38
ANNEX 3 – CASE STUDY 3: Estonian Case ............................................................................................ 47
List of Figures

Fig. 1. The methodological framework of the Targeted Analysis……………………………..5

List of Tables

Table 1. Conceptual framework to analyse the potentials of big data for integrated territorial development in the European growth corridors. The framework is used in mapping different data sources to ensure comprehensive analysis of different dimensions of corridor development……………………………………………………………………………………………………………………………………………………………………4

Table 2. Categorisation of available data sources…………………………………………….19
1 Introduction

European spatial development has been largely about normative envisioning for a more economically competitive, socially cohesive and environmentally sustainable spatial structure. One example of the spatial visions are European growth corridors that cross several administrative boundaries and follow the trans-European transport network (see, e.g., Kalliomäki 2012; Moilanen 2012; de Vries & Priemus 2003). Corridor development focuses on the more efficient use of the existing infrastructure, which is seen to enable both the economic integration of the area and sustainable transportation as well as spatial planning and mobility based on a connecting infrastructure. However, there has also been criticism towards the abstract nature of these spatial visions and their inability to engage with regional development practices at the local and regional level (e.g. Kalliomäki 2012; Davoudi 2009). The utilisation of spatial data that adequately captures the functionalities of corridors related for example to daily mobilities and practices has been so far rather weak in planning processes.

This Inception report presents the first phase of the Big Data & EGC targeted analysis project that attempts to address the challenge of big data utilisation by analysing the potentials of big data for integrated territorial policy development in the European growth corridors. It is now widely understood that due to the rapid growth in data sources and analytical tools there are going to be big changes in the ways of policymaking and its effects on citizens. Recently the potentials of big data have been emphasised in policymaking due to the fast developments in both the amount of data and the ways it is being handled in policymaking.¹ The combination of the trend of digitizing administrative data, collecting data through diverse devices and rapid developments in data storage has led to the establishment of numerous big and open data initiatives at diverse governmental scales of policymaking (Giest 2017), including corridor development.

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). In addition to referring to the size of the data, big data refers also to its variety, velocity and veracity: This 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).

The themes arising from the academic literature related to the role of big data in policymaking can be categorised into three topics (Giest 2017: 1) the role that institutional capacity has within government to utilise big data analytics; 2) government use of big data analytics in the context
of digital public services, and; 3) the way that big data information enters the policy cycle. All of these three aspects are highly relevant also from the perspective of integrated territorial policy development in the European growth corridors. The role of institutional capacity – and the data culture – in the context of growth corridors has to be looked at several levels, starting from the capacity of individual civil servants as well as the organisations at the local, regional, national levels to finally transnational and corridor levels of policymaking. Furthermore, intertwined to the discussions related to big data, there has been a growing demand and interest on opening public sector information all over the world and especially in Europe (see OECD 2017). In recent years, numerous public organizations have made initiatives to publish public sector information as open data.

The aim of the Big Data & EGC project is to strengthen the knowledge-base for evidence-based 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. In addition, the corridor includes the northern parts of ScanMed TEN-T corridor. The NGZ aims at boosting the region’s competitiveness in the global arena through the creation of a single, internationally recognized market and commuter belt. However, a central challenge in the NGZ is that it has been developed through partial solutions and organizational silos that are not optimally serving the development of the corridor as a whole. For example, transportation planning has long been too separate from land use planning even though the location of jobs and housing is a crucial factor for sustainable mobility. In its planning and development processes, the corridor is still mostly leaning on more traditional data sources.

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. By examining big data related policies and needs in Finland, Sweden and Estonia, the project answers to the needs of stakeholders in Finland (Regional Council of Southwest Finland), Sweden (Region Örebro) and Estonia (Ministry of Economic Affairs and Communications) by generating new knowledge about the functionalities of the NGZ and European growth corridors and solutions that can be utilised in the corridor development.

The report is organised as follows. After the introduction, the second chapter describes the applied conceptual and methodological framework as they form the starting point and backbone for the analysis. Third, an overview of stakeholder’s public policies is presented at growth corridor levels that could benefit from new big, open and soft data provision. Fourth chapter presents the first draft categorisation of available data sources for evidence-based policy making as well as a plan for conducting prospective analysis of new data sets at European level. Fifth, a plan for identifying good practice examples of big data sets provision is described. Finally, conclusions are drawn from the first phase of the project and ideas are presented for the next steps.
2 Description of conceptual and methodological framework

2.1 Conceptual framework for addressing the research need

The approach taken in the project to explore the potentials of big data for integrated territorial policy development in the European growth corridors builds on the conceptualisation of growth corridors as meta-governance frameworks for spatial development. Reviewing corridors as meta-governance frames emphasises the objective of integrated territorial policy development as they aim at integrating diverse development practices sectorally, spatially, and between multiple actors involved in spatial development (e.g. Haughton et al. 2010).

Corridor-based development practices under the meta-governance framework have often been categorized to emphasise three thematic approaches around connecting infrastructure (Zonneveld & Tripp 2003): 1) an *infrastructure axis*, emphasising transportation and logistics; 2) an *economic development axis*, highlighting the impacts of connecting infrastructure on creating functional economic cooperation; and 3) an *urbanisation axis*, focusing on directing the future urbanisation and public transportation. These dimensions of corridor development were also raised as most important ones in the needs analysis of stakeholders. In addition, a fourth dimension has been often added to refer to the *institutionalisation of the corridors* after their physical formation. The institutional axis sheds light on the issues of governance and capacity building, implying the creation of “institutional conditions for collective action” (de Vries 2003).

In the conceptual framework, these dimensions are taken as a starting point to build a comprehensive understanding of different flows and interactions characterising functionalities along growth corridors. However, these basic dimensions are analysed more in detail in the form of different identified flows and interactions. The conceptualisation of corridors based on their different development dimensions is important, because the different dimensions and interactions have also different implications on evidence-based policy development.

The different spatial characteristics of corridor development have to be also present in the conceptual framework as they refer to the different implications on governing spatial development. In contemporary Europe, strategies for bounded, networked, and fluid territories are simultaneously present in spatial policy and planning (Jauhiainen & Moilanen 2011). In the corridor-based development framework these three modes are brought together for contemporary spatial development. Building around the physical infrastructure and the material basis of the corridor, namely its geographical characteristics, and connecting actors to diverse cross-sectoral, cross-scalar, and cross-temporal activities corridors bring together different dimensions of spatiality. Therefore, spatial development and policy should build on the physical, bounded notions of space with more relative and relational understandings that is, to make use of existing material context and networks in building the relational dimension.
In this targeted analysis this conceptualisation of corridors is developed further for the purposes of analysing the potentials of big data for integrated territorial policy development in the European growth corridors. Three spatial categories are emphasised because of their different policy implications: physical, social and digital. In the needs analysis (described in the next section), the following interactions were identified that can be explored by utilising various big data sources: physical, digital, administrative, organisational, intellectual, and social. In the conceptual framework (Table 1), administrative, organisational and intellectual networks are seen to be taking place in networked practices of social space.

Table 1. Conceptual framework to analyse the potentials of big data for integrated territorial development in the European growth corridors. The framework is used in mapping different data sources to ensure comprehensive analysis of different dimensions of corridor development.

<table>
<thead>
<tr>
<th>TYPES OF FLOWS/INTERACTIONS</th>
<th>AVAILABLE DATA</th>
<th>SCALE</th>
<th>METHODS</th>
<th>RESTRICTIONS</th>
<th>FUTURE POTENTIALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHYSICAL SPACE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Commuter flows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Transportation flows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Tourism flows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOCIAL SPACE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Administrative interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Organisational networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Intellectual networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Social networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIGITAL SPACE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Service provision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the context of exploring the potentials of big data for integrated corridor development, the conceptual frame is utilised to categorize big data-based potentials. For example, the analyses related to mobile positioning data have different implications on developing growth corridors as an axis for infrastructure development, economic development, urbanisation, and finally, institutionalising corridor development. In a similar manner, the results of the research related to automated traffic measurement data have different implications on the different dimensions of corridor development. Increased understanding about the intercity functionalities related to traffic flows and the daily mobilities of citizens can provide a more detailed evidence-base not only for infrastructural development but also for land use planning at diverse scales. In addition, increased evidence about corridor functionalities can increase the institutional acceptance of
corridors as frameworks for governing spatial development as well as provide new business opportunities for private actors.

