

# Data at a crossroads: use of linked administrative data for labour market analysis and policy making

Tom Bevers

**TAIEX Workshop on the use of administrative data for evidence-based policy making**

**Madrid, 11 December 2019**



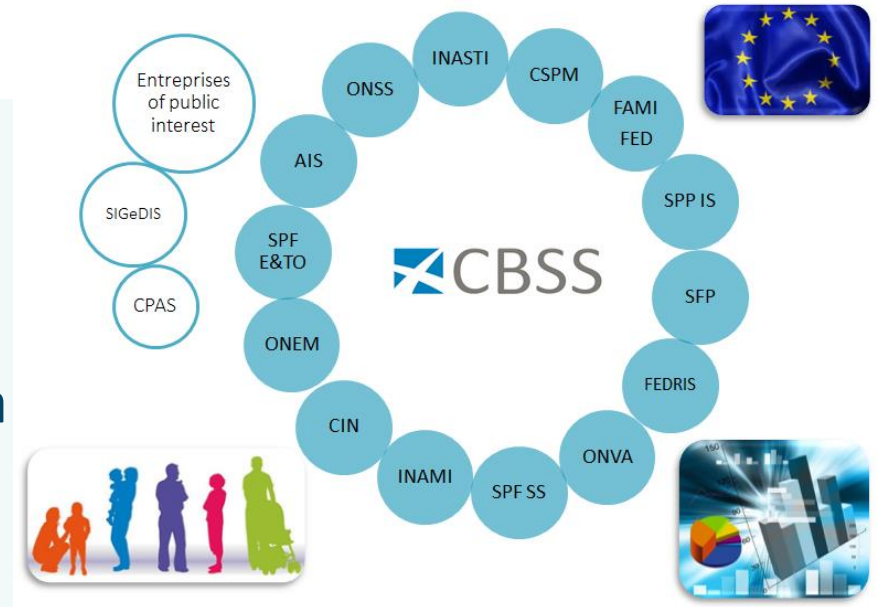
# | Contents

1. The Crossroads Bank for Social Security
2. The Datawarehouse Labour Market and Social Protection
3. Using DWH data for analysis and policymaking
4. Risks and challenges
5. Future work



# The Crossroads Bank for Social Security

- e-Government agency for the whole Belgian social security
- CBSS organises data exchange and integrated procedures for a large network of social security institutions that continue to manage their own data
- More information :  
<https://www.ksz-bcss.fgov.be/nl/information-english>



- > Set up in 1990 by the social affairs ministry
- > Autonomous since 1997
- > Managed by the institutions & social partners
- > 17 million EUR / year (and 1 billion reduction of administ. costs)

# DWH Labour Market & Social Protection

- Maintaining the DWH LMSP is a legal task of the Crossroads Bank, but is a relative small part of its activity
- Permanent dataset (exception to the CBSS rule)
- Created to efficiently process data requests from research institutions and the government
- Constructed with data from
  - the social security institutions
  - other government institutions
  - the national register (population)
  - and complemented by self-defined notions
- Linkable with other data sources
- Documentation
  - <https://www.bcsc.fgov.be/nl/dwh/homepage/index.html>
  - <https://www.bcsc.fgov.be/fr/dwh/homepage/index.html>



# | DWH functioning and governance

- CBSS
  - Daily monitoring and coordination
  - Handling custom data requests and developing web applications
- University of Leuven and Université libre de Bruxelles
  - Permanent scientific support
- SMALS (ICT agency of the federal government)
  - Hosting of the Datawarehouse
- Users board
  - Researchers
  - Public institutions
- Management board
  - Data suppliers
- Sponsors
  - Consortium of federal and regional authorities



# | Sources of the DWH

The datawarehouse includes a very wide range of variables, delivered by: the **National Register**, the 4 regional **PESs**, National Institute for Sickness and Disability Insurance (RIZIV), National Office for Family Benefits for Employees (FAMIFED), National Institute for Social Insurance for the Self-employed (RSVZ), National Office for Social Security (RSZ), National Office for Employment (NEO), FPS Social Security, the Occupational Accidents Fund, the Occupational Diseases Fund, the National Intermutualist College (NIC), the Pension Service for the public sector, the FPS Social Integration (POD MI), the National Agency for Annual Holidays (RJV), the National Agency for Pensions, SIGeDIS, the Flemish Agency for Higher Education & Adult Education, the Académie de recherche et d'enseignement supérieur (ARES), the Conseil des Recteurs (CREF),...

