2021

WorkSafe New Zealand Scoping Report:

Exploring the development of a New Zealand occupational exposure measurement database

DR AMANDA ENG

Research Centre for Hauora and Health, Massey University in collaboration with WorkSafe New Zealand and the New Zealand Occupational Hygiene Society (NZOHS)

13 December 2021

Executive Summary

The lack of information on the distribution and level of workplace exposures in New Zealand limits the ability to develop, target, and evaluate interventions to reduce harmful exposures. WorkSafe contracted the Research Centre for Hauora and Health (a recent merger between the Centre for Public Health Research and the Research Centre for Māori Health and Development) at Massey University to explore the development of a New Zealand occupational exposure measurement (OEM) database in collaboration with WorkSafe and the New Zealand Occupational Hygiene Society (NZOHS). An OEM database is a centralised repository of objectively collected workplace exposure measurements. This scoping report includes:

- A description of the advantages and disadvantages of OEM databases;
- An overview of international OEM databases, including a literature review and a survey of database managers/contacts to explore their experiences with key international databases;
- The advantages and disadvantages of an OEM database for New Zealand in the context of current legislation and key players who collect exposure monitoring data;
- Stakeholder consultation, through discussion at a NZOHS members meeting and an online survey of NZOHS members, to explore the willingness of occupational hygiene providers and industry in New Zealand to contribute to an OEM database;
- An overview of issues that need to be considered prior to developing an OEM database such as database host, sustainability, intellectual property, privacy, user access, and consent;
- Recommendations for whether a New Zealand OEM database should be established as well as specific recommendations for potential implementation;
- An outline of options and possible next steps required to pilot and establish a database.

The systematic collection of OEM data is recognised as the most effective way to target interventions to reduce harmful exposure with the centralisation of this data in a database being critical to the development of effective work-related disease prevention interventions. However, the shift in responsibility for collecting exposure measurement data from regulatory bodies to employers in New Zealand and in many jurisdictions around the world has meant that up-to-date OEM data are harder to access. Section 1 of this report discusses the importance of occupational exposure measurement data as well as the advantages and disadvantages of OEM databases, with some data use examples from overseas. International OEM databases have been used for a wide range of purposes including the development and targeting of policy and prevention activities, assessing trends in exposure levels over time, informing the review of occupational exposure limits, and for exposure assessment in epidemiological studies investigating the causes of work-related disease.

Section 2 of the report presents a literature review and summary of international OEM databases. Twelve major databases from 11 countries that were considered of relevance to New Zealand were identified. The database managers/contacts for each of these databases were approached and asked to complete a questionnaire, which was completed for all but two countries. Section 2 also describes some examples of exposure-specific databases and multi-country (mostly European Union) collaborative databases.

For many of the large successful national databases, the collection of exposure monitoring is a mandatory requirement. Germany's MEGA and France's COLCHIC are the largest data sources and both operate under national statutory insurance systems. They have been operating for 48 and 33 years, respectively and continue to add new data. The MEGA database is part of a wider quality management system where all processes are highly standardised and data flow is direct from the worksite to the database. In France, in addition to COLCHIC, a database of mandatory compliance measurements called SCOLA was established in 2007. The two French databases have different objectives but share the same software, extensive ancillary information, standardised coding, and source population. However, access to these three databases is restricted to parties outside of the national insurers. In contrast, in the United States (U.S.), the data from Occupational Safety and Health Administration's (OSHA) enforcement inspections and consultation visits were previously centralised in the Integrated Management Information System (IMIS), which could be accessed by external parties via the Freedom of Information Act. As a result, a considerable number of scientific papers have been published from the IMIS data. The OSHA Information System (OIS) replaced the IMIS around 2014; however, it is unclear whether the level of access for external parties is the same as for the IMIS.

Databases from Asia are underrepresented in the international review; however, examples from Singapore and South Korea are presented in the report. In both countries, the reporting of OEM by companies to the regulator is either mandatory or recommended. In South Korea, it is mandatory for companies to conduct exposure monitoring annually or bi-annually if hazardous materials are present and to report the results via an electronic collection system to the regulator's Work Environment Measurement Database (established in 2000). In Singapore, the Ministry of Manpower (MOM) established the National Database for Noise and Chemical Exposure in the early 1990s; however, it was redeveloped in 2010 to enable companies to submit their data online. Companies with workers likely to be exposed to excessive noise or toxic substances are required by law to conduct exposure monitoring and to retain the results for inspection. However, only a fraction of workplaces with potential exposure report their results to the database.

In countries/jurisdictions where regulatory agencies are no longer responsible for exposure monitoring such as the United Kingdom (U.K.) and in most Canadian states, the decline in occupational exposure measurements being collected and centralised is a significant challenge. The U.K.'s National Exposure Database (NEDB) was established in 1986 initially as a centralised repository of hygiene measurements for airborne substances collected by the Health and Safety Executive (HSE) occupational hygiene inspectors. The NEDB is still operational; however, data collection appears to have declined significantly over the past two decades due to operational changes at HSE, which changed the focus to asking companies to conduct exposure monitoring themselves as well as the increasing cost of data collection. The HSE U.K. has expressed an intention to redevelop the NEDB as part of their science plan. The Canadian Workplace Exposure Database (CWED) is the most recently established OEM database and was initially developed for research purposes. The database contains historical information from eight (out of fourteen) Canadian jurisdictions; however, the majority of measurements were taken for compliance purposes from two jurisdictions and the planned inclusion of data from other jurisdictions is hampered by declining data collection by regulatory agencies across Canada.

Whilst the lack of a legislative requirement is a significant limitation, several countries have ongoing databases where the collection and reporting of exposure monitoring data is not mandatory. In Scandinavian countries such as Finland and Norway, national occupational health institutes provide occupational hygiene services to companies and centrally store OEM data. In Finland, the OEM database has evolved into a laboratory information management system (LIMS) and in Norway, the long-running EXPO database was supplemented by the introduction of EXPO Online in 2017, which is a voluntary online registration service open to all companies in Norway. The representativeness of these databases is unclear at present; however, there are ongoing efforts in Norway to increase the representation of all industries. In Finland, the national occupational health institute maintain several databases and information sources and has a strong history of combining data sources to provide a more comprehensive picture of occupational exposure.

There are also several other databases/projects operating without a regulatory requirement which serve as useful examples for New Zealand, although the representativeness of these systems is unclear. The sustainability and further development of the CWED is the focus of the CWED 2.0 project, which includes improving sustainability and data access and encouraging the deposition of new data. Other useful examples include the aforementioned EXPO Online from Norway, ongoing efforts to centralise Workers Compensation data in the U.S., and the industry-led Industrial Minerals Association-Dust Monitoring Programme (IMA-DMP), which consists of prospectively collected measurements from ~500 mineral companies operating >700 mines and quarries and 750 plants throughout Europe.

The common components of successful international databases include standardised processes, quality control systems/reviews, efficient data flow into the database (including laboratory management information systems), reporting of results (i.e. value-added benefits) to data generators and users, and regular use of the data for both policy and research. However, OEM databases also have important limitations. Whilst the measurement of occupational exposure is objective, the selection of workers and workplaces that are sampled is often subjective or non-random, which makes exposure assessment for the general working population difficult. There has been much debate about the representativeness of data from OEM databases, especially compliance and enforcement systems. Other important limitations include lack of coverage for some occupations and industries, inconsistent data quality, data acquisition challenges, and the inconsistent collection and analysis of potential exposure determinants (i.e. contextual information).

Despite these limitations, the general consensus is that exposure measurement databases are valuable provided that effort is put into understanding and characterising their limitations and biases. Many of the international databases are still operating and the most recent effort to create a national centralised database was in 2008 with the development of the CWED. In addition, the U.K.'s NEDB is potentially being redeveloped, all of which reinforces the continued value of OEM databases in today's modern society.

Section 3 includes a description of the occupational health and safety context in New Zealand and the major players who collect exposure measurement data. There is currently a glaring lack of collective knowledge on the actual levels of workplace exposures to inform policy and evidence-based solutions and no legislative requirements for any Government agency, including WorkSafe, to measure or collect exposure data. Section 3 also outlines the benefits of an OEM database for New Zealand, including uses for informing targeted interventions, assessing trends over time, providing evidence for compliance assessment, allowing for policy impact assessment, helping evaluate compliance with and inform reviews of Workplace Exposure Standards as well as benefits for industry benchmarking, research, and informing occupational hygiene assessments. Exposure monitoring in New Zealand is largely carried out by private industrial hygienists and researchers and thus an exposure measurement database developed for New Zealand would involve centralising data collected and stored by industry and private occupational hygienists and academic institutions.

The willingness of occupational hygiene providers and industry in New Zealand to contribute to a database was explored through discussion at a NZOHS members meeting and an online data holders' survey. It was requested that one occupational hygienist per organisation complete the survey. The survey included 21 NZOHS members and the overall view was positive. All respondents thought that a database should be set up in New Zealand and the average willingness to contribute rating was 7.1 (on a scale from 1-10 with 10=very willing). The majority supported for the database to be open access. The main barriers that were expressed were about privacy concerns, fear of WorkSafe using the data to prosecute companies, and time available to anonymise and input data.

Section 4 of the report discusses the potential issues/barriers for a New Zealand OEM database including data acquisition, representativeness, data quality, privacy and intellectual property, user access, consent, funding and sustainability, and IT system requirements. For each issue, experiences from international databases and responses from the data holders' survey are summarised. The potential risks to be considered for each issue are also outlined. The advantages and disadvantages of options for a database host are also presented.

Sections 5 and 6 of the report present the recommendations and options for a New Zealand OEM database. Based on the lack of New Zealand-specific occupational exposure data, successful examples from overseas, and results from the data holders' survey, an occupational exposure measurement database for New Zealand is recommended. A database would be an invaluable resource for a wide variety of stakeholders in New Zealand and ultimately for the prevention of work-related harm. Several factors such as New Zealand's smaller size (i.e. there are not multiple states/jurisdictions), NZOHS as collaborators on this scoping project, and the current momentum for the widespread sharing and analysis of anonymised data managed by Statistics New Zealand's Integrated Data Infrastructure will facilitate the development of a New Zealand OEM database. The recommendation from this report is to establish a database with a focus on prospectively collected measurement data from NZOHS members, and to standardise all processes pertaining to data collection from the outset. The addition of historically collected data from both NZOHS members and researchers can be investigated at a later date. A further recommendation would be to pilot test the database using respirable crystalline silica as an example. Specific recommendations are also made for database development, governance, data acquisition, data management and collection, data access, and future use and sustainability. An outline of the steps required to pilot and establish a database, including general indicative costs are presented and some future opportunities and considerations are briefly discussed.

Table of Contents

Background and scope

The lack of information on the distribution, frequency, and level of workplace exposures in New Zealand limits the ability of WorkSafe to develop and target interventions to reduce exposures as well as evaluate the impact of prevention activities. WorkSafe contracted the Research Centre for Hauora and Health (a recent merger between the Centre for Public Health Research and the Research Centre for Māori Health and Development), Massey University to explore the development of a New Zealand occupational exposure measurement (OEM) database in a scoping report, including:

- A description of the advantages and disadvantages of OEM databases;
- An overview of international OEM databases, including a literature review and a survey of database managers/contacts to explore their experiences with key international databases;
- The advantages and disadvantages of an OEM database for New Zealand in the context of current legislation and key players who collect exposure monitoring data;
- Stakeholder consultation, through discussion at a NZOHS members meeting and an online survey of NZOHS members, to explore the willingness of occupational hygiene providers and industry in New Zealand to contribute to an OEM database;
- An overview of issues that need to be considered prior to developing an OEM database such as database host, sustainability, intellectual property, privacy, user access, and consent;
- Recommendations for whether a New Zealand OEM database should be established as well as specific recommendations for potential implementation;
- An outline of options and possible next steps required to pilot and establish a database.

The project has been conducted in collaboration with WorkSafe and the New Zealand Occupational Hygiene Society (NZOHS).

This scoping report includes six sections:

Section 1: Introduction to occupational exposure measurement databases
This section includes a discussion of the importance of occupational exposure measurement data as well as the advantages and disadvantages of occupational exposure measurement databases, with some data use examples from international databases.

Section 2: Literature review and summary of international occupational exposure measurement databases

The review of international databases was based on a variety of data sources including the previous National Occupational Health and Safety Advisory Committee (NOHSAC) 2006 report: '*International Review of Surveillance and Control of Workplace Exposures*', formal scientific literature, grey literature (including reports and database websites), and informal contacts. The databases included as part of this review do not represent an exhaustive and systematic list of databases from around the world. The focus is primarily on large-scale databases that were cited in English-language literature with an overrepresentation of European databases and an underrepresentation of databases from Asia. Twelve major databases from eleven countries were identified for this report and the relevant database contacts were approached to complete a questionnaire. The majority of database managers/contacts completed the questionnaire, with the exception of Finland (partially completed) and Denmark, where the database manager was unable to locate any former employees who could answer questions about the now defunct database. In addition, regular data users also completed the survey for databases from the United States (U.S.), France, and Switzerland. The common components of successful databases and the common challenges are also summarised.

Section 3: A New Zealand occupational exposure measurement database

This section includes a description of the occupational health and safety context in New Zealand and a brief description of private industrial hygienists and researchers as the major players who collect exposure measurement data. The benefits of a New Zealand database for specific stakeholders are presented as well as a what an OEM database would involve for New Zealand. The likely willingness of data holders to contribute to a database was gauged from a discussion at a NZOHS members meeting and an anonymous online survey that was sent to all members of the NZOHS via the fortnightly newsletter during February to March 2021.

Section 4: Overview of major barriers/issues

This section presents an overview of barriers/issues relevant to the establishment of a New Zealand OEM database including data acquisition, representativeness, data quality, privacy and intellectual property, user access, consent, cost/funding and sustainability, database host, and information technology (IT) system. For each issue, an overview of the problem is outlined as well as a summary from the international review, a summary from the NZOHS data holders' survey, and an outline of the likely risks.

Section 5: Recommendations

An overall recommendation is made as well as specific recommendations for database development, governance, data acquisition, data management and collection, data access, and future use and sustainability.

Section 6: Options for a New Zealand occupational exposure measurement database

This section describes options for a New Zealand OEM database including for the timing of data collection (i.e. prospective vs. historical) and priority vs. all exposures/industries. Options for a pilot database are also presented. This section also outlines the steps required to pilot and establish a database as well as general indicative costs to consider. Finally, this section briefly considers some future opportunities and considerations.

1. Introduction to occupational exposure measurement databases

Occupational exposure surveillance is defined as "*the ongoing and systematic collection, analysis, and interpretation of data related to occupational exposures, and the use of this information for prevention and control purposes*".[1] The benefits of occupational exposure surveillance have been described extensively.[1, 2] Briefly, the comprehensive collection of occupational exposure data allows for a greater understanding of the distribution, frequency, and level of exposures which is essential for assessing and managing health risks to workers and for the effective prevention of work-related disease and injury.[3]

Occupational exposure databases are one method of exposure surveillance and can be defined as "*the information needed to prospectively or retrospectively define worker exposures to hazards in the workplace and the circumstances that give rise to the exposure*".[4] More specifically, an occupational exposure measurement (OEM) database is a centralised repository of objectively collected workplace measurements and it is this type of database that is the focus of this report. Other types of occupational exposure surveillance methods include surveys of working conditions, observational surveys, registers of exposed workers, substance registers, and expert information systems such as a Job Exposure Matrix (JEM);[1] however these methods will not be discussed further.

1.1 Why do we need exposure measurement data?

Occupational exposure measurement data provides information on the actual level of worker exposures, which are often compared to compliance threshold levels (i.e. occupational exposure limits (OELs)) set by local regulatory bodies. The OELs are based on levels at which adverse health effects may occur, although epidemiological studies have observed adverse health effects at levels below OELs. The systematic collection of information on actual levels of exposure is the most effective way to target interventions to reduce exposure and, in turn, prevent work-related disease.[5] In a large number of jurisdictions around the world (including New Zealand), the responsibility for collecting exposure measurement data has shifted from regulatory bodies to employers. As a consequence, access to up-to-date exposure data for policymakers and researchers has significantly declined, which has been described as a "*negative development for occupational exposure assessment science and beyond*".[6]

In recent years, a 'hygiene without numbers' approach has emerged[7] with an increased emphasis on alternative methods to collecting exposure data such as control banding (a generic risk assessment method that prescribes a control method for a range or 'band' of hazards/exposures)[8] and advanced statistical modelling (e.g. Bayesian methods). However, Hall et al.[6] argue that alternative methods such as control banding tend to have an enforcement focus and are thus inadequate for other uses of the data. In a 2016 critique of the 'hygiene without numbers' approach, Kromhout argued that actual exposure measurements are required to develop meaningful interventions to reduce exposures and to "*confidently monitor progress in preventing hazards from becoming risks*".[7] Such data are also required to account for intrinsic (e.g. temporal and personal) variability in exposure and especially in situations where exposure is not obvious (e.g. respirable crystalline silica (RCS)).[7]

Exposure measurement data has also been used to develop exposure assessment tools. For example, under the Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH) legislation, companies that manufacture or import high volumes of potentially dangerous chemicals in the European Union (EU) are required to estimate exposure levels under different scenarios for risk assessment reporting. The Advanced REACH tool (ART) is a mechanistic model (currently only calibrated for inhalable dust, vapours, and mists) which was developed using a Bayesian modelling approach combined with exposure measurement data. The Stoffenmanager model is another exposure assessment tool used for REACH developed in the Netherlands.[9] In addition to informing their development, occupational exposure measurement data are essential to the continued accuracy and validation of these tools.[10] The assignment of workers to homogenous exposure groups for exposure assessment will also continue to require validation by quantitative exposure measurements.[11, 12]

1.2 What are the advantages of exposure measurement databases?

As described above, an OEM database is a centralised repository of objectively collected workplace measurements. The benefits of OEM databases have been discussed previously in the scientific literature[13-16] and are summarised in Table 1. Several scientific conferences have also been dedicated to the topic, including a 1995 conference on occupational exposure databases.[17] A workshop from this conference identified the top five users as industrial hygienists, epidemiologists, occupational medicine professionals, employers, and regulators[18] and therefore the type of stakeholder that could potentially benefit from an OEM database is wide-ranging (discussed further in section 3.2.2). The visibility of a national exposure database could also increase general awareness of occupational hazards and work-related health in the general population.[19]

Large-scale international OEM databases have been operating since the 1970s and 80s and the benefits and numerous data use examples are described in Section 2. Whilst they have been a part of the exposure surveillance landscape for decades now, many of the databases are still operating today, which is a testament to their value and reinforces that an exposure database is not an outdated surveillance tool. In addition, the most recent effort to create a national centralised database was in 2008 with the development of the Canadian Workplace Exposure database (CWED). The database was created out of recognition that the centralisation of data is critical to the development of work-related disease prevention activities.[6] There are ongoing efforts to further develop the CWED (see section 2).

The advantages for both primary prevention and surveillance and research are described in Table 1 below:

Table 1: Advantages of occupational exposure measurement databases

Primary prevention:	**Surveillance and research:**
OEM Databases can:	
Set priorities for prevention and inform the development and targeting of policy and prevention activities by: • Identifying 'high-risk' occupations • Identifying 'high-risk' industries • Estimating the number of workers exposed • Identifying workers at higher risk of exposure, including situations of overexposure • Evaluating compliance with occupational exposure limits *Example 1: an analysis of France's COLCHIC database identified that occupational formaldehyde exposure was most common in human health activities and in the manufacture of wood products;*[20] *Example 2: the U.S. IMIS data was used to develop an algorithm to estimate the proportion of workers exposed to RCS. The industry with the highest proportion of workers exposed to the recommended exposure limit level or above was the 'cut stone and stone products' industry*[21]	Provide a benchmark of current and historical exposures[6]
Present broad 'portraits' of specific occupational exposures, including priority hazards such as RCS or classes of agents such as carcinogens *Example: the CWED can be used to identify which occupations are exposed to carcinogens; estimate how many workers are exposed; where exposure happens; as well as levels of exposure to carcinogens.*[22]	Highlight data gaps for future policy and research

Evaluate impacts of exposure control regulations and other policies *Example: analysis of exposure levels of mineral fibres from France's COLCHIC during 1986-2004 found that levels were broadly influenced by changes in exposure limit values*[23]	Inform exposure assessment for epidemiological research e.g. providing data for: • Exposure levels at which adverse health effects may occur • Dose-response associations • Developing exposure prediction models;[24] • Developing information systems and tools e.g. JEMs, the Advanced REACH tool[10]
Facilitate the characterisation of baseline exposures and assess trends in exposure levels over time *Example 1: a study of U.S. IMIS data observed a decline in occupational lead exposures for U.S. industry over the period 1979-1997, but not for the construction sector;*[25] *Example 2: an analysis of CWED data observed an annual decrease in wood dust exposure of 3% p.a. between 1978-2008*[26]	Provide data to estimate the current and future disease burden of occupational exposures
Inform the review of occupational exposure limits	Provide quantitative data for risk assessment
Inform occupational hygiene assessments	Encourage the standardisation of data and thus improve comparability of data from different sources and between countries
Provide a knowledge base from which to educate industry and the wider public	Encourage the preservation of data which might otherwise be lost
Improve exposure information for worker compensation	Encourage the digitisation of data and therefore ease of use and reporting

As mentioned above, the collection of actual exposure level data is the most effective way to develop and target interventions to reduce workplace exposures and the advantages outlined in Table 1 demonstrate that the centralisation of this data has wide-ranging uses for primary prevention, surveillance, and research.

The content and quality of data collected for OEM databases can vary widely and the database attributes for high and limited utility have been previously outlined by Matonski et al.[27] (see Table 2).

Table 2: Utility of exposure databases

Attribute	**High utility**	**Utility limited to specific case**
Availability	All users	Restricted to owner organisation
Access	Digital files with user friendly search functions	Hard copy only
Level of aggregation	Individual data points	Summary statistics only
Toxicants	Multiple	Single
Data presentation	All measured values	Out of compliance only
Geographic coverage	All measured values	Haphazard samples
Method of measurement	Consistent	Not standardised
Time coverage	Historic and ongoing	One time only

Source: adapted from Matonski et al 1992[27]

1.3 What are the disadvantages of exposure databases?

The limitations of OEM databases have also been described previously[14, 28] and the general limitations are briefly summarised below. The shortcomings for each of the major international databases are discussed further and summarised in Section 2.

Resource-intensive

In general, the collection of exposure measurements is resource-intensive, particularly given that a sufficient number of measurements are required to provide a representative picture of occupational exposure. OEM databases are expensive to set-up and maintain and the ascertainment of data is generally difficult and time-consuming e.g. challenges include obtaining consent, different data storage types, and lack of data standardisation.[6, 29]

Representativeness of exposure data

A major limitation of OEM databases is that the extent to which the data represents exposure levels in the wider workforce is often unknown. The data that is collected is influenced by ease of access (e.g. which regulatory agency or companies are willing to share data) as well as the cultural, legal, and industrial structures of the host country. Many of the international systems are based on compliance data collected from non-randomly selected workplaces and are therefore more likely to capture poorer performing companies and consequently higher exposure levels. There are also a vast number of factors that can influence exposure levels including sampling strategies, sampling methods, the competency of those measuring exposures, and other contextual information all of which increases the risk of invalid claims being made from the data. The risk of misinterpretation is one of the reasons that access is restricted to owner organisations for some of the large databases such as Germany's MEGA and France's COLCHIC (discussed further in Section 2). The range of occupations and industries included in a system also depends on priorities or coverage (in the case of insurance) of the host organisation. In addition, databases where the majority of the data was collected decades ago may not be fully relevant to current occupational health and safety (OHS) contexts.

Inconsistent data quality

The lack of quality control is another of the major problems of OEM databases[30] and the aggregation of data can be difficult and time-consuming depending on data quality, which can include inconsistencies in type of variables included/reported, missing data, and in coding of the data.

2. Review of international exposure measurement databases

Literature review methodology

The scope of the report was to review OEM databases from countries with OHS systems that may be relevant to New Zealand. Therefore, the databases included as part of this review do not represent an exhaustive and systematic list of databases from around the world. The focus is primarily on large-scale databases that were cited in English-language literature with an overrepresentation of European databases and an underrepresentation of databases from Asia.

Unless stated specifically, the review of international databases is based on a variety of data sources:

a. The National Occupational Health and Safety Advisory Committee (NOHSAC) 2006 report: '*International Review of Surveillance and Controls of Workplace Exposures*'. The report included a survey of exposure surveillance system managers[1]
b. Formal scientific literature (e.g. using PubMed)
c. Grey literature (including reports and database websites)
d. Informal contacts

For the major databases that were identified, the relevant database contacts were approached via email and asked to complete a questionnaire (December 2020-April 2021), which was adapted from the 2006 NOHSAC report[1] (Appendix 1). Information from literature/web-based searches was combined with responses from the questionnaire survey.

The availability of English-language publications and reports for each database varied widely (and thus the discussion of benefits and limitations). For example, a large number of scientific publications are available for the U.S. Integrated Management Information System (IMIS) but the information for South Korea's Work Environment Measurement Database is based on the database contact survey only.

Results

This review identified 12 major overseas databases from 11 countries (France has two major databases), which are listed in Table 3. The U.S. IMIS has been replaced by the Occupational Safety and Health Administration (OSHA) Information System (OIS) and this is described in a separate row as the database survey was completed for the OIS system. The questionnaire was completed by the majority of database contacts (also listed in Table 3), with the exception of Finland (partially completed) and Denmark, where the database contact was unable to locate any former employees who could answer questions about the now defunct database. In addition, regular data users also completed the survey for databases from the U.S., France, and Switzerland.

Table 3: List of reviewed international databases								
#	**Country**	**Name of database**	**Owner of database**	**Is the database still operating?**	**Survey contact**	**Survey contact's organisation**	**Survey completed**	**Additional survey contact**
1	Canada	Canadian Workplace Exposure Database (CWED)	Hosted by the University of British Columbia	Yes	A/Prof Hugh Davies (HD)	The University of British Columbia	Yes	
2	United States of America	Intergrated Management Information System (IMIS)	Occupational Safety and Health Administration (OSHA)	Replaced by OIS ~2014	-	-	Yes for OIS	Jérôme Lavoué (JL), Université de Montréal [data user]
2	United States of America	OSHA Information System (OIS)	Occupational Safety and Health Administration (OSHA)	Replaced IMIS ~2014	Gina Scott (and multiple colleagues)	OSHA		
3	United Kingdom	National Exposure Database (NEDB)	Health and Safety Executive (HSE) UK	Yes but data collection has stalled	Peter Baldwin (PB)	Health and Safety Executive (HSE) UK	Yes	
4	Germany	Measurement Data on Exposure to Hazardous Substances in the Workplace (MEGA)	Institute for Occupational Safety and Health of the German Social Accident Insurance (IFA)	Yes	Dr Dorothea Koppisch (DK)	Section 1.3 Exposure Monitoring - MGU, IFA	Yes	
5	France	COLCHIC	The French National Institute of Safety and Health (INRS)	Yes	Gautier Mater (GM), Head of the ERE Laboratory	Pollutant metrology, INRS	Yes	Jérôme Lavoué (JL), Université de Montréal [data user]
5	France	SCOLA	The French National Institute of Safety and Health (INRS)	Yes				
6	Finland	Finnish Database of Occupational Exposure Measurements	Finnish Institute of Occupational Health (FIOH)	Now Laboratory Information System (LIMS; used comprehensively since 2010)	Dr Tapani Tuomi (TT), Chief Specialist	FIOH	Partially	
7	Denmark	ATABAS	Danish National Institute of Occupational Health (AMI)	No	Lars Andrup (LA), Head of Research, Chemical Working Environment	National Research Center for the Working Environment	No; survey contact was unable to locate someone with knowledge of ATABAS	
8	Norway	EXPO	National Institute of Occupational Health (STAMI)	Yes + Expo Online since 2017	Ragnhild B. Strand Ostrem (RO), Senior engineer	Chemical Work Environment, STAMI	Yes	
9	South Korea	Work Environment Measurement Database	Occupational Safety and Health Research Institute (OSHRI) of the Korea Occupational Safety and Health Agency (KOSHA)	Yes	Dr Seunghyun Park (SP), Director	Work Environment Research Bureau, OSHRI, KOSHA	Yes	
10	Singapore	National Database for Noise and Chemical Exposure	Ministry of Manpower	Yes	Evelyn Koh (EK), Senior Assistant Director (Occupational Hygiene) (and multiple colleagues)	Specialist Department, Occupational Safety and Health Division, Ministry of Manpower	Yes	
11	Switzerland	SUVA	Swiss Insurance Fund (SUVA)	Yes	Patrick Steinle (PS), Head of Analytics Section	SUVA	Yes	Nenad Savic (NS), Unisante [data user]
Exposure-specific (examples)								
	France	Evalutil						
	Denmark	CODUST						
EU Collaborations								
	Finland/EU	IARC Registry of Industrial Hygiene measurements in the Pulp and Paper industry	FIOH and University of British Columbia					
	EU	IMA-DMP: Industrial Minerals Sector in Europe	Utrecht University					
	EU	WOODEX (International information system on occupational exposure to wood dust)						
	EU	Asphalt Worker Exposure (AWE) database						
	EU	Improved Exposure Assessment for Prospective Cohort Studies and Exposure Control in the Rubber Manufacturing Industry: EXASRUB						
	EU	EXPOSYN: Chromium, RCS, nickel, PAH, asbestos						

The detailed characteristics of these databases are presented in Appendix tables A2.1 to A2.7, including a description, data collection, data uses, resource requirements, data governance, barriers and enablers, and challenges and advice. Table 3 also lists exposure-specific examples including CODUST (Denmark) and Evalutil from France, as well as collaborative databases from multiple European Union (EU) countries. The exposure-specific examples and EU databases are not presented in Appendix tables A2.1 to A2.7 but are described briefly at the end of section 2.

The major international databases (#1-11) are described below:

1. Canada: The Canadian Workplace Exposure Database (CWED)

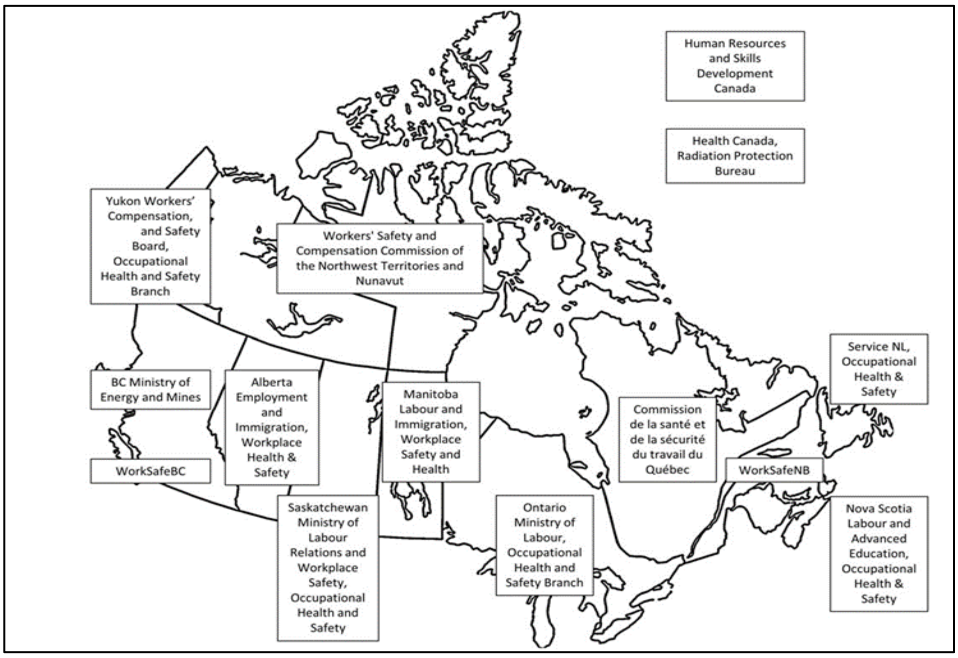
Background and aims:

The Canadian Workplace Exposure Database (CWED) was established in 2008 by the CAREX Canada project (an occupational carcinogen exposure information system) to support the estimation of the prevalence of occupational carcinogen exposure. The development of the database was funded by grants from WorkSafe British Columbia (BC) and the Ministry of Labour in Manitoba with further financial support from CAREX Canada.[26] The goal of the CWED is "*to collect exposure measurements from government agencies, researchers, and other sources to create a centralised database of both current and historical exposure measurements across workplaces in Canada*".[22] The database is currently hosted by the University of BC.

Data collection:

In Canada, there are 14 federal and provincial regulatory agencies responsible for health and safety (see Figure 1), including specialist divisions in some provinces (e.g. the BC Ministry of Energy and Mines). The CWED initially consisted of historical data from existing electronic databases held by WorkSafe BC, the Ontario Ministry of Labour, The National Dose Registry, and Quebec's Institut de Recherche Robert-Sauve en Sante et en Securite du Travail[6] through a data acquisition agreement process in partnership with PopdataBC.

