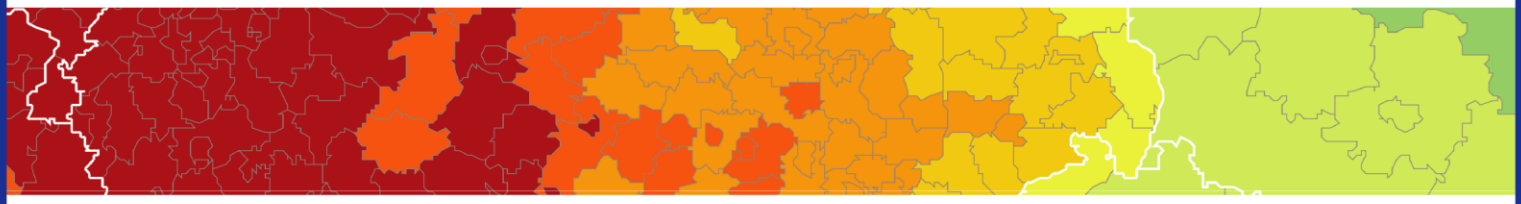
ESPON

EUROPEAN UNION

Co-financed by the European Regional Development Fund

Inspire policy making by territorial evidence



Potentials of big data for integrated territorial policy development in the European growth corridors (Big Data & EGC) – Steps towards data-driven corridor governance

Targeted Analysis

Practical Guide

28/06/2019

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, Ville Keränen, The Department of Information Technology, University of Turku (Finland)
Siiri Silm, Anto Aasa, Mobility Lab, University of Tartu (Estonia)

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

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

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

This delivery exists only in an electronic version.

Contact: info@espon.eu

Table of content

List of Figures

List of Maps

List of Tables

1 Purpose and intended audience for the practical guide

This practical guide is intended to support actors interested in contributing to corridor development and applying the results of the ESPON targeted analysis, Big Data and European Growth Corridors, in their work. It is intended especially for public sector authorities interested in finding new data sources and approaches to data analytics to support their policymaking. The guide presents tools that result from the targeted analysis, in which the aim was to generate understanding about the various potentials that big data brings to the field of territorial development, and particularly to European growth corridors. These tools can be used in evaluating new data sources and boosting data-driven policymaking in functional regions.

The potentials of big data for integrated territorial policymaking relate first and foremost to seeing things differently: big data can act as an eye-opener when it comes to spatial connectivities. The utilisation of new datasources can reveal new sides of functionality, revealing both strong and weak, or even unexpected, connectivities that require different policy measures. Versatile understanding of various flows and interactions is needed to legitimise publicly funded corridor-development policies and tailor corridor-specific policy measures.

Both collective and individual actions are needed to support the utilisation of new data sources. For example EU and statistical organisations could take a more active role in exploring and supporting the possibilities related to utilisation of new data sources and in supporting enabling legislation. In addition, corridors as networked and soft governance bodies can take several actions when it comes to networking and lobbying activities as well as developing strategies and practices for data-driven corridor development.

The guide is especially targeted at public sector authorities as a specific need has been identified to increase and broaden the capacities related to data literacy. In practice this means skills related to traditional data management plus skills in data analytics, data science, data literacy, and practical know how regarding how to interpret and use data in policymaking. In addition, a general awareness of the data economy and advanced data analytics is necessary to be able to recognise potentials related to big data in territorial policymaking. The organisation-wide development of these capacities and understandings regarding how to use big data requires active involvement from all levels of a public sector organisation for it to make its fullest impact.

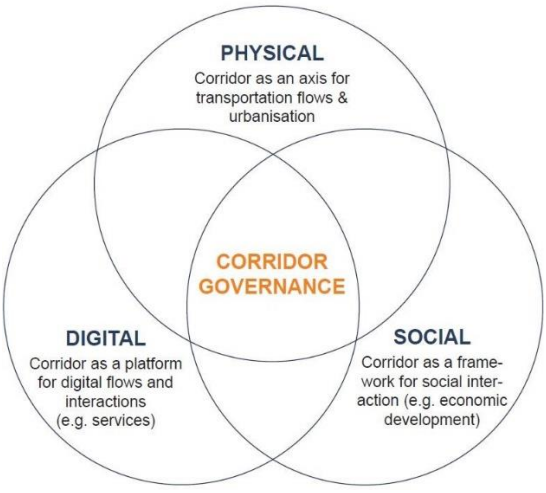
2 Corridor development is based on one-sided understanding of functionality – how to broaden the perspective?

Growth corridors have been at the centre of European territorial development already for a long time as their development is essentially tied to the development of the TEN-T network. Recently, the potentials of big data for corridor-based development have been discussed for example in the context of testing and deployment of 5G technology, which provides various possibilities for application areas such cross-border mobility[1]. However, there is a need for a more comprehensive understanding of functional corridors as areas featuring various overlapping flows and interactions that cannot be captured by using traditional data sources.

Conventionally, corridor functionality has been understood and measured mostly through physical mobility of people and goods. However, a broader understanding of flows and interactions is needed to support comprehensive corridor governance. For example, digitalisation and technological development have many spatial and social effects that are not sufficiently understood in policy contexts. To better understand the potentials of corridor policies in contributing to socially, economically and environmentally sustainable development, a transportation-based understanding of functionality can benefit by the addition of alternative data sources capable of broadening the perspective of flows and interactions.

