







국가 화학물질 유해성정보 데이터베이스 구축 과정의 신뢰도 제고 방안에 관한 연구

이소민  · 이민혁  · 강미진*  · 권순광**  · 나진성†  · 박백수† 

한국생산기술연구원, *환경부 화학물질정책과, **한국화학물질관리협회

Improving the Reliability of the National Database for Chemical Hazard Information

Somin Lee, Minhyeok Lee, Mijin Kang*, Soon-Kwang Kwon**,
Jin-Sung Ra†, and Beaksoo Park†

Korea Institute of Industrial Technology,

**Ministry of Environment*

***Korea Chemicals Management Association*

ABSTRACT

Objectives: According to the Act on Registration, Evaluation, Etc. of Chemicals, new and existing chemicals must be registered by 2030. In addition, industries need to submit hazard data as an attachment during the registration process. Therefore, we constructed a nationwide chemical database to support small industry by providing hazard data and original sources. During the process, we developed a new standard procedure for minimizing errors and increasing reliability.

Methods: We analyzed the categories of errors and the cause of the errors through the verification results of the 2019 project. We present an improved database construction methodology and system.

Results: Errors are categorized according to their causative factors into simple, technical, and structural type errors. Simple errors arise simply because of decreased concentration or negligence in following the instructions. Technical errors are caused by a discrepancy between the professional field and the type of data. Structural errors indicate systemic errors such as incomplete forms on the excel database or ambiguity in the guidelines. Lessons from the errors collected in the 2019 project are used to update the procedures for database authorization and technical guidelines. The main update points are as follows; 'supplementation of review process', 'giving regular training to external reviewers', 'giving additional information to authors, like physico-chemical properties of substances, degradability, etc.', 'amendment of excel form', and 'guideline upgrades'.

Conclusions: We conducted this study with the aim of improving the accuracy and reliability of the database of hazard information for chemical substances. The new procedures and guidelines are now being used in the 2020 project for construction of a hazard information database for Korea.

Key words: Hazard information database, Korea REACH, human health, environmental health, national database

†Corresponding authors: Jin-Sung Ra: Eco-testing & Risk Assessment Center, Korea Institute of Industrial Technology, 143 Hanggaulro, Sangnok-gu, Ansan-si, 15588, Republic of Korea, Tel: +82-31-8040-6077, Fax: +82-31-8040-6210, E-mail: jinsungra@gmail.com

Beaksoo Park: Korea National Cleaner Production Center, Korea Institute of Industrial Technology, Hanshin Intervalley 24 East B/D 18F, 06211, Republic of Korea, Tel: +82-2-2183-1510, Fax: +82-2-2183-1519, beaksoo@kncpc.re.kr

Received: 10 August 2020, Revised: 19 August 2020, Accepted: 19 August 2020

I. 서 론

유럽연합(EU)은 2007년 Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) 제도를 시행하여, 화학물질을 시장에 출시하고자 하는 기업은 용도별 등록을 통해 해당 화학물질이 위해하지 않다는 것을 증명하도록 하고 있다.¹⁾ 즉, 그 동안 국가가 지고 있던 화학물질 위해성에 대한 책임을 기업에 나누어주고 책임 있는 시장진입을 유도한 제도이다. 우리나라도 가습기살균제 사고(2011년) 등에서 야기된 화학물질에 대한 국민의 불신과 불안을 해소하기 위하여 「화학물질의 등록 및 평가 등에 관한 법률(이하 화학물질등록평가법)」을 시행하였다.²⁾

화학물질등록평가법에 따라 기존화학물질을 연간 1톤 이상 제조·수입하려는 자(신규화학물질은 0.1톤)는 법 제10조에 따른 등록을 이행해야 하며, 이때 같은 법 시행규칙 제5조에 따른 물리적·화학적 특성 및 유해성에 관한 시험자료를 제출해야 한다.^{2,4)} 특히, 우리나라는 2018년 기준 중소기업이 630만 개로 전체기업의 99.9%를 차지하고 있어,⁵⁾ 화학물질 등록 의무대상자의 대다수가 중소기업에 해당한다. 이에 따라 환경부는 중소기업의 화학물질등록평가법 이행 지원을 목적으로 2017년부터 기존화학물질의 유해성정보 조사·확인사업을 수행하고 있다. 이를 통해 구축한 기존화학물질 유해성정보는 2020년 1월부터 화학안전산업계지원단 홈페이지를 통해 단계적으로 공개하고 있으며,⁶⁾ 화학물질 유해성정보 구축에는 신뢰도 제고를 위한 단계적 검증방안을 도입하였다.

국외에서도 화학물질 유해성정보를 평가보고서 또는 웹 사이트 형태로 구축 및 운영 중이다. 예로, 미국의 NTP (National Toxicology Program),⁷⁾ 일본의 NITE (NITE Chemical Risk Information Platform),⁸⁾ 호주의 NICNAS (The National Industrial Chemicals Notification and Assessment Scheme),⁹⁾ EU의 ECHA (European Chemicals Agency)¹⁰⁾ 등이 국가단위로 화학물질 유해성정보를 제공하고 있다.

유해성정보 데이터베이스 구축 대상은 화학물질 등록평가법 제10조 제3항에 따라 신고한 만 여종 이상의 화학물질을 대상으로 하고 있다. 또한 데이

터베이스로 구축하려는 유해성정보 항목은 각 화학물질별로 물리화학적 특성 13개 항목, 인체 유해성 15개 항목, 환경 유해성 19개 항목이다. 47개의 유해성정보 항목 이외에도 해당 물질의 분류 및 표시 정보, 화학식, 구조식, UN 번호, 발암등급, 유럽화학물질청(ECHA) 등록 여부, 면제 확인사항 정보 등 매우 방대한 정보를 포함하도록 하고 있다. 대상 정보는 화학물질 등록 시 유해성 시험자료 확인 및 활용을 위한 필요정보이지만, 기업이 직접 조사·분석하는 것은 전문성의 한계가 존재한다. 이에 정부가 신뢰성 있는 화학물질 유해성정보를 제공함으로써, 국내 기업의 등록 이행을 지원하고자 하였다. 유해성정보 데이터베이스의 신뢰성을 확보하기 위해, 데이터베이스 구축 프로세스 표준화는 필요한 과정이다. 구축 프로세스의 표준화에는 국제적으로 통용되는 정보소스의 정립과 수집된 정보의 완전성과 신뢰성 검증, 대외적인 평가까지 포함된다.^{11,12)} 고 하였다. 유해성정보 데이터베이스 구축은 화학물질등록평가법에 따른 등록이나 유해성 평가 경험을 보유하고 있거나, 독성 시험, 자료 분석 등 관련 전문가가 진행한다. 그럼에도 불구하고 통일된 용어사용의 한계나, 전문 분야의 차이 등 작성자에서 기인할 수 있는 오류의 최소화도 데이터베이스의 신뢰성 확보에 매우 중요한 인자이다. 또한, 데이터베이스 구축 시에는 대상의 종합적인 정보 파악과 수집을 위해 조사양식을 통일화하여 정보 공유의 효율성을 제고해야 한다고 하였다.¹³⁾

따라서 본 연구에서는 국가 수준의 대규모 화학물질 데이터베이스 구축 시 신뢰성을 향상시킬 수 있는 방안을 제시하기 위해, 다음과 같은 연구목표를 설정하였다. 첫 번째, 화학물질 유해성정보의 정확도 향상을 위해 개발된 데이터베이스 구축 표준방법론을 제안하고자 하였다. 두 번째, 개발된 방법론의 실증연구를 통해 대규모 데이터베이스 구축 시 발생 가능한 오류의 유형화 및 원인분석을 실시하였다. 세 번째, 오류 유형 및 원인을 최소화할 수 있는 방법론을 모색하여, 화학물질 유해성정보 데이터베이스 구축 방법의 고도화 방안을 제시하고자 하였다.

