ESPON Big Data for Territorial Analysis and Housing Dynamics

Monitoring and tools

Interim Report 1
Interim Report 1

This monitoring and tools activity is conducted within the framework of the ESPON 2020 Cooperation Programme.

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
Renaud Le Goix, UMS RIATE – University Paris Diderot (France) – Project coordinator
Ronan Ysebaert, UMS RIATE – University Paris Diderot (France)
Timothée Giraud, UMS RIATE – CNRS (France)
Lambert Nicolas, UMS RIATE – CNRS (France)
Marianne Guérois, UMS RIATE – University Paris Diderot (France)
Malika Madelin, University Paris Diderot (France)
Thomas Louail, Géographie-cités (France)
José Ravier Ramasco, FISC-CSIC (Spain)
Pere Colet, FISC-CSIC (Spain)
Thierry Teurillat, Haute-Ecole Arc (Switzerland)
Alain Seggesseman, Haute-Ecole Arc (Switzerland)
Szymon Marcinczak, University Lodz (Poland)
Bartosz Bartosiewicz, University Lodz (Poland)
Elisabete Silva, Cambridge University (UK)
Sølve Baerug, Norwegian University of Life Sciences (Norway)
Terje Holsen, Norwegian University of Life Sciences (Norway)
Guilhem Boulay, University of Avignon (France)
Marco Kellenberg, Office Fédéral du Développement Territorial (Switzerland)

Advisory Group
ESPON EGTC: Marjan Van Herwijnen (project expert), Caroline Clause (financial expert)

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.

© ESPON, 2019

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
Interim Report 1

ESPON Big Data for Territorial Analysis and Housing Dynamics

Version 1.2
06/11/2018, amended 25/01/2019

Disclaimer:
This document is a interim report.

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

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

List of Maps ........................................................................................................................................... ii
List of Maps ........................................................................................................................................... ii
List of Figures ....................................................................................................................................... ii
List of Tables ......................................................................................................................................... ii
Abbreviations ......................................................................................................................................... iii
1 A policy-oriented project ......................................................................................................................... 1
2 Geographical scope of the project ........................................................................................................... 3
3 Detailed plan of tasks ............................................................................................................................ 6
3.1 Task 1: Development of a methodological framework ................................................................. 6
3.2 Task 2: Gathering data sets ................................................................................................................. 7
3.3 Task 3: Analysing data sets ................................................................................................................ 8
3.4 Task 4: Reporting results ................................................................................................................... 9
3.5 Task 5: Communicating methodologies ............................................................................................ 9
4 Reminder of the organisation of the first delivery (6 November 2018) ............................................. 11
References ............................................................................................................................................. 13
List of Maps
Map 2-1 Presentation of case-study FUAs ................................................................. 3
Map 2-2 - 1km population grid on Geneva cross-border case-study .......................... 5

List of Figures
Figure 2.1 - Gantt Diagram of the project .................................................................. 6

List of Tables
Table 2-1 - Socio-economic characteristics of case-studies as regards to the other FUAs of the ESPON Area .............................................................................. 4
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESPON</td>
<td>European Territorial Observatory Network</td>
</tr>
<tr>
<td>ESPON EGTC</td>
<td>ESPON European Grouping of Territorial Cooperation</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>FUA</td>
<td>Functional Urban Area</td>
</tr>
<tr>
<td>IDS</td>
<td>Internet Data Sources</td>
</tr>
<tr>
<td>OECD</td>
<td>Organisation for Economic Co-operation and Development</td>
</tr>
<tr>
<td>TOR</td>
<td>Terms of reference</td>
</tr>
</tbody>
</table>
1 A policy-oriented project

Housing is a major policy issue regarding urban well-being, cohesion and sustainability in Europe: “in 2014, approximately 7 % of the EU-28 population faced the situation where housing costs accounted for more than half of their disposable income”\(^1\). Housing prices have increased faster than the income of residents and buyers in major post-industrial city-regions, and real estate has become an important driver of socio-economic inequalities.

