



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

Valerie Dewaelheyns¹, Isabelle Loris^{2,3}, Anuja Dangol¹ and Thérèse Steenberghen¹

- 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?



Smart planning



Micro-dynamics of housing market?



Monitoring by policy-support indicators



Research objective and question

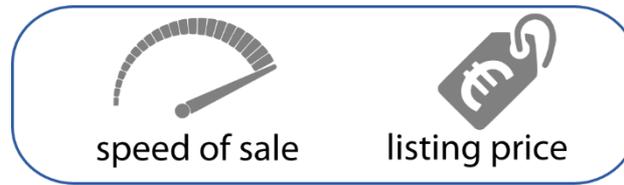
- **Objective:** development of a proof-of-concept of two 'open' (e.g. freely available and accessible) policy-support indicators on micro-dynamics of the housing market
- **Focus:** micro-data of real estate listings
- **Research question:** *What are suitable methods or strategies to collect data on 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 (Clauretje 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 f t p Zimmo code NL

PUBLICEER UW PAND | MEER - MIJN ZIMMO Aanmelden | Registreren ★ 0

TE KOOP | TE HUUR | NIELWBOUWPROJECTEN | 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



€ 345.000
Huis te koop »
Mechelissenweg 41
3000 Leuven
Zimmo Code 83984



€ 350.000
Huis te koop »
Suisstraat 29
3000 Leuven
Zimmo Code 83982



€ 345.000
Huis te koop »
Dagbortstraat 1
3000 Leuven
Zimmo Code 83983



€ 255.000
Huis te koop »
Nervierstraat 7
3000 Leuven
Zimmo Code 83985
IN PRIJS VERLAAGD

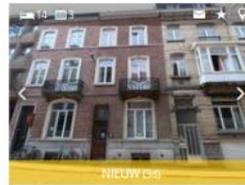


€ 1.200.000
Huis te koop »
Mechelissenstraat 85
3000 Leuven
Zimmo Code 83979



€ 290.000
Huis te koop »
Brouwerstraat 80
3000 Leuven
Zimmo Code 83980

Leuven - Vaartkom
I love Hungaria
www.zimmojanstas.be
ONTDEK DIT PROJECT



€ 880.000
Huis te koop »
Koning Leopold I-Straat 25
3000 Leuven
Zimmo Code 83986
NIEUW LSH



€ 335.000
Huis te koop »
Burgemeesterstraat 39
3000 Leuven
Zimmo Code 83987



€ 455.000
Huis te koop »
Vamingenstraat 94
3000 Leuven
Zimmo Code 83974



€ 405.000
Huis te koop »
Tiersenstraat 167
3000 Leuven
Zimmo Code 83988



€ 355.000
Huis te koop »
Hogestraat 30
3000 Leuven
Zimmo Code 83989
YOURHOUSE



- **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

Grid of real estate listings. Each listing includes a photo, price, and basic details. The first listing is highlighted with a yellow box around its price.

- Listing 1: € 199.000, 77 dagen, 80m², 2350 Bierbeek
- Listing 2: € 499.500, 77 dagen, 200m², 3053 Haasrode
- Listing 3: € 349.500, 84 dagen, 140m², 3010 Kessel-Lo
- Listing 4: € 495.000, 87 dagen, 210m², 3001 Heverlee
- Listing 5: € 630.000, 89 dagen, 239m², 3053 Haasrode
- Listing 6: € 270.000, 95 dagen, 120m², 3001 Heverlee
- Listing 7: € 245.000, 96 dagen, 190m², 3150 Haacht
- Listing 8: € 265.000, 97 dagen, 132m², 3051 Sint-Joris-Weert
- Listing 9: € 239.000, 98 dagen, 96m², 3000 Leuven

	prijs	dagen	woonvlak	adres
1	€ 199.000	77	80m²	2350 Bierbeek
2	€ 499.500	77	200m²	3053 Haasrode
3	€ 349.500	84	140m²	3010 Kessel-Lo
4	€ 495.000	87	210m²	3001 Heverlee
5	€ 630.000	89	239m²	3053 Haasrode
6	€ 270.000	95	120m²	3001 Heverlee
7	€ 245.000	96	190m²	3150 Haacht
8	€ 265.000	97	132m²	3051 Sint-Joris-Weert
9	€ 239.000	98	96m²	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...