Furthermore, from the perspective of different data sets, understanding the different spatial dimensions of corridors related to physical, social and digital spaces is important for the appropriate framing of policies and the efficient utilisation of research results. For example, the results of the mobile positioning research have different implications on developing the corridor as a physical space (infrastructure planning), as a social space (balanced development, equity etc.) and as a digital space (e.g. service development).

### 2.2 Methodological framework

The methodological framework of the project is presented in Figure 1. The methodological framework presents the overview about the execution of tasks from 1 to 4.

![Diagram of the methodological framework](image)

*Fig. 1. The methodological framework of the Targeted Analysis.*

First, stakeholders’ knowledge needs are explored and defined in Task 1 to formulate a solid basis for further actions. In addition, gaps in the current evidence-base are identified based on the academic and policy literature as well as the concrete practices in stakeholder territories. Needs analysis as well as identification of gaps in the knowledge-base are conducted by utilising the conceptual frame presented in the chapter 1 to generate a comprehensive picture about the knowledge needs and gaps related to the different dimensions of corridor development. Further actions are planned so that they provide answers to, and improved understanding of, the concrete policy needs.
Next, an analysis of new datasets is conducted at the European level in Task 2 to create a comprehensive overview about the possibilities and challenges related to big data utilisation in different contexts. An important part of this task is the categorisation of available data that is later utilised in formulating the policy proposals for corridor development and diverse scales of territorial development.

In the Task 3, explorative case studies are conducted in the stakeholder territories to demonstrate the potentials of big data for integrated territorial policy development. Case studies utilise various new data sources, most important being the automated traffic measurement data and mobile positioning data. In parallel with the case studies also a data hackathon is organised to explore more in detail the possibilities and challenges related to cross-border utilisation of big data and the questions related to public-private interface and governance in data utilisation.

Based on the previous tasks, good practices and policy proposals are then formulated in Task 4 to guarantee the transferability of the results of the Targeted Analysis to the European regions.
3 Overview of stakeholders´ public policies and needs

3.1 Public policy considerations for big data

3.1.1 Better governance
The concepts of Digital-era Governance (DEG), Data Readiness, Evidence-based Policymaking and Policy Design all link to public use of new technologies and big data streams (Giest 2017). Data-driven process improvements are not limited to the private sector, but it can help planners make decisions about how to apportion and optimize public resources. (Hemerly 2013). It has been estimated that big data has potential to create 250 000 M€ productivity growth per year in Europe (Manyika et al. 2011).

Many case studies have shown that big data can be utilized in different policy processes: agenda setting, alternative policy development and policy evaluation (Shi 2014). Big data analytics can provide information and thus help policy makers targeting the important and primary agendas. In addition, big data can be implemented to find alternative policy solutions. It is evident that structured and unstructured forms of data can aid policy-making. Big data can provide accurate, proactive and participatory policy analysis (Shi et al. 2017). Big data enables decision-makers to decide based on evidence rather than just intuition (Giest 2017; Shi et al. 2017). Decision-makers are often confronted with the need to collect and analyse data across an array of different sectors such as economic, education and health.

3.1.2 Data driven innovations
By use of big data, it can be easier to make strategic decisions and drive innovations. Data-driven innovations is a vital area of innovation and economic growth. Data-driven innovation has high potential for improving economies and societies. This is vital not just in private sector, but data driven innovations have also potential in public sector, such as enhancing economic growth in EU-level. For instance, in 2010 the European Commission initiated its “Digital Agenda for Europe” to address how to deliver sustainable economic and societal benefits to EU citizens from a single digital market through fast and interoperable Internet applications (European Commission 2014). The European Commission has also published a big-data strategy emphasizing the economic potential of public data (European Commission 2014).

In addition, big data can expose people’s subconscious behavioural patterns, and behaviour changes in lifestyles. For instance, it can bridge the gap between what people say they want to do and what they actually do, as well as how they interact with others and their environments. This information is useful to public sector organizations as well as private companies to support decision-making and data driven innovations. By gaining better understanding of citizens’
behaviour, e.g. passenger needs and behaviour in transportation, big data can drive new service and infrastructure innovations. (Michael & Miller 2013; Strickland 2013.)

3.1.3 Different needs in public and private sectors

Big data needs, processes and time-scales differ in public and private sectors. In businesses, big data is typically used for short-term decision-making processes for maximizing self-interest and minimizing costs, whereas in public sector decision making the goal is in long-term decision-making processes for maximizing self-interest and promoting the public interest (Kim et al. 2014). In public sector, decision-making usually takes much longer and is conducted through mutual consent of a large number of actors, including officials, interest groups and citizens. Big data has been helping to facilitate social and civic empowerment, enhance and expand stakeholder participation in the policy and planning processes. In addition, it can be used to promote transparency and accountability in decision-making. Many well-defined steps are therefore required to reduce risk and increase the efficiency and effectiveness of public policy-making, as big data and methodological development in the area of sentiment analysis have potential to find the weight for different criteria like interests, preferences, priorities and values of a diverse and sometimes competing set of stakeholders. (Stone 2002; Birkland 2011; Brennan et al. 2014; Schintler & Kulkarni 2014.)

3.2 Big data and corridor development

Big data has also potential in policy-making and corridor-based development. Studies show that big data can help public policy-making concerning growth corridors. Big data has begun to derive insight to help support public territorial decision making, for instance, policy areas typically related to promoting infrastructure plans, economic growth, sustainable development, transportation, smart grid, energy efficiency, education and security (Kim et. al. 2014; Shi et al. 2017).

Portraying European growth corridors as frameworks for meta-governing spatial development highlights the role of evidence – and data – in their planning processes. The role of data is central in understanding the complex, often place-based development challenges. Nowadays the arguments about functional corridors are too often based on assumptions and strategic objectives without an adequate evidence-base. Therefore, understanding both micro- and macro-level dynamics and mobilities becomes increasingly important in transnational corridor development that in the end aims at affecting the behavioural patterns of individual citizens and companies. So far, the knowledge-base of corridor development has leaned mostly on the statistical data related to administrative areas at diverse scales, and also the functionalities and flows along the corridors have been described mostly between the administrative areas (Kotilainen et al. 2016). This is a major information gap for growth corridors that should be
addressed in the analyses. As far as traffic is concerned, data is standardized at the EU level. Data on transnational traffic and transport rates are available through Eurostat, but more detailed analyses at a regional level are often based on cross-border and small sample size surveys. For example, total freight rates give little information on the functionality of the corridors and the networking of cities.

There are very few detailed analyses about the functionality of growth corridors. The reviews lack, inter alia, the analysis of people’s day-to-day mobility, for example, in terms of running errands and commuting. Concerning the analyses conducted in the projects of ESPON program, most functionality-related research has focused on urban areas (Kotilainen et al. 2016). So far, corridors have been studied very little, although these are major frameworks for territorial development. Thus the existing gaps in the knowledge-base could be significantly complemented by utilising new data sources and methods of analysis.

However, there are special challenges related to European growth corridors and public policy-making for big data driven decisions. One of those is the spatial information. Policy analysts are increasingly using geographic information system (GIS), spatial data and spatial modelling, but this spatial dimension is often missing from big data components (Shi et al. 2017). It is also challenging to collect data for the use of corridor development, as the data typically has to come from multiple channels (social networks, the Web, crowdsourcing), but also from different sources (countries, institutions, departments). According to Kim et al. (2014) sharing data and information between countries is a special challenge. Governments may not be willing, allowed or technically prepared to share information, or there may be challenges related to language translation and interpretation of text semantics and sentiment, causing originally intended meanings to be obscured. In addition, 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 a data structure is siloed is there is not institutional setup and routine of sharing and collecting data. In addition, private companies have non-public data such as customer data that could be useful, but it is not open data. This can change only if companies find economic advantage in opening their data for the use of different networks.