Most data is available from **2003** onwards and is **longitudinal** !



# | Contents of the DWH

- Data about the labour market situation
- Data about family allowances
- Data about pensions and pension build-up
- Data about incapacity for work and disabilities
- Data about social assistance (poverty)
- Data about education and certificates
- Personal characteristics
- Data about origin
- Self-defined notions (e.g. socio-economic position)



# | Access to the data

1. Ready-made tables: available on DVD/CD-rom
2. Web applications: interactive local and federal data per year, socio-economic mobility, households
3. Tailor-made data requests: at the request of research institutes and government institutions
  - Combining variables from several social security institutions
  - Sample or entire population
  - Deliberation by the Social Security and Health Chamber of the Information Security Committee
  - Authorisation by the data owner institution (if requested)
  - Possibility to link external data (for example census data, more detailed variables, other institutions)
  - Standard variables, but also variables designed by the requester





# | Data and meta-data in DWH

- Data follow an administrative logic: can be confusing (“allowance” instead of “participation”) and incomplete (no administration wants to know your occupation)
- Description of variables is available: main task of the scientific support unit
- However: no complete “users’ guide” and no “helpdesk”
  - Examples: how to construct working time, how to identify people in certain ALMPs (10 variables, which one to use?)
  - > work in progress
  - > users’ group, exchange of experiences,...
  - > almost impossible for a real outsider without “inside” assistance (e.g. OECD!)



# Conditions for the use of DWH data (1)

- In compliance with
  - privacy regulation (GDPR, privacy law 30-07-2018,...)
  - decision of the Information Security Committee
  - contract between data supplier and data recipient
- Proportionality
  - limitation of the amount of data
  - limitation of the amount of data subjects (a sample of the population)
  - communication of data in classes (age, nationality, income,...)

However: exact criteria not always crystal clear
- No data communication of data to third parties  
(unless identified in the demand )
- No re-identification
  - all possible measures to avoid re-identification of the data subject
  - no attempts to convert the pseudonymised into non-pseudonymised data
  - publication of the research results exclusively in the form of anonymous data



# Conditions for the use of DWH data (2)

- Possibility to work in the DWH on the most sensitive data  
However: not the best developed procedures
- Only for the duration of the research
  - Data must be destroyed afterwards
  - A trusted third party can keep the data for a longer period
    - for the justification of the research results
    - for follow-up
- Separation of responsibilities
  - If the data recipient is also (partially) a data supplier
  - Strict separation between
    - the service that processes the non-pseudonymized data for operational purposes (input)
    - the service that processes the pseudonymized data for research purposes (output)



# | Examples

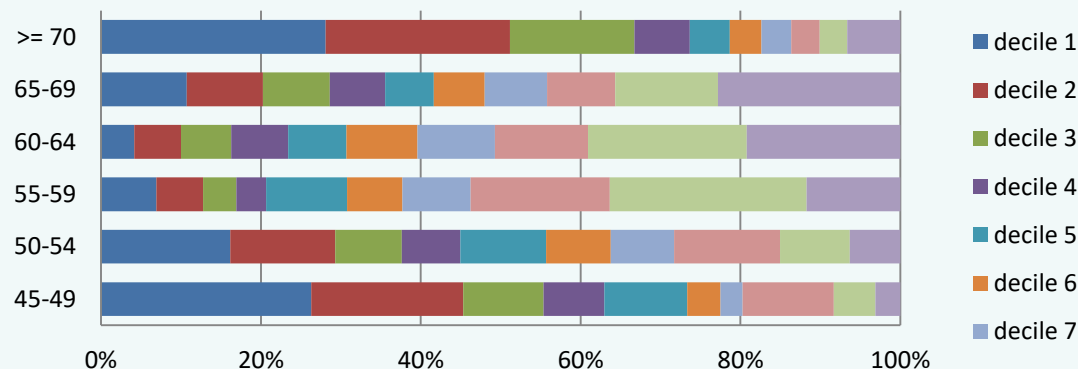
1. Policy evaluation older workers
2. Impact of company restructurings on workers' careers
3. Socio-economic Monitoring: Labour Market and Origin



