

# Ontology-Based Integration of Occupational Health Data: Method and Case Studies

Cassandra BARBEY<sup>1,2</sup>, Malika Smaïl-Tabbone<sup>3</sup>, Nathalie Bonvallot<sup>2</sup>, Frédéric CL

<sup>1</sup> Department of Pollutant Metrology, French Research and Safety Institute for the Prevention of Occupational Accidents and Health (INRS), 54547 Vandœuvre-lès-Nancy, France

<sup>2</sup> Univ Rennes, Inserm, EHESP, Irset- UMR\_S 1085, Rennes, France

<sup>3</sup> LORIA Université de Lorraine, CNRS, LORIA, UMR 7503, Vandœuvre-lès-Nancy, France



**dataPOST**  
Multiple exposure analysis and occupational health



# Presentation

- The work presented here was developed as part of my PhD thesis
- This thesis is the result of a joint agreement between INRS and EHESP.
  - INRS: French Research and Safety Institute for the Prevention of Occupational Accidents and Diseases (funder)
  - EHESP: French School of Public Health
- The work and the presentation will focus on public health data.

# Introduction

- Workers are exposed to a number of risks that can affect their health.
  - These risks are caused by *occupational exposures*
  - The study of their effects is known as *occupational health*

- In France, several data sources exist to describe workers' health.
  - Information on occupational diseases
  - Specific objective : to describe occupational diseases to describe
  - Various collection methods



Workers' health.  
occupational diseases  
to describe

Sumer	C2P	Colchic / Scola
National surveys	Regulatory declarations	Sampling and analysis of workplace air by specialised chemistry laboratories
Worker	Worker	Measurement

and sampling	information
Measurement	Occupational group

AT-MP	Evrest	MCP	RNV3
Medical consultation	Systematic occupational health interviews	Compulsory professional medical consultation	Medical consultation specialist o
Worker	Worker	Worker	Health pro

- It is difficult to use all data together.

# Introduction

- DataPOST : methodology for pooling information from several databases
  - Several studies or projects have tried to develop a methodology
    - L. Rollin et al., « Complementarity of 4 data bases in occupational health », Arch. Mal. Prof. Environ., vol. 82, 261-276, May 2021, doi: 10.1016/j.admp.2020.11.002.
    - <https://data.risquesautravail.be/fr>
  - Applying this methodology to all types of industry
  - Allowing for a detailed analysis of individual and combined exposures

**Objective : Bring together information from 10 French databases to improve knowledge about occupational exposures and their effects on health**

1. Integrate data with a structured methodology
2. Use integrated data in two cases studies

Step 1:

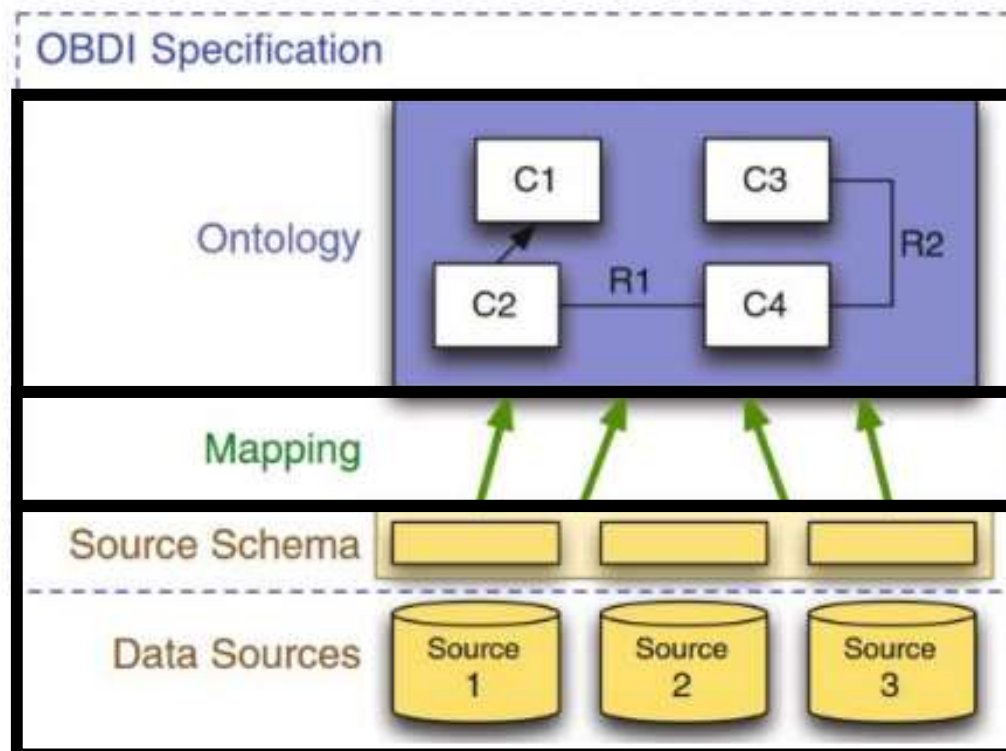
Integrate data with a structured methodology