II. 연구 방법

데이터베이스 구축 신뢰성 향상 방안은 환경부에서 2019년부터 2020년까지 진행 중인 “기존화학물질 유해성정보 확인사업(이하 유해정보 확인사업)”을 대상으로 하였다. 유해정보 확인사업은 구축 자료의 정확도 향상을 위해 데이터베이스 구축 표준방법을 개발·운영하였다.¹⁴⁾ 그럼에도 불구하고 작성 결과물의 검증에서 원문 확인, 신뢰도,

입력 값, 표기방식 등 다양한 오류가 확인되었다. 확인된 오류는 오류 유형과 발생원인 도출을 통해 기존의 표준방법론을 보완하는데 활용되었으며, 최종적으로 신뢰성 있는 화학물질 유해성정보 데이터베이스 구축 방법론을 도출하고자 하였다. 보완된 방법론은 환경부(주관부처), 국립환경과학원(화학물질등록평가법 등록 기관), 한국환경공단(한국화학물질관리협회(사업 관리기관), 한국생산기술연구원(수행기관) 및 외부 전문가의 의견수렴을 통해

Table 1. Hazard Items Gathered at the Database

Category	Items	Category	Items
General information (8)	Chemical name	Physico-chemical hazard (13)	Physical state
	Substance ID		Water solubility
	Chemical/Constitutional formula, Substance description		Melting/freezing point
	Classification information		Boiling point
	Carcinogenesis grade		Vapor pressure
	Registration dossier found at ECHA		Octanol/water partitioning coefficient
	Domestic registration		Density
	Unclassified hazard, Polymer compound, UVCB		Particle size analysis
Environmental Hazard (19)	Acute Fish toxicity		Flammability
	Chronic Fish toxicity		Explosiveness
	Acute Daphnia toxicity		Oxidation
	Chronic Daphnia toxicity		Viscosity
	Degradability		Dissociation constant
	Fresh water algae growth inhibition test	Acute oral toxicity	
	Biodegradation as a function of pH	Human Hazard (15)	Acute dermal toxicity/Acute inhalation toxicity
	Inherent biodegradability		Ames test
	Identification of degradation products		Skin irritation/corrosion
	Acute Terrestrial plants toxicity		Skin sensitization
	Chronic Terrestrial plants toxicity		Eye irritation/corrosion
	Acute Terrestrial invertebrate toxicity		Chromosome abnormality test
	Chronic Terrestrial invertebrate toxicity		Genotoxicity test
	Activated sludge respiration inhibition		Repeated dose toxicity study (28 d)
	Adsorption/desorption		Repeated dose toxicity study (90 d)
	Additional information on environmental behavior and dynamics		Screening for reproductive/developmental toxicity
	Additional information on adsorption/desorption	Additional genetic toxicity	
	Chronic toxicity of benthic organisms	Bioaccumulation	Teratogenicity
			Two generation reproduction toxicity
Carcinogenicity			
		Data analysis	Exemptions from K-REACH
		Classification and Labelling	Hazard classification

최종 확정하였다.

데이터베이스 구축대상 2천여 종의 기존화학물질은 2024년 12월 31일까지 등록할 수 있도록 유효기간이 부여된 물질, 중소기업 지원사업 대상물질, 유해성이 낮은 것으로 신고한 물질, 유해성정보 확보가 시급한 것으로 환경부에서 판단한 물질을 대상으로 작성하였다. 데이터베이스는 화학물질 등록평가법 제14조 제1항 제5호와 제6호에 따라 제출이 요구되는 유해성정보를 대상으로 하였다 (Table 1). 데이터베이스는 'Microsoft Excel (이하 엑셀)' 프로그램을 사용하여 작성양식(Template)을 제공하였으며, 작성자 운영 프로그램의 차이점을 고려하여 EXCEL 2010 버전 이상에서 상호 호환이 가능하도록 구성하였다. 데이터베이스는 항목별

로 구분하여 상이한 엑셀 시트에 구축할 수 있도록 하였다. 각 항목별 시트는 하위 항목으로 구분하여, 시험항목별 참고문헌 정보, 유해성 시험정보 (종말점, 부등호, 값/결과, 단위, 주요영향, 시험중, 계통/서브계통, 노출시간, 노출빈도, 기간, 시험용량, 노출경로, 시험물질, 결과 도출방법, 시험지침, 기타 정보), 참고 데이터베이스 확인여부, 무료자료 확인여부 등을 작성할 수 있도록 구성하였다. 데이터베이스에 구축하는 항목은 각 시트별로 참고 데이터베이스를 제시하고 하이퍼링크를 연결하여 물질 검색 시간과 정확성을 높일 수 있도록 구성하였다. 각 시트별로 제시한 참고 데이터베이스 이외의 자료는 구축하지 않도록 규정하였다.