A conceptual tool was produced in the targeted analysis to help policymakers consider what kinds of data could be used to support comprehensive corridor governance (Figure 2.1). It aims to broaden the perspective of corridor functionalities.

Figure 2.1: Conceptual tool for broadening the perspective of corridor functionalities.



The tool describes three overlapping functional dimensions of corridor development: 1) the physical dimension that describes the corridor as an axis for transportation flows and

[1]https://ec.europa.eu/digital-single-market/en/cross-border-corridors-connected-and-automated-mobility-cam

urbanisation; 2) the social dimension that describes the corridor as a framework for social interaction, and; 3) the digital dimension that describes the corridor as a platform for digital flows and interactions. The three aspects of corridor development are often interrelated and overlapping. These relative and relational views change the perspective of planning as both social and digital realities affect the ways in which the physical space is organized.

Corridor functionalities consist of multiple flows and interactions, which can be categorised as being physical, social and digital, or a combination of some or all of these dimensions. The targeted analysis produced a categorisation of key flows and interactions related to these dimensions, and examples of what kinds of datasets can convey something about the flows and interactions (Table 2.1). This categorisation can be used to stir ideas for what kinds of datasets can be included in a study, recognise new data sources, and design data-driven inquiries to support corridor-wide policymaking.

Table 2.1: Data categorisation for data-driven corridor governance and example data sources.

	Interactions/Flows	***Example data sources***
Physical	*Cargo Flows*	*traffic management systems, waste and recycling metrics*
	Customer Flows	*retailer bonus programs*
	Commuting	*vehicle-produced data, mobile positioning data, private GPS services*
	Student Mobility	*student train/bus ticket sales, university student registers, exchange programs*
	Migration & Relocation	*postal service records, population registers, issued residence permits, home purchases, rental listings*
	Tourism & Leisure Flows	*national border crossing statistics, hotel night stays, device languages on wireless networks, social media posts of tourist sights*
	Business Travel Flows	*border crossing statistics, tradeshow 'booth lists', customer data from booking agencies*
Social	*Research Cooperation*	*funded research projects & consortiums, bibliographic datasets (e.g. co-authors & institutions), open science portals, patent data*
	Trade Interaction	*trade statistics, B2B transactions, industry conference lists, tradeshow listings of registered booths, speakers, or attendees*
	Inter-Firm Cooperation	*supply chains, business consortiums, shared patents, board data, event attendees*
	Family/Social Interactions	*mobile positioning data, social graphs (e.g. Facebook Social Graph, LinkedIn connections, Twitter followers)*
	Investor to Business Interaction	*stock market data, crowd-investing platforms, AngelList, Crunchbase, investor call transcripts, start-up investors*
	Student Interactions	*course lists, student union member lists, exchange student lists, joint degree programs, student groups on social media, registers of student organisations*
	Planning Cooperation	*minutes of official planning meetings, verbatim transcripts of public meetings, vendor bids*
	Purchase flows	*B2B invoices, price scanner data, import/export metrics*
Digital	*Network traffic flows*	*IP Pairing Site Network Traffic (IPX), wireless data traffic; satellite signals*
	Monetary flows	*IPO trading, corporate bonds, public bonds, business loans, consumer loans & purchase data*
	Communication flows	*phone utility metadata, social media posts*

3 Big data utilisation is about balancing potential gains and resourcing costs – what has to be taken into account?

When considering new data sources and their applicability for corridor governance, several variables ought to be taken into account. Being aware of these variables not only helps determine the suitability of a dataset to a particular analytic process, but also indicates the strengths and limitations of a dataset regarding what kinds of insights can be produced from it. Such variables to be considered when evaluating the overall value and effort of using a new data source are presented in Table 3.1.

Table 3.1: Various aspects related to evaluating new datasets.

Data Categorization Variables	**Range of Attributes**
Availability	Open data ↔ Available only under special contract ↔ Purchasable proprietary data ↔ Unavailable proprietary data
Comprehensiveness	The data sample is limited (only some subset of data subjects or entities of all possible ones is included) ↔ Dataset is complete (includes all possible data subjects or entities, e.g. a population register)
Level of Processing	Raw (e.g. direct from sensors) ↔ Pre-Processed Data ↔ Processed Data ↔ Highly Processed Data
Intended Audience	Machines ← Programmers of Machines → People
Observational Qualities	Direct Observation ↔ Synthetic
Level-of Detail	Fine-Grained (e.g. vehicle journeys) ↔ Rolled-up (e.g. average household income by zip code)
Level of Structure	Highly Structured ↔ Semi-Structured ↔ Unstructured
Refresh Frequency	Instant ↔ Milliseconds ↔ Daily ↔ Weekly ↔ Monthly ↔ Quarterly ↔ Annually ↔ Every few years
Confidence in Updates	Low Confidence (updates are uncertain) ↔ High Confidence (updates are highly likely to occur)
Extraction Effort	Great effort is required to extract data ↔ Little effort is required
Analytical Effort	Great effort is required to analyse data ↔ Little effort is required
Clarity of Ownership	Ownership is clear and singular ↔ Ownership is clear but shared ↔ Ownership is clear on paper and unclear in practice (e.g. “Do I or Facebook or a 3rd party app maker own my Facebook data?”) ↔ Ownership is unclear or no longer traceable
Spatial Resolution	Exact location, Local, Neighbourhood, Municipal, Regional, National, Multi-national, Continental, Global
Temporal Resolution	Milliseconds ↔ Years

For example, in the context of *mobile positioning data*, the value for policymakers is the potential for very up-to-date insights into a nation or city’s mobility patterns. Overall, mobile positioning data has proved itself as a very good and promising data source for studying the

mobility of society and/or smaller groups. In countries like Estonia and Indonesia mobile positioning data is already used as input for official statistics. The data has fine and deep temporal resolution, meaning it includes frequent observations and is available going back many years (see Table 3.2). Because of these characteristics, the dataset can reveal short-term or seasonal differences. It can indicate changes in the numbers of people moving among key places in their lives, including homes, workplaces, and stores.