The study is framed in the Territorial Agenda of the European Union for 2020 by dealing with the impacts of housing on territorial cohesion, and how to tackle with risks of exclusion. The OECD database on affordable housing shows that increased price and income inequalities exacerbate unequal access to affordable housing (OECD, 2018), a situation that spreads across property markets (André and Chalaux, 2018; Kemeny, 2001). A continuous increase in property prices and a steady increase of homeownership was analyzed as a new price regime (Tutin, 2013). This situation is linked with financial and macroeconomic parameters like monetary policies and credit affordability (national policies); and how advanced economies have shifted toward an asset-based welfare model, yielding a regime linking an ideology of ownership, credit affordability and house price often subsidized by the State and local governments (Ronald, 200; Rolnik, 2013).

Therefore, real estate has become an important driver of inequalities, for at least two reasons. First, increased property prices sort out buyers and renders access to housing highly dependent of assets and access to credit. Second, the flows of household real estate investments are instrumental to the dynamics of asset capitalization (Piketty, 2013). Therefore, both access and assets directly affect social inequalities and spatial segregation patterns of residents and buyers. The dimensions of these socio-spatial inequalities derive from a decrease of housing affordability, i.e. a widening gap between property prices and households’ income (Friggit, 2017).

In this research, ESPON interest in the methodology is also policy oriented: to “what degree new ‘big data’ collection approaches can be used to enrich existing territorial policies and provide up-to-date evidence”\(^2\). Our proposal seeks at bringing new insights on how to use unconventional data to gather valuable information, to provide a cost-effective and harmonized data collection and contribute to the analysis of socio-economic cohesion.

Our contribution also aims at developing a methodology to make housing-data more widely available at a European level, as our survey of available data showed that current Europeans databases lack information. First, in 2019, adapted indicators delivered by Eurostat on housing issues cover Functional Urban Areas and Core Cities. Unfortunately, this harmonised information allows only to deliver a “global picture” of the housing phenomena. All studies

\(^1\) 2016 Eurostat Urban Europe report.
\(^2\) ESPON EGTC, 2017, ToR Big Data for Territorial Analysis and Housing Dynamics
stands that the appropriate level of observation of housing inequalities is located at infra-urban micro-scales, within the cities. This is the reason why this project aims at enhancing the geographical granularity to collect data (X/Y location for real estate values, LAU2 units or 1 km grid for income data). Secondly, even if the housing thematics has been identified relevant by the European Commission and Eurostat (some indicators are available for FUAs), it seems however that the data collection process is not yet implemented since the Eurostat data portal returns no values for these indicators.

Therefore, by the means of bringing together conventional and unconventional data, the policy relevance of the case-study can be described according to the following goals:

- To monitor the spatial effects of pro-ownership policies on socio-economic inequalities, and the attendant risks of market-based exclusion.
- To analyze the spatial patterns of inequalities stemming from unequal capitalization of housing wealth some areas, vs. vulnerability of households in others.
- To better inform and map the increased affordability gap, a critical issue for social cohesion sustainability in metropolitan areas in Europe.
- To make available relevant data sources, ad-hoc methodologies and analysis to build harmonized indicators, and propose a reproducible framework of analysis, in order to extend the study at a European scale.
2 Geographical scope of the project

This project focuses on 10 case-study cities, covering 4 countries of the ESPON Area: Geneva (Switzerland), Annecy, Avignon and Paris (France), Madrid, Barcelona and Palma de Majorca (Spain) and Warsaw, Lodz and Cracow (Poland), highlighted in blue on the Map 2-1. These case-studies correspond to the geographical field of expertise of partners involved in the project, in term of data availability and analysis of housing dynamics properties.

In the choice of case-study cities, a cross-border area has been identified: Geneva-Annemasse-Annecy, where housing stakes are really important to be considered, and not yet studied in term of data harmonisation.