Figure 1: Regulatory agencies and specialist organisations responsible for health and safety in Canada



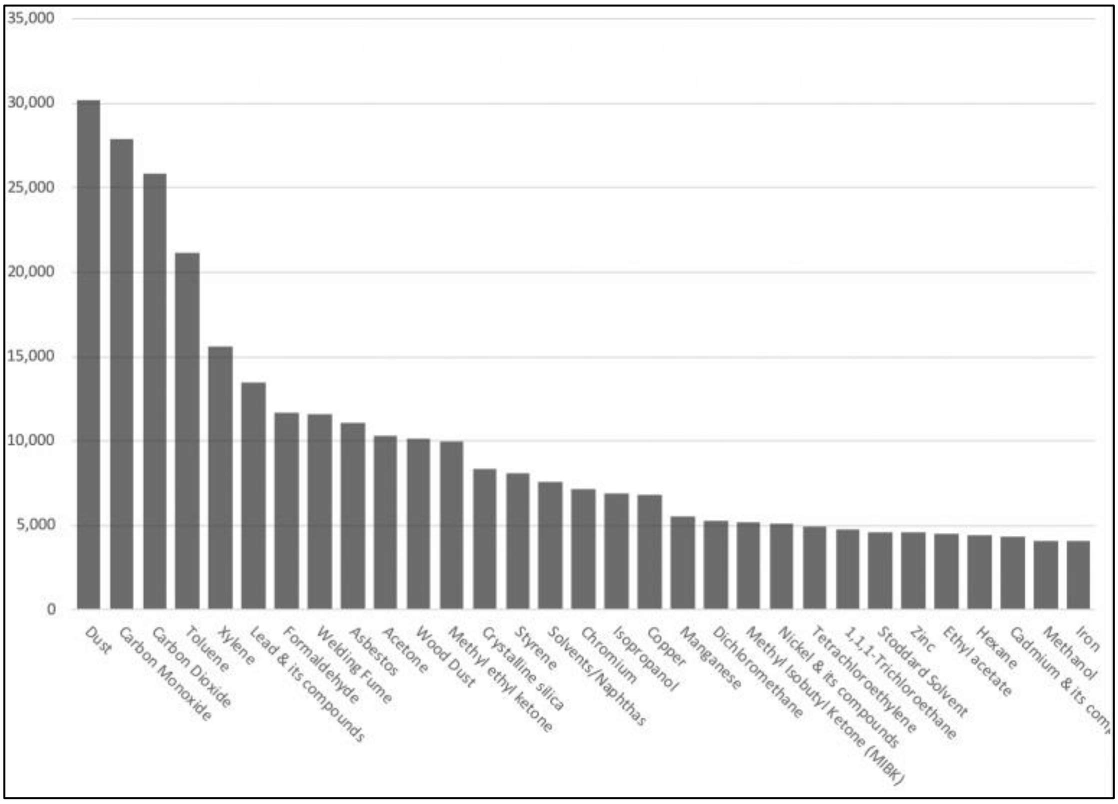
Source: Hall et al. 2014[6]

In 2009, a data holders' survey conducted by CAREX Canada of all relevant agencies identified a downward trend in exposure measurements by regulatory agencies since the 1990s (Quebec is the only large province in Canada that still collects exposure measurements and has an indefinite data retention policy). Of the 15 agencies surveyed, 4 reported virtually no data, 7 used their data for compliance purposes only, and 4 actively used their data for research or reporting purposes.[6] Most agencies were no longer responsible for collecting the majority of exposure measurements in their jurisdiction.[31] The survey also found that only two of fourteen agencies contacted held their data in electronic databases. Many agencies stored their measurement records in paper form or individual electronic reports (e.g. PDFs).

Following the data holders' survey, funding was provided for 2 years from WorkSafe BC to transfer paper-based records into an electronic database for data held by the BC Ministry of Energy and Mines (BCMEM) and the regulatory agencies for Saskatchewan (SK), Yukon, and Manitoba. The data abstraction process for BCMEM and the Government of SK was detailed in a 2014 capacity development report, which estimated that 7.5% more data would be added.[31] Students/a research assistant were employed to abstract data (on-site for SK) supervised by occupational hygienists. By 2014, data had also been added to CWED from the regulatory agencies of Manitoba and Yukon.

The CWED currently contains data across eight federal, provincial, and territorial jurisdictions including Labor Canada.[22] As at 2019, the CWED contains approximately 500,000 measurements of 350 substances,[26] the majority (80%) of which are from British Columbia and Ontario. The data collection period spans 1960-2010 but the majority of measurements were taken mid-1970s to the late 1990s.[26] The most commonly sampled exposures (with >4,000 observations) are presented in Figure 2.

Figure 2: The most commonly sampled exposures in the CWED



Note: the y-axis is the number of measurements *Source: http://cwed.spph.ubc.ca*[22]

Benefits/data use examples:

A considerable amount of effort went into the creation of the CWED; however, its use to date has been limited to informing CAREX Canada as well as a few other specific projects *(Hugh Davies, personal communication, January 2021)*:

The CWED has been used:

- Primarily for the surveillance of exposure to carcinogens in the Canadian population and for setting priorities for prevention.[32] The CWED has been used by CAREX Canada to estimate the number of workers exposed to carcinogens, identify occupations with potential exposure, and identify exposure levels of concern by industry and occupation;[6]

- To contribute to quantitative estimates for the burden of occupational cancer in Canada by industry and occupation;[33]
- To contribute to the CANJEM general population JEM;[26]
- To document historical trends in isocyanate exposure.[34]

***Limitations and challenges*:**

- The majority of the historical data in the CWED was taken by regulatory agencies for compliance purposes and therefore measurements may reflect higher than average exposure levels;[31]
- The available ancillary (i.e. contextual) information varies by regulatory agency.[26] Investigations of potential biases appear to be limited;
- The majority of data is from two jurisdictions (Ontario and BC) and efforts to include data from more jurisdictions have been hampered by the downward trend in regulatory agencies collecting OEM data as well as variable data retention policies e.g. some jurisdictions have destroyed their historical data;
- There is a lot of missing data e.g. the 2014 capacity report described the data extraction process for data held by BCMEM and the Government of SK and reported that only 40% of records had occupation information, 29% for tasks, and 63% for duration of sample collection.[31]
- Sampling methods (e.g. the size fraction of dust particles) are inconsistently recorded[26] and there is limited information available on control methods or personal protective equipment (PPE).

A considerable amount of effort was expended to create the CWED, which involved the extraction of historical OEM data from different regulatory agencies across Canada. This data acquisition process was time-consuming and involved negotiating data access agreements with existing data holders, designing and implementing data access and security protocols, and extensive data cleaning. For several jurisdictions, the transfer of data from paper to electronic records was a 2-year project. For one of the jurisdictions (BCMEM), this involved CWED investigators reviewing archived files and an occupational hygienist reviewing all records.[31] The quality and content of the existing data differed widely and thus minimum data requirements were specified and standardised definitions and protocols to handle missing data were developed *(Hugh Davies, personal communication, January*

2021). The major challenges that were identified included the lack of harmonisation of coding (particularly for non-carcinogens which were not a focus until recently), the difficulties in identifying data sources, and lack of funding *(Hugh Davies, database contact survey, February 2021)*.

Ongoing development:

Work is ongoing to include other provinces and to collect prospective measurements, harmonise data, and encourage the deposition of new data. In 2018, WorkSafe BC granted the University of BC funding for a **CWED 2.0** project to "help sustain and grow" the CWED. The aims of the project are to: "*a) secure the existing CWED; b) make data more widely available to researchers and knowledge users; c) improve data management; e) improve stakeholder knowledge of the database; f) explore database potential; and g) explore sustainable funding models*".[22] Other aims include examining exposures other than carcinogens and fostering the development of new access tools *(Hugh Davies, personal communication, January 2021)*. The website[i] also currently contains a form for potential partners to contribute data.

The long-term goal of CWED 2.0 is to develop it for data uses beyond CAREX Canada and make it more widely available for research and policy *(Hugh Davies, personal communication, January 2021)*. The currently proposed "*data mobilisation phase*" involves building new infrastructure including: 1) renewed governance and stewardship; 2) improved data management (including quality control); and 3) knowledge translation and mobilisation. The CWED will also contribute to two major upcoming research projects which will provide some funding. Part of the mobilisation project involves investigating whether the database should be housed elsewhere to serve its function as a national resource *(Hugh Davies, personal communication, January 2021)*. The report will be made available online by WorkSafe BC but was not publicly available at the time of the completion of this report.

2. United States: IMIS and WC Insurers

In the U.S., several government agencies collect exposure data for regulatory purposes including the Occupational Safety and Health Administration (OSHA), the Mine Safety and Health Administration, and the Department of Defence. OSHA's information system was formerly the Integrated Management Information System (IMIS) but since ~2014 was replaced by the OSHA Information System (OIS).

[i] http://cwed.spph.ubc.ca/

The Center for Disease Control's (CDC) National Institute for Occupational Safety and Health (NIOSH) also collects exposure data through its Health Hazard Evaluation (HHE) Program whereby companies can request an evaluation be carried out at their workplace. There are currently two exposure databases from the HHE Surveys that were developed to provide researchers and other interested stakeholders with measurement results: one for noise (data collected between 1996-2013) and one for lead (work in progress).[35]

Exposure measurements are also collected by Workers' Compensation (WC) insurers; however, these are currently not available in a centralised database. Ongoing efforts to centralise the data are described further below.

The Integrated Management Information System (IMIS)

Background and aims:

Inspectors from OSHA have been collecting occupational hygiene data during their enforcement and consultation visits of worksites since 1972 to verify compliance with Permissible Exposure Limits (PELs).[36] Since 1979, these measurements were entered into the Integrated Management Information System (IMIS; since 1984 for consultation data), which was designed as an information retrieval system for penalty assessment, arbitration, and monetary collection regarding OSHA's inspections.[37]

Data collection:

OSHA's enforcement efforts are conducted through inspections of workplaces whereas consultation visits are conducted at the request of employers seeking assistance on controlling hazards and complying with regulations at their workplaces. A 2018 paper reported that there were more than 1.6 million samples in the IMIS database.[3]

Benefits/data use examples:

The IMIS represented the largest database of occupational exposure measurements in the U.S.[38] and a considerable number of research papers have been published. A 2013 literature review identified 29 publications that had used IMIS data for a variety of purposes, including for exposure surveillance and tracking trends in exposure levels over time.[36]

IMIS data (mainly enforcement data, fewer studies have included consultation data to preserve anonymity) has been used to:

- Provide general portraits of specific exposures including the estimation of exposure levels, the examination of time trends, and the identification of exposure determinants e.g. levels of formaldehyde exposure have been estimated in various industries;[39]
- Identify specific industries with the highest exposure levels, which is useful for surveillance purposes e.g. Henn et al.[37] examined personal airborne lead results for 1979-2008 and observed that whilst the majority of lead samples were taken in the manufacturing sector, the highest lead results came from the construction sector;
- Estimate the proportion of workers exposed (including situations of overexposure) to a specific exposure e.g. Linch et al.[21] developed an algorithm to estimate the proportion of total workers exposed to RCS at various levels above the Recommended Exposure Limit (REL) in 1993 of 0.05mg/m^3. The study also found that the industry with the highest proportion of workers exposed to the REL level or above was the '*cut stone and stone products*' industry;
- Track trends in exposure levels over time e.g. Okun et al.[25] observed a decline in occupational lead exposures for U.S. industry over the period 1979-1997; although this decline was not observed for the construction sector;
- Inform expert information systems e.g. a JEM for beryllium exposure was developed based on over 12,000 personal measurements from the IMIS between 1979-2005.[40]

Limitations and challenges:

The IMIS/OIS is a regulatory compliance database and the selection of workplaces for its enforcement and consultation visits is non-random and therefore may overrepresent workplaces and tasks with potentially higher or lower than average exposures.[41] For example, employers who request consultation visits may have difficulty controlling exposures.[37] In addition, the mining and agricultural sectors were not covered in the IMIS data.[21]

The IMIS database captured data on some ancillary variables such as inspection type and scope, company size, and penalty history, and thus several studies have investigated whether these ancillary variables were associated with exposure levels.[28, 42] A literature review of 29 studies using IMIS data by Lavoue et al.[36] included 16 studies that examined the association between exposure levels and workplace selection variables and concluded that there were no associations of "*appreciable magnitude*". However, a later study of IMIS data by Sarazin et al.[38] examined 77 exposure agents and observed higher odds of having a sample result exceed the threshold limit value (TLV) for follow-up inspection (vs. planned), federal (vs. state), and penalty history suggesting that OSHA's process for selecting worksites for inspection influences the exposure measurement levels captured in IMIS. Exposure levels were also marginally lower for large companies (151+ employees) compared to small (1-35) companies.

The IMIS database did not capture information on important variables such as sampling duration, sampling method, work processes or PPE/engineering controls and several researchers have advocated for the inclusion of such variables to increase the utility of IMIS data for exposure assessment.[39, 43] Furthermore, information on occupation was only available in free text and therefore has been inconsistently analysed.

Another important limitation of the IMIS data was the reporting of non-detectable (ND) measurements to and within the database which includes multiple facets: 1) if a measurement showed no or low detectable exposure, an OSHA inspector may not have submitted the results to the IMIS database;[21, 41] 2) the coding of a ND measurement was classified within the same variable as TWA or STEL and thus it was potentially unclear which of these categories (which may have different limits of detection (LOD)) the ND referred to;[36] 3) the proportion of sample results below the LOD varied widely from 3% (wood dust) to 91% (antimony).[38] It was not possible to distinguish between 'present but not detected' and 'not present' measurement results. In addition, multiple agents have been sometimes measured on the same sample media.[36] As a result, several investigators have examined the probability of exposure over certain thresholds rather than quantitative exposure levels.[40]

Ongoing development:

Since ~2014 (2011 in some regions), the IMIS was replaced by the OSHA Information System (OIS), which is an online platform that is linked to the analytical laboratory.[41] The OIS operates under the Occupational Safety and Health Act 1970 and was designed to better meet the operational and strategic needs of OSHA *(Gina Scott, database contact survey, April 2021)*. It serves as a case management system used to "*manage enforcement safety and health inspections, accident investigations, consultation visits, and compliance assistance/cooperative program and training outreach activities each year*" *(ibid.)*. The OIS contains > 99,000 measurements from enforcement sampling visits and > 65,000 measurements from consultation visits for ~500 exposure agents. According to the OIS database contact, the data collection (although it is unclear whether this refers to the type of variables collected) is the same as for IMIS; but the IMIS data has not been migrated to the OIS *(ibid.)*. It is also unclear whether the data is still easily accessed by external parties.

Workers' Compensation (WC) Insurers

In the U.S. exposure measurements are also collected by WC insurers (state-based and private) in order to evaluate compliance as part of the service to their customers. Taylor Shockey and colleagues from NIOSH (the Center for Workers' Compensation Studies) have been engaged in ongoing efforts to promote the use of WC insurers data which included a 2018 survey of WC insurers that collected information on whether insurers conduct air and noise sampling for their policyholders, data elements, data storage, and data usage.[3] A core list of data elements has also been developed from a review of air and noise insurance survey forms.[44]

The three main challenges to using WC insurers data identified by Taylor Shockey were: 1) standardisation of data: the type and quality of data elements collected by insurers varies which makes aggregation difficult; 2) data storage: a recent American Industrial Hygiene Association (AIHA) '*IH Data Standardisation*' survey of industrial hygienists revealed that a large amount of data are stored in hard copy or individual electronic reports (e.g. PDFs); 3) privacy and intellectual property concerns: U.S. companies have concerns about data privacy and the sharing of potentially commercially sensitive (e.g. trade secret) data *(Taylor Shockey, personal communication, November 2020).*

Other databases:

Chemical Exposure Health Data (CEHD)

In 2010, a dataset of over 1 million personal sample results analysed at OSHA's central laboratory in Salt Lake City – the Chemical Exposure Health Data (CEHD) was made available online[ii]. The central laboratory processes most of the samples collected by federal OSHA inspectors and some of the samples collected by State inspectors. In contrast, the IMIS analytical results were first interpreted by OSHA inspectors and then recorded into the IMIS. As at 2013, the CEHD contained 1,450,836 records from 1984-2009 and an analysis of both data sources using lead as an example reported ~40% overlap with IMIS data.[36] The proportion of NDs was 71% for CEHD-only data compared to 46% for the IMIS-only dataset suggesting differential under-reporting of NDs in IMIS. A later study

[ii] https://www.osha.gov/opengov/healthsamples.html

comparing IMIS and CEHD data found that ND results are less likely to be recorded into IMIS.[41] Both studies recommended the use of both IMIS and CEHD datasets as complementary data sources.

Synthetic vitreous fibers (SVF) database

In 1995, OSHA developed a list of 18 priority agents, 5 of which were addressed through regulatory action and the remaining 13 were addressed through voluntary measures due to the lack of OSHA resources. Synthetic vitreous fibers (SVF) were one of the exposure categories designated as a non-regulatory priority and as a result, the North American Insulation Manufacturers Association (NAIMA) approached OSHA and subsequently the Health and Safety Partnership Progam (HSPP) was developed. The HSPP included the creation of an exposure database and its development, challenges, and key elements of success are described in a 2011 paper.[45] The elements of success included stringent quality assurance/control (QA/QC) procedures and a committee, a comprehensive data dictionary, and the commitment of the trade association and its members. The QA/QC committee was comprised of industry experts, which enhanced the external credibility of the database, and as such they were also responsible for resolving any potential trade secret issues. The challenges included consistency issues (e.g. different analytical methods), data quality issues, and the lack of incentive to companies outside of NAIMA. The database was also endorsed by OSHA as a reliable data source. It is unclear whether this database is still operational, but it provides a useful example of an industry-regulator partnership.

3. United Kingdom: The National Exposure Database (NEDB)

Background and aims:

The National Exposure Database (NEDB) was established in 1986 initially as a centralised repository of hygiene measurements for airborne substances collected by the Health and Safety Executive (HSE) occupational hygiene inspectors. The objectives as stated in 1989 were: "*1) to provide detailed and comprehensive exposure data for use in the setting of the new occupational exposure limits; 2) to provide a major source of exposure data for use in epidemiological studies; and 3) to facilitate dissemination of information on occupational exposures*".[46]

Data collection:

The data was gathered by HSE occupational hygiene inspectors during surveys, inspections, and investigations in United Kingdom (U.K.) workplaces. In addition to inspection visits, representative surveys were also carried out to ascertain information on working practices and exposure levels for specific industries.[47] The majority of measurements for the NEDB were collected between 1985-

1990.[1] The database was upgraded to an Oracle database in 1999 and at the time included 80,000 exposure samples from the existing NEDB, abstracts of all occupational hygiene visits, and two data files containing 20,000 asbestos exposure samples.[48] The NEDB is still operational; however data collection appears to have declined significantly over the past two decades. In 2001, Tickner[48] reported that the decline in data collection was likely due to operational changes at HSE which changed the focus to asking companies to conduct the monitoring themselves as well as the increasing cost of data collection. It is estimated that the NEDB currently contains between 50,000-100,000 measurements *(Peter Baldwin, database contact survey, April 2021)*.

Benefits/data use examples:

The NEDB has been described internally as a useful resource to support hygiene inspectors and for HSE policy development *(Peter Baldwin, database contact survey, April 2021)*, although limitations in data quality and database functionality may have impeded its full potential for use (see limitations). Only one research report has been generated from the NEDB over the past 15 years *(Peter Baldwin, database contact survey, April 2021)* and there does not appear to be any output from the database online after 2006.[49]

Other data use examples:

- One of the initial aims of the NEDB was to inform the standard setting of the Advisory Committee on Toxic Substances;[46]
- An examination of NEDB wood dust measurements (n=1,459) found a decline in exposure levels of 8% p.a. between 1985-2005;[47]
- NEDB data has been used for ExpoSYN (see section on EU databases);
- The data has been used for the development of the Exposure Assessment and Substance Estimation (EASE) model for regulatory exposure assessment[50] (which reportedly needs to be revised).[51]

Limitations and challenges:

The NEDB is a database of measurements largely collected during HSE enforcement duties and therefore higher than average exposure levels may be overrepresented.[48] The data has also been described as "*patchy*" and highly dependent on the users that entered measurement data.[49] The database survey contact reported that the representativeness of the data (i.e. selection of sites for inspection and HSE focus areas) and difficulty in accessing the data are two frequently cited barriers *(Peter Baldwin, database contact survey, April 2021)*.

In 2006, the Health and Safety Laboratory was commissioned by the HSE Occupational Hygiene Unit to ascertain how the NEDB was being utilised and how it could be improved.[52] Interviews (n=17) were carried out with representatives from HSE Units and Divisions. Only one of the 10 users accessed the database regularly. The general view was that the information was valuable, particularly for contextual information, however the data quality was inconsistent (e.g. incomplete records and inaccuracies in the information) and the database search functionality was inflexible. A better-quality control system was suggested in order to avoid misuse of the data. Further suggestions for increasing the power of the NEDB included links to international databases, greater involvement from industry, and targeted campaigns. The 2006 survey also revealed that HSE staff thought that there was a lack of willingness for industry to provide information because companies were concerned about being identified and targeted by HSE.[52] It is unclear whether any recommendations from this 2006 report were implemented.

In 2001, Cherrie et al.[29] conducted a feasibility survey for the retrospective collection of OEM data from industry in the UK and found that few organisations had readily accessible measurement data. The study utilised five strategies to identify data sources for toluene, ethylene oxide, and acrylonitrile: 1) using trade associations to identify companies; 2) hazardous substance users from two geographic areas identified from a commercial marketing database; 3) major users identified from a variety of sources (e.g. online database, industry advisory groups, personal contacts); 4) occupational hygiene consultants identified from the British Institute of Occupational Hygienists; and 5) Government organisations. The study found that of more than 800 organisations contacted, ~9% had data and of these, 74% were willing to provide data. The costs of identifying and collecting exposure data ranged from £7 to £380 per valid measurement depending on the data source.[29] For occupational hygiene consultants, 34 organisations were approached, 24 answered the questionnaire, and 12 were willing to provide data. The consultants and major users had the greatest data availability and were also the most likely to contribute data to the project. The consultants were also the only source of data for the limited small to medium enterprise (SME) data, which represented ~6% of the total measurements collected.

Ongoing development:

The HSE U.K. has expressed an intention to redevelop the NEDB as part of their science plan but planning has not yet commenced (*Yiqun Chen, personal communication, March 2021*). The database survey contact outlined several factors that are required to maximise the potential of the NEDB: *"improving accessibility and availability of the database both internally and externally; increasing* *awareness of the data and its potential use; and identification of new applications to increase data gathering and resource to deliver this." (Peter Baldwin, database contact survey, April 2021)*.

The HSE U.K. have funded the Center for Occupational and Environmental Health (COEH) at Manchester University to conduct a feasibility study to scope an occupational exposure control intelligence system (OccECIS; using RCS as a working example), which is due for completion in April 2021 and will partly inform the NEDB redevelopment (*Damien McElvenny, personal communication, December 2020)* . Any recommendations/developments from the OccECIS would be of interest to New Zealand.

4. Germany: MEGA

Background and aims:

The Measurement Data on Exposure to Hazardous Substances in the Workplace (MEGA) database was established in 1972 in order to provide a data pool for prevention, epidemiological issues, and investigations of occupational diseases.[53]

Data collection:

In Germany, social accident insurance is statutory and health and safety at work is monitored by statutory accident insurance institutions, including Berufsgenossenschaften (BGs) for the private sector and Unfallkassen (called "UK") for the public sector (excluding the military). Coverage is optional for self-employed workers and freelancers *(Dorothea Koppisch, database contact survey, January 2021)*. The amount of the wage-related contribution to the national insurance scheme is dependent on the risk class in a given industry.

Workplace measurements are carried out as part of the supervisory duties of the accident insurers or from measurement services campaigns (e.g. formaldehyde in pathology institutes of Universities, welding fume and metal exposure during welding of stainless steel *(Dorothea Koppisch, database contact survey, January 2021)*). The data collected by the inspectorates of the accident insurers are analysed by the *Institute for Occupational Safety and Health of the German Social Accident Insurance* (IFA; formerly BGIA).

The monitoring system for the measurement, analysis, and assessment of chemical and biological agents is called the *Measurement System for Exposure Assessment of the German Social Accident Insurance* (MGU; formerly the BGMG). All processes within the MGU system, including data measurement, analysis, and data entry are standardised and conform to the requirements of a quality management system (DIN EN ISO 9001).[20] Air samples are collected using standardised measurement systems (including techniques and devices for sampling) developed in-house and all sampling and analysis must be compliant with the standard EN482. In addition, the German Technical Rules for Hazardous Substances (TRGS) 402 require the collection of multiple contextual parameters of the workplace.

The data storage system used is the purpose-designed *Organisation System for the collection and use of measured data on exposure to hazards at the workplace* (OMEGA). The system is based on coding indices (including work area), specialised files, and record files and links data from on-site sampling to analysis at IFA through to documentation in MEGA, thus reducing data entry error.[54] The data are checked for plausibility and completeness both during data input on-site and in the central OMEGA software at IFA and other MGU laboratories. The MEGA database software is called MEGA^{Pro} which can be used for the extraction and statistical analysis of data *(Dorothea Koppisch, database contact survey, January 2021)*.

As at the end of 2020, MEGA contained 3.6 million records including 910 chemical agents and 750 biological agents. Data collection began in the 1960s and the number of measurements increased until ~2000 *(Dorothea Koppisch, database contact survey, January 2021)*. The number of data fields has grown from ~30 items from 1972-1989 to >200 from 1990-2007.[54]

Benefits/data use examples:

A number of reports/publications of MEGA data are available on the IFA website[iii] (some in English) and these can be searched by industry or substance. Some selected data use examples are listed below.

Data use examples:

The database has been used to:

- Support accident insurers to concentrate their prevention activities;[55]
- For risk assessment in the EU existing substances programme;[1]

[iii] https://dguv.de/ifa/gestis/expositionsdatenbank-mega/expositionsdaten-aus-mega-in-publikationen/index-2.jsp

- Estimate exposure levels for specific substances e.g. Clerc et al.[20] estimated formaldehyde exposure in various industries, including a multi-country comparison with France's COLCHIC database;
- Identify situations with potential for overexposure e.g. Pesch et al.[56] identified circumstances with high exposure levels for hexavalent chromium in welders and other occupations;
- Contribute to discussions about the establishment of EU limit values along with other European exposure database holders;[57]
- Inform evaluations for the preparation of REACH exposure scenarios;
- Validate exposure models e.g. Stoffenmanager Model;[9]
- Contribute data to ExpoSYN (see section on EU databases).

Limitations and challenges:

The data stored in MEGA are primarily collected for insurance purposes and thus the exposure measurements may be unrepresentative of average exposure levels[55], although the nature and direction of this bias is unclear. The accident insurers act as inspectorates[55] and the data they collect are also used for prevention purposes as well as for legal proceedings and compensation related to occupational disease claims and thus it is unclear how the multiple (potentially conflicting) functions may affect its operations. In addition, access to data is restricted to the statutory accident insurance institutions. One of the arguments for this is that data analysis needs to be carried out by "*skilled specialists*" with an awareness of historical developments in data acquisition and appropriate background knowledge in order to avoid misinterpretation of the data.[54] Successful attempts at obtaining MEGA data for international collaborations has generally been limited. For example, IFA did not contribute to the European Commission (EC) HazChem@Work project (described at the end of this section) citing data ownership reasons.[58]

Ongoing development:

MEGA is still going strong; the number of measurements has declined in recent years but the number of exposure agents is still increasing *(Dorothea Koppisch, database contact survey, January 2021)*.

5. France: COLCHIC and SCOLA

Background and aims:

France has two major exposure measurement databases hosted by the French National Institute of Safety and Health (INRS): COLCHIC and SCOLA, which represent the same population of workers but have different objectives. COLCHIC was established in 1987 and contains industrial hygiene measurements of chemical and biological agents from the INRS and the eight regional health insurance funds (CARSATs and CRAMIF) and their associated laboratories. The aims of the database are to: 1) centralise measurement data from various laboratories; 2) harmonise workplace sampling and air analysis methods; and 3) support chemical risk assessment in France.[24]

In France, exposure measurements made during assessment of compliance to regulatory occupational exposure limits values (OELV) are recorded in a national register called SCOLA (established in 2007). Under French law, compliance to OELVs must be assessed where a contaminant is thought to be present in the workplace.

Data collection:

COLCHIC:

COLCHIC contains data from companies covered by the national insurance scheme, which includes eight interregional chemical laboratories assigned to defined geographical areas. The national insurance scheme defines the priority industries for measurement campaigns. The three main reasons for a visit from a CARSAT/CRAMIF hygienist are based on: 1) prevention priorities (which can differ between regional laboratories who will identify and contact employers to be visited); 2) research priorities in the French prevention network (e.g. research surveys of specific industries); and 3) a request by a CARSAT/CRAMIF safety controller or engineer, the reason for which can vary widely. Employers cannot refuse a visit but they are scheduled in advance.[24] The majority (~70%) of exposure measurements that are stored in COLCHIC occur where CARSAT/CRAMIF hygienists suspect over-exposure.[20]

During 1987-2019, data for 1,197 exposure agents and ~1,150,000 exposure measurement results were recorded in COLCHIC *(Gautier Mater, database contact survey, December 2020)*. In 2016, it was reported that the three exposure agents most frequently measured in COLCHIC were inhalable particulates, toluene, and acetone.[24] Figure 3 from Mater et al. shows the number of samples collected up until 2012 was relatively stable over time (median=31,917 measurements). The most frequently sampled industries as reported in 2016 were manufacture and metal products, manufacture of rubber and plastic products, and manufacture of chemicals and chemical products, which represented nearly a third of the data.[24]

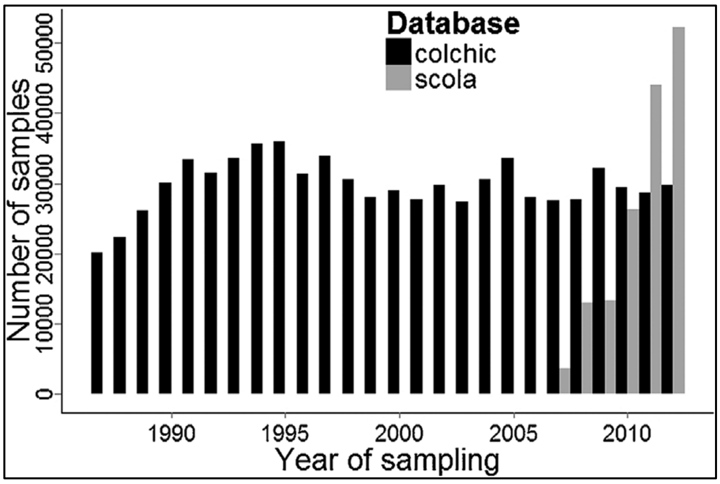
SCOLA:

In France, the company director is legally responsible for workplace risk assessments. In situations with the potential for overexposure and where an OELV exists, a sampling strategy must be developed, which may include compliance measurements conducted by independent accredited laboratories. A minimum of three and up to nine representative samples are legally required in order to assess compliance to regulatory OELVs.

For the period 2007-2019, data in SCOLA included 150 chemicals with more than 1 million measurements *(Gautier Mater, database contact survey, December 2020)*. In 2016, it was reported that the three agents most frequently measured in SCOLA were asbestos, wood dust, and alveolar particulate fraction.[24] The three most common industries were specialised construction activities, remediation activities and other waste management services, and manufacture of other non-metallic mineral products, representing ~45% of the data.[24]

COLCHIC and SCOLA were set up for different purposes and a comparison of the two databases including exposure levels has been described.[24] While COLCHIC data contains a mix of area and personal samples, SCOLA contains a larger proportion of personal samples,[24] the duration of sampling was generally shorter in COLCHIC, and overall higher levels have been observed for COLCHIC. Comparisons of exposure levels within common industry, task or occupational groups showed higher proportions of measurements above the OEL and higher geometric means for results contained in COLCHIC.[24]

Figure 3: Number of measurement samples in COLCHIC and SCOLA between 1987 and 2012



Source: Mater et al. 2016[24]

Benefits/data use examples:

Direct access to the data is not available; however, several information tools have been developed, including SOLVEX (chemicals), FIBREX (fibres except asbestos), and SCOLAMIANTE (asbestos). The data has been used for a variety of purposes, some examples are described below. There are a number of publications available from the INRS website.[iv]

Data use examples:

- General portraits of exposure (including trends over time) have been published for formaldehyde[59] and lead;[60]
- High-risk industries for specific exposures have been identified e.g. an analysis of the COLCHIC database identified that formaldehyde exposure was most common in human health activities and in the manufacture of wood products;[20]
- Inform OELVs e.g. the data has been used to provide a portrait of occupational exposures to chemical substances covered by the updated OELVs effective from July 1st 2020 *(Gautier Mater, database contact survey, December 2020);*
- Policy impacts have been assessed e.g. an analysis of exposure levels of mineral fibres from COLCHIC during 1986-2004 found that levels were broadly influenced by changes in OELVs;[23]
- Multi-country comparisons have been carried out for formaldehyde levels between COLCHIC and IMIS (U.S.)[59] and with MEGA (Germany);[20]
- The data has been used to inform the development of a software Tool for Exposure Assessment (TEXAS) which can be used to obtain estimates of level of exposure control (i.e. well-controlled to poorly controlled) as a function of determinants of exposure;[61]
- COLCHIC data has been used to provide retrospective portraits of occupational exposure to nickel; biological agents (2008-2018); crystalline silica (2009-2018); and cobalt (2007-2017) *(Gautier Mater, database contact survey, December 2020).*

Limitations and challenges:

As with most national databases, the representativeness of exposure levels is unknown as the database was not originally designed with the aim of being representative of all workers and this has been identified as the most difficult challenge by the database contact *(Gautier Mater, database contact survey, December 2020)*. The data from COLCHIC is from companies covered by the national insurance scheme (~55% of companies in France) excluding public services, agriculture, mining,

---

[iv] https://www.inrs.fr/publications/hst/base-de-donnees-colchic.html

energy production, national mass transit, and small retail and artisans.[24] In addition, the majority of measurements are taken upon suspicion of overexposure by CARSAT/CRAMIF hygienists. It is also possible that only companies proactive in occupational hygiene submit results to SCOLA *(Jerome Lavoue, personal communication, November 2020)*. The investigation of biases from the two databases is limited; however, the comparison of COLCHIC and SCOLA data[24] reported that for COLCHIC, 62% of records were for "*determining the exposure for an operation or task*" and 30% for compliance. For SCOLA, 91% of records were taken for compliance. In addition, a previous study by Lavoue et al.[62] identified several determinants of formaldehyde exposure levels from COLCHIC (e.g. sampling duration, year of sampling, and local exhaust ventilation (LEV)), which were incorporated into statistical models to provide a multi-industry portrait of formaldehyde exposure.

A new system of coding for COLCHIC was introduced in 2002 based on recommendations for minimum data requirements by specific work groups.[63, 64] This involved the modification and/or replacement of existing variables and the addition of new variables such as occupation,[24] and thus information on occupation is only available after 2002. In COLCHIC, sequential samples are entered separately and linkage of this data would require manual checking.[24] In addition, there is no identification of individuals which prevents estimation of within- and between- worker variances.[62]

Ongoing development:

COLCHIC has been operating for over 30 years and is still going strong and the measurements in SCOLA, which is used to define priorities for national prevention programmes, have grown exponentially since its inception in 2007. The data contained in SCOLA has grown from 11 substances with more than 1,000 measurements reported in 2013 to ~1 million results for 150 chemicals last year. In 2016, Mater et al.[24] predicted that SCOLA will be the major source of exposure information in France in the coming decades.