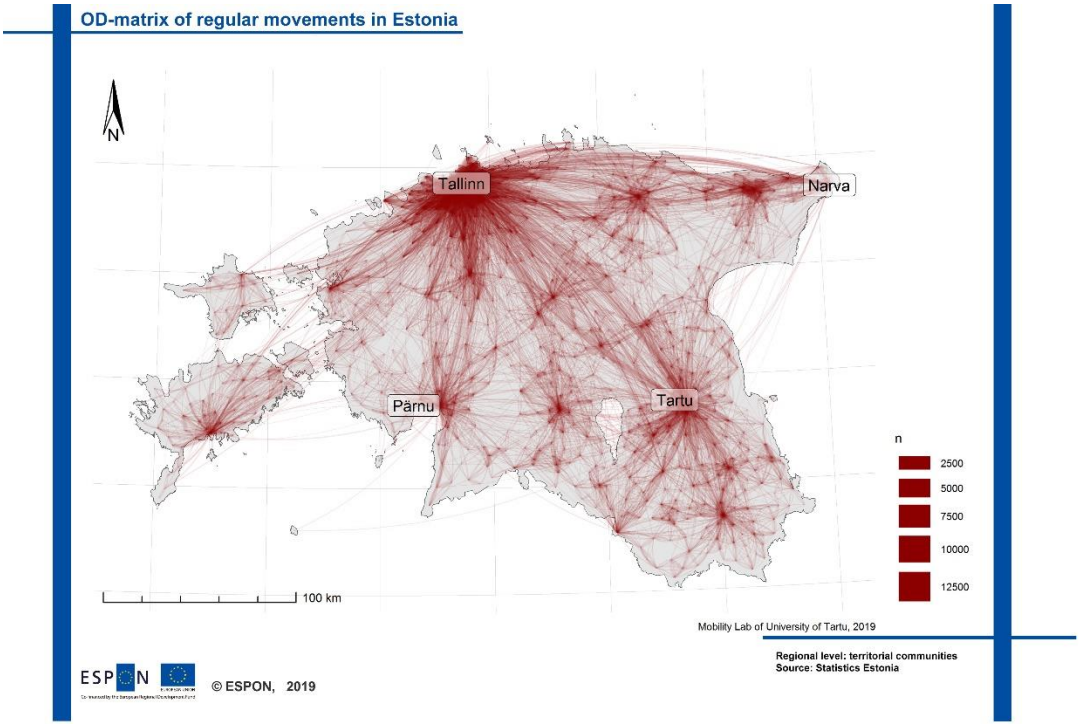
Table 3.2: Key attributes of mobile positioning data.

Dataset Call Detail Records, Telia (in Estonia)	
Data Characteristics	
Availability	Purchasable Proprietary Data
Level of Processing	Raw
Comprehensiveness	Partial, can be increased by adding CDR data from other carriers
Intended Audience	Machine Readable
Observational Qualities	Direct Observation
Level-of Detail	Fine-Grained
Level of Structure	Highly Structured
Refresh Frequency	Monthly
Confidence in Updates	High Confidence (updates are highly likely to occur)
Extraction Effort	Requires some effort to prepare data for use
Clarity of Ownership	Ownership is clear and singular
Spatial Resolution	Local, sizes of network cells vary
Temporal Resolution	A range from minutes to years

In the Big Data & EGC Targeted Analysis, a methodology for everyday mobility database was developed which contains the OD-matrices of movements between territorial communities (Map 3.1). The database will be used to support mobility-related policymaking in Estonia. Therefore, the Estonian example functions as a best practice example in Europe, which could be widely applied in diverse scales of territorial development. The Estonian experiences related to the utilisation of mobile positioning data in national transportation planning have to be further studied to identify future potentials, best practices, and challenges related to utilising this dataset. Furthermore, the possibilities of mobile positioning data for cross-border settings should be investigated, as the data is increasingly available in neighbouring nations of Finland, Estonia and Sweden. This kind of data exists in all European nations which means it would be suitable for cross-border analysis of functionalities of Europe-wide corridors.

In the future, measures need to be taken in legislation and ecosystem development to support and ease the usage of mobile positioning data in both public sector and academic research supporting territorial development. The utilisation of this data usually implies long value chains requiring expertise from several fields, thus indicating the need for public-private partnerships and ecosystem building to foster public value creation. In addition, strong protections for consumer privacy must be maintained while public benefits of utilising this data are pursued.

Map 3.1: Regular movements in Estonia according to the analysis based on mobile positioning data.