Considering big data and corridor development, there are also challenges related to usability or other qualities of data. For instance, the data is not up-to-date or the data can experience structural changes over time (Santala 2016). Or, for example, data is not suitable for reuse as some data portals publish data in PDF tables (Hemerly 2013). There is a lack of standardized solutions, software and cross-agency solutions for extracting useful information from discrete
datasets in multiple agencies. On the other hand, much governmental data is global in nature and can be used to prevent and solve global issues. (Kim et al. 2014)

### 3.3 Big data and public policy needs - cases Finland, Estonia and Sweden

Finland, Sweden and Estonia are seen as digital forerunners in many areas of societal development, and everybody seems to be familiar with the term “big data”. It is often mentioned in conference presentations and written into the different strategies and agendas. Nevertheless, the handling of the topic often remains on the level of slogans in the context of territorial development, and there is a need for deeper analysis on the practical development possibilities and challenges. The following sections 3.3.1 – 3.3.3 present short overview on the future plans, strategies and public policy needs related to usage of big data and related services in three countries to deepen an understanding of their public sector big data policy needs.

#### 3.3.1 Big data and public policy needs in Finland

In Finland, one of the main strategies related to big data and public policies has been formulated by the Ministry of Transport and Communication. The Ministry of Transport and Communication started to formulate national big data strategy for Finland in 2013, which were later published in 2014 (Liikenne- ja viestintäministeriö 2014). The Ministry of Transport and Communications’ working group on the use of big data noticed that big data sources are growing rapidly, but the extensive exploitation of data sets is still in its infancy. Big data can change operations and decision-making in nearly all sectors of the society. The working groups suggested measures on how to promote the use of big data and big data solutions in Finland to support decision-making. The benefits supporting e.g. decision-making and economic development were categorized into four categories: public sector, business life, citizens and researchers. In public sector, big data can increase the efficiency of processes and increase quality and transparency of decision-making. It can also bring cost savings and help in allocating and prioritizing resources.

According to the Ministry of Transport and Communication, as well as the Ministry of Economy and Employment (Työ- ja elinkeinoministeriö 2013; Liikenne- ja viestintäministeriö 2014), one of the biggest potentials related to big data and territorial development is related to smart infrastructure, especially on community planning and smart grid related to energy supply and housing. For instance, mobile data can be useful for community planning as the data shows mobility patterns of citizens, as well as mobility differences in different districts and regions such as city centres, neighbourhoods and countryside. This knowledge related to mobility can be useful, for instance, in land-use planning. In addition, big data has great potential related to
transportation. The digital traffic data can support decision-making related to traffic infrastructure planning, as well as support creating new mobility services and applications.

Also, other big data possibilities were brought up in the national big data strategy of Finland (Liikenne- ja viestintäministeriö 2014). For instance, big data creates a great possibility to promote sustainable development. Sustainability indicators can be used to measure sustainable development internationally. Secondly, big data can be useful for promoting economic development. The data can e.g. visualize money flows and interactions between different zones. Open data published by the public sector can produce knowledge that improves competitiveness of companies (Koski et al. 2012). This in turn, can create new businesses and jobs increasing economic development of territories.

One of the national strategies related to big data and public sector territorial development is The Six City Strategy – 6Aika (see website http://6aika.fi). The strategy is implemented through cooperative and experimental projects between the six largest cities (Helsinki, Espoo, Tampere, Vantaa, Oulu and Turku) in Finland. The 6Aika development projects have been based on solution-oriented thematic collaboration rather than geographical region. These cities shared projects, platforms, co-creation models, standards and open data aiming to go towards interoperability and shared services. The Six City Strategy co-operates with national programs related to Big Data, MyData and MaaS (Mobility as a Service). Typical projects and test cases are related to big data and smart mobility, transport, circular economy, education and employment (Antikainen et al. 2016). In the now ended 6Aika Open Data and Interfaces spearhead project, the six participating cities opened up their data while developing and following shared operating models. The target group of the open data spearhead project was companies. The aim was to make the opening up of data a natural part of the normal operation of cities, thus helping them facilitate the creation of data-based services and business.

The results showed that one city and many times even the six largest cities in Finland are too small to act as well functioning innovation platforms. Therefore, there should be co-operation with private sector and with cities from other countries. One of the cases testing this idea is the so-called Open and Agile Smart Cities network (see website http://oascities.org/) that was founded by the 6Aika cities in 2015. The network has already 117 member cities from 24 countries and regions. Most of the cities are from Europe, but there are also members from Latin America and Australia. This network aims at data-based co-operation promoting interoperable standards, solutions and smart services among the member cities. According to Antikainen et al. (2016), 6Aika test cases showed that in public decision-making typical benefits from big data are connected with cost saving and more efficient planning processes. The data supports productivity and measurability of processes, as well as transparency of decision-making. Big data can also increase customer understanding that is useful knowledge in product and service development, as well as in territorial development resource allocation.
One concrete case combining big data and regional development is the Helsinki Region Infoshare service – HRI (see website https://hri.fi/fi/). The HRI service distributes municipal data from the Helsinki Metropolitan Area cities for free use by anyone. For the most part, the data produced by the cities is already public because of the Openness of Government Activities Act. The project aims to make regional information quickly and easily accessible to all. The information is provided for the use of citizens, businesses, universities, research institutions and local governments and may be used freely and at no cost. Procurement data, historical aerial photos and postal code areas are among the most downloaded open data on the HRI service.

According to Holopainen (2016), in Finland, big data has been typically used in modelling unequal regional development; e.g. unemployment and negative net migration, health care and economic development. It seems that in Finland big data and territorial development are mostly focusing on strategies and projects focusing on national or regional level development. At the moment, there are only few policies focusing on international corridor development and interactions of actors between borders. In Finland the corridor development linking with big data and open data, is mostly focusing on the Northern Growth Zone (Santala 2017), as well as Growth Corridor Finland between Tampere and Helsinki. Cities of Stockholm, Turku, Helsinki and cities from the Helsinki Metropolitan Area are active providers of open data for the corridor.

3.3.2 Big data and public policy needs in Estonia
In “Digital Agenda 2020 for Estonia” four goals have been established to support the development of Estonian information society:

- ICT infrastructure that supports economic growth, the development of the state and welfare of the population
- Larger number of jobs with higher added value, improved international competitiveness and higher quality of life
- Smarter governance
- Enhanced awareness of Estonia as an e-state all over the world

The focus for the future will be on creating an environment that facilitates the use of ICT and the development of smart solutions in Estonia in general. The ultimate goal is to increase the economic competitiveness, the well-being of people and the efficiency of public administration. (Ministry of Economic Affairs and Communications, 2013). Measure 2 of the Digital Agenda 2020 is focused on the development of a common service space for the public and the private sector. It is stated in section 4: “Innovative technologies will constantly be analysed and their uptake will be piloted. Actions in this area include analysis of technology trends and carrying
out pilot projects to keep the state information system and common service space up-to-date. This would include, for instance, Internet of Things, advanced analytics, big data, linked open data, augmented reality and privacy enhancing technologies”. Measure 3 in the same agenda examines how to achieve better decision-making with ICT. In section 2, it is stated: “ICT solutions that advance data analytics capacity in the public sector will be developed and taken into use. For example, the potential use of advanced analytics and big data for real-time monitoring of data and making predictive policy decisions will be researched and relevant pilot projects carried out.”

Ministerial Declaration on eGovernment – the Tallinn Declaration signed by all EU Member States and EFTA countries (in Tallinn on 6 October 2017) states: “We will in our countries prepare and implement initiatives to widen and deepen the use of data and analytics (including big data) in our countries to move to data-driven public services and make full use of data for better decision-making” (Ministerial Declaration 2017).