6. Finland: LIMS

The various databases of the Finnish Institute of Occupational Health (FIOH) have been described previously (most recently in 2014[65]); however, there do not appear to be any recent descriptions in English-language published literature.

Background and aims:

FIOH is an independent research and specialist organisation which operates under the Ministry of Social Affairs and Health in Finland. FIOH has offered occupational hygiene and biomonitoring services (which they are required to provide by law) for chemical, physical, and microbiological agents since the 1950s[66] and workplace measurements were previously stored in the Finnish Database of Occupational Exposure Measurements (FDOEM). Although data had been collected since 1950, electronic data was available for chemical agents only from 1994 onwards.[66]

Data collection:

The FDOEM database has since been replaced by a customised laboratory information system (LIMS)[v], which is used to manage and store workplace measurement data from FIOH's hygiene and biomonitoring services. This system was established in 1999 but not used comprehensively until ~2010 *(Tapani Tuomi, personal communication, February 2021)*.

FIOH also maintains a number of different (not specifically OEM) databases including the mandatory Registry of Employees Occupationally Exposed to Carcinogens (the ASA Register) and the FINJEM information system, which has been used extensively throughout the world. The FINJEM combined data from multiple FIOH databases, including the FDOEM, with expert assessment and national labour force statistics, in order to provide exposure estimates for each occupational group. In 2014, it was reported that the total number of exposure measurements used to assess chemical exposure for FINJEM was 157,035 and the most commonly measured agents in the database were styrene, chromium and chromium VI, formaldehyde, and other mineral dusts.[65]

Benefits/data use examples:

- The OEM data collected by FIOH has been used in combination with the ASA Register, national work and health surveys, and expert assessment to inform the FINJEM, which has been used extensively in Finland and internationally for the development of other JEMs and for numerous epidemiological studies;
- The LIMS is used to:
    - Track trends in exposure;
    - Facilitate the efficient production of client reports;
    - Conform to the quality standard ISO-17025 for testing and calibration laboratories *(Tapani Tuomi, personal communication, February 2021)*.

Limitations and challenges:

[v] A LIMS is a software system that supports a clinical laboratory's operations and workflow and is used to track, manage, store, and report results for laboratory samples and associated data.

The representativeness of companies approaching FIOH for occupational hygiene services is unknown but medium to large firms may be overrepresented. Customers vary by industry e.g. most companies in the mining sector are likely to engage FIOH's services; however, few companies within the construction sector engage FIOH's services *(Tapani Tuomi, personal communication, February 2021)*.

7. Denmark: ATABAS

The Danish Institute of Occupational Health (AMI; now called the National Research Centre for the Working Environment) used to host the ATABAS database of chemical exposure measurements, which was established in 1983. The data collected were from all air samples performed by AMI as a result of visits to workplaces with OHS problems[67] as well as measurement data from the Danish Working Environment Authority; however, ATABAS has not been used for many years *(Lars Andrup, personal communication, April 2021)*. The database has been transferred to an archive and is reportedly difficult to access due to both technical reasons and restrictions imposed by the General Data Protection Regulation (Regulation (EU) 2016/679) *(Lars Andrup, personal communication, April 2021)*. The survey contact was unable to locate any personnel that could answer questions about ATABAS.

The Doc*X Danish Occupational Cohort is a nationwide database hosted by Statistics Denmark including occupational information for all Danish residents for the period 1964-2013.[68] The central data source is the national job register which contains occupation and industry data for the entire employed population every year. The job register is linked to other data sources including labour market statistics, education, deaths, and migration. Doc*X is an open research resource; however, exposure assessment is primarily carried out with JEMs.

8. Norway: EXPO and EXPO Online

Background and aims:

EXPO is the Norwegian national database for chemical and biological exposure measurements (established in 1984) administered by the National Institute of Occupational Health (STAMI). The aim of EXPO is to collect airborne occupational exposure measurement data to build a knowledge base of exposures in different occupations and industries in order to ultimately prevent illness and promote good health *(Ragnhild Ostrem, database contact survey, January 2021)*.

In 2014, STAMI received funding to change the focus and purpose of EXPO to a nationwide database and in 2017, EXPO Online[vi] was developed as an online registration service open to all companies in Norway. The aim of EXPO Online is to serve as an archive for companies' measurement data *(Ragnhild Ostrem, database contact survey, January 2021)* and the long-term objective is for the database to be representative of all industries. The system was initially designed as a mandatory registration system but was changed (for political reasons) to a voluntary registration tool designed to help companies fulfil their legal obligations to document a safe work environment *(Ragnhild Ostrem, database contact survey, January 2021)*.

Data collection:

Prior to 2014, the database contained measurements from the analyses carried out by the STAMI service analysis laboratory through: a) STAMI research projects; and b) companies contacting STAMI to request exposure monitoring. From 1984-2019, the database contained about 100,000 air samples (which can contain one or more components) corresponding to ~600,000 exposure results for 1,400 different components (i.e. agents and different fractions) *(Ragnhild Ostrem, database contact survey, January 2021)*.

Benefits/data use examples:

The EXPO database has been running for 35 years and is one of the few OEM databases specifically aiming to increase representation of all industries.

Data use examples:

- Data has been used in the Nordic Occupational Cancer Study (NOCCA);
- Data has been used for ExpoSYN (see section on EU databases);
- The distribution and skewness of exposure measurements for certain agents collected up until 1997 have been examined;[69]
- An analysis of organic solvent exposure measurements between 1985-1994 reported a downward trend;[70]
- A general portrait of styrene exposure has been reported, including average measurements, time trends, and results by occupational group.[71]

Limitations and challenges:

[vi] https://expodemo.stami.no/ (In Norwegian)

The database is not representative of all companies and industries as the data represents companies that contacted STAMI to conduct exposure monitoring and the policy/research focus areas at the time. It has been previously estimated that STAMI received ~20% of all airborne exposure measurements conducted in Norway between 2001-2005 *(Ragnhild Ostrem, database contact survey, January 2021)*.

There are currently no data or methods to assess the representativeness of companies reporting to EXPO Online but potential solutions are being sought through industry-specific projects *(Ragnhild Ostrem, database contact survey, January 2021)*.

Ongoing development:
In addition to EXPO Online, STAMI are conducting industry-specific projects in order to collect exposure measurements in industries where information is lacking as well as companies with less resources for exposure monitoring *(Ragnhild Ostrem, database contact survey, January 2021)*.

9. South Korea: Work Environment Measurement Database

The information described below is solely from the survey of the database manager *(Seunghyun Park, database contact survey, January 2021)*.

Background and aims:
In 2000, the Occupational Safety and Health Research Institute (OSHRI) of the Korea Occupational Safety and Health Agency (KOSHA) established the Work Environment Measurement Database with the aim of "*monitoring workplace hazards, research and producing annual reports of company exposure data*".

Data collection:
In South Korea, employers are required to regularly evaluate their workplace if hazardous materials are present. It is mandatory for companies to conduct exposure monitoring once every 6 months or once every year if the concentration level of all chemical substances is below the occupational exposure limit. These workplace exposure measurements are typically conducted by consulting institutes who transfer the results to the Work Environment Measurement Database via an electronic collection system. The database contains data for 190 different exposure agents and ~75,000 companies report annually or bi-annually.

Benefits/data use examples:

The main users of the database are the Ministry of Labour, researchers, and workers with pending occupational disease claims.

Data uses include:

- An annual report of company exposure data is produced;
- The exposure database has recently been linked to the workers' health data to form a central worker health database;
- OSHRI researchers have established a cohort study from the data.

Limitations and challenges:

Access to the database is restricted to third parties; however, researchers can access the data through an application process at OSHRI. The challenges described by the database manager concerned the difficulties with linking exposure data to health data.

10. Singapore: National Database for Noise and Chemical Exposure

Background and aims:

The Ministry of Manpower (MOM) established the National Database for Noise and Chemical Exposure in the early 1990s,[72] which is currently managed by the Occupational Safety and Health Division (OSHD) of the MOM. The system was upgraded in 2010 to enable companies to submit their data electronically. The aims of the database are to enable MOM to: 1) identify high-risk workplaces; 2) assess trends in exposure levels; and 3) provide advice to employers on control measures and appropriate monitoring programmes *(Evelyn Koh, database contact survey, March 2021)*.

Data collection:

In Singapore, companies with workers likely to be exposed to excessive noise or toxic substances are required by law to monitor occupational noise levels under the Workplace Safety and Health (Noise) Regulations and toxic substances (of which there are 600+) under the Workplace Safety and Health (General Provisions) Regulations and to retain the results for inspection. All hygiene monitoring must be conducted by a competent person (i.e. who has undergone the requisite training). The measurement data contained in the database are from hygiene assessments conducted by OSHD inspectors and reports submitted by companies *(Evelyn Koh, database contact survey, March 2021)*.

For toxic substances monitoring, the frequency of monitoring depends on the level of exposure against the PEL (e.g. <10% of the PEL: no monitoring is required; 10-50% of the PEL: at least once a year). For noise monitoring, the frequency is dependent on the number of employees likely to be exposed to excessive noise levels and the reporting of results is required once every 3 years.[73] The reporting of both noise and toxic substances can be submitted online using a standard form. In 2020, 1,561 chemical exposure results were received and 672 workplaces submitted noise measurement results between 2018-2020.[74]

Benefits/data use examples:

The database has enabled OSHD to develop targeted workplace health programmes, which has led to reductions in occupational disease cases over time. For example, the number of cases of "noise-induced deafness" has reduced over the years.[75]

Limitations and challenges:

In 2017, MOM reported that noise and chemical exposure reports were received from ~320 workplaces which is a "*fraction*" of all workplaces where workers may be exposed to excessive noise and toxic substances levels.[75]

Ongoing development:

The number of workplaces submitting their exposure reports is increasing over time. The MOM is planning to increase the coverage of its hygiene surveillance programme, which was a recommendation under the Workplace Safety and Health National Strategy to 2028.[75]

11. Switzerland: SUVA

Background and aims:

The Schweizerische UnfallVersicherungsAnstalt (SUVA) is Switzerland's largest provider of compulsory insurance cover against accidents and occupational diseases. One of its mandates is the prevention of occupational disease for all employees in Switzerland; however, exposure monitoring is only carried out for employers insured and therefore supervised directly by SUVA which equates to about half of Switzerland's workforce, including the construction and production sectors.[76] Measurements are usually carried out at the request of inspectors (*Patrick Steinle, database contact survey, April 2021)*.

The OEM database was established in 2007 and the aims include: a) ensuring that chemical exposure does not exceed OELs for individual substances under concern; b) informing reviews of OELs; c) identifying high-risk industries to target for control activities; and d) enabling comparisons with published literature (*Patrick Steinle, database contact survey, April 2021)*.

Data collection:

The database includes data for ~400 chemical and biological agents and ~5000 measurements are collected each year (*Patrick Steinle, database contact survey, April 2021)*. Some (a few thousand measurements) retrospectively collected data prior to 2007 is also included in the database. The results are contained in reports which also include a description of the company's activities, measurements results (typically from multiple measurements at different times and locations), and contextual information.

Benefits/data use examples:

Researchers from the University Institute for Public Health (Unisante) have used data from the SUVA database to validate the Advanced REACH tool.[77]

Limitations and challenges:

Some industry sectors are better represented than others in the SUVA database and measurements are usually carried out at the request of inspectors i.e. when there is suspicion of overexposure. It is also likely that companies try to reduce exposure prior to a visit by an inspector (*Patrick Steinle, database contact survey, April 2021)*. The data entry process and associated quality assurance has also been described as cumbersome *(ibid.)*.

12. Examples of exposure-specific databases:

CODUST (Denmark)

The CODUST programme includes two exposure databases, one for wood dust and the other for endotoxin and organic dust measurements as well as two quantitative JEMS which were developed for use in a large epidemiological study of respiratory diseases.

The database for wood dust includes occupational hygiene measurements from Denmark, Finland, France, Germany, the Netherlands, Norway, and the UK. During the period 1978-2007, 25,450 wood dust measurements were used to develop a JEM.[78] The database for organic dust exposure measurements were collected from Europe and Canada, including >3000 endotoxin measurements from workers in the animal and crop production and related industries, which were collected between 1992 and 2008.[79]

Evalutil (France)

Evalutil was established in 1992 by the French Institute of Public Health Surveillance (InVS) to document occupational exposure to asbestos and man-made mineral fibres (one database for each exposure).[80] In the early 2000s, the database was also extended to nanoscale particles.

The Evalutil data is primarily from Europe and North America and data sources include: 1) published scientific literature; 2) measurement campaigns conducted by laboratories for prevention and regulatory checks, particularly the regional laboratories of the risk prevention division (CARSAT) of the French Health Insurance Fund; and 3) industrial hygiene reports from international organisations (e.g. FIOH and the US National Institute of Occupational Health) and voluntary dust control programmes from industry.

The data are stored in an ACCESS database with indexing criteria for bibliographic references (author/laboratory, year of publication) and exposure situation (occupation, industry, material containing fibres, and operations performed on material). The indexing criteria and some of the exposure data are coded from thesauruses developed by expert industrial hygienists and the occupations and industries are coded to international and French standard classifications. Expert industrial hygienists also assign an information quality score. Three JEMs (asbestos, mineral wool fibres, and refractory ceramic fibres) have been developed from the data.

Evalutil contributes to epidemiologic studies of inhaled particles and their health effects[80] and in 2015 it was reported that the database was consulted an average of 400 times per month mainly by occupational physicians and industrial hygienists. Since 2000, Evalutil can be accessed for free in both French and English[vii]. The data are from a variety of diverse sources and therefore there are differences in data quality, reporting, and availability of ancillary variables. The representativeness of the data is also unknown.

13. European Union Collaborative Databases:

A number of industry-specific databases have been developed over the last two decades, such as PAPDEM for the pulp and paper industry,[81] AWE for the asphalt industry[82] and EXASRUB[83] for the

[vii] https://ssl2.isped.u-bordeaux2/fr/eva003/

rubber manufacturing industry, all of which assemble/collate exposure measurement data retrospectively. Most of these databases also identified previously unpublished sources of data. Multi-national exposure databases include WOODEX for wood dust exposures,[84] DERMDAT for dermal exposures,[85] the European Industrial Minerals Association (IMA) Dust Monitoring Programme (DMP),[30] and the ExpoSYN database.[86] The latter two programmes are described below.

European Industrial Minerals Association - Dust Monitoring Programme (IMA-DMP)

In 2000, the European IMA which represents ~500 mineral companies operating >700 mines and quarries and 750 plants throughout Europe initiated a Dust Monitoring Programme (DMP) of prospectively collected measurements. Since 2006, the database has been hosted by Utrecht University in the Netherlands. The aim of the programme is to "*collect representative exposure data from a wide variety of workplaces to determine existing exposure levels and to monitor exposure trends over time*".[30]

At the initiation of the project, a detailed standardised protocol including sampling strategy (including number and duration of measurements), sampling methods, analytical methods, and data management was given to all participating companies. The IMA-DMP protocol specifies minimum information requirements for inclusion in the database: 1) sampling of pre-defined exposure groups; 2) personal sampling; 3) availability of a unique worker code (to distinguish between measurements performed on different individuals); and 4) sampling duration.[30]

In 2007, a quality control review was conducted[87] and a quality control system was implemented, which included the verification and checking of data completeness and consistency before inclusion in the database.[30] In addition, post-2006 information on field blanks were also included and the definition of full shift sampling was expanded. The quality of the data improved after the strict quality control procedures were introduced.[87] The IMA-DMP has >35,000 RCS measurements collected over a 15-year period *(Hans Kromhout, personal communication, October 2020)*.

Participating companies receive a report of sampling results after each campaign which can be used to identify jobs with high exposure levels and inform the development of interventions to reduce exposure. This company-specific feedback is thought to be one of the main reasons for companies to participate.[30] An industry-wide report of pooled anonymous data is produced 1-2 times per year.

EXPOSYN

Another EU collaboration is the ExpoSYN exposure measurement database which was created from the SYNERGY study, which is a large pooled analysis of epidemiological case–control studies on occupational risk factors for lung cancer.[86] The ExpoSYN database comprises exposure measurements for asbestos, chromium, nickel, polycyclic aromatic hydrocarbons (PAH), and RCS from 18 European countries and Canada covering a 50-year time period (1951-2009). In 2012, the database included a total of 356,551 measurement results distributed by agent as follows: respirable crystalline silica (42%), asbestos (20%), chromium (16%), nickel (15%) and PAHs (7%). The majority (90%) of the data is from national exposure databases (mainly MEGA, COLCHIC, NEDB, and EXPO), 1% from industry-specific databases (AWE and PAPDEM) and the remaining 8% from research institutes. Peters *et al.*[86] noted in 2012 that ExpoSYN is intended to be used to develop a JEM for the SYNERGY study. The challenges of creating ExpoSYN included time-consuming data collection (coding data into a standardised format), language barriers, and the storage of non-digitised data.[86]

HazChem@Work database

The aim of the HazChem@Work project is to harmonise information from available national databases and data sources (including non-OEM data sources) for occupational exposure to chemical agents in EU member states. A pilot project was conducted by the European Commission in 2016 which included the development of a pilot database (using MySQL open source technology with data being delivered in pre-designed Excel data sheets).[58] The pilot study identified 80-100 data sources (~10 million measurements) across the EU member states; however, a limited number of data holders were willing to contribute data. The key barriers included national legal constraints, concerns about treatment and processing of data, different data formats, and language issues. A further feasibility study was conducted in 2018 to investigate the transfer to and further development of the database by EU-OSHA[88].

Table 4: Summary of international databases under review

Number	Country	Name of System	Still operating?	Regulatory requirement to conduct and report OEM?	Third party access?	Summary
1	Canada	Canadian Workplace Exposure Database (CWED)	Yes, plans to develop CWED 2.0	Historically yes but currently less so (depends on State)	Not at the moment, but plans to improve access	The CWED is one of the more recently established databases and its further development is being championed by researchers. The database contains valuable historical information from eight (out of fourteen) jurisdictions; however, the majority of measurements were taken for compliance purposes from two jurisdictions and therefore may not be representative of exposure levels in the general workforce. The expansion to other jurisdictions is hampered by declining data collection by regulatory agencies. The CWED 2.0 project, which includes improving sustainability and data access, is still in its infancy but will be of great interest to New Zealand in the near future, particularly as it aims to obtain prospective data and encourage the deposition of new data.
2	US	Integrated Management Information System (IMIS)	Yes, now OIS	Yes	Yes via Freedom of Information Act. Unclear for OIS	The IMIS represented the largest database of occupational exposure measurements in the U.S. and was also one of the few databases that granted access to external parties through the Freedom of Information Act. As a result, the data has been used for a wide variety of purposes and a considerable number of research publications have been generated. The limitations section of this review for the IMIS database is more detailed than for other international databases but this could be the result of the limitations, particularly the potential selection bias of the compliance data, being more thoroughly investigated and published in the scientific literature. For example, over half of the papers in the Lavoue et al.[36] review examined or included investigations of potential selection bias for the limited ancillary variables that were collected. The lack of empirical investigation of bias and the treatment and interpretation of NDs have been described as the two major limitations of using IMIS data to assess exposure in the general population; however, use of the data, providing biases are adequately characterised, has been encouraged[36]. The OIS replaced the IMIS around 2014 to better meet the operational needs of OSHA; however, it is unclear whether the type of data collected and the level of data access granted to external parties is the same as for the IMIS.
3	UK	National Exposure Database	Data collection has stalled; but plans to re-develop	No	NA	The NEDB was established in 1986 initially as a centralised repository of hygiene measurements for airborne substances collected by the HSE occupational hygiene inspectors. The NEDB is still operational; however data collection appears to have declined significantly over the past two decades. In 2001, it was reported that the decline in data collection was likely due to operational changes at HSE as well as the increasing cost of data collection. However, the HSE UK has expressed an intention to redevelop the NEDB as

						part of their science plan. Any recommendations/developments from the OccECIS and re-development plans for the NEDB would be of great interest to New Zealand.
4	Germany	MEGA	Yes	Yes	No	MEGA represents the largest and most comprehensive of the international occupational exposure databases. All processes pertaining to data collection and storage are highly standardised and the specialised bespoke software and data flow from worksite to the MEGA database is an "*essential element of quality assurance*".[54] A considerable amount of effort is made to collect and describe all factors that can affect a measurement result and thus more than 200 data items can be collected per measurement record (including task, room volume, ventilation, LEV). However, access to the database by external parties is restricted and successful attempts to access the data for international collaborations has been limited.
5	France	COLCHIC	Yes	Yes	No	France has two large, ongoing exposure measurement databases that share the same software, extensive ancillary information (e.g. task, type of workplace, PPE), standardised coding, and source population and thus can provide a more complete picture of exposure levels compared to other countries (e.g. with one compliance database); however, caution has been advised in their joint interpretation due to their different objectives.[24] Direct access to the data is not available; however, several information tools have been developed. The representativeness of exposure levels is unknown and the national insurance scheme that provides data to COLCHIC does not cover all industries.
6	France	SCOLA	Yes	Yes	No	
7	Finland	Finnish Database of Occupational Exposure Measurements	Yes, now LIMS	No	No	The centralised exposure measurement data in Finland stored in the relatively recent LIMS database is from the occupational hygiene and biomonitoring services provided by FIOH and it is unclear how representative the exposure levels/companies are. However, FIOH have a number of databases and information sources and has a strong history of combining data sources to provide a more comprehensive picture of occupational exposure. In particular, the FINJEM has been a leading exposure assessment tool for several decades.
8	Denmark	ATABAS	No	No	NA	The Danish Institute of Occupational Health (AMI; now called the National Research Centre for the Working Environment) used to host the ATABAS database of chemical exposure measurements, which was established in 1983. The data collected were from all air samples performed by AMI as a result of visits to workplaces with OHS problems[67] as well as measurement data from the Danish Working Environment Authority; however, ATABAS has not been used for many years.
9	Norway	EXPO	Yes, +EXPOOnline	No	Yes with restrictions	The EXPO database has been running for 35 years and Norway is one of the few countries included in this review making an ongoing effort to increase representation of all industries and focus exposure monitoring efforts (though industry-specific projects) in industries with less available data. EXPO Online is still relatively recent and therefore the representativeness of industries voluntarily reporting data is unclear at present but the concept will be of interest to New Zealand.
10	South Korea	Work Environment Measurement Database	Yes	Yes	Yes to researchers	Unlike other compliance databases where the frequency of data collection is dependent upon inspection visits, the data collection of the Work Environment Measurement Database is consistent and regular i.e. annual or bi-annual according to monitoring results

						is mandatory. The database also has efficient data flow from results collected by the consulting institutes to KOSHA and is linked to worker health data. The annual report is an important component and was the main impetus for the creation of the database.
11	Singapore	National Database for Noise and Chemical Exposure	Yes	Yes	No	The information from the database has informed targeted programmes to reduce work-related health risks and a reduction in occupational disease cases, such as noise-induced deafness, has been attributed to these programmes. There is also a legal requirement for the data to be collected by a competent person. It appears that the reporting of noise monitoring is mandatory whereas toxic substances monitoring is recommended and thus the number of reports received represents only a fraction of workplaces where workers may be exposed to excessive levels. However, MOM have an ongoing campaign to increase its hygiene surveillance “footprint”.
12	Switzerland	SUVA	Yes	No	No	The data collected in SUVA is from an insurance system which covers about half of Switzerland’s workforce, and some sectors are better represented than others. The database has been used to answer complex occupational health questions, for benchmarking, and to identify problem areas. Companies are regularly provided with summary reports.

Common components of successful databases:

1. Regulatory or insurance law requirement

Many of the international databases comprise data collected by regulatory agencies where the collection and reporting of exposure monitoring is a legislative requirement. The countries where there is a mandatory requirement to report exposure monitoring data or Government inspectors are collecting data include Germany, France, the U.S., South Korea, and Singapore (see Table 4).

In a 2018 report reviewing options for occupational exposure and disease surveillance in Ontario, Canada, Demers et al.[5] concluded that the reporting of workplace hazards in the absence of a regulatory requirement is a significant challenge as employers may be unwilling to voluntarily divulge information.

2. National OHS institutes

In several of the Scandinavian countries, there is no regulatory requirements to conduct exposure monitoring but long-standing national occupational health institutes (often operating under a Government Ministry) provide exposure monitoring services to companies. In Finland, the national OHS institute is required under law to provide exposure monitoring services and therefore has permission to maintain a database although it is not mandatory for companies to use their services *(Tapani Tuomi, personal communication, February 2021)*.

3. Standardised processes and quality control system/reviews

The collection and maintenance (e.g. regular quality reviews) of high-quality data increases the value and utility of the data. Several of the European databases conform to the requirements of European standards for measurement strategies (EN689) and performance of procedures (EN482). Germany's MEGA database is part of a wider measurement system (i.e. MGU) where all processes pertaining to data collection and analysis are standardised and conform to the requirements of a quality management system. The industry-led IMA-DMP has a standardised protocol which was prescribed to all participating companies at the beginning of data collection. A quality assurance check of IMA-DMP data was conducted in 2007 and a stricter quality control procedure was subsequently implemented.[87] A review was also conducted for France's COLCHIC in 2002 and a new system of coding was introduced based on recommendations for minimum data requirements by specific work groups.[24] Several systems include an assignment of a quality control score e.g. Evalutil in France.[80] In

Singapore, hygiene monitoring must be carried out by competent persons that have undertaken the requisite training courses.

4. Efficient data flow into the database

Data entry can be a cumbersome process which is prone to errors and can be highly user-dependent (a weakness identified for the U.K.'s NEDB). The systems in Germany and South Korea allow for efficient data flow from worksite to OEM database, which for Germany's MEGA has been described by the Head of Unit responsible for MEGA as an "*essential element of quality assurance*".[54] Several systems have evolved into laboratory information management systems where the OEM database is linked to the analytical laboratory as part of an overall system, such as the OIS in the U.S. and the LIMS in Finland. In Switzerland, the quality assurance of data entry was described as tedious and the database manager reported that a LIMS system would be desirable. These LIMS systems reduce the need for excessive "data handling" and associated entry errors.

5. Reporting of results and value-added benefits

For the IMA-DMP, the company-specific feedback received by participating companies is thought to be one of the main reasons for participation in the programme.[30] In South Korea, the annual report of company exposure data is an important component and was the main impetus for the creation of the Work Environment Measurement Database. For the development of the CWED, the approach to data holders described "value-added" raw data to the original data holder such as assistance with data management, standardisation, and coding of data and advice/assistance of conversion of paper files into electronic format. Norway's EXPO Online provides free and secure data storage for companies *(Ragnhild Ostrem, database contact survey, January 2021)*. One international researcher who regularly uses OEM data from large databases emphasised that a database needs to be useful to the individuals who generate and enter the data.

6. The data are regularly used for policy and research

Both systems that have restricted access (e.g. Germany) and more open access (the U.S.) have been used for a variety of different purposes:

> *-Identifying 'high-risk' groups such as occupations and industries*
>
> One of the more common uses of OEM data for policy and prevention is to identify priority areas (including workers, occupations, industries, and situations) to focus prevention efforts. This includes highlighting industries where hazardous exposures (e.g. RCS) are highly

prevalent or industries with a high proportion of workers exposed to levels over occupational exposure limits.

-Providing general portraits of exposure for specific substances or substance groups
In addition to identifying priority groups for intervention, broad portraits of specific exposures, including priority hazards such as RCS or classes of agents such as carcinogens, can be provided from OEM databases including the estimation of exposure levels, time trends (see below), and exposure determinants.

-Monitoring trends in exposure levels over time
Many of the databases that include historical data have reported trends in exposure levels for various substances over time. Even trends in exposure based on compliance and enforcement information (i.e. exposure levels that may be potentially skewed) are also useful to observe if "*trends (and activity) are moving in the right direction*".[52]

-Assess policy impacts
Evaluating whether an intervention or policy had the desired impact can be difficult without the consistent availability of data at baseline and post-intervention.

-Contributing to expert information systems such as JEMs
Some of the limitations of databases (discussed below) make exposure assessment for the general working population difficult. The combination of exposure measurements with expert information systems such as JEMs can address some of these limitations and data gaps.[26] A prominent example from this review is the CWED which was created for the CAREX Canada information system to "*increase the accuracy and robustness of exposure estimates*".[6]

The identified enablers from the database contacts survey for the establishment of international OEM databases are reported in Appendix Table A2.6 and generally related to the need for exposure data for both prevention and research.

Common challenges:

1. Representativeness of exposure levels

Many of the reviewed systems were administrative databases held by regulatory agencies and a common theme was that given the purpose (e.g. for compliance) and the nature of the data collection (i.e. the non-random selection of workers/workplaces) measurements may be biased towards highly exposed occupations and short-term measurements of peak ‘worst-case’ scenario exposures. The measurement strategies (i.e. representative or worst-case) varied widely or were unknown.

The range (or lack thereof) of occupations and industries represented in a given database can depend on a number of factors including, for example, the priority areas set by a regulator or accident insurance institution, or the type of clients that engage the monitoring services of an OHS institute (e.g. in the cases of Finland and Norway). In addition, not all industries may be covered by a national insurance or OHS scheme e.g. IMIS data did not cover the mining and agricultural sectors and France’s insurance scheme does not cover public services, agriculture, mining, energy production, national mass transit, and small retail and artisans.[24]

The representativeness of SMEs in the database as well as the recording of company size also varies. Larger companies may collect and submit more data as well as have more resources for controlling exposure.[45] Several studies of IMIS data have observed associations between company size and exposure levels.[37, 41, 43]

Potential biases must be considered when using any database and the collection of contextual information and ancillary variables are important for assessing the factors (e.g. reason for sampling, sampling duration etc.) that can affect measurement results. Although, Germany’s MEGA can document more than 200 data items per sample, the risk of misinterpretation is one of the reasons that MEGA and France’s COLCHIC are not readily accessible to external parties. In contrast, the U.S. IMIS could be accessed through the Freedom of Information Act and a large number of published papers have assessed the likely effect of bias. Although the number of ancillary variables collected is limited, Lavoue et al.[62] argued that the application of “refined” statistical tools to ancillary data will improve the interpretation of OEM data.

Previous attempts to validate against external sources have been limited; however, a limited number of comparisons have been made between exposure levels from databases in different countries:

- In 1995, Vinzents et al.[89] compared exposure levels in ATABAS, COLCHIC, EXPO, MEGA, and NEDB and reported overall little variation in geometric means (EXPO was the exception) for xylene measurements. The results for MEGA and COLCHIC (i.e. both insurance databases) were generally lower than for the other databases. The paper reported that all database administrators stated that the national sampling strategy was "*biased toward a worst-case strategy*".
- Lavoue et al.[59] conducted a comparison of formaldehyde measurement data between France's COLCHIC and the U.S. IMIS collected between 1986-2001. The results showed similar multi-industry portraits of formaldehyde exposure, with the exception of the health-related activities where COLCHIC levels were higher.
- A comparison of formaldehyde measurements from Germany's MEGA and France's COLCHIC found that exposure levels collected over 2002-2011 did not differ significantly for five studied sectors despite the differences in overall purpose of data collection.[20]

2. Data acquisition

In several jurisdictions, such as in the U.K. and most of Canada, there has been a downward trend in in the number of exposure measurements conducted by regulatory agencies and therefore it is harder to access exposure measurement data for policy and research. The data acquisition process of historical data for the CWED was time-consuming and involved negotiating data access agreements with existing data holders, designing and implementing data access and security protocols, and extensive data cleaning and coding. The lack of harmonisation of coding (particularly for non-carcinogens) was a major challenge that was identified *(Hugh Davies, personal communication, January 2021)*.

There are few examples of voluntary data being provided by industry. There was a suggestion of the inclusion of industry for U.K.'s NEDB but the survey of HSE staff revealed that companies were concerned about being identified and targeted by HSE.

3. Determinants of exposure

There are a large number of variables that could potentially affect exposure levels. In addition, the collection and reporting of these variables varies widely.

a. *Measurement reason and strategy*

The collection of information about measurement reason and strategy i.e. representative or worst-case is vital information for understanding the exposure results.[86]

b. *Sampling method*

Aspects of the sampling method can affect the measurement of exposure levels. Personal (in the breathing zone) versus area (or stationary) samples has been shown to affect exposure levels. Sauve et al.[26] reported that personal samples were associated with exposure levels that are 72% higher than area samples. Sampling duration has also been associated with exposure levels in large databases[20, 59, 90] with longer sampling time associated with lower exposure levels.

c. *Between- and within-worker variance*

For many of the databases, it is unclear how repeated measures are handled or even if they can be identified from the data. The importance of between and within worker variance was highlighted in a 1993 paper by Kromhout et al.[91] which advocated for including repeated measurements from the same worker in measurement strategies. Since then, specific measurement strategies, models, and estimation procedures for exposure variability have been developed.

The availability of repeated measurements has been described in several of the EU collaborative databases. A 2008 analysis of IMA-DMP data found that one third of measurement data included repeated measurements that would enable assessment of within and between worker variability in exposure,[87] although this is likely to have increased. In the EXASRUB database for the rubber manufacturing industry, identification numbers were assigned to individual workers and locations to identify repeated measurements from the same worker/location and thus enable the within- and between-worker variance components to be analysed, without the need for collecting personal information.[83]

d. *Non-detectable results*

The problem of (under) reporting and recording of non-detects has been discussed extensively for the U.S. IMIS data and handled differently by different researchers. Comparisons between IMIS and the CEHD suggest the OSHA inspectors are likely to underreport NDs to IMIS. Furthermore, how missing values are imputed (i.e. averaged or assigned a zero) is important for the interpretation of the data.

e. *Contextual information*

The content and quality of contextual information collected for exposure sampling varies widely, particularly for systems based on historical data. In particular, the recording of PPE and other control information is generally lacking.

Two widely referenced guidelines were developed in the 1990s, which proposed a list of minimum data requirements to enable the valid interpretation of quantitative measurements (discussed further in 5.4). However, the collection of contextual information to facilitate proper interpretation of exposure results needs to be balanced against practicality in the field.