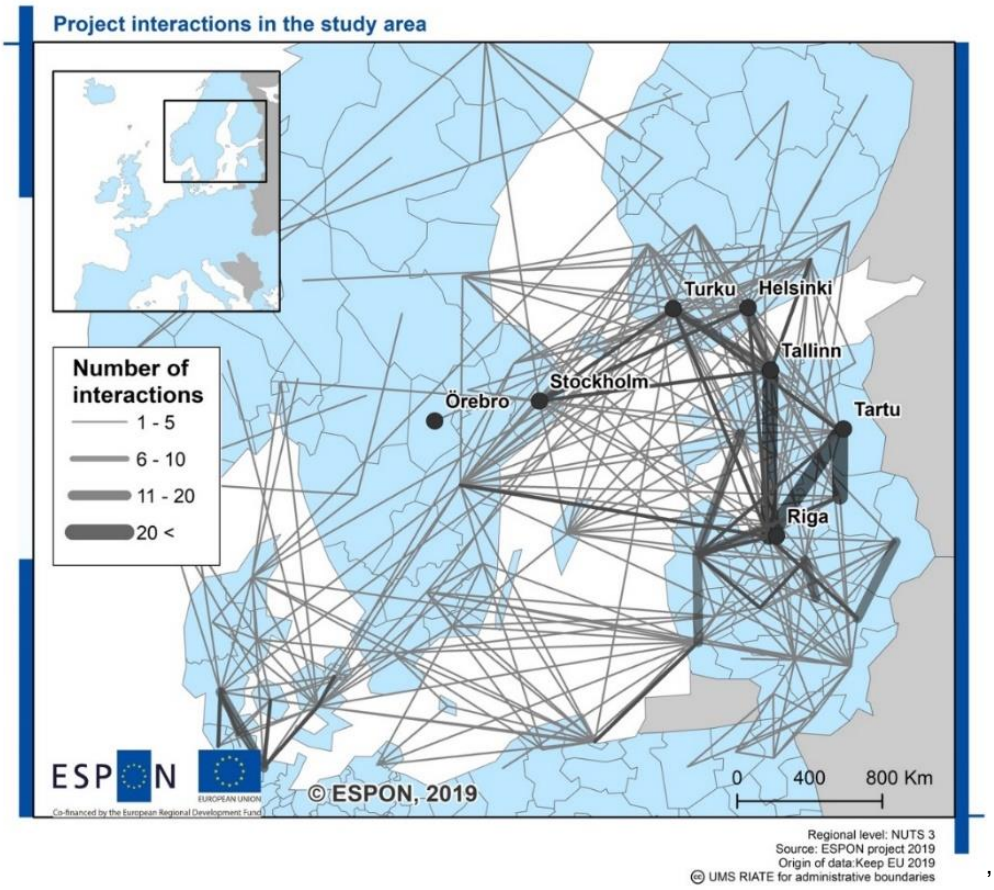
Social network analysis can also provide interesting insights on territorial development as it can be done by using versatile datasets such as consortium, social media, or board member data. In this targeted analysis, ERDF project partnership data (Table 3.3) was analysed to understand the collaboration dynamics in the Northern Growth Zone.

Table 3.3: Key attributes of a dataset used to map collaborations

Dataset European Regional Development Funding (ERDF) data	
Data Characteristic	**Attribute**
Availability	*Open data, through interface*
Level of Processing	*Highly processed*
Intended Audience	*Humans*
Observational Qualities	*Direct observation*
Level-of Detail	*Rolled up*
Level of Structure	*Highly structured, but gaps found*
Refresh Frequency	*Updates weekly, programming data quarterly*
Confidence in Updates	*High confidence* *(Keep.eu is updated constantly, but the data is added to the database a few days after delivery from the programs)*
Extraction Effort	*Requires some effort*
Clarity of Ownership	*Ownership is clear, it is property of the Interact III*

The resulting visualisations clearly showed a spatial imbalance of collaboration activity, with most of collaboration happening in the eastern parts of the corridor (Map 3.3). These insights, could be used as a basis for making policies intended to foster networking activities, especially between Finland and Sweden, or Sweden and Norway. An increased understanding of collaboration dynamics and social interactions can help policymakers to formulate interventions and policy-measures to support transnational corridor development.

Map 3.2: Project interactions in the study area of the Big Data & EGC Targeted Analysis.