# Policy evaluation older workers

- Traditional area of concern for Belgian policy makers: need for detailed follow-up
- Survey data (LFS, SES, EU-SILC, EWCS) gives us an indication of our progress, but is insufficient to monitor the results of specific measures
- Analysis includes demographic (region, gender, origin, age) and career related variables (job mobility, wages, recruitment, career breaks, etc.)

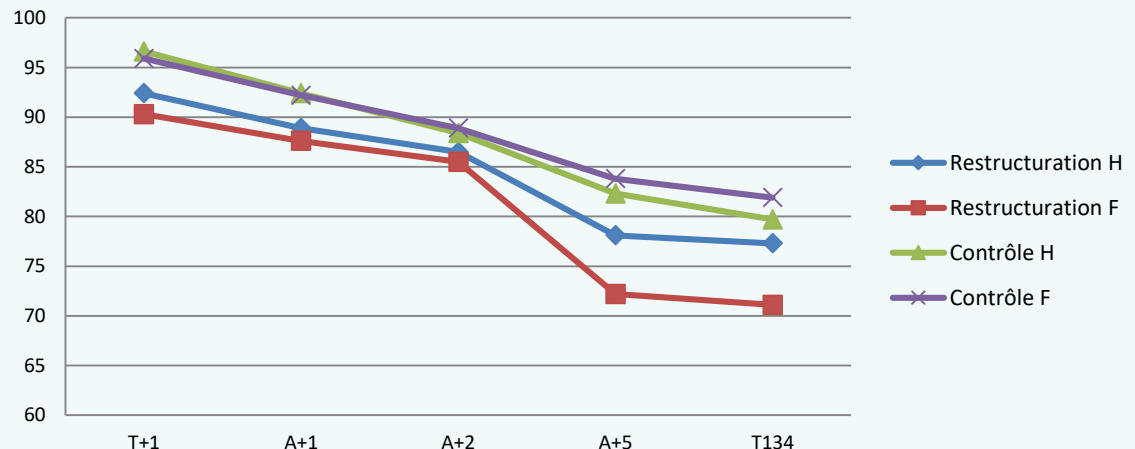
**distribution according to the wage decile in which persons are located in the trimester preceding retirement (2015)**



# Impact of company restructurings on workers careers

- Analysis of workers who have faced company restructuring or collective dismissal (at the request of a government working group)
- Individual worker trajectories were needed -> pseudonymized data
- Linking of data from the DWH LM&SP to the [Crossroads Bank for Enterprises](#) (size of the enterprise, number of employees, date of restructuring)

Percentage of persons employed in the 4<sup>th</sup> quarter of 2007 still employed in the following quarter, 1 year after, 2 years after, 5 years after and the 4<sup>th</sup> quarter of 2013



# Socio-economic Monitoring: Labour Market and Origin

- Two-yearly report on the position of people of different origins on the Belgian labour market, published by our ministry and Unia, the Belgian equality body
- Two new indicators were developed ('origin' and 'migration background'), combining separate variables from the DWH -> now available to other users on request
- Variables on education: linking of data from the 2011 CENSUS (Statbel) with the DWH LM&SP (data from the PES and different regional education institutions) -> involved a lot of manual research due to lack of uniform 'key'