The selected case-studies correspond to various positions in the urban hierarchy: the spectrum ranges from 3rd tier cities functional urban areas (276 000 inhabitants for Annecy, 320 000 for Avignon or 670 000 for Palma de Majorca), up to larger cities (3.1 millions inhabitants for Warsaw, 4.9 for Barcelona, 6.6 for Madrid) and includes on global city (11.9 M in Paris FUA).
Our selection of case studies also covers variegated socio-economic characteristics, to include a wide range of the spectrum of European cities. Some variables available through Eurostat are summarized Table 2-1, that describe relevant aggregated information to analyze housing dynamics. In term of socio-economic characteristics and as regard to the data currently available on Eurostat, selected case-studies correspond to diverse typologies. From a demographic perspective, some are FUAs where young populations are over-represented as regards to the others (Geneva, Annecy, Paris); some are FUAs where young active population are over-represented (all the FUAs excepted Avignon); other are FUAs more characterized by an over-representation of older populations (Lodz, Avignon and Barcelona, to some extent). From a housing perspectives, Spanish FUAs for instance are more characterized by the average size of households (nb of persons), and by an over-representation of homeowners. As regards to socio-economic characterization, all FUAs selected have a relative high level of employment in real estate activities, that highlights the interest of an analysis of housing market activities. However, many case studies are also characterized by a high unemployment rates (Annecy and Geneva set aside).

Table 2-1 - Socio-economic characteristics of case-studies as regards to the other FUAs of the ESPON Area

One of the challenge of this project will be to go beyond this aggregated territorial level to understand intra-urban inequalities between the cities. So as to deliver harmonized statistics for all the selected cities, 2 geographical levels will be used to aggregate collected data on housing dynamics: the LAU2 level, and the 1km European reference grid. The other challenge of the project will be to cover the entire Functional Urban Areas, despite missing data and incomplete datasets, as displayed for a very basic statistic s.a. population density. Where data is found missing, case studies will be conducted at list within the delineation of the core cities (surrounded in black on the figure below).
Map 2-2 - 1km population grid on Geneva cross-border case-study

Geneva transnational and Annecy FUAs - Population grid 1km 2011

Population density, 2011
1km square-grid

- 3000
- 200
- 100
- 50
- 10
- 1

No data

Source: ESPON Big Data for Territorial Analysis and Housing Dynamics, 2019

© EuroGeographics for the administrative boundaries
3 Detailed plan of tasks

As agreed during the kick-off meeting, and during the first meeting with partners, five tasks are currently conducted. The following section describes the tasks, the advancement of the project, as planned according to our initial proposal. This section, designed as a plan of tasks, describes all subtasks to be conducted for the duration of the project, as agreed with ESPON, as well as the state of completion according to the Gantt chart (Figure 3.1), and refers to the sections of deliverable #1 in which the results of the tasks are delivered (milestones for D1).

We propose in this first delivery a methodological framework and a proposal for the case study cities framed as stipulated in our technical proposal. The two documents are delivered as two commented R code, to meet ESPON requirements in terms of complete transferability of the methodology and full reproducibility. Details regarding each documents are provided in section 3.

Figure 3.1 - Gantt Diagram of the project

3.1 Task 1: Development of a methodological framework

The main purpose of Task 1 is to frame a methodological framework to conduct a feasibility study and a survey the possible data sources in the countries (France, UK, Poland, Spain, Norway, Sweden and Switzerland) and cities of expertise (cf. case studies).

<table>
<thead>
<tr>
<th>Subtask</th>
<th>Milestone for D1</th>
<th>File</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) to describe the data collection process, both conventional and unconventional.</td>
<td>Work in progress</td>
<td>Delivery_listing_data_providers.xlsx</td>
</tr>
<tr>
<td>(2) to describe and document the methodology employed to harvest datasets, using APIs, and R packages s.a. rvest and other associated packages. Partners with experience of IDS harvesting / scrapping using these methodologies are CISC, Spain ; Lodz Univ., Poland and UMS RIATE</td>
<td>X</td>
<td>D1_Draft_outline_guidance_document.html</td>
</tr>
</tbody>
</table>
(4) To describe a set of harmonized variables in Europe, which data sources are used, what methods have been applied. Harmonized variables should be made comparable across European cities, and over time. Ratios and standardized indices, such as affordability ratio will be considered as valuable alternative to rough data (s.a. price, surface), that are structurally contingent to each country, city and local market context.