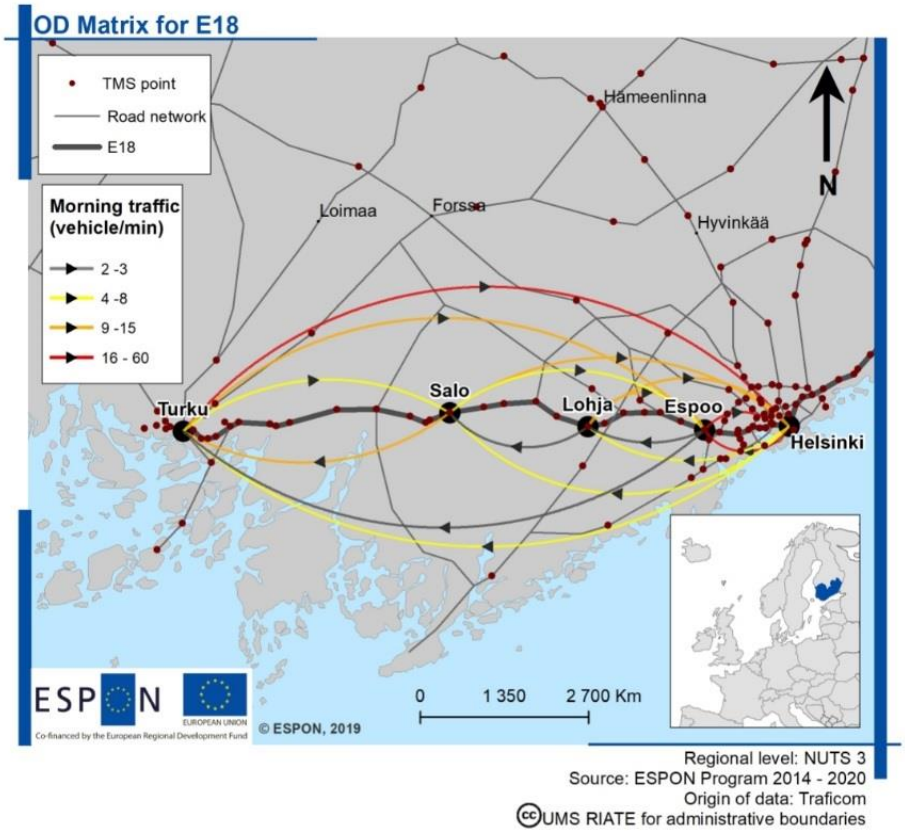
Efforts to derive new insights from *automated traffic measurement (ATM) data* are also an avenue for supporting corridor development using big data. In this targeted analysis, a rich dataset collected from a network of inductive loop sensors in the Finnish segment of highway E18 was analysed. The dataset contained detailed automated traffic intensity measurement data provided by the Finnish Transport and Communications Agency (Traficom) (Table 3.4). Similar forms of data are passively and automatically gathered in other European nations, including Swedish Transport Administration (Trafikverket) and Estonian Road Administration (Maanteeamet), thus indicating a potential for cross-cutting value in developing new analytical tools for such data. In this case study, the data model was used to produce an O-D matrix that could be scaled to any time parameter or segment of the highway. O-D matrices are excellent for illuminating dependencies, similarities, synchronization and differences between districts, which can help regional decision-making related to spatial planning, economic growth, or

infrastructure improvements. The model is geographically generic, meaning that it is applicable anywhere where corresponding traffic data and similar economic indicators are available.

Table 3.4: Key attributes of traffic measurement data

Dataset: Traffic Management System Data from Traficom	
Variable	**Attribute**
Availability	Open Data
Comprehensiveness	Nearly complete: All vehicles crossing over sensors are recorded except in rare cases when a sensor may malfunction.
Level of Processing	Raw
Intended Audience	Machines
Observational Qualities	Direct Observation
Level-of Detail	Fine-Grained
Level of Structure	Highly Structured
Refresh Frequency	Daily
Confidence in Updates	High Confidence
Extraction Effort	Requires little effort to extract data
Clarity of Ownership	Ownership is clear
Size	20 Bytes / (vehicle passage) or 160 kB / (year x km) or 0.5 TB (total E18)

Map 3.3: The OD Matrix for E18 for a business week in 2017



4 Steps forward – what can be done to support big data utilisation in territorial development?

This chapter presents steps which could be taken to advance data-driven development of European growth corridors. In addition, it summarizes a key rationale for pursuing comprehensive corridor governance based on a versatile big data supported evidence-base. In a data-driven future of territorial development, there is an increasing need for skills related not only to traditional data management, but increasingly also to data analytics and data interpretation for use in policymaking. In the context of European growth corridors, such capacity building can be developed in the many public organisations found in growth corridors as more advanced organisations share their best practices, experiences, and tools those just starting to use big data. Because corridor governance is done by many actors working simultaneously, cross-corridor cooperation is preferred over competition or data hoarding among municipal and regional governments.

4.1 What should be done collectively?

When it comes to creating a basis for purposeful and ethical data utilisation, the EU plays a key role in creating enabling legislation. While the GDPR provides a solid basis for protecting consumer rights to privacy and data ownership, it has also introduced challenges when it comes to knowing what can and cannot be done with data in what contexts. Furthermore, a lack of clarity regarding legal requirements and best practices can often dissuade public organisations involved in territorial development from beginning to explore what they can do using big data sources. Due to the costs related to different phases of data extraction, handling and analysis, bodies such as statistical offices could take a more active role in highlighting the most fruitful opportunities to utilise new data sources and combine them with traditional statistical data.

When it comes to the further use of mobile positioning data, recent changes in the legal environment have introduced new challenges for using data in scientific research. Questions have arisen regarding what levels of data anonymization and other privacy supporting practices are necessary to comply with the law. Therefore, clarification is highly needed regarding under what conditions mobile positioning data can be used in scientific research. At least the following issues need to be considered:

1. Legislation should provide clear and unambiguous exemptions for service providers or communication undertakers to allow transferring mobile positioning data for research and statistical purposes.
2. The supervisory authority responsible for monitoring the application of GDPR should offer verification services to inform researchers whether their plans are compliant with GDPR. For example, in Estonia, laws before GDPR went in force provided the nation's supervisory authority the possibility to verify the compliance of a research design. That option was removed for most cases in the nation's post-

GDPR legislation. Therefore, service providers now have little to any assurance from authorities that they are doing the right thing, leaving them to figure out for themselves if they are exposed to legal risks arising from non-compliant research projects and possible misuse of data.

3. It would also be helpful, if a specific guidelines for researchers are added to the Code of Conduct (Article 40) of the GDPR. Such guidelines would help reduce legal uncertainty for all parties involved in big data research for public benefit. These guidelines could help researchers and data providers assess their choices and the extent of measures and procedures needed to prepare data transfers or conduct research.