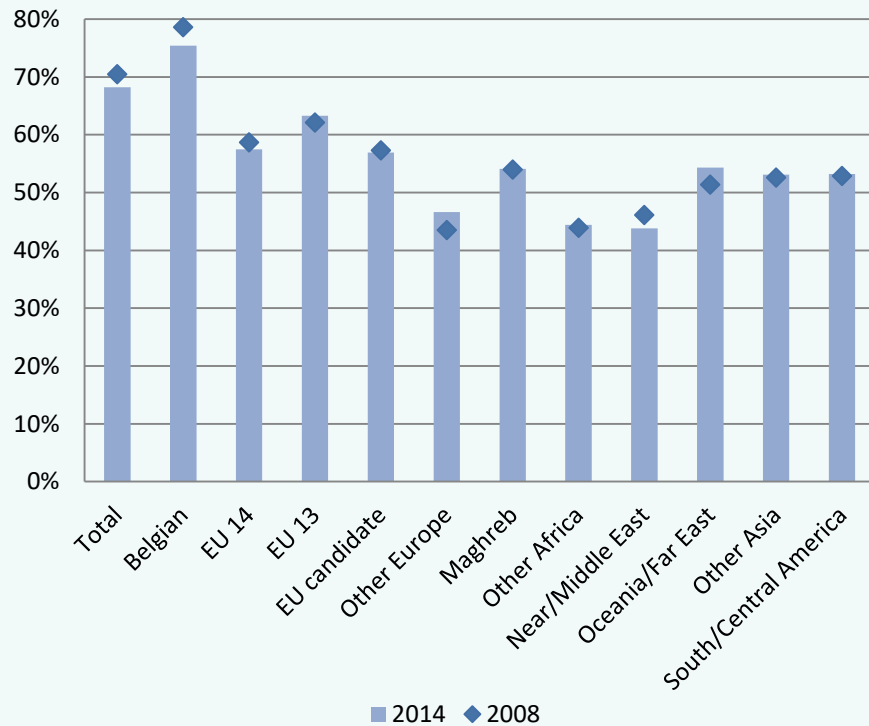


# Socio-economic Monitoring.

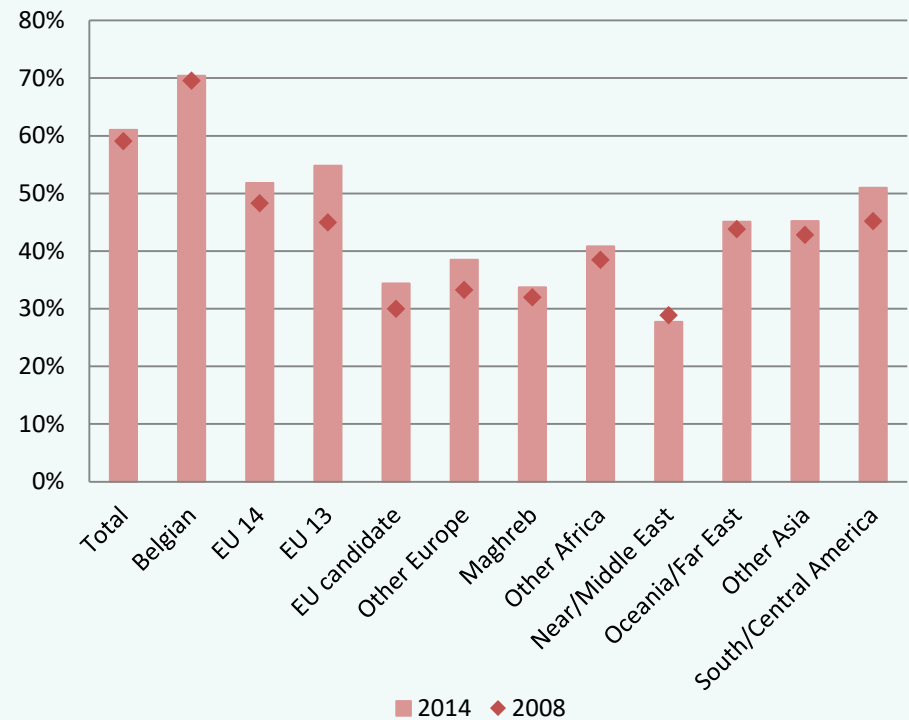
## examples

Employment rate by origin, 20-64 (2008/2014)

