



HAL
open science

Toxicological and Exposure Database Inventory: A review

Pascal Petit

► **To cite this version:**

Pascal Petit. Toxicological and Exposure Database Inventory: A review. International Journal of Hygiene and Environmental Health, 2022, 246, pp.114055. 10.1016/j.ijheh.2022.114055 . hal-04139621

HAL Id: hal-04139621

<https://hal.science/hal-04139621v1>

Submitted on 21 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Title**

2 Toxicological and Exposure Database Inventory: a review

3

4 **Author**

5 *Pascal Petit*^{1,2,3}

6 ¹ Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, CHU Grenoble Alpes, TIMC,
7 38000 Grenoble, France

8 ² CHU Grenoble Alpes, Centre Régional de Pathologies Professionnelles et Environnementales,
9 Grenoble, France

10 ³ Univ. Grenoble Alpes, AGEIS, 38000 Grenoble, France

11

12 ORCID ID: 0000-0001-9015-5230

13 Corresponding author at: Laboratoire TIMC (UMR 5525 CNRS–UGA), Equipe EPSP, UFR de
14 Médecine de Grenoble, Domaine de La Merci, 38706 La Tronche Cedex, France.

15 E-mail address: pascal.petit@univ-grenoble-alpes.fr (P. Petit).

16

17 **Highlights**

- 18 • A review of databases that can support human risk assessment to chemicals is needed
- 19 • 715 databases (DBs) were identified and described from a literature review
- 20 • The chemicals studied and the summary information reported was often very limited
- 21 • A free searchable online inventory was set up to collect and share meta-data on DBs
- 22 • This inventory of toxicological and exposure DBs remains open for further additions

23 **ABSTRACT**

24

25 **Background:** Knowledge of hazards and exposure data about chemical pollutants are essential for
26 epidemiological studies for assessing health risks and preventing the development of diseases. Access
27 to this type of information is clearly needed and although numerous databases (DBs) have been
28 compiled, they are often extremely difficult to find and navigate.

29 **Objectives:** This paper presents the creation of an online inventory of toxicological and exposure DBs
30 to support human health risk assessment resulting from the chemical exposome.

31 **Methods:** A free searchable online inventory, called TEDI (Toxicological and Exposure Database
32 Inventory), was set up to collect meta-data on toxicological and exposure DBs resulting from a review
33 of the literature conducted on PubMed. Only chemical agents, including drugs, natural and
34 manufactured/synthetic chemicals and air pollutants, were considered. Regarding exposure DBs, only
35 human exposure to chemicals were included. Because of time and resource constraints, only an
36 inventory was performed, and no effort was made to rate, assess or rank DBs on quality criteria.

37 **Results:** A total of 715 DBs were identified, catalogued, described and made available on TEDI. The
38 number of chemicals ranged from one single chemical to 90 million for toxicological DBs and from 1
39 to 103,817 chemicals for exposure DBs. Some DBs were restricted to specific chemicals, with for
40 instance 71 exposure DBs reporting polycyclic aromatic hydrocarbon measurements. Toxicological DBs
41 (n=362) provided a broad range of different information such as toxicity data (57%), physicochemical
42 properties (38%), experimental study results (28%) or prediction data (16%). A total of 382 exposure
43 DBs were found, including exposure measurements DBs (71%), web application/tool (21%), exposure
44 and risk assessment models (19%) and epidemiological DBs (13%).

45 **Discussion:** TEDI is of value to the broader community and could support human health risk assessment
46 to chemicals in various contexts. This inventory remains open for further additions, to enlarge its
47 coverage, increase the meta-data collected and include newly developed DBs.