The Statistics Estonia Agenda 2016–2021 states that as demand for statistics in the society is increased, customers expect information about more and more phenomena and they expect to get it faster and faster. Next to the development of the existing products, Statistics Estonia plans to start using new data sources (data generated by activities of private and governmental sector, including big data like mobile positioning, card payments, sale systems of shops, Internet postings, speed cameras etc). This assumes better cooperation with research institutions, participation in international cooperation networks, and shall result in new methods and technologies. In the near future, Statistics Estonia plans to start using the big data to develop methodology for using the data about electricity consumption in households. Furthermore, it is planned to start using mobile positioning as one possible option to get additional data for population census. (Statistics Estonia 2017a). Plans to start using big data in governmental statistics is described more thoroughly in Quarterly bulletin of Statistics Estonia (Statistics Estonia 2017b): The countries participating in the European Statistical System (ESSnet) have initiated the ESSnet Big Data project to find opportunities for using big data in the production of statistics.² The participants of the project include 22 partners from 20 countries, incl. Statistics Estonia. The results are published on the website of the project (ESSnet 2017). The project was launched in February 2016 and it ended in May 2018. The project includes eight subprojects related to a particular data source or methodology, and, in addition, management and communication subprojects. The first five subprojects are pilot projects that deal with the processing of real big data. (ESSnet Big Data 2018)

E-Health (The agenda of the Estonian E-health Foundation) is a cooperative network of health services based on digital solutions, which supports the promotion of consistency of the treatment, timeliness of the services, and patient satisfaction. The Digital Health Information


ESPON 2020
System contains health data entered by the hospitals, family physicians and other providers of health services. Health data concerning each person is collected in the system and everyone has an opportunity to see their health data from the patient's portal, digilugu.ee. (Ministry of Social Affairs 2017). Measurable goals of the activities planned in the E-Health strategy are: reduction of the number of illnesses, more effective time use by health professionals, reduction of the number of avoidable visits, as well as avoidable and repeated hospital treatment. According to the agenda, the E-Health plans to start using the big data to assess the efficiency and performance of the health care system in Estonia (Sotsiaalministeerium 2017 - link).

There are also other agendas and strategies emphasising public policies and big data. The agenda of the Estonian Unemployment Insurance Fund (Töötukassa) 2018 – 2021 mentions as one of the objectives the enhancement of the use of large-scale unemployment funds in forecasting and evaluating performance developments in the workforce (Eestitöötukassa 2017). The Ministry of Rural Affairs is funding the establishment of the Electronic System of Agricultural Big Data forecast in the Estonian Rural Development Plan (ERDP) for 2014–2020. The planned system supports linking existing databases for data analysis, analytical models and practical applications. This big data system also supports simplifying data collection and accessibility. (Ministry of Rural Affairs 2014.)

Agenda 2019 – 2022 of the Ministry of Justice emphasises that the protection of privacy is becoming increasingly important in the information society. On the other hand, it is necessary to find solutions to support the knowledge-based economy and innovation with data mining and big data analysis. Data collection and use in both the private and public spheres must be clearly regulated. Ensuring and developing the regulation of effective protection of personal data suitable for the information society is the responsibility of the Ministry of Justice. To this end, it is necessary to modernize the general data protection rules and to introduce uniform data protection rules in various areas of life. (Ministry of Justice 2018). The 2017 annual report of the Bank of Estonia notes that in compiling its statistics on travel services, the bank has for the past decade used data from mobile positioning that are collected using a methodology developed jointly by the central bank, the University of Tartu, and OÜ Positium LBS. (Eesti Pank 2017.)

The Government Office and the Ministry of Economic Affairs and Communications decided that Estonia will have an Artificial Intelligence strategy. Its term is April 2019 (Government Office 2018). Estonian open-data government initiative, the Open Data Portal of Estonia, provides a single point of access for general public to unrestricted public-sector data with the permission to re-use and redistribute such data for both commercial and non-commercial purposes. On the goal of opening data it is to encourage migration to future technologies such as Linked Data, Big Data and Internet of Things. (Ministry of Economic Affairs and Communications 2018.)
3.3.3 Big data and public policy needs in Sweden

According to the Global Open Data Index by Open Knowledge International (2016), Sweden is ranked as 21st among the listed 94 countries in the index and is therefore considered to be opening its data sources to the public at a good pace. The country is in the forefront of the digital transformation, and the above-mentioned rather slowly updated ranking does not depict the current status or aspirations of the data-driven policies in Sweden. The ambitions of the Swedish government encompasses both strategic work harnessing big data throughout policymaking together with concrete measures in opening new sources of information to the public (Government Offices of Sweden 2017). In 2016 the Swedish government assigned the National Archives to step up measures to further develop digital channels, gather open data interfaces for various users and give support to authorities opening up public sector information (PSI) even more. This has been carried out since by the new initiative called Open Data and Public Sector Information. The development of broader access and support in publishing relevant digital information sources to users has included the new national open data portal (see https://www.oppnadata.se). This new platform represents a foundation for anyone to find access to relevant data sources about all fields of the society e.g. economy, healthcare, civics, transport, regions, governance and environment.

The Swedish government is the owner of a leading ICT-research institute on applied information and communication technology called Swedish Institute of Computer Science (RISE SICS, see https://www.sics.se/). The research of the institute comprises the most relevant scientific issues and application purposes in a modern ‘big data’-driven environment. Focus of the research that SICS conducts include such areas as data analytics, machine learning and optimization, data intensive computing and cloud platforms, network technology and architectures, security and UX to name a few. Hosted by SICS, the Swedish Big Data Analytics Network (2013) published a joint report called Big Data Analytics - A Research and Innovation Agenda for Sweden. This report takes a look at the new value chains and evolving technology both in global and Swedish contexts and makes proposals for how Sweden can remove barriers to big data activity. It lists also the members of their open network which is a consortium of key actors from the field of research, governance and business.

As stated above, there are numerous programmes and projects by Swedish authorities about open data and related infrastructure in progress. In addition to the already mentioned, there is the Digital Strategy for Sustainable Digital Transformation in Sweden by the nation’s Ministry of Enterprise and Innovation (2017). Moreover, the ministry is working to achieve its Completely Connected Sweden by 2025 (Ministry of Enterprise and Innovation 2016), a national broadband strategy. Sweden also has a National Geodata Strategy and workplan for 2017-2020 with the goals of making Geodata open, usable, accessible, and having well-developed collaboration. (Lantmäteriet 2017).
3.4 Big data needs defined by the stakeholders

It is now widely understood that due to the rapid growth in data sources, there are going to be big changes in policy-making and corridor-based development. Therefore, already at the beginning of the project the aim was to specify, what are the concrete policy needs defined by the stakeholders to formulate a solid basis for further actions. Firstly, this was done by framing the concrete needs with stakeholders of the project. Secondly, policy needs were identified by reviewing the research and policy literature on data-related needs in territorial development.

In a workshop, organized in Brussels on 8th May 2018, project stakeholders were asked to think about the public policies that could benefit most from the big data (at least three most important policies in territorial development), and the types of big data that could be useful for serving territorial policy development (big data needs, gaps and challenges). As a result, the stakeholders identified three concrete policy contexts, which will serve as a starting point for the targeted analysis. The most important dimensions that the stakeholders identified in the workshop were:

- Infrastructure planning, especially transportation
- Regional economic development
- Land-use planning

3.4.1 Infrastructure and connectivity planning

Considering infrastructure, stakeholders stressed that promotion of connectivity and investments in transportation infrastructure is an important policy area in territorial development that could benefit from big data. The investments can be more efficiently directed based on the more accurate knowledge e.g. about the flows of people and goods, which are at the core of corridor development. Especially, mobility data is needed at the corridor level, not just in a city or region. This is in line with policies in regional development, where it is becoming more typical to combine physical infrastructure with so-called soft infrastructure like networks and knowledge to increase innovation capacity (Morgan 1997). A corridor ties a region together while connecting it to other regions, thus fostering regional development.

3.4.2 Economic development

Policies related to economic development could benefit greatly from big data. Big data can enable new business and jobs. New types of data are needed to support regional and territorial economic development, enhancing partnerships with private companies and more efficient allocation of investments. Knowledge about different territorial networks related to, for example, organizational interlinkages and project networks, as well as interregional collaboration in education and between companies would be useful for actors at various organisations and scales. In addition, development of education and culture contribute to economic development. For instance, knowledge on the mobility of students in the corridor would be useful for planning collaboration in higher education.
3.4.3 Land-use planning
Land use is another policy area that could benefit substantially from big data. Knowledge about the mobility of people, information and goods is again needed for future land use planning. In addition, safety related policies, as well as environmental sustainability policies were brought out as policies that could benefit from big data. On the other hand, those policies are strongly interlinked with land-use and infrastructure promoting connectivity. Stakeholders also brought out that there is enough data on migration, but there is a data gap on detailed data on international migration. While much is known about how many people migrate to a state, much less is known about in migration and out migration. A challenge is also that as the operational environment is developing so quickly, the data is rapidly outdated. For instance, when using mobile positioning data, the age cohorts should be taken into account as they might have different habits in using mobile phones (e.g. differences in the usage of social media applications or making actual phone calls by age groups). They noted that a challenge in using big data for policymaking is low technical know-how for utilizing big data in public administration.