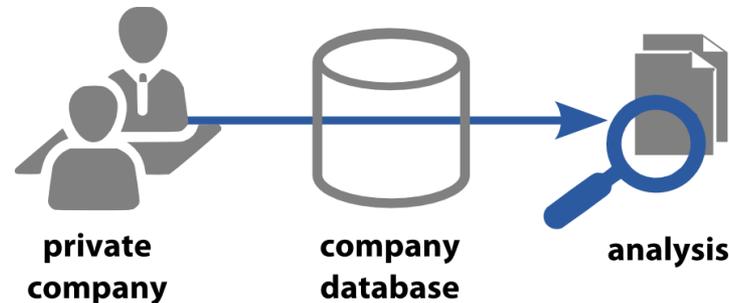
Transdisciplinary research collaboration

- **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

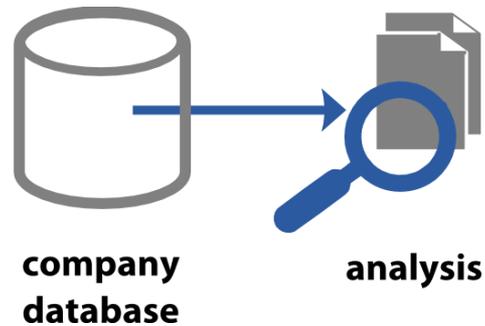
- **Initiate dialogue with two private companies through a qualitative research approach:**
 - semi-structured explorative interviews: 2/8 portal sites
 - purposive sampling



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

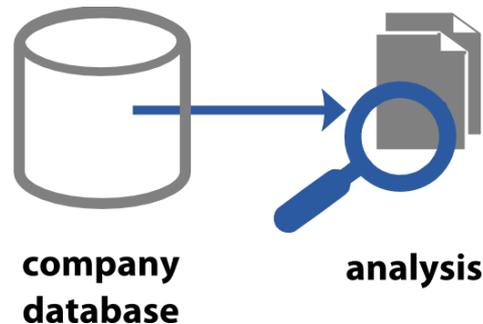
(Strauss and Corbin 1998; Maxwell 1997; Patton 2002; Guarte and Barrios 2006; Teddlie and Yu 2007; Creswell 2008)

