Integrating and Analysing Occupational Health Data Using a Multi-Ontology Approach

Cassandra BARBEY
Department of Pollutant Metrology, INRS
Univ Rennes, Inserm, EHESP, Irset- UMR_S 1085
Vandœuvre-lès-Nancy, France
Email: cassandra.barbey@inrs.fr

Nathalie BONVALLOT
Univ Rennes, Inserm, EHESP, Irset - UMR_S 1085
Rennes, France
Email: nathalie.bonvallot@ehesp.fr

Malika SMAÏL-TABBONE
Université de Lorraine, CNRS, LORIA, UMR 7503,
Vandœuvre-lès-Nancy, France
Email: malika.smail@loria.fr

Frédéric CLERC
Department of Pollutant Metrology, INRS
Vandœuvre-lès-Nancy, France
Email: frederic.clerc@inrs.fr

***Abstract* - A variety of occupational data are collected by health organisations to investigate workplace exposures encountered by workers in their occupational activities and the potential health effects that may arise. These datasets have diverse characteristics and are not inherently designed to interoperate. However, they contain complementary information, which, when analysed collectively, can provide a broader perspective on high-risk occupational scenarios and inform targeted prevention strategies. The objective of this study is to develop a methodology to integrate and analyse heterogeneous French data. For this, ten French occupational databases, provided by six French institutes were used. An Ontology-Based Data Integration approach was employed, involving the mapping of data sources to a domain-specific ontology, namely the Adapted Occupational Exposure Ontology. Four additional ontologies were utilised: the Occupational Exposure Thesaurus, which categorises occupational exposures and hazards; the International Classification of Diseases, which classifies health disorders and diseases; the French Nomenclature of Activities, which identifies activity sectors in France; and the Professions and Socio-professional Categories, which defines occupational classifications. Data integration is primarily achieved through the concept of the "occupational group", defined as a group of individuals sharing the same sex, occupation, and activity sector. Two case studies derived from the integrated dataset are presented: (1) a quantitative analysis identifying occupational groups at highest risk and most affected by diseases; and (2) a qualitative analysis evaluating the consistency of exposure and disease-related information. The construction sector was selected for these case studies due to its significance in occupational health research and the availability of substantial, relevant data. This methodological approach structures all the data and enables various analysis methods to be designed and implemented, making it possible to envisage targeted responses to current and emerging occupational health problems using specialised tools and queries.**

Keywords - ontologies; data integration; heterogeneous data; occupational health; occupational exposures; data analytics.

I. INTRODUCTION

This article constitutes an extended version of the international conference paper entitled "Ontology-Based Integration of Occupational Health Data: Method and Case Studies" [1], which was presented at the Thirteenth International Conference on Data Analytics in October 2024. The indicator calculation method has been updated, and the use of case studies has been expanded to the occupation and sector of activity.

Workers are routinely exposed to various occupational hazards that may affect their safety and health, potentially leading to occupational accidents or diseases. These risks result from occupational exposures, which can be present in any work situation linked to the specific tasks performed in a given occupation. Occupational exposures are categorised into five main groups: chemical, physical, biological, organisational, and psychosocial. Chemical exposures encompass all chemical products or substances encountered in occupational settings, regardless of their form, such as paints, cleaning products, and dusts. Physical exposures refer to all exposures that may have a physical effect on the worker (*i.e.*, temperature, vibrations, or noise). Biological exposures involve contact with living organisms encountered in occupations, including microorganisms, animals, or plants. Organisational exposures relate to structural aspects of the workplace (*i.e.*, hours, workload, or lack of resources). Psychosocial exposures refer to workplace dynamics that impact psychological and social well-being (*i.e.*, stress, hierarchical relations, or bullying). The interactions among these various exposures can have complex effects on workers' health, potentially reducing the efficacy of risk mitigation measures often designed for individual exposures. The implementation of relevant preventive actions requires an understanding of these interactions and their effects for effective occupational risk management. However, this remains an understudied area despite ongoing research efforts [2].

In France, several health institutions collect data on occupational exposure and workers' health. Each database is designed for a specific objective, such as characterising work environments or monitoring occupational disease development. Consequently, each dataset possesses its own characteristics, including its collection method and target population. This diversity of characteristics enables the collection of a substantial amount of information; however, it also hinders the ability to share this information, as the databases were not designed to interact with one another. While some databases provide a representative overview of the French workforce, others have a more limited scope. For instance, the Sumer database, derived from cross-sectional surveys on workplace exposures and worker perceptions, provides a nationally representative dataset. By contrast, the Scola database, which compiles exposure measurements conducted under regulatory chemical substance monitoring, has a different scope and unit of measurement, focusing on companies subject to regulatory controls. Despite their differences, both databases contribute valuable insights into chemical exposure in occupational settings.

The French DataPOST project (analysis of multiple exposure and workers' health) aims to provide a general joint use of French occupational health data. It aims to develop a methodology for extracting knowledge on occupational exposure and health effects using an ontological approach, by integrating data from ten occupational health databases. To achieve this objective, we propose to rely on the Ontology-Based Data Integration (OBDI) approach [6] to qualify and quantify occupational exposures and associated health effects. The statistical unit for analysis is the "occupational group", which constitutes a group of individuals of the same sex sharing the same occupation and working in the same activity sector. Such occupational groups are defined across all databases, facilitating data integration and exploitation. The integrated dataset is then utilised for multiple analyses, two of which are presented (Sections V and VI):

- A **quantitative analysis**, which quantifies various exposures and diseases for each occupational group by generating relevant indicators;
- A **qualitative analysis**, which assesses data consistency across databases for each type of exposure and disease.

The remainder of the article is structured as follows: Section II presents the current status of research on the joint use of several data sources. Section III provides an overview of OBDI and its main components (ontologies, data schemas, and mappings) used in this study. Section IV details the general data representation framework. Sections V and VI present the case studies, describing the methodologies and results. Finally, Section VII concludes with a discussion of findings and potential future research directions.

Specific Author Contributions:

- Cassandra BARBEY: Writing – original draft, Methodology, Investigation, Formal analysis;
- Malika SMAIL-TABBONE: Writing – review & editing, Methodology, Investigation, Formal analysis, Validation;
- Nathalie BONVALLOT: Writing – review & editing, Validation, Supervision;
- Frédéric CLERC: Writing – review & editing, Methodology, Validation, Supervision.

II. RELATED WORK

While the datasets under consideration are diverse in their structure and scope, their combined analysis holds the potential to provide a more comprehensive view of work-related risks and their impacts. To enable the joint use of these varied data sources, multiple analytical approaches have been developed, although their specificity limits their reproducibility. Some methodologies have been applied in case studies, such as those conducted by L. Rollin *et al.* [3], [4]. The first study compares data from multiple occupational health surveillance and monitoring systems using homecare workers as an example [3]. The second study examines musculoskeletal disorders in male professional drivers and their associated risk factors, leveraging multiple occupational health databases [4]. The use of the databases was based on the objective, the population, and the target disease. The joint analysis therefore represents a specific subset of the active working population. Other methodologies are being developed within broader projects, within the European context, such as the Datamining project, which aims to promote the monitoring of populations of workers at risk, by centralising data, in order to improve occupational sensitisation and prevention measures (https://data.risquesautravail.be/fr). This initiative integrates administrative records and survey data collected at multiple levels (European, federal, regional) to ensure that Belgian populations are represented. A related approach was employed by Dandan *et al.* [5]; however, their work was not conducted within occupational health, instead studying the monitoring of elderly health, where ontologies were used to integrate data from sensors, surveys, and medical records.

III. ONTOLOGY-BASED DATA INTEGRATION (OBDI) APPROACH

The OBDI system aims to achieve semantic integration of heterogeneous data sets by leveraging an existing descriptive ontology. This system establishes relationships between theoretical concepts in a specific field and real-world data using formal definitions crafted by experts in the domain. The OBDI framework consists of three key components: the domain ontology, which encapsulates the descriptions, definitions, and concepts relevant to the domain; the data sources, which represent diverse storage systems containing real-world data; and the mapping, which provides the structure for reconciling raw data with theoretical concepts in the ontology through formal definitions (Figure 1). The OBDI system enhances the understanding and utilisation of data, offering reproducibility and adaptability across various domains through a wide range of semantic integration approaches [7], [8]. An illustrative example is provided in R. Dandan's thesis [9], where the aim was to design health profiles for the elderly to improve the associations of health recommendations using a knowledge-based integration approach (DIKG2). In this context, an ontology of activities for older individuals was developed, incorporating preventive strategies and recommendations for nutrition and physical activity. Data were gathered from patients' medical records, connected measurement devices, and online questionnaires. The integration of these data was facilitated through vocabulary correspondence (*i.e.*, What pathology do you have? → Diabetes = Disease class: Diabetes) as well as a link between two entities (*i.e.*, John has diabetes → Diabetes recommendation).

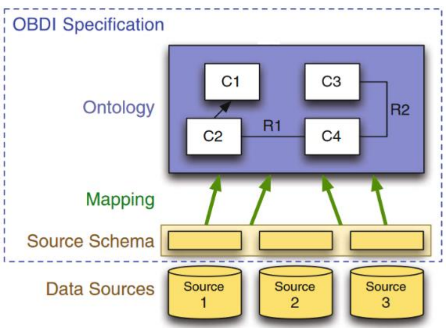

Figure 1. Ontology-based data integration adapted from Calvanese *et al.* (2017) [6].

The OBDI approach serves as the methodological foundation for the present study. However, this study's context is more complex, necessitating the use of multiple ontologies to integrate the available data and their interrelationships. For this purpose, we decided to start with the Exposure Ontology (ExO) proposed by Mattingly *et al.* [10].

A. *Occupational health ontologies*

The ExO ontology comprises four core exposure concepts: receptor, stressor, event, and outcome. Each of these concepts is further detailed through child terms and attributes. The "receptor" refers to an individual worker or a population of workers who may be exposed to a stressor. The "stressor" represents exposure to an agent, activity, or event that has the potential to impact the receptor. Stressors are categorised into chemical products, microbiological agents, physical elements, postural constraints, organisational or psychosocial factors, or combinations of these. The interaction between the stressor and the receptor is termed an "exposure event", which may result in a health "outcome", such as disease (see Figure 2, left-hand part).