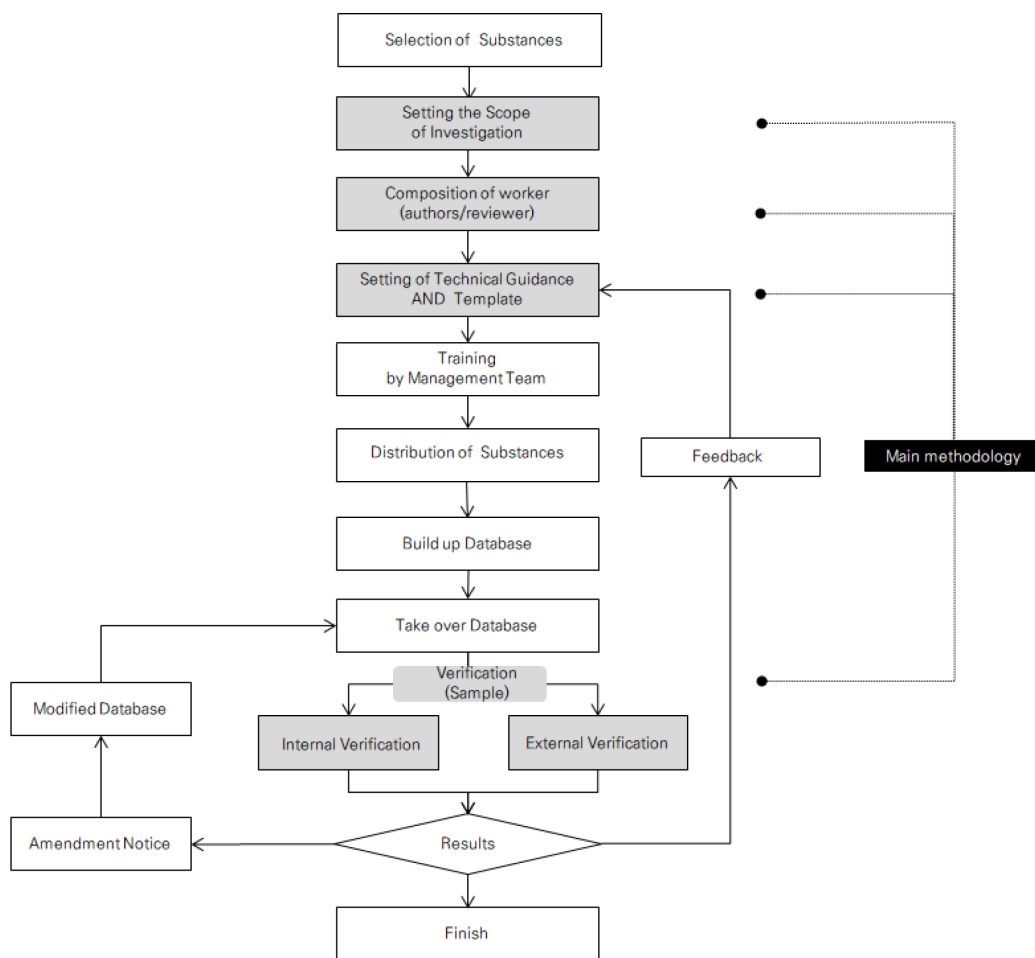


Fig. 1. Standard Procedure for Database Construction on Chemical Hazards (2019)

III. 연구 결과

1. 데이터베이스 구축 표준방법 개발

유해성정보 데이터베이스 정확도 향상을 위해 작성자의 구성, 작업분량 및 기간, 데이터베이스의 유형을 고려하여 표준 데이터베이스 구축 방법론을 개발·운영하였다. Fig. 1에서 확인할 수 있는 바와 같이 우선순위 대상물질을 선정한 후, 각 항목별로 확인 및 구축해야 하는 정보를 한정하였다. 정보의 유형을 바탕으로 관련 작성자 및 전문가를 구성하였으며, 작성자의 수준을 고려한 작성 지침서를 개발하는 절차로 구성하였다. 물질 배포 후 작성한 데이터베이스는 내부 및 외부 검증의 이원화된 검

증체계를 적용하였으며, 해당 검증과정에서 확인된 오류는 작성자에 수정/보완을 요청하는 방식으로 절차를 개발하였다. 각 단계별 세부적인 내용은 다음과 같다.

1.1. 조사범위 설정

유해성 자료의 조사범위는 국제적으로 신뢰성이 높다고 인정되는 과학 핸드북, 국가·국제기구 보고서 및 국가에서 운영하는 웹 사이트를 참고 데이터베이스로 선정하였다(Table 2). 각 참고 데이터베이스는 해당 데이터베이스의 특성에 따라 제공하는 정보가 제한적인 경우가 많다. 따라서 참고 데이터베이스의 정보 유형을 확인하고, 확인 가능

Table 2. List and priority of Reference Database

Visiting order	Group	Reference Database	Physico-chemical hazard	Human Hazard	Environmental Hazard
1	Registration status	Registration dossier found at ECHA	○	○	○
2		Merck Index- chemicals, drugs and biologicals	○	-	-
3		CRC Handbook of Chemistry and Physics	○	-	-
4		EU Risk Assessment Report (RAR)	○	○	○
5		ECHA Biocidal Assessment Report	○	○	○
6		IARC Monographs on the Identification of Carcinogenic Hazards to Humans (IARC Monograph)	○	○	-
7	Group 1	US National Toxicology Program (NTP)	○	○	-
8		US Registration Eligibility Decision	○	○	○
9		NITE Chemical Risk Information Platform	○	○	○
10		The National Industrial Chemicals Notification and Assessment Scheme (NICNAS)	○	○	○
11		Agency for Toxic Substances and Disease Registry (ATSDR)	○	○	○
12		OECD Screening Information DataSet (SIDS)	○	○	○
13		Integrated Risk Information System (IRIS)	○	○	-
14		International Programme on Chemical Safety Environmental Health Criteria (IPCS EHC)	○	○	○
15		American Conference of Governmental Industrial Hygienist (ACGIH)	-	○	-
16	Group 2	Japan CHEmicals Collaborative Knowledge database (J-CHECK)	-	-	○
17		ChemIDplus Advanced	○	○	-
18		Central Credit Reference Information System (CCRIS)	-	○	-
19		Genetic Toxicology Data Bank (GENE-TOX)	-	○	-
20		The Ecotoxicology Database (ECOTOX)	-	-	○

한 유해성정보의 매칭을 통해 조사 범위를 한정하였다. REACH 제도에 따라 기 등록된 화학물질의 정보 확인이 가능한 유럽화학물질청(ECHA) 등록자료 데이터베이스는 화학물질등록평가법에 따른 등록 이행자의 참가가 용이하게 구축되어 있어, 단순히 물질의 등록여부만을 제시함으로써 필요시 ECHA 등록자료 데이터베이스를 활용하도록 하고 데이터베이스 구축에서는 제외하였다. 데이터베이스 구축 대상 물질의 수와 항목을 고려하여 사업기간 내 모든 가용자료의 확인이 불가능하기 때문에, 자료가 충분한 경우에는 15개 이하의 자료만 확인하여 구축하는 것으로 한정하였다. 따라서 활용성 측면에서 데이터베이스의 검색 우선순위를 정하였으며, Table 2에 나타내었다. 검색은 방문순위 1번을 우선으로 하고 해당 참고 데이터베이스에 충분한 자료가 없는 경우 다음순위 검색을 원칙으로 하였다.

각 참고 데이터베이스 목록 중에서 대상 항목에 해당하는 정보를 확인한 경우에는 ‘○’, 정보가 확인되지 않는 경우에는 ‘×’ 표기를 하도록 지침을 정하였으며, 확인 과정에서 항목별로 15개를 초과하여 확인되는 경우 15개까지만 구축하도록 하였다. 확인하지 않은 나머지 참고 데이터베이스는 ‘미확인’으로 표기함으로써, 사용자가 나머지 참고 데이터베이스도 필요한 경우 살펴볼 필요가 있음을 인지하도록 하였다.