4. Data quality

One of the major challenges of aggregating data is the inconsistent data quality, which includes differences in type of variables included (described above), missing data, and inconsistencies in coding of data. Data quality was identified as a major challenge for the CWED, IMIS (particularly the treatment of NDs and free-text occupation variable), and a contributor to the under-utilisation of the U.K.'s NEDB. For the EXASRUB database in the rubber manufacturing industry, the lack of occupational hygiene information in the original datasets was the reason for poorer agreements in coding.[30]

Identified barriers

The identified barriers from the database contacts survey for the establishment and further development of international OEM databases are reported in Appendix Table A2.6. Funding was a common barrier for both the establishment and further development of databases. Other barriers for establishment included difficulty in identifying data sources (CWED); data protection and security issues as well as harmonising the data collection for multiple accident insurers (MEGA); poor knowledge of source data and transformation of data to a usable format (Singapore); high costs of measurements and exposure variability (Switzerland); and technological issues (OIS and EXPO Online). The identified barriers to further development generally related to resources, both in terms of funding and staff. For the U.K.'s NEDB, the decline in data collection and the applicability of the data were significant barriers.

3. A New Zealand OEM database

3.1 New Zealand context:

The principal legislation governing occupational health and safety in New Zealand is the Health and Safety at Work Act (HSWA) 2015. This legislative framework is underpinned by self-regulation where a '*person conducting a business or undertaking*' (PCBU) has an obligation to identify, assess, and manage risks in the workplace. Specifically, under S36(3)(g) of the HSWA 2015 there is a primary duty of care placed on PCBUs to monitor the conditions at the workplace and the health of workers for the purpose of preventing work-related illness or injury; however, there is not an explicit requirement to meet a certain Workplace Exposure Standard (WES) or Biological Exposure Indices (BEI).

Prior to 1988 when occupational health resources were transferred to the Department of Labour (DoL; formerly the Occupational Safety and Health (OSH) Service), the then Department of Health proactively carried out workplace exposure monitoring.[92] This historical monitoring data is either not publicly available or has been lost through the transfer of responsibilities from one agency to another. In particular, the Institute of Environmental Science and Research Limited (ESR) have not retained any historical OEM data (*Rob Lake, personal communication, March 2021*). With the introduction of the previous Health and Safety in Employment (HSE) Act 1992, the onus was shifted to employers to carry out exposure monitoring, and therefore there was limited exposure data available in the public domain.

There are currently no legislative requirements for any Government agency, including WorkSafe, to measure or collect exposure data. Under the HSWA Act 2015, inspectors can enter a place of work and conduct, or direct the PCBU to conduct, "*examinations, tests, inquiries, and inspections*". The inspectors can also require PCBUs to produce documents; however, there is no extraction of this data for further analysis *(Kerry Cheung, personal communication, March 2021)*. Exposure monitoring in New Zealand is largely carried out by private occupational hygienists and researchers:

Occupational hygienists

Occupational hygienists include consultants (who provide occupational hygiene services to a range of companies) and in-house occupational hygienists who typically provide services to one employer. The New Zealand Occupational Hygiene Society (NZOHS) is a network of occupational hygiene professionals. There are ~115 New Zealand members (including associate members) and around 40 are practising members. NZOHS members conduct the majority of exposure monitoring in New Zealand (*Robert Murray and Kerry Cheung, personal communication, April 2021*).

In the mining industry, some quarries contract occupational hygienists to carry out monitoring whereas some have ventilation officers undertake monitoring (*Paul Hunt, personal communication*, *April 2021*).

Academic institutions

Academic institutions involved in occupational health and exposure research have also collected occupational exposure measurement data and have invested in exposure measurement equipment.

For example, the CPHR has been collecting exposure measurement data for their occupational health and exposure epidemiological studies for over a decade. These studies have been mainly of specific occupational groups and includes measurements of airborne samples (e.g. inhalable and respirable dust, solvents, bacterial endotoxin), biological samples (e.g. persistent organic pollutants in blood serum, pesticide metabolites in urine), video exposure monitoring, and more recently dermal exposure monitoring. The CPHR's current programme of intervention research involves collecting measurements of solvent exposure in collision repair workers, dust and silica exposure in construction workers and engineered stone benchtop manufacturers, and pesticide exposure in pesticide applicators.

A limited number of other academic institutions collect exposure data e.g. Dr David Welch from the School of Audiology, Faculty of Medical and Health Sciences, University of Auckland and Associate Professor David McBride from the Department of Preventive and Social Medicine, University of Otago collect workplace noise measurements for research studies.

In New Zealand there has never been a coordinated effort to centrally record exposure monitoring data. Therefore, there is limited information available on the distribution, frequency, and level of workplace exposures in the public domain which hinders efforts to inform New Zealand-specific prevention efforts for reducing work-related disease and injury. The National Occupational Health and Safety Advisory Committee (NOHSAC) was established (2003-2009) to provide independent advice to the Minister of Labour on issues relating to occupational health and safety. The committee produced a series of technical reports including a review of Australian and New Zealand workplace exposure surveillance systems,[93] which involved a review of current exposure surveillance systems based on consultation with relevant organisations and agencies. The report identified one industry-based surveillance system in the forestry industry but concluded that there was no national exposure surveillance system in New Zealand.[93] Recommendations were made for New Zealand based on the available evidence,[94] the majority of which have not been implemented.[95]

The different types of occupational exposure surveillance carried out in New Zealand to date include surveys of working conditions, a register of exposed workers, a substance register, and researcher-developed expert information systems (described in Appendix 4). There is currently no centralised occupational exposure database in New Zealand.

3.2 A New Zealand occupational exposure measurements database

3.2.1 Why should we have a database of New Zealand-specific measurements?

Section 1.1 outlined why actual exposure measurement data is important for prevention and to inform policy and evidence-based solutions. There has been a historical shift in New Zealand and other countries of the responsibility for exposure data being collected by employers rather than regulators, which means that aggregated and up-to-date data is harder to access for policymakers to make decisions about prevention and for researchers to investigate causes of disease.

A common theme from the public submissions for the '*Health and Safety at Work Strategy 2018-2028*' was that the collection of health and safety data in New Zealand needs to be improved and shared across the wider system and thus the Strategy states that quality data is a priority.[96] The current estimates of the burden of work-related harm used by WorkSafe and the currently available research JEMs mostly rely on measurement data collected overseas. It is unclear to what extent these overseas measurements reflect the situation in New Zealand workplaces given likely differences in legal, economic, and social infrastructures between countries. Furthermore, there are many historical examples of New Zealand being slower than other countries to change (through legislation or otherwise) their working practices. For example, New Zealand was the last country in the world to halt the production of the phenoxy herbicide 2,4,5-T in 1987 and the importation of asbestos-containing products was only banned recently in 2016.

The priority areas for WorkSafe include carcinogens and airborne risks and a New Zealand OEM database could provide valuable data on the level of exposures for these priority areas as well as complement other data sources. For example, WorkSafe's recent 'OccIdeas' survey of carcinogen exposure will provide information on the distribution and frequency of carcinogen exposure in certain industries and an OEM database could potentially provide information on whether workers are exposed to levels above current standards.

3.2.2 What are the benefits of a New Zealand occupational exposure measurement database?

An OEM database for New Zealand could provide collective knowledge of what hazards are present in New Zealand workplaces and thus provide an invaluable national resource that could be used to inform targeted interventions (as well as determine what is reasonably achievable), assess trends over time, provide evidence for compliance assessment, allow for policy impact assessment, and help evaluate compliance with and inform reviews of Workplace Exposure Standards (WES). The data to inform decision-making will help contribute towards WorkSafe's mission to be a world-class regulator.

An exposure measurement database could also be complementary to the existing (and any future) workforce surveys and could also be used to vastly improve New Zealand-specific information systems such as the NZJEM[97] and to update the NZ-CAREX.[98]

The advantages for specific stakeholders with a vested interest in the prevention of work-related disease and injury in New Zealand are outlined in Table 5 (also see Table 1 for general advantages). The benefits for each of the major international databases were discussed and summarised in Section 2.

Table 5: Benefits for specific stakeholders in New Zealand

Stakeholder:	**Benefits:**
WorkSafe New Zealand	• Inform the development and targeting of interventions to reduce exposures e.g. this could include targeting occupations and industries exposed to RCS for exposure control and/or education of health risks • Inform which hazards are priorities for prevention • Evaluate whether interventions have the desired impact • Evaluate the impact of specific policies or regulatory controls • Evaluate the achievement of and inform the review of WES and BEI • Assess trends in exposure levels over time • Compare what exposures are being monitored within WorkSafe's focus areas

	• Indicate competency of professionals undertaking exposure measurements • Highlight data gaps that may be important for future policy • Complement other WorkSafe data collection methods (e.g. the new OccIdeas survey) • Facilitate broad comparisons with other countries (i.e. benchmarking) • Raise general awareness of occupational hazards and work-related health
Occupational hygienists	• Enhance professional judgement and inform occupational hygiene assessments • Provide knowledge of the major risks/hazards in a given industry • Facilitate professional development • Industry benchmarking and therefore added value for clients • Potential for standardised data entry (and therefore less errors) and better reporting capabilities • Help increase awareness of risks among clients • Facilitate the planning of future sampling strategies • Generate new ideas for using OEM data to promote their businesses and help clients
New Zealand Occupational Hygiene Society (NZOHS)	• Review consistency in exposure measurement sampling • Indicate competency of professionals undertaking exposure measurements • Quality assurance for members • Provide evidence for WES review submissions • Help increase awareness of risks and the importance of exposure monitoring • A rigorous standardisation process will improve the overall quality of data
Researchers	• Epidemiological research (e.g. the level and intensity of exposures in relation to levels at which adverse health effects may occur) • Highlight data gaps that may be important for future research • Improve exposure assessment for epidemiological studies and contribute to the development of exposure models and information systems e.g. JEMs
Industry	• Facilitate industry benchmarking • Increase awareness among businesses of risks to be managed/prioritised (particularly useful for SMEs) • Provide knowledge of the major risks/hazards in a given industry • Encourage the preservation of data • Implement exposure reduction controls where needed

3.2.3 What would a New Zealand occupational exposure measurement database look like?

The measurement of workplace exposure levels is resource-intensive and since there is no regulator data (historical or present), an exposure measurement database developed for New Zealand would involve centralising data collected and stored by industry and private occupational hygienists and academic institutions. A New Zealand exposure measurement database would capitalise on the time and expense already expended to collect measurement data (which is otherwise underutilised) with the aim of building a body of exposure measurements over time.

3.3 Stakeholder consultation with NZOHS

As mentioned above, NZOHS members conduct the majority of exposure monitoring in New Zealand and therefore their views and ultimately their buy-in are crucial for the success of a New Zealand database. The consultation with NZOHS members through discussion at a members' meeting and an online data holders' survey are presented in this section.

3.3.1 NZOHS Members' Meeting – 11 December 2020

The scoping project was presented to a virtual meeting of NZOHS members on 11 December 2020 which included the scope of the problem, the benefits of a database, a description of the international review, and the planned stakeholders' survey. The issues that were raised in the discussion that followed included potential data quality issues (e.g. the handling of poor sampling and poor reporting of results) and whether there should be a competency requirement to contribute data; whether the data might potentially overrepresent proactive employers and thus underestimate exposure levels; concerns about WorkSafe as a host; who the end users will be and how to mitigate the risk of a database replacing occupational hygiene services if it was accessible to the public. Finally, some recommendations by members were also made for good examples of databases from the mining industry in Australia.

3.3.2 Data holders' survey – NZOHS members

The data holders' survey (see Appendix 3) was an anonymous online survey that was sent to members of the NZOHS via the fortnightly newsletter from the 18 February to 28 March 2021. Within WorkSafe, the industry engagement leads and the High Hazard Unit were approached for relevant contacts but no further potential data holders were identified.

The results for the survey are presented here or under specific barriers/issues discussed in section 4. We requested that one hygienist per organisation completed the questionnaire. The full results are presented in Appendix 5.

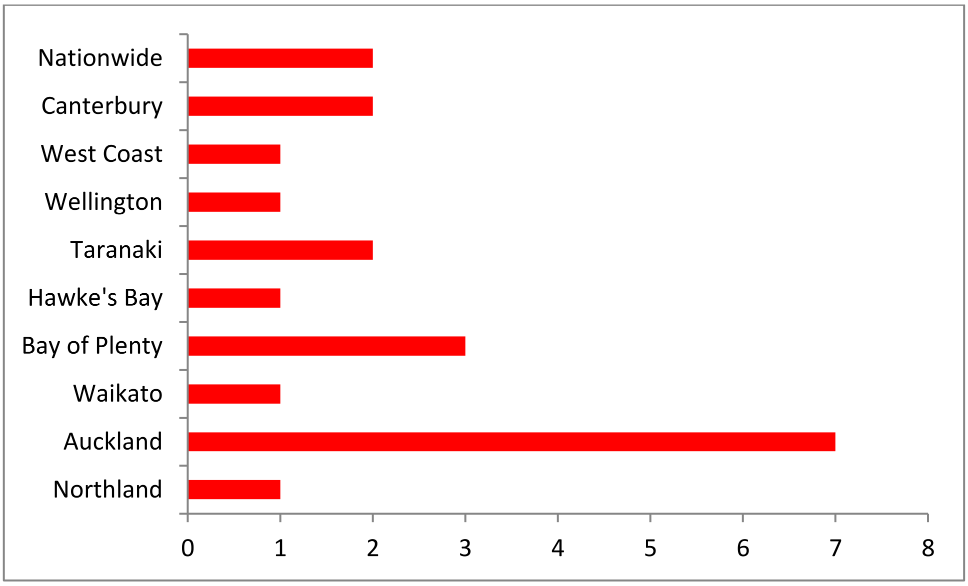
Description of respondents:

The survey was completed by 21 participants; however as mentioned above, we requested one hygienist per organisation.

Of the respondents that completed the online survey, 57% (n=12) provided OH services to a range of companies, 14% (n=3) provided OH services to one company (i.e. in-house), 14% (n=3) provided OH services but a third-party company conducts the sampling, and 14% (n=3) were Government employees. Two out of three Government employees did not complete the remainder of the questionnaire (the majority of which was not relevant to them).

The region(s) in which the respondents' organisation was *located* in is presented in Figure 4 and the regions for which the respondents' organisation provides OH services for is presented in Figure 5.

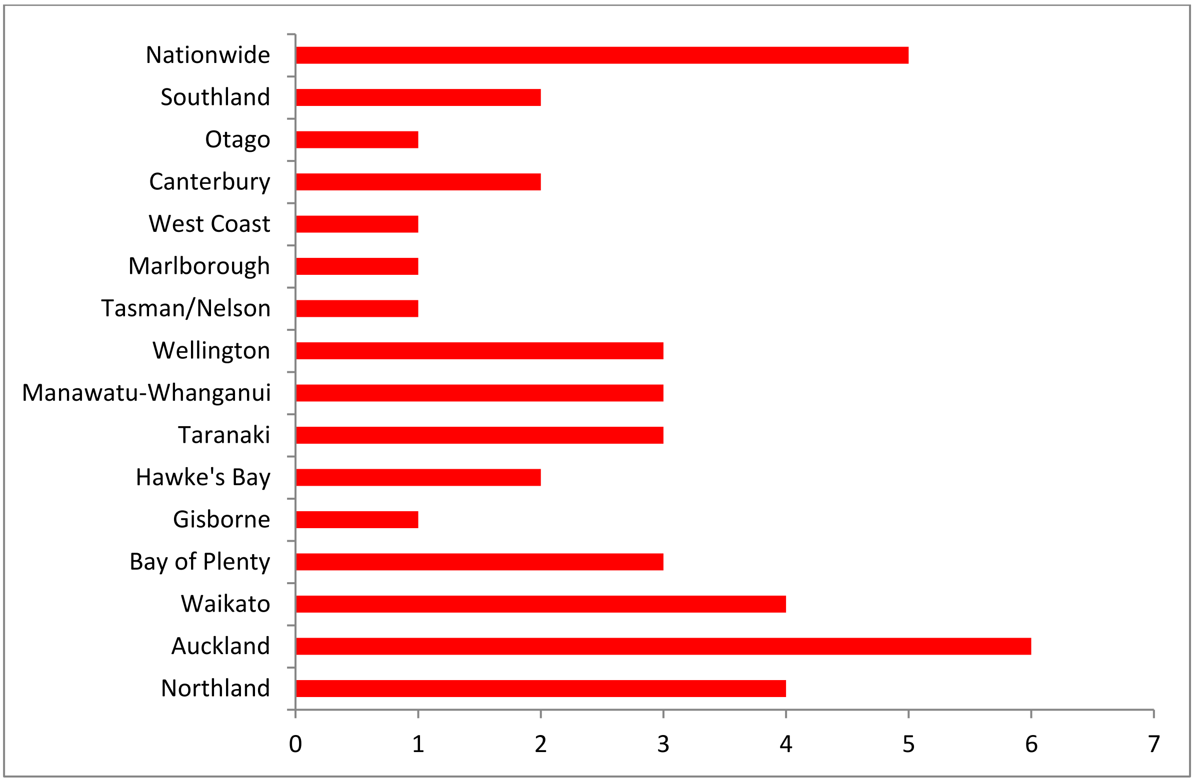
Figure 4: The region(s) where respondents' organisations are located:



Number of respondents

(21 responses from 18 respondents (3 respondents reported two locations))

Figure 5: The region(s) serviced by respondents' organisations:



Number of respondents

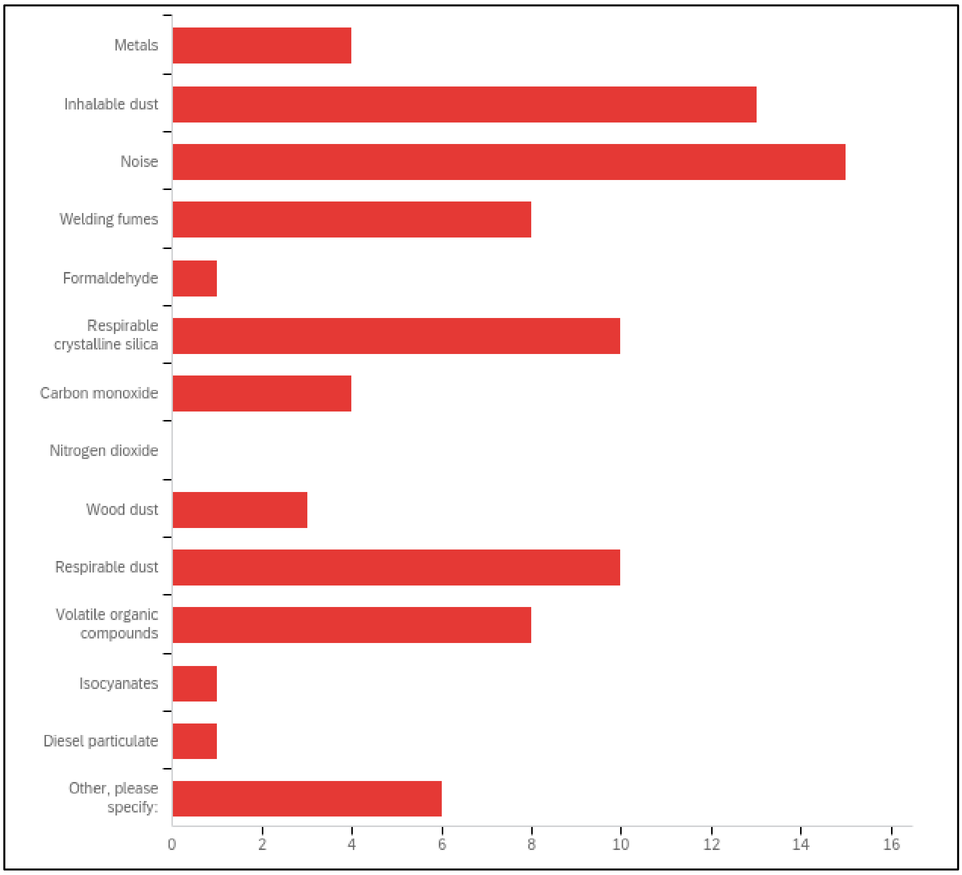
(42 responses from 15 respondents (12 respondents reported >1 location))

Of 16 respondents, 38% (n=6) were the sole hygienist in their organisation, 44% (n=7) reported 2-5 hygienists, and 19% (n=3) reported 6-10 hygienists in their organisation.

Occupational hygiene data

Just over half (56%) of respondents reported that their organisations use standard collection forms for each OH visit. Respondents were asked to report the five most commonly sampled agents by their organisations (Figure 6). The most commonly sampled agents were noise (83% of respondents), inhalable dust (72%), RCS (56%), and respirable dust (56%). The agents specified under the "other" category were asbestos (n=2), fungal spores (n=2), heat, and fibres.

Figure 6: The five most commonly sampled agents reported by respondents

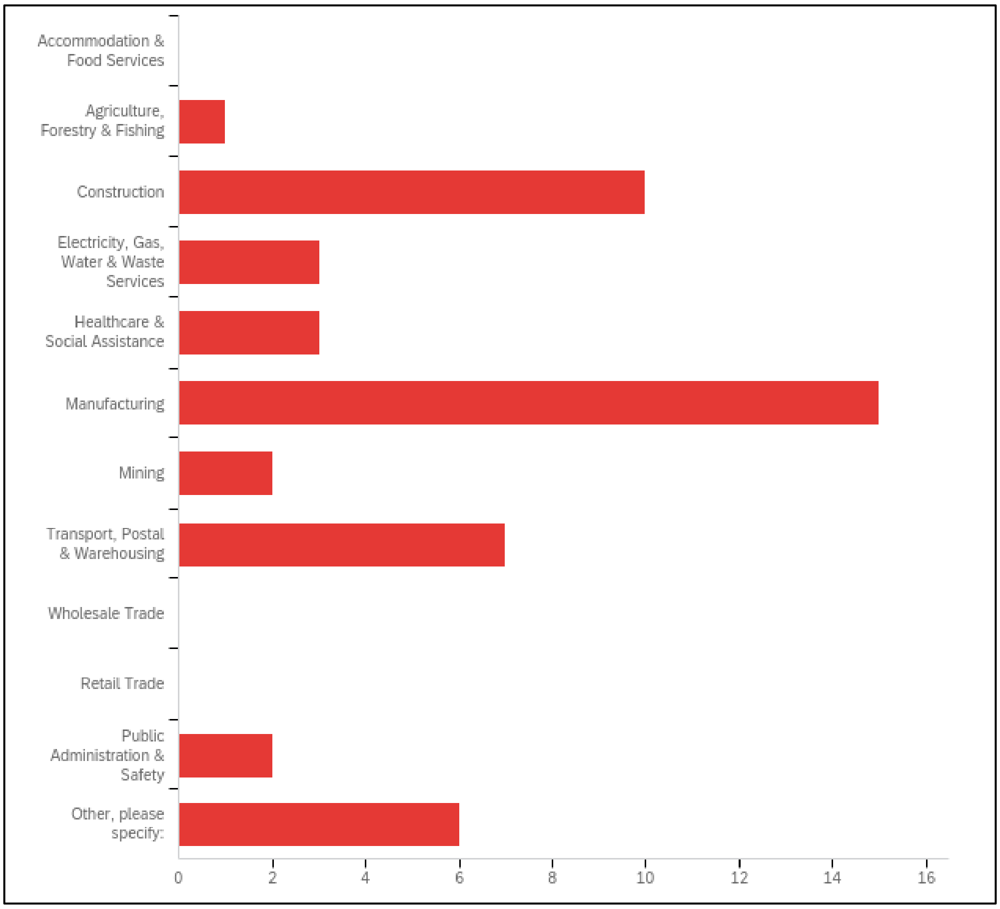


Number of respondents

(84 responses from 18 respondents (NB: some answered > 5, some answered less))

Respondents were also asked to report the three most commonly sampled industries that their clients were from (Figure 7). The most commonly sampled industries were manufacturing (83% of respondents), construction (56%), and transport, postal and warehousing (39%). The industries specified under "other" were oil and gas (n=2), office, entertainment, and Local Government.

Figure 7: The three most commonly sampled industries reported by respondents



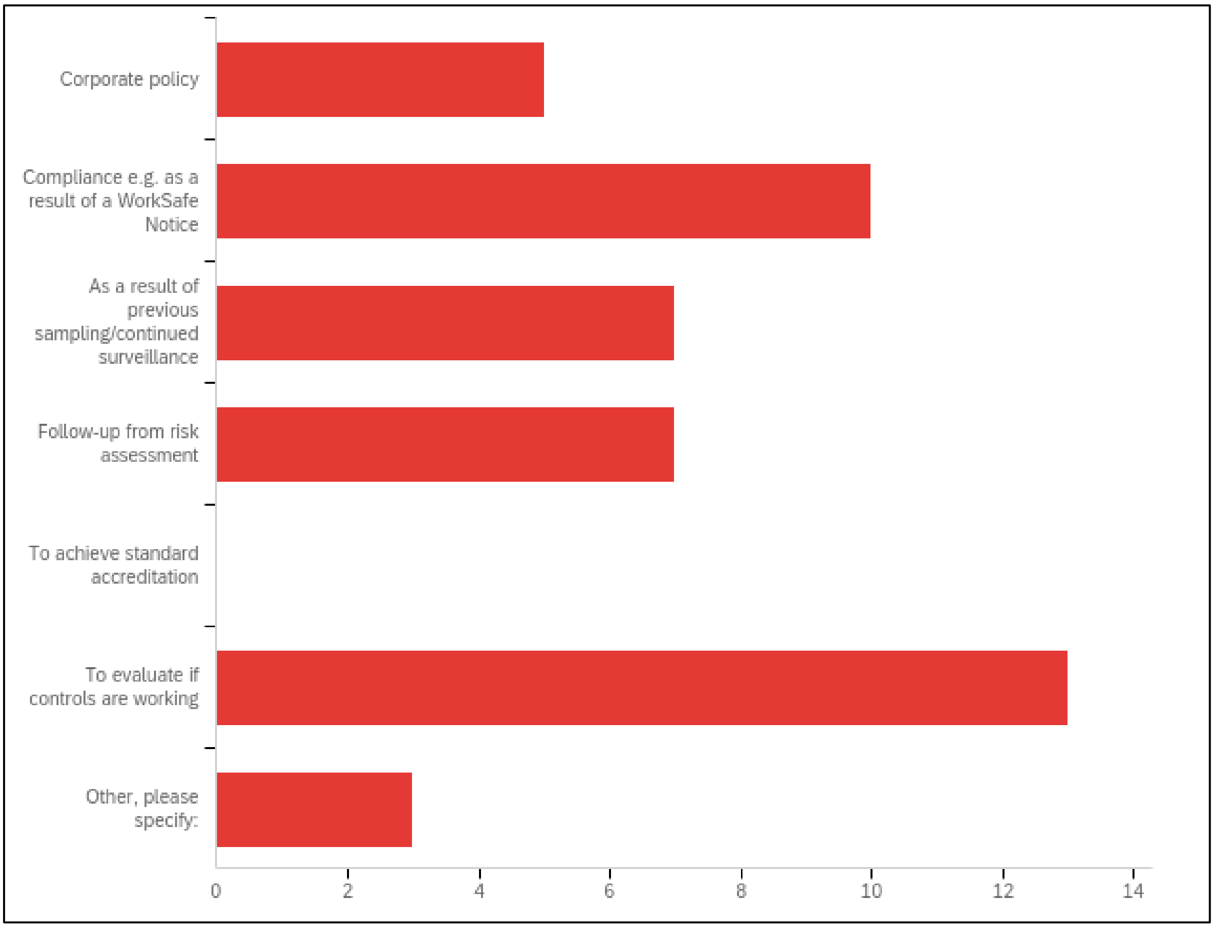
Number of respondents

(49 responses from 18 respondents)

Half of respondents (n=8) reported that the average size of the companies that OH services are provided for varied widely and 38% (n=6) reported that the average size of their client's companies is 100+ employees. The average number of measurements collected per year was 1,383 (from 12 respondents) and ranged from 100-10,000. The median was 275 (25^{th} quartile: 100; 75^{th} quartile: 1250).

Figure 8 presents the main reasons for clients engaging the exposure monitoring services of the respondents' organisations with multiple answers possible. The main reasons for clients engaging OH services were to evaluate if controls are working (81% of respondents) and compliance (63%). "Other" included review of systems and in-house service. The in-house occupational hygienists all listed corporate policy and 2/3 answered to evaluate if controls are working.

Figure 8: The main reasons for clients engaging the services of the respondents' organisations



Number of respondents

(45 responses from 17 respondents)

Of 16 respondents, 44% (n=7) reported storing their exposure data in individual electronic reports only (e.g. PDFs), 13% (n=2) reported hard copy and electronic reports, and 44% (n=7) reported an electronic storage system (some including individual electronic reports). Electronic formats included Microsoft Excel, Google Drive, and one "in-house" system. The length of data retention ranged from 5 to 50 years with an average of 18 years (median: 10; 25th quartile: 10; 75th quartile: 30) from 13 respondents, the remaining 3 answered indefinitely or "*depending on requirements*".

The views on the database are described throughout Section 4.

Limitations of the data holders' survey

We did not have access to descriptive information for the non-respondents and thus we cannot rule out the possibility that there may be differences between respondents and non-respondents. For example, there is a possibility that those that took part in the survey may be the ones who are keen to contribute to a database. The survey was completed by 21 participants (3 were Government employees who did not answer the majority of the survey) and the number of respondents that answered each question varied but was generally between 16-18 respondents and thus caution is advised when generalising results to the wider occupational hygienist population.

4. Outline of barriers/issues

This section discusses the potential barriers including an outline of the issue, experiences from international databases, responses from the data holders' survey, and the likely risks relevant to each issue.

Relevant sources for specific recommendations:

As mentioned in section 2, the development of CWED 2.0 project in Canada which is currently underway, is a useful example for New Zealand to follow given the historical decline of the availability of regulator data in both countries. At present, the "*data mobilisation phase*" is being conducted which focuses on building a new governance structure distinct from the former CAREX Canada partnership and new infrastructure (primarily administrative) to facilitate the evolution of the CWED *(Hugh Davies, personal communication, January 2021)*. Several of the recommendations presented here are based on the report to the WorkSafeBC Research Secretariat on the data mobilisation phase, including the development of data sharing agreements (a template is available from the report), the data governance group, the concept of the three-tiered access system, and the development of specific protocols. The principles and strict confidentiality protocols developed for the Statistics New Zealand's Integrated Data Infrastructure (IDI) will also serve as a useful practical example, particularly for user access.

4.1 Data acquisition

"Development and ongoing maintenance of an exposure assessment database requires, at a minimum, consultation with the persons who have a vested interest in the collection and use of exposure data. Involvement by these "stakeholders" is necessary to improve the quality of the database, as well as to assure commitment of the organisations to the database's success" (Cole 1995)[99]

Overview of issue:

Data acquisition issues include identifying relevant data sources, convincing data holders to contribute data, negotiating data access/sharing agreements with individual data holders, and dealing with different data storage formats (i.e. hard-copy vs. electronic). The standardisation of data is covered in *data quality* below.

In New Zealand, in the absence of mandatory requirements for the collecting and reporting of exposure monitoring results, data acquisition would depend on voluntary reporting of privately collected data. Gaining 'buy-in' from occupational hygienists and industry will be a key challenge.

Experiences from international databases regarding data acquisition:
The data acquisition process for many of the overseas databases involves a regulatory requirement (see Section 2). Data acquisition for the CWED involved the collection of historical regulator data, which has been described in detail in relatively recent reports.[6, 31] The creation of the database required considerable effort and involved negotiating memorandums of understanding and data access agreements with regulatory agencies across Canada, designing and implementing data access and security protocols, and developing standardised definitions, core requirements, and protocols to handle missing data. The 2009 data holders' survey found that only two of fourteen contacted agencies held their data in electronic databases and many agencies stored their measurement records in paper form or individual electronic reports (e.g. PDFs). For several jurisdictions, the transfer of data from paper to electronic records was a 2-year project funded by WorkSafe BC. The CWED team have published documents such as letters to data holders and data access agreements, which will serve as useful templates for a New Zealand database.

There are several examples of ongoing efforts to acquire privately collected data, including CWED 2.0, which will involve identifying new data partners. In Norway, EXPO Online was created in 2017 as a voluntary online registration service open to all companies in Norway. Similarly, in Singapore, companies are encouraged to submit their OEM results online. However, the likely representativeness of these systems is currently unclear. In both Norway and Singapore, efforts to increase coverage of companies have been indicated but the exact details are unclear. In the U.S., efforts to centralise Workers' Compensation (WC) insurers data are ongoing but key challenges include different data storage types, lack of data standardisation, and the privacy and intellectual property concerns of private companies. The earlier UK feasibility survey conducted by Cherrie et al.[29] found that few UK organisations had readily accessible measurement data.

There are limited well-established examples of systems that describe the incentives for private companies to contribute data. For the industry-led IMA-DMP, participating companies receive a report of sampling results after each campaign containing statistical evaluations of the measurements against OELs, which can be used to identify jobs with high exposure levels and inform the development of interventions to reduce exposure.[30] Norway's EXPO Online is a registration tool designed to help companies fulfil their legal obligations to document a safe work environment and provides a means for companies to archive their OEM data. For the development of the CWED, the approach to data holders described “value-added” raw data to the original data holder, such as assistance with data management, standardisation, and coding of data and advice/assistance of conversion of paper files into electronic format.

Finally, the CWED 2.0 and the HazChem@Work projects have both highlighted the importance of stakeholder engagement. One of the criticisms of the European Commission pilot project for the HazChem@Work database was the lack of stakeholder engagement early in the process concerning database requirements, shape, and content.[88]

Responses from stakeholder survey regarding data acquisition:

The data holders’ survey asked respondents if they thought a centralised database of occupational exposure measurements should be set up in New Zealand and all respondents that answered the question (n=16) responded in the affirmative. Respondents were asked to rate how willing they were to contribute data on a scale from 1 (unwilling) to 10 (very willing). The average rating was 7.1 (median: 7.5; 25th quartile: 6; 75th quartile: 8.5; n=16). Three respondents also indicated caveats which included board approval, a well-designed system that does not take up too much time, a clear benefit to the hygienist’s company, a “*disconnect from the regulator*”, and that information is strictly confidential/anonymous. Three respondents (plus 2 Government employees) did not answer the question.

Respondents were asked which data types (i.e. prospective and/or historical) they would be willing to contribute to the database. Of the 15 respondents that answered the question, 60% (n=9) reported that they were willing to contribute prospective data only, 40% (n=6) reported that they were willing to contribute both prospective and historical data, and none reported that they were willing to contribute historical data only.