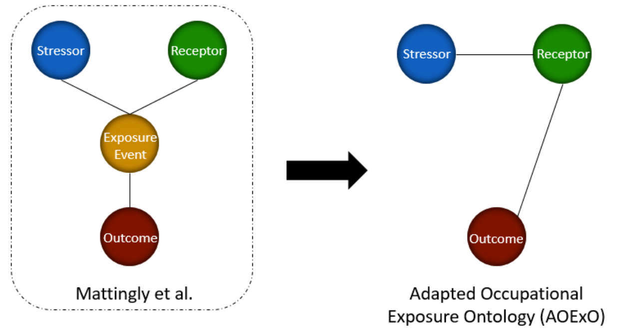

Figure 2. Main concepts of the exposure science ontology from Mattingly *et al.* (2012) [10] and their adaptation to our context.

In this study, the receptor corresponds to the occupational group, and the stressor represents occupational exposure. Unlike ExO, the "exposure event" concept is not explicitly defined and, therefore, is not used in our context. However, the available data allow the direct association of occupational groups with health outcomes. This leads to the Adapted Occupational Exposure Ontology (AOExO) (see Figure 2, right-hand part).

Four ontologies from the field of occupational health have been linked to AOExO concepts to create a more comprehensive occupational health database:

- The Occupational Exposure Thesaurus (TEP) [11] is a reference system designed in 2014 by the French agency for health safety to uniformly collect data on occupational exposures. The TEP is structured across eight hierarchical levels, encompassing approximately 8,300 exposure concepts, including "industrial product or process", "quality of the work space", "equipment tools machines and work machinery", "chemical agents", "biological agents", "rocks and other mineral substances", "physical agents", "biomechanical factors", and "organisational interpersonal and ethical factors". The system classifies exposures from general to specific concepts (*i.e.*, exposure to physical agents in general → exposure to biomechanical factor → exposure to work with strength → exposure to handling loads and handling people → exposure to carrying a load (horizontal movement) → exposure to carrying loads above the level of the shoulder).
- The International Classification of Diseases, tenth revision (ICD-10) (https://www.who.int), designed and maintained by the World Health Organisation, is an international compilation on the causes and consequences of human disease. ICD-10 provides a common health language by using approximately 150,000 codified clinical terms [12]. The classification is divided into 22 chapters, which are further subdivided into blocks of three-character categories and, in some cases, four-character subcategories. These diseases are categorised from broad concepts to more specific ones such as "Diseases of the musculoskeletal system and connective tissue" (Chapter XIII),

"Dorsopathies" (M40-M54), "Other dorsopathies" (M50-M54), "Dorsalgia" (M54), and "Sciatica" (M54.3).

- The statistical classification of activity sectors in the European Community is used for the organisation of information pertaining to economic and social activities. In the context of this study, the French version of this classification, referred to as NAF, is employed to define occupational groups (https://www.insee.fr). NAF is organised into five nested levels, such as "Construction" (Section F), "Specialised construction activities" (Division 43), "Demolition and site preparation" (Group 43.1), and "Demolition" (Class 43.11 and Sub-class 43.11Z).
- The Professions and Socio-professional Categories (PCS) (https://www.nomenclature-pcs.fr), a statistical classification developed by the French National Institute of Statistics and Economic Studies, groups occupations by social background. This ontology is used to define occupational groups. PCS is organised into four hierarchical levels of job designations, such as "Manual workers" (Aggregate category 6), "Skilled manual workers" (Category 61), "Skilled artisanal manual workers" (Detailed category 63), and "Skilled bricklayers" (Occupation 632a). Additionally, an intermediate level not officially included in the nomenclature is available, such as "Skilled craftsmen in building" (632). The latter is used to create occupational groups as it provides an optimal balance between a sufficiently small number of occupations and an appropriate level of detail.

The development of the AOExO ontology has been guided by a set of competency questions arising directly from the problems encountered in occupational health. These questions guided the choice of relevant vocabularies and ontologies to be integrated. In particular, they cover occupational exposure, work activities (occupation and sector) and health effects. These questions made it possible to precisely define the scope of the ontology, by identifying the essential concepts to be represented as well as the levels of granularity required to respond to the issues. Based on these questions, the relationships between the concepts were constructed in such a way as to reflect the complexity of real work situations, taking into account the links between occupational groups (modelled on the basis of NAF and ISCO nomenclatures), exposures (defined on the basis of TEP) and health effects (classified by ICD-10). They have also made it possible to test and validate the ontology's ability to respond to a wide range of competency questions, including the following main ones:

- "What are the differences in occupational diseases between men and women?" (Using ICD-10 ontology)
- "What are the exposures encountered by bricklayers in the workplace?" (Using TEP and ISCO ontologies)
- "Are painters in the specialised construction sector exposed to benzene?" (Using ICSO, NAF and TEP ontologies).

B. Data sources and their schema

A total of ten data sources were used in this study. Six of them are focused on occupational exposures: SUMER [3], C2P [13], COLCHIC and SCOLA [14], COLPHY [15] and MatGene [16]; three are focused on occupational diseases: AT-MP [17], MCP [3] and RNV3P [3]; and one provides both data: Evrest [3]. A comprehensive description of all data sources, collection methods, statistical units, and institutions is provided in Table I.

TABLE I. TABLE DESCRIBING THE DIFFERENT DATA SOURCES AND THE INFORMATION THEY CONTAIN

Data source	Institution	Collection method	Original statistical unit
Sumer	Dares	National surveys	Worker
C2P	CNAM	Regulatory declarations	Worker
Colchic / Scola	INRS	Sampling and analysis of workplace air by specialised chemistry laboratories	Measurement
Colphy	INRS	Historical measurements and sampling	Measurement
MatGene	SPF	Historical and census information	Occupational group
AT-MP	CNAM	Medical consultation	Worker
Evrest	Gis Evrest	Systematic occupational health interviews	Worker
MCP	SPF	Compulsory professional medical consultation	Worker
RNV3P	Anses	Medical consultation with a specialist of CCPPE	Health problem
Data source	**Target population**	**Content**	**Example**
Sumer	All French workers	340 columns representing exposures to which workers are exposed to.	Exposure to lead [yes ; no]
C2P	All workers employed by private companies	10 columns representing exposures to which employers declared the worker are exposed to.	Exposure to repetitive movements [yes ; no]
Colchic / Scola	All French companies with exposure measurements / All French companies with regulatory controls	460 columns representing the measurement of the intensity of the concentration of 230 substances in the air with regards to the regulatory limit value.	Lead Intensity [moderate ; high ; very high]
Colphy	French companies with exposure measurements in national campaign or volunteering	Four columns representing the measurement of the intensity of the emissivity of two physical exposures with regards to the regulatory limit value.	Whole body vibration Intensity [moderate ; high ; very high]
MatGene	All French workers regardless of their status or employer (salaried or self-employed, private or public)	Four columns representing the nature and intensity of exposure according to occupation.	Night work [yes ; no]
AT-MP	All French workers affiliated to the general health care system	86 columns representing occupational recognised diseases.	Spondylopathies [yes ; no]
Evrest	All employees born in October over 18 years old with at least two months of seniority	45 columns representing the percentage of workers concerns by exposure. 15 columns representing clinical signs. 15 columns representing first treatment.	Noise [yes ; no] Treatment for hearing problems [yes ; no]
MCP	All employees reported by a physician as sick due to occupational exposure	720 columns representing exposures associated with reported occupational diseases and 56 columns representing these work-related diseases.	[Allergic contact dermatitis] [Chemical agents]
RNV3P	All registered people (craftsmen, apprentices, students, disabled, self-employed, unemployed, retired…)	129 columns representing the occupational diseases identified and 908 columns representing the exposures probably linked to these diseases.	[Scoliosis] [Heavy loads ; Awkward postures]

Abbreviations: Anses: French Agency for Food, Environmental, and Occupational Health & Safety; CNAM: National Fund for Health

Insurance; Dares: Directorate of Research, Economic Studies, and Statistics; Gis Evrest: Scientific Interest Group for a monitoring system; INRS: The French National Research and Safety Institute for the Prevention of Occupational accidents and Diseases; SPF: Santé publique France.

These databases are confidential, and access to and use of them is restricted by agreement with the various health organisations that supply them.

C. *Mapping between data schemas and ontologies*

The mapping between data schemas and ontologies is performed in three stages (Figure 3):

- Defining Occupational Groups: Combinations of activity sector (NAF division), occupation (PCS intermediate level), and sex are used to create unique occupational groups, such as "Skilled craftsmen in building (PCS 632, sex male) in specialised construction activities (NAF 43)".
- Standardising Health Outcomes: Variables from the AT-MP, MCP, and RNV3P data sources are mapped to disease codes in the ICD-10 classification, grouping diseases into blocks such as "Other dorsopathies" (M50-M54).
- Identifying Occupational Exposures: Exposure variables are mapped to the TEP by comparing descriptions in the data source collection protocols with standardised descriptions of exposures in the TEP. Expert working groups were convened to validate these mappings. For example, the variable exposure "manual handling of heavy loads" in C2P, which refers to any activity necessitating the utilisation of human force (lifting, lowering, transporting object), is mapped to the TEP category "handling loads and handling people". Similar mappings were performed for all data sources (SUMER, MatGene, C2P, Evrest, COLCHIC/SCOLA, COLPHY, MCP, RNV3P).

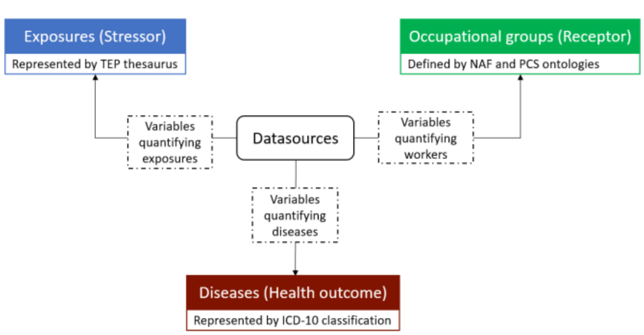

Figure 3. General representation of data and ontologies used, linked to the AOExO ontology.