3.4.4 Data gaps defined by the stakeholders
Also, data gaps were identified by the stakeholders of the project. The main challenge is that there is only a very limited amount of data on interactions, like company or social media networks, municipal cooperation or email traffic. This data has typically poor spatial or temporal resolution or has poor data accuracy. Project stakeholders emphasized that they need more information on the following six interaction categories: physical, digital, administrative, organizational, intellectual, and social. Datasets related to these six categories will be emphasized in the further analyses, as well as in the categorization of available data sources.
4 First categorisation of available data sources

The first categorisation of available data builds on the tentative mapping of available data sources, which is done by utilising the conceptual framework developed in the project. The first categorization takes into account both the many governmental levels of policy making and the nature of data and its potential contexts of utilization.

The available data can either enable territorial analysis at the level of the whole European territory, or it can deal with only parts of it, emphasizing the significance of scalar categorization of data. This categorization enables diverse stakeholders involved in territorial policymaking to easily perceive the utilization possibilities related to different datasets. In the categorization based on the characteristics of the data, the conceptual framework is utilized to evaluate the potential uses of the data (e.g. from social media platforms, mobile phones and traffic) in the particular context of corridor development. Furthermore, categorisation is done in a manner that takes into account the existing statistical data and its deficiencies in integrated territorial development, which is further elaborated in the needs analysis. The availability of data is influenced by a number of factors such as the ownership and storage of the data.

The development of big data has been mainly influenced by technological, social and economic factors. For example, the growth of smartphones has contributed significantly to the number of sensors carried by people. The current basic smartphone can measure and collect data from its acceleration to local atmospheric pressure and the magnitude of the magnetic field. The GPS receiver on the phone also makes it possible to locate it accurately. The popularity of smartphones and the applications built for the metrics have contributed to the collection of versatile data at a global level. Furthermore, the accumulation of big data has also been influenced by the transition to social media, which enables it to be recorded, tracked and analysed. Many companies monitor users’ feedback on their products in social media and seek to fix or improve potential shortcomings. Social media can also be used as a tool for societal development. For example, Google’s data from used search terms can be used to predict, for example, the occurrence of seasonal influenza (Yang et al. 2015).

Social media data as well as mobile positioning data are examples of new available big data sets that can be utilised in exploring the functionalities of European growth corridors. Utilisation of positioning data widens understanding about the daily mobilities of citizens along the corridors that can be utilised in land use and transportation planning, as recognized in the needs analysis. In addition, social media data can be used in exploring the interrelations of diverse actors along the corridors. Overall, mass and sensor data can allow the exploration of cross-border developments across Europe. Opening up publicly produced information is one of the processes that makes society’s information available to everyone.

Altogether, development of the growth corridor is a product of complex systems in interaction with each other. Therefore, a variety of data is needed to build robust models and meaningful insights useful to policy-making contexts. A guiding principle for identifying useful data sources is to seek ones that represent interactions among, for example, people, businesses,
organizations, vehicles, goods, and transportation infrastructure. There are selections to be made within any datasets used, in order to identify and narrow down the relevant indicators suitable to the exact research questions of the data analysis.

In the further comprehensive analysis of available data sources, the conceptual framework presented in the chapter 2 is utilised to map and categorise the different sources of data. The following typology can also serve as a guide for searching for available data based on the data objects found in each layer (table 2).

Table 2. Categorisation of available data sources.

<table>
<thead>
<tr>
<th>Types of Interactions (ToI)</th>
<th>Data Objects relevant to ToI</th>
<th>Data Sources relevant to Data Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physical</strong></td>
<td>Highway E18, Roads Intersecting E18, Land, Buildings, Parking Lots, Vehicles (cars, trucks, buses, etc), Freight/Cargo, Materials, Individuals</td>
<td>Transport Maps; Land use Zoning Maps (e.g. could every parcel between St. Petersburg and Oslo be retrieved?) Vehicle Registries; Port/trucking Manifests; Individual Level Spatial Movement; Vehicle geo-spatial movement.</td>
</tr>
<tr>
<td><strong>Digital Communication</strong></td>
<td>Wired Infrastructure; Wireless Towers; Wireless Internet Access Points; Internet Connected Devices; Internet Connected Vehicles; Navigation Systems; Websites for companies, organizations, or governments in corridor.</td>
<td>Internet Providers; Telecom Companies; Webmail Providers; Information Technology Departments; Infrastructure Maps; Geolocation transponders; Geolocation satellites.</td>
</tr>
<tr>
<td><strong>Administrative</strong></td>
<td>EU Administrative Entities, National Governments and Agencies, Provincial/Regional Governments, Municipalities, Park Services (national, regional, municipality), Neighbourhood Associations (if existing)</td>
<td>Open data from government and agencies; open data from civic-, housing-, environmental- (park), and economic services; news and blogs from all of the above.</td>
</tr>
<tr>
<td><strong>Organizational</strong></td>
<td>Large Businesses, SMEs, Non-profit Organizations, Societies, Public Sector</td>
<td>Business Registries (for business names and ownership); Real Estate Information (for office and facility locations); Business Websites; City planning.</td>
</tr>
<tr>
<td><strong>Intellectual</strong></td>
<td>Universities; Research Institutions, Research Groups, Think tanks; Consortiums; Academic Writings (e.g. articles, essays, theses); Academic Networks (e.g. ResearchGate).</td>
<td>DBs of research funders; Bibliographic DBs; University websites; Academic Blogs.</td>
</tr>
<tr>
<td><strong>Social</strong></td>
<td>Homes; Inhabitants of Homes (e.g. families, couples, residents); Workplaces; Relationships among People (who interacts with who, and how).</td>
<td>Population Registry (Finland) or equivalent; Tax Office; Social Media (connections).</td>
</tr>
</tbody>
</table>

3 Notable new restrictions exist for these first three potential data sources in the EU because the GDPR is now in force, providing significant protections to individuals.
4.1 Plan for conducting prospective analysis of new datasets at European level

Data ecosystems continually evolve and new datasets become available as various factors change or emerge. This evolution is occurring at many levels. At the technological level, new capacities to sense, transmit, and store data emerge as technologies mature and become readily available. At a governance level, new programs to deploy such technologies in pursuit of other goals – for example, becoming a ‘smart city’ – actualize new data sources in conjunction with other governance aspects that seek to control or open access to the data source. Trends in this area include the ‘open data’ movement. At the consumer level, the uptake of various Information Communication Technologies and the integration of ICT with other types of technology such as automobiles, bicycles, currencies, infrastructures, and buildings produce new varieties of data streams. This is only a brief list; other levels can be active as well. The goal of prospective analysis of new datasets at the European level is to identify already existing datasets that may become more available or usable as well as emerging datasets that may become commonly available in the future.

A key to searching for European level and transnational data will be to seek out data sources that cut across the specific peculiarities of datasets produced in various nation states, regions, and municipalities. Examples of such cross-cutting data sources could include those:

- maintained by large corporations (e.g. Uber, Google, or Facebook);
- available in open data and standardized data formats;
- and maintained by (or for) EU agencies or actors.

At this point of the research project, a key task is to also look into any potentially useful datasets which have significant differences in format or quality between nations, regions, and municipalities. It is important to look not only for the already available transnational and EU-level datasets, but also to compare datasets from multiple regions while looking for ways to harmonize, connect, or otherwise utilize datasets that are slightly different in their characteristics. An approach for doing this will be to search data portals managed by the countries represented by stakeholders (Estonia, Sweden, Finland), as well as regional and municipal data portals along the Northern Growth Corridor. An expected outcome of this activity is to identify opportunities for greater data harmonization and repeatable methods for utilizing datasets with significant differences in terms of their characteristics. The repeatability of these methods can be tested by assessing data compatibility among a few countries outside of the targeted growth corridor. For example, a sampling could be taken from the TEN-T Atlantic Corridor assessing the availability and compatibility of data from France (a leader in open government data according to OECD 2017), Spain and Portugal.
The analysis will be linked to the conceptual model proposed in the above chapter. It will begin by mapping available data and categorizing it according to the model. During this data mapping process, attention will be paid to how well the found data sources fit the model and how suitable the model is for evaluating the data sources – i.e., it can be further developed. An expected result is that specific details of the model – such as various types of restrictions or types of flows and interactions – will evolve during the process. The conceptual model will help the data mapping team locate a wider variety of data.