1.2. 데이터베이스 작성 및 검증자 기준

데이터베이스 작성자는 항목별 유해성정보의 확인 및 분석이 가능하고, 화학물질의 물리·화학적 특성 및 유해성에 대한 전문성을 보유한 관련 전공의 박사급 인력으로 구성하였다. 검증자는 화학/보건/환경분야 경력 10년 이상을 갖춘 전문가로 구성하였으며, 검증자의 전문 분야에 한하여 작성 데이터베이스를 검증하였다. 작성과 검증 후 해당 엑셀시트에 작성자 및 검증자의 이름과 작성일을 명시하는 「데이터베이스 실명제」로 업무의 책임감과 효율성을 제고하였으며, 오류 발견 시에는 작성자의 추적이 용이하도록 하였다. 이외 행정 및 IT 지원업무를 수행하는 약 20명의 인력은 별도로 구성하였다.

1.3. 데이터베이스 작성 기준 정립

데이터베이스 작성 지침서는 ‘화학물질 정보 데이터베이스 구축 일반기준’, ‘화학물질 정보 데이터베이스 구축 방법’, ‘화학물질 유해성정보 확인 결과 데이터베이스 구축 방법’ 3개의 단원으로 구성하였다(Table 3). 작성 지침서는 일반적인 데이터베이스 작성 절차를 고려하여, 참고 데이터베이스 검색, 독성 자료 확인, 확인된 독성 자료 수집, 데이터베이스 작성양식 등 순으로 세부내용을 구성하였다. 또한, 항목별 세부 작성기준 및 예시(글꼴, 단위, 소수점 표기, 위첨자 표기, 용어 정리 등),

Table 3. Main contents in the technical guidance

Category	Contents
General criteria for the database	Details on the general criteria
	List and priority of the reference database
	How to search in the reference database
	Reliability check
Methods for drawing up a database for each items	Draw up method of hazard information (x axis)
	Notice to the general information sheet
	Notice to physicochemical item
	Notice to Human hazard item
Methods for drawing up a database that reflects chemical hazard classification	Notice to environmental hazard item
	General criteria
	Summary table of physicochemical and degradability results
	Draw up method of hazard classification
	Draw up method of data submission exemption

시험항목의 정의, 참고 데이터베이스의 방문 및 검색 방법, 원하는 정보의 선정 등을 명확히 제시하여, 작성자 판단과 의견에 따른 정보 가공을 최소화할 수 있도록 하였다(예: 영어는 번역하지 않음, 단위 환산 없음 등). 데이터베이스 작성 전 미리 배포하여, 작성자들이 반드시 숙지할 수 있도록 하였다.

작성양식의 경우, 1종의 화학물질은 엑셀파일 1개 생성을 원칙으로 유해성정보 분야별로 6개의 시트로 구성하였다. 엑셀의 자동화 프로그램 기능인 매크로(Macro), 하이퍼링크(hyperlink), 드롭다운 메뉴(drop-down menu) 등을 활용하여, 작업의 효율성 및 정확성을 극대화할 수 있는 작성양식을 개발하였다. 예로, 매크로 기능을 사용하여, 일반정보 시트에서는 구축대상 화학물질의 CAS NO.를 검색하면 해당 화학물질의 명칭과 CAS NO., 기존화학물질 고유번호 등 기본정보가 자동 입력될 수 있도록 하였다. 하이퍼링크 기능을 활용하여, 물리·화학적 특성 및 인체·환경 유해성정보 시트에서는 참고 데이터베이스 사이트의 검색창을 확인할 수 있는 URL을 작업파일에 적용하였다. 드롭다운 메뉴 기능으로 작성정보 중 출처, 작성일자, 신뢰도, 시험결과 도출방법 및 참고 데이터베이스 확인여부 등 작성 값이 목록화 가능한 항목에 적용하였다. 이에 작성자가 해당 정보를 선택하는 형태로 작성하였다.

1.4. 데이터베이스 검증체계 이원화

검증은 1차 작성된 결과물을 사업관리팀에서 인수하고, 60여명의 작성자가 데이터베이스 작성과 동시에 교차검증을 위해 재배포하여, 즉각적인 피드백이 이루어지도록 하였다. 1단계 검증대상은 작성자별 담당물질의 20%로 정하고, 검증체계에서 오류사항이 확인된 작성자의 경우에는 나머지 80% 물질까지 작성자가 직접 확인 및 수정하여 신뢰성을 제고할 수 있는 검증체계를 적용하였다. 2단계 검증에서는 2천여 종 물질 중 무작위 30종 추출하여, 외부 검증자에 의해 표본검증을 실시하였다. 오류 발견 시 해당 작성자가 작성한 모든 결과물에 대해 작성자가 직접 수정·보완하여, 작성자가 직접 작성 결과물에 책임지는 검증체제로 작성 시 정확성을 높이도록 운영하였다.

2. 오류 유형화 및 원인분석

개발된 데이터베이스 구축 표준방법의 실증연구를 수행하였으며, 검증단계에서 다양한 정보의 오류를 확인하였다. 이렇게 확인된 오류를 유형화 및 원인분석을 실시하였으며, 각 단계의 세부 내용은 아래와 같다.

2.1. 데이터베이스 작성 오류

작성대상 화학물질 2천여 종의 검증과정에서 빈도가 높은 오류를 중심으로 아래와 같이 정리하였

Table 4. Error category and their examples

Examples of Error	Type of Error
<ul style="list-style-type: none"> · Wrong information on the structural formula of substance. · Inconsistency in hazard data through the data file (i.e. across the sheet) · Wrong reference information 	▶ Simple Error
<ul style="list-style-type: none"> · Wrong data given at the data submission exemption (missing in deleting the sentence in a cell matched to the particle size for a liquid substance) · Wrong data or data are not matched to the original paper · Full data set are not given (i.e. species are missing) 	▶ Technical Error
<ul style="list-style-type: none"> · Link or match error between CAS NO. and chemical name in system · No guideline is given to the free of charge data · No cell is allocated to the reviewer for their opinion · Missing of sentence that can explain the data submission exemption, i.e., to inorganic substance 	▶ Structural Error

Type of error	Cause of Error	Supplementation of system
Simple Error	<ul style="list-style-type: none"> · Mistake arise from excessive workload of a worker · Failure in following a guidance 	<ul style="list-style-type: none"> · Additional reviewer process
Technical Error	<ul style="list-style-type: none"> · discrepancy in expertise 	<ul style="list-style-type: none"> · Regular training to external experts · Sharing of additional information on substance
Structural Error	<ul style="list-style-type: none"> · Incomplete form of excel sheet · Ambiguity in guideline 	<ul style="list-style-type: none"> · Form modification · Guideline update

Fig. 2. Revision of Methodology by Categories and Cause of Errors

다. 오류내용은 작성대상 유해성정보 시트별로 구분하였으며, 확인된 오류사항은 여러 차례의 검증 단계를 통해 수정·보완하였다.