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

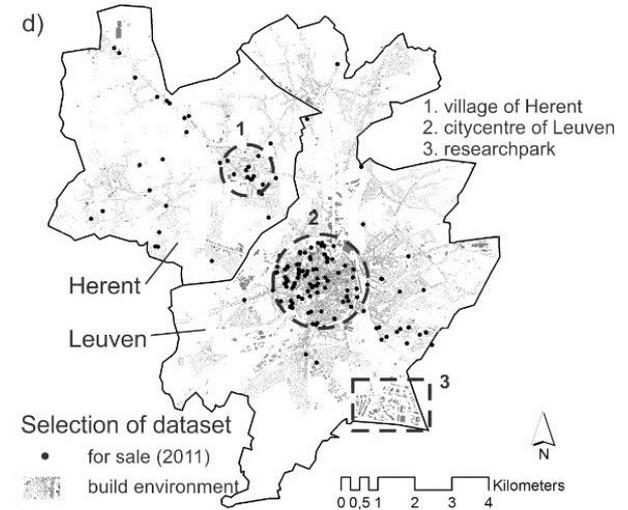
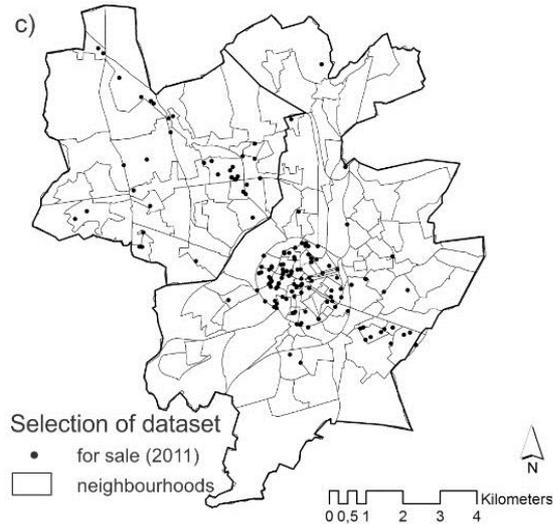
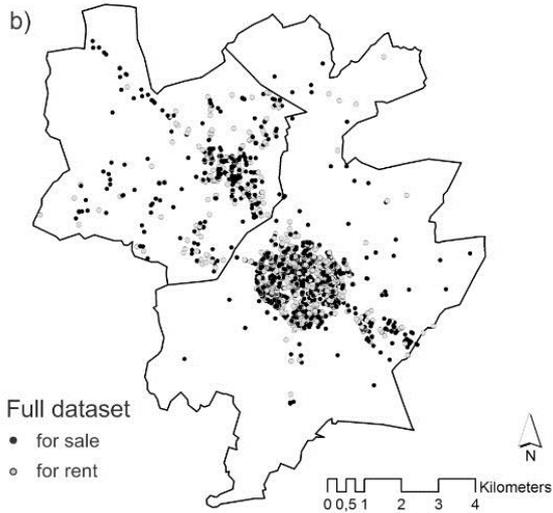


apartment
€336,027



building plot
€682/m²

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



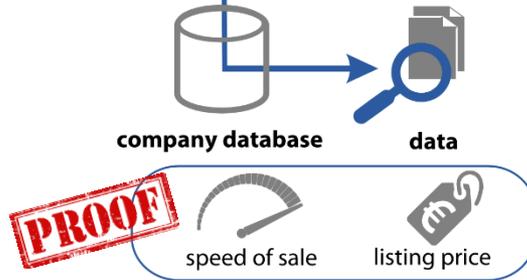
2. Exploration web data and testing prototype indicators