IV. GENERAL REPRESENTATION OF THE HEALTH OCCUPATIONAL DATA USING ODBI AND ADAPTED OEXO

The proposed methodology was applied to construct a total of 11,331 occupational groups, identify 88 occupational exposures, and document 191 occupational diseases. The resulting integrated dataset can be represented as a complex database comprising over 3 million exposure-disease pairings across various occupational groups, as illustrated in Figure 4. Due to the inherent limitations of the data sources, the database inevitably contains missing data. The extent of occupational risk exposure and disease development among workers is influenced by their respective occupations and the specific tasks they perform.

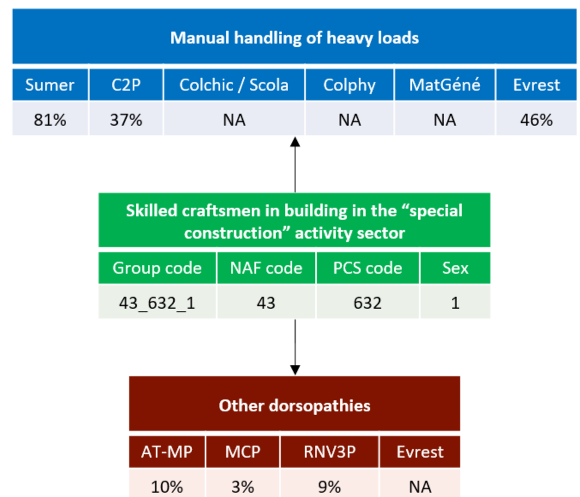

Figure 4. Example of pairing of exposure (blue – stressor) and disease (red – health outcome) for an occupational group (green – receptor).

The case studies discussed in the next two sections demonstrate practical applications of the integrated data. Each case study is defined by distinct objectives and methodological approaches including appropriate indicators. The methodologies developed can be applied to individual occupational groups or extended across multiple occupations and activity sectors.

The development of methodologies and analyses was facilitated by RStudio, an integrated development environment (IDE), using version 4.3.2 (https://docs.posit.co/ide/user/).

V. CASE STUDY ONE: QUANTITATIVE ANALYSIS

The integration of multiple data sources can be used to facilitate the visualisation of occupational exposures and diseases of greatest concern for each occupational group. By bringing different sources of data together, it becomes possible to identify significant risks to occupational health, irrespective of the perspective from which the work situation is analysed. The availability of this information enhances the targeted implementation of safety measures aimed at mitigating exposures associated with specific work environments and reducing the risk of occupational diseases. To this end, an indicator has been constructed for each exposure pair and disease pair to each occupational group.

A. *Indicator construction method for simple exposure or disease*

Due to the heterogeneity of the data, the recorded values for each exposure and disease vary in scale. For example, the Sumer dataset is representative of French workers, whereas the Scola dataset originates from companies that are subject to regulatory controls (see Table 1). To standardise the data for indicator construction, values from each original source were transformed into non-parametric values and discretised on a scale from 1 to 10, where 1 represents "very low" and 10 denotes "very high". The discretised values were then averaged based on the type of information they represent, *i.e.*, exposure or disease. The exposure indicator was calculated as the mean of the exposure intensity values across all data sources, while the disease indicator was derived as the mean of the disease intensity values from all sources.

As an illustrative example, skilled craftsmen in the "special construction" activity sector (43_632_1) may be exposed to manual handling of heavy loads, with some workers experiencing dorsopathies. For this occupational group, exposure data were available from the Sumer, C2P, and Evrest sources (Table II), while disease data were obtained from the AT-MP, MCP, and RNV3P sources. Each original value was discretised according to the data available for manual handling of heavy loads and dorsopathies. The exposure indicator was determined by averaging the values from exposure data sources, yielding a score of 7.3, whereas the disease indicator, calculated as the mean of the disease data sources, resulted in a score of 5.5.

TABLE II. SUMMARY TABLE OF DATA ON MANUAL HANDLING OF HEAVY LOADS AND DORSOPATHIES FOR THE OCCUPATION GROUP "43_632_1"

	Data source	***Original data source value (%)***	***Discretised value***	***Indicator***
Manual handling of heavy loads	Sumer	81	8	Exposure indicator: 7.3
	MatGene	/	/	
	C2P	37	4	
	Evrest	46	10	
	Colchic /Scola	/	/	
	Colphy	/	/	
Other dorsopathies	MCP	3	6	Disease indicator: 5.5
	RNV3P	9	6	
	AT-MP	10	5	

The presence of missing data across different data sources is interpreted as a lack of available information rather than an absence of exposure or disease. The Sumer dataset is the sole source that distinguishes between non-exposed workers and non-respondents. When constructing indicators, missing data are disregarded, but the information on non-exposed workers provided by Sumer is taken into account.

The constructed indicators are then represented in the form of a heatmap, where indicator values are visualised using a colour gradient. Cool colours, such as blue and green, represent lower indicator values, while warm colours, such as orange and red, indicate higher values Each heatmap focuses on a specific category of information—exposure or disease—forming the y-axis, while the x-axis represents occupational groups, activity sectors, or occupations. Three types of heatmap are presented in this case study: (1) a heatmap of the most exposed occupational groups in the construction sector, (2) a heatmap of the most exposed activity sectors, and (3) a heatmap of the most exposed occupations. In the latter two heatmaps, the indicators represent the average values for all occupational groups within the same activity sector or occupation. For example, in the case of activity sector 43, labelled "specialised construction activities", indicators were calculated as the mean values of all occupations within this sector, including directors, secretaries, and manual workers.

B. *Visualisation of exposure and disease indicators in the construction sector*

As illustrated in Figure 5, the left-hand panel presents the ten most exposed occupational groups within the construction sector—a critical case study in France—while the right-hand panel displays the disease indicators for these same groups. The x-axis represents various occupational groups arranged in ascending order according to the number of exposures with those experiencing fewer exposures positioned on the left and those with greater exposure levels on the right. The y-axis illustrates the exposures or diseases, ranked according to the cumulative sum of their respective indicators. Exposures or diseases affecting a significant number of occupational groups appear towards the upper portion of the y-axis, while those impacting fewer groups are positioned lower.

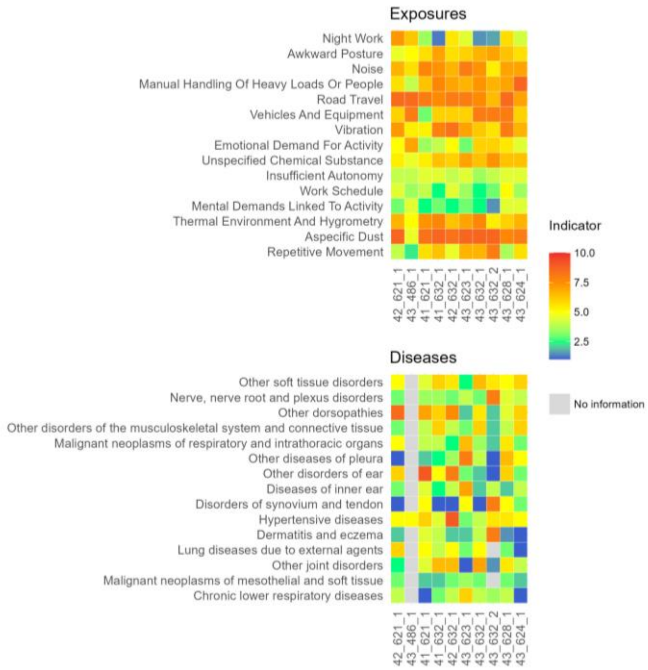

Figure 5. Heatmap of indicators by occupational group, exposure and disease: focus on 10 most exposed occupational groups and 15 exposures and diseases.

For instance, skilled craftsmen in building in the "special construction" activity sector (43_632_1 see above) were identified as the fourth most exposed and the most affected by occupational diseases, primarily due to the presence of numerous high-value indicators.

The most exposed occupational groups are particularly vulnerable to physical exposures such as noise, manual handling of heavy loads, road travel, vehicle use, and vibration. They also face exposure to certain chemical hazards, including non-specific dust, asbestos, silica, and construction products, commonly encountered in the construction industry. Furthermore, some groups are significantly affected by psychosocial and organisational risks, such as high emotional demands, limited autonomy, and mental strain. Musculoskeletal disorders (*e.g.*, soft tissue disorders, nerve root disorders), and hearing impairments (*e.g.*, inner ear diseases), are prevalent in this sector.

Although the present study does not quantify the direct correlation between exposures and diseases, it provides a qualitative assessment that underscores the need to enhance preventive measures to reduce worker exposure. This prompts reflection on the development of safety interventions tailored to work environments characterised by multiple high-risk exposures. Several ongoing studies seek to propose more effective preventive solutions [18], [19], [20].

C. *Visualisation of exposure and disease indicators across all sectors*

Figures 6 and 7 helps to identify the most exposed activity sectors alongside the most concerning exposures across all sectors. The x-axis displays all French activity sectors, labelled with letters below the heatmap, along with their corresponding divisions, represented by numbers. For example, the construction sector (labelled "F") includes divisions 41, 42, and 43. The y-axis represents the exposures identified through the data integration method, structured according to the hierarchical TEP classification, as indicated to the left of the heatmap. A detailed breakdown of the activity sector divisions and exposures can be found in Annexes A and B.

