Web data extraction systems versus research collaboration in sustainable planning for housing: smart governance takes it all

Valerie Dewaelheyns\(^1\), Isabelle Loris\(^2,3\), Anuja Dangol\(^1\) and Thérèse Steenberghen\(^1\)

\(^1\) Spatial Applications Division Leuven (SADL), KU Leuven
\(^2\) Center for Mobility and Spatial Planning, Ghent University
\(^3\) Housing and Urban Studies, KU Leuven

REALCORP2016, Hamburg
introduction

Smart cities, open data, sustainable planning and housing
Smart cities: open data

WHAT MAKES A SMART CITY SMART?

Open Data

http://europe-ul.com/makes-smart-city-smart/
Smart planning
Micro-dynamics of housing market?
Monitoring by policy-support indicators
Research objective and question

• **Objective:** development of a proof-of-concept of two ‘open’ (e.g. freely available and accessible) policy-support indicators on micro-dynamics of the housing market

• **Focus:** micro-data of real estate listings

• **Research question:** What are suitable methods or strategies to collect data on micro-dynamics of the housing market?
material and methods

Housing market indicators, web data extraction, transdisciplinary research collaboration, case study Leuven
Indicators

- **Speed of sale**: duration that houses are listed for sale on the market (‘time-on-market’)
  - Proxy = time that a listing is published online

**Interacts with** (Clauretie and Thistle 2007; Johnson, Benefield et al. 2007)

- **Listing price**: price of the real estate listing, this differs with the realized price
Web Data Extraction
• **Web Data Extraction System**: extract and collect unstructured or semi-structured data that are stored or published on Web sources

(Laender, Ribeiro-Neto et al. 2002; Sarawagi 2008; Ferrara, De Meo et al. 2014)
<table>
<thead>
<tr>
<th>Prijs</th>
<th>Dagen</th>
<th>Woonvlak</th>
<th>Adres</th>
</tr>
</thead>
<tbody>
<tr>
<td>€ 199.000</td>
<td>77</td>
<td>60m²</td>
<td>3350 Berchem</td>
</tr>
<tr>
<td>€ 499.500</td>
<td>94</td>
<td>100m²</td>
<td>3003 Haasrode</td>
</tr>
<tr>
<td>€ 349.500</td>
<td>84</td>
<td>140m²</td>
<td>3010 Kussen-Lu</td>
</tr>
<tr>
<td>€ 495.000</td>
<td>87</td>
<td>270m²</td>
<td>3001 Heverlee</td>
</tr>
<tr>
<td>€ 630.000</td>
<td>89</td>
<td>238m²</td>
<td>3033 Haasrode</td>
</tr>
<tr>
<td>€ 270.000</td>
<td>95</td>
<td>120m²</td>
<td>3001 Heverlee</td>
</tr>
<tr>
<td>€ 345.000</td>
<td>90</td>
<td>190m²</td>
<td>3150 Heist</td>
</tr>
<tr>
<td>€ 285.000</td>
<td>97</td>
<td>180m²</td>
<td>3051 Sint-Jans-Meiwer</td>
</tr>
<tr>
<td>€ 239.000</td>
<td>98</td>
<td>95m²</td>
<td>3000 Leuven</td>
</tr>
</tbody>
</table>
Transdisciplinary research collaboration

- **Involved partners:** university researchers on geosciences and –technology, the spatial planning policy department of the Flemish government, and private businesses active on the real estate market

- **Sharing of system knowledge:** knowledge on the current state of given system and its ability to change
  
  (Hirsch Hadorn, Bradley et al. 2006; Brandt, Ernst et al. 2013)

- Situated at the ‘consultation’ level, but aims at initiating a process that can **evolve towards the level of ‘collaboration’**
  
  (Pretty 1994; Leeuwis 2000; Aarts and Leeuwis 2010; Lang, Wiek et al. 2012; Brandt, Ernst et al. 2013)
Transdisciplinary research collaboration

- Initiate dialogue with two private companies through a qualitative research approach:
  - semi-structured explorative interviews: 2/8 portal sites
  - purposive sampling

- three goals
  - (1) initiate the transdisciplinary research collaboration
  - (2) gain insights in the housing market;
  - (3) explore the development of housing dynamics indicators

(Strauss and Corbin 1998; Maxwell 1997; Patton 2002; Guarte and Barrios 2006; Teddlie and Yu 2007; Creswell 2008)
Micro-data real estate listings