Respondents were asked if there were perceived barriers to contributing data (other than privacy and intellectual property, discussed below). Of the 14 respondents who answered this question:

- 2 reiterated that privacy concerns were important;
- 3 respondents said that fear of WorkSafe using the data to prosecute companies was a barrier;

- 4 mentioned the time taken to input data would be a potential barrier (1 mentioned that an interface system that would allow uploads from existing software would be key);
- Other barriers raised by single respondents included lack of incentive (would there be some form of rebate/recognition to the companies that participate?), the risk of data security breaches, misinterpretation of results, and unwillingness of clients. One respondent mentioned that there would be less barriers if the database were a Government mandated requirement.

Potential risks to consider regarding data acquisition:

- In the absence of mandatory requirements, few occupational hygienists/industries submit their data;
- The data access negotiation process is time-consuming e.g. navigating client's non-disclosure agreements etc.
- The extraction of data is expensive and time-consuming. Fifty-six percent of respondents reported that their data were stored in individual electronic reports (i.e. PDFs) rather than electronic databases and 80% of respondents (n=10) responded "*yes*" or "*maybe*" to charging for the contribution of newly collected data;
- NZOHS members do not fully represent the exposure monitoring landscape in New Zealand i.e. we are not capturing the poorer performing companies.

4.2 Representativeness: how do we ensure that the data represents all New Zealand workplaces?

"Exposure measurements tend to be objective, but identification and selection of workplaces and workers to be measured involve substantial subjectivity" (Peters 2012)[86]

Overview of issue:

The measurements captured in the database may over- or under-estimate average exposure levels in the wider workforce due to the type of companies/hygienists that voluntarily report to the database e.g. proactive employers with generally well-controlled exposures may be more likely to contribute data. It will also be important to understand whether certain industries and SMEs are more or less likely to engage monitoring services and therefore be represented in the database.

Experiences from international databases regarding data representativeness:

The representativeness of exposure levels is one of the major issues which affects all exposure databases and one database manager identified it as the most difficult challenge *(Gautier Mater,* *database contact survey, December 2020)*. Many of the overseas databases are administrative databases held by regulatory agencies and therefore the selection of workplaces for sampling is non-random and may overrepresent workplaces with higher or lower than average exposures. Even in compulsory systems such as France's SCOLA database of compliance measurements, there is a possibility that proactive companies are more likely to submit results *(Jerome Lavoue, personal communication, November 2020)*. The databases vary widely in their collection and investigation of ancillary variables but the general consensus is that exposure databases are invaluable resources provided effort (e.g. using advanced statistical tools) is put into characterising potential biases.

Responses from stakeholder survey regarding data representativeness:
Half of respondents reported that the average size of the companies that OH services were provided for varied widely and 38% reported that the average size of their clients' companies were 100+ employees. Very few specifically stated 100 or less and therefore it remains unclear whether SMEs will be adequately represented.

The issue of representativeness was brought up at the NZOHS members' meeting, specifically whether the data might potentially overrepresent proactive employers and thus underestimate exposure levels. One respondent from the data holders' survey commented that a database would need to be compulsory to be meaningful and expressed concern over the potential for only certain industries to opt-in to the database.

Considerations regarding representativeness:
There are multiple factors that could potentially contribute to selection bias and therefore skew the exposure data:

- Only certain types (e.g. proactive) of employers/industries engage monitoring services;
- Only certain types of hygienists/industry submit their data to the database;
- Too few samples collected;
- SMEs are less likely to engage monitoring services and are therefore underrepresented in the database

4.3 Data quality:

"Without decent data quality, there is no point in trying to aggregate." (Taylor Shockey, personal communication, November 2020)

Overview of issue:

The data from different sources will vary in content, detail, completeness (i.e. missing data fields), and coding practices, which makes it difficult to aggregate. The standardisation of data will likely require considerable effort.

Experiences from international databases regarding data quality:

The review of international databases found that inconsistent data quality was a common challenge for exposure databases (see Section 2). This included differences in type of variables included, missing data, and inconsistencies in coding of data. Data quality was identified as a major challenge for the CWED, IMIS (particularly the treatment of NDs and free-text occupation variable), and a contributor to the under-utilisation of the UK's NEDB. One of the crucial improvements of the new CWED 2.0 is to harmonise the coding of data *(Hugh Davies, personal communication, January 2021)*.

Responses from stakeholder survey regarding data quality:

At the NZOHS members' meeting, concern was expressed about how the handling of poor sampling and bad reporting of results would be handled and whether there would be a competency requirement to contribute results to the database.

The data holders' survey revealed that data was routinely collected on job title, industry, number of employees, reason for sampling, work process, sampling duration, exposure duration or frequency, measurement strategy, engineering controls, and use of PPE. However, only 56% of respondents reported that their organisations use standard collection forms for each OH visit.

In the general comments section of the survey, several respondents expressed concern about the wide range in data quality of OEM collected in New Zealand. One respondent commented that any historical data would need to be checked as there is some poor quality data "*out there*" (both in terms of data collection and interpretation). Another respondent expressed concern about samples being taken by inexperienced people as well as the risks of incorrect data entry and incorrect calculations ("*will it be 8-hour TWA*?"). Another respondent commented that guidelines to ensure similar methods are being used would be helpful.

Potential risks to consider regarding data quality:

- The quality of measurement results is highly variable due to collection, analysis, data entry or reporting;
- The standardisation and coding of data is time-consuming and labour-intensive.

4.4 Privacy and intellectual property:

Overview of issue:

The database will contain privately collected data from consultant occupational hygienists and industry and issues include concerns about potentially handing over personal identifiable information to a third party who was not part of the initial contract, commercial sensitivity (i.e. trade secrets), and fear of being identified and targeted by WorkSafe. Given New Zealand's small size, it is plausible that companies could be identified depending on the level of reporting of aggregate data. The privacy issues will also depend on the database host (discussed in 4.8).

Experiences from international databases regarding privacy and intellectual property:

The majority of overseas databases do not include personal identifiers (and no company names for some databases) and therefore privacy issues are not a concern according to some database contacts. The CWED project addressed privacy concerns in MoU agreements with data holders specifying issues of confidentiality, use, and access. In the U.K., the intention was for the NEDB to include industry data; however, the 2006 survey of HSE staff thought that there was a lack of willingness for industry to provide information because companies were concerned about being identified and targeted by the HSE.[52]

The survey of database contacts did not reveal many intellectual property concerns. In Norway, clients who engage STAMI's monitoring services own the collected data but STAMI has the right to use the de-identified data for research and to create summaries and statistical overviews *(Ragnhild Ostrem, database contact survey, January 2021)*. In South Korea, trade secret issues were identified as a barrier. The potential disclosure of trade secrets was also identified as a barrier for centralising the WC Insurers data *(Taylor Shockey, personal communication, November 2020).*

Responses from stakeholder survey regarding privacy and intellectual property:

Respondents were asked if there are specific privacy requirements that would need to be addressed. Of 13 respondents who answered the question, 4 mentioned that the inclusion of identifier (e.g. worker or company names) information would be a concern or would need to be removed, whereas

1 respondent didn't see any privacy issues for contributing anonymised data. Two respondents mentioned the Privacy Act, 2 said that non-disclosure agreements would be a barrier, 2 said "yes" but with no further information and 1 said that they would not contribute due to clients' fear of being prosecuted.

Respondents were also asked if there are intellectual property issues that would need to be addressed, and in general this appeared to be less of an issue compared to privacy issues. Of 12 respondents, 3 said there were no issues, 6 said "*maybe*", 2 said that company intellectual property would be an issue and 1 commented that New Zealand was so small that even if data were anonymised, it would be possible to still identify companies. One respondent queried whether there will be contractual agreements with data holders that covers issues of consent, uses of the data, and whether the data will be "*sold*" overseas.

Potential risks to consider regarding privacy and intellectual property:

- Data holders and their clients will not contribute data due to privacy or intellectual property concerns;
- The process governing user access and ensuring confidentiality is preserved is onerous;
- The lack of data security provisions decreases trust in the database.

4.5 User Access:

"In a time of scarce resources, any new database must meet or exceed a variety of user needs to receive support, especially financial support" (Cole 1995)[99]

Overview of issue:

The limitations discussed above such as selection bias and inconsistent data quality all increase the risk of the data being misinterpreted and therefore the type of user who can access the data is an important consideration. User access also affects other issues such as privacy and intellectual property. Data contributors need to be considered as key data users.

Experiences from international databases regarding user access:

For the majority of databases, access is restricted to parties outside the owner organisation. The risk of misinterpretation and potential lack of consideration for contextual factors are reasons that MEGA and COLCHIC are not readily accessible to external parties. In South Korea, OSHRI grants access to researchers through a "data handling process". The U.S. IMIS data was able to be accessed via the Freedom of Information Act and one of the aims of CWED 2.0 is to make the database more widely available to researchers and policymakers.

Responses from stakeholder survey regarding user access:
Of 16 respondents, 75% thought that the database should be open access with restrictions (either choosing the open access option or all options). At the NZOHS members' meeting, concern was expressed about the risk of a database replacing occupational hygiene services if it was accessible to the public as well as the potential for misinterpretation of the data.

Risks to consider regarding data access:

- The database is not user friendly or does not meet user needs and is therefore under-utilised;
- The database is not utilised by a wide variety of end users;
- Data is misused or misinterpreted;
- Data is not of use to those that generated the data.

4.6 Consent:

Overview of issue:
The data will be privately collected measurements from consultant hygienists and industry and thus consent has not been explicitly obtained to hand over data to a third party to be stored in a national database.

Experiences from international databases regarding consent:
For the majority of administrative databases of regulator data, consent for OEM data to be included in the database is not required. Even in Finland and Norway, where institutes provide monitoring services, it does not appear that customer consent is required to store OEM data.

Responses from stakeholder survey regarding consent:
Of 12 respondents, 100% reported that they would need to seek consent from their clients to contribute newly collected data. Respondents were asked if they would need to seek consent to contribute anonymised historical data. Of 7 respondents who answered the question, 100% said that they would need to seek consent from their clients.

Potential risks to consider regarding consent:

- Clients do not consent for their data to be included in the database;
- Data holders and their clients are unclear what they are consenting to.

4.7 Funding/costs and sustainability:

Overview of issue:

OEM databases are generally costly to establish and maintain. In addition, the database will not belong to the regulator or insurance provider and therefore short and long-term funding will be a critical issue. The sustainability of the database will also depend on continued buy-in from data holders.

Experiences from international databases regarding cost, funding and sustainability:

The overseas databases are funded by Government funds, insurance organisations or specialist OHS institutes. The exception is the most recent database, the CWED which was funded by research grants. A long-term stable funding source for CWED 2.0 is yet to be established *(Hugh Davies, personal communication, January 2021)*.

The resource requirements in terms of both financial costs and human resourcing for the international databases are presented in Table A2.4. The indicative costs are hard to compare due to the vastly different requirements of the databases (e.g. for the CWED, a large amount of historical paper records had to be abstracted and entered into an electronic database) and few database managers provided actual cost estimates; however the few that did are mentioned below.

a) Establishment

Only a few database managers gave responses to the establishment costs for the databases, which reflect costs at different points in time. At the time of set-up, France's COLCHIC establishment cost was ~400,000 Euros and the CWED was estimated between 500,000-1million Canadian dollars. The 2006 NOHSAC report previously presented an estimate of Singapore $1 million for their National Database for Noise and Chemical Exposure.[1]

The human resource requirements for establishment of the CWED, which initially involved retrospective data collection, was described as "*high*" and the largest efforts were for data abstraction from paper records and coding. The set-up of both Germany's MEGA and France's

COLCHIC involved 4 staff, including 1-2 IT developers, although these are very large databases. The set-up of the SUVA database in Switzerland involved 3 laboratory staff and 4 software engineers totalling ~900 hours.

The UK feasibility study conducted by Cherrie et al.[29] in 2001 estimated that the costs of identifying data (for three target exposures) from consultants was £4,250 and data collection was £4,040 which equates to ~£41 unit cost per measurement (n=203 measurements); however, this study was conducted 20 years ago.

In Canada, the capacity development report to add historical data from BCMEM and the Government of SK to the CWED estimated that the abstraction cost was less than $4 per measurement;[31] however, CWED researchers had direct access to the data.

b) Maintenance

For COLCHIC, the current estimate for database maintenance is ~30,000 Euros (~$50,700 NZD) and for SUVA the estimate is 35,000 CHF (~$54,000 NZD). The 2006 NOHSAC report presented an estimate of 10,000 Euros per year for Germany's MEGA and $100,000 Singaporean dollars for the National Database for Noise and Chemical Exposure;[1] however, this was ~15 years ago.

The human resource requirements for maintenance also vary widely. For Germany's MEGA, the human resource requirement is 0.5 FTE per year; in France, maintenance involves an IT developer and 2 INRS staff; in Norway, a part-time person is responsible for the maintenance of EXPO, and for SUVA, an estimate of 40 hours per year each for occupational hygienists and software engineers.

Responses from data holders' survey regarding costs

On the assumption that data holders would not allow external parties to directly extract the data, respondents were asked if they would charge for the ongoing contribution of newly collected data. Of 10 respondents who answered the question, 30% (n=3) said yes they would charge, 50% said maybe (n=5), and 20% (n=2) said no. For the extraction of historical data, of the 7 respondents who answered the question, 43% (n=3) said yes, 14% (n=1) said no, 14% (n=1) said "don't know" and 29% (n=2) said "maybe".

Potential risks to consider regarding funding and sustainability:

- The charge-out rate for the extraction of data is significant;
- A considerable amount of effort is expended to set-up the database but a sustainable funding source is not identified to guarantee its longevity.

4.8 Host:

Overview of issue:

The choice of database host will affect the willingness of data holders to contribute to a database i.e. they may be more likely to contribute to a host they trust or regard as representative/independent. Other important considerations include the long-term funding structure of the host as well as the available support infrastructure.

Table 6: Advantages and disadvantages for different options for an OEM database host

Potential host	% of data holder respondents choosing this option	Advantages	Disadvantages
WorkSafe New Zealand	38% (n=6)	-The highest proportion of respondents chose this option -Direct access could result in wider use of the data for the development of policy and prevention activities -More likely to have adequate support infrastructure (e.g. administrative, IT, legal etc.) -There is multi-disciplinary expertise available across WorkSafe	Data holder/clients may not contribute due to fear of action from WorkSafe, which could be a significant limitation e.g. for the question about potential barriers in the data holders' survey, 3 respondents said that fear of WorkSafe using the data to prosecute companies would be a barrier
Academic institution/University	13% (n=2)	-Independent from Government -Researchers have experience with collating and analysing data sets, and coding etc. -More likely to be used for research -More likely to have international contacts -More likely to have adequate support infrastructure (e.g. administrative, IT, legal etc.)	-The longevity of academic institutes (research centres in particular) is not guaranteed due to the reliance on external funding -Only one of the overseas databases in the review is hosted by a University; however options for an alternative host are being explored
Statistics New Zealand	-	-Experience with hosting the IDI, potential to use the same processes and infrastructure	-Not within their mandate
NZOHS (see 3.1)	19% (n=3)	-Independent from Government -Data holders (i.e. NZOHS members) may be more likely to participate	-Potentially lacking support infrastructure (e.g. administrative, IT, legal etc.)
Health and Safety Association of New Zealand (HASANZ)*	19% (n=3)	-Independent from Government	- Potentially lacking support infrastructure (e.g. administrative, IT, legal etc.)
Private company	6% (n=1)	-Independent from Government	-Clients may be unwilling to contribute data with a private company as a host (e.g. intellectual property issues)

*HASANZ is an umbrella organisation representing health and safety professions in New Zealand.

Experiences from international databases regarding the host:

The overseas databases are hosted by OHS regulators, insurance organisations, and specialist OHS institutes. The Scandinavian institutes such as FIOH in Finland and STAMI in Norway are specialist organisations operating under Government Ministries. They have existed since the late 1940s[66] and 1980s, respectively. The CWED is currently hosted by the University of BC but part of CWED 2.0 involves investigating options for a host suitable for the evolution of the database into a national resource *(Hugh Davies, personal communication, January 2021).*

Responses from stakeholder survey regarding the host:

The data holders' survey asked respondents who should host the database and 38% (n=6) said WorkSafe, 19% (n=3) chose NZOHS, 19% (n=3) chose HASANZ and 13% (n=2) thought an academic institute/University should host the database. One respondent indicated "*other*" but did not have a view on who. Statistics New Zealand was not an option in the survey.

Despite this, concerns about WorkSafe as a host were brought up in the discussion at the NZOHS members' meeting and throughout the data holders' survey; 3 respondents said that fear of WorkSafe using the data to prosecute companies was a barrier and one respondent described a "*disconnect from the regulator*" as a caveat for contribution of their data.

Potential risks to consider regarding the host:

- Data holders don't contribute due to the choice of host e.g. the fear of WorkSafe using the data for prosecution;
- The host's financial sustainability and resources for infrastructure is uncertain.

4.9 IT system

Experiences from international databases regarding IT systems:

The database software systems used by the international databases are presented in Table A2.4. Off the shelf software is used in Norway ("SQL database"), South Korea (MS ORACLE and ACCESS), Switzerland (SAP CRM), Singapore (Power Pivot), and France (IBM DB2). In Germany, the purpose-designed OMEGA (Organisation System for the collection and use of measured data on exposure to hazards at the workplace) is used for data collection and MEGA^{PRO} is used for data storage. The U.S. and Finland have laboratory information management systems but the software is unknown. In the

U.K., the NEDB uses bespoke software written in Massachusetts General Hospital Utility Multiprogamming System (MUMPS). Both Norway and Singapore also have online forms for companies to submit their data. In Canada, a new IT platform is being investigated for CWED 2.0.

Responses from stakeholder survey regarding IT system:

Respondents were not asked specifically about IT systems but several respondents expressed the need for a well-designed system that would minimise the time involved with data extraction. One respondent mentioned that an interface system that would allow uploads from existing software would be key.

5. Recommendations

Overall recommendation for a New Zealand occupational exposure measurements database

Two major tenets of exposure surveillance are the *timely* collection of data and the analysis of data that translates to action, both of which are lacking in New Zealand (Section 3). A common theme from the submissions for the 'Health and Safety at Work Strategy 2018-2028' was that the collection of health and safety data in New Zealand needs to be improved and shared across the wider system. The other methods for occupational exposure surveillance in New Zealand (described in Appendix 4) are not suitable for ongoing surveillance.

As described in Section 2, many countries have occupational exposure measurement databases, most of which have been operating successfully for decades and utilised for a wide variety of purposes. The more relevant examples for New Zealand (i.e. where monitoring is not a regulatory requirement) include the CWED 2.0 and Norway's EXPO Online; however, these are recent or ongoing examples and thus any lessons learnt from their development are unclear at present. In addition, the U.K. has plans to re-develop their national database.

New Zealand has the advantage of its smaller size (i.e. there are not multiple states/jurisdictions), NZOHS as collaborators on this scoping project, and the current momentum for the widespread sharing and analysis of anonymised data promulgated by the advent of the IDI.

Whilst it is possible that the NZOHS members that took part in the data holders' survey are the ones that are keen to contribute to a national database, the overall response was positive. All respondents thought that a database should be set up in New Zealand and the average willingness to contribute rating was 7.1 (from 16 respondents). The majority were keen for the database to be open access suggesting that respondents view the database as a resource to be used for the common good in New Zealand.

Based on the lack of New Zealand-specific occupational exposure data, successful examples from overseas, and results from the data holders' survey, an occupational exposure measurement database for New Zealand is recommended. The proposed OEM database primarily focuses on dust/chemical factors and noise exposure. Several overseas database contacts mentioned that hygienist expertise is required for data entry and coding. A New Zealand database would capitalise on already collected data which would be extracted by occupational hygienists.

Specific recommendations for development and operation of a New Zealand occupational exposure measurements database

Several respondents from the data holders' survey emphasised that the database would need to be part of a well-designed system in order for data holders to feel confident in handing over their monitoring results. Specific recommendations are outlined below for different facets of a "system", including an indication of the barriers/issues from section 4 in [*brackets*] where relevant.

5.1 Development of the database

1. **Clearly define main use(s)/main users**

   As illustrated in Table 5, a wide variety of stakeholders can benefit from an OEM database; however, a few key uses/users need to be prioritised to guide the data collection and data access agreements. The development of a database proof of concept will enable potential data holders to better understand the purpose, functionality, and how the data will be presented.

2. **Involve stakeholders (especially end-users) throughout the entire process**

   The engagement of stakeholders (e.g. NZOHS) throughout the entire process is crucial for several reasons including determining core database requirements, ensuring the output meets user needs, and for fostering buy-in for participation *[data acquisition and data quality]*. A theme from the database contacts survey was that input from experts for both the development and data entry was important.

3. **Involve managers/contacts of international databases in the development process**

   The lessons learnt from the establishment and ongoing development of overseas databases are invaluable for the development of a New Zealand database and further engagement and involvement of international contacts would be extremely useful. Several of the international database managers that responded to the questionnaire expressed an interest in the scoping project or in providing further assistance. Dr Dorothea Koppisch from MEGA has expressed interest in cooperation regarding the establishment of a measurement, documentation, and storage system for exposures to hazardous substances and biological agents. Other recommended advisors from systems under parallel development include Hugh Davies from CWED, Taylor Shockey from the U.S., and Norway's EXPO. It will also be useful to keep abreast of developments by the HSE U.K. (especially from a technical standpoint) who are investigating options to re-develop the NEDB. In particular, an analytical platform is being designed as part of the OccECIS project which aims to integrate structured and unstructured data sources, statistical inference (with the aim of creating an alert system for emerging risks from statistical patterns in the data), and data visualisation *(Damien McElvenny, personal communication, December 2020)*.

4. **The IT infrastructure of the database needs to be flexible and user-friendly**

   The system needs to be designed to be flexible and sustainable for longitudinal data collection as well as user-friendly. The majority of respondents from the data holders' survey reported that they would or potentially would charge for the extraction of data and thus the data interface/entry system needs to be easy and quick to use.

5.2 Governance

5. **Host**

   The recommendation for a host is not straightforward and further consultation is recommended. However, one option could be for an academic institute to host the pilot/conduct a feasibility study and for NZOHS to be the database host if adequate resources are provided.

6. **Establish a data governance group with two sub-committees**

   a. Steering committee (strategic guidance)

   b. Data Management committee (operational oversight)

   These are based on recommendations made for the CWED's data mobilisation project *(Hugh Davies, personal communication, January 2021)*. For the CWED 2.0, the mandate of the data steering committee includes identifying sources of funding *[funding and sustainability]*, increasing awareness of the database amongst potential data use partners *[data acquisition]*, and identifying opportunities to link to other agencies *[funding and sustainability]*. The mandate of the data management committee includes the development and oversight of protocols, policies, and guidelines *[representativeness, data quality, user access, consent, privacy]*. Draft terms of reference templates for the two committees are available from the mobilisation project. There is potential to include international advisors in the data governance group.

5.3 Data acquisition

7. **Formally identify data holders**

   Conduct a more thorough environmental scan of potential data holders.

8. Collect data from NZOHS members in the first instance

The reasons for this are two-fold: 1) as members of a professional association, they are more likely to generate good-quality data; 2) they are more accessible as the NZOHS are collaborators on this project *[data acquisition and data quality]*.

9. Identify opportunities to increase participation from data holders

The aim should be to include as many data holders as possible *[representativeness].* This could include outreach event(s) to raise awareness and encourage participation (e.g. NZOHS annual conference in May). Potential incentives for contributing data should also be explored e.g. accreditation programmes, ACC rebates, or other value-added benefits. More than half of respondents in the data holders' survey stored their data in individual electronic reports and thus the digitisation of results in one data file could be offered as a value-added benefit. A website should be created for generating awareness as well as for knowledge translation and exchange.

10. Develop a standard template for a data sharing agreement and data acquisition guidelines

Develop a data sharing agreement template specifying confidentiality, authorised users, and access etc. [*data acquisition and privacy*]. The data sharing agreement from the CWED is a useful template. This could also include an example of a clause to be included in consent forms for clients.

11. Approach relevant research groups that collect OEM data

This would involve developing an example consent clause for prospectively collected research data and investigating ethics requirements for the ascertainment of anonymised historically collected data.

5.4 Data management and collection

12. Employ a dedicated data manager (at least part-time)

A dedicated data manager would provide crucial oversight and coordination of incoming data. Quality control of incoming data is also crucial [87] and the data manager could potentially carry out quality control checks as well as approve applications to use the data.

13. Develop protocols for data collection and management as well as appropriate data security provisions

The protocols developed for overseas databases (e.g. CWED, IMA-DMP) will be useful guides. Personal identifying information should not be stored in the database *[privacy]*. If practical, allocate a unique code to each individual worker to allow for the analysis of repeated measurements.

14. Develop a list of core variables including ancillary variables

Develop a list of core data requirements *[data quality]* including ancillary information to aid interpretation of the data and to help understand the variability of exposure levels *[representativeness]*. In particular, the collection of measurement reason (e.g. for compliance or investigative (i.e. solving a specific problem) or for evaluating a control measure) and information about the sampling strategy are key variables. Consult with hygienists with regards to the collection of the number of ancillary variables versus practicality in the field.

Two widely referenced guidelines were developed in the 1990s, which proposed a list of minimum variables to enable the proper interpretation of exposure measurements.

1) The Working Group on Exposure Registers in Europe[63] compared exposure databases (MEGA, EXPO, COLCHIC, ATABAS, NEDB, Montreal) in the mid-1990s. The working group identified 10 key categories (39 data elements) to describe and assess exposure.

2) A joint task group of the American Conference of Governmental Industrial Hygienists (ACGIH) and the American Industrial Hygiene Association (AIHA) in 1996 recommended specific standardised data elements for air and noise sampling.[64] They proposed 13 key categories (134 data elements; see Table 7).

Babik et al.[44] recently developed a list of core fields in consultation with experts (including NIOSH and WC Insurers industrial hygienists) after reviewing the fields contained in WC Insurer's air and noise survey forms. The list was compared to the ACGIH-AIHA recommendations which shared 85-90% of data with the ACGIH-AIHA Task Group list from 1996 suggesting that the recommended data elements are still relevant today.

Table 7: Minimum data requirements proposed by two working groups

Working Group on Exposure Registers in Europe (Rajan 1997)[63]	**Joint ACGIH-AIHA Task group (1996)[64]**
Premises	Facility/Site information
Workplace	Work area information
Worker activity	Employee information
Product	Process and operation information
Chemical agent	Chemical agent information
Exposure modifiers (exposure pattern and pattern of control)	Exposure modifier information
Measurement strategy	Sample information
Measuring procedure	Sampling device information
Results	Chemical exposure results
Reference	Noise exposure results
	Survey tracking information
	Engineering controls information
	Personal protective equipment information

These guidelines have been widely used for international databases (e.g. the data structure of COLCHIC was overhauled in 2002 in consultation with these guidelines). The investigation of ancillary variables, for example using multivariable statistical analysis, is commonly carried out to attempt to address the issue of representativeness of OEM databases.

15. Ensure standardisation of data and harmonisation of coding *[data quality]*

- Develop a standardised data collection online form with in-built validation checks
- Develop clear guidelines for data collection and entry e.g. reporting of NDs and the reporting of both high-level and low-level exposures *[representativeness]*
- Develop data dictionaries
- Develop clear guidelines for coding of data, including specified classifications for occupation, industry, and task
- Potentially engage occupational hygienists to assign quality scores
- Run training workshops for data collection and entry
- Conduct regular quality audits

16. Build-in regular reporting

The ongoing benefits from the database (particularly for those who generated the data) need to be clear as the sustainability of the database will depend on continued buy-in from data holders. Regular feedback or reporting could be a strong incentive to contribute data (as is the case for the IMA-DMP). There are also possibilities to develop information access tools of aggregate data.

5.5 Data access:

17. Develop data access rules and user guides

Consult with end users to ensure that user needs are met and develop user guides and rules (including privacy and ethical issues) with oversight provided by the data management committee.

18. Three-tiered access system *[privacy]*

The proposed three-tiered system is based on the CWED 2.0 data access system who have developed policies and procedures for each tier:

Tier 1: Raw data release for researchers, required to submit application similar to IDI process.

Tier 2: Moderately aggregated data e.g. the CWED have developed a 'CWED Prototype Aggregate interface' which allows for data queries using a secure interface for agencies and institutions with "pre-determined reporting needs".

Tier 3: Highly aggregated data for the general public e.g. an annual report of summary data.

Some of the elements of the IDI system governing use and outputting of data could be implemented such as mandatory confidentiality training and output checking (e.g. all counts <6 are suppressed). The development of the IDI has also been accompanied by various user forums and data sharing networks. For example, the Virtual Health Information Network is a network of IDI users from multiple Government departments, academic institutions, and private companies who share advice, code, aggregate results etc.

5.6 Future use and sustainability

19. View the database as a complementary tool in a toolbox of exposure surveillance methods

There are many possibilities to use the database in combination with other exposure surveillance methods (e.g. the 'OccIdeas' survey) or exposure assessment tools. The problem of OEM databases being unrepresentative of all workers and workplaces led to the development of the FINJEM which was designed to overcome some of the limitations of traditional exposure databases.[66] The utility of the proposed database could be further increased by combining the data with locally developed JEMs for occupations/industries that may have a lower representation in the database *[representativeness]*. Each exposure surveillance method has its limitations and thus it is important to consider an exposure database as a complementary method. An integrated system for exposure surveillance in New Zealand has been previously recommended by NOHSAC.[94]

20. Conduct a preliminary scan of potential funding sources (e.g. ACC, WorkSafe)

Examples of potential funding sources include WorkSafe and ACC. As mentioned above, the mandate of the CWED 2.0 data steering committee includes identifying sources of funding, identifying opportunities to link to other agencies, and increasing awareness of the database amongst potential data use partners. With regards to the latter, the more users that benefit from the data, the easier it will be to sustain the database.

6. Options for a New Zealand occupational exposure measurement database

This section includes options for a New Zealand occupational exposure measurements database including prospective vs. historical and priority vs. all agents/industries.

6.1 Prospective vs. historical data

The options proposed below are not all mutually exclusive and could potentially be implemented in phases. The advantages and disadvantages of historical vs. prospectively collected data are outlined in Table 8.

Option 1: historical data collection only

Option 2: prospective data collection only– active or passive?

Option 3: both historical and prospective data collection, either together or phased

Table 8: The advantages and disadvantages of historical vs. prospective data collection

Option 1: historical data collection only	
Advantages	**Disadvantages**
Encourages the preservation of irreplaceable historical data including digitisation, standardisation, and coding	Information may not reflect current exposure patterns and therefore has limited use for the current OHS context
Useful for studying occupational diseases of long-latency e.g. cancer	Historical data is likely to vary widely in storage type (e.g. paper-based vs. electronic) and data quality
Useful for examining long-term trends in exposure levels	The extraction of historical data could be expensive.
Useful for evaluating the impact of regulatory controls over time e.g. historical changes to WES	It is unclear whether retrospectively collected measurement data can be included given research ethics and customer consent requirements. This would require extra work to retrospectively seek consent or determine whether it is required.
Multiple New Zealand data sources are available e.g. industrial hygiene data, research data.	Data retention policies will vary
Useful for epidemiological research involving historical exposures	Verification of reliability will be difficult if not impossible.
	Expert judgement may be required to assess/review historical data and its suitability for use
Option 2: prospective data collection only	
This option provides the opportunity to ensure standardisation of all protocols and data collection and coding from the outset and	Facilitate the collection of baseline data but will not be able to study trends in exposure levels over time for several years

theoretically less time spent extracting data. There is also a lower risk of variable data quality	
All data collection will be electronic;	The impact of regulatory controls will not be observed for several years
It is unclear whether retrospectively collected research measurement data can be included given research ethics requirements. This option can include provision for future research studies to include a clause in their consent forms	Not useful for studying occupational diseases of long latency in the immediate future
The costs of extracting historical data can be high. The recommendation from the findings of the UK feasibility study by Cherrie et al.[29] was to obtain anonymised prospective measurements through occupational hygiene consultants and industrial organisations;	Data will need to be collected for the database at specified intervals which will be resource-intensive
Enable the *timely* identification of determinants of exposure as targets for exposure prevention efforts[34]	
The ability to associate trends in occupational illness and injury claims with measured exposure levels[34]	

Option 1: historical data collection only

This option would be useful from a research perspective but not for the current data needs of WorkSafe. Furthermore, none of the respondents from the data holders' survey were willing to contribute historical data only. One respondent commented that there will be less available samples for data older than 5 years old and on average, historical data were retained for 9 years. A small number of respondents answered a question about what the likely time/cost commitment of data extraction would be and 60% (n=3) reported that the likely time/cost extraction would be significant, 20% (n=1) reported moderate, and 20% (n=1) reported minimal. Forty-three respondents (n=3) reported that they would charge for extraction of historical data, 29% (n=2) said maybe, and 29% (n=2) said no or "don't know".

Option 2: prospective data collection only– active or passive?

In the data holders' survey, 60% of respondents were willing to contribute prospective data only. The collection of prospective data would be cheaper, the data collection and reporting processes can be standardised from the outset therefore improving data quality, and obtaining client consent would be easier for prospectively collected data. This option was also recommended from the UK feasibility survey conducted by Cherrie et al.[29], which concluded that the most cost-effective option would be to acquire prospective anonymised data from occupational hygiene consultants and industrial organisations.

Within option 2, there are two sub-options:

> Sub-option 1: Active data collection: develop data acquisition agreement to access data at specified intervals.
> Sub-option 2: Passive data collection: e.g. both EXPO Online and CWED have online forms for industry to submit data

The active data collection option would require a high degree of compliance from data holders to contribute at regular intervals; however, the passive data collection option may exacerbate issues of representativeness and data quality.

Option 3: both historical and prospective data collection
This option would include the advantages described above for both options and provide a benchmark for both current and past exposures. The CWED initially comprised of historical data from regulatory bodies across Canada and was established as part of the CAREX Canada project. However, CWED 2.0 is currently being developed and aims to include prospective data from industry.

In the data holders' survey, 40% (n=6) reported that they were willing to contribute both prospective and historical data. Of the respondents who were also willing to contribute historical data, 100% (n=7) said that they would need to seek consent from their clients and 60% (n=3/5 respondents that answered the question) reported that the likely time/cost extraction would be significant. The average years of retention of historical data was also just under a decade. Thus, the collection of both historical and prospective data would be the most expensive and resource-intensive option; however this will depend on consent requirements for historical data and the data acquisition process.

6.2 Options for specific agent(s)/industries
Most of the large international databases in the review do not target specific exposures or industries; however, several exposure-specific (e.g. Evalutil) and industry-specific databases (e.g.