Men



Women



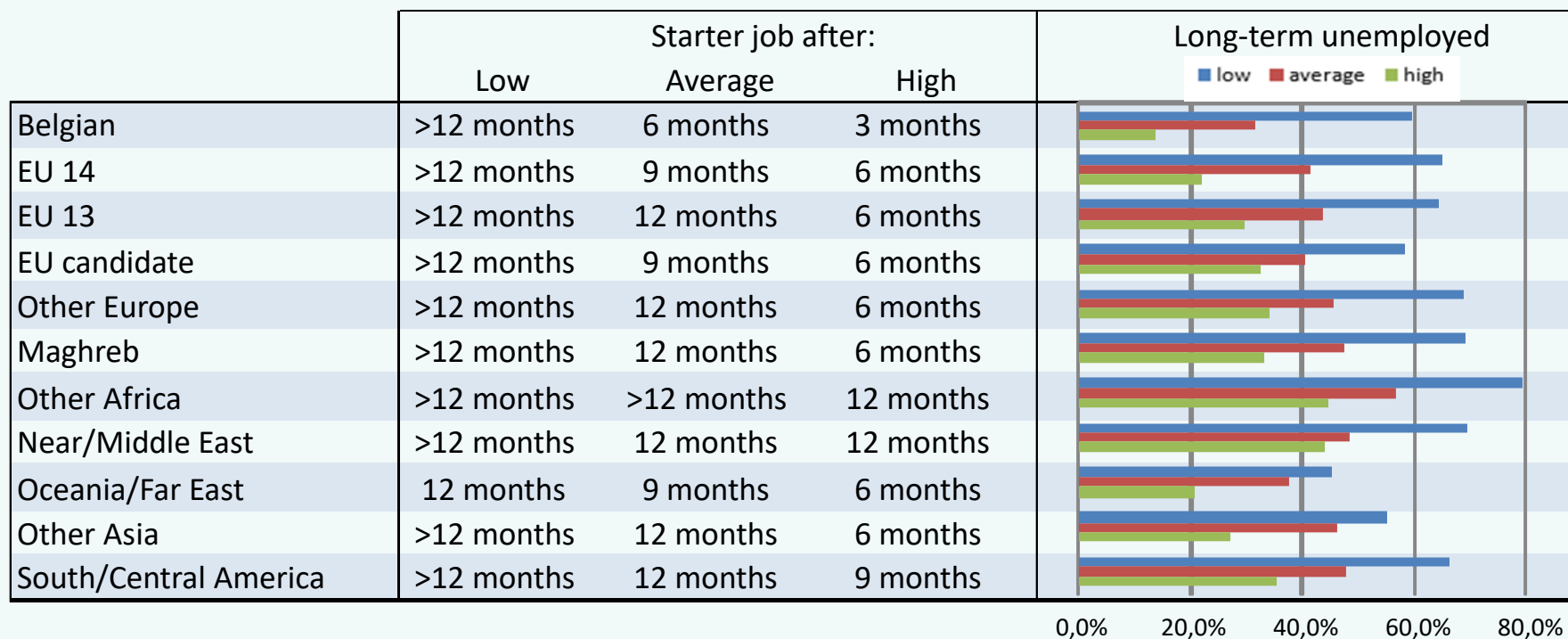


# Socio-economic Monitoring.

## examples

Young people (18-29) in insertion period, by origin and education level (2014)