(5) To analyze for each datasets and variables how it can be used to document the dynamics of markets, over time. Spatio-temporal information is sensitive to two types of sampling issues: in space, and in time, therefore requiring the use of various interpolation and estimations procedures.

(7) The framework will also characterize and document the missing informations, that would be of value to undertake a full-scale survey of housing and living conditions. While some tax data (local tax, property tax, propriety asseessment) on households may be obtained through fiscal administration, a full characterization of living conditions would include a debt-to-value ratio or a characterization of households/ individual assets invested on the property markets. To our knowledge, no banking / credit database would allow data mining and harvesting from unconventional data sources or IDS (cf. credit scoring, income data, outstanding debt of households).

| (4) | To describe a set of harmonized variables in Europe, which data sources are used, what methods have been applied. Harmonized variables should be made comparable across European cities, and over time. Ratios and standardized indices, such as affordability ratio will be considered as valuable alternative to rough data (s.a. price, surface), that are structurally contingent to each country, city and local market context. | X | D1_Draft_outline_guidance_document.html Harmonized and standardized variables are proposed, down to the local level (LAU2). |
| (5) | To analyze for each datasets and variables how it can be used to document the dynamics of markets, over time. Spatio-temporal information is sensitive to two types of sampling issues: in space, and in time, therefore requiring the use of various interpolation and estimations procedures. | In progress | D1_Draft_outline_guidance_document.html Spatial interpolation procedures are described using for instance the Spatialposition R Package. |
| (7) | The framework will also characterize and document the missing informations, that would be of value to undertake a full-scale survey of housing and living conditions. While some tax data (local tax, property tax, propriety asseessment) on households may be obtained through fiscal administration, a full characterization of living conditions would include a debt-to-value ratio or a characterization of households/ individual assets invested on the property markets. To our knowledge, no banking / credit database would allow data mining and harvesting from unconventional data sources or IDS (cf. credit scoring, income data, outstanding debt of households). | Work in progress | Delivery_listing_data_providers.xlsx |

3.2 Task 2: Gathering data sets

Available data by partners for a European-wide analysis

To address the request of the ToR for a European-wide study, we focus on a feasibility study, based upon the expertise of partners in a selection of european countries, and case studies. The expertise of partners builds upon previous work and a good knowledge of databases already existing (conventional and institutional), that have been used in many of their scholarly work and studies. The service will elaborate on this knowledge of conventional database to inventory webservices and unconventional IDS.

In the framework of this study, our understanding of the tasks and feasibility command that data will be gathered thoroughly for case studies, and exploratory harvesting will be conducted exclusively in FUAs cities in the countries of expertise. The final report will issue recommendations and quality assessment of data collected to envision a full deployment of harvesting / storage / analysis in every FUA cities. We will gather data in order to complement
administrative data from sources such as the ESPON database and Eurostat should cover cities (FUAs) included in the Urban Audit database.