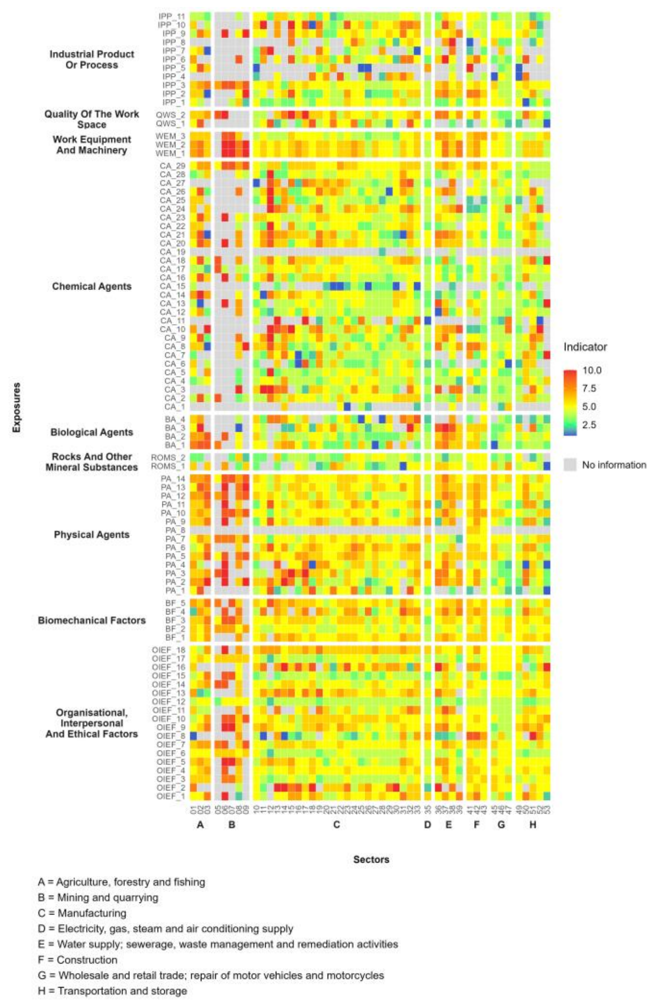

Figure 6. Heatmap of averaged indicators of all occupational groups for activity sectors A to H and by exposure.

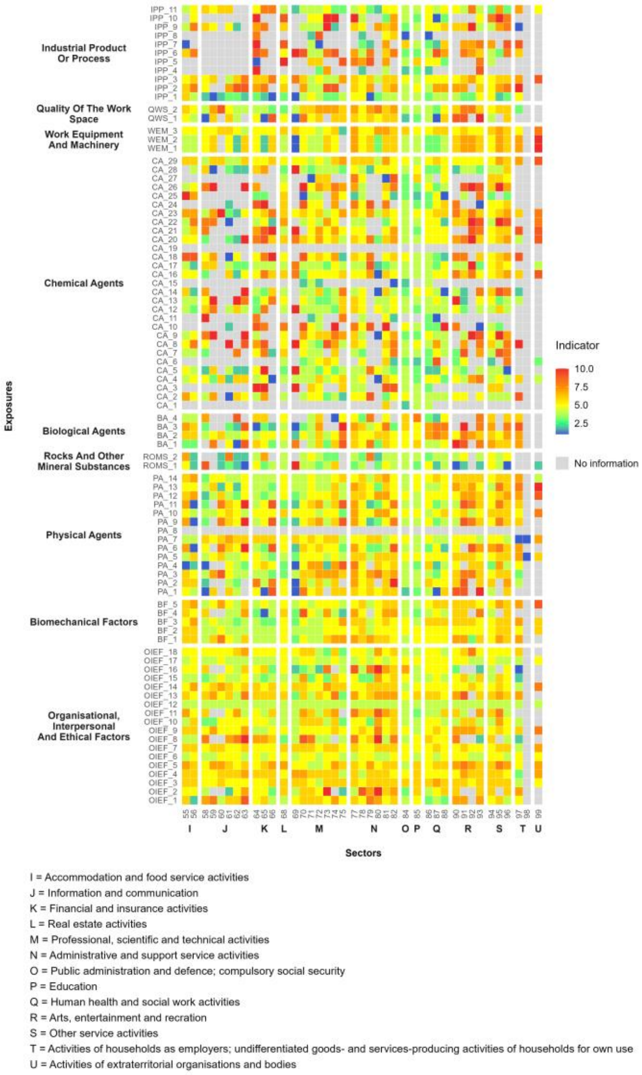

Figure 7. Heatmap of averaged indicators of all occupational groups for activity sectors I to U and by exposure.

The heatmap also reveals gaps in data availability, particularly in extractive industries (B), household employer activities (T), and extra-territorial activities (U), which include work conducted in diplomatic organisations such as embassies and consulates. Furthermore, it highlights sector-specific exposures, demonstrating the necessity for tailored preventive strategies.

For instance, workers in the agriculture sector (A) demonstrated elevated indicators of exposure to industrial materials and products (IPP_3), exposure to work equipment and machinery and some chemical agents (CA_10, CA_14, or CA_20), as well as exposure to biological agents, a preponderance of physical and biomechanical exposures, and several psychosocial-organisational factors. These observations are consistent with the findings of previous research that explored the identification of exposure profiles [21] and identified that agricultural workers are mainly exposed to organisational, chemical, physical, and biological constraints.

Another example is the sanitation and waste management sector (E), which presents high indicators for exposure to industrial products and processes, work equipment and machinery, some chemical agents (CA_3, CA_10, CA_20, and CA_21), biological agents, in particular microbiological agents (BA_3), as well as a majority of physical and biomechanical exposures, and some psychosocial-organisational factors. The results of this example are also consistent with the work of C. Fourneau *et al.* [21], who identify sanitation workers as being largely exposed to organisational, chemical, physical, and biological constraints.

The heatmap further elucidates the sectors most vulnerable to such hazards, underscoring the necessity for preventive measures tailored to the specific exposures encountered.

D. Visualisation of exposure and disease indicators across all occupations

The third heatmap of indicators (divided into Figures 8 and 9) helps to identify the occupations with the highest levels of exposure, as well as the exposures of greatest concern across all occupational categories. It highlights the disparity in available information between different occupations and their specificity. The x-axis represents all existing occupations (numbered below the heatmap) along with their corresponding intermediate levels (numbered below the heatmap). For instance, farm operators (number 1) are associated with intermediate levels 111, 122, and 131. The y-axis follows the same structure as the previous heatmap, depicting exposures identified through the data integration method, grouped according to the eight hierarchical levels of TEP. A detailed breakdown of the intermediate levels of occupations can be found in Annex C.

The figure reveals discrepancies in the availability of information across various occupations, particularly for farm operators (1) and craftsmen, shopkeepers, and company managers (2), which are associated with very few indicators. Each occupation is characterised by specific exposures linked to the tasks performed by workers, as illustrated by the heatmap.

For example, medium-sized farmers (122) exhibit elevated indicators for exposure to work equipment and machinery, as well as exposure to animals (BA_1) and biological agents (BA_2). Additionally, they face exposure to a range of physical, biomechanical, and psychosocial-organisational agents. These findings align with existing literature, which identifies farming as a high-risk occupation involving significant physical effort, frequent use of heavy machinery, and exposure to chemical substances, all of which are inherent to agricultural activities such as cultivation and livestock breeding [22], [23].

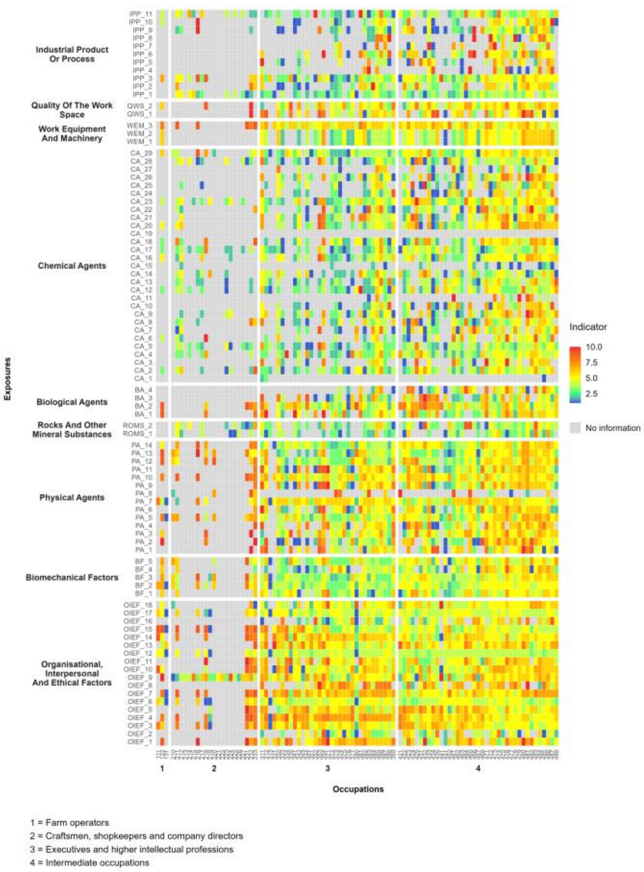

Figure 8. Heatmap of averaged indicators of all occupational groups for occupations 1 to 4 and by exposure.

Another notable example is that of police officers and prison wardens (531), who demonstrate high levels of exposure to biological agents (BA_2), as well as certain physical agents such as road travel (PA_10) and thermal environments (PA_12). Moreover, they are frequently exposed to psychosocial-organisational stressors, including external violence (OEIF_5), night work (OEIF_9), and workstation-related factors (OEIF_18). These observations are corroborated by existing research, which highlights that police officers and prison wardens routinely encounter challenging work environments. Their duties, which involve surveillance and maintaining authority, often expose them to various psychosocial stressors, including verbal and physical aggression, as well as organisational challenges such as atypical working hours and night shifts. Furthermore, they are at risk of exposure to biological substances such as blood [24], [25].

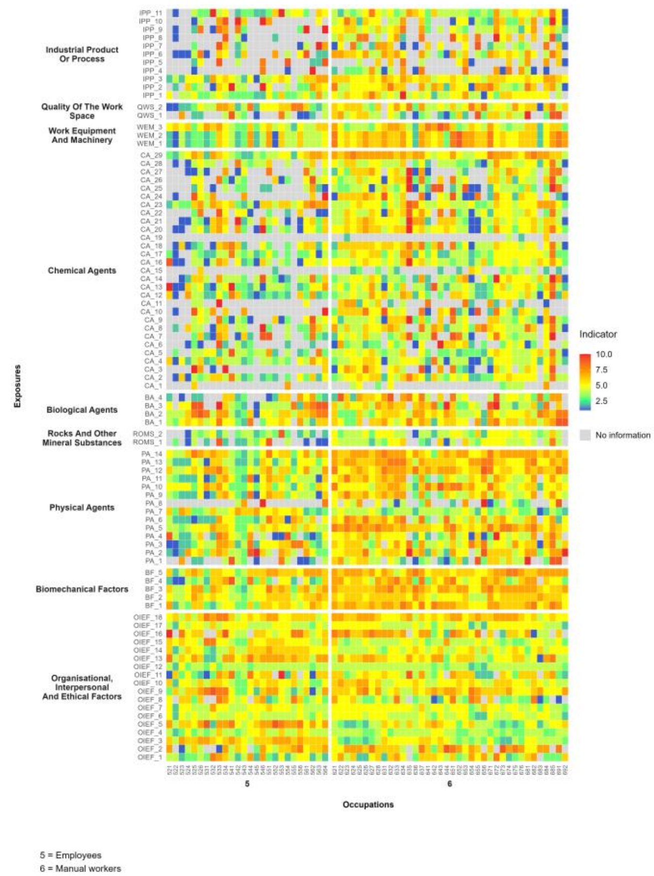

Figure 9. Heatmap of averaged indicators of all occupational groups for occupations 5 to 6 and by exposure.