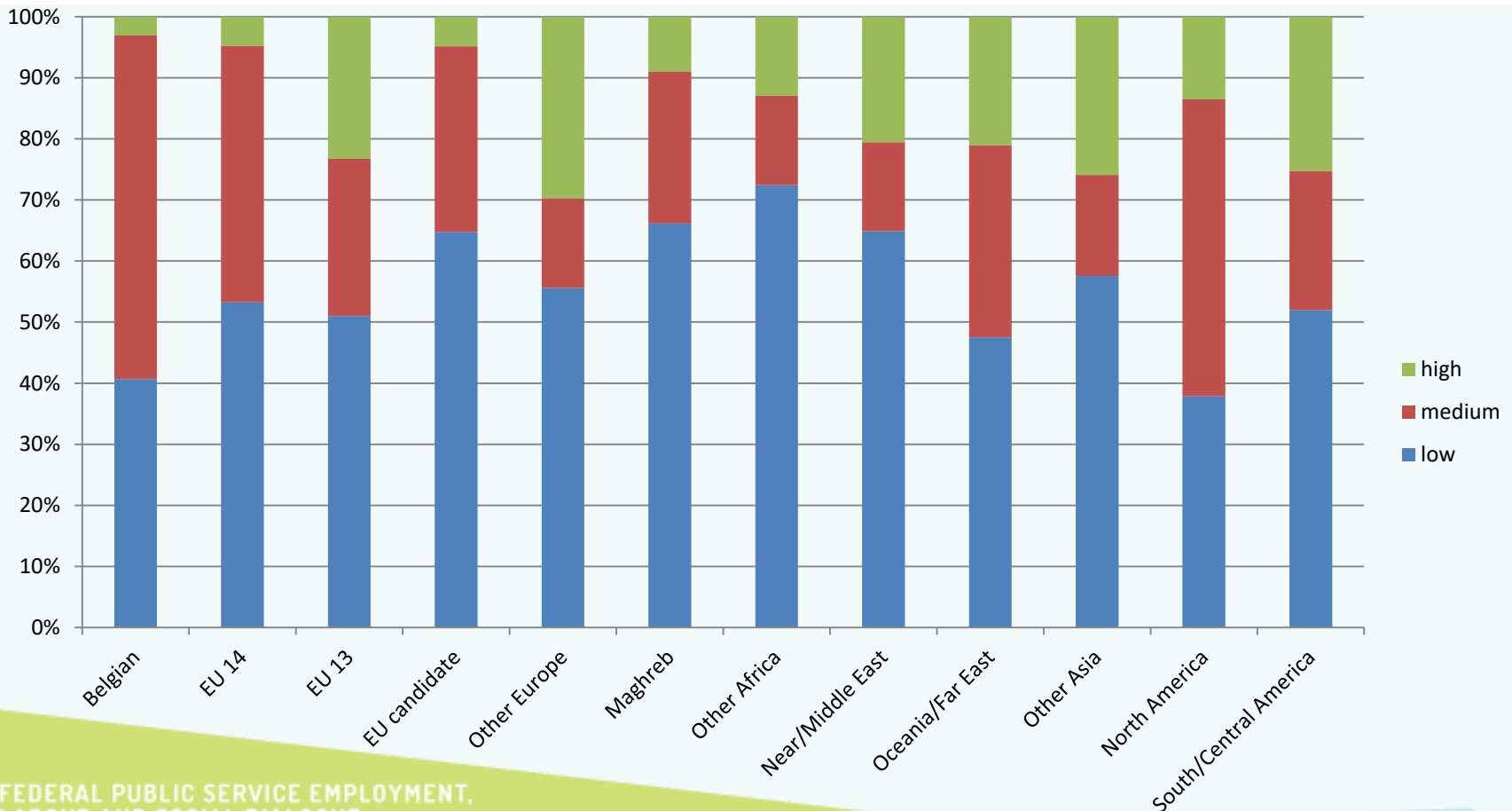
Median duration of the search for a first job  
and the percentage of long-term unemployment among young people aged 18 to 29