일반정보 시트에서는 CAS NO. 기재 시 작성 데이터베이스 양식에 의해 자동 작성되는 화학물질 명의 오류 그리고 구조식 표기 부적절 또는 누락하는 형태의 오류를 확인하였다. 예로, 고분자의 경우 단량체 표기는 적절치 않으나, 단량체 단위의 구조식으로 작성되었다.

데이터 확인분석 시트에서는 ‘면제 확인사항’ 항목에서 작성자 전문성 부족 및 데이터베이스 양식 오류로 해당되지 않는 면제조건이 기재되었거나, 해당 면제조건 누락, 면제사유 표기 부적절 등의 오류사항을 발견하였다. 예로, 옥탄올-물 분배계수 값(Log kow)이 6임에도 불구하고, ‘생물농축성’ 항목에 (면제조건)생물농축 가능성이 낮은 물질(예, Log kow <3)이라는 정보가 기입되어 있었다. ‘물리화학적 성질’ 항목에서는 물리화학적 특성자료 시트의 값과 다른 유해성 시험값 입력, 정보 표기방법 부적절, 폭발성 관련 구조 미확인 등을 확인하였고, 무료자료 유무항목에서는 잘못된 정보를 기입하는 등의 오류를 확인하였다.

물리화학적 특성 자료 시트에서는 ‘출처’ 항목에서 원문 내 정보의 기입이 잘못되거나 인용을 잘못하거나, 출처명 및 URL 기재 오류 등 출처의 작성지침 미준수에 해당하는 오류를 확인하였다. 해당 내용은 물리화학적 특성 자료시트 뿐만 아니라, 인체유해성 자료 시트, 환경유해성 자료 시트에서 공통되는 사항으로 확인하였다. 이외 ‘단위’ 항목에서는 상대밀도, 옥탄올-물 분배계수 등 특성자료 단위에 대한 작성오류, ‘시험조건’ 항목에서는 끓는점, 물 용해도, 증기압 등 시험조건을 작성해야 하는 항목 누락, 옥탄올-물 분배계수의 시험조건

미기재, ‘기타정보’에는 출처가 Merck Index, CRC handbook일 경우 ‘등록정보 활용가능’을 기입해야 하나, 미기입하는 형태의 오류를 확인하였다. 전반적으로 서로 다른 시험항목에 동일 내용을 중복 기입하거나, 복수의 참고 데이터베이스 자료가 확인됨에도 불구하고 단일 정보만 기입하는 형태의 오류를 확인하였다.

인체유해성 자료 시트에서는 ‘종말점’ 항목에서 시험자료 종말점 표기 오류, ‘부등호’ 항목에서는 독성자료 부등호 표기 오류, ‘값/결과’ 항목에서는 표기 누락, ‘시험종’과 ‘시험물질’ 항목에서는 복수의 시험 종이 있으나 한 종만 기입된 오류를 확인하였다.

환경유해성 자료 시트에서는 ‘부등호’의 표기 오류 및 ‘계통’ 항목의 경우, 시험종 정보를 오기입, ‘결과도출방법’ 항목에서는 K_{OC}값, 생물농축계수의 경우 계산값이 대부분이나 시험값으로 잘못된 형태를 확인하였고, 시험항목 중 ‘육생식물 독성’에서는 해당 독성에 해당하는 생물종 그룹이 혼재되어 있는 등의 오류를 확인하였다.

2.2. 데이터베이스 오류의 유형화

데이터베이스 작성 결과물의 오류 유형을 분석한 결과, 단순오류, 기술오류, 구조적 오류 3가지 유형으로 구분할 수 있다(Table 4). 첫 번째, 단순오류 유형은 물질명 또는 구조식 오기입, 물리화학적 성질 표기 부적절, 출처 기입 오류 등 작업자의 단순 실수로 인한 오타, 미확인 등이 해당된다. 예로, 작성한 CAS NO.와 다른 화학물질의 명칭 기입 또는 물질명 오타, 구조식이 존재함에도 불구하고 미기입하는 등 표기의 부적절 또는 누락, 데이터 확인분석 시트와 물리화학적 특성자료 시트의 특성값 불일치, 유해성 자료의 원문이 아닌 참고 데이

터베이스명을 출처에 기입하거나, 출처명/ URL 기재의 지침을 미준수한 경우이다. 두 번째, 기술오류 유형은 작성자의 능력에 따라 잘못된 값의 기입, 신뢰도 평가, 작성 방법, 면제 확인사항 등의 오류에 해당된다. 예로, 화학물질의 상태가 액체인 경우 ‘입도분석’ 시험항목이 면제되에도 불구하고 면제조건 미기입, 폭발성 화학구조를 보유한 물질임에도 불구하고 유해성 자료가 없을 경우 폭발성이 없는 물질로 간주하는 등 폭발성 관련 구조 미확인, 복수의 시험종이 있으나 한종만 기입, 독성 결과에 기반한 면제조건 확인 미흡으로 육생식물 독성, 만성독성, pH에 따른 가수분해, 이분해성 등 특정 시험항목 오기입 또는 누락을 확인하였다. 세 번째, 구조적 오류유형에는 데이터베이스의 작성양식에 의한 오류, 지침 모호에 따른 무료자료 유무 표기 등 미흡한 관리 기준에 의한 것으로 판단하였다. 작성양식 오류의 예로, CAS NO. 기재 시 자동으로 불러오는 화학물질명에서 띄어쓰기, ‘x’, ‘-’, ‘;’의 누락, 작성양식에 기본정보로 제공되는 면제조건 항목 중 피부 부식성으로 분류되는 물질과 무기물 등 일부 면제조건 누락, 확인분석 시트에 별도의 검증의견 작성양식 부재 등이 있다. 또한, 지침 모호의 예로는 Merck Index, CRC handbook 등 참고 데이터베이스의 무료자료 해당여부를 판단할 수 있는 명확한 지침이 부재한 경우이다.

2.3. 오류 유형별 발생 원인분석

3가지 유형별 오류에 대한 원인을 분석하였다. 첫 번째, 단순오류 유형의 경우에는 물질명 또는 구조식 오기입, 물리화학적 성질 표기 부적절, 출처 기입 오류 등 작성자의 단순실수로 발생한 것으로 구분하였다. 이는 작성자가 기존 지침서를 제대로 확인하지 못했거나, 작성 시 방대한 정보를 검색 및 작성하는 과정에서 작성자의 실수로 인한 휴먼에러(Human Error)가 발생된 것으로 확인되었다. 두 번째, 기술오류 유형은 작성자가 화학물질 특성을 이해하지 못하거나, 또는 유해성정보 기반 면제조건 해당여부 판단 부족, 독성 시험 값의 해당 종을 제대로 파악하지 못한 경우 등 작업자 역량 부족에 따른 작성오류로 판단되었다. 마지막으로 구조적 오류는 지침 모호에 따른 무료자료의 유무 표기 오류, 작성양식에 면제조건 일부항목을 제시하지 않아 발생한 오류 등 데이터베이스 구축 지침서와 양식 등 작성방법 표준화 부족 등으로 발생하였다. 특히, 최초 작성한 데이터베이스의 오류 사항이 방대할 경우, 해당 작성자가 작성한 데이터베이스에 대한 검증작업 범위, 횟수 등에 관한 명확한 기준이 없어, 지나치게 많은 검증시간이 소요되어 비효율적인 상황이 발생하였다. 또한, 사업이 진행되는 과정에서 작성자와 검증자의 의견을 지침서에 수시로 반영함에 따라 작성기준이 자주 변