The heatmap provides a visual representation of occupational exposure profiles, which can be directly linked to the specific tasks performed in each profession. It serves as a valuable tool for identifying high-risk exposure profiles and triggering the necessary protective measures to mitigate occupational hazards.

VI. CASE STUDY TWO: QUALITATIVE ANALYSIS

The integrated dataset facilitates the verification of information concordance across different data sources, thereby enabling the identification of the number of sources reporting on each exposure and disease. This approach makes it possible to visualise both the most and least frequently studied exposures and diseases. When combined with the first case study, this methodology helps to confirm which exposures pose the greatest risks to occupational health—particularly those with high indicator values, studied across multiple sources, and exhibiting a high degree of consistency. To evaluate this concordance, a consistency score has been developed for each exposure and disease based on occupational group.

A. *Consistency score construction method for simple exposure or disease*

In the data sources, information—excluding missing data—can either confirm or refute the presence of an exposure or disease for a given occupational group. Confirming data sources refer to those that contain values greater than zero for an exposure or disease, indicating that at least one worker in the occupational group has been exposed or affected by the disease. Conversely, refuting data sources contain zero values, signifying no recorded exposure or disease occurrence within the occupational group. The consistency score is determined by comparing the number of sources that confirm an exposure or disease with those that refute it.

If all data sources either confirm or refute a particular exposure or disease, the consistency is categorised as strong, indicating complete agreement among all sources. If an equal number of sources confirm and refute the presence of an exposure or disease, the consistency is categorised as weak, indicating a complete disagreement among all sources. In cases where the number of confirming and refuting sources is unequal, the consistency is categorised as medium. This methodology for calculating the consistency score can be generalised to any number of sources greater than one.

An illustrative case involves skilled craftsmen in building in the "special construction" sector, who are exposed to manual handling of heavy loads and suffer from inner ear diseases. Data on exposure are derived from the Sumer, C2P, and Evrest data sources, while disease-related data come from the AT-MP, MCP, and RNV3P sources (Table III). All exposure-related values in the dataset are greater than zero, and no data sources refute exposure. Consistency is therefore classified as strong. However, in the case of disease, the AT-MP source confirms the presence of disease (positive value), while the MCP and RNV3P sources refute it (zero values). This imbalance results in medium consistency.

An additional example concerns higher-level female secretaries in the "special construction" sector who are exposed to night work. Four data sources provide information on this exposure: MatGene, C2P, Sumer, and Evrest. The first two sources confirm exposure (non-zero values), while the latter two refute it (zero values). As the number of confirming and refuting sources is equal, there is weak consistency in this example.

TABLE III. EXAMPLE OF CONSISTENCY SCORES APPLIED TO DATA

Occupational group	*Exposure / Disease*	*Confirming data sources*	*Refuting data sources*	*Consistency*
Skilled craftsmen in building in the "special construction" activity sector	Manual handling of heavy loads	Sumer: 81 C2P: 37 Evrest: 46	/	Strong
	Diseases of inner ear	AT-MP: 2	MCP: 0 RNV3P: 0	Medium
Higher-level female secretaries in the "special construction" activity sector	Night work	MatGene: 1 C2P: 13	Sumer: 0 Evrest: 0	Weak

Overall, these findings highlight the need for a thorough assessment of the health implications of manually handling heavy loads, as multiple sources indicate that a significant number of workers are exposed to this risk. However, further clarification is required to establish the threshold for night work exposure. Although current methodologies do not enable direct linkage with occupational diseases, safety measures should be reviewed to ensure adequate worker protection.

The consistency scores calculated for each exposure and disease are subsequently visualised using two types of graphical representation: a heatmap and a bar chart. The heatmap depicts the consistency scores of occupational groups for each exposure or disease, while the bar chart illustrates the distribution of consistency scores for exposures across occupational groups based on the number of data sources available.

B. *Visualisation of exposure and disease consistency scores in the construction sector*

Consistency scores are represented using a blue gradient, where dark blue indicates high consistency and light blue signifies low consistency. The x-axis represents the same occupational groups as in Figure 5 affected by exposures or diseases, while the y-axis displays the exposures or diseases under consideration (see Figure 10).

For example, skilled craftsmen in building in the "special construction" activity sector (group 43_632_1 discussed above) exhibit high consistency scores for manual handling of heavy loads. This occupational group is characterised by a high proportion of strong consistency scores for exposures, with a more moderate representation for diseases. The exposures and occupational groups align with those identified in Case Study 1, where concerns were categorised according to their severity. Notably, the consistency of data sources is strong for most exposures and occupational groups, suggesting that exposures of significant concern are well-documented across all data sources. However, the occupational group 43_486_1 corresponding to foremen and supervisors (excluding administrative supervisors) in maintenance in the "special construction" activity sector lacks data, likely due to the under-recognition of diseases as occupational or caused by workplace exposure. The available information for this group primarily derives from self-reports during medical interviews, revealing a deficiency in how it is monitored.

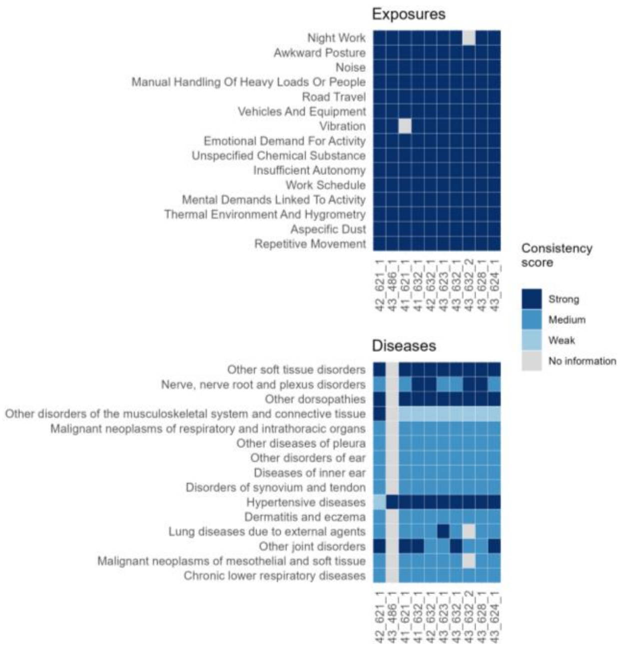

Figure 10. Heatmap of the consistency score by occupational group and by exposure and disease: focus on 10 exposed occupational groups and 15 exposures and diseases.

Overall, consistency scores are weaker for diseases than for exposures. This discrepancy arises from the lower visibility of occupational diseases—some of which, such as hearing loss, are often not officially recognised as occupational illnesses. As a result, such conditions may not be included in the main occupational disease database (AT-MP) but could still be recorded in other databases. This underscores the necessity of integrating diverse disease-related data sources to establish a more comprehensive understanding of occupational health risks.

C. *Visualisation of exposure consistency scores in all occupational groups*

The bar chart depicted in Figure 11 presents the distribution of consistency score values for exposures across all occupational groups. It is structured into three sections, based on the number of data sources reporting on each exposure. For example, three databases provide information on "wood dust and other plant-based products" (first bar of the graph), while two databases report on "welding fumes" (second bar). The consistency score is only calculated for exposures with information available from at least two data sources. Nearly half of the exposures (n=40) were reported by only one data source and were therefore excluded from the graph. Of the remaining exposures, the majority (n=38) were assessed using data from two sources. Only ten exposures were identified in three sources, and just two exposures were reported across four sources, representing the maximum.

For each section, the x-axis indicates the percentage of occupational groups affected by an exposure, categorised as having high, medium, or low consistency scores. The y-axis represents exposures, grouped according to the eight hierarchical levels of the TEP. The consistency scores are visually represented using four colours: red for high consistency, yellow for medium consistency, green for low consistency, and blue for non-exposed occupational groups. This visualisation highlights the most frequently studied exposures, as represented by a high number of data sources.

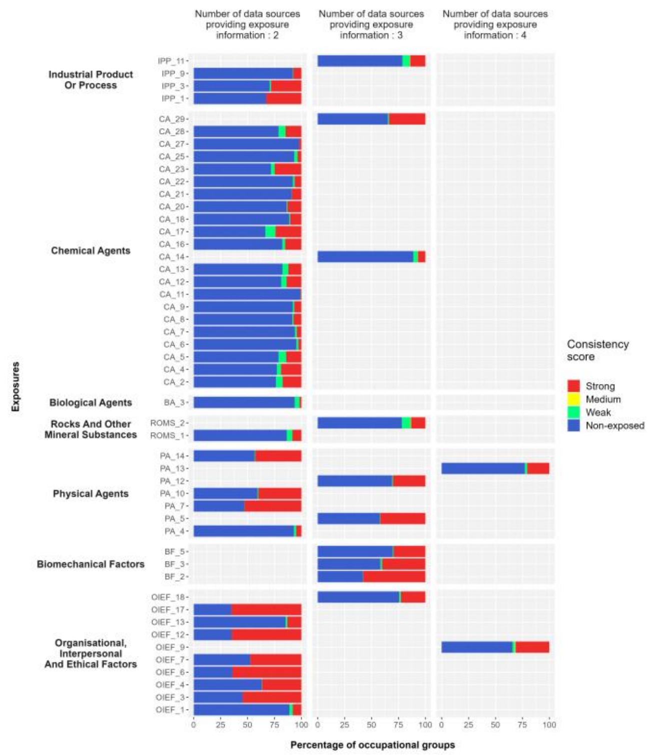

Figure 11. Distribution of exposure consistency scores in all occupational groups.

