

Industrial Hygiene Data Standardization

Past Lessons, Present Challenges, and Future
Directions

BY TAYLOR M. SHOCKEY, MATTHEW M. DAHM,
STEVEN J. WURZELBACHER, AND JOHN BAKER

critical task at the center of the IH profession. Industrial hygienists usually collect occupational exposure measurements using sampling methods that have been validated by OSHA or NIOSH to protect workers from chemical, biological, and physical hazards. But beyond the information required on a standard chain of custody form (sampler/media type, analysis method, sample time, sample volume, and so on), any additional data elements and variables collected are often left to the discretion of the hygienist.

Over the last several years, the profession has seen renewed discussions regarding the standardization of IH data collection. The principal goal of IH data standardization is the widespread adoption of a set of well defined, exposure-relevant variables to harmonize the collection of occupational exposure data across different worksites, companies, agencies, and other entities (see Figure 1). Within an organization, standardized IH data are less prone to error and are more consistent, allowing for easier, more efficient analyses with increased potential for benchmarking. Across organizations or worksites, standardized IH data creates the opportunity for data aggregation, which can help identify exposure patterns or trends that may lead to improvements in occupational safety and health.

CLOSE

ADVERTISEMENT

CHEM DISK™ AND TRACE AIR™ SAMPLERS

Best Overall Value

Personal Monitoring Badges
Smart air sampling solutions

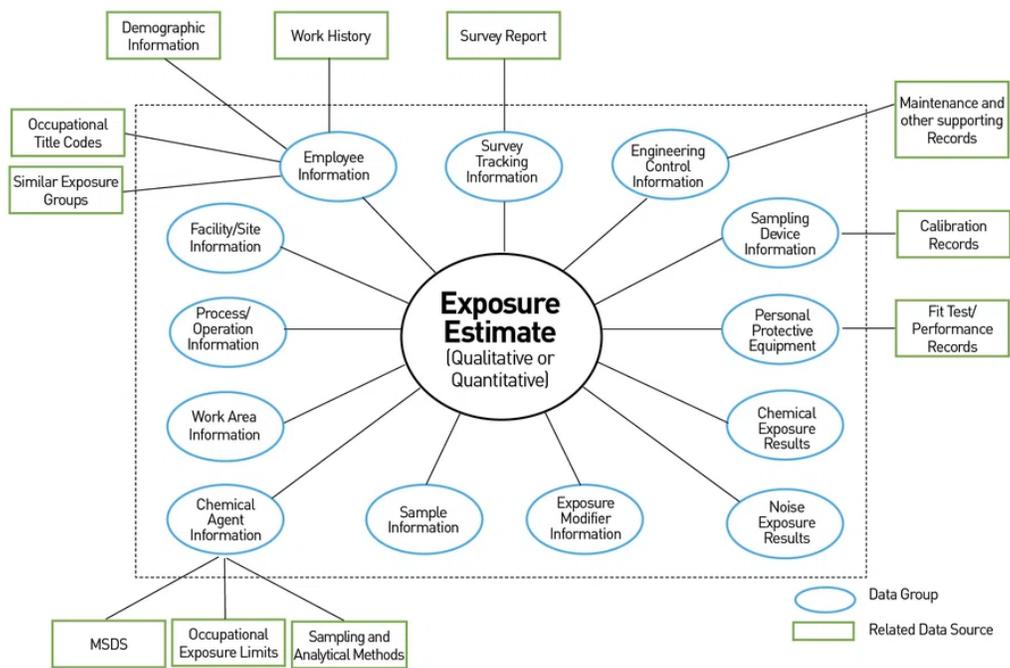
Highest Sampling Rate

Highest Sampling Capacity

assay technology
www.assaytech.com
925-854-4039

The IH field is rapidly evolving, forcing hygienists to keep up with changes in monitoring technology and in the materials or agents commonly measured. While data collection remains at the core of the field, the use of direct-reading instruments and sensor technology has shifted IH data into the realm of Big Data systems. Big Data are commonly defined by parameters known as the “five Vs”—volume, velocity, variety, veracity, and value. Volume refers to the amount of data created, while velocity refers to the speed of data acquisition and processing. Variety refers to the types of data being generated, including both structured numerical data and increasingly unstructured narrative descriptions that provide context. Veracity refers to the reliability, precision, and quality of the data, while value is defined by whether the data can be generalized and used to generate or test hypotheses (see Figure 2).

