



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

Valerie Dewaelheyns<sup>1</sup>, Isabelle Loris<sup>2,3</sup>, Anuja Dangol<sup>1</sup> and Thérèse Steenberghen<sup>1</sup>

<sup>1</sup> Spatial Applications Division Leuven (SADL), KU Leuven

<sup>2</sup> Center for Mobility and Spatial Planning, Ghent University

<sup>3</sup> Housing and Urban Studies, KU Leuven



# introduction

Smart cities, open data, sustainable  
planning and housing

# Smart cities: open data



## WHAT MAKES A SMART CITY SMART?



# Smart planning



# *Micro-dynamics of housing market?*



# Monitoring by policy-support indicators



# Research objective and question

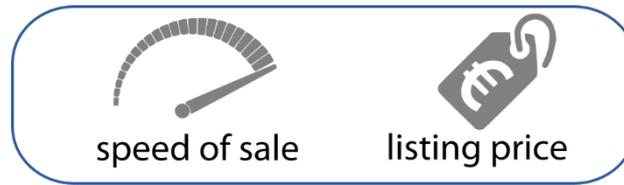
- **Objective:** development of a proof-of-concept of two 'open' (e.g. freely available and accessible) policy-support indicators on micro-dynamics of the housing market
- **Focus:** micro-data of real estate listings
- **Research question:** *What are suitable methods or strategies to collect data on 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 (Clauret 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

Uw zoekresultaten



Vind ik leuk 14k | Facebook | Twitter | Zimmo code NL

PUBLICIEER UW PAND | MEER - | MIJN ZIMMO Aanmelden | Registreren

TE KOOP | TE HUUR | NIELWOUWPROJECTEN | BEDRIJFSVASTGOED | BELEGGINGSVASTGOED | TE FLIRT | WOONINSPIRATIE

## Huizen te koop in Leuven (3000)

GALERIJ & KAART | GALERIJ | KAART

OPSLAAN | ZOEKCRITERIA AANPASSEN

1 - 30 van 69 resultaten

Pagina 1

Sorteer op AANBEVOLEN

145m<sup>2</sup> 3 34

€ 345.000

Jan Stas

Huis te koop »  
Mechelsesteenweg 41  
3000 Leuven

Zimmo Code 83284

166m<sup>2</sup> 5 37

€ 350.000

Jan Stas

Huis te koop »  
Suijstraat 29  
3000 Leuven

Zimmo Code 83482

180m<sup>2</sup> 3 91

€ 345.000

Jan Stas

Huis te koop »  
Dagbortestraat 1  
3000 Leuven

Zimmo Code 83503

91m<sup>2</sup> 2 142

€ 255.000

Jan Stas

Huis te koop »  
Nervierstraat 7  
3000 Leuven

Zimmo Code 83493

IN PRIJS VERLAAGD

342m<sup>2</sup> 5 140

€ 1.200.000

Jan Stas

Huis te koop »  
Mechelsestraat 85  
3000 Leuven

Zimmo Code 83113

135m<sup>2</sup> 4 85

€ 290.000

Jan Stas

Huis te koop »  
Brouwersstraat 80  
3000 Leuven

Zimmo Code 84008

Leuven - Vaartkom  
I love Hungaria

ONTDEK DIT PROJECT

www.janstas.be

143m<sup>2</sup> 3

€ 880.000

Jan Stas

Huis te koop »  
Koning Leopold I-Straat 25  
3000 Leuven

Zimmo Code 81974

NIET WILLEN

4 29

€ 335.000

Jan Stas

Huis te koop »  
Burgemeestersstraat 39  
3000 Leuven

Zimmo Code 83480

8 51

€ 455.000

Jan Stas

Huis te koop »  
Vlaamingerstraat 94  
3000 Leuven

Zimmo Code 84374

2 70

€ 405.000

Jan Stas

Huis te koop »  
Tierssestraat 167  
3000 Leuven

Zimmo Code 83090

166m<sup>2</sup> 3 71

€ 355.000

YOURHOUSE

Huis te koop »  
Hogestraat 30  
3000 Leuven

Zimmo Code 83810

250m<sup>2</sup> 3 7

350m<sup>2</sup> 7 8

175m<sup>2</sup> 4 19

195m<sup>2</sup> 3 29

600m<sup>2</sup> 8 48

- **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)