Comprehensive flow analytics could be developed in collaboration with public and private organisations involved in territorial development, evidence production, and service provision in functional corridors, which could be promoted by the national and EU authorities. For example the automated traffic measurement (ATM) data mentioned in the previous chapter could be combined with data depicting wider corridor dynamics. Private actors could be instrumental in developing such data insights as a service offered to public authorities. If there were a thriving data ecosystem in the corridor, the ease with which datasets could be combined would support public and private actors in working together toward reliable approaches to providing data. An improved understanding of flows and interactions would support policymakers in making land-use and transportation planning decisions as well as allocating resources and service provisions more efficiently.

4.2 What can corridors as collaborative governance networks do?

In practice, organisations in European growth corridors can take several actions to boost data-driven governance at diverse scales of formal territorial administration and informal networked settings. Some of these measures are more facilitative in nature, whereas others, such as setting agendas and taking specific actions are more targeted. Such networks could be organised around functional areas aimed at bridging gaps between data-driven visions and practices.

The following steps (Figure 4.1) are presented to support wider utilisation of big data in the Northern Growth Zone (NGZ).[2] However, these steps are applicable to other European growth corridors as well as to territorial administrations more generally. A reason is that many public actors face similar challenges such as lack of capacities and skills for utilising big data while often facing inertia from internal policies and external legislative restrictions on wider utilisation.

[2] The NGZ is a transnational collaboration network organised to boost the development of the transnational corridor from Oslo via Stockholm, Turku and Helsinki to St Petersburg.

Figure 4.1 and the list below present four steps territorial administrations can take to develop a more versatile evidence-base regarding the flows and interactions in corridors and more efficiently use new data sources:

1) **Strategizing**: Envisioning and strategizing for data-driven governance of corridors is needed to understand the potentials of such data-driven approaches. Here, for example benchmarking and environmental scanning could be used to clarify one's strategic positioning. Furthermore, action planning is needed for short and long term time horizons.
2) **Understanding**: Corridor organisation can coordinate and boost capacity building and awareness raising efforts to become data-driven. This can mean for example organizing events as well as training and professional education regarding new approaches, data sources, and best practices for both public and private authorities. This capacity building should aim to boost awareness and skills in data-analytics and foster data literacy. By having increased skills in analysing and interpreting data, territorial administrations will be better prepared to support corridor development based on versatile and more current evidence. These efforts are more likely to succeed if they are equally supported across corridors by management who recognise the value and need for these capacities and cooperation among regions and municipalities. In addition, supporting research on new data sources, data integration and data-driven governance is important to deepen the understanding about possibilities and challenges related to a future of data-driven governance.
3) **Collaborating**: Often the most promising datasets describing flows and interactions are privately owned. Therefore, public organizations in the corridor would benefit from be involved in both building and facilitating public-private partnerships and collaborative business models to promote data utilisation and incentivize data provision. Furthermore, partnerships in research and development for example between research organisations and companies should be supported to create public value from new data sources and analytical approaches. Overall, corridor organisations should take an active role in developing data ecosystems oriented toward producing impactful insights from data.
4) **Enabling**: Lobbying for legislation both nationally and internationally that enables data utilisation across administrative boundaries is key to corridor development. Especially in the context of mobile positioning data, measures need to be taken in both legislation and ecosystem development to support and ease the usage of such data – while simultaneously protecting consumer privacy – in research conducted by the public sector and academics to support territorial development. Typically, the utilisation of this kind of data implies long value chains requiring expertise from several fields. Therefore, also lobbying for long-term theoretical research on new data sources and methodological development is needed to support the wider utilization of

data sources. In addition, supporting data standardization, quality benchmarking and data harmonization across administrative boundaries – work already underway – is needed to promote transnational corridor development.

Figure 4.1: Steps towards data-driven corridor governance.

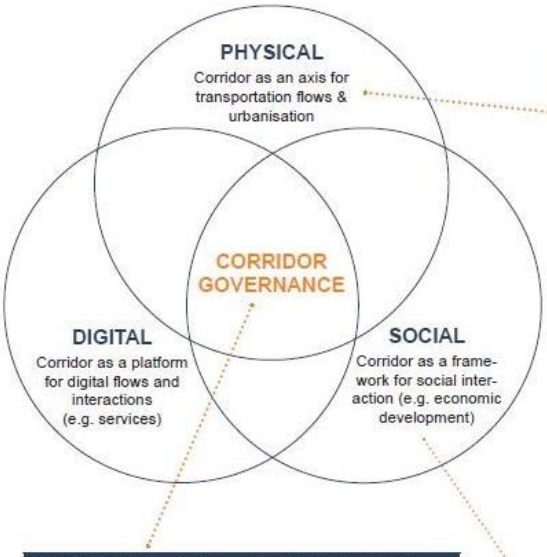
Corridors per se function as a framework for sectorally and spatially integrative policy-making. An integrated approach is fundamentally related to combining datasets to cover physical, social and digital aspects of interactions and flows. Using conventional and unconventional datasets in combination holds a high potential to create new insights and reveal patterns over time. However, not everything has to be – nor can be – integrated under comprehensive corridor governance. Rather, comprehensive and integrative approaches should acknowledge the existence of parallel and overlapping functionalities that require diverse policy measures.

In the context of the NGZ, which was the geographical context of the Big Data & EGC targeted analysis, the results of the project partnerships case could be complemented by a dataset describing physical mobility which would further illuminate transnational corridor functionality. In addition, the proposed conceptual tools such as the three dimensions – physical, social, and digital – can be used to select datasets which emphasize different corridor development perspectives. This targeted analysis proposes that a policymaker's comprehension of corridor functionality can be broadened when insights are produced using analysis processes that are applied to data from all three dimensions.