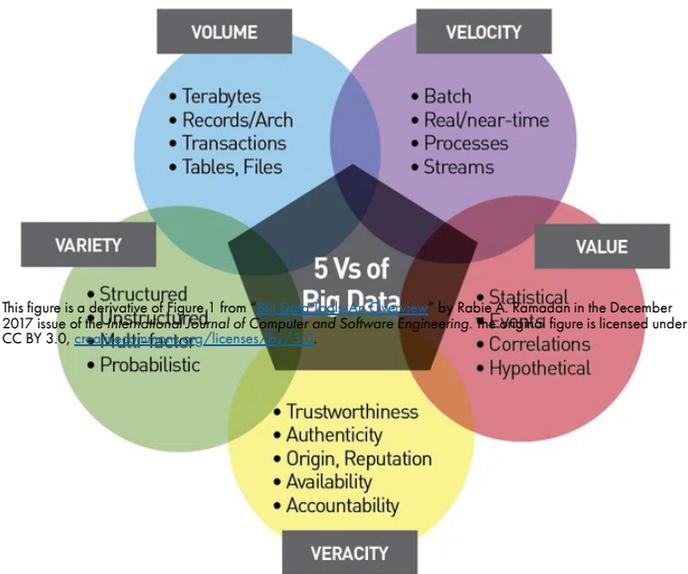
Figure 1. Focus and scope of recommended data elements for capture of standardized IH data.



From "Special Report: Data Elements for Occupational Exposure Databases: Guidelines and Recommendations for Airborne Hazards and Noise" by Morton Lippmann, Manuel R. Gomez, and Gregorie M. Rawls in the November 1996 issue of *Applied Occupational and Environmental Hygiene*. Reprinted by permission of the publisher, Taylor & Francis Ltd, tandfonline.com.

Figure 2. Description of the five dimensions or characteristics of Big Data (known as the “5 Vs”).

Tap on the graphic to open a larger version in your browser.



Occupational Safety and Health Act of 1970 (OSH Act) sparked the demand for improvement of IH data quality as it greatly increased the amount of IH data collected in the public and private sectors. The OSH Act magnified the scope of the entire IH field. It was, and still is, widely believed that these data should be used beyond solely determining compliance with set standards. To read more about this topic, see the special report “Data Elements for Occupational Exposure Databases: Guidelines and Recommendations for Airborne Hazards and Noise” in the November 1996 issue of *Applied Occupational and Environmental Hygiene*.

IH professionals have recommended that IH data be used in other exposure assessment activities related to the management of workplace risk, such as epidemiologic research, exposure surveillance, benchmark development, population-level and quantitative risk assessments, and the creation of retrospective and prospective exposure models. However, for IH data to be meaningful for such activities, hygienists must include additional information related to the conditions of the exposure measurement. Raw exposure measurements have limited use without the appropriate context in the form of sampling, employee, environmental, and process information (see Figure 1).

A 1993 international conference on occupational exposure databases led to the formation of a joint task group between ACGIH and AIHA. To harmonize IH data collection and improve data quality, the task group created a uniform list of 134 data elements with corresponding definitions, then organized the elements into 13 data groups (as described in the 1996 special report in the *Applied* journal archives). These data groups include information about the sampled employee’s job title and work history as well as the site, process, work area, chemical agent, exposure modifiers, sample, sampling device, engineering controls, and personal protective equipment. Chemical and physical hazard exposure results are also included (see Table 1). The data elements are flexible and can be adapted to different uses and levels of detail while maintaining the standardization necessary to allow for data aggregation and comparisons.

Although the task group’s list of IH data elements and variables remains relevant and useful, it has not been widely adopted. This lack of adoption is likely due to a variety of reasons, including technological or software-based constraints, insufficient organizational capability (for example, if adoption is cost prohibitive or a shortage of time or personnel prevents implementation), inaccessibility to the list of data elements, and the absence of a champion to support its use.