The distribution of exposure consistency scores reveals that most occupational groups are not exposed to the hazards under study. A substantial proportion of occupational groups with strong consistency scores is observed in relation to organisational, interpersonal, and ethical factors, as well as biomechanical factors and physical agents. A smaller proportion of groups with weak consistency scores is also present, albeit dispersed among different exposure types. Notably, exposures associated with medium consistency scores are nearly invisible on the graph (11 exposures). Importantly, weak consistency does not necessarily reflect poor quality of the data; rather, it may indicate that the information from multiple sources is complementary. This holds true for all exposure types except chemical agents and substances. Additionally, the analysis highlights disparities in data availability for certain exposures, such as biological agents, where only microbiological exposures are documented by more than one data source.

VII. CONCLUSION AND FUTURE WORK

The application of the Occupational Health Data Integration (OBDI) methodology in structuring the dataset has enabled us to implement two distinct analytical approaches for examining occupational health data. To the best of our knowledge, this study represents the first attempt to integrate ten heterogeneous data sources, utilising four domain ontologies. We constructed indicators in order to visualise information as well as underscore occupational groups exposed to multiple hazards and affected by various diseases. Furthermore, consistency scores have been useful in validating these indicators and assessing the complementarity of the data. This methodological approach has facilitated the confirmation of a common body of knowledge, as evidenced by a comparison of our results with those presented in the existing literature.

The two case studies presented here allow for both in-depth analyses at the level of specific occupational groups and broader evaluations encompassing occupations and activity sectors. These case studies offer a comprehensive and precise depiction of exposure and disease profiles across occupational groups, highlighting exposures that impact a significant proportion of the workforce and necessitate ongoing or enhanced safety measures. The visualisation of diseases aids in identifying those profiles that cause the greatest concern, thereby helping in discerning which preventive strategies should be prioritised.

The indicators and consistency scores, although relatively simple and not exhaustive, provide a versatile analytical framework due to the adaptability of their construction methodology. While the case studies serve as illustrative examples of the potential applications of integrated data, we contend that our methodological approach to data integration could be expanded to incorporate additional data sources. An avenue for future research may involve enhancing the first case study by integrating data on the total number of workers within each occupational group, thereby allowing for a more precise estimation of the proportion of exposed and affected workers.

The use of ontologies—particularly the ExO (Exposure Ontology)—provides a powerful framework within which to enhance the integration, interoperability, and value extraction of data in the field of occupational health. National stakeholders in this domain—including public agencies, occupational health services, and research institutes—routinely collect diverse datasets related to occupational exposures, job activities, diseases, and intervention contexts. However, these datasets are often gathered for different purposes, by different actors, and stored in heterogeneous systems, which are rarely connected. This fragmentation severely limits the ability to perform global analyses, conduct longitudinal monitoring, or identify emerging risks.

In this context, the Ontology-Based Data Integration (OBDI) approach, as formalised by Calvanese *et al.*, offers a robust framework for semantically linking these disparate data sources through a shared ontology. The ExO ontology, developed by Mattingly *et al.*, provides an explicit conceptual foundation for describing entities, processes, and relationships in the exposure science domain. It helps standardise terminology, it formalises relationships (*e.g.*, between chemical agents, exposure routes, and occupational activities), and makes data semantically interoperable.

The combined use of ExO and the OBDI approach would enable occupational health stakeholders to answer complex, cross-domain questions that isolated datasets cannot address—for example, identifying multi-source exposures, linking surveillance data with reported occupational diseases, or mapping emerging risks based on various datasets. This approach contributes not only to better data governance, but also—and more importantly—to informed, coordinated decision-making aimed at improving workplace health and safety.

The proposed contextualisation of the data introduces novel perspectives for its application in risk assessment.

For instance, the analysis of correlations between exposure and/or co-exposure indicators could facilitate the identification of key occupational risks, ultimately contributing to the development of a risk assessment tool [26]. The use of these indicators could highlight risks that are little known or studied, and for which preventive measures still need to be developed. Another possibility would be to evaluate or question the effectiveness of measures already put in place for common combined risks to which workers declare themselves to be exposed, in line with the number of accidents at work.

Alternatively, these indicators could enable the creation of worker exposure profiles, consolidating information on all potential exposures and co-exposures associated with a given occupation [21]. Such profiles could be subsequently employed in order to anticipate future occupational diseases and implement occupation-specific safety measures. When calculating the indicator, the tasks performed, as well as biological characteristics such as sex should also be considered in order to better represent the diversity of work situations and workers, and to compare differences in exposure with a view to improving individual safety.

Although our article is focused on the field of occupational health, the proposed method of use and case studies could be generalised to the broader healthcare sector. Current research in the health sector has increasingly focused on the contextualisation of heterogeneous data to establish a common semantic framework aimed at enhancing knowledge sharing [27], [28]. Generalising our approach could facilitate the development of targeted responses to both existing and emerging health concerns through specialised tools and analytical queries. Moreover, it would promote interdisciplinary collaboration and facilitate the generation of knowledge grounded in practical, real-world scenarios.

ACKNOWLEDGEMENTS

The authors would like to thank the various healthcare organisations that agreed to participate in the DataPOST project and to share their data. The work presented was based on data from the Dares Sumer survey, data supplied by the GIS EVREST and made available on 01/08/2024, and data provided by the Anses, CNAM, SPF, and INRS.

The authors would also like to thank E. Algava and M. Duval (Dares), L. Meunier (CNAM), C. Pilorget, J. Chatelot, and J. Homère (SPF), C. Nisse and A. Aachimi (Anses), and L. Rollin and A. Leroyer (Evrest) for their expertise.

REFERENCES

[1] C. Barbey, N. Bonvallot, M. Smaïl-Tabbone, and F. Clerc, 'Ontology-Based Integration of Occupational Health Data: Method and Case Studies', *DATA Anal.*, 2024.

[2] C. Barbey, N. Bonvallot, and F. Clerc, 'Health Outcomes Related to Multiple Exposures in Occupational Settings: A Review', *Saf. Health Work*, vol. 15, no. 4, pp. 382–395, Dec. 2024, doi: 10.1016/j.shaw.2024.10.004.

[3] L. Rollin *et al.*, 'What complementarity for french occupational health databases (Evrest, MCP, Sumer, RNV3P): an example from the women home help workers?', *Arch. Mal. Prof. Environ.*, vol. 82, no. 3, pp. 261–276, May 2021, doi: 10.1016/j.admp.2020.11.002.

[4] L. Rollin *et al.*, 'Les conducteurs de poids lourd, transport en commun et livreurs sont-ils à risque d'affections du membre supérieur ?', *Arch. Mal. Prof. Environ.*, vol. 85, no. 6, p. 102765, Nov. 2024, doi: 10.1016/j.admp.2024.102765.

[5] R. Dandan, S. Despres, and J. Nobecourt, 'OAFE: An Ontology for the Description of Elderly Activities', in *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, Nov. 2018, pp. 396–403. doi: 10.1109/SITIS.2018.00068.

[6] 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_80667-1.

[7] A. Z. E. Qutaany, O. M., and A. H., 'A Mapping Approach for Fully Virtual Data Integration System Processes', *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 12, 2018, doi: 10.14569/IJACSA.2018.091216.

[8] H. Zhang *et al.*, 'An ontology-guided semantic data integration framework to support integrative data analysis of cancer survival', *BMC Med. Inform. Decis. Mak.*, vol. 18, no. S2, p. 41, Jul. 2018, doi: 10.1186/s12911-018-0636-4.

[9] R. Dandan, 'Intégration de données guidée par une ontologie : Application à l'assistance personnalisée aux personnes âgées', Université Paris-Nord - Paris XIII, 2022. [Online]. Available: https://theses.hal.science/tel-03945528

[10] 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.

[11] J. Bloch *et al.*, 'National Network for Monitoring Prevention of a Occupational Disease (RNV3P) - 2022 Annual Report', Oct. 2023.

[12] J. A. Hirsch *et al.*, 'ICD-10: History and Context', *Am. J. Neuroradiol.*, vol. 37, no. 4, pp. 596–599, Apr. 2016, doi: 10.3174/ajnr.A4696.

[13] C. Brossard and I. Falinower, 'Salariés déclarés exposés à des risques de pénibilité en 2016 : portrait':, *Retraite Société*, vol. N° 77, no. 2, pp. 139–163, Oct. 2018, doi: 10.3917/rs1.077.0139.

[14] G. Mater, C. Paris, and J. Lavoué, 'Descriptive analysis and comparison of two French occupational exposure databases: COLCHIC and SCOLA', *Am. J. Ind. Med.*, vol. 59, no. 5, pp. 379–391, May 2016, doi: 10.1002/ajim.22569.

[15] F. Bonthoux and N. Schmitt, 'Colphy : base de données dédiée aux nuisances physiques', *Hyg. Sécurité Trav.*, no. 270, pp. 61–65, 2023.

[16] J. Févotte *et al.*, 'Matgéné: A Program to Develop Job-Exposure Matrices in the General Population in France', *Ann. Occup. Hyg.*, vol. 55, no. 8, pp. 865–878, Sep. 2011, doi: 10.1093/annhyg/mer067.

[17] Assurance Maladie - Risques professionnels, 'Rapport annuel 2023 de l'Assurance Maladie - Risques professionnels', Paris, Dec. 2024.

[18] Z. Zhu, 'Exoskeletons for manual material handling – A review and implication for construction applications', *Autom. Constr.*, 2021.

[19] J. J. Devereux, I. G. Vlachonikolis, and P. W. Buckle, 'Epidemiological study to investigate potential interaction between physical and psychosocial factors at work that may increase the risk of symptoms of musculoskeletal disorder of the neck and upper limb', *Occup. Environ. Med.*, vol. 59, no. 4, pp. 269–277, Apr. 2002, doi: 10.1136/oem.59.4.269.

[20] L. L. Andersen, J. Vinstrup, E. Sundstrup, S. V. Skovlund, E. Villadsen, and S. V. Thorsen, 'Combined ergonomic exposures and development of musculoskeletal pain in the general working population: A prospective cohort study', *Scand. J. Work. Environ. Health*, vol. 47, no. 4, pp. 287–295, May 2021, doi: 10.5271/sjweh.3954.