import **io**

121 - 150 van 460 resultaten Pagina 5 Sorteer op AANBEVOLEN

IN PRIJS VERLAAGD

- € 199.000 - Wijnstraat 3, 3360 Bierbeek
- € 499.500 - Konynghoekstraat 63, 3003 Haastrode
- € 349.500 - Blatte Loostraat 37 L, 3010 Kessel-Lo
- € 495.000 - Kessel-Lo

	prijs	dagen	woonvlak	adres
1	€ 199.000	77	80m <sup>2</sup>	3360 Bierbeek
2	€ 499.500	77	200m <sup>2</sup>	3053 Haastrode
3	€ 349.500	84	140m <sup>2</sup>	3010 Kessel-Lo
4	€ 495.000	87	270m <sup>2</sup>	3001 Heverlee
5	€ 630.000	89	239m <sup>2</sup>	3053 Haastrode
6	€ 270.000	95	120m <sup>2</sup>	3001 Heverlee
7	€ 245.000	96	190m <sup>2</sup>	3150 Haacht
8	€ 265.000	97	132m <sup>2</sup>	3051 Sint-Joris-Weert
9	€ 239.000	98	96m <sup>2</sup>	3000 Leuven

Train rows

Teach us where the results are...

Train by example - highlight a result on the page and click the Train button

Repeat until you're happy

I've got all 30 rows!

A bit stuck here? Here's a tutorial on how to train rows

Cancel

Huisen te koop in uw g... Huisen te koop in uw g...

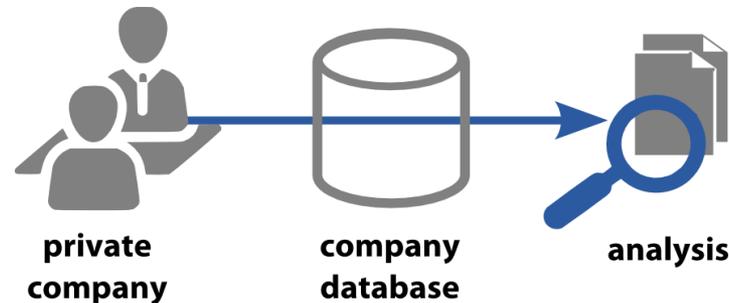
# Transdisciplinary research collaboration

- **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

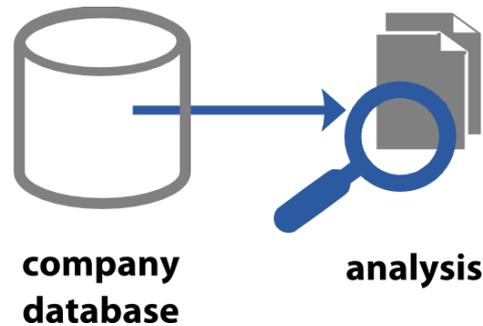
- **Initiate dialogue with two private companies through a qualitative research approach:**
  - semi-structured explorative interviews: 2/8 portal sites
  - purposive sampling



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

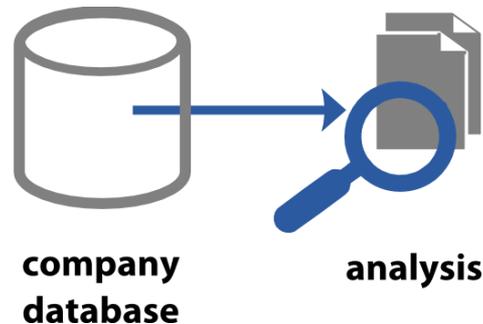
(Strauss and Corbin 1998; Maxwell 1997; Patton 2002; Guarte and Barrios 2006; Teddlie and Yu 2007; Creswell 2008)

# 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

# Micro-data real estate listings



- **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 (%) =  
$$\left[ \frac{\text{realized price} - \text{listing price}}{\text{listing price}} * 100 \right]$$

# 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



37% vs 63%  
(weighted average, 2011)