- 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



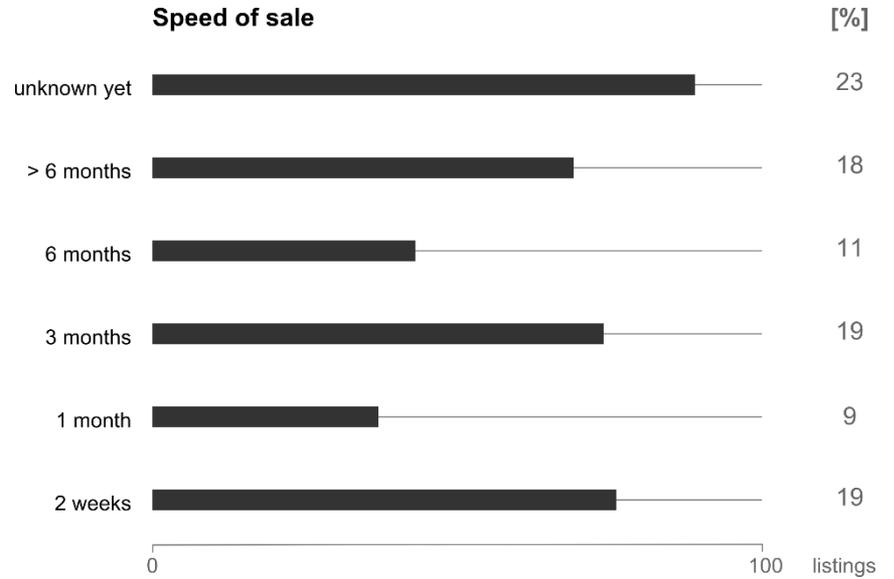
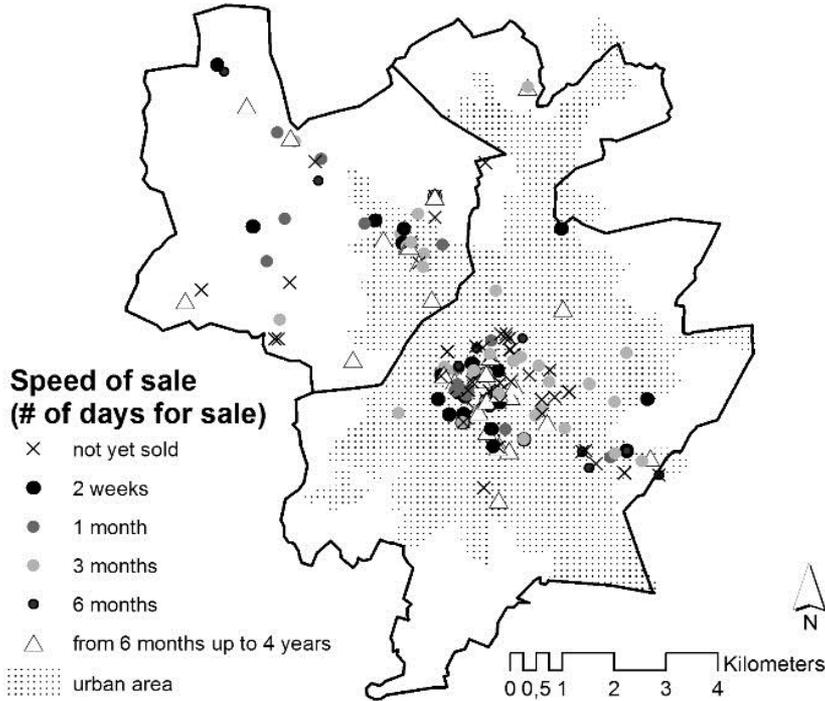
speed of sale