In practice, a spreadsheet will be used for the data-mapping process. The primary data entity in the spreadsheet will be the datasets. Each row will contain one dataset, some metadata about the dataset (e.g. datasource, URL, format) followed by the characteristics of Space, Interactions/Flows, Scale, Methods, Restrictions, and Future Potentials. The proposed fields for this spreadsheet are:

- **Data Source**
  - Name
  - Brief Description
  - URL
  - Format(s)
- **Space** (checkboxes -- some data may be in more than one space)
  - Physical
  - Digital
  - Social
- **Flows/Interactions**
- **Scale** (EU, Transnational, National, Regional, Municipal)
- **Methods**
- **Restrictions**
  - Ownership-related (Open Data; Pay-to-use; Closed Data (unavailable, [why]))
  - Law-related (Prohibited by national law; Prohibited by EU directive)
  - Owner policy-related (Prohibited by company/org policy)
  - Practicality-related (not available because organization hasn’t prioritized making it available)
- **Future Potentials** (ways it can be imagined this data might be used)
  - As applied to Stakeholder Policy Contexts
  - As applied to other corridor development tasks & territorial development

When a significant number of data sources have been mapped, the rows will be sorted by various characteristics as researchers look for patterns in data themes as well as gaps. The resulting data map will support the three cases of the overall research project and could serve as a resource for Hackathon participants.
5 Plan for identifying good practice examples

Good practice identification is used in this project to support policy learning, which, according to a definition from the Innovation Policy Platform (World Bank & OECD 2013), \textit{refers to the ways in which policy systems generate and use knowledge about the rationales, design, operation and impacts of policies and policy mixes. It takes place at all points in the policy cycle: direction-setting and orientation, policy mix definition, policy mix delivery, and monitoring and evaluation. As such, policy learning demands the organisational capabilities to frame issues in terms of problems and solutions, to take a holistic view and to anticipate future needs.}

Good practices related to utilising big data in territorial policy and in particular corridor development are identified in the project to support policy innovation and the continuous enhancement of quality in policy development. Learning from “what works” has for long been a central approach in policy development (e.g. Sanderson 2002). The good practices in data utilisation can be used to support policy learning in different parts of Europe and in different scales of policy making, regardless of the fact that data utilisation is also restricted by the context specific preconditions related e.g. to data availability and legislative framework. Experiences of other regions and actors can be used to stimulate one’s own thinking and tailored policy formulation in diverse European contexts. For example, the case study in Estonia benefits big data related policy making in Finland and Sweden due to the wide utilisation experiences about the possibilities of mobile positioning data. Overall, learning from good practices increases actors’ understanding about policy alternatives and foreshadowing of forthcoming issues, as well as critical evaluation of one’s own policy development practices (see, e.g. Bulkeley 2006).

However, attention has to be paid to the evaluation methods in good practice identification, as good practice reporting does not necessarily lean on long-term impact evaluation nor comprehensive understanding about the mechanisms that bring about change in social systems. As good practice lessons are increasingly drawn from policy piloting a critical evaluation of their evidence base is needed to support effective policy learning (see Sanderson 2002). Altogether, challenges related to policy learning between European regions have been identified (e.g. Radaelli 2008) that have to be taken into account in good practice dissemination.

Databases and platforms such as European Commission’s policy learning database are utilised together with research and policy literature review to get a comprehensive understanding about the current situation in big data utilisation in territorial development. The database has been developed to promote policy learning, and contains case studies, ex-post evaluation reports and other project summaries that can be utilised in identifying good practice examples. Furthermore, The Interreg Europe Policy Learning Platform has been recently launched to promote policy learning from the expert validated cases funded by the Interreg programme.
The conceptual framework is utilised to guarantee the comprehensive approach in good practice identification. Good practices are identified at different scales of territorial development, dealing with local, regional, national and European level practices as well as cross-border and corridor related questions. Even though case studies utilise different data sets in different institutional contexts, good practices can be identified that can benefit other countries. An important part of the case studies and the data hackathon is to investigate also the cross-border possibilities and challenges in data utilisation, which acts as a basis for the identification of good practices also at the level of transnational corridor development at the European level.

The Northern Growth Zone is by Scandinavian and Baltic measures densely populated, and the challenges are similar in stakeholder areas (e.g. pressure to develop sustainable mobility solutions, increasing cross-border freight flows). It is important to build cooperation to develop transnational policy learning and universal solutions, and to benefit from good practices. It is crucial to strengthen the cooperation between public authorities, research organizations and enterprises along the corridor to develop evidence-informed solutions for sustainable transportation that cannot be planned in organizational or disciplinary silos.

After identifying good practices in data utilisation, policy proposals are formulated for growth corridors and regions across Europe in general. Policy proposals focus on promoting the use of new “big data” sources and to overcome policy and legal challenges in their endeavours to meet strategic policy and development objectives related to growth corridors and cross-border cooperation. The conclusions of this task form an important part of the “Practical guide” that ensures transferability of the Targeted Analysis results to other European regions.
6 Conclusions

The inception report presented the results of the first phase of the Big Data & EGC project, which will continue by further analysing and categorising available big data as well as by conducting three big data related case studies in stakeholder territories.

The stakeholders of the project represent the countries – Finland, Sweden and Estonia – that are forerunners in many areas of digital transformation. Countries have launched several initiatives to widen the use of big data and analytics. However, as stated in the needs analysis, although the importance of big data is well identified in different policy areas in the stakeholder territories, there is still a need to move from strategies to practices in big data utilisation in many policy areas relevant to territorial and corridor development. In all stakeholder territories, big data related strategies exist, and potentials are emphasised, for example in the following areas relevant also for corridor development: more efficient planning processes supporting decision-making; data supporting smart solutions and infrastructure, mobility and transport; data indicating and promoting sustainable development and circular economy; data creating better customer understanding, useful for product and service development, and data describing the constantly changing developments in the economy as well as education and employment. In the particular context of corridor development, new data sets are needed to increase the understanding about the interregional and transnational functionalities.

However, there is still a great amount of work to be done in standardizing the open and big data practices. Various factors have been identified to act as barriers to release open data in the corridor development, the main barriers being the lack of policies, practices and resources. In addition to the lack of standardization, the biggest challenges in re-using existing open data are the lack of real-time data, insufficient number of actors, as well as the lack of knowledge and poor awareness (Santala 2016). In addition, a challenge with big data is that there may be significant distortions that need to be corrected. In some cases, the partial normalization of the incomplete or inaccurate big data set can act as a useful correction measure. Moreover, the use of statistical comparisons, peer sets or other viable points of reference may facilitate the analysis and use of incomplete big data sets. Correction and normalization of the existing open data represent common and useful tools to harness big data, before more accurate data is published or new interfaces are opened.

In the context of transnational growth corridors, the different regulatory frameworks in different countries also challenge the ways in which big and open data can be used by both public and private entities. More and more accurate data gathering causes problems with privacy. One of the biggest challenges associated with sensor data is respect for privacy, without losing the benefits of data. The differences in privacy laws in the area of the Northern Growth Corridor place limitations to integrated territorial policymaking due to the restrictions in collecting, sharing and utilizing personal information. Altogether, although big data has certainly established its place in policy development, its wide utilization by the government takes time due to various institutional barriers and bottlenecks in capacities (Giest 2017).
The next phases of the Big Data & EGC project aim at further analysing the challenges as well as the possibilities to overcome them to promote big data utilisation in territorial policy development. A key question is, for example, “How could big data support the planning of land-use and infrastructure that has long-term impacts on the future?” Predictive models used to analyse big data are largely locked to data describing the present or the past behaviour and interlinkages. As most policy decisions concerning the development of a corridor have long impacts on the future, futures approaches that go further than the past-bound limitations of big data will be essential. Viable solutions to this challenge of producing futures-oriented insights are important to pursue in the next phases of the project. For instance, simulations built from big data may be one such way to bring futures perspective to decision making. Another means would be to include participation by stakeholders, inhabitants, and users of a growth corridor imagining potentials for the corridor based on futures thinking. The project will take on the task of producing recommendations for how to overcome past- and present-bias in big data for future-oriented policy making.
References


ESPON 2020

Santala, V. (2016) Avoimen datan kehitys pohjoisella kasvuvyöhykkeellä. Haasteet ja mahdollisuudet sekä merkitys kasvuvyöhykkeen kehittämiselle. (Translation: Open data development in the Northern Growth Zone: Challenges and potential of open data in the development of regional development zone) University of Turku Faculty of Mathematics and Natural Sciences, Department of Geography and Geology.