## PROPERTY PRICES



house  
€449,000



villa  
€614,131



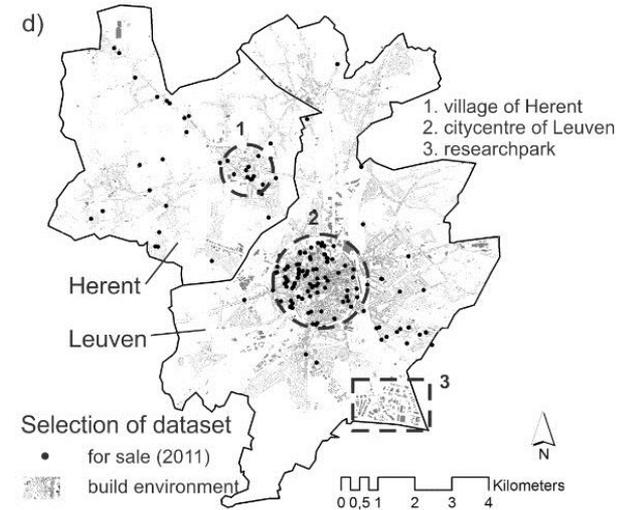
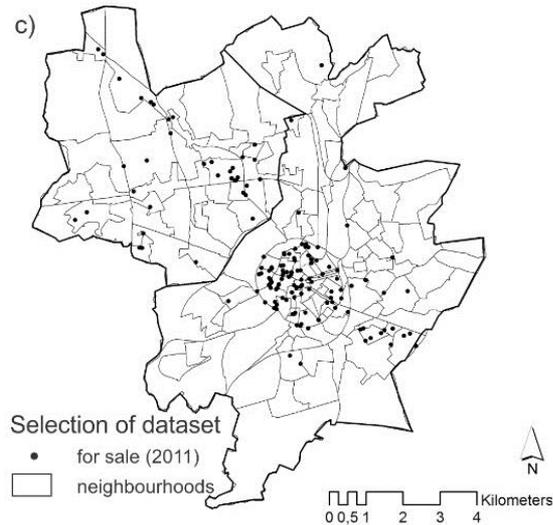
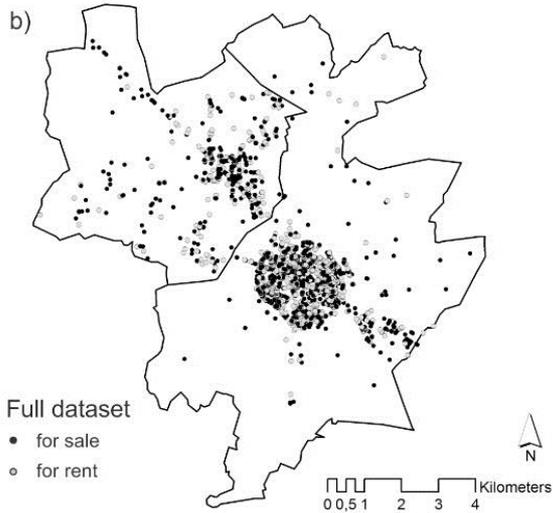
apartment  
€336,027

(average, 2014)



building plot  
€682/m<sup>2</sup>

# Micro-data real estate listings



results



## 1. Web data extraction



## 2. Exploration web data and testing prototype indicators



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

*“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



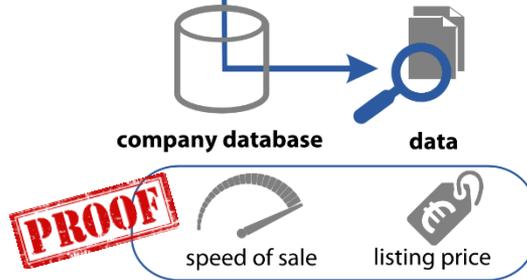
## 2. Exploration web data and testing prototype indicators