EXASRUB, PAPDEM) exist. The options for targeting specific exposure(s) or industries, either for a pilot or full database, are discussed below.

Options A: One or few priority exposure(s)

Examples of priority agents include RCS (a priority focus area for WorkSafe and the example being used by OccECIS in the UK), carcinogens, or the ExpoSYN agents (i.e. RCS, asbestos, chromium, nickel, and polyaromatic hydrocarbons). Information on the most frequently measured agents for some of the international databases was available (Table A2.2).

The 2006 NOHSAC report reviewing Australian and New Zealand exposure surveillance systems outlined criteria for prioritising which exposures should be the focus of exposure surveillance, including long latency disease, magnitude of the problem (i.e. number of people affected), absence of adequate data, and whether support exists to act on this issue.[93]

The HazChem@Work pilot project described in section 2 assessed the available data sources for occupational exposure to chemical agents in EU member states. As part of the project, a list of chemicals was developed based on three stages: Stage 1: scores we assigned based on exposure potential (based on typical working tasks etc.), nature of the hazard to human health, and assumed consumption (and therefore high exposure prevalence); stage 2: existence of an OEL (and therefore previous targeting by regulation) and refined step 1 revisions; and step 3: data availability. The final list included 100 substances and the top 5 scoring substances were benzene, formaldehyde, quartz (RCS), lead, and chloroform. [58]

Advantages:

- A targeted approach would focus data acquisition and collection efforts;
- Carcinogens (including RCS) are a priority focus area for WorkSafe;
- For ExpoSYN, Peters et al.[86] noted that the five agents selected were relatively prevalent in workplaces and also frequently measured and that "*it would be difficult to find sufficient measurement data to study exposure to less prevalent agents*".

Disadvantages:

- The targeting of priority exposures will be dictated by what data is collected by the data holders' that contribute to the database. There may not be enough samples for specific agents collected to achieve critical mass for analysis. In the data holders' survey, the most commonly sampled agents were noise (83% of respondents), inhalable dust (72%), RCS (56%), and respirable dust (56%).

- New or emerging hazards may be missed;
- This approach will not contribute to the knowledge of the range of risks across all New Zealand workplaces;
- Multiple hazards are present in many workplaces;
- The effort expended to negotiate access agreements with data holders, set up data extraction etc., may mean that this a less cost-effective option.
- The UK feasibility study conducted by Cherrie et al.[29] concluded that it is not cost-effective to collect data for a limited number of target substances.

Option B: One or few industries e.g. selected high-risk industries.

Examples include WorkSafe's priority industries including agriculture, forestry, construction, manufacturing, and transport.

Advantages:

- A targeted approach would focus data acquisition and collection efforts;
- The data holders' survey revealed that the most commonly sampled industries were manufacturing, construction, and transport, postal and warehousing. Thus, the majority of measurements are taken from these industries anyway.

Disadvantages:

- New or emerging industries of risk may be missed;
- Will not provide a full picture of where the problem areas/occupations/industries are.

6.3 Options for a pilot:

Options for a pilot also include options A and B described above - option A as a pilot is described below.

One or few priority exposure(s): RCS

The targeting of a specific exposure or exposures is a practical option for testing the data acquisition and collection process and for the development and testing of the database. In particular, RCS is a priority focus area for WorkSafe and the example being used by OccECIS in the UK as well as an ExpoSYN agent. In addition, it is a commonly collected agent according to the data holders' survey.

This would also fit in with other proposed WorkSafe programmes including RCS and accelerated silicosis in engineered stone workers.

Sub-group of NZOHS e.g. North Island only

A significant amount of work will go into negotiating access agreements. The process could be fine-tuned by focusing on a sub-group of NZOHS.

Recommendations for options

Given the cost and resources required to extract historical data, the option of collecting prospective measurements, preferably using the active collection approach, is recommended with the added advantage of establishing standardised collection and coding protocols from the outset. The data acquisition process will involve negotiating access with data holders as well as obtaining client consent and given this effort, a database of all agents and industries is recommended but with a pilot/feasibility study carried out for respirable crystalline silica.

6.4 General indicative costs

The likely costs for the development of an occupational exposure measurement database in New Zealand include:

- Extraction of data by occupational hygienists potentially at charge-out rates, although the time commitment will depend on the data entry design
- A data/project manager
- Administrative costs e.g. for developing protocols, organising governance meetings, negotiating access agreements etc. including legal advice
- IT costs including the development and testing of the database and website development
- Training workshops and outreach events

6.5 Feasibility study

The proposed next step is a feasibility study for implementing an OEM database in New Zealand, including options for a database and the issues to be addressed. The recommended steps for a feasibility study include:

1. Conduct further consultation for database host options
2. Conduct a scan of data holders, including the likely representativeness of the data

3. Conduct further consultation with data holders to ascertain information on what data variables are available, data storage, privacy and client consent requirements.
4. Investigate options for efficient data capture and storage e.g. automated insertion scripts, text mining/information extraction algorithms for PDFs, webapp extraction capabilities etc.
5. Investigate incentives for contributing data e.g. for the development of the CWED, the approach to data holders described "value-added" raw data to the original data holder such as assistance with data management, standardisation, and coding of data and advice/assistance of conversion of paper files into electronic format. Norway's EXPO Online provides free and secure data storage for companies. From the data holders survey, 56% of respondents reported that their data were stored in individual electronic reports (i.e. PDFs) rather than electronic databases and 56% reported that they use standard forms which is an opportunity to add value.
6. Explore potential funding sources (e.g. ACC, WorkSafe).

6.6 Next steps and timeframe

The steps outlined below are proposed for a 12-18-month timeframe following the feasibility study. Stakeholders/potential end users should be consulted/engaged throughout the entire process.

Administrative and IT:

1. Finalise and set terms of reference for database host
2. Establish and set terms of reference for a data governance group ideally comprised of two sub-committees
3. Develop a database proof of concept including purpose, functionality, and potential data uses
4. Develop a data sharing agreement template and data acquisition guidelines
5. Employ a data/project manager
6. Develop a data collection and management protocol
7. Develop a data security protocol
8. Develop a list of core data requirements, data dictionaries, and coding classifications in consultation with end users (e.g. hygienists, researchers)
9. Develop data user access rules and procedures (including application form and approval process)
10. Investigate options for long-term IT platform and storage. The system needs to be flexible and sustainable for longitudinal data collection and user-friendly

11. Commission the development and testing of the database including an online data collection form that guides data extraction with built-in validation checks. Considerations include high data security and potential external access to multiple users with different levels of access privileges
12. Develop website (either stand-alone or as part of host organisation)

Data collection:

13. Organise outreach event(s) to raise awareness and gain interest in contributing to the database in conjunction with NZOHS;
14. Approach data holders and negotiate data sharing agreements, piloting the agreement process with a small number of data holders;
15. Approach relevant research groups to contribute data;
16. Run training workshops

6.7 Future possibilities

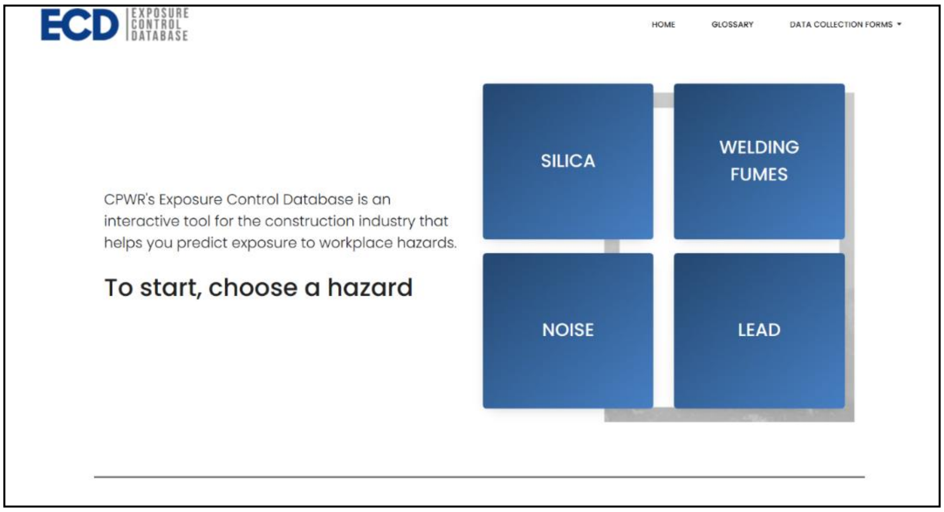
The system must be flexible to be able to incorporate other hazards e.g. psychosocial and other physical hazards as well as emerging hazards (e.g. nanotechnology). There are also several examples of control databases including the NIOSH Engineering Controls Database[viii]. The proposed OccECIS system in the U.K. also intends to collect comprehensive control information which is also a future possibility for the proposed New Zealand database.

The database manager from South Korea was a strong advocate for linking exposure and health surveillance systems. A New Zealand exposure database could potentially be linked to routinely collected health data within the IDI.

There is also potential to develop information access tools. For example, the Center for Construction Research and Training (a non-profit organisation in the U.S. which aims to reduce occupational illness and injuries in the construction industry) has developed an exposure control database where external companies can enter their exposure data (Figure 9) and an interactive tool that provides aggregate results for specific exposures.[ix]

[viii] https://www.cdc.gov/niosh/engcontrols/ecd/default.html
[ix] https://ecd.cpwrconstructionsolutions.org/

Figure 9: The Center for Construction Research and Training's Exposure Control Database



6.8 Final recommendation:

A wider system of mandatory exposure monitoring and reporting to a centralised OEM database should ultimately be implemented for New Zealand. For example, the South Korean model of mandatory exposure monitoring once every 6 months (or once every year if the concentration level of all chemical substances is below the occupational exposure limit) and the centralisation of this data using an electronic collection system is a useful example. In the meantime, an occupational exposure measurement database would be an invaluable resource for a wide variety of stakeholders in New Zealand and the centralisation of data already being collected is a cost-effective approach. The recommendation from this report is to establish a database with a focus on prospectively collected data from NZOHS members initially and to standardise all processes pertaining to data collection from the outset. The addition of historically collected data from both NZOHS members and researchers can be investigated at a later stage. A further recommendation would be to conduct a feasibility study and then pilot test the database using respirable crystalline silica as an example.

Glossary:

ACGIH	American Conference of Governmental Industrial Hygienists, a non-governmental scientific organisation with the aim of advancing occupational and environmental health
AIHA	American Industrial Hygiene Association
BEI	Biological Exposure Indices: the guidance values used in New Zealand for assessing biological monitoring results
EN482	the European Standard specifying the general requirements for the performance of procedures for the measurement of workplace chemical exposure
EN689	the European Standard specifying a strategy to perform representative measurements to demonstrate compliance with occupational exposure limit values
HSWA	Health and Safety at Work Act 2015, New Zealand's principal legislation governing workplace health and safety
JEM	Job Exposure Matrix: a type of expert information system which typically uses multiple data sources to create an exposure assessment tool that can be applied to a wider population. JEMs are cross-classifications between a list of job titles and a list of agents (and sometimes a third axis of time) to which persons carrying out the job may be exposed, and usually involve expert assessment of exposure based on multiple data sources, including exposure measurement data
LOD	Limit of Detection: the lowest concentration of an analyte that can be reliably detected
ND	Non-Detects: the amount of analyte collected is too little to be detected by the analytical method
PCBU	Person conducting a business or undertaking
PEL	Permissible Exposure Limits: the legal limits for maximum exposure to an airborne substance published by the Occupational Safety and Health Administration (OSHA) in the U.S.
OEL	Occupational Exposure Limit: the upper acceptable concentration level of a hazardous (typically airborne) substance. Sometimes referred to occupational exposure limit value (OELV)
SMEs	Small to Medium Enterprises, currently defined as a company with 19 or fewer employees
STEL	Short-term Exposure Limit: a limit value above which exposure to a chemical substance should not occur and usually relates to a 15-minute reference period. The aim of a STEL is to prevent adverse health effects and other unwanted effects due to peak exposure that may not be controlled by the application of an 8-hour TWA limit
TLV	Threshold Limit Value: health-based guidance values recommended by the American Conference of Governmental Industrial Hygienists. The values indicate exposure levels at or

below which nearly all workers may be repeatedly exposed without experiencing adverse health effects

TWA Time Weighted Average: the average airborne concentration of a substance calculated over an 8-hour working day

WES Workplace Exposure Standards: the health-based guidance values used in New Zealand for airborne substances. The values indicate the concentration below which nearly all workers should not experience adverse health effects

Acknowledgements:

Thank you to the following people who contributed to the report: *WorkSafe:* Sue Cotton, Simon Bidwell, Kerry Cheung, Al Threlfall, and Simon Buckland; *NZOHS*: Robert Murray, Kim Pearson-May, and all of the participants of the data holder survey as well as those that attended the NZOHS members' meeting for their valuable feedback; *Research Centre for Hauora and Health*: Andrea 't Mannetje, Jeroen Douwes, and Dave McLean for their useful comments on the report; *International*: Lothar Lieck from the *European Agency for Safety and Health at Work* for his comments on the review of international databases; the database managers and associates who kindly took the time to complete the questionnaire for international databases: Hugh Davies, Gina Scott, Dorothea Koppisch, Gautier Mater, Tapani Tuomi, Lars Andrup, Ragnhild Ostrem, Seunghyun Park, Evelyn Koh, Patrick Steinle, Nenad Savic, Peter Baldwin, and Chris Keen; and the following international experts who provided useful advice and contacts: Hans Kromhout, Amy Hall, Paul Demers, Jerome Lavoue, Damien McElvenny, Martie van Tongren, Ioannis Basinas, and Yiqun Chen.

References:

1. VIOSH Australia at the University of Ballarat. International Review of Surveillance and Control of Workplace Exposures. National Occupational Health and Safety Committee (NOHSAC) Wellington, New Zealand. 2006;Technical Report 5.
2. Lemen R. Role of Exposure Databases in Disease Surveillance and Occupational Epidemiology. Appl Occup Environ Hyg. 1995;10:400-1.
3. Shockey TM, Babik KR, Wurzelbacher SJ, Moore LL, Bisesi MS. Occupational exposure monitoring data collection, storage, and use among state-based and private workers' compensation insurers. J Occup Environ Hyg. 2018;15:502-9.
4. Armstrong A. Past and Current Approaches to Occupational Exposure Databases in the Private Sector. Appl Occup Environ Hyg. 1995;10:257-63.
5. Demers P, DeBono N, Arrandale V, Keefe A. Options for Tracking Occupational Disease and Exposure in Ontario. Occupational Cancer Research Centre, Cancer Care Ontario, Toronto, Canada. 2018.
6. Hall AL, Peters CE, Demers PA, Davies HW. Exposed! Or not? The diminishing record of workplace exposure in Canada. Can J Public Health. 2014;105:e214-7.
7. Kromhout H. Hygiene Without Numbers. Ann Occup Hyg. 2016;60:403-4.
8. Zalk DM, Nelson DI. History and evolution of control banding: a review. J Occup Environ Hyg. 2008;5:330-46.
9. Koppisch D, Schinkel J, Gabriel S, Fransman W, Tielemans E. Use of the MEGA exposure database for the validation of the Stoffenmanager model. Ann Occup Hyg. 2012;56:426-39.
10. Schinkel J, Ritchie P, Goede H, Fransman W, van Tongeren M, Cherrie JW, Tielemans E, Kromhout H, Warren N. The Advanced REACH Tool (ART): incorporation of an exposure measurement database. Ann Occup Hyg. 2013;57:717-27.
11. Burdorf A, Van Tongeren M. Commentary: variability in workplace exposures and the design of efficient measurement and control strategies. Ann Occup Hyg. 2003;47:95-9.
12. Cherrie JW. The beginning of the science underpinning occupational hygiene. Ann Occup Hyg. 2003;47:179-85.
13. LaMontagne AD, Herrick RF, Van Dyke MV, Martyny JW, Ruttenber AJ. Exposure databases and exposure surveillance: promise and practice. AIHA Journal. 2002;63:205-12.
14. Stewart P, Rice C. A Source of Exposure Data for Occupational Epidemiology Studies. Appl Occup Environ Hyg. 1990;5:359-63.
15. Botkin A, Conway H. Relevance of Exposure Data to Regulatory Impact Analyses: Overcoming Availability Problems. Appl Occup Environ Hyg. 1995;10:383-90.
16. Lippmann M. Exposure Data Needs in Risk Assessment and Risk Management: Database Information Needs. Appl Occup Environ Hyg. 1995;10:244-50.
17. Gomez M, Rawls G. Conference on Occupational Exposure Databases: A Report and Look at the Future. Appl Occup Environ Hyg. 1995;10:238-43.
18. Creek K, Schinkel J. Workshop on Key Data Needs for an Occupational Exposure Database, Session II. Appl Occup Environ Hyg. 1995;10:408-10.
19. Gomez M. A Proposal to Develop a National Occupational Exposure Databank. Appl Occup Environ Hyg. 1993;8:768-74.
20. Clerc F, Steinhausen M, Bertrand N, Vincent R, Gabriel S, Van Gelder R. Comparison of Formaldehyde Exposure Measurements Stored in French and German Databases. Gefahrstoffe - Reinhaltung der Luft. 2015;75:119-26.
21. Linch KD, Miller WE, Althouse RB, Groce DW, Hale JM. Surveillance of respirable crystalline silica dust using OSHA compliance data (1979-1995). Am J Ind Med. 1998;34:547-58.
22. The University of British Columbia. Canadian Workplace Exposure Database. http://cwed.spph.ubc.ca/ Accessed 30 November 2020.

23. Kauffer E, Vincent R. Occupational exposure to mineral fibres: analysis of results stored on colchic database. Ann Occup Hyg. 2007;51:131-42.
24. Mater G, Paris C, Lavoué J. Descriptive analysis and comparison of two French occupational exposure databases: COLCHIC and SCOLA. Am J Ind Med. 2016;59:379-91.
25. Okun A, Cooper G, Bailer AJ, Bena J, Stayner L. Trends in occupational lead exposure since the 1978 OSHA lead standard. Am J Ind Med. 2004;45:558-72.
26. Sauvé JF, Davies HW, Parent M, Peters CE, Sylvestre MP, Lavoué J. Development of Quantitative Estimates of Wood Dust Exposure in a Canadian General Population Job-Exposure Matrix Based on Past Expert Assessments. Ann Work Expo Health. 2019;63:22-33.
27. Matanoski G, Selevan SG, Akland G, Bornschein RL, Dockery D, Edmonds L, Greife A, Mehlman M, Shaw GM, Elliott E. Role of exposure databases in epidemiology. Arch Environ Health. 1992;47:439-46.
28. Gómez MR. Factors associated with exposure in Occupational Safety and Health Administration data. Am Ind Hyg Assoc J. 1997;58:186-95.
29. Cherrie JW, Sewell C, Ritchie P, McIntosh C, Tickner J, Llewellyn D. Retrospective collection of exposure data from industry: results from a feasibility study in the United Kingdom. Appl Occup Environ Hyg. 2001;16:144-8.
30. Zilaout H, Vlaanderen J, Houba R, Kromhout H. 15 years of monitoring occupational exposure to respirable dust and quartz within the European industrial minerals sector. Int J Hyg Environ Health. 2017;220:810-9.
31. Davies HW, Peters CE, Hall A, Demers P, Nicol A. Capacity Development for a Canadian Workplace Exposure Database: Report to the WorkSafeBC Research Services. British Columbia, Canada. 2014.
32. Peters CE, Ge CB, Hall AL, Davies HW, Demers PA. CAREX Canada: an enhanced model for assessing occupational carcinogen exposure. Occup Environ Med. 2015;72:64-71.
33. Demers P, Peters CE, Davies HW, Kim J, Pahwa M, McLeod C, Nicol A, Labreche F, Lavoue J, Hutchings S, Rushton L. Incorporating More Detailed Exposure Assessment with Quantitative Estimates is Assessing the Burden of Occupational Cancer (abstract). Occup Environ Med. 2014;71:A51.
34. Hon CY, Peters CE, Jardine KJ, Arrandale VH. Historical occupational isocyanate exposure levels in two Canadian provinces. J Occup Environ Hyg. 2017;14:1-8.
35. The National Institute for Occupational Safety and Health (NIOSH). Health Hazard Evaluations (HHEs) Exposure Databases. https://www.cdc.gov/niosh/hhe/resourceshtml.Last Updated 2018.
36. Lavoue J, Friesen MC, Burstyn I. Workplace measurements by the US Occupational Safety and Health Administration since 1979: descriptive analysis and potential uses for exposure assessment. Ann Occup Hyg. 2013;57:77-97.
37. Henn SA, Sussell AL, Li J, Shire JD, Alarcon WA, Tak S. Characterization of lead in US workplaces using data from OSHA's integrated management information system. Am J Ind Med. 2011;54:356-65.
38. Sarazin P, Burstyn I, Kincl L, Lavoué J. Trends in OSHA Compliance Monitoring Data 1979-2011: Statistical Modeling of Ancillary Information across 77 Chemicals. Ann Occup Hyg. 2016;60:432-52.
39. Lavoue J, Vincent R, Gerin M. Formaldehyde exposure in U.S. industries from OSHA air sampling data. J Occup Environ Hyg. 2008;5:575-87.
40. Hamm MP, Burstyn I. Estimating occupational beryllium exposure from compliance monitoring data. Arch Environ Occup Health. 2011;66:75-86.
41. Sarazin P, Burstyn I, Kincl L, Friesen MC, Lavoué J. Characterization of the Selective Recording of Workplace Exposure Measurements into OSHA's IMIS Databank. Ann Work Expo Health. 2018;62:269-80.
42. Melville R, Lippmann M. Influence of Data Elements in OSHA Air Sampling Database on Occupational Exposure Levels. Appl Occup Environ Hyg. 2001;16:884-99.
43. Middendorf PJ. Surveillance of occupational noise exposures using OSHA's Integrated Management Information System. Am J Ind Med. 2004;46:492-504.

44. Babik KR, Shockey TM, Moore LL, Wurzelbacher SJ. Standardizing industrial hygiene data collection forms used by workers' compensation insurers. J Occup Environ Hyg. 2018;15:676-85.
45. Marchant GE, Crane A. The benefits and challenges of a voluntary occupational exposure database. J Occup Environ Med. 2011;53:S52-6.
46. Burns DK, Beaumont PL. The HSE National Exposure Database--(NEDB). Ann Occup Hyg. 1989;33:1-14.
47. Galea KS, Van Tongeren M, Sleeuwenhoek AJ, While D, Graham M, Bolton A, Kromhout H, Cherrie JW. Trends in wood dust inhalation exposure in the UK, 1985-2005. Ann Occup Hyg. 2009;53:657-67.
48. Tickner J. Occupational Exposure Databases. Occup Health Rev. 2001;89:27-31.
49. ter Burg W. Inventory of Databases Containing Worker Exposure Data on Non-Threshold Carcinogens in Europe. RIVM Letter Report, National Institute for Public Health and the Environment, The Netherlands. 2014;2014-0083.
50. Tickner J, Friar J, Creely KS, Cherrie JW, Pryde DE, Kingston J. The development of the EASE model. Ann Occup Hyg. 2005;49:103-10.
51. Tielemans E, Warren N, Schneider T, Tischer M, Ritchie P, Goede H, Kromhout H, Van Hemmen J, Cherrie JW. Tools for regulatory assessment of occupational exposure: development and challenges. J Expo Sci Environ Epidemiol. 2007;17 Suppl 1:S72-80.
52. Bell J. Survey of views on Occupational Hygiene Data Collection and use of the National Exposure Database. Health & Safety Laboratory, Derbyshire, United Kingdom. 2006;HSL/2006/18.
53. Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung (IFA). Exposure Database MEGA. https://www.dguv.de/ifa/gestis/expositionsdatenbank-mega/index-2jsp Accessed January 2021.
54. Gabriel S, Koppisch D. The MGU - a Monitoring System for the Collection and Documentation of Valid Workplace Exposure Data. Gefahrstoffe - Reinhaltung der Luft. 2010;70:43-9.
55. Gabriel S. The BG measurement system for hazardous substances (BGMG) and the exposure database of hazardous substances (MEGA). Int J Occup Saf Ergon. 2006;12:101-4.
56. Pesch B, Kendzia B, Hauptmann K, Van Gelder R, Stamm R, Hahn JU, Zschiesche W, Behrens T, Weiss T, Siemiatycki J, Lavoué J, Jöckel KH, Brüning T. Airborne exposure to inhalable hexavalent chromium in welders and other occupations: Estimates from the German MEGA database. Int J Hyg Environ Health. 2015;218:500-6.
57. Stamm R. MEGA-database: one million data since 1972. Appl Occup Environ Hyg. 2001;16:159-63.
58. European Commission. Service Contract to Create a Database and Develop a Model to Estimate the Occupational Exposure for a List of Hazardous Chemicals in the Member States of the European Union and in the EFTA/EEA Countries. European Commission, Brussels. 2016;No VT/2013/079.
59. Lavoué J, Gérin M, Vincent R. Comparison of formaldehyde exposure levels in two multi-industry occupational exposure databanks using multimodel inference. J Occup Environ Hyg. 2011;8:38-48.
60. Carton B. COLCHIC Chemical Exposure Database: Information on Lead and Formaldehyde. Appl Occup Environ Hyg. 1995;10:345-50.
61. Clerc F, Bertrand N, Vincent R. TEXAS: a Tool for EXposure ASsessment-Statistical models for estimating occupational exposure to chemical agents. Ann Occup Hyg. 2015;59:277-91.
62. Lavoué J, Vincent R, Gérin M. Statistical modelling of formaldehyde occupational exposure levels in French industries, 1986-2003. Ann Occup Hyg. 2006;50:305-21.
63. Rajan B, Alesbury R, Carton B, Gérin M, Litske H, Marquart H, Olsen E, Scheffers T, Stamm R, Woldbaek T. European Proposal for Core Information for the Storage and Exchange of Workplace Exposure Measurements on Chemical Agents. Applied Occupational and Environmental Hygiene. 1997;12:31-9.

64. Joint ACGIH-AIHA Task Group on Occupational Exposure Databases. Data Elements for Occupational Exposure Databases: Guidelines and Recommendations for Airborne Hazards and Noise: Special Report. Appl Occup Environ Hyg. 1996;11:1294-311.
65. Kauppinen T, Uuksulainen S, Saalo A, Mäkinen I, Pukkala E. Use of the Finnish Information System on Occupational Exposure (FINJEM) in epidemiologic, surveillance, and other applications. Ann Occup Hyg. 2014;58:380-96.
66. Kauppinen T. Finnish Occupational Exposure Databases. Appl Occup Environ Hyg. 2001;16:154-8.
67. Simonsen L, Lund SP. A strategy for delineating risks due to exposure to neurotoxic chemicals. Am J Ind Med. 1992;21:773-92.
68. Flachs EM, Petersen SEB, Kolstad HA, Schlünssen V, Svendsen SW, Hansen J, Budtz-Jørgensen E, Andersen JH, Madsen IEH, Bonde JPE. Cohort Profile: DOC*X: a nationwide Danish occupational cohort with eXposure data - an open research resource. Int J Epidemiol. 2019;48:1413-k.
69. Osvoll P, Woldbaek T. Distribution and skewness of occupational exposure sets of measurements in the Norwegian industry. Ann Occup Hyg. 1999;43:421-8.
70. Osvoll P, Woldbaek T. Trends in Organic Solvent Exposure in the Norwegian industry from 1985 to 1994. Occup Hyg. 1998;4:85-94.
71. Lenvik K. Occupational Exposure to Styrene in Norway, 1972-1996. Appl Occup Environ Hyg. 1999;14:165-70.
72. Tang TK, Siang LH, Koh D. The development and regulation of occupational exposure limits in Singapore. Regul Toxicol Pharmacol. 2006;46:136-41.
73. Ministry of Manpower. Requirements for hygiene monitoring. https://www.mom.gov.sg/workplace-safety-and-health/monitoring-and-surveillance/hygiene-monitoring/requirements Accessed March 2021.
74. Ministry of Manpower. Workplace Safety and Health Report 2020: National Statistics. Singapore. 2020.
75. Ministry of Manpower. Workplace Safety and Health 2028, Singapore. https://www.mom.gov.sg/-/media/mom/documents/safety-health/publications/wsh2028-report.pdf. 2019.
76. Schweizerische UnfallVersicherungsAnstalt (SUVA). SUVA. https://www.suva.ch/en Accessed February 2021.
77. Savic N, Gasic B, Schinkel J, Vernez D. Comparing the Advanced REACH Tool's (ART) Estimates With Switzerland's Occupational Exposure Data. Ann Work Expo Health. 2017;61:954-64.
78. Basinas I, Liukkonen T, Sigsgaard T, Andersen NT, Vestergaard JM, Galea K, Wiggans R, Vincent R, Eduard W, Kolstad HA, Vested A, Kromhout H, Schlünssen V. P096 Statistical modelling and development of a quantitative job exposure matrix for wood dust in the wood manufacturing industry. Occupational and Environmental Medicine. 2016;73:A152-A3.
79. Basinas I, Wouters IM, Sigsgaard T, Heederik D, Spaan S, Smit LA, Bønløkke JH, Eduard W, Radon K, Straumfors A, Omland Ø, Duchaine C, Nowak D, Schlünssen V, Kromhout H. O46-4 Development of a quantitative job exposure matrix for endotoxin exposure in agriculture. Occupational and Environmental Medicine. 2016;73:A88-A.
80. Orlowski E, Audignon-Durand S, Goldberg M, Imbernon E, Brochard P. EV@LUTIL: An open access database on occupational exposures to asbestos and man-made mineral fibres. Am J Ind Med. 2015;58:1059-74.
81. Kauppinen T, Teschke K, Savela A, Kogevinas M, Boffetta P. International data base of exposure measurements in the pulp, paper and paper product industries. Int Arch Occup Environ Health. 1997;70:119-27.
82. Burstyn I, Kromhout H, Cruise PJ, Brennan P. Designing an international industrial hygiene database of exposures among workers in the asphalt industry. Ann Occup Hyg. 2000;44:57-66.

83. De Vocht F, Straif K, Szeszenia-Dabrowska N, Hagmar L, Sorahan T, Burstyn I, Vermeulen R, Kromhout H. A database of exposures in the rubber manufacturing industry: design and quality control. Ann Occup Hyg. 2005;49:691-701.
84. Kauppinen T, Vincent R, Liukkonen T, Grzebyk M, Kauppinen A, Welling I, Arezes P, Black N, Bochmann F, Campelo F, Costa M, Elsigan G, Goerens R, Kikemenis A, Kromhout H, Miguel S, Mirabelli D, McEneany R, Pesch B, Plato N, Schlünssen V, Schulze J, Sonntag R, Verougstraete V, De Vicente MA, Wolf J, Zimmermann M, Husgafvel-Pursiainen K, Savolainen K. Occupational exposure to inhalable wood dust in the member states of the European Union. Ann Occup Hyg. 2006;50:549-61.
85. Kromhout H, Vermeulen R. Temporal, personal and spatial variability in dermal exposure. Ann Occup Hyg. 2001;45:257-73.
86. Peters S, Vermeulen R, Olsson A, Van Gelder R, Kendzia B, Vincent R, Savary B, Williams N, Woldbæk T, Lavoué J, Cavallo D, Cattaneo A, Mirabelli D, Plato N, Dahmann D, Fevotte J, Pesch B, Brüning T, Straif K, Kromhout H. Development of an exposure measurement database on five lung carcinogens (ExpoSYN) for quantitative retrospective occupational exposure assessment. Ann Occup Hyg. 2012;56:70-9.
87. Houba R. Building an industry-wide occupational exposure database for respirable mineral dust – experiences from the IMA dust monitoring programme. Journal of Physics: Conference Series. 2009;151.
88. ICF Consulting Services Limited. Feasibility study: Resources and needs of EU-OSHA related to a potential takeover, further development, extension and maintenance of the HazChem@Work database. EU-OSHA. 2018.
89. Vinzents P, Carton B, Fjeldstad C, Rajan B, Stamm R. Comparison of Exposure Measurements Stored in European Databases on Occupational Air Pollutants and Definition of Core Information. Appl Occup Environ Hyg. 1995;10.
90. Clerc F, Vincent R. Assessment of occupational exposure to chemicals by air sampling for comparison with limit values: the influence of sampling strategy. Ann Occup Hyg. 2014;58:437-49.
91. Kromhout H, Symanski E, Rappaport SM. A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. Ann Occup Hyg. 1993;37:253-70.
92. Allen and Clarke. Occupational Health and Safety in New Zealand. National Occupational Health and Safety Committee (NOHSAC) Wellington, New Zealand. 2006;Technical Report 7.
93. Driscoll T. Review of Australian and New Zealand Workplace Exposure Surveillance Systems. National Occupational Health and Safety Committee (NOHSAC) Wellington, New Zealand. 2006;Technical Report 6.
94. Pearce N, Dryson E, Feyer A, Gander P, Wagstaffe M. Surveillance and Control of Workplace Exposures in New Zealand: Report to the Minister of Labour. National Occupational Health and Safety Committee (NOHSAC) Wellington, New Zealand. 2006.
95. National Occupational Safety and Health Advisory Committee. Fifth Annual Report to the Minister of Labour. Wellington, New Zealand. 2008.
96. WorkSafe New Zealand. Health and Safety at Work Strategy 2018-2028. Wellington, New Zealand. 2018.
97. t Mannetje AM, McLean DJ, Eng AJ, Kromhout H, Kauppinen T, Fevotte J, Pearce NE. Developing a general population job-exposure matrix in the absence of sufficient exposure monitoring data. Ann Occup Hyg. 2011;55:879-85.
98. 't Mannetje A. Workplace Exposure to Carcinogens in New Zealand: Study Report for the Deparment of Labour. Centre for Public Health Research, Massey University, Wellington, New Zealand. 2013.
99. Cole C, Arthur J. Workshop on Practical Guidelines. Appl Occup Environ Hyg. 1995;10:417-9.

Appendices

Appendix 1: Questionnaire for international database contacts

1. Exposure database description

- Who and what type of organisation owns the database (e.g. is it a regulator)?
- Who hosts the database (if different from above)?
- What year was it established?
- Who manages/oversees the database?
- What is the geographical location of the database?
- What are the aims/objectives of the database?
- Which occupational hazard type(s) are captured in the database e.g. physical, chemical, biological?
- How many different types of agents are included in the database and over what time period?
- How many exposure measurements have been included to date?