Table 1. Data Groups with Summary Description of Each Group's Contents of Associated Data Elements

	Data Group	Summary Contents
A.	Facility/Site Information	Basic descriptive information regarding the facility or site.
B.	Survey Tracking Information	Summary survey identification and tracking information.
C.	Work Area Information	Information about the area(s) where the exposure measurements/assessments were performed (location, type, ambient conditions).
D.	Employee Information	Information about the employee(s) whose exposure was assessed/measured, nature of job, exposure groups.
E.	Process and Operation Information	Information about the process(es), operation(s), exposure source(s).
F.	Chemical Agent Information	Information about the chemical agent(s) monitored.
G.	Exposure Modifier Information	Information about modifiers of exposure (representativeness, pattern, frequency, work load).
H.	Sample Information	Information about the sample(s) (reason, date, duration, type).
I.	Sampling Device Information	Information about the sampling device used.
J.	Engineering Controls Information	Information about engineering controls (type, condition, usage).
K.	Personal Protective Equipment Information	Information about respirators; gloves; protective clothing; hearing, eye, face, and foot protection.
L.	Chemical Exposure Results	Concentration measured, TWA estimates, limit of detection.
M.	Noise Exposure Results	Noise exposure estimates (frequencies, intensities, intermittences).

CURRENT STATUS

Advancements in technology have ushered in a wide range of data collection and storage options via mobile devices and database/analytic software. The volume, velocity, and variety of IH data being collected continues to increase, particularly with the development of direct-reading instrumentation and cloud-based analytics. However, IH data standardization that may ensure data veracity and support data value remains a challenge.

A survey of IH data practices among a sample of workers' compensation insurers published in the June 2018 issue of the *Journal of Occupational and Environmental Hygiene* (JOEH) revealed that the collection and storage of IH data varied, even though the majority of insurers provided IH services and employed IH staff. The insurers mostly used the data to provide recommendations to their policyholders. Their data storage capabilities were reflective of this, as most insurers stored the data in disparate, non-standardized file types (for example, PDFs and Word documents) that could not be easily grouped or searched collectively to identify patterns, or they did not retain any IH data after providing a report to their clients. The data elements collected by these insurers also differed greatly. While these insurers represent a small subset of organizations that conduct IH tasks, this survey highlights the elusive nature of IH data standardization.

Another study collected the IH data forms used by a similar sample of workers' compensation insurers and government agencies and used the forms to derive a core list of essential data fields to determine the feasibility of standardization and pooling of IH data (see "Standardizing Industrial Hygiene Data Collection Forms Used by Workers' Compensation Insurers" in the September 2018 JOEH). The authors compared the core list to the list of data elements developed by the 1993 ACGIH-AIHA task group and found that the two lists had 85 to 90 percent of the data elements in common. Results from the 2018 study indicate that the recommendations of the 1993 task group are still relevant to standardization in today's IH field, but that progress toward adoption remains slow.

AIHA has increased efforts to promote data quality and standardization by including an appendix on data quality (Appendix IX) in the latest edition of *A Strategy for Assessing and Managing Occupational Exposures*, which was published in 2015. This appendix introduces literature on how to evaluate historical or current IH data quality and includes a checklist of data elements adapted from the literature, including the 1996 special report in the *Applied* journal. The checklist serves as a source for comparison, providing a method for evaluating current IH forms to examine which data fields are being captured and how.

A consensus study report published in 2018 by the National Academies of Sciences, Engineering, and Medicine, "[A Smarter National Surveillance System for Occupational Safety and Health in the 21st Century](#)," represents additional efforts to encourage data standardization. A major recommendation from the report is for NIOSH and OSHA to prioritize development of a comprehensive approach for occupational exposure surveillance. The report refers to the collection and analysis of data on occupational hazards and exposures as a "key component missing from U.S. occupational safety and health surveillance" and identifies several short- and long-term recommendations for achieving a cohesive and "smarter" surveillance system to protect the current and future workforce. IH data standardization is needed for overall improvement of occupational safety and health surveillance.

The literature, IH professional organizations, and those employed within the IH field agree on the importance of data quality and completeness. Current and historical exposure monitoring data sources, such as OSHA's Integrated Management Information System (IMIS) and workers' compensation data, have been valuable in providing exposure trends over time, by agent or job type, and by severity of exposure. However, many of these data sources are lacking contextual information that could allow them to serve even greater

purposes. For example, the OSHA IMIS records do not include information on the sampling method, sampling instrument, or site or work area.

Those few examples that have successfully cleaned and aggregated IH data illustrate the value of IH data on a larger scale. The [Exposure Control Database](#) (ECD), created by CPWR – The Center for Construction Research and Training, is an interactive tool that predicts exposure to silica, welding fumes, noise, and lead within the construction industry. The source data are continually updated from reliable sources, including government reports and peer-reviewed literature. The ECD showcases the predictive capabilities of standardized and aggregated occupational exposure monitoring data for the implementation of effective hazard controls.