Table 5. Annexes and Appendices adopted to the guidance

Category	Contents
Annexes	How to check hazard information for unclassified substances
	How to check the hazard information of polymers and UVCB materials
	How to update database of the existing chemical substance
	How to review the completeness of a database of hazard information
	How to validate a database of hazard information
Appendices	NTP 14th Report on Carcinogens (RoC)
	IARC classification on carcinogenicity and corresponding report volume
	Substance list on EU risk assessment report
	File naming method across the database construction procedure
	Question and Answer
	How to use the 'General criteria' sheet in the excel form
	Decision of original source and the reference recording method
Exemption criteria for each test items	

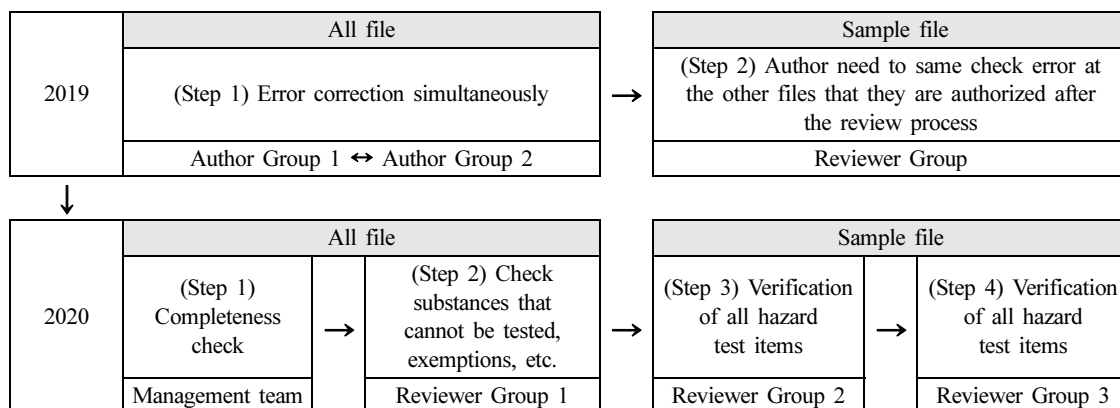


Fig. 3. Diagram of the Review Process for the Database

경되었는데, 이러한 잦은 변경이 작성자 및 검증자 혼란 가중을 초래하였다.

3. 데이터베이스 구축 표준방법 고도화

오류 유형 및 원인을 최소화할 수 있는 방법론을 모색한 결과, 데이터베이스 구축 지침서 개정, 분야별 유해성 전문가 교육, 화학물질 카테고리 분류, 데이터베이스 검증의 체계화 등 주요 방법론을 개선하였다(Fig. 2).

3.1. 개선 방법의 지침서 반영

작성자들의 문의가 많았던 참고 데이터베이스 검색방법에 대해서는 1차 작성되었던 지침서를 기반으로 시험방법 확인에 관한 내용을 요약·정리하여 부속서와 부록의 형태로 추가 보완하였다. 유해성 미분류 물질, 고분자화합물 및 UVCB (Unknown or Variable Composition, Complex Reaction Products and Biological Materials) 물질의 유해성정보 확인 방법, 완전성 검토 및 검증 매뉴얼 등의 내용은 부속서 형태로 구성하였다(Table 5). 예로, 유해성 미분류 물질 유해성정보 확인 방법에는 시험항목별 유해성정보 검색 키워드(keyword) 목록 및 웹사이트 내 정보의 검색 위치 등 작성자의 오류를 최소화하기 위하여 세부적으로 작성방법을 정리하였다. 또한, 부록에는 작성자들의 반복적인 질의 및 관련 응답정보는 Q&A로 정리하고, 많은 오류가 발생하는 유해성정보의 원문 출처 확인절차 및 정리방법 등을 위주로 구성하였다. 작성된 부속서 및 부록은 지침서에 첨부하여 작성자 및 검증자들의 참고자

료로 활용하도록 공유하였다.

3.2. 데이터베이스 작성자 및 검증자 교육 프로그램 운영

데이터베이스 구축의 정확도와 전문성 확보를 위해서는 작성자·검증자 대상 분야별 주기적인 교육이 필수적인 절차라고 판단하였다. 약 60명의 전문인력이 작성하는 데이터베이스의 표준화를 위해 「데이터베이스 구축 지침서」에 기반한 교육을 실시하고, 물리·화학적 특성 및 인체·환경분야별 이해도 제고를 위한 정기 워크숍을 개최·운영(분기 1회/3회 이상)하였다. 작성자 및 검증자 전원이 교육받을 수 있도록 참석여부에 대해 별도확인 후 미참석자에 대해서는 추가교육을 실시하였다. 교육 이후, 이메일 및 유선전화를 통하여 작성자의 문의사항을 확인하고, 답변하였다. 확인한 내용은 지침서의 별책 부록 형태인 Q&A 자료집에 반영하여 지속적으로 업데이트하는 절차를 따랐다. Q&A 자료집과 같은 방식을 적용하는 이유는 작성 과정에서 지속적인 작성 지침서의 업데이트가 작성자의 혼란을 초래할 수 있으므로 Q&A 자료집만 수시로 업데이트하도록 한 것이며, 향후 Q&A 자료집에 구축된 결과는 다음 단계의 데이터베이스 구축 시 작성 지침서 업데이트에 활용하도록 하였다.