- 10 year database of listings
- scraped small(er) independent real estate agents websites across Belgium
- more than 110,000 listings on the Web
**Data cleaning:**
- *selection* of relevant and sufficiently documented records (houses and apartments for sale, municipalities of Leuven and Herent)
- *transformation* of data to usable formats (listing price, housing numbers, dates, geographic coordinates, etc.).
- open source data cleaning tool “OpenRefine”
  (Verborgh and De Wilde 2013)
Indicator calculation

- Year 2011
- Speed of sale = difference in number of days between the first and last date of publication on the portal website
  - Variables “Pub1 start” and “Pub1 stop”
- Listing price = variable ‘initial asking price’
  - compared with
    - the average selling price tracked by Belgian censuses
    - the average realized price for the entire dataset (2005-2014)
    - Pricing difference (%) = 
      
      \[
      \frac{(realized \ \text{price} - \ \text{listing \ price})}{\ \text{listing \ price}} \times 100\].
Case study: greater Leuven (Flanders, Belgium)

**INHabitants**
- 119,410 inhabitants (2015)
- 58,636 households (2015)
- Increase pop. of 9.6% (2005-2015)

**Built Environment**
- Population density 13 inh/ha (2015)
- Av. 2.3 persons per dwelling (2011)
- 51,683 buildings (2011)
- 31% built-up space (2015)

**Ownership**
- Tenant: 37%
- Owner: 63%
  (weighted average, 2011)

**Property Prices**
- House: €449,000
- Villa: €614,131
  (average, 2014)
- Apartment: €336,027
- Building plot: €682/m²

Isabelle Loris, based on data from Statistics Belgium (beSTAT.be) and censusdata (Census2011.be).
Micro-data real estate listings
“We don’t prefer it [web scraping], because it is a very suboptimal way to transfer data. We translate our data to html, made for browsers, and then you translate through html this information into data again. Many data gets lost and is very computer intensive. We have to generate all those web pages at our server side, you have to collect all the data and process it. There are just better ways to collect data.” Company B
1. Web data extraction

2. Exploration
   web data
   and testing
   prototype indicators

3. Transdisciplinary collaboration
   users' agreement
   between research partner
   and one company

- technical & juridical barriers
- suboptimal data transfer
- collaboration is a better option

- shared challenge
- data exchange
- support
- data quality control
- data preparation for analysis

PROOF
speed of sale
listing price
Speed of sale (number of days for sale):

- **Not sold yet**: 23%
- **> 6 months**: 18%
- **6 months**: 11%
- **3 months**: 19%
- **1 month**: 9%
- **2 weeks**: 19%

Legend:
- ×: Not yet sold
- ●: 2 weeks
- ●: 1 month
- ●: 3 months
- ●: 6 months
- △: from 6 months up to 4 years
- Urban area
conclusion
• **Goal:** development of policy-support indicators for spatial planning

• **Proof-of-concepts:** explicit housing market dynamics in space and time at a detailed scale level

• **Points of attention**
  o calculated for 2011 and for one case
  o actual interpretations of the data and comparisons with other cities or benchmarks are not possible
• **Way forward**
  o map dynamics of the housing market in time and space  
    - bench marks for a fast or slow speed of sale?  
  o evaluate in which neighborhoods the housing market is very dynamic and where it is dull  
• The importance of monitoring over longer time series: housing dynamics can change rather fast, where spatial planning is characterized by a certain level of slowness.

  “Where would it be opportune to stimulate or limit the housing supply? Are we planning at the right places? What if vibrant housing market dynamics appear there where we are not looking?”
1. Web data extraction

2. Exploration web data and testing prototype indicators
   - technical & juridical barriers
   - suboptimal data transfer
   - collaboration is a better option

3. Transdisciplinary collaboration
   - shared challenge
   - data exchange
   - support
   - data quality control
   - data preparation for analysis
   - users' agreement between research partner and one company

4. Future co-production
   - policy
   - science
   - companies
   - past
   - policy
   - science
   - companies
   - current
   - policy
   - science
   - companies
   - near future collaboration
   - long-term users' agreement between government and one company

Smart governance for smart cities

far future self-mobilisation
The Flemish Policy Support Centre on Space (2012-2015) funded this research. The views expressed in this paper are not necessarily those of the Flemish Government.

Detailed real estate data were kindly provided by Zimmo.
thank you

questions?

therese.steenberghen@kuleuven.be
isabelle.loris@ugent.be