2. Data collection

2.1 Data sources

- What are the data sources/who provides data to the exposure database?
- How is the source population defined i.e. occupation, industry sector, or other?
- What is the geographical capture of the database?
- Is data collection passive or active? i.e. is data actively sought or passively received?
- What *type* of data is included e.g. qualitative, quantitative, semi-quantitative?
- What data variables are included? (e.g. sampling method, LOD, 8h TWA, work task etc.)

2.2 Frequency of data collection

- What is the frequency of data collection? i.e. is the data collection continuous or periodic? If periodic, what are the time intervals?

2.3 Medium

- In what format is the data collected e.g. hard copy or electronic?

2.4 Coding

- How is the data coded e.g. for retrieval?
- How is the data from the database extracted and analysed?

2.5 Reliability

- Is the reliability of data assessed e.g. reliability/reproducibility of exposure measures, are data sources assessed or verified, are confidence levels determined?
- Variability – does the data collection method account for irregular exposure patterns e.g. itinerate or seasonal workers?

2.6 Completeness

- Is there any sense of the degree to which data pertaining to the target population is captured? Is the data is collected in the form of a sample survey or on the whole population of interest?

2.7 Ethical/privacy requirements

- What ethics/privacy issues had to be dealt with to meet privacy requirements in its local jurisdiction?
- What consent issues had to be dealt with to meet privacy requirements in its local jurisdiction?
- Were there any intellectual property-related issues that had to be dealt with in its local jurisdiction?

3. Data uses

3.1 Users

- Who is the data aimed at i.e. who are the intended prime users of information collected?
- Who or what are the main users of the information?

3.2 Links/interrelated

- Is the data linked to other databases?

3.3 Presentation

- How is the data presented e.g. raw data in database form, via reports/hard copy, electronic?
- If reports are generated, what is the frequency?
- Are the data freely available for third parties or can they be made available with restrictions?
- Can trends in exposure levels be assessed over time?

3.4 Transferability

- Is it intended that the data might be available for transfer to other databases?

4. Resource requirements

4.1 Funding

- How was the set-up of the database funded?
- How is the ongoing maintenance and development of the database funded?

4.2 Data establishment

- What were the human resource requirements for establishment?
- What were the non-salary expenditure costs?

4.3 Data collection

- What are the human resource requirements (skills and time) for data collection/entry?
- What are the ongoing, non-salary costs associated with data collection and entry requirements?

4.4 Data maintenance

- What are the human resource requirements (skills and time) for maintenance of the database?
- What are the ongoing, non-salary costs associated with maintaining the database?

4.5 Data storage

- What data storage system is used – off the shelf or purpose-designed?
- If off the shelf, which software package is used?

5. Data governance

- How are decisions made on its use, development, expansion to other hazards etc?

6. History/future

6.1 Drivers for establishment

- What led to the establishment of the database?

6.2 Barriers/enablers for the establishment of the database?

- What were the identified barriers to the establishment of the database?
- What were the identified enablers to the establishment of the database?

6.3 Barriers/enablers for the further development?

- What are the identified barriers to further development of the database?
- What are the identified enablers to further development of the database?

6.4 Subjective assessment

- Can a subjective assessment of the system usefulness be made?
- Do you have any further comment on the challenges faced/limitations of the database?
- What advice do you have if you were to start the database again?

Appendix 2: Review of international databases, tables of characteristics A2.1-2.7

Table A2.1 Description of databases

#	Country	Name of database	a. Owner of database	b. Host of database	c. Year established	d. Database manager	e. Database location	f. Aims/objectives of database	g. Which occupational hazard type(s) are captured in the database	h. How many different types of agents are included?	i. How many exposure measurements have been included to date?	j. Information on included industries
1	Canada	Canadian Workplace Exposure Database (CWED)	Hosted by the University of British Columbia		2008	Hugh Davies	Vancouver, Canada	The database was created with the goal to collect exposure measurements from government agencies, researchers, and other sources to create a centralised database of both current and historical exposure measurements from workplaces across Canada (22)	Chemical	350, as at 2019 (26)	500,000 measurements, as at 2019 (26)	-
2	U.S.	Integrated Management Information System (IMIS)	Occupational Safety and Health Administration (OSHA)		1979	NA	NA	IMIS was designed to collect, process, retrieve, and communicate penalty assessment, arbitration, and monetary collection information regarding OSHA's inspections (37)	Chemical and physical (noise)	-	>1.6 million, as at 2018 (3)	All industries except mining and agriculture (21)
2	U.S.	OSHA Information System (OIS)	The Office of the Chief Information Officer (OCIO) is the system owner and OSHA is the business owner (GS)	Data center hosted by the U.S. Department of Agriculture, National Information Technical Center (NITC; GS)	2011 for Federal and 2014 for State	-	U.S. Department of Agriculture, National Information Technical Center (NITC)	OIS was designed to support OSHA's 'mission critical system' to support the enforcement of the Occupational Safety and Health Act of 1970. It serves as a *"case management system used to manage enforcement inspections, consultation visits, accident investigations, compliance assistance/cooperative programmes, and training outreach activities"* (GS)	Chemical and physical (noise)	500 (GS)	The OIS database contains >99,000 measurements from enforcement sampling and >65,000 measurements from consultation visits. IMIS data was not migrated to OIS (GS)	-

3	U.K.	National Exposure Database (NEDB)	Health and Safety Executive (HSE) U.K.	HSE U.K.	1986	NA	The database was developed and managed by the occupational hygiene unit in the technology division of HSE in Bootle, U.K.	The objectives as stated in 1989 were: 1) to provide detailed and comprehensive exposure data for use in the setting of the new occupational exposure limits; 2) to provide a major source of exposure data for use in epidemiological studies; and 3) to facilitate dissemination of information on occupational exposures (46)	Chemical, some biological (PB)	*"Not known. In the hundreds"* (PB)	50,000-100,000 (PB)	During mid-1980s to early 1990s, ~5000 workplaces were visited during which an average of 10 samples were collected (1)
4	Germany	Measurement Data on Exposure to Hazardous Substances in the Workplace - MEGA	Institute for Occupational Safety and Health of the German Social Accident Insurance (IFA)	IFA	1972	IFA	IFA, Sankt Augustin, Germany	The database was set up to provide a data pool for IFA to solve a range of problems including: prevention, epidemiological issues, investigations of occupational disease claims, and determination of pollutant exposure (53)	Dust (1961-), organic compounds (1982-), and biological (1995-) (DK)	910 different chemical agents (including dust and fibres) and 750 biological agents (2020; DK)	3.6 million (2020; DK)	72,400 companies; 850 sectors; 5,000 industrial workplaces, as at 2018 (53)

5	France	COLCHIC	French National Health Insurance Fund for Salaried Workers (CNAM)	The French National Institute of Safety and Health (INRS)	1987 (online since 2009)	Risk and Exposure Assessment Laboratory, INRS	INRS in Vandoeuvre-Lès-Nancy	COLCHIC comprises data from prevention and research activities. The aims of the database are to: 1) centralise measurement data from various laboratories; 2) to harmonise workplace sampling and air analysis methods; and 3) to support chemical risk assessment in France (24)	Chemical and biological (GM)	For the period 1987-2019, ~1,150,000 air measurement results from 1,197 agents are recorded in COLCHIC (GM)	The data from COLCHIC is from companies covered by the national insurance scheme (~55% of companies in France) excluding public services, agriculture, mining, energy production, national mass transit, and small retail and artisans. In 2016, it was reported that the three most common industries are manufacture and metal products, manufacture of rubber and plastic products, and manufacture of chemicals and chemical products (=27% of data) (24)
5	France	SCOLA	French Directorate-General for Labour (DGT)	The French National Institute of Safety and Health (INRS)	2006	Risk and Exposure Assessment Laboratory, INRS	INRS in Vandoeuvre-Lès-Nancy	In France, exposure measurements made during assessment of compliance to regulatory occupational exposure limits are recorded in a national register called SCOLA. The measurements are performed by independent certified laboratories and the data are used to define priorities for national prevention programmes (24)	Chemical	For the period 2007-2019, 1 million air measurement results from 150 agents are recorded in SCOLA (GM)	In 2016, the three most common industries are specialised construction activities, remediation activities and other waste management services, and manufacture of other non-metallic mineral products (=45% of data) (24)

6	Finland	Finnish Database of Occupational Exposure Measurements (FDOEM)/Laboratory Information System (LIMS)	Finnish Institute of Occupational Health (FIOH)	FIOH	Data had been collected by FIOH since 1950 but electronic data was available for chemical agents from 1994 onwards (66)	-	FIOH	-	FIOH hygienists collect chemical, physical (e.g. noise) & microbiological agents but only chemical and microbiological are stored in LIMS (TT)	>100 (TT)	In 2014, it was reported that the total number of exposure measurements used to assess chemical exposure for FINJEM was 157,035 (65). The number of measurements stored annually is a few thousand (TT).	-
7	Denmark	ATABAS	Danish National Institute of Occupational Health (AMI)		-	-	-	-	Chemical	-	-	-
8	Norway	EXPO	Statens arbeidsmiljøinstitutt (STAMI; National Institute of Occupational Health). STAMI is a research institute funded by the Ministry of Labour and Social Affairs		1984	National Institute of Occupational Health (STAMI)	Oslo, Norway	The aim of EXPO is to collect airborne occupational exposure measurement data in order to build a knowledge base of exposures in different occupations and industries and to prevent illness and promote good health. The aim of EXPO Online is to serve as an archive for companies' measurement data (RO)	Airborne chemical and biological agents (RO)	~1400 different components (agents and different fractions, inhalable, respirable etc.) (RO)	From 1984-2019, the database contained about 100,000 air samples (which can contain one or more components) and ~600,000 exposure results. For the period 2000-2019, the database contained~45,000 air samples and ~250,000 exposure results (RO).	Different industries are represented including mining and quarrying, construction, manufacturing, and waste management (RO)

9	South Korea	Work Environment Measurement Database (WEMD)	Occupational Safety and Health Research Institute (OSHRI) of the Korea Occupational Safety and Health Agency (KOSHA)	2000	OSHRI	Ulsan City, South Korea	Monitoring workplace hazards, research, and producing annual reports of company exposure data (SP)	Physical, chemical (SP)	~190 (SP)	No available figure on individual exposure measurements but 75,000 companies report annually or bi-annually (SP)	-
10	Singapore	National Database for Noise and Chemical Exposure (NDNCE)	Occupational Safety and Health Division (OSHD), Ministry of Manpower (MOM)	Early 1990s, upgraded in 2010	MOM	Singapore	The aims of the database are to enable MOM to: 1) identify high-risk workplaces; 2) assess trends in exposure levels; and 3) provide advice to employers on control measures and appropriate monitoring programmes (EK)	Chemical, noise	600+ chemical agents listed in the toxic substances schedule (73)	In 2020, 1,561 chemical exposure results were received and 672 workplaces submitted noise measurement results between 2018-2020 (74)	Chemical: the largest proportion of results were from the manufacture of petrochemical products and the metalworking industries (74)
11	Switzerland	SUVA	Swiss Insurance Fund (SUVA)	2007	Swiss Insurance Fund (SUVA)	Lucerne, Switzerland	The aims include: a) ensuring that chemical exposure does not exceed occupational exposure limits (OEL) for individual substances under concern (NS); b) informing reviews of OELs; c) identify high-risk industries to target for control activities; and d) enable comparisons with published literature (PS)	Chemical and biological (PS)	~400 (PS)	~5000 measurements each year. Some (a few thousand measurements) retrospectively collected data (prior to 2007) is also included in the database (PS)	-
NB: Quote marks reflect direct quotes from the survey responses of the database managers; initials in brackets indicate the database contact from Table 3											

Table A2.2A Data collection												
#	**Country**	**Name of database**	**k. What are the data sources/who provides data to the exposure database?**	**l. How is the source population defined?**	**m. Geographical capture**	**n. Is data collection passive or active?**	**o. Data type e.g. quantitative or qualitative?**	**p. Frequency of data collection**	**q. Medium: in what format is the data collected e.g. hard copy or electronic?**	**r. How is the data coded e.g. for retrieval?**	**s. How is the data extracted and analysed?**	**t. Is the reliability of data assessed e.g. reliability/reproducibility of exposure measures, are data sources assessed or verified. Other quality control/assurance procedures?**
1	Canada	CWED	Canada has 14 federal and provincial regulatory agencies responsible for health and safety, including specialist divisions in some provinces. The CWED initially comprised of historical data from existing electronic databases held by WorkSafe British Columbia (BC), the Ontario Ministry of Labour, The National Dose Registry, and Quebec's Institut de Recherche Robert-Sauve en Sante et en Securite du Travail (6) through a data acquisition agreement process in partnership with PopdataBC. By 2014, data had also been added from the regulatory agencies of Manitoba, Yukon, Saskatchewan (SK), and the BC Ministry of Energy and Mines (BCMEM). The CWED currently contains data from across eight federal, provincial and territorial jurisdictions (22) including Labor Canada. Approximately 80% of the measurements are from British Columbia and Ontario (26)			Historical data collection to date	Quantitative, qualitative	The data collection period spans 1960-2010 but the majority of measurements were taken mid-1970s to the late 1990s (26)	The medium varies according to jurisdiction. The CWED initially comprised of historical data from existing electronic databases for four agencies. Following a nationwide data holders' survey in 2009, funding for 2 years was received to transfer paper-based records into an electronic database for data held by the BCMEM and the regulatory agencies of SK, Yukon, and Manitoba (6)	North American Industry Classification System (NAICS) 2002 industry codes and National Occupational Classification for Statistics (NOC-S) 2006 occupation codes, both from Statistics Canada (26). The 2014 Capacity Development Report (31) that investigated the addition of data from BCMEM and the Government of SK reported that new codes were developed to group agents by toxicological category. The lack of harmonisation of coding (particularly for non-carcinogens) was a major challenge that was identified (Hugh Davies, personal communication, January 2021)	-	Data collection to date has involved historical data from regulatory agencies and therefore data sources have not been verified further (HD)
2	U.S.	IMIS	OSHA safety inspectors have been collecting hygiene data during its enforcement and consultation visits of worksites since 1972 to verify compliance with PELs (36). Since 1979, these measurements were entered into the IMIS (since 1984 for consultation data). OSHA's enforcement efforts are conducted through inspections of workplaces and consultation visits are conducted at the request of employers seeking assistance on controlling hazards and complying with regulations at their worksite		All States (JL)	Actively sought by OSHA inspectors (JL)	Quantitative (JL)	Continuous (enforcement and consultation visits)	-	Industry is coded using the Standard Industrial Classification (SIC) codes and the North American Industry Classification System (NACIS) (38)	Freedom of Information Act requests - excel files (JL)	Analytical methods are certified (JL)

2	U.S.	OIS	The data collection is the same as for the IMIS above. The system supports over 4,000 users nationwide (GS)	The database serves Federal and State users nationwide (GS)	-	Quantitative, qualitative	Continuous throughout the year (GS)	The data is stored electronically (GS)	*"Data is coded by key variables that aids in retrieval including case number, case type, assigned office, establishment name, address, key dates, etc."* (GS)	*"OIS includes a reports module for generating reports which can be downloaded to Excel for analysis"* (GS)	-
3	U.K.	NEBD	Reports of visits by HSE inspectors during surveys, inspections, and investigations in U.K. workplaces. In addition to inspection visits, representative surveys were also carried out to ascertain information on working practices and exposure levels for specific industries (47)	Nationwide	-	Quantitative, qualitative (1)	The majority of measurements were collected between 1985-1990; ~80,000 samples were collected between 1986-2001 (1)	-	-	-	No (PB)
4	Germany	MEGA	In Germany, social accident insurance is statutory and health and safety at work is monitored by statutory accident insurance institutions, which include Berufsgenossenschaften (BGs) for the private sector and Unfallkassen (called "UK") for the public sector (excluding the military). Coverage is optional for self-employed workers and freelancers (DK). Workplace measurements are carried out as part of the supervisory duties of the accident insurers or from campaigns carried out by measurement services	Nationwide	Active data collection by the accident insurers	Quantitative, qualitative	Data collection for the accident insurers' supervisory activities is continuous. Data collection began in the 1960s and the number of measurements increased until ~2000. The number of measurements has declined in recent years but the number of agents is still increasing (DK)	Electronic	Code lists (~30 lists, developed in-house) are provided to ensure a standardised and harmonised data collection all over the MGU. Training is available for measurement services to ensure uniform coding (20)	IFA-developed software called MEGA^{Pro} is used for the extraction and statistical analysis of data (DK)	All of the processes within the MGU system, including data measurement, analysis, and data entry are standardised and conform to the requirements of a quality management system according to DIN EN ISO 9001 (20). Air samples are collected using standardised measurement systems (including techniques and devices for sampling) which are in-house developments and all sampling and analysis must be compliant with the standard EN482

5	France	COLCHIC	COLCHIC contains industrial hygiene measurements made by the laboratories of the INRS and the eight regional health insurance funds (CARSATs and CRAMIF) and their interregional chemical laboratories (24)	The data in COLCHIC are from measurements undertaken in companies covered by the national health insurance scheme, which determines who the CARSAT/CRAMIF hygienists target based on prevention and research priorities (24)	The eight interregional chemical laboratories are assigned to defined geographical areas (24)	Active data collection determined by prevention and research campaigns set by the national insurance scheme	Quantitative, qualitative	Continuous since 1987. A median of 31,917 measurements per year was recorded for the period 1987-2012 (24)	Electronic (JL)	The majority of data are coded using standardised classifications specific to COLCHIC (such as task or ventilation) or using national classifications (e.g. job and industry). A new coding system was introduced in 2002 with the modification and/or replacement of existing variables and the addition of new variables such as occupation (24)	Data analyst at INRS (JL)	Most of the data stored in COLCHIC are compliant with the European Standard EN482 (20)
5	France	SCOLA	Independent certified laboratories take measurements at the request of companies in order to fulfil their regulatory obligations. The company director is responsible for risk assessment and the development of a hazard measurement strategy (24)		Nationwide	Passive data collection: compliance measurements are submitted to SCOLA	Quantitative	Continuous since 2007. The number of measurements started with 3,598 in 2007 and steadily increased to 52,202 in 2012, to now >1 million in 2019 (GM)	Electronic (JL)	The coding of data is similar to that in COLCHIC (24)	Data analyst at INRS (JL)	*"Compliance measurements are governed by strict sampling guidelines / standards (NF, EN, ISO, MetroPol(INRS))"* (GM)
6	Finland	FDOEM)/LIMS	FIOH offers industrial hygiene services and biomonitoring services to workplaces and the measurement results are stored in their database. The database has since been replaced by a customised laboratory information system (LIMS) which is used to manage and store workplace measurement data (TT)		Nationwide	Measurement data is from FIOH's hygiene and biomonitoring services	Quantitative, qualitative	Measurement data is from FIOH's hygiene and biomonitoring services	-	-	-	-
7	Denmark	ATABAS	Inspectorates and AMI visited workplaces where there were OHS problems (1). ATABAS also included measurement data from the Danish Working Environment Authority (LA)		-	-	Quantitative (1)	-	-	-	-	-

8	Norway	EXPO	STAMI results from research projects and results of companies contacting STAMI to conduct exposure monitoring. EXPO Online: companies register and submit their data online (RO)	Nationwide	Measurement data is from STAMI's exposure monitoring services and research projects. EXPO Online: passive data collection (RO)	Quantitative	Measurement data is from STAMI's exposure monitoring services and research projects	Hard copy but will soon be electronic. EXPO Online is electronic (RO)	Certain variables e.g. work task and work area are coded according to lists developed by STAMI with input from the advisory board and the companies. Occupation and industry are coded according to national standard classifications. The database contains dropdown menus and predefined lists (RO)	*"Depends on the purpose"* (RO)	No (RO)
9	South Korea	WEMD	In South Korea, employers are required to regularly evaluate their workplace if hazardous materials are present. In particular, companies are required to conduct exposure monitoring once every 6 months or once every year if the concentration level of all chemical substances is below the occupational exposure limit (SP)	Nationwide	Passively received from employers (SP)	Quantitative, qualitative (SP)	Bi-annually or annually if results are below the OEL	Electronic (SP)	Employers transfer their data through consulting institutes via an electronic system with pre-determined variables. The majority of consulting institutes are provided with training on how to record variables in standardised manner (SP)	OSHRI downloads the electronic database in Excel or MS Office (SP)	No (SP)
10	Singapore	NDNCE	The data contained in the database are from hygiene assessments conducted by MOM and results submitted by companies (EK). For the latter, companies with workers likely to be exposed to excessive noise or toxic substances are required by law to monitor occupational noise levels under the Workplace Safety and Health (Noise) Regulations and toxic substances (of which there are 600+) under the Workplace Safety and Health (General Provisions) Regulations and to retain the results for at least 5 years. The reporting of both noise (compulsory) and toxic substances (recommended) can be submitted to the database online using standardised templates (73)	Nationwide	Mostly passive but companies may be required to submit results during MOM inspections (EK)	Quantitative	For toxic substances monitoring, frequency of monitoring depends on the level of exposure against the PEL. For noise monitoring, the frequency is dependent on the number of employees likely to be exposed to excessive noise and the reporting of results is required once every 3 years (73)	Electronic	Singapore Standard Industrial Classification (SSIC) codes (EK)	Data mining tools (EK)	Hygiene monitoring must be carried out by competent persons (e.g. who have undertaken the necessary training courses) (73). Reliability of data is not assessed by OSHD (EK)

11	Switzerland	SUVA	SUVA carries out exposure monitoring for employers insured and therefore supervised directly by SUVA (mainly industrial and construction sectors) which equates to about half of Switzerland's workforce (76). Measurements are usually carried out at the request of inspectors (PS)	Nationwide	Actively collected by SUVA (NS)	Quantitative, qualitative	Continuous (NS) but no fixed intervals (PS)	Hard copy individual measurement reports. The database is electronic (PS)	Substances are coded according to *"IUPAC/Swiss OEL-Listing; industry sectors are coded usis NOGA-Codes; and internal coding is used for work processes and activities"* (PS)	*"Manual interpretation and coding into the input parameters of different REACH exposure models (Excel format)"* (NS)	The SUVA laboratory is accredited according to ISO 17025 and therefore the reliability of sampling and most of the analytical methods are assessed. The measurement strategy used conforms to the requirements of EN689 (PS)

Table A2.2B Data collection (continued….)									
#	**Country**	**Name of database**	**u. Variability – does the data collection method account for irregular exposure patterns e.g. itinerant or seasonal workers?**	**v. Is there any sense of the degree to which data pertaining to the target population is captured? Is the data collected in the form of a sample survey or on the whole population of interest?**	**w. What ethics/privacy issues had to be dealt with to meet privacy requirements in its local jurisdiction?**	**x. What consent issues had to be dealt with in its local jurisdiction?**	**y. Were there any intellectual property-related issues that had to be dealt with in its local jurisdiction?**	**z. Relevant data included: If quantitative, what type of data e.g. 8h-TWA, 15-minute STEL, other task duration? Availability of ancillary information?**	**aa. Main exposures**
1	Canada	CWED	*"No, we have no control over sampling campaign motivation. Some of the data may very well have been collected for these purposes but if that information is not in the dataset we receive, then we would not know."* (HD)	*"Whole population, but as data is mostly collected by regulatory bodies, we assume it is primarily compliance sampling where there is belief there may be over-exposure."* (HD)	No personal identifiers are recorded and therefore *"no privacy issues arise"* (HD)	Data sharing agreements are signed with all data owners (HD)	No (HD)	Analyte name, exposure concentration (ppm), exposure units, sampling method, sampling date, company name, location, sampling duration, sample type (area or personal), occupational title (NOCS) and industry title (NAICS), LOD, work area/task, flow rate, analysis method, and particle size fraction (22)	The most frequently measured agents are dust, carbon monoxide, carbon dioxide, toluene, xylene, lead, and formaldehyde (22)
2	U.S.	IMIS	-	The process for which OSHA selects workplaces for its enforcement and consultation visits is non-random and may overrepresent workplaces and workers with potentially higher or lower than average exposures. A recent study of ancillary variables suggests that OSHA's process for selecting worksites for inspection influences the exposure levels captured in IMIS (38)	Some data variables for consultation visits such as company name and address are not included to preserve confidentiality	-	-	Measured levels including TWA, STEL, ND, sample type (i.e. area, bulk, personal, screening), and limited ancillary variables including inspection type (i.e. complaint, referral, follow-up, planned), OSHA plan (federal or state), inspection scope, union status, OSHA region, establishment size, penalty, and industry	The most frequently measured chemical agents during the period 1979-2011 were inorganic lead, iron oxide fume, and crystalline silica (38)
2	U.S.	OIS	-	-	OIS is accessible to authorised users only (GS)	*"Users must acknowledge and confirm that they have received, read and signed the OIS Rules of Behavior in order to receive an OIS account. To login to OIS and use the system, user consents to monitoring, recording, disclosure and be subject to audit*	-	*"OIS is a much more complete record"* (GS)	-

						by authorized personnel." (GS)			
3	U.K.	NEBD	-	-	*"Consent forms for participation in surveys (since General Data Protection Regulations – GDPR was enacted) and for biological monitoring sampling (all samples)"* (PB)	*"Requirements for GDPR since it was enacted. Also, individuals are not named in the database. The name and address of employers are held"* (PB)	No (PB)	The recorded information included: occupier name and address, date of visit, industry, process and job, substances sampled and their concentration, sample type (breathing zone; static or biological), sample duration, 8-hour TWA, visit type, respirator use, and analytical methods. Abstracts (free text) of all occupational hygiene visits were included after the 1999 upgrade (48)	In 2001, the most frequently measured agents were quartz (~6000), total dust (~4,500 samples), toluene (~3,200), and styrene (~3,000) (48)
4	Germany	MEGA	The job title is the only information collected about the worker (no information on age or sex) and thus no information is available on whether workers are itinerant or seasonal. Information is collected on irregular exposure patterns and the following categories are used: a) usual working situation: normal condition; b) usual working situation: assembly, disassembly, start-up procedures, preparations, set-up work; c) simulated working situation in the context of an occupational disease	Selection of sampling locations is conducted in a non-random manner as determined by the measurement services of the accident insurer (DK)	No personal information is stored and name and address of companies are not stored in MEGA (DK)	*"Because no personal information about workers is stored, we did not have to ask for their consent. We work in the framework of a statutory insurance system and law approves exposure measurements"* (DK)	*"Because we use IFA-developed software, we don't have to deal with intellectual property-related issues"* (DK)	The number of data fields has grown from ~30 items from 1972-1989 to ~200 from 1990-2007 (54). Examples of variables include: exposure measurement, products, type of machine or tool used, area or personal sample, effect of other emission sources, sector, occupation, sample method, sampling duration, size of the production hall/room, duration of exposure, shift or peak exposure, local exhaust ventilation, analytical method, climatic conditions, day, season and year of sampling (54)	-

			investigation; d) simulated working situation, to verify improved conditions; e) simulated working situation, to verify unfavourable conditions; f) reasonable worst case working situation; g) unforeseeable operational disruption, accident. *"The variability of exposure is documented through many different measurements for one task or one industrial branch. One report is used to describe the working situation on that day. We do not collect several samples from one worker to describe his or her exposure situation during the whole year"* (DK)						
5	France	COLCHIC	The type of employment is specified in COLCHIC but not whether a measurement relates to itinerant or seasonal workers (GM)	The national health insurance scheme does not cover ~55% of companies in France including public services, agriculture, and small retail and artisans. The number of workers by activity sector in the population and the number of measurements in COLCHIC for activity sectors is known. Some work has been conducted to ascertain the degree to which data reflects the target population but it has not been published (GM)	COLCHIC follows the EU regulation 2016/679 of the European Parliament and of the council of April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (GM)	Individuals leaving their laboratory (i.e. because of retirement) must be anonymised in the database (GM)	No (GM)	Administrative (laboratory, year, postcode, # employees, origin and reason for request, workplace area); sampling conditions (sampling date, method, device, flow rate, area vs personal, representativeness of sample indicated by sampler, task, occupation, PPE, sampling duration, observations on sampling); exposure conditions (type of workplace, global ventilation, collective protection, estimated production activity, ambient temperature, type of process, frequency of exposure, category of product); and analytic conditions (analysis laboratory, chemical agent, analysis technique) (24)	In 2016, the three agents most frequently measured in COLCHIC were inhalable particulate fraction (38,304 personal, 23,476 area), toluene (20,274 personal, 10,935 area), and acetone (19,790 personal, 8,678 area) (24)

5	France	SCOLA	Sampling strategy is used to assess representative exposure (GM)	*"There is a potential that only companies proactive in IH submit results"* (JL)	COLCHIC follows the EU regulation 2016/679 of the European Parliament and of the council of April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (GM)	Individuals leaving their laboratory (i.e. because of retirement) must be anonymised in the database (GM)	No (GM)	Administrative (laboratory, year, postcode, # employees); sampling conditions (sampling date, method, device, flow rate, area vs personal, representativeness of sample indicated by sampler, task, occupation, PPE, sampling duration, observations on sampling); exposure conditions (type of workplace, global ventilation, collective protection); and analytic conditions (analysis laboratory, chemical agent, analysis technique) (24)	In 2016, the three agents most frequently measured in SCOLA were asbestos (45,001 personal, 25,205 area), wood dust (16,055 personal, 485 area), and alveolar particulate fraction (6,697 personal, 173 area) (24)
6	Finland	FDOEM)/LIM S	-	The representativeness of companies approaching FIOH for occupational hygiene services is unknown but larger firms may be overrepresented. The customers vary by industry e.g. most companies in the mining sector are likely to engage FIOH's services; however, few companies within the construction sector engage FIOH's services (TT)	Published reports do not contain identifiable information for companies/workers (TT)	Not required from clients (TT)	-	Measured concentration, type of measurement (personal vs area), PPE used, repeatability and duration of exposure, occupation, working task etc. Modelled on information included in Stoffenmanager programme (TT)	In 2014, the most commonly measured agents in the database were styrene, chromium and chromium VI, formaldehyde and other mineral dusts (65)
7	Denmark	ATABAS	-	The data contained in ATABAS were from inspections of workplaces where there are OHS problems, and thus the results were not representative of all workplaces (1)	-	-	-	-	-

8	Norway	EXPO	NA	It has been estimated that STAMI received ~20% of all airborne exposure measurements conducted in Norway between 2001-2005 (RO)	In general, individual workers cannot be identified but they could be indirectly identified in special cases (RO)	A data protection impact assessment (DPIA) has been performed (RO)	Companies own the collected data but STAMI has the right to use the de-identified data for research and to create summaries and statistical overviews (RO)	EXPO: *"Sampling strategy, the conditions when sampling compared to normal (subjective), inside or outside, wind, work area (where the task is performed, e.g. laboratory, production hall), sampling time, sampling method, method of analysis, laboratory of analysis, gender, year of birth, country of origin, occupation, work task, use of respiratory protection, frequency and duration of the task, number of other employees with the same job type (exposure)"*. EXPO Online: *"Free text field (250 characters) for the company to add information not added elsewhere"* (RO)	Exposures include silica dust, wood dust, solvents, and welding fumes (RO)
9	South Korea	WEMD	No (SP)	No (SP)	Trade secret issues (SP)	-	*"Parliament mandated the definition and process of trade secret last year"* (SP)	Work process, industry, occupation, sampling method, TWA, working time, department, company, date and time, measurement duration, analytic method (SP)	-
10	Singapore	NDNCE	No (EK)	*"The number of workplaces who submit noise and toxic substance exposure reports make up a fraction of all workplaces where workers are likely subjected to excessive noise and toxic substances. MOM is working towards increasing its surveillance footprint to cover more workplaces where workplace health hazards are likely to arise"* (EK)	MOM does not release data on individual workers or workplaces (EK)	-	-	Measurement results are submitted using online form templates. Noise: type of monitoring, equipment, work process, number of persons exposed, measured decibel level, measurement time, duration of exposure. Toxic substances: name of substance, number of persons exposed, type of sample (area vs. personal), sampling method, process, duration of process, control measure, monitoring duration, PEL standard, concentration measured, TWA (EK)	-

11	Switzerland	SUVA	*"No. Data cannot be attributed to individual workers. Irregular patterns of exposure are sometimes accounted for in the OELVs"* (PS)	SUVA covers about half of Switzerland's workforce but the measurements carried out by sector is highly variable (PS)	The database has to comply with the Privacy Act. Individuals are anonymised in the database (PS)	SUVA has the right to retain records of the exposure measurements that they conduct (PS)	*"Not for the database, but we accept these issues on a case-by-case basis for the documentation of measurements in companies"* (PS)	Name of company, SUVA code for industry, NOGA code for industry, date of measurement, general ventilation, temperature, humidity, air pressure, ID of measurement location, sampling method, analysis method, personal vs. area sampling, measurement strategy (worst-case/TWA/background etc), measurement location (free-text), process, activity, control measures, measurement concentration, duration of measurement (PS)	-