# OBDI approach

- OBDI: Ontology-Based Data Integration

- Bring together heterogeneous data by linking them to an existing ontology
- D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati, « Ontology-Based Data Access and Integration », in Encyclopedia of Database Systems, 2017, pp. 1-7. doi: 10.1007/978-1-4899-7993-3\_80667



Ontology = formalised description of the concepts of a domain

Mapping = link between data and ontology

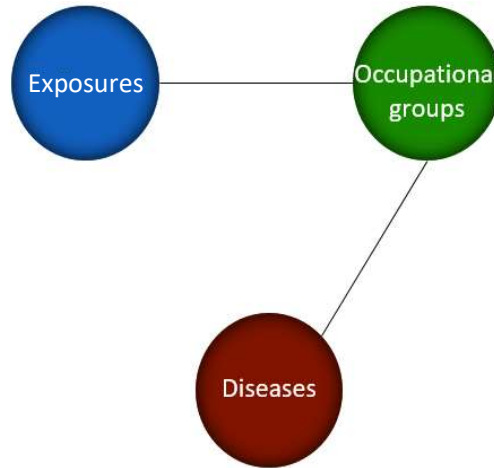
Data sources = heterogeneous data

# General representation of the concept

Exposures

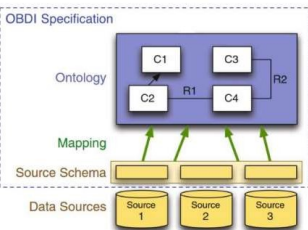
Skilled male workers in the  
"special construction"  
activity sector  
EXO