ANNEX 1 – PRESENTATION OF INCEPTION REPORT TO PROJECT STEERING COMMITTEE

Big Data & European Growth Corridors (BIG DATA)

Potentials of big data for integrated territorial policy development in the European growth corridors

Targeted Analysis activity for ESPON EGTC.

// 2nd Steering Committee Meeting (Inception Report)

- 11 July 2018, 10:00-16:00, Clarion Hotel Helsinki

Turun yliopisto University of Turku UNIVERSITY OF TARTU
Agenda

1. Welcome and presentation of participants
2. Presentation of the inception report by the research team
3. Discussion and feedback from stakeholders
4. Feedback from ESPON EGTC on the report submitted
5. Lunch break
6. Presentation of case studies
7. Discussion about the next steps
8. AOB
9. Bilateral discussions between SC members and the research team to prepare next steps

Big Data & EGC – Potentials of big data for integrated territorial policy development in the European growth corridors
Inception Report, Version 29/6/2018

1 Introduction ........................................ 3
2 Description of conceptual and methodological framework .................. 5
   2.1 Conceptual framework for addressing the research need ............... 5
   2.2 Methodological framework ..................................... 7
3 Overview of stakeholders’ public policies and needs ......................... 9
   3.1 Public policy considerations for big data ................................ 9
   3.1.1 Better governance ................................................. 9
   3.1.2 Data driven innovations ......................................... 9
   3.1.3 Different needs in public and private sectors ..................... 10
3.2 Big data and corridor development ...................................... 10
3.3 Big data and public policy needs – cases Finland, Estonia and Sweden ......................................................... 12
   3.3.1 Big data and public policy needs in Finland .......................... 12
   3.3.2 Big data and public policy needs in Estonia ....................... 14
   3.3.3 Big data and public policy needs in Sweden ...................... 14
3.4 Big data needs defined by the stakeholders .................................. 17
   3.4.1 Infrastructure and connectivity planning .......................... 18
   3.4.2 Economic development ......................................... 18
   3.4.3 Land-use planning ................................................. 19
   3.4.4 Data gaps defined by the stakeholders ........................... 19
4 First categorisation of available data sources .................................. 20
   4.1 Plan for conducting prospective analysis of new datasets at European level .......................... 22
   4.2 Plan for identifying good practice examples .......................... 24
5 Conclusions ............................................................................. 26

References .................................................................................. 28
Conceptual framework

<table>
<thead>
<tr>
<th>TYPES OF FLOWS/INTERACTIONS</th>
<th>AVAILABLE DATA</th>
<th>SCALE</th>
<th>METHODS</th>
<th>RESTRICTIONS</th>
<th>FUTURE POTENTIALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g. Commuter flows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Transportation flows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Tourism flows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Administrative network</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Organisational networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Interact networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Social networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Service provision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overview of stakeholders’ public policies and needs (1/2)

Public policy considerations:
- Better governance
- Data driven innovations
- Different needs in public and private sectors

Big data and corridor development – lots of potential:
- Support public policy making concerning growth corridors, adequate evidence-base help support public territorial decision making
- Typical areas: infrastructure plans, economic growth, sustainable development, transportation, smart grid...
Overview of stakeholders’ public policies and needs (2/2)

Big data and corridor development – challenges:
- Corridors have been studied very little – data gaps – need of new data sources and methods of analysis
- Data from multiple channels and multiple sources, spatial dimensions often missing from big data components
- Governments may not be willing, allowed or technically prepared to share information
- Language translation and interpretation of text semantics and sentiment

Big data and public policy needs – Finland, Estonia, Sweden:
- Lot’s of big data related strategies
- Only few policies focusing on international corridor development and interactions of actors between borders

First categorisation of available data sources

- The first categorisation takes into account:
  - Many governmental levels in policy making
  - Key characteristics of the data
  - Potential contexts for using the data
  - Scalar aspects (all of Europe, or only parts of it)
  - Socio-technical change has led to a world with a vast number of sensors that collect all kinds of data (e.g. smart phones; smart city sensors, etc.)
  - A data mapping process will be used to identify what data exists as well as data deficiencies.
Plan for conducting prospective analysis of new datasets at European level

- **Goal** is to identify:
  1. existing datasets that may become more available or usable.
  2. emerging datasets that may become available in the future.
- Seeking existing cross-cutting data sources:
  - Maintained by large corporations
  - Available as open data and/or in standardized formats
  - Maintained by (or for) EU agencies and actors
- Seeking regional and municipal datasets:
  - Determine ways to harmonize, connect, or utilize potentially different datasets.
  - Test these approaches by assessing a different corridor (e.g. France – Spain – Portugal)

Plan for conducting prospective analysis of new datasets at European level

- Apply data model to locate a wide variety of data, evolving the model through use.
- In practice datasets will be mapped in a spreadsheet tracking their: metadata, space (physical, digital, social), flows/interactions, scale, methods, restrictions, and future potentials.
- Once a critical mass of datasets are mapped, the research team will analyse the total set, looking for patterns and gaps.
Plan for identifying good practice examples

- Databases and platforms (e.g. European Commission’s policy learning database)
- Research and policy literature review
- Conceptual framework
- After identifying good practices in data utilisation, policy proposals are formulated for growth corridors and regions across Europe
- Attention has to be paid to the evaluation methods in good practice identification

Environmental Scanning

Goal: Identify key driving forces, weak signals, future potentials, and obstacles affecting the operational environment in which big data can be applied to policy making in European Growth corridors.

We propose we will focus first on the Northern Growth Corridor with an aim of producing a repeatable process for other European growth corridors.
Defining Operational Environment as Northern Growth Corridor

Environmental scanning focusses on the territory surrounding NGC. The scanning will systematically focus on the links (A-F). Variations are expected in terms of what driving forces, future potentials, and obstacles are most relevant to each link.

Conclusions (for now...)

- The inception report presented the results of the first phase of the Big Data & EGC project
- The stakeholders of the project represent the countries – Finland, Sweden and Estonia – that are forerunners in many areas of digital transformation
- However, there is still a need to move from strategies to practices in big data utilisation in many policy areas relevant to territorial and corridor development
Presentation of Case Studies

Research teams presented their own slides. (See separate annexes.)

---

Next steps

- Comments on Delivery 1
- Delivery 2, dd 8 October 2018
- Case studies + hackathon
- 3rd steering committee meeting? In London (13 or 15 Nov)?
- ESPON Scientific Conference 2018, 14 November 2018
- What else?
Delivery 2: Interim report

- the final overview of stakeholder’s public policies at growth corridor levels that could benefit from new ‘big, open or soft’ data provision
- the final typology of new available data sources
- first overviews of: the European analysis on data collection, processing and analysis challenges; state-of-the-art of data governance (legal, cultural, economic, technological and any other relevant aspects), and of; potentials, challenges and obstacles for big data driven future spatial policies at growth corridor level, and more generally, at the level of European regions.

// Thank you

Helka Kalliomäki, Centre for Collaborative Research, University of Turku
Ira Ahokas, Finland Futures Research Centre, University of Turku
Nicolas Balcom Raleigh, Finland Futures Research Centre, University of Turku

This presentation will be made available at: www.espon.eu
ANNEX 2 – FRAMEWORK FOR CASE STUDIES 1 & 2

Big Data & European Growth Corridors (BIG DATA)

Potentials of big data for integrated territorial policy development in the European growth corridors

Targeted Analysis activity for ESPON EGTC.