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

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

## 3. Transdisciplinary collaboration

users' agreement  
between research partner  
and one company



# PROOF



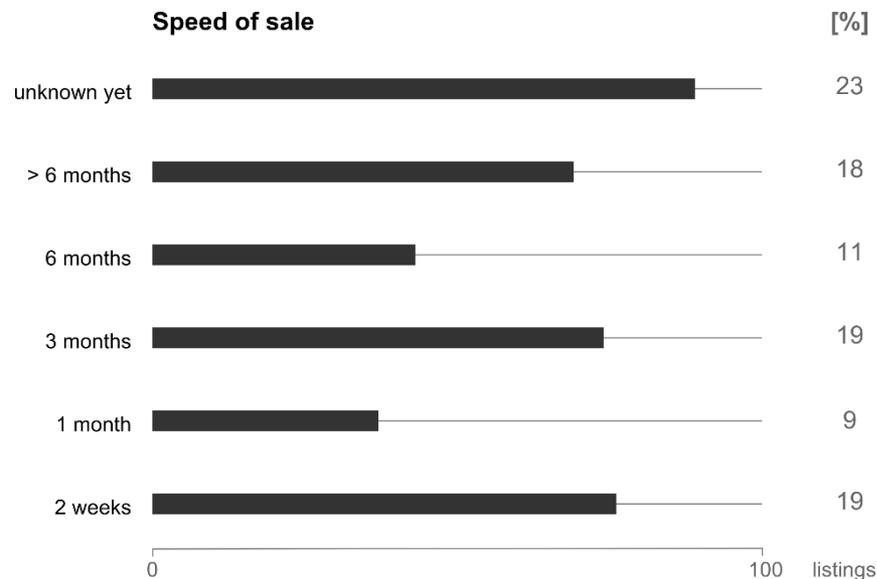
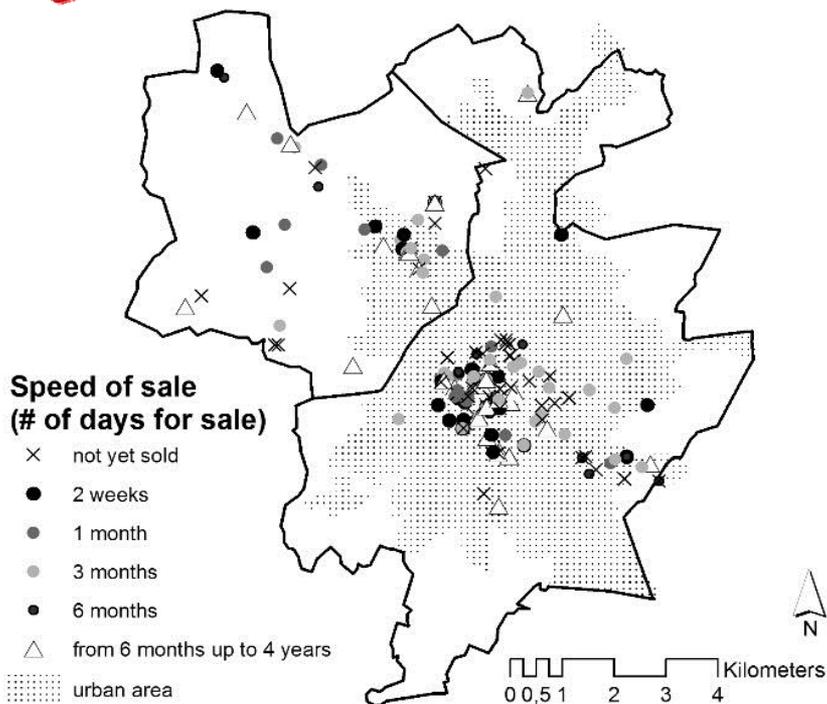
speed of sale