listing price



Leuven & Herent



PROOF



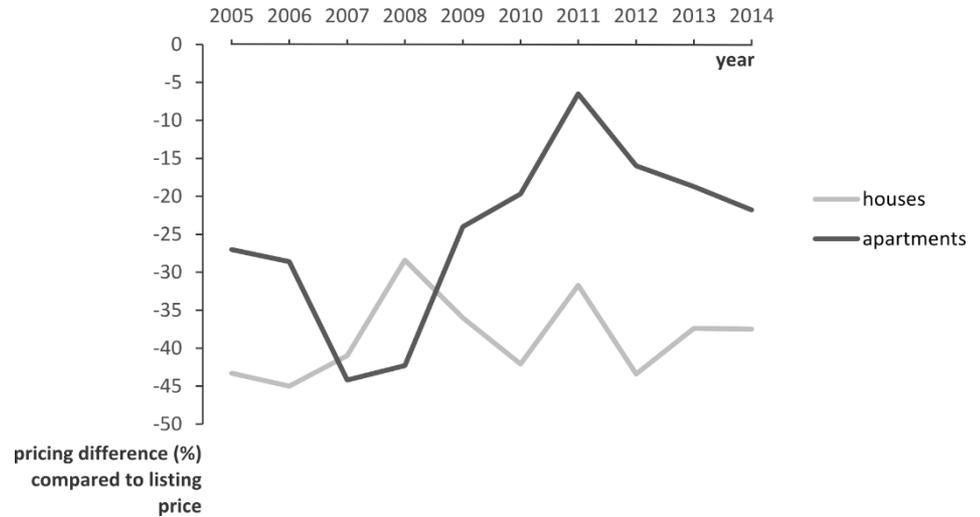
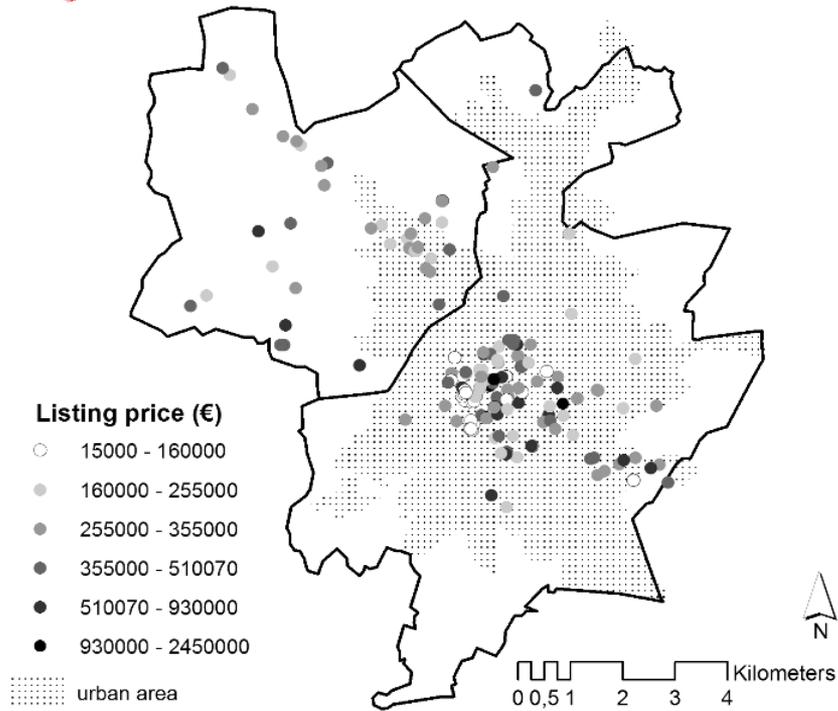
speed of sale



listing price

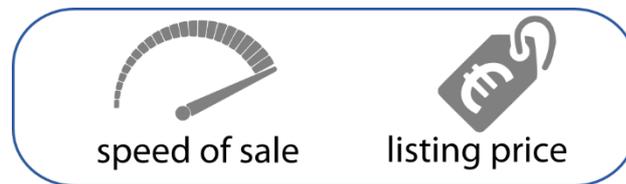


Leuven & Herent



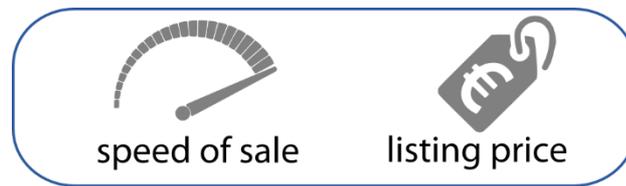
conclusion





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



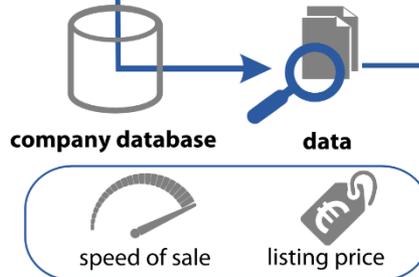
- shared challenge
- data exchange
- support
- data quality control
- data preparation for analysis

2. Exploration web data and testing prototype indicators

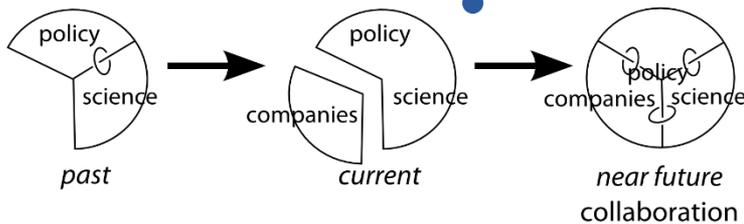
- technical & juridical barriers
- suboptimal data transfer
- collaboration is a better option

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
isabelle.loris@ugent.be