<table>
<thead>
<tr>
<th>Subtask</th>
<th>Milestone for D1</th>
<th>File</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) We will <strong>detail availability, access, and status</strong> (commercial/proprietary,.opendata) of data available. Available data will serve as a guide to determine harvested data and standardize and harmonized variables suitable for the project, in order to make informed proposal in Delivery 1. We will analyze the <strong>feasibility of collecting data and construct harmonized variables for the case studies</strong>, based on the methodological framework.</td>
<td>Work in progress</td>
<td>Delivery_listing_data_providers.xlsx</td>
</tr>
<tr>
<td>(2) Alongside with the <strong>countries of expertise</strong>, some case study should be <strong>carefully chosen</strong>, for (1) their interest and feasibility in terms of data harvesting and scrapping possibilities; (2) the significance of the case study; (3) the current availability of knowledge and data on the case studies, to assess the quality of data harvested. but also for their significance for each of the case studies. This selection of case studies shall be confirmed in delivery 1, after assessment of data harvesting and scrapping possibilities for each of the case studies.</td>
<td>X</td>
<td>D1_case_study_presentation.html</td>
</tr>
</tbody>
</table>

All proposed case studies are listed, for ESPON EGTC approval, in the document, alongside with a description and basic descriptive statistics.

### 3.3 Task 3: Analysing data sets

An empirically grounded and informed policy making process requires a set of variables that are to be standardized and harmonized. Usefulness will be assessed, as well as availability. The delivered table [Delivery_listing_data_providers.xlsx] and the methodological framework [D1_Draft_outline_guidance_document.html] give an overview of some of the variables that we have identified as possible to obtain for case studies either using conventional data, or unconventional data. Rents may be obtained through different unconventional websources: real estate agent listings (long term) and Airbnb (short term and tourism). Living environment (households income, composition of households, population development) will be obtained using publicly available census datasets. Harmonized variables should be made comparable across European cities, and over time. Ratios and standardized indices, such as affordability ratio will be considered as alternatives to standard variables (s.a. price, surface), that are structurally contingent to each country, city and market context. The iterations for harmonisations, representativeness and quality have been assigned to Task 1, and the process fully described, in a reproducible way, in the methodological framework [D1_Draft_outline_guidance_document.html], as part of delivery #1.

<table>
<thead>
<tr>
<th>Task</th>
<th>Milestone for D1</th>
<th>File</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) <strong>An empirically grounded and informed policy making process</strong></td>
<td>Work in progress</td>
<td>- Delivery_listing_data_providers.xlsx</td>
</tr>
</tbody>
</table>
requires a set of variables that are to be standardized and harmonized.  

- An additional proof of concept has been delivered to ESPON as an additional participation, besides the original contract, as a paper prepared for the London Conference, Nov. 2018.

(2) Detailed analysis for case studies, with harmonized variables, allowing comparisons between and across cities and countries.  

Work in progress  

3.4 Task 4: Reporting results

We are committed to make the results available and transferable for a better informed policy making process. The following table presents the current state of workplan for task 4.

<table>
<thead>
<tr>
<th>Subtasks</th>
<th>Milestone for D1</th>
<th>File</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) a report written in short, clear, grounded language, describing the wellbeing of European citizens focused on affordable housing and their living situation.</td>
<td>X</td>
<td>Table of content for the report describing the wellbeing of European citizens according to “territorial analysis and housing dynamics”.</td>
</tr>
<tr>
<td>=&gt; Delivery #3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) data, maps, graphs and other resources used as main input for the report mentioned in the report.</td>
<td>Work in progress</td>
<td>D1_case_study_presentation.html and D1_Draft_outline_guidance_document.html</td>
</tr>
<tr>
<td>Findings will be grounded in results presented as maps, table, designed according to ESPON standards and templates.</td>
<td></td>
<td>NB : after ESPON’s approval of Delivery#1, we will prepare the layout of graphs and maps describing harmonized variables according to ESPON standards and templates</td>
</tr>
</tbody>
</table>

3.5 Task 5: Communicating methodologies

The aim of the service is to make methodologies (methodological framework) available and applicable to others.