listing price



## Leuven & Herent



# PROOF



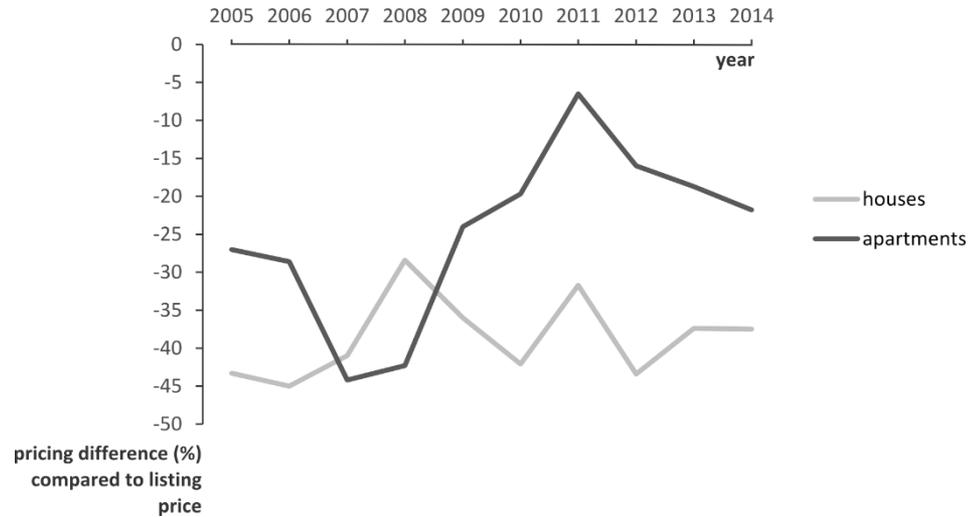
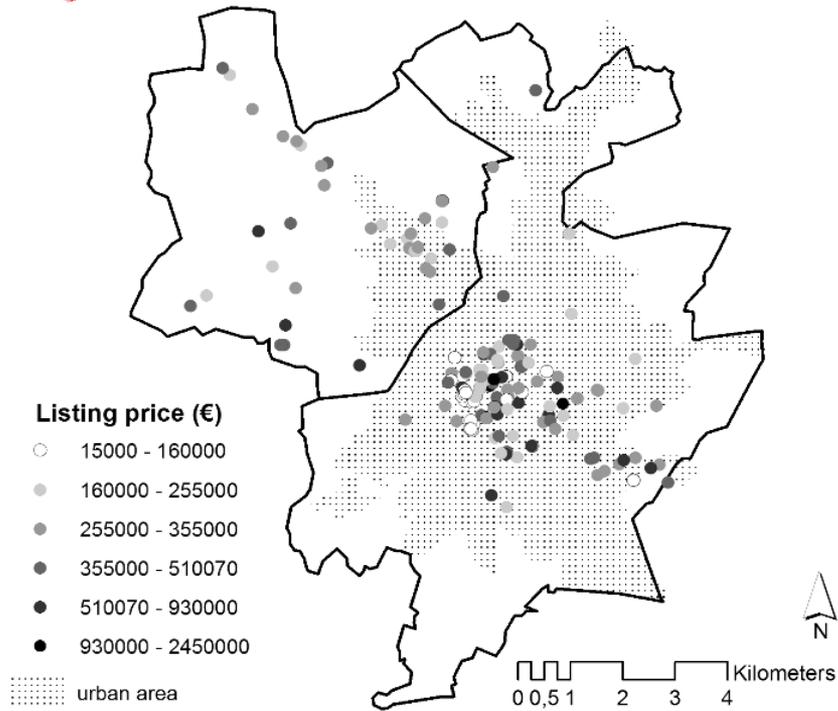
speed of sale



listing price

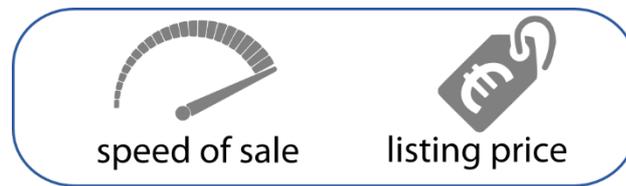


## Leuven & Herent



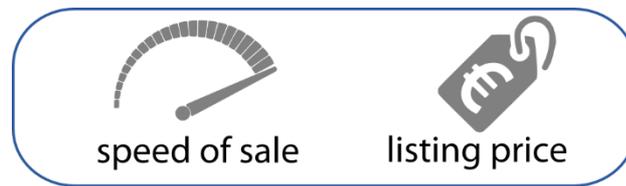
conclusion





Leuven &  
Herent

- **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**
  - calculated for 2011 and for one case
  - actual interpretations of the data and comparisons with other cities or benchmarks are not possible



Leuven &  
Herent

- **Way forward**
  - map dynamics of the housing market in time and space
    - bench marks for a fast or slow speed of sale?
  - 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



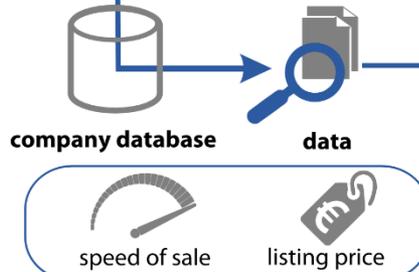
- shared challenge
- data exchange
- support
- data quality control
- data preparation for analysis

## 2. Exploration web data and testing prototype indicators

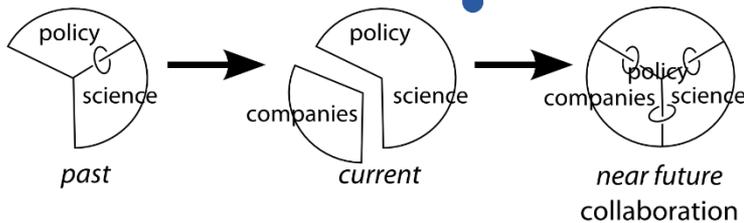
- technical & juridical barriers
- suboptimal data transfer
- collaboration is a better option

## 3. Transdisciplinary collaboration

users' agreement between research partner and one company



## 4. Future co-production



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](mailto:therese.steenberghen@kuleuven.be)  
[isabelle.loris@ugent.be](mailto:isabelle.loris@ugent.be)