3.3. 화학물질 카테고리 분류 정보 제공

금속화합물 등 구조나 특성에 따라 유해성정보의 유형이 유사한 경우가 있는데, 모든 작성자가 이러한 특성에 대한 전문성과 독성시험자료 등에

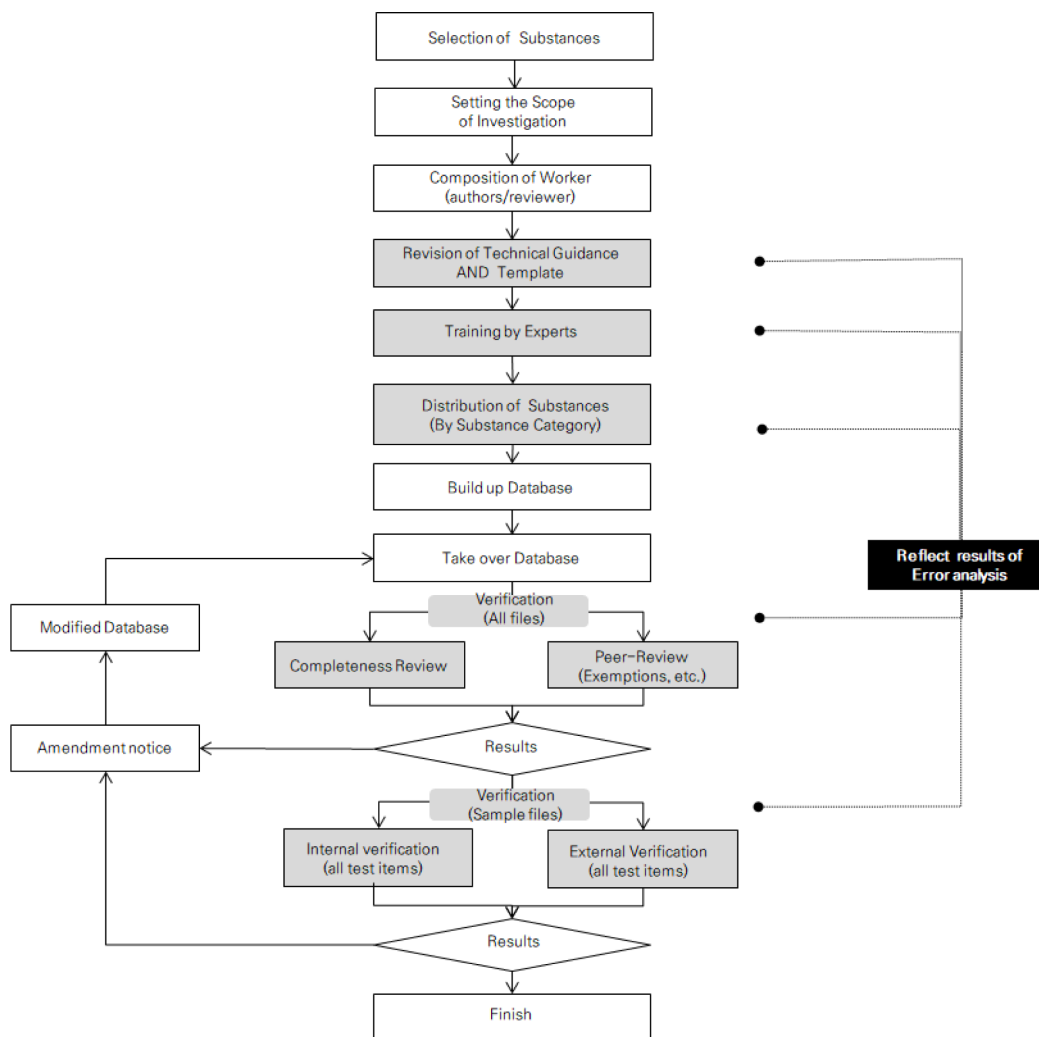


Fig. 4. Advanced Standard Procedure for Database Construction on Chemical Hazards (2020)

대한 전문성을 모두 갖추기는 어렵다는 한계를 확인하였다. 이에 사전에 화학물질의 구조·특성분야의 특화된 전문가가 전체 데이터베이스 대상물질에 대해 카테고리별로 분류하고, 그에 따른 데이터베이스 작성 특이사항을 안내하였다. 카테고리는 폭발성물질, 산화성물질, 가수분해성 물질, UVCB 물질, 고분자화합물, 중금속화합물로 구분하고, 일부 항목에 대해서는 기술적인 시험불가 물질, 무기물 등 해당물질의 구조, 성상, 물성 등을 기준으로 분류하였다. 이에 작성자는 해당물질의 물리·화학적 특성을 고려하여 작성하도록 하였고, 특히, 유해성정보 미분류 물질, 고분자화합물 및 UVCB 물

질은 일반물질과 달리 기존의 유해성정보가 많지 않을 것으로 예측됨에 따라 기존의 참고 데이터베이스 조사범위 외에 US EPA Dashboard와 Google scholar를 추가하여, 유해성정보를 작성하도록 하였다.

3.4. 데이터베이스 검증 체계 개선

데이터베이스 구축 시, 전문가가 작성하더라도 특정 분야의 지식기반 판단을 통한 정보 확인 및 작성 단계를 거치기 때문에 오류가 발생할 수 있다. 이에 단계별 검증을 통하여 신뢰도 높은 데이터베이스를 구축하고자 하였다. 1단계는 사업관리팀에 의한 데이터베이스 스크리닝 및 완전성 검토

이다. 이는 작성 지침서에 따라 필수적으로 작성해야 할 항목에 대해 작성되었는지 여부를 확인하는 단계이다. 2단계는 Peer-review를 통하여 기술적인 시험불가 여부, 시험자료 생산 면제 여부 및 등록 면제 물질 여부 확인 등 제출 면제조건을 제대로 판단하였는가를 확인하는 단계이다. 3단계는 작성자 간 데이터베이스 작성 오류를 검증·보완하는 단계이고, 4단계는 무작위로 추출된 일부 결과물에 대해 외부전문가에 의한 검증단계이다(Fig. 3). 매 단계마다 부적합 판단 시, 작성자에게 검토의견을 송부하고, 최종적으로 검토의견이 모두 반영되어 수정·보완된 경우에 사업관리팀에 의해 작성 완료를 확정하도록 하였다.

기존의 화학물질 유해성정보 데이터베이스 구축 체계(Fig. 1)에 대하여 앞서 언급한 주요 방법론을 보완하여, 최종적으로 정확성과 신뢰성이 고려된 화학물질 유해성정보 데이터베이스 구축 체계로 고도화하였다(Fig. 4).

IV. 고 찰

화학물질 데이터베이스 작성을 관련분야 박사급 인력으로 제한하여 유해성정보의 오류를 최소화하고자 하였다. 그럼에도 화학공학, 보건학, 환경공학 등 전문분야가 구축해야할 정보 중 특정분야에 한정되고, 물리·화학적 특성 및 인체·환경 유해성정보 전반을 아는 전문가는 국내 제한적이다. 2020년도 사업에서는 관련 분야 박사 또는 동등 경력의 전문가들까지 포함하여 약 60명의 인력풀을 구성하였다. 그럼에도 불구하고 발생될 수 있는 오류는 그간의 검증과정에서 발견한 오류 유형별로 개선대책을 마련하고, 주기적인 전문교육과 4단계로 세분화한 검증절차를 실시하여 보완할 수 있도록 노력하였다.