The three examples described in Chapter 3 demonstrate the complex nature of corridor development. For instance, mobile positioning data can be used in numerous contexts such as planning cross-border services and developing transportation infrastructures and services at various administrative and functional scales. Figure 4.2 summarises how the different datasets

used in the experiments conducted for the Big Data & EGC Targeted Analysis fit into the conceptual tool and its dimensions.

Figure 4.2. Three datasets from case studies and their relation to the corridor governance framework.



DATASET: Traffic Management System Data from Trafi	
Data Characteristic	Attribute
Availability	Open Data
Comprehensiveness	Nearly complete: All vehicles crossing over sensors are re-corded except for in rare cases in which a sensor is malfunctioning
Level of Processing	Raw
Intended Audience	Machines
Observational Qualities	Direct Observation
Level-of Detail	Fine-Grained
Level of Structure	Highly Structured
Refresh Frequency	Daily
Confidence in Updates	High Confidence
Extraction Effort	Requires little effort to extract data
Clarity of Ownership	Ownership is clear
Size	20 Bytes / (vehicle passage) or 160 kB / (year x km) or 0.5 TB (total E18)

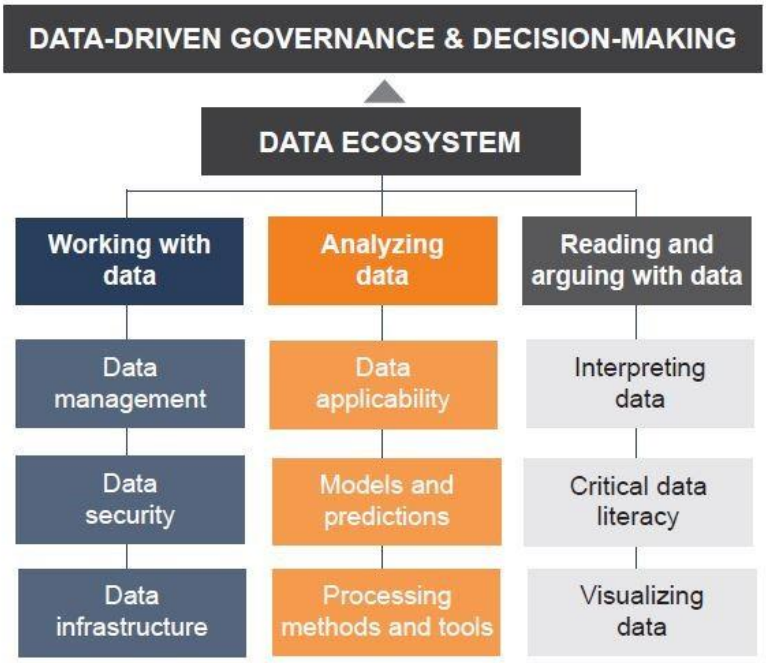
DATASET: Call Detail Records, Telia (in Estonia)	
Data Characteristics	
Availability	Purchasable Proprietary Data
Level of Processing	Raw
Comprehensiveness	Partial, can be increased by adding CDR data from other carriers
Intended Audience	Machines
Observational Qualities	Direct Observation
Level-of Detail	Fine-Grained
Level of Structure	Highly Structured
Refresh Frequency	Daily
Confidence in Updates	High Confidence (updates are highly likely to occur)
Extraction Effort	Requires some effort to prepare data for use
Clarity of Ownership	Ownership is clear and singular
Spatial Resolution	Local
Temporal Resolution	Minutes ⟷ Years

DATASET: European Regional Development Funding (ERDF) data	
Data Characteristic	Attribute
Availability	Open data, through interface
Level of Processing	Processed data
Intended Audience	Humans
Observational Qualities	Direct observation
Level-of Detail	Rolled up
Level of Structure	Highly structured, but gaps found
Refresh Frequency	Updates weekly, programming data quarterly
Confidence in Updates	High confidence (Keep.eu is updated constantly, but the data is added to the database after few days from the delivery)
Extraction Effort	Requires some effort
Clarity of Ownership	Ownership is clear, property of the Interact III

4.3 What is the role of individual organisations?

Striving for fundamental change towards big data driven governance requires collective efforts from the EU and national level authorities. In addition, however, individual organisations are important change agents and can influence the overall evolution of territorial development in growth corridors. Therefore, it is important to aim capacity building practices at both the organisational level and corridor level. Developing a data ecosystem for data-driven corridor governance requires improving data literacy in targeted skill areas (see Figure 4.3). A mental shift from a focus on data management to an aspiration toward data-driven governance is needed. Overall, organisations need versatile capacities to enable a data-infused organisational culture.

Figure 4.3: Recommendations for capacity building.



To establish a big data ecosystem for a growth corridor, multiple data literacy aspects must be addressed:

Working with data – One key aspect for developing a corridor level data ecosystem is to promote data literacy – skills to work with data. Working with data includes:

- *Enhance data management:* These practices support the usability and discoverability of data. Efforts also include setting policies and standards for how and when various datasets can be accessed or used that are clear and support dynamic utilisation. Data management includes operational models to support rapid development of using big data in policy-making.
- *Pay attention to data security:* Organisations should have a data protection framework that enables big data utilisation while keeping the data secure. Technology and tools must be in place to enforce organisational policies on data protection, data ownership and privacy to comply with laws and internal rules.
- *Improve data infrastructure:* Invest in IT infrastructure or services that support performing operations on big data. Some key infrastructure components include hardware and software that can handle machine learning, large-scale data analytics, rapid storage and retrieval, and processing of real-time data.