Occupational groups



Individuals of  
sex sharing  
occupation and  
the same sector

OBDI

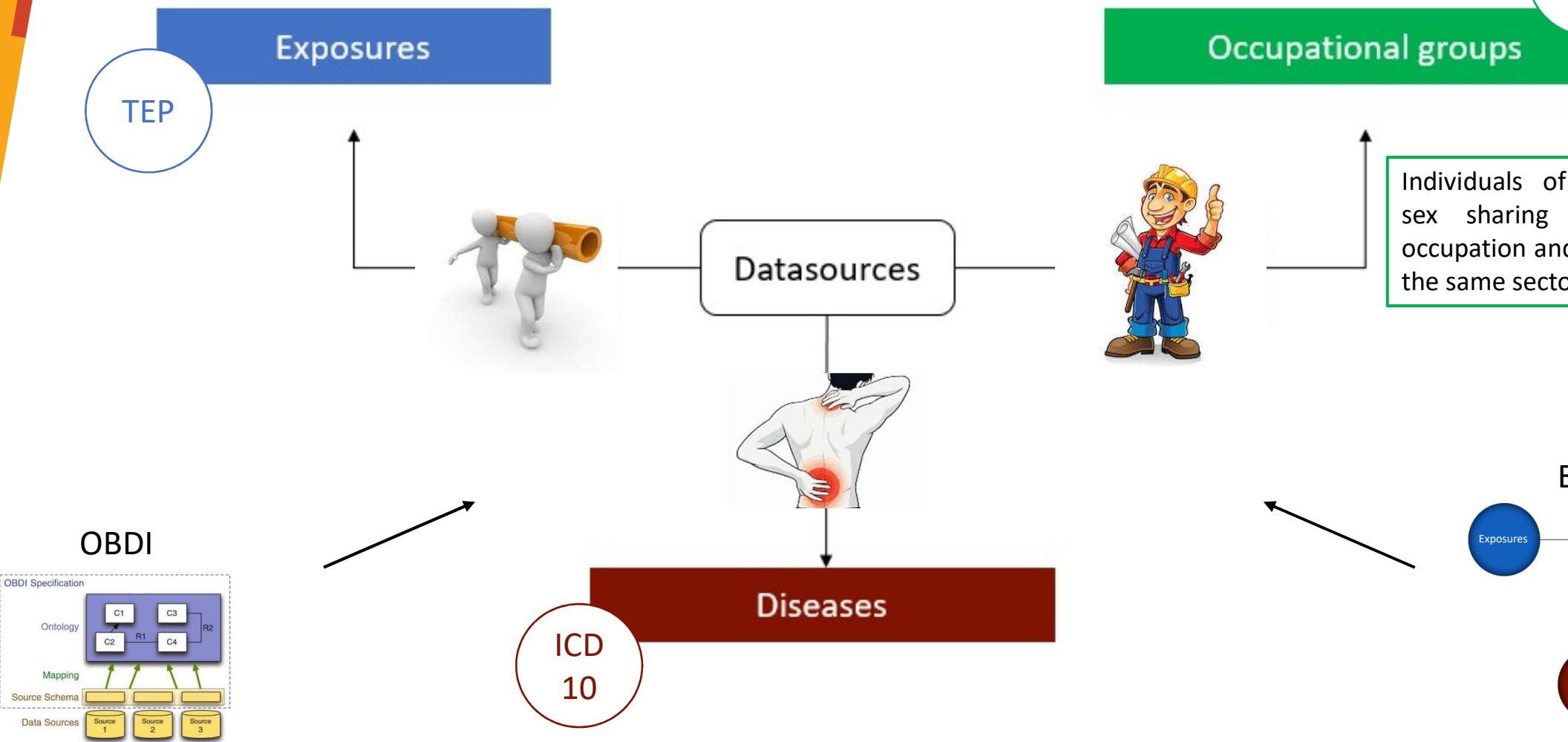


Diseases

C. J. Mattingly, T. E. McKone, M. A. Callahan, J. A. Blake, and E. A. C. Hubal, « Providing the Missing Link: the Exposure Science Ontology ExO », Environ. Sci. Technol., vol. 46, no. 6, pp. 3046-3053, Mar 2012, doi: 10.1021/es2033857.



# General representation of the health occupational data





# General representation of the health occupational data

- Integration of 10 databases using the OBDI method
  - Enable us to structure our data
  - 12,835 occupational groups
- First focus on the construction sector:
  - 816 occupational groups
  - 308 occupational exposures
  - 174 occupational diseases
- Second focus:
  - 10 most exposed occupational groups
  - 15 most serious exposures and diseases

Step 2:

Use integrated data into two cases studies

# Case study 1: Quantitative analysis

- Quantify several exposures and diseases for each occupational group
- Create an indicator for each exposure and each disease
  - Group together databases with similar characteristics

Example of the skilled male workers in the “special construction” activity sector		Database 1	Database 2	Database 3	Database 4	Database 5	Database 6	Database 7
<b>Exposure:</b>  <b>Manual handling of heavy loads</b>	<b>Value</b>	A	<i>No data</i>	B	<i>No data</i>	<i>No data</i>	<i>No data</i>	C
	<b>Similar characteristics</b>	Number of workers exposed		Number of declaration of exposure		Intensity of exposure		Number of cases caused by
	<b>Score (0-1)</b>	Score 1		Score 2		Score 3		Score 4
	<b>Indicator (0-4)</b>	<b>Sum of the four scores</b>						