<table>
<thead>
<tr>
<th>Subtasks and deliveries</th>
<th>Milestone for D1</th>
<th>File</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) a guidance document describing how the methods can be used.</td>
<td>Work in progress</td>
<td>D1_Draft_outline_guidance_document.html</td>
</tr>
<tr>
<td>With the user and the reader in mind, the documents, reports, and guidance documents will explain how to use and apply in practice the methodological framework related to the data gathering and indicators calculation and analysis. Maps and Maps and graphs will convey direct messages and will be complemented with</td>
<td></td>
<td>Detailed code is provided and documented. The framework includes an overview of possible data source, data harvesting platforms.</td>
</tr>
</tbody>
</table>
short explanations for policy makers. Guidance documents will be delivered as ‘technical reports’ (30-40 p.). A technical report is a synthetic document of no more than 30-40 pages. It provides step-by-step guidance and can easily be consulted when a project group or members of the ESPON EGTC would like to clarify a specific point. In all guidance documents, screenshots, code examples, with explanatory comments will occupy a prominent place as well as figures describing the different steps to be followed to produce maps and analysis of the housing dynamics and wellbeing of European dwellers. The guidance document, Data and metadata will follow the metadata model of the ESPON 2020 Database Portal which follows the international (ISO-19115) and European (INSPIRE) standards. UMS RIATE has expertise in delivery documents, maps and data in the required ESPON formats. The ESPON 2020 MapKit templates will be implemented, and the ESPON corporate identity and the rules of representations will be strictly abide to. The guidance documents will issue technical recommendations and feasibility assessment if a full-scale project, covering a large set of European FUAs were to be launched.

(2) proposals for further developments in relation to the use of big data within the ESPON 2020 Programme.
=> delivery #3
4 Reminder of the organisation of the first delivery (6 November 2018)

The delivery includes 4 files:

1) A proposal for the case study cities. A proposal for the case-studies. This html file (D1_case_study_presentation.html) combines R code and findings (to ensure the full reproducibility of the elements displayed in the file). It presents the spatial extent of the case-studies and some basic indicators derived from the Eurostat database related to the housing dynamics. The case-studies have been selected taking into account the following considerations: (1) Presence of at least one expert/partner knowing the relevant data sources and having the thematic expertise for analyzing housing dynamics in her/his country. (2) Stock of relevant indicators for analyzing the housing dynamics market. (3) Complementarity of case-studies for analyzing several European patterns (housing market, size of the city, geographical diversity, socio-economics characteristics, etc.)

2) A table of content for the guidance document. This document (D1_Draft_outline_guidance_document.html) is structured as a R Markdown document in order to ensure a full methodological reproducibility of the results. This document elaborates on the prefigurative case of Paris as an ideal case study because of the availability of institutional data. We consider a variety of datasets: property-level data from the Paris Chamber of Notaries (1996-2012, a sample of 1 million rows), public data, compared to possible harvested big data sources (real estate websites; opendata resources and warehouses; Airbnb…). Harmonized and standardized variables are proposed, down to the local level, such as price-to-income ratio, sq.meters per month of income, changes in the balance between inward and outward flows of sellers and buyers, relative density of Airbnb rental. To characterize the local conditions of housing market, theses variables are used to map and visualize price dynamics, inequalities, social change and flows of sellers and buyers, unequal pressure of vacation rental markets. Methods are applied to subset of data available to test and demonstrate to ESPON the feasability of the methodology.

3) A proposal for a set of indicators to measure the housing dynamics, affordable housing and the overall living environment of European citizens: This table (Delivery_listing_data_providers.xlsx) lists all the available resources (institutional data, relevant sources for Web-Scrapping). It describes the availability of data sources according to the geographical extent (is it possibile to deliver data for the whole FUA or not?), the geographical granularity (LAU2, points location, etc.), the temporal extent (how long?) and the thematic (what kind of indicators?). This table enables us to propose a set of harmonized indicators for the overall project, and propose some specific analysis according to the data sources and the thematic relevance regarding to the housing market.
4) **A table of content for the report describing the wellbeing of European citizens according to “territorial analysis and housing dynamics”**. Following the ESPON template (please, confirm we use the current template), the table of content describes how this document will be structured. It aims at summarizing in a written report all the inputs derived from our project. The final document will be synthetic (30-40 pages max.)
References


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.