Table A2.3 Data uses										
#	**Country**	**Name of database**	**ab. Who is the data aimed at?**	**ac. Who or what are the main users of the information?**	**ad. Is the data linked to other databases?**	**ae. How is the data presented?**	**af. If reports are generated, what is the frequency?**	**ag. Are the data freely available for third parties?**	**ah. Is it intended that the data might be available for transfer to other databases?**	**ai. Ability to assess longitudinal trends?**
1	Canada	CWED	The CWED was developed as part of CAREX Canada and therefore has been limited to informing CAREX Canada as well as a few other specific research projects. The long-term aim of CWED 2.0 is to make the database more widely available for research and policy		No	-	-	One of the aims of CWED 2.0 is to make data available to third parties (HD)	No	Yes, e.g. documenting historical trends to isocyanates (34)
2	U.S.	IMIS	In addition to use by OSHA, a considerable number of research papers have been published. A 2013 literature review identified 29 publications that had used IMIS data for a variety of purposes, including exposure surveillance and tracking trends over time (36)		The IMIS has about ~30% overlap with the Chemical Exposure Health Data (CEHD) which comprises measurements from samples analysed at the central Salt Lake City laboratory (36)	-	-	Data are available by request and for a processing fee through the Freedom of Information Act. NIOSH has an inter-agency agreement to use the data	-	Yes; many authors have reported that temporal trends are comparable to other sources of data, implying that at least extrapolation of relative time trends from these data may be reliable (36)
2	U.S.	OIS	OSHA		*"OIS has interface with the U.S. Treasury, Lockbox and Pay.gov for penalty payment and debt collection activities. There are also seven interface states where case data are transmitted to OIS from state systems. In addition, health sampling data is transmitted from both Salt Lake City Technical Center (SLTC) and the Wisconsin Occupational Health Laboratory (WOHL)"* (GS)	*"OIS data can be viewed within the system. A variety of letters, documents and reports can be generated for printing or*	Reports can be generated by users (GS)	OIS is accessible to authorised users only (GS)	The OIS data can be transferred to other Department of Labour databases (GS)	-

						exported for viewing and editing electronically." (GS)				
3	U.K.	NEDB	The data was used by HSE inspectorates and policy makers (including the Advisory Committee on Toxic Substances responsible for standard-setting) for prioritisation of interventions and as a training tool for inspectors (1)		No	-	-	Restricted to HSE (1); however it is possible that the data is available to researchers on request	NA	-
4	Germany	MEGA	The database is used by IFA and the accident insurers for prevention through identification of hazards, assessing the efficiency of risk control measures, and the determination of the technical criteria of exposure limit values. It is also used for epidemiological purposes and for the investigations of occupational disease claims (57)		No	Reports and publications: https://www.dguv.de/ifa/gestis/expositionsdatenbank-mega/expositionsdaten-aus-mega-in-publikationen/index-2.jsp	~1 large report per year, plus several smaller reports (DK)	No direct access to raw data outside of the institutions for accident insurance (DK)	No	*"Yes, as we already have several decades of data for many hazardous substances"* (DK)
5	France	COLCHIC	INRS for the prevention of work-related injury and disease	*"Industries, hygienists, legislator…"* (GM). Data are used to define priorities for national prevention programmes	No (GM)	*"It depends how the extraction is made and who makes the extraction. If we make an extraction from the software,*	Reports are available on request, ~15 reports per year (GM)	Access to the data is restricted (GM)	Statistical data is available via online tools such as SOLVEX: https://www.inrs.fr/publications/bdd/solvex.html and	Yes, but not directly from the database (R software is used) (GM)

						the data were presented in a excel spreadsheet or a word form. If we used a programming language (sql and R in our case), we can extract the raw data in a raw file, or an excel spreadsheet" (GM)			FIBREX for fibre measurements: https://www.inrs.fr/publications/bdd/fibrex.html	
5	France	SCOLA	INRS for the prevention of work-related injury and disease	Data are used to define priorities for national prevention programmes	No (GM)				Statistical data is available via online tools such as SCOLAMIANTE http://scolamiante.inrs.fr/Scolamiante/Accueil	
6	Finland	FDOEM)/LIMS	The main data user is FIOH; the data is used to track trends in exposure, facilitate the efficient production of client reports; and to conform to the quality standard ISO-17025 for testing and calibration laboratories (TT)		Used for FINJEM (65)	Reports often require "manual re-working of spreadsheets derived from the database" (TT)	Agent-specific and task-specific reports are available for specified timeframes or working environments (TT)	Not freely available	-	Yes
7	Denmark	ATABAS	-	-	-	-	-	-	-	-

8	Norway	EXPO	The main users of EXPO include the Department of Occupational Health Surveillance (NOA) at STAMI for national statistics and trends, researchers, and occupational physicians. For EXPO Online the main users are the *"companies, to ensure that this data is stored and available for future HSE-work"* (RO)	No	*"Electronic, aggregated level"* (RO)	-	Can be made available with restrictions (RO)	No	Yes
9	South Korea	WEMD	Ministry of Labour, researchers, and workers who claim occupational disease. OSHRI researchers have also established a cohort study from the data (SP)	The exposure database has recently been linked to the workers' health data to form a central worker health database (SP)	Electronic files (SP)	Annual report (SP)	Data access by third parties is limited due to trade secrets. Researchers can access data through a "data handling process" at OSHRI (SP)	The exposure database has recently been linked to the workers' health data to form a central worker health database	Not previously but there is an intention to start from the present (SP)
10	Singapore	NDNCE	OSHD staff (EK)	No	Aggregated data is presented in electronic reports (EK)	Annual Workplace Safety and Health National Statistics reports (https://www.mom.gov.sg/workplace-safety-and-health/wsh-reports-and-statistics)	Only aggregate data is published in annual reports	No (EK)	Yes (EK)

11	Switzerland	SUVA	SUVA (including labour inspectors), occupational medicine specialists (PS), and institutes of occupational health (NS)	No (PS)	Reports (hard copy and electronic) and SAP-based database (PS)	A company report is generated after each measurement campaign and an annual report on the numbers of measured substances (PS)	No, SUVA must grant access (PS)	No (NS)	Yes, if the sector is large enough (PS)

Table A2.4 Resource requirements														
						Data Collection		Data Maintenance		Data Establishment			
#	**Country**	**Name of database**	**aj. Who funds the database?**	**ak. How was the set-up of the database funded?**	**al. How is the ongoing maintenance and development of the database funded?**	**am. What are the human resource requirements (skills and time) for data collection/entry?**	**an. What are the ongoing, non-salary costs associated with data collection and entry requirements?**	**ao. What are the human resource requirements (skills and time) for maintenance of the database?**	**ap. What are the ongoing, non-salary costs associated with maintaining the database?**	**aq. What were the human resource requirements for establishment?**	**ar. What were the non-salary expenditure costs?**	**as. What data storage system is used – off the shelf or purpose-designed?**	**at. If off the shelf, which software package is used?**
1	Canada	CWED	Research grants (HD)	Research grants (HD)	Research grants: in 2018, WorkSafeBC funded the University of BC for a CWED 2.0 project to "help sustain and grow" the CWED. The aims of the project are to: "a) secure the existing CWED; b) make data more widely available to researchers and knowledge users; c) improve data management; e) improve stakeholder knowledge of the database; f) explore database potential; and g) explore sustainable funding models" (22)	*"Coding requires occupational hygiene expertise"* (HD). For the 2014 Capacity Development Report (31) to add data from SK and BCMEM, students were employed to abstract data (on-site for SK) and were supervised by occupational hygienists. All data were cleaned, and occupation, industry, and analyte variables were coded	For the 2014 Capacity Development Report (31) to add data from SK and BCMEM, the abstraction was less than $4 per measurement record	*"Variable – we do what we can with available funds"* (HD)	*"Variable. We do not have a set budget"* (HD)	*"High!! Difficult to answer it has been a long process. But biggest effort is in (1) data abstraction from paper records and (2) coding jobs and industries"* (HD)	Estimated in the range of 500,000-1m Canadian dollars (HD)	A database that guided data abstraction and performed primary validation tests was front-end built for the Capacity Development Report (31). A new IT platform is being investigated for CWED 2.0.	
2	U.S.	IMIS	OSHA	OSHA	OSHA	Each OSHA inspector was responsible for documenting the outcome of each inspection, including entry of exposure measurements into IMIS	-	-	-	-	-	-	-
2	U.S.	OIS	OSHA	OSHA	OSHA	-	-	*"OSHA awarded a contract to a contractor for maintaining the database"* (GS)	*"Not sure"* (GS)	*"Not sure"* (GS)	*"Not sure"* (GS)	Purpose-designed	NA

3	U.K.	NEDB	HSE	HSE	NA	*"Early data was collected by inspectors with limited occupational hygiene knowledge and experience. More recent data collection is through suitable qualified occupational hygienists. Data entry is by administrative staff. There is also some data collected from specific sites visited as part of HSE exposure surveys"* (PB)	Unknown (PB)	*"Not known. Limited maintenance of the database by one full time administrative staff on an ad-hoc basis"* (PB)		Unknown (PB)		*"The database was configured on a VAX 11/750 mini computer housed in an HSE headquarters building in Bootle. The data now is on a SQL Server 2008 R2 database instance and this gets backed up daily. We retain the full backups for 66 hours locally, but the backup files will also go to tape and be subject to the standard retention policy. The front end application is built on the .NET Framework (version 3.5). The NEDB application software and the link to the standard graphics software (DEGRAPH) were provided by Computer Design Systems Ltd. The application software is a bespoke system chosen following consideration of this option and standard packages. The language is Digital Standard MUMPS (DSM) which originated in the USA"* (PB)	
4	Germany	MEGA	IFA internal funds	IFA internal funds	IFA internal funds	*"Skills: knowledge about occupational hygiene and the workplaces/techniques to select places for the sampling; engineer knowledge about the sampling devices at the measuring services; analytical knowledge for the chemical analytic; IT-knowledge for the data base development and programming; statistical and occupational hygiene knowledge for the running of the database and the statistical evaluations. Time: multiple processes are involved so too difficult to estimate"* (DK)	Difficult to estimate (DK)	*"Skills: Many years of knowledge of the whole system (sampling, analysis and storage). Time: Difficult to say, but more than you would think at the beginning. We estimate about ½ a person per year"* (DK)	Difficult to estimate (DK); a previous estimate from the 2006 NOHSAC reported an estimate of Euro 10,000 per year (1)	Two IT staff, two staff for the *"competence on workplace exposure"* (DK)	*"IT equipment, e.g. data storage capacity. We can't estimate the amount of money needed"* (DK)	Purpose-designed OMEGA (Organisation System for the collection and use of measured data on exposure to hazards at the workplace) is used for data collection. The system is based on coding lists, specialised files and record files and links data from measuring services to analysis at IFA through to documentation in MEGA. IFA-developed software is used for the documentation of exposure-related variables in the field (up to 200 variables can be collected for each sample including task, room volume, ventilation, LEV) and for the database software called MEGAPro which can be used for the extraction and statistical analysis of data (54). For data storage, purpose-designed MEGA software is used. Both are based on Firebird (DK)	-

5	France	COLCHIC	INRS	The French National Health Insurance Fund for Salaried Workers (CNAM-TS)	INRS funds the maintenance of COLCHIC	Each laboratory collects its own data. On file completion, the laboratory manager certifies the file and proceeds with its storage, which leads to automatic recording on the COLCHIC database (24)		An external IT developer is used and testing is carried out by 2 INRS staff (GM)	~30 000 Euros for annual maintenance and development including correcting bugs and improving the database (GM)	3 personnel and an external IT developer (GM)	The development of the database cost ~400 000 Euros (GM)	The storage system is IBM DB2 (GM)	-
5	France	SCOLA	INRS	French Directorate-General for Labour (DGT)	INRS	Each laboratory collects its own data. On file completion, the laboratory manager certifies the file and proceeds with its storage, which leads to automatic recording on the SCOLA database (24)							
6	Finland	FDOEM)/LIMS	FIOH	-	-	-	-	-	-	-	-	-	-
7	Denmark	ATABAS	-	-	NA	NA	NA	NA	NA	-	-	-	-
8	Norway	EXPO	STAMI	The Norwegian Government	STAMI	-	-	A part-time person with the responsibility to maintain and secure the integrity of the database (RO)	-	-	-	SQL database (RO)	
9	South Korea	WEMD	OSHRI	OSHRI	OSHRI	3 full-time staff (SP)	-	-	-	-	-	MS Oracle and MS ACCESS (SP)	
10	Singapore	NDNCE	MOM (i.e. Government)	According to the 2006 NOHSAC report, "the establishment costs were Singapore $1m and the ongoing maintenance is estimated to be Singapore $100,000 payable to the external operator" (1)		Knowledge in data extraction and data mining tools (EK)	*"Not available"* (EK)	Knowledge in data extraction and data mining tools (EK)	*"Not available"* (EK)	IT knowledge and data extraction (EK)	*"Not available"* (EK)	*"Both"* (EK)	Power Pivot, BI tool
11	Switzerland	SUVA	Swiss insurance fund (NS), which is funded by employers and employees	Swiss insurance fund (NS)	Swiss insurance fund (NS)	*"Measurement technicians under surveillance of occupational hygienist, about half an hour for a measurement campaign consisting of 10 – 50 values"* (PS)	None (PS)	*"Occupational hygienists, software engineers, about 40 h/year each"* (PS)	*"About 35'000 CHF annual fee for hosting and maintenance. Rather high, due to high quality and security standard and technical complexity of Suva's overall IT-*	*"Project team consisting of 3 members of the lab and 4 software engineers, total 900 hours"* (PS)	*"Establishment almost none"* (PS)	Off the shelf	SAP CRM for entry/storage, SAP BW for the reports

									infrastructure" (PS)				

Table A2.5 Data Governance			
#	**Country**	**Name of database**	**au. How are decisions made on its use, development, expansion to other hazards etc?**
1	Canada	CWED	Revised governance structure under development as part of CWED 2.0
2	U.S.	OIS	*"Decisions on its use, development and expansion are made by OSHA officials based on changes to OSHA regulations, changes in program policy, operational and maintenance needs, technical enhancements and available funding"* (GS)
3	U.K.	NEDB	Unknown (PB)
4	Germany	MEGA	All decisions are made in cooperation with the accident insurers within the MGU Monitoring System (DK)
5	France	COLCHIC	INRS in accordance with the French National Health Insurance Fund for Salaried Workers (CNAM-TS)
5	France	SCOLA	INRS in accordance with the Labour Ministry
6	Finland	FDOEM)/LIMS	-
7	Denmark	ATABAS	-
8	Norway	EXPO	There is an advisory board with representatives from the Labour union (Norwegian Confederation of Trade Unions), Labour Inspection Agencies (Petroleum Safety Authority Norway and The Norwegian Labour Inspection Authority) and Employers organisation (The Confederation of Norwegian Enterprise) and STAMI. New agents are added when *"new hazardous airborne agents emerge (method of analyses is established)"* (RO)
9	South Korea	WEMD	OSHRI
10	Singapore	NDNCE	*"The use, development and management of data is guided by data governance policies for the whole-of-government in Singapore"* (EK)
11	Switzerland	SUVA	SUVA's Department of Health Protection (PS)

Table A2.6 Barriers and enablers								
				Establishment		Further development		
#	**Country**	**Name of database**	**av. What led to the establishment of the database?**	**aw. What were the identified barriers to the establishment of the database?**	**ax. What were the identified enablers to the establishment of the database?**	**ay. What are the identified barriers to further development of the database?**	**az. What are the identified enablers to further development of the database?**	**ba. Can a subjective assessment of the system usefulness be made?**
1	Canada	CWED	The CWED was established by the CAREX Canada project to support the estimation of exposure prevalence to occupational carcinogens	*"Funding and identifying data sources"* (HD)	*"Perseverance by researchers"* (HD)	In 2009, a data holders' survey of all relevant agencies was conducted which identified a downward trend in exposure measurements by regulatory agencies since the 1990s. Most agencies were no longer responsible for collecting the majority of exposure measurements in their jurisdiction (6)	Hugh Davies (Associate Professor at the University of BC) is lead investigator on the project to develop and seeking funding for CWED 2.0	*"Yes – contributions to research, publications"* (HD)
2	U.S.	IMIS	IMIS was designed to collect, process, retrieve, and communicate penalty assessment, arbitration, and monetary collection information regarding OSHA's inspections (37)	-	-	-	-	-
2	U.S.	OIS	*"OIS was developed to replace an outdated legacy application and to expand the support for other program and strategic needs of OSHA"* (GS)	*"A completely new system based on new technology needed to be developed for use by over 2,800 users nationwide. There is also challenges in transferring cases from the legacy system to the new system or maintaining reporting on cases in dual systems (legacy and OIS) during the transition period"* (GS)	*"Basic format and requirements of the legacy system can be duplicated in OIS and thereby saving time and effort in development"* (GS)	*"Finding funding and resources, and prioritizing development items are major barriers"* (GS)	*"Having a good working system supported by knowledgeable and skilled staff creates a good foundation for further enhancement and development of the system"* (GS)	*"OSHA's position is that the system meets the operational needs of the agency and has the ability to continue meeting operational needs into the future given continued funding for system maintenance and enhancements"* (GS)

3	U.K.	NEDB	The NEDB "was intended to be an institutional resource on the actually occurring levels of industrial exposures, and of the situations in which these levels can arise" (1)	*"Suitable database to host the information and Windows compatible software interface. Access to industry data"* (PB)	None (PB)	Changes in the structure at HSE led to a decline in the collection of exposure measurements and a shift in focus from measuring to controlling (48). *"Applicability of the data. Lack of new data. Lack of industry data"* (PB)	-	*"The National Exposure database (NEDB) has been successfully used for over 20 years to support HSE policy and enforcement activity. The data is also important for research and development. Examples include 'Trends in inhalation exposure – Mid 1980s till present' and 'development of the EASE model for exposure estimates. These applications all contribution and have a major impact to health risk management in the UK and other countries. This view is supported by the referencing of this work in other research papers and application of the EASE model for EU chemical regulation"* (PB)
4	Germany	MEGA	*"The idea that data, which were collected for consultation and guidance of companies, could be used for general preventive guidance, if collected in a database"* (DK)	*"Data protection and data security issues. At the time of the establishment of the database, about 30 accident insurers existed. They all had to agree on the process e.g. finding of uniform criteria for data collection"* (DK)	*"The need for exposure data for prevention and cost reduction using already collected data"* (DK)	See (aw.)	*"The need for data if limit values are new established or old ones lowered. The need for data in case of new medical knowledge about work related diseases"* (DK)	*"As far as we know, MEGA is the biggest database on workplace exposure to hazardous substances. The many national and international requests for statistical evaluations and cooperation show us the usefulness"* (DK)
5	France	COLCHIC	COLCHIC was set up at the request of the Commission of Work Accidents and Occupational Diseases (24)	-	-	Economic barriers (GM)	*"Improve knowledge on occupational exposure to chemicals and biologicals"* (GM)	-

5	France	SCOLA	In France, exposure measurements made during assessment of compliance to regulatory occupational exposure limits are recorded in a national register called SCOLA. The measurements are performed by independent certified laboratories and the data are used to define priorities for national prevention programmes (24)	-	-		-	The data contained in SCOLA has grown from 11 substances with more than 1,000 measurements reported in 2013 to ~1 million results for 150 chemicals in 2019 (GM). In 2016, Mater et al predicted that SCOLA will be the major source of information in the coming decades (24)
6	Finland	FDOEM)/LIMS	-	-	-	-	-	-
7	Denmark	ATABAS	-	-	-	-	-	-
8	Norway	EXPO	*"The need to know more about the exposure in different occupations and industries to prevent work-related disease"* (RO)	*"It is important to establish technical solutions for a user-friendly registration solution, EXPO Online, especially since it is not mandatory to register the data"* (RO)	*"The benefits of cost free and secure data storage for the companies"* (RO)	-	*"Improvement of the benefits of using the system and user-friendly interface. Continuously, maintain and develop according to technical developments and expectations"* (RO)	*"It is dependent on companies in Norway to continuously add their exposure measurements to achieve the goal of a representative national database on exposure. Since registration is not mandatory, it might take more time to achieve this goal in all industries. Otherwise, the database will be skewed"* (RO)
9	South Korea	WEMD	*"Government annual reporting was the main impetus"* (SP)	*"The awareness of the leadership in Government and KOSHA" (SP)*	*"Request from academics, NGOs". The requirement for a cohort study by researchers" (SP)*	*"Further input of full-time workers is limited" (SP)*	*"Request from academics, NGOs, Parliament" (SP)*	The database has the potential to contribute to effective and evidence-based policy as well as research (SP)
10	Singapore	NDNCE	The monitoring of exposure to noise and toxic substances being a regulatory requirement (EK)	*"Manpower resources, poor knowledge of source data and transformation of data to readable format/code"* (EK)	*"Knowledge in the new tool to be used"* (EK)	-	-	*"The hygiene surveillance system allows OSHD to formulate targeted workplace health programmes based on the data submitted, to help reduce workplace health*

								risks. This can be seen from the reduction in occupational diseases cases, in particular, noise induced deafness (NID) cases over the years" (EK)
11	Switzerland	SUVA	Prior to the establishment of the database, the data were held in archived paper records and therefore there was a need for better accessibility for different evaluations. *"Complex measurement campaigns with lot of measured substances were hard to classify/refind"* (PS)	*"High costs of measurements and inter- and intra-worker variability" (NS). "Knowledge, resources, conservatism"* (PS)	*"Availability of funding and resources of occupational hygienist, external pressure"* (PS)	See (aw.)	See (ax)	*"Database query has become standard to answer complex OH-questions, benchmarking of individual results and to identify fields of actions"* (PS)

Table A2.7 Challenges and Advice				
#	**Country**	**System**	**bb. Challenges**	**bc. Advice**
1	Canada	CWED	-	-
2	U.S.	IMIS/OIS	-	-
3	U.K.	NEDB	*"The use of NEDB data for HSE research is limited with only one report making significant use of the data in the past 15 years. The representativeness of the data and difficulty in access is frequently stated. This is partly due to bias in selection of the sites where data has been collected and is determined by HSE interest in specific substance or process"* (PB)	*"There is a clear potential for the use of exposure data to support HSE business needs. Amongst other things the data is extremely useful for research in exposure trends and development of exposure models. The more data placed into NEDB the more applicable the database will be as current and historic drivers for obtaining data introduced a bias that needs to be taken into consideration. It has been shown that industry data can be successfully integrated into NEDB and could address this. However, more needs to be done to exploit the potential of NEDB. This requires: • Improving accessibility and availability of the database both internally and externally • Increasing awareness of the data and its potential use. • Identification of new applications to increase data gathering and resource to deliver this"* (PB)
4	Germany	MEGA	-	*"Before planning the database, it is important to ensure, that you will have access to data, you can later store in the database; It is important to ensure a structured information collection during the whole process, e.g. through a purpose-designed software; It is important to have specialists from different fields in the development team, e.g. occupational hygienists, engineers/technicians, chemists, statisticians; In order to pass on experience and knowledge about the database it is very important to have a balanced age structure in the team"* (DK)
5	France	COLCHIC	*"The representativeness of the exposure recorded in a database is the most difficult challenge. Knowing the limitations of the information collected is important for conducting an analysis"* (GM)	*"1) define the need; 2) identify purpose; 3) use reference tables to characterise information; 4) record information in a database takes a lot of time; database must also be useful for the user"* (GM)
5	France	SCOLA	-	-
6	Finland	FDOEM)/LIMS	-	-
7	Denmark	ATABAS	-	-
8	Norway	EXPO	-	*"Development is not a sprint and a thorough planning phase is very important"* (RO)

9	South Korea	WEMD	*"The ultimate purpose of exposure data might be the measure of the health result of the exposed workers. If you have a plan to construct exposure data, you would better to expect the connection with the health examination or death certificate, or cancer registries, so that you do not have to revise it again"* (SP)	*"Exposure database system development is not optional, but compulsory in modern occupational safety and health management, specifically in national OSH level. Each country has developed their own system comparable with the social structure (legal and industrial system). Thus, analyse first the data you already have, and then structure them and then reorganize of revise your system" (SP)*
10	Singapore	NDNCE	-	*"We would suggest to have a wider coverage of surveillance data collection for purposes of policy formulation and intervention programmes" (EK)*
11	Switzerland	SUVA	*"Insufficient measurements to cover full exposure distribution" (NS) "Quality assurance of the entries is quite tedious; Integration into a LIMS would facilitate a lot and limit changes in medium; Consistent and meaningful entries need a lot of instruction of the technicians"* (PS)	*"More contextual data must be collected regarding the details on the activity tasks performed and controls applied" (NS). "Overall, we are quite happy with our database. I think, it was worth keeping it simple, as the experience of others with more sophisticated coding (including job descriptions, materials handled and so on) shows that in real life these entries are either avoided or (if compulsory) filled out with poor quality"* (PS)

Appendix 3: Data holders' survey questionnaire

Start of Block: Questions about your organisation

Q1 Does your organisation provide occupational hygiene (OH) services?

- Yes, we provide OH services to one company (i.e. in-house) (6)
- Yes, we provide OH services to a range of companies (7)
- Yes, but a third-party company conducts the sampling (8)
- No, I am a Government employee (9)
- Other, please specify: (10) ____

Skip To: End of Block If Does your organisation provide occupational hygiene (OH) services? = No, I am a Government employee

Q2 In which region(s) is your organisation located?

- Northland (1)
- Auckland (2)
- Waikato (3)
- Bay of Plenty (4)
- Gisborne (5)
- Hawke's Bay (6)
- Taranaki (7)
- Manawatū-Whanganui (8)
- Wellington (9)
- Tasman/Nelson (10)
- Marlborough (11)
- West Coast (12)
- Canterbury (13)
- Otago (14)
- Southland (15)

Q3 What region(s) does your organisation provide OH services for? (if different from above)

- Northland (4)
- Auckland (5)
- Waikato (6)
- Bay of Plenty (7)
- Gisborne (8)
- Hawke's Bay (9)
- Taranaki (10)
- Manawatū-Whanganui (11)
- Wellington (12)
- Tasman/Nelson (13)
- Marlborough (14)
- West Coast (15)
- Canterbury (16)
- Otago (17)
- Southland (18)

Q4 How many occupational hygienists work for your organisation?

- 1 (1)
- 2-5 (2)
- 6-10 (3)
- 10+ (4)

End of Block: Questions about your organisation

Start of Block: Questions about your OH data

Q5 Does your organisation use a standard data collection form for each OH visit?

- Yes (1)
- No (2)

Q6 What are the five most commonly sampled agents by your organisation?

- Metals (4)
- Inhalable dust (5)
- Respirable dust (13)
- Noise (6)
- Welding fumes (7)
- Formaldehyde (8)
- Respirable crystalline silica (9)
- Carbon monoxide (10)
- Nitrogen dioxide (11)
- Wood dust (12)
- Volatile organic compounds (14)
- Isocyanates (15)
- Diesel particulate (16)
- Other, please specify: (17) ____

Q7 What are the three most common industries your clients are from?

- Accommodation & Food Services (4)
- Agriculture, Forestry & Fishing (5)
- Construction (6)
- Electricity, Gas, Water & Waste Services (7)
- Healthcare & Social Assistance (8)
- Manufacturing (9)
- Mining (10)
- Transport, Postal & Warehousing (11)
- Wholesale Trade (12)
- Retail Trade (13)
- Public Administration & Safety (14)
- Other, please specify: (15) ____

Q8 What is the average size of the companies that you mainly provide OH services for?

- 1-5 employees (1)
- 6-19 employees (2)
- 20-49 employees (3)
- 50-100 employees (4)
- 100+ employees (5)
- It varies widely (7)
- Don't know (6)

Q9 Approximately how many samples does your organisation collect per year?

____

Q10 What are the main reasons for clients engaging your services? (multiple options possible)

- Corporate policy (1)
- Compliance e.g. as a result of a WorkSafe Notice (2)
- As a result of previous sampling/continued surveillance (3)
- Follow-up from risk assessment (4)
- To achieve standard accreditation (5)
- To evaluate if controls are working (6)
- Other, please specify: (7) ____

Q11 How does your organisation store OH data?

- Hard copy paper files (1)
- Individual electronic reports (e.g. PDF) (2)
- We don't retain OH data after reporting to clients (3)
- In an electronic database (e.g. Microsoft Excel or Access), please specify electronic format: (4) ____
- Other, please specify: (5) ____

Q12 How long is OH data stored for (please specify months or years)?

____

Q13 Do you collect data on (please select all that apply):

- Job title of workers (1)
- Industry of workers (2)
- Gender of workers (3)
- Ethnicity of workers (4)
- Number of employees (5)
- Reason for sampling (6)
- Work process related to sampling (7)
- Sampling duration (8)
- Exposure duration or frequency (9)
- Measurement strategy (i.e. representative or worst-case) (10)
- Engineering controls (11)
- Use of Personal Protective Equipment (12)

End of Block: Questions about your OH data

Start of Block: Your thoughts on a New Zealand worker exposure database

Q14 If an exposure measurement database was set-up in New Zealand, who do you think should host the database?

- WorkSafe New Zealand (1)
- An academic institute/university (2)
- New Zealand Occupational Hygiene Society (NZOHS) (3)
- Health and Safety Association of New Zealand (HASANZ) (4)
- A private company (5)
- Other, please specify: (6) ____

Q15 Who do you think should have access to an exposure measurement database if data were anonymised? (select all that apply)

- Government (i.e. WorkSafe) (1)
- Occupational hygienists (2)
- Researchers (3)
- Businesses (4)
- Open access (i.e. general public) with restrictions (5)

Q16 If you were asked to contribute data, would you need to seek consent from your clients to contribute newly collected data to a centralised database?

- Yes (1)
- No (2)
- Don't know (3)

Q17 If you were asked to contribute data, are there specific privacy requirements that would need to be addressed?

____
____
____
____
____

Q18 If you were asked to contribute data, are there intellectual property issues that would need to be addressed?

____
____
____
____
____

Q19 Are there any other perceived barriers to you/your clients contributing data to a centralised database?

____
____
____
____

____

Q20 Do you think a centralised database of occupational exposure measurements should be set up in New Zealand?

- Yes (1)
- No (2)

Q21 How willing are you to contribute to a New Zealand centralised database of exposure measurements on a scale from 1 (unwilling) to 10 (very willing)?

	1	2	3	4	5	6	7	8	9	10
Willingness to contribute ()										

Q21a OR I would only be willing to contribute if:

____
____
____
____
____

Q22 If you were willing to contribute data, would you contribute (please select one):

- Prospective (newly collected) data only (1)
- Prospective and historical data (2)
- Historical data only (3)

Skip To: Q27 If If you were willing to contribute data, would you contribute (please select one): = Prospective (newly collected) data only

Q23 Would you need to seek consent from your clients to contribute anonymised historical data?

- Yes (1)
- No (2)
- Don't know (3)

Q24 How many years do you have historical OH data for?

____

Q25 What would be the likely time/cost commitment for extracting historical data?

- Minimal (1)
- Moderate (2)
- Significant (3)
- Don't know (4)

Q26 Would you charge for the extraction of historical data?

- Yes (1)
- Maybe (4)
- No (2)
- Don't know (3)

Q27 Would you charge for the ongoing contribution of newly collected data?

- Yes (1)
- Maybe (2)
- No (3)
- Don't know (4)

Q28 Any other comments (e.g. on the likely representativeness of the data):

____
____
____
____
____

End of Block: Your thoughts on a New Zealand worker exposure database

Appendix 4: Description of other exposure surveillance methods in New Zealand

a. *Surveys of working conditions*
Prevalence data for a wide range of self-reported occupational exposures is available from the Centre for Public Health Research's (CPHR) two previous workforce surveys: one of the general population (2004-2006; 3,003 participants) and one of the Māori population (2009-2010; 2,107 participants). In 2018, a similar survey (WorkSafe RFP) was also carried out in seven targeted occupational groups (~100 in each for community-based nurses, collision repair, construction, hospitality, clerical, sawmill, and agricultural workers). WorkSafe's '*OccIdeas*' survey of carcinogen exposure prevalence is currently being conducted and will also contribute to this knowledge base.

Whilst workforce surveys are valuable (particularly for certain risk factors such as psychosocial factors) and can tell us where exposures are present, they do not provide information on the actual level and intensity of dust/chemical exposures in relation to levels at which adverse health effects may occur. The previous workforce surveys included self-reported data which is inevitably subjective and respondents may not necessarily be aware of or know the names (e.g. specific chemical components) of exposures in their environment. In addition, workforce surveys are often cross-sectional and therefore represent a 'snapshot' of exposure at one point in time. An ongoing exposure surveillance system based on workforce surveys would require a long-term commitment to regular surveys, which would be invaluable but expensive and time-consuming.

b. *Registers of exposed workers or products*
The only register of exposed workers in New Zealand is the Asbestos Exposure Register which was established in 1992 by the former OSH service of the DoL. Notifications to the

Register are voluntary and are made by those who felt that they had been exposed to asbestos or by people acting on their behalf (e.g. relatives, an employer, OSH nurse or union representative). When a notification is made to the Register, a short questionnaire is sent out by the Registrar to be completed by the exposed individual. From 1992-2011, there were ~19,000 registrants.

The only exposure register in New Zealand was the voluntary AER. Examples of successful exposure registers from overseas include ASA in Finland and SIREP in Italy; however, reporting to these registers is a mandatory requirement in these countries.

Product register: Certified handler compliance certificates for hazardous substances
Under the Health and Safety at Work (Hazardous Substances) Regulations 2017, people handling (including using and storing) acutely toxic substances must hold a certified handler compliance certificate or work under the supervision of a person holding a certified handler certificate. Certified handler certificates are issued by WorkSafe authorised compliance certifiers. The certified handler requirement replaced the approved handler test certificate that operated under the Hazardous Substances and New Organisms Act 1996.

Agrichemicals that are classified as acute toxicity (oral, dermal, inhalation) categories 1 and 2 (Globally Harmonised System 7) require the user to be a certified handler. For a certified handler compliance certificate to be issued, the user must demonstrate their knowledge of the regulatory controls for safely handling the agrichemical and have practical experience in the handling of the particular agrichemical(s).

WorkSafe maintains a national register of compliance certificates. Since 1 December 2017 [when the hazardous substances regulations commenced] to March 2021, 9,371 certified handler compliance certificates had been issued, of which 2,116 were issued specifically for agrichemicals. Certified handler compliance certificates must be renewed every 5 years.

c. *Expert information systems*
The CPHR has previously developed a wood dust JEM[97] and more recently developed JEMs for night shift, noise, and sedentary work, all of which involved NZ experts assigning exposure levels for each occupation. All JEMs used NZ-based prevalence data from the two workforce surveys described above; however, the exposure measurement data for noise and wood dust were largely based on overseas data. In 2008, a New Zealand version of CAREX (NZ-CAREX) was developed by CPHR. The NZ-CAREX provides for each industry, an estimate of the fraction of workers within that industry potentially exposed to a list of carcinogens expected to be present in that industry. The estimates are based on the international CAREX evaluations combined with local expert assessment. The NZ-CAREX is available in an ACCESS database but would require updating for current use.

As described above several JEMs and a NZ-CAREX have been previously developed; however these tools are heavily reliant on measurement data from overseas, which may not accurately represent exposure levels in New Zealand workplaces. For example, for wood dust exposure in the NZJEM, only 23% of measurements are from New Zealand worksites.[97]

The Integrated Data Infrastructure

Statistics New Zealand's Integrated Data Infrastructure (IDI) is a longitudinal meta-dataset consisting of information from routine datasets administered by Government agencies (e.g. health, social, economic), Statistics New Zealand surveys (including the 2013 Census), and non-government organisations. The different data sources are linked to individuals in a central spine using births, visa, and tax information. The IDI was established in 2011 and is increasingly used for a wide variety of both policy and research uses. Access to the IDI is governed by strict confidentiality rules and all data are output checked by Statistics New Zealand before release. There is very limited occupational exposure information in the IDI except for occupation, industry and working hours.