[21] C. Fourneau *et al.*, 'The French 2016-2020 National Occupational Health Plan: a better understanding of multiple exposures', *Environ. Risques Santé*, vol. 20, no. 4, pp. 377–382, Aug. 2021, doi: 10.1684/ers.2021.1570.

[22] T.-H.-Y. Nguyen, M. Bertin, J. Bodin, N. Fouquet, N. Bonvallot, and Y. Roquelaure, 'Multiple Exposures and Coexposures to Occupational Hazards Among Agricultural Workers: A Systematic Review of Observational Studies', *Saf. Health Work*, vol. 9, no. 3, pp. 239–248, Sep. 2018, doi: 10.1016/j.shaw.2018.04.002.

[23] M. Ekmekci and S. Yaman, 'Occupational health and safety among farmers: a comprehensive study in Central Anatolia, Turkey', *BMC Public Health*, vol. 24, no. 1, p. 2732, Oct. 2024, doi: 10.1186/s12889-024-20249-7.

[24] R. Anders, A. Frapsauce, C. Sauvezon, and D. Gilibert, 'Police officer occupational health: a model of organizational constraints, trauma exposure, perceived resources, and

agency', *J. Occup. Med. Toxicol.*, vol. 19, no. 1, p. 46, Dec. 2024, doi: 10.1186/s12995-024-00444-3.

[25] S. St. Louis, N. A. Frost, C. E. Monteiro, and J. Trapassi Migliaccio, 'Occupational Hazards in Corrections: The Impact of Violence and Suicide Exposures on Officers' Emotional and Psychological Health', *Crim. Justice Behav.*, vol. 50, no. 9, pp. 1361–1379, Sep. 2023, doi: 10.1177/00938548231177710.

[26] A. J. Williams, J. C. Lambert, K. Thayer, and J.-L. C. M. Dorne, 'Sourcing data on chemical properties and hazard data from the US-EPA CompTox Chemicals Dashboard: A practical guide for human risk assessment', *Environ. Int.*, vol. 154, p. 106566, Sep. 2021, doi: 10.1016/j.envint.2021.106566.

[27] R. R. Boyles, A. E. Thessen, A. Waldrop, and M. A. Haendel, 'Ontology-based data integration for advancing toxicological knowledge', *Curr. Opin. Toxicol.*, vol. 16, pp. 67–74, Aug. 2019, doi: 10.1016/j.cotox.2019.05.005.

[28] R. R. Rao, K. Makkithaya, and N. Gupta, 'Ontology based semantic representation for Public Health data integration', in *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, Mysore, India: IEEE, Nov. 2014, pp. 357–362. doi: 10.1109/IC3I.2014.7019701.

ANNEXES

Annex A. Correspondence between codes and labels for activity sector divisions.

A - Agriculture, forestry, and fishing	
01	Crop and animal production, hunting, and related service activities
02	Forestry and logging
03	Fishing and aquaculture
B - Mining and quarrying	
05	Mining of coal and lignite
06	Extraction of crude petroleum and natural gas
07	Mining of metal ores
08	Other mining and quarrying
09	Mining support service activities
C - Manufacturing	
10	Manufacture of food products
11	Manufacture of beverages
12	Manufacture of tobacco products
13	Manufacture of textiles
14	Manufacture of wearing apparel
15	Manufacture of leather and related products
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
17	Manufacture of paper and paper products
18	Printing and reproduction of recorded media
19	Manufacture of coke and refined petroleum products
20	Manufacture of chemicals and chemical products
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
22	Manufacture of rubber and plastic products
23	Manufacture of other non-metallic mineral products
24	Manufacture of basic metals
25	Manufacture of fabricated metal products, except machinery and equipment
26	Manufacture of computer, electronic, and optical products
27	Manufacture of electrical equipment
28	Manufacture of machinery and equipment
29	Manufacture of motor vehicles, trailers, and semi-trailers
30	Manufacture of other transport equipment
31	Manufacture of furniture
32	Other manufacturing
33	Repair and installation of machinery and equipment
D - Electricity, gas, steam, and air conditioning supply	
35	Electricity, gas, steam, and air conditioning supply
E - Water supply; sewerage, waste management, and remediation activities	
36	Water collection, treatment, and supply
37	Sewerage
38	Waste collection, treatment, and disposal activities; materials recovery
39	Remediation activities and other waste management services
F - Construction	
41	Construction of buildings
42	Civil engineering
43	Specialised construction activities
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	
45	Wholesale and retail trade and repair of motor vehicles and motorcycles
46	Wholesale trade, except of motor vehicles and motorcycles
47	Retail trade, except of motor vehicles and motorcycles
H - Transportation and storage	
49	Land transport and transport via pipelines
50	Water transport
51	Air transport
52	Warehousing and support activities for transportation
53	Postal and courier activities
I - Accommodation and food service activities	
55	Accommodation
56	Food and beverage service activities
J - Information and communication	
58	Publishing activities
59	Motion picture, video, and television programme production, sound recording, and music publishing activities
60	Programming and broadcasting activities
61	Telecommunications
62	Computer programming, consultancy, and related activities
63	Information service activities
K - Financial and insurance activities	
64	Financial service activities, except insurance and pension funding
65	Insurance, reinsurance, and pension funding, except compulsory social security
66	Activities auxiliary to financial services and insurance activities
L - Real estate activities	
68	Real estate activities
M - Professional, scientific, and technical activities	
69	Legal and accounting activities
70	Activities of head offices; management consultancy activities
71	Architectural and engineering activities; technical testing and analysis
72	Scientific research and development
73	Advertising and market research
74	Other professional, scientific, and technical activities
75	Veterinary activities
N - Administrative and support service activities	
77	Rental and leasing activities
78	Employment activities
79	Travel agency, tour operator, and other reservation service and related activities
80	Security and investigation activities
81	Services to buildings and landscape activities
82	Office administrative, office support, and other business support activities
O - Public administration and defence; compulsory social security	
84	Public administration and defence; compulsory social security
P - Education	
85	Education
Q - Human health and social work activities	
86	Human health activities
87	Residential care activities
88	Social work activities without accommodation
R - Arts, entertainment, and recreation	
90	Creative, arts, and entertainment activities
91	Libraries, archives, museums, and other cultural activities
92	Gambling and betting activities
93	Sports activities and amusement and recreation activities
S - Other service activities	
94	Activities of membership organisations
95	Repair of computers and personal and household goods
96	Other personal service activities
T - Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	
97	Activities of households as employers of domestic personnel
98	Undifferentiated goods- and services-producing activities of private households for own use
U - Activities of extraterritorial organisations and bodies	
99	Activities of extraterritorial organisations and bodies

Annex B. Exposure code details.

IPP - Industrial product or process	
IPP_11	Wood dust and other plant-based product
IPP_10	Steel and metal
IPP_9	Product released in welding processes
IPP_8	Product released in foundry processes
IPP_7	Product of organic origin
IPP_6	Product of inorganic origin
IPP_5	Pharmaceutical products
IPP_4	Nanomaterial nanoparticle
IPP_3	Miscellaneous industrial equipment and products
IPP_2	Construction products
IPP_1	Non-specific dust
QWS - Quality of the work space	
QWS_2	Constrained workspace
QWS_1	Clean room and workspace with a special situation
WEM - Work equipment and machinery	
WEM_3	Vehicles and equipment
WEM_2	Machine tool
WEM_1	Construction machine
CA - Chemical agents	
CA_29	Unspecified chemical substance
CA_28	Transition metal (cadmium, chromium, nickel, cobalt, etc.)
CA_27	Sulphoxide and sulphone
CA_26	Sulphonic acid and thioacid
CA_25	Phenol and derivatives
CA_24	Organic metal compounds
CA_23	Non-metal
CA_22	Nitrile cyanate, isocyanate, and cyanurate
CA_21	Natrium, potassium, lithium, and other alkaline metals
CA_20	Metalloid
CA_19	Lanthanide and rare earths
CA_18	Lactone and lactam
CA_17	Hydrocarbons and drift
CA_16	Halogen
CA_15	Glycol
CA_14	Formaldehyde and other aldehydes
CA_13	Ether thioether and derivatives
CA_12	Ester
CA_11	Epoxy
CA_10	Carboxylic acid salt
CA_9	Carboxylic acid and peracid
CA_8	Beryllium, magnesium, and other alkaline earth metals
CA_7	Amine imine and drift
CA_6	Amide sulphonamide phosphoramide imide and thiuram
CA_5	Aluminium, lead and other poor metal
CA_4	Alcohol and polyalcohol and derivatives
CA_3	Actinide
CA_2	Acetone and other cetone, quinone, cetene, and derivatives
CA_1	Acetal and derivatives
BA - Biological agents	
BA_4	Vegetal
BA_3	Microbiological
BA_2	Biological agents
BA_1	Animal
ROMS - Rocks and other mineral substances	
ROMS_2	Silica and other sedimentary rocks
ROMS_1	Asbestos and other silicate minerals
PA - Physical agents	
PA_14	Walk on the same level
PA_13	Vibration
PA_12	Thermal environment and hygrometry
PA_11	Static field
PA_10	Road travel
PA_9	Radiation and electromagnetic fields
PA_8	Pressure above atmospheric pressure
PA_7	Other lighting and visual constraints
PA_6	Non-ionising radiation
PA_5	Noise
PA_4	Ionising radiation
PA_3	Insufficient ventilation
PA_2	Constraint linked to humidity
PA_1	Artificial light
BF - Biomechanical factors	
BF_5	Repetitive movement
BF_4	Other biomechanical factor
BF_3	Manual handling of heavy loads or people
BF_2	Awkward posture
BF_1	Another position
OIEF - Organisational, interpersonal, and ethical factors	
OIEF_18	Workstation
OIEF_17	Work schedule
OIEF_16	Work experience contrary to its principles
OIEF_15	Teleworking
OIEF_14	Special working arrangements
OIEF_13	Quality of working relationships
OIEF_12	Quality of life prevented
OIEF_11	Other organisational, relational, and ethical factors
OIEF_10	Other functional organisation of activity
OIEF_9	Night work
OIEF_8	Methods that increase psychosocial risk
OIEF_7	Mental demands linked to activity
OIEF_6	Insufficient autonomy
OIEF_5	External violence
OIEF_4	Experience of overwork or underwork
OIEF_3	Emotional demand for activity
OIEF_2	Company-related bonus factor
OIEF_1	Business travel

Annex C. Correspondence between codes and labels for intermediate professions.