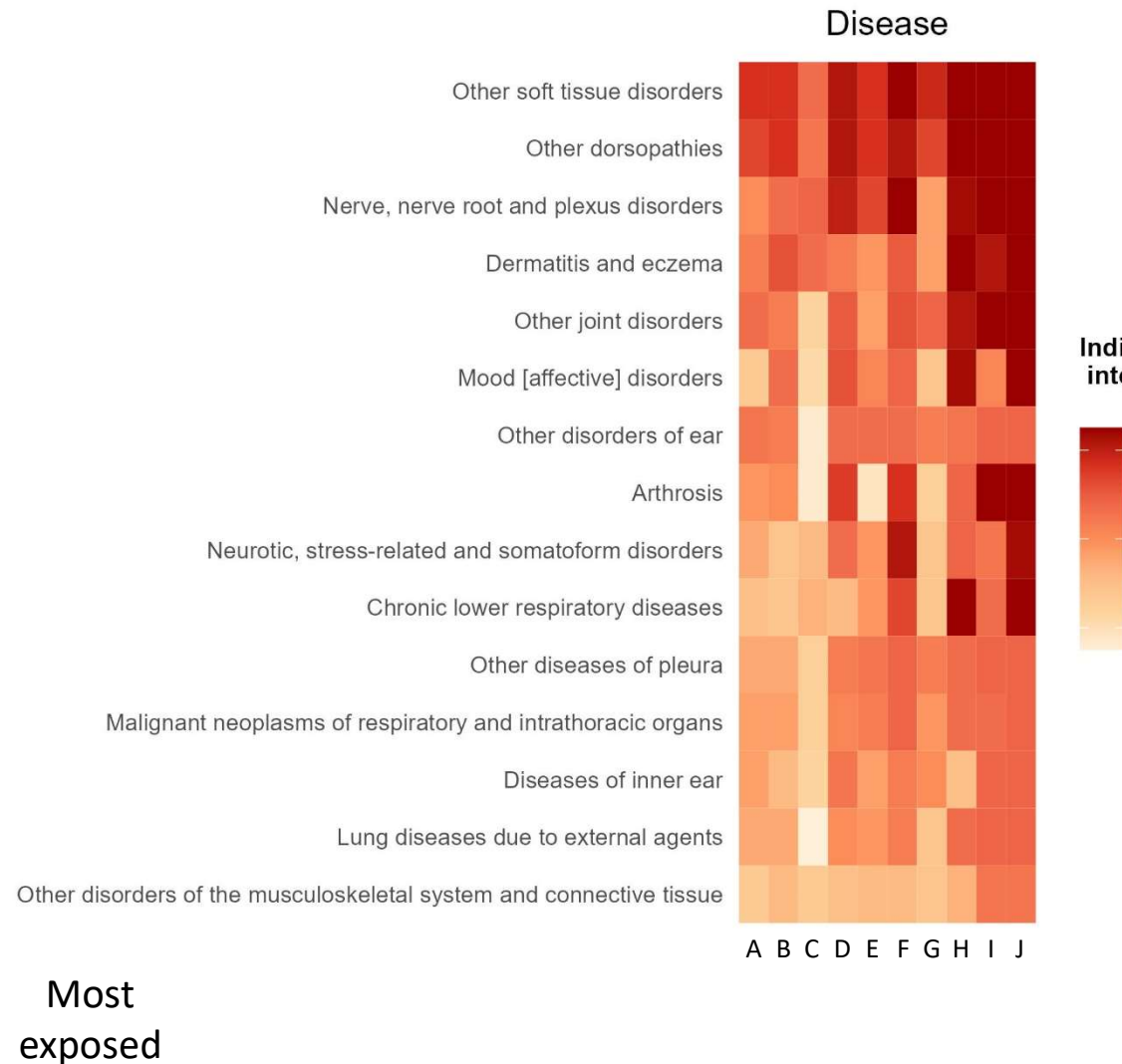
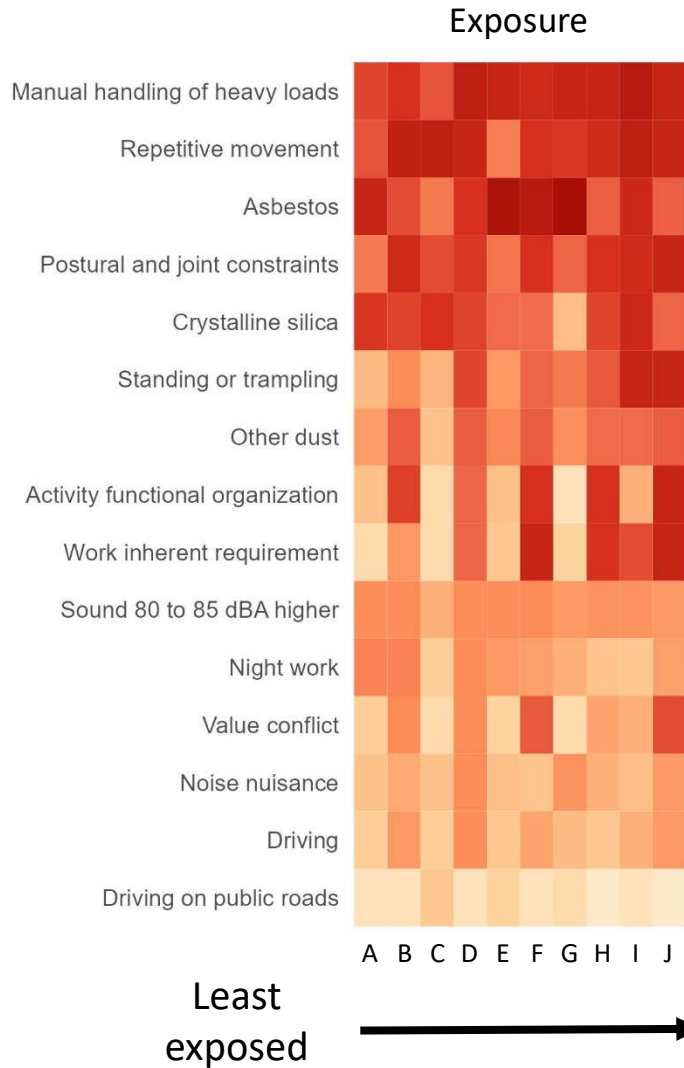
- Many indicators have been created.
- Exposure and disease indicators were represented using a heatmap.