본 데이터베이스는 한정된 참고 데이터베이스에서 확인할 수 있는 정보에 대해서만 원문자료를 조사·정리한 것이다. 따라서 화학물질등록평가법에 따라 화학물질을 “등록”하기 위해서는 본 데이터베이스에서 확인하지 않은 다른 문헌을 조사하고 REACH 법에 따라 등록된 등록자료를 확인하는 등 추가적인 작업이 필요할 수 있다. 또한 본 데이터베이스에서 제공하는 정보는 화학물질등록평

가법, 같은 법 시행령과 시행규칙 및 관련된 행정규칙에서 정하는 사항을 종합적으로 고려하여 해석 및 활용되어야 한다. 또한, 정보의 참고 데이터베이스 등 업데이트 상황에 따라 유해성정보가 추가 또는 삭제되거나 분류결과 및 정보출처 URL의 변경 등이 발생할 수 있다. 제공하는 정보는 CAS NO. 기준으로 참고 데이터베이스와 해당 원문을 조사·정리한 결과이다. 그러므로 사용자는 자료 활용 전 반드시 해당물질과의 동질성 여부 확인이 필요하며, 원료와 제조공정의 차이 및 부산물·불순물 등으로 인해 정보 제공물질과 상이(UVCB, 고분자화합물 등)할 수 있으므로, 추가 확인이 필요하다. 이러한 내용은 데이터베이스 공개 웹사이트에 사용자의 사전 확인사항으로 명시하였다.

참고 데이터베이스 목록은 해외 국가·국제 유해성 평가보고서 및 국가에서 운영하는 웹사이트를 기반으로 하였다. 본 데이터베이스를 통해 기 확보되어 있는 화학물질의 유해성 자료를 활용할 수 있으므로, 자료의 중복생산을 방지하고, 시간 및 비용 절감할 수 있을 것으로 기대한다.

데이터베이스 구축사업은 다양한 출처로부터 많은 전문가들이 직접 정보를 확인하고, 데이터베이스로 옮기는 수작업으로 진행되었다. 이로 인하여, 발생된 오류의 유형이 대부분이다. 4차 산업혁명이 도래한 이 시점에서 인공지능(Artificial Intelligence, AI)을 활용한 데이터베이스 구축은 더 이상 비현실적인 일은 아니다. 정보 수집을 위한 명령어 알고리즘을 배열하고, 체계적으로 운용할 수 있게 논리적인 시퀀스를 나열한다면 정보를 수집하는 과정에서 발생하는 작성자의 단순오류 없이 데이터베이스를 마련할 수 있을 것이다. 그러나 다양한 유형의 수많은 정보들 중 목적에 부합하는 정보 선택에는 전문가의 판단이 필연적이다. 그러한 이유로, 향후 연구에서는 AI 활용과 전문가 판단을 아우를 수 있는 방법론이 개발되어야 할 것으로 사료된다.

V. 결 론

본 연구에서 데이터베이스 구축은 국내 제조·수입하는 중소기업의 화학물질등록평가법 이행 지원을 목적으로 진행되었다. 이에 정확하고, 전문성이

고려된 유해성정보를 제공하기 위하여 데이터베이스 구축 방법론의 신뢰성 제고 방안을 모색하였다.

한해 2,000~4,000여개의 화학물질에 대한 47개 항목의 유해성정보를 포함한 관련정보를 수집 및 정리하여 현재까지 데이터베이스를 구축한 결과, 대규모의 데이터베이스 작성 전문가가 필요한 상황이었다. 다양한 참고 데이터베이스로부터 각국의 원문으로 작성된 정보를 많은 작성자들이 표준화된 양식으로 작성함으로써, 발생 가능한 오류를 줄일 수 있도록 최대한 표준화된 정보를 구축하고자 노력하였다. 데이터베이스 구축은 많은 시간과 비용이 소용되는 작업이지만, 신뢰성이 낮은 정보는 오히려 사용자에게 혼란만 가중하는 상황을 초래할 수 있으므로, 정확성과 전문성을 고려한 신뢰성 높은 유해성시험 정보 데이터베이스를 구축하는 것이 요구된다.

작성된 데이터베이스 결과물은 화학안전산업계지원단 홈페이지를 통해 제공·운영될 수 있도록 지원하여, 1차적으로는 국내 화학물질등록평가법 등록 대상 기업이 자체적으로 유해성시험자료 확인 및 활용으로 등록 부담을 경감을 기대하며, 더 나아가 정부기관 및 연구기관 등에서 화학물질에 노출되어 있는 일반시민 및 근로자의 건강과 안전을 위해 데이터베이스 결과물이 활용되기를 기대한다.

본 연구는 국가 수준의 대규모 화학물질 데이터베이스 구축 시 신뢰성을 향상시킬 수 있는 방안을 제시하기 위해 진행되었으며, 이는 ‘2020년 기존화학물질 유해성정보 확인사업’에 적용하고 있다. 향후, 제안한 방법론 적용 전후에 대한 효과분석 등 후속연구를 기대한다.

감사의 글

본 연구는 환경부, 한국환경공단(2019년), 한국화학물질관리협회(2020년)의 연구용역 ‘기존화학물질 유해성정보 확인사업’의 지원을 받은 과제입니다.

References

1. The European Parliament and of the Council(EU),

Regulation of Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), EC No. 1907/2006

2. Ministry of Environment, The Act on Registration, Evaluation, etc. of Chemicals, Act No. 17326.; 2020
3. Ministry of Environment, Enforcement Decree of The Act on Registration, Evaluation, etc. of Chemicals, Presidential Decree No. 30849.; 2020
4. Ministry of Environment, Enforcement rule of The Act on Registration, Evaluation, etc. of Chemicals, Ordinance Of the Environment No. 840.; 2020
5. Ministry of SMEs and Startups, “Basic Statistics of Small and Medium Enterprises” 2019, from: URL: <https://www.mss.go.kr/site/smba/foffice/ex/statDB/MainSubStat.do>
6. Ministry of Environment, Industry Support Center 2020, from: URL: <https://www.chemnavi.or.kr/>
7. U.S. Department of Health and Human Services, National Toxicology Program, from <https://ntp.niehs.nih.gov/>
8. National Institute of Technology and Evaluation, NITEL, from: URL: <https://www.nite.go.jp/en/>
9. Australian Government Department of Health, Australian Industrial Chemicals Introduction Scheme (NICNAS), from: URL: <https://www.industrialchemicals.gov.au/>
10. European Chemicals Agency, ECHA, from: URL: <https://echa.europa.eu/>
11. Rim KT, Kim HO, Kim YK, Cho HW, Ma YS, Lee KS, Lim CH, Kim HY, Kim HY, Yang JS. Development and use of data for chemical risk assessment. *EHT* 2007; 22(1): 91-101.
12. Park YS, Hwang MS, Roh HR, Cheun GS, Kim SB. Development of chemical database and information system for safety management. *KIHM* 2015; 1(1): 58-64.
13. Kim YJ, SUE SY, The study on construction and utilization:management of suwon institute database, Suwon Research Institute.; 2015. p. 1-104
14. Korea Environment Corporation, Establishment Hazard information Database for existing chemical substance.; 2020

<저자정보>

이소민(학생연구원), 이민혁(전문위원), 강미진(사무관), 권순광(처장), 나진성(수석연구원), 박백수(실장)