# Socio-economic Monitoring.

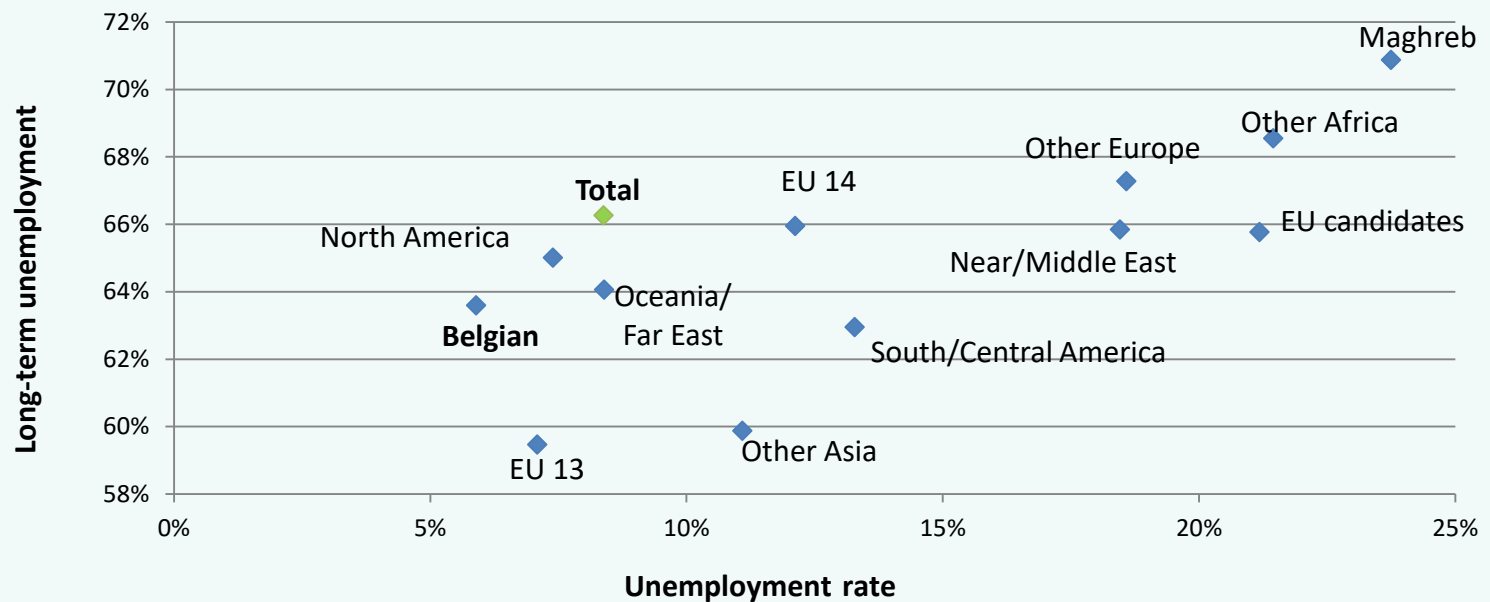
## examples

Distribution of female wage earners in the system of service vouchers by level of education, per origin (2014)



# Socio-economic Monitoring. examples

Overall unemployment and long-term unemployment by origin (2014)



# | Risks and challenges

- Legal/privacy related issues: manage the risk of (re)identifying the persons to whom the data relate
  - Solutions: anonymization (aggregating variables in large numbers), pseudonymization, clear legal framework regarding data transfer and storage (on-site analysis in certain sensitive cases)
- Organizational issues: large variety of themes within the DWH -> insufficient background info, quality of the (meta)data, time lag (most recent data: 2017), quickly growing interest
  - Solution: more personnel
- Coordination: different institutions working on the same data without deliberation (not enough scale benefits)
  - Solution: central Datawarehouse coordinated by central strategic committee
- Epistemological issues: Does the availability of data direct the research?



While definitely vast, the available data still has many limitations. This might lead to tunnel vision. Can data ensuing from existing administration be used to inform future policy needs?



# | Future work

- Increase quality of existing variables and metadata
- Increase accessibility
- Harmonize variables (for example from the different regions and communities): conceptual and technical work
- Development of new variables and compound variables (for example socio-economic background) to improve usability
- Assist (international) users to access Belgian data