# Case study 1: Quantitative analysis

groups  
d by the  
e/disease



groups  
d by the  
e/disease

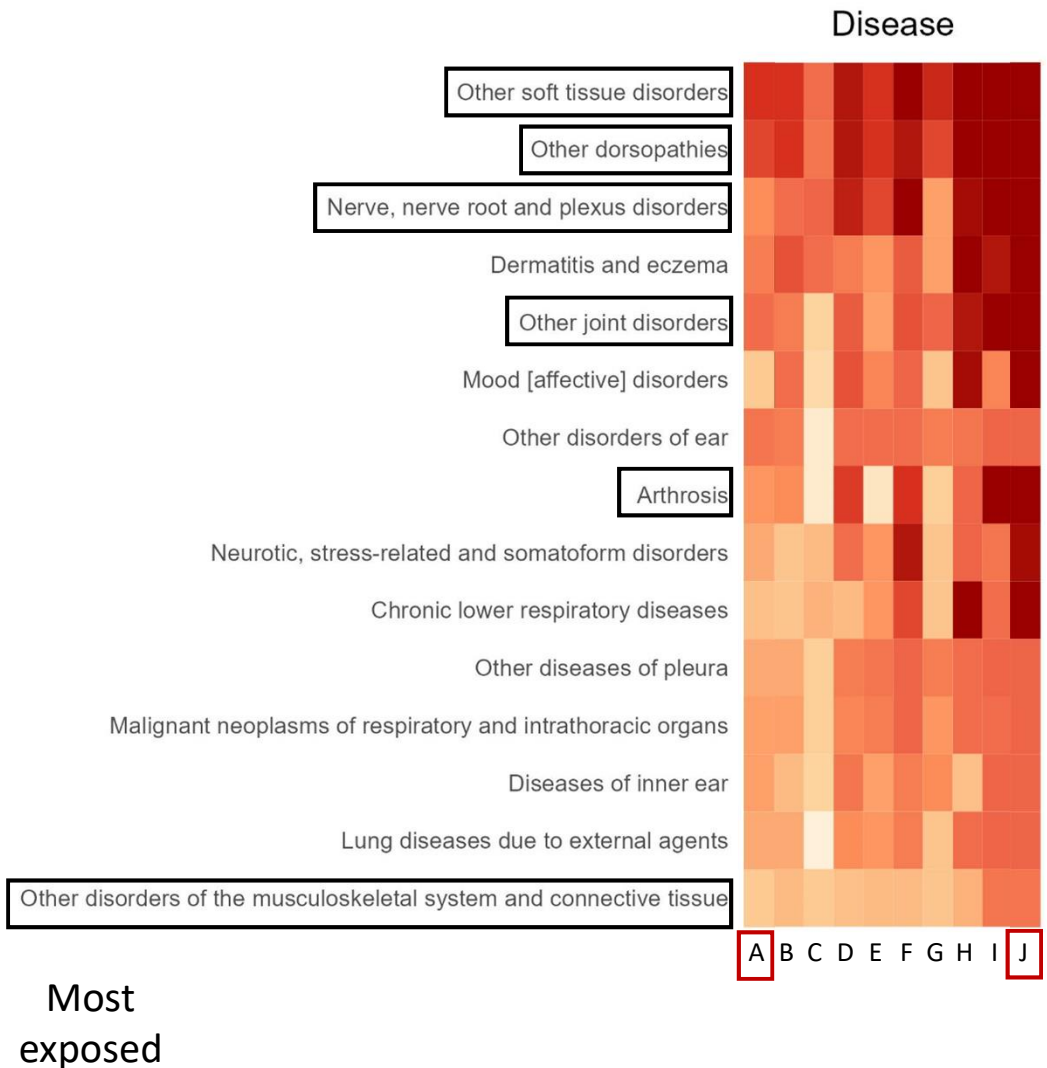
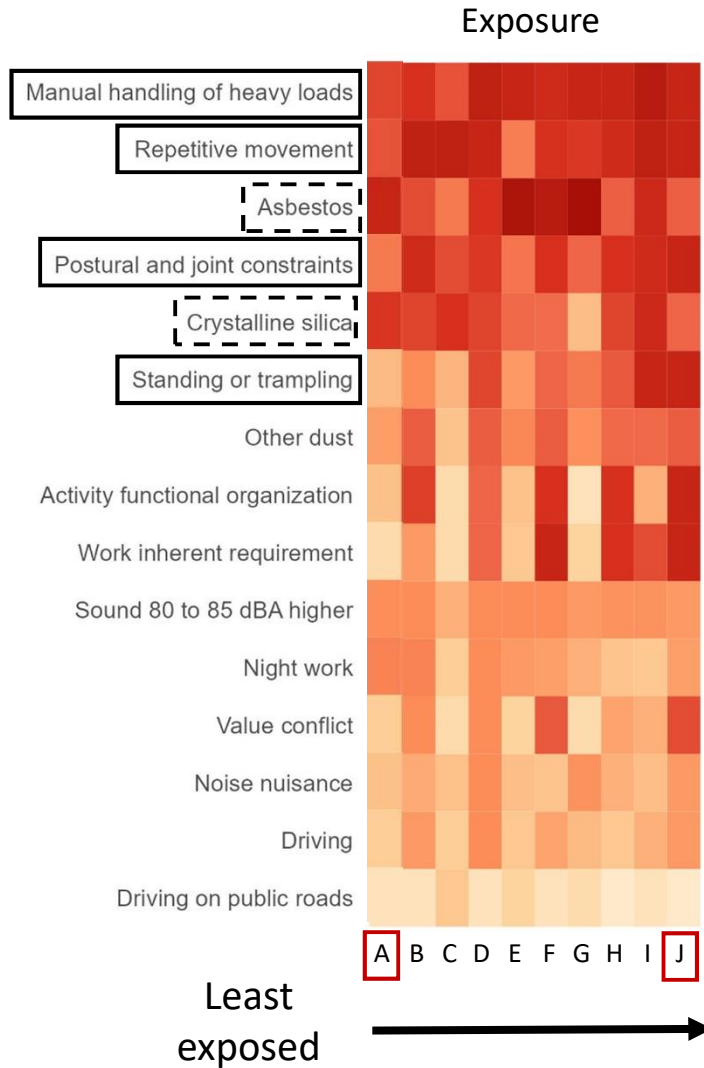