A project similar to the ECD, the [Canadian Workplace Exposure Database](#), is a large, centralized database for current and historical occupational exposure data from workplaces across Canada. It currently contains approximately 480,000 air-sampling measurements for chemical carcinogens. A project team is now engaged in expanding this database to include non-carcinogenic exposures, and to ensure the data is properly collected and coded to inform research and policy projects. The team also plans to enhance the governance structure, data stewardship and management, and user access. Several groups—including the British Columbia Construction Safety Alliance, the British Columbia occupational health and safety workers' compensation insurer WorkSafeBC, and exposure scientists from the University of British Columbia—used this database as a source of “objective data” on respirable crystalline silica (RCS). The RCS data in the Canadian construction sector was used to develop a web-based tool that allows end users with varying expertise, including non-OHS experts, to generate a single-task exposure control plan (ECP), which is required under regulation in British Columbia for those undertaking work potentially exposing workers to RCS. (The development and implementation of the web-based tool is described in an [article](#) published in *Frontiers in Public Health*.) Users generate up to 3,900 ECPs per year with the tool.

The final example is provided by the Washington State Department of Labor and Industries, which was able to aggregate nine years (2008–2016) of worker exposure assessment results collected by the state's IH compliance safety and health officers into a summary report ([PDF](#)) and corresponding database. The report allowed the state of Washington to identify the most frequently sampled substances and industries and determine the substances with the highest severity exposure (which represents the ratio of the measured concentration to its permissible exposure limit).

LOOKING AHEAD

IH data standardization will be critically important for the development of risk assessments, epidemiologic research, and exposure modeling of emerging hazards to understand potential workplace health effects associated with exposure. This is especially true for emerging chemicals that do not yet have established occupational exposure limits (OELs) or standard sampling methods.

One example of the need for IH data standardization is from the continually developing field of nanotechnology and advanced materials (NAMs). Innovations in manufacturing and nanotechnology have led to an entirely new slate of agents known as “advanced materials,” and NAMs have already had an impact within the IH field. Many rapid, significant changes with which the field of industrial hygiene is currently contending have made data standardization increasingly important to ensure data veracity and value to help in the prevention of hazardous occupational exposures.

Currently, few NAMs have an OEL or standard sampling method. These materials exemplify some of the challenges for data standardization due to:

- varied morphologies of the aerosol constituents (such as shape and relative number of primary particles, agglomerates, and functionalization elements)
- types of exposure metrics and methods collected for a given exposure assessment resulting from limited standardized sampling methods (for example, integrated filter samples for chemical analysis or electron microscopic image analysis, X-ray energy dispersive spectra, real-time particle number concentrations, surface charge, and surface area)
- application to exposure or control banding schemes for interpretation

Collection and storage of standardized IH data in the present will allow future researchers to pool data for health-related studies, which could include retrospective cohorts and exposure modeling. The lack of consensus on sampling methods can be overcome by providing detailed information within these standardized variables and data groups that will be critical for future studies.

IH data standardization is a massive undertaking, not easily accomplished, but those in the IH field can take small steps to promote data quality. AIHA recently circulated two surveys to gather relevant feedback from the IH professional community: one focused on current IH data applications and standardization among organizations, and the second focused on the list of data elements specific to standardization. At the time this issue of *The Synergist* went to press, results of these surveys had not yet been released. Ultimately, communication across the IH field is needed to ensure a better understanding of common practices related to data collection, storage, and use. Sharing best practices and identifying data gaps will help the finalization and distribution of a list of standardized data elements.

The future of IH data is both promising and daunting with continuous technological and analytical advancement (such as real-time monitoring and cloud-based computing), the expansion of Big Data, and the increase in emerging chemicals. IH data standardization can ease the burden of these changes and help guide the field in protecting the health and safety of the evolving workforce.

TAYLOR M. SHOCKEY, PHD, MPH, is an associate service fellow at CDC/NIOSH in Cincinnati, Ohio.

MATTHEW M. DAHM, PHD, MPH, is a research industrial hygienist at CDC/NIOSH in Cincinnati, Ohio.

STEVEN J. WURZELBACHER, PHD, directs the Center for Workers' Compensation Studies at CDC/NIOSH in Cincinnati, Ohio.

JOHN BAKER, MS, MBA, CIH, FAIHA, is a principal consultant and south regional practice lead (IH) for BSI Consulting Services, Inc. in The Woodlands, Texas.

Disclaimer: The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention.

Send feedback to [*The Synergist*](#).