Analyzing data – Before any data analysis process, many steps are required, including selecting a dataset and which of its parts to analyse. To develop an organisational capacity for analysing data pay attention to:

- *Determining the applicability of datasets*: Take care that datasets used for an analysis process fulfil usability criteria and hold potential to add value for corridor level decision-

making. The data can be, for example, open data from public data sources such as governments and agencies accessed directly or from open data portals. In addition, big data can come from non-open datasets, for instance, customer data and data from search engine terms or social media. Input data can be both from conventional data sources, like census data, population registers, or from unconventional data sources like space-based remote sensing.

- *Modelling and trend predictions*: Models combining different datasets can provide significant value in describing interactions and connectivity in the growth corridors and in some cases provide predictions of future circumstances. Models can also produce outputs that combine many datasets or restructure data in new formulations. Capacities to model data are beneficial across an organisation and at all levels.
- *Processing methods and tools*: Organisations must begin learning and using new tools for making sense of big data. These advanced tools for data analytics – including those based in artificial intelligence, machine learning and deep learning – can be useful for supporting national, regional, and municipal governments in strategically developing their big data capacities for corridor development.

Reading and arguing with data - Using big data for evidence-based decision-making requires staff members with deep technical skills and highly informed non-technical actors. Data literacy is needed at all levels of organisations to put data into action. Experts who are capable of transforming data into appropriable data and concrete action pathways are helpful to all parts of an organisation. Generally, capacities to understand how data can be read and questioned are essential for all members of public organisations. The following pathways are recommended for the following capacities:

- *The capacity to interpret data*: Promote training and professional education for improving capacities on the skills and knowledge to process and interpret big data. Develop data literacy skills to interpret data, and apply and contextualize the results of data analysis. Learn by doing how to transform interpretations into policy actions.
- *Critical data literacy*: Develop skills assessing data critically. Promote training and professional education for improving capacities to consider the properties and limits of different datasets. Gain knowledge on what questions to ask about how datasets are produced and used.
- *Visualizing big data*: Emphasis and praise visualisations that clearly convey results of big data analytics in a very understandable way that can be used by policy-makers and the public to make sense of complicated matters. Assess potential big data visualisation tools and applications for the needs of your organisation. Focus on learning visualisation tools that map data geospatially.

4.4 What next?

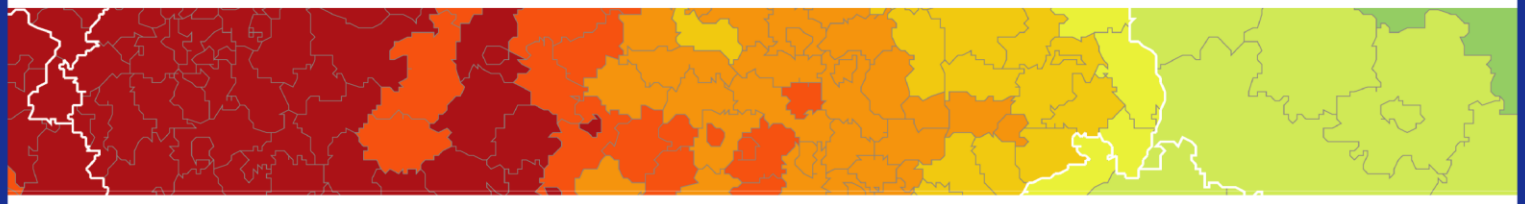
The rapid increase in new data sources related for example to sensor and customer data is seen to be changing the scheme of spatial analytics. Several ongoing trends such as digitalisation, rapid development of new data analytic tools, and increased opening and sharing of data combine to make big data a significant source of new insights for territorial development. In the context of European growth corridors, big data is already key to novel concepts for sustainable mobility (e.g. Mobility as a Service).

However, broadening the perspective is still needed to increase the understanding about constantly evolving territorial dynamics that challenge the existence of various administrative boundaries. The Big Data & EGC targeted analysis took a step towards this direction and provides a good basis for future research and development to broaden the understanding of functional growth corridors. Collective efforts should be made for example by the EU and statistical offices to further explore the utilisation potential of big data together with research organisations.

The presented conceptual tool in this project can be utilized to widen the horizon of big data utilisation. Using it can help shed light on the three different yet inherently interconnected physical, social and digital dimensions of corridor development. In addition, the conceptual framework can be used as a practical tool in selecting datasets from different corridor development perspectives.

Big data has a clear potential in renewing or even disrupting cross-border corridor governance and generating systemic change based on new collaborative business models between public and private actors. New knowledge about functional geographies arising from such collaborations could be used to challenge traditional territorially bounded playing fields. However, integrating various data sources necessitates dynamic structures for integrated data management.

Integrative and collaborative platforms should be developed for data-driven corridor governance to create public value from new data sources. The development of corridors depends heavily on the capacities of public organisations to utilize new data sources and develop their skills in advanced data analytic tools and data literacy. To implement corridor-wide data-driven governance, learning approaches to produce evidence and insights from big data must be coupled with organisational capacity building related to interpreting, contextualising and using evidence from big data.



ESPON 2020 – More information

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

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