// framework for case studies (data & analysis tools)
Presentation of the ongoing case studies 1/2

General observations

- UTU FT-team focus set towards filtering the usable data sets and preliminary analysis
- Main criteria of the viable data sources: accessibility to comparable crossborder data
  - e.g. national authorities' data on economy, traffic, freight transport etc.
- Novel examination of the potentials of big data via the NGC
  - Measurability of volumes, movement and economic factors combined with new analysis and machine learning
  - NGC's role as a part of the Pan-European society by means of measurability
- Key driver is future usability – to provide knowledge and base for applications to public policy-making
- Tasks conducted and ongoing as planned: Overview; Analysis; Categorisation.

Presentation of the ongoing case studies 2/2

About the tasks and phases

- Key task: targeting most promising and useful data sets \( \rightarrow \) APIs, open big data
- Basic selection of the themes made based on assignment \( \rightarrow \) obvious big data & machine learning potential found
- Preliminary data gathered for experimenting \( \rightarrow \) some analysis made on selected data
- Ongoing next steps: enriching the selected data sets by combining cases 1 and 2

Challenges

- Accessibility and availability issues encountered (e.g. age and format of the data)
- Significant differences in the data sources and sets between EU-countries, national regions, and municipalities

Helinj, 11.7.2018
BIG DATA & EGC

- Unifying view to NGC and other European corridors based on big/open data
- Understanding the current situation and needs for the development
- Data-driven spatial visualization
- Experiences on quality of the data and elaboration of generalization possibilities
- Possibility to simulate the effects of decision making on growth corridors

Helsinki, 11.7.2018
Case study 1: *Automated traffic measurement data*

Focus
- Gathering and examining automatically created data sets (from EST, FI, SE)
- Harnessing available traffic measurement system (TMS) data to corridor analysis
- Mathematical and spatial analysis (theory section to follow)

Description of the data
- Vast number of measurement points → Wide-scale and long trend analysis possible
- Large-scale data set produces comprehensive view of the traffic patterns
- Currently collecting data via selection of the E-18, between 2010-2017
- APIs and other sources of information available

---

**Preliminary analysis of the case studies**

The plot of count of F10 to F4. Color shows data is about F10. The data is in a range of F13, which ranges from 1 to 1.
Case study 2: Analysis of regional growth and impact to corridor development

Focus
- Exploiting the existing (traditional) data sources by enriching the data sets with big data analysis (from EST, FI, SE)
- Experimenting on, and modelling the economical impact and growth potential via NGC -> provides experience of theory/method transferrability and generalization quality
- Strong future oriented development focus -> smart city applications
- Mathematical and spatial analysis; attempt to combine regional to cross-border economic big data

Description of the data
- Up-to-date data in applicable form
- Data sets provided by the national authorities (and Eurostat)
- APIs and other sources of information available

Helsinki, 11.7.2018

Case study 2: Analysis of regional growth and impact to corridor development

Description of the viable data, ctd.
- Zip-code based socio-economic data, e.g.:
  - Age cohort, life-span data
  - Education
  - Professional transit
  - Income distribution
  - Economic dependence
- Regional open data, e.g.:
  - Business sector development
  - Statistics of housing, land use, startups
  - Targeted EU funding, other regional investments

Helsinki, 11.7.2018
Preliminary experiments of the case study 2

Example: Structuring the base for economic relevance oriented experiments. Mapping the relevant infrastructure and economic clusters via the NGC.

Focusing on cross-border traffic, custom data. Evaluating the emphasis of economic transactions among domestic traffic (TMS).

---

Preliminary experiments of the case study 2

Example: Close-up/selection of the E-18.

Visualization of the NGC route and it’s imminent economic sphere of influence via Stockholm – Orebro – Oslo.

Example: X/Y-locations of TMS points brought from a matrix to map. E-18 route from FIN-RUS border to Naantali harbour.
Case studies: Theoretical approach

- 1) Hidden Markov models (HMM):
  - Unobservable (hidden) state $x(t) \in \mathbb{R}^d$, $t \in \mathbb{N}$ in only one location!
  - Can be e.g. a ratio of working age population “at home”
  - Observable (measured) state $y(t)$ (e.g. vehicles passing a LAM point)
- Markov property: only $x(t)$ is needed to predict $x(t+1)$
  - Also: no memorization of the history ..., $x(t-2), x(t-1)$ needed
  - In our case (tested by prediction performance) the Markov property may not hold

![Diagram showing Hidden Markov Model](example.png)

Extensions of HMM

- 2) State space models (SSM)
  - HMM with a continuous hidden state $x(t)$ ($x(t)$ has real distributions)
- 3) Spatially distributed SSM == spatial linear dynamical system (SLDS)
  - Only a limited locality coupling (filtering windows, radial basis functions etc.)
- 4) Stochastic differential equations (SDE)
  - Like SLDS but more options as physical analogies for locality coupling
Extensions of HMM

**Approach:**
- Smoothing by radial basis kernel → elimination of noise → generalization potential
- Difference grid and simple assumptions of some econometric vars being constant over a district (not all vars)

**NOISY VAN DER POL OSCILLATOR**

Effect of increase in noise

Spatial discretization by locality areas

- **green:** the road network
- **circles:** function space basis of RBF
- **polygons:** localities
- **squares:** LAM points

**z(x,t): hidden variables**

<table>
<thead>
<tr>
<th>time t</th>
<th>z(p1)</th>
<th>z(p2)</th>
<th>z(p3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>time t=1</td>
<td>z(p2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**y(x,t): observed traffic flow, economic vars.**

\[
\frac{dz}{dt} = f(z,p) + A(z(p1),z(p2),z(p3))\]

**Locality coupling:**
- The function \(z(p1)\) satisfies some condition, e.g.:
  - Changes are smooth
  - Values are extrapolated far
  - Changes are sharp
  - Flows are incompressible
  - Mass balance holds
- etc.

No good combination of locality coupling condition yet selected

Helsinki, 11.7.2018
Thank you

Jukka Heikkonen
Pekko Lindblom
Paavo Nevalainen
Ville Keränen

Future Technologies Lab research team.
ANNEX 3 – CASE STUDY 3: Estonian Case

Big Data & European Growth Corridors (BIG DATA)

Potentials of big data for integrated territorial policy development in the European growth corridors

Targeted Analysis activity for ESPON EGTC.

// Estonian Case (phone data)

UNIVERSITY of TARTU

Mobility Lab,

http://mobilitylab.ee/ena/
Examples of using the Big Data to study spatio-temporal behaviour

- Mobility Lab of University of Tartu has using Call Detail Record (CDR) data since year 2005.

Call Detail Records (CDR)

- Foreign roaming clients in Estonia
- Estonian clients in Estonia
- Estonian clients abroad

<table>
<thead>
<tr>
<th>user_ID</th>
<th>timestamp</th>
<th>site_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>9664422</td>
<td>2017/07/..</td>
<td>135</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sited_id</th>
<th>lon</th>
<th>lat</th>
</tr>
</thead>
<tbody>
<tr>
<td>135</td>
<td>27.845655</td>
<td>57.5466</td>
</tr>
</tbody>
</table>
Research topics

- Mobility in Estonia
  - Meaningful locations (anchor points), (Ahas et al. 2010 – link)
  - Central places (Novak et al. 2013 - link)
- Movement to / from Estonia
  - Tourism (Ahas et al. 2007 - link)
  - Working
  - Transit
  - ...


CDR in mobility studies

- Population statistics
- Census
- Migration
- Commuting
- Central places & catchment areas
- Tourism statistics, (Nilbe 2014 – link)

Additional data sources

- Public Transport Information System, [link]
- Traffic Counters, [link]
- Functionality of space (landuse, services, ...)
  - Open Street Map (OSM), [link]
  - Google
  - ...
- Crime data
- Meteorological data
- Events

Additional data sources

- University of Tartu
- Positium LBS
- ...

- Bank of Estonia, [link]
- Transport planning

Eesti Pank (2014) LIST OF STATISTICAL ACTIONS OF Eesti Pank in 2015-2020, under the official statistical programme. Available at:
Movement flows

- Antenna based
- Network graph?

Network of flows
Movement during Christmas Eve, [link]

Migration [link]
Central places & hinterlands (commuting),

Routing
Finnish people in Estonia

Functionality of space

- Residential
- Business
- Transit
  - Leisure
  - Household
callgraph

// Thank you
Anto Aasa (anto.aasa@ut.ee)
Siiri Silm (siiri.silm@ut.ee)

This presentation will be made available at: www.espon.eu
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.