1 - Farm operators	
111	Smallholding farmers
122	Activities similar to farmers on medium-sized holdings
131	Farmers on large holdings
2 - Craftsmen, shopkeepers, and company directors	
210	Craftsmen in general
211	Craftsmen in building, public works, parks, and gardens
212	Craftsmen in metalworking, mechanics, electromechanics, and electrical equipment
213	Craftsmen in textiles, clothing, and leather
214	Craftsmen in furniture, woodworking, and other manufacturing
215	Food craftsmen
216	Craftsmen in repairs and maintenance
217	Craftsmen in other services
218	Craftsmen-like
219	Craftsmen's family helpers
220	Retailers and similar in general
221	Small and medium-sized wholesalers (0 to 9 employees)
222	Small and medium-sized food retailers (0 to 9 employees)
223	Small and medium-sized specialist retailers (by field) (0 to 9 employees)
224	Operators of cafés, restaurants, and hotels (0 to 9 employees)
225	Retail intermediaries (0 to 9 employees)
226	Agents of insurance, transport, and tourism (0 to 9 employees)
227	Other service providers (0 to 9 employees)
231	Managers of large companies (500 employees or more)
232	Managers of medium-sized companies (50 to 499 employees)
233	Managers of companies with 10 to 49 employees
3 - Executives and higher intellectual professions	
311	Self-employed health professionals
312	Self-employed legal and technical professions
313	Self-employed caregivers
331	Public service: Executive staff
332	Public service: Engineers and senior technical staff
333	Public service: Administrative staff (excluding teaching and heritage)
334	Public service: Army and Gendarmerie officers
335	Public service: Political and trade union representatives
341	Professors and scientific occupations in secondary education
342	Professors and scientific occupations in higher education and public research
343	Professors and scientific occupations in educational and vocational guidance
344	Professors and scientific occupations in health
351	Information, arts, and entertainment professions: Civil service in documentation, heritage
352	Information, arts, and entertainment professions: Journalism, creative writing
353	Information, arts, and entertainment professions: Managers in the press, publishing, audiovisual, and entertainment industries
354	Information, arts, and entertainment professions: Artists
371	Executive managers (large companies with 500 or more employees)
372	Administrative and financial specialists
373	Other administrative and financial managers
374	Commercial administration, commercial function
375	Administrative and commercial managers in advertising, public relations, communication
376	Administrative and commercial managers in banking, insurance and real estate, financial market managers
377	Administrative and commercial managers in hotels, restaurants
380	Engineers and technical managers: Executive staff (large companies)
381	Engineers and technical managers in agriculture, water, and forestry
382	Engineers and technical managers in building, public works
383	Engineers and technical managers in electricity, electronics
384	Engineers and technical managers in mechanics, metalworking
385	Engineers and technical managers in processing industries (agri-food, chemicals, metallurgy, and heavy materials)
386	Engineers and technical managers in other industries (printing, soft materials, furniture and wood, energy)
387	Engineers and technical managers in related production functions: Industrial purchasing, logistics, methods, quality control, maintenance (excluding IT), environment, etc.
388	Engineers and technical managers in IT, telecommunications
389	Engineers and technical managers in transport (excluding logistics)
4 - Intermediate occupations	
421	School teachers and similar staff in nurseries and elementary schools
422	Other school teachers and similar staff
423	School teachers and similar staff in continuing education
424	School teachers and similar staff in sports education and professional sport
425	School teachers and similar staff in documentation and heritage
431	Intermediate health and social work professions: Nurses, midwives, and related professions
432	Intermediate health and social work professions: Physiotherapists and rehabilitation specialists
433	Intermediate health and social work professions: Medical technicians and medical equipment specialists
434	Intermediate health and social work professions: Specialists in socio-educational intervention
435	Intermediate health and social work professions: Specialists in socio-cultural and leisure activities
441	Clergy and religious
451	Intermediate administrative professions in the civil service: Administrative staff
452	Intermediate administrative professions in the civil service: Police and military personnel
461	Intermediate administrative and commercial professions: Higher-level secretaries, corporate administrative services supervisors
462	Intermediate administrative and commercial professions: Sales shop managers, purchasing function, sales administration
463	Intermediate administrative and commercial professions: Sales force technicians, sales representatives
464	Intermediate administrative and commercial professions: Advertising, public relations, communication
465	Intermediate administrative and commercial professions: Artistic expression, entertainment
466	Intermediate administrative and commercial professions: Transport, tourism
467	Intermediate administrative and commercial professions: Technical and commercial services in banking, insurance, and social security organisations
468	Intermediate administrative and commercial professions: Hotels and restaurants
471	Technical staff in agriculture, water, and forestry
472	Technical staff in building, public works
473	Technical staff in electricity, electromechanics, and electronics
474	Technical staff in mechanics, metalworking
475	Technical staff in processing industries (agri-food, chemicals, metallurgy and heavy materials)
476	Technical staff in other industries (printing, soft materials, furniture and wood)
477	Technical staff in related production functions: logistics, maintenance (excluding IT), environment
478	Technical staff in informatics and telecommunications
479	Technical staff: Experts and public research technicians
480	Foremen and supervisors (excluding administrative supervisors) in agriculture, water and forestry, maritime professions
481	Foremen and supervisors (excluding administrative supervisors) in building, public works
482	Foremen and supervisors (excluding administrative supervisors) in electricity, electronics
483	Foremen and supervisors (excluding administrative supervisors) in mechanical engineering, metalworking
484	Foremen and supervisors (excluding administrative supervisors) in processing industries (agri-food, chemicals, metallurgy, heavy materials)
485	Foremen and supervisors (excluding administrative supervisors) in other industries (printing, flexible materials, furniture and wood, energy, water)
486	Foremen and supervisors (excluding administrative supervisors) in maintenance, new works
487	Foremen and supervisors (excluding administrative supervisors) in warehousing, storage, and handling
488	Foremen and supervisors (excluding administrative supervisors) in restaurants
5 - Employees	
521	Civil servants and service agents in the public service: Post Office and France Télécom employees (public status)
522	Civil servants and service agents in the public service: Specialised tax, treasury, and customs agents
523	Civil servants and service agents in the public service: Administrative assistants in the public service (including teaching)
524	Civil servants and service agents in the public service: Administrative officers in the public service (including teaching)
525	Civil servants and service agents in the public service: Service agents in the public service (including teaching) and similar (hospitals and private clinics)
526	Civil servants and service agents in the public service: Caregivers and related professions in the public or private sector
531	Police and military: Police and prison guards

532	Police and military: Army
533	Police and military: Firefighters, nature, and heritage conservation officers
534	Police and military: Security and surveillance officers
541	Company administrative employees in reception and information
542	Company administrative employees in secretarial and typing duties
543	Company administrative employees in accounting and administrative services
544	Company administrative employees in information technology
545	Company administrative employees in banking, insurance, and social security technical and commercial services
546	Company administrative employees in transport, tourism
551	Commercial employees in procurement, labelling
552	Commercial employees: Cashiers
553	Commercial employees in non-specialised sales
554	Commercial employees in specialised sales (by field)
555	Commercial employees in mail order, telesales
556	Commercial employees in wholesale of capital and intermediate goods
561	Direct services to individuals in hotels, cafés, restaurants
562	Direct services to individuals in personal care services
563	Direct services to individuals in social intervention and domestic help
564	Direct services to individuals (Miscellaneous)
6 - Manual workers	
621	Skilled industrial workers in building, public works, quarries, extraction
622	Skilled industrial workers in electricity, electronics
623	Skilled industrial workers in metal working
624	Skilled industrial workers in mechanics, mechanical engineering
625	Skilled industrial workers in food, chemical, and related industries (chemicals, plastics, pharmaceuticals, water, energy)
626	Skilled industrial workers in processing industries (metallurgy, glass production, building materials)
627	Skilled industrial workers in other industries (textiles, haberdashery, clothing, industrial leatherworking, woodworking, furniture, paper and cardboard, printing)
628	Skilled industrial workers in maintenance, industrial equipment maintenance, adjustment, laboratory work
631	Skilled craftsmen in gardening
632	Skilled craftsmen in building
633	Skilled craftsmen in electricity, electronics
634	Skilled craftsmen in mechanics, metalworking
635	Skilled craftsmen in textiles, clothing, leather
636	Skilled craftsmen in food, catering
637	Skilled craftsmen (miscellaneous)
641	Lorry drivers
642	Taxi and private car drivers
643	Delivery drivers, couriers
644	Drivers related to environmental purposes (waste collection, sanitation, cleaning)
651	Skilled workers in handling, warehousing, and transport: Conductors of heavy lifting and shunting equipment
652	Skilled workers in handling, warehousing, and transport: Skilled handling workers, forklift truck drivers, forklift truck operators
653	Skilled workers in handling, warehousing, and transport: Warehousemen
654	Skilled workers in handling, warehousing, and transport: Qualified drivers of guided transport vehicles
655	Skilled workers in handling, warehousing, and transport: Other qualified transport workers
656	Skilled workers in handling, warehousing, and transport: Merchant seamen, captains, and helmsmen on inland waterways
671	Unskilled industrial workers in building, public works, quarries, extraction
672	Unskilled industrial workers in electricity, electronics
673	Unskilled industrial workers in forging, metalworking, mechanics
674	Unskilled industrial workers in processing industries (chemicals, pharmaceuticals, plastics, food processing, metal processing, glass, building materials)
675	Unskilled industrial workers in other industries (textiles, clothing, leather, wood, furniture, paper and cardboard, printing, press, publishing)
676	Unskilled industrial workers in handling, sorting, packaging, shipping, miscellaneous
681	Unskilled craftsmen in building
682	Unskilled craftsmen in mechanics
683	Unskilled craftsmen in power supply
684	Unskilled craftsmen in cleaning, sanitation, waste treatment
685	Other unskilled craftsmen
691	Agricultural workers in agriculture, forestry
692	Agricultural workers in fishing, aquaculture