# Case study 1: Quantitative analysis

groups  
d by the  
e/disease



groups  
d by the  
e/disease



Indi  
inte



ed industrial building male workers in the "civil engineering" sector  
male workers in the "special construction" activity sector



## Improved preventive measures



Powered exoskeletons could help reduce the risks associated with manual handling of heavy loads, repetitive movements, and also awkward postures.

## Case study 2: Qualitative analysis

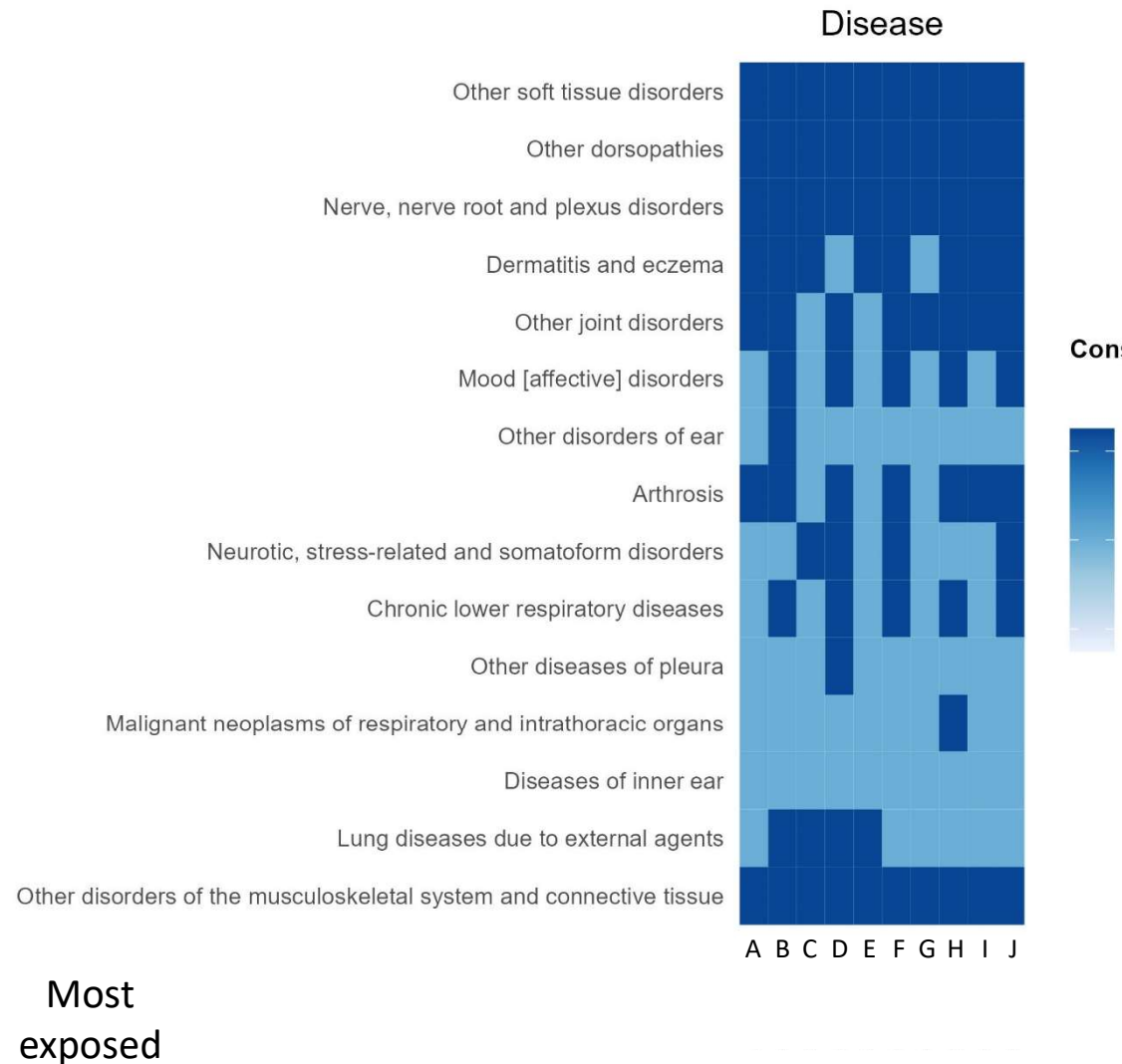
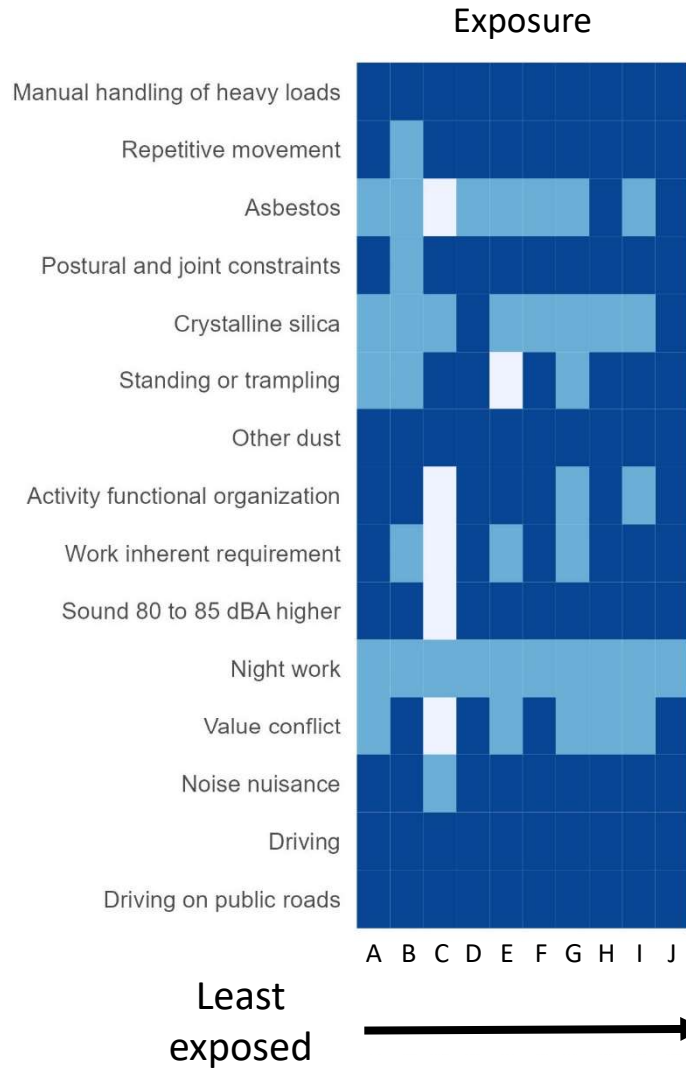
- Assess the consistency of data across the databases for each exposure and disease
  - Number of databases confirming an exposure or a disease
- The greater is the number of databases with information reporting an exposure or a disease, the higher is the consistency.
- Exposure and disease consistency were represented using a heatmap.

# Case study 2: Qualitative analysis

groups  
d by the  
e/disease



groups  
d by the  
e/disease



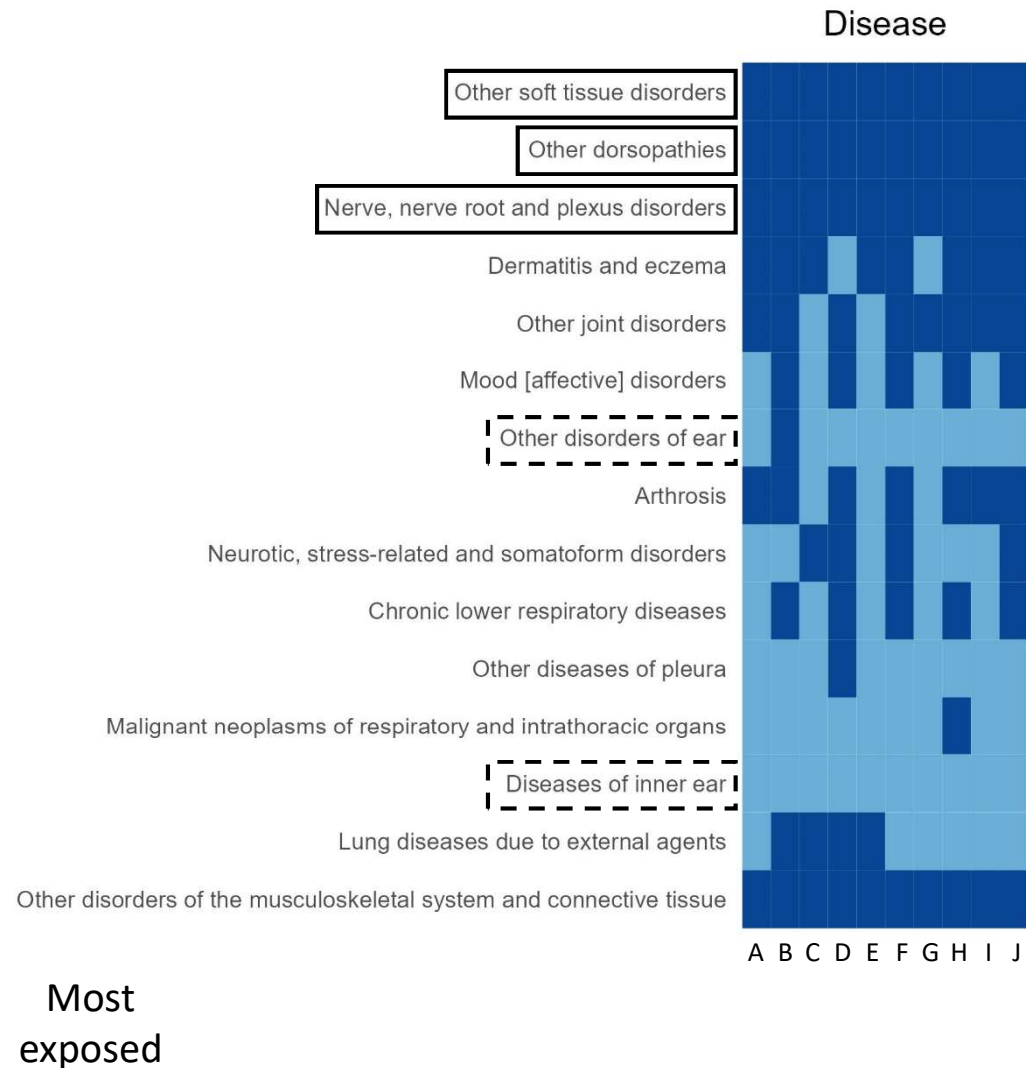
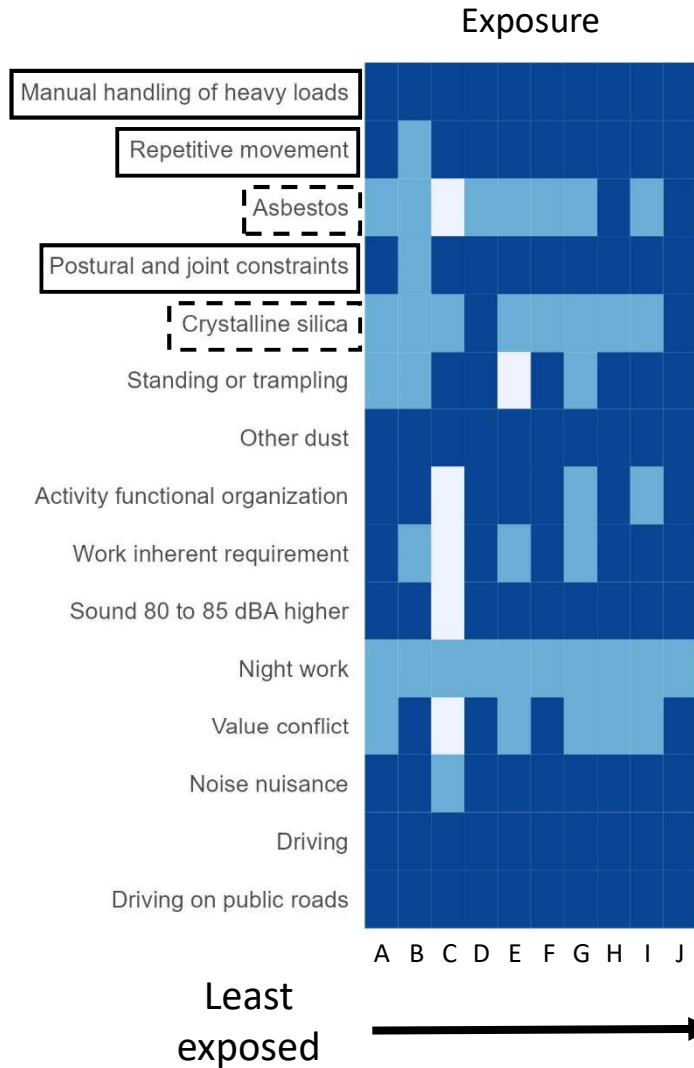


# Case study 2: Qualitative analysis

groups  
d by the  
e/disease



groups  
d by the  
e/disease



## Case study 2: Qualitative analysis



The presence of moderate consistency highlights

- The differences between the databases
- The complementary of the information provided

## Conclusion and future work

- Structuring data and reconciling occupational health ontologies has enabled to improve knowledge about occupational exposures and their effects on health.
- Several perspectives are considered with integrated data:
  - Have a better adaptation of our data to the ExO ontology
  - Search for correlations between exposure and disease indicators
  - Create a visualisation tool
- A future project would be to generalise the methodology to integrate other databases.

# Thanks for your attention

## Any question ?