48

49 **KEYWORDS:** toxicological database; exposure database; chemical exposome; health risk assessment;
50 inventory

51 **Abbreviations**

52 ADME: absorption, distribution, metabolism, excretion

53 DB: database

54 EU: European Union

55 GBD: Global Burden of Disease

56 GHS: Globally Harmonized System of Classification and Labeling of Chemicals

57 JEM: job-exposure matrices

58 LB: lower bound

59 MS: Microsoft

60 OMEGA-NET: Network on the Coordination and Harmonization of European Occupational Cohorts

61 PAHs: polycyclic aromatic hydrocarbons

62 QSAR: structure-activity relationship

63 QSPR: quantitative structure-property relationship

64 REACH: Registration, Evaluation, Authorization and restriction of CHemicals

65 SAR: structure-activity relationship

66 TEDI: Toxicological and Exposure Database Inventory

67 UB: upper bound

68 1. Introduction

69 The introduction of the exposome concept emphasizes on the multiplicity, interaction and complexity
70 of health risks. The exposome can be defined as the study of the impact of all potential stressors on
71 human health (including chemical, biological, physical, psychological, social, etc.) from conception
72 onwards (Dennis et al. 2017; Kim and Hong 2017; Siroux et al. 2016; Wild 2005, 2006). This concept,
73 first formulated in 2005, was developed to highlight the need for better and more complete exposure
74 data related to all life-stages for various populations (Siroux et al. 2016; Wild 2005).

75 Because chemicals are among the major environmental threats to human health, knowledge of hazards
76 and exposure data about chemical pollutants (chemical exposome) are essential for assessing health risks
77 and preventing the development of diseases. Risk assessment can be viewed as a four-step process
78 including hazard identification, dose-response assessment, exposure assessment, and risk
79 characterization (Persad and Bobst 2020). Since the REACH regulation, knowledge about hazards has
80 increased, but there is still a scarcity of data pertaining to thousands of chemicals, in particular for
81 chronic hazards, chemical mixtures and the condition of use of many chemicals (ECHA 2015; Wignall
82 et al. 2018). Regarding exposure data, the knowledge is much more limited and highly variable. Some
83 chemicals have been intensively monitored, such as asbestos, while others have been well characterized
84 in various media (water, sediment, food) but poorly in humans. However, for most chemicals, there is a
85 paucity of knowledge and even occasionally no data, with only a small fraction of the chemical
86 exposome covered (Vineis et al. 2017; Wignall et al. 2018).

87 Technological advancement has led to evolving exposure environments/conditions and to a drastic
88 increase in the health data generated and captured. To be used for research purposes, the generated data
89 have to be collected, recorded, stored and properly handled (Gaudet et al. 2011; Zheng et al. 2017). A
90 great diversity of databases (DBs) exists, whose nature and scope, as well as the degree of complexity
91 vary immensely (Andreu-Perez et al. 2015; Frelinger 2015; Sowe and Zettsu 2014; Steckling et al. 2018;
92 Zheng et al. 2017). The true challenge is to identify what data are truly relevant/useful and concentrate
93 efforts on gathering it. Access to this type of information is clearly needed and although numerous DBs
94 have been compiled, they are often extremely difficult to find and navigate. A concerted effort is needed

95 to catalogue available toxicological and exposure DBs that can support human health risk assessment of
96 the chemical exposome.

97 This paper attempts to catalogue relevant DBs that might benefit human health risk assessment resulting
98 from the chemical exposome into an easily accessible format. This inventory could be a useful
99 tool/resource for epidemiologist, risk assessors, risk managers, practitioners, industry practitioners,
100 decision makers and sanitary agencies to obtain and find relevant information from available resources
101 when conducting risk assessment to chemicals.

102

103 **2. Material & Methods**

104

105 *2.1 Scope of the study*

106 The scope of this study was to create an online inventory to collect meta-data on various toxicological
107 and exposure DBs into an easily findable, accessible, interoperable, and re-useable format. No attempt
108 to conduct an analysis of the scientific content of the identified records was undertaken because it was
109 outside the scope of this study. In addition, no data quality analysis (qualitative or quantitative) of the
110 DBs compiled from this work was made because it was also outside the scope of this work.

111 For this study, a toxicological DB was defined as a collection of data which included at least one of the
112 following information: physicochemical properties, results of experimental studies, reference/limit
113 values (e.g., cancer slop factor), prediction data, prediction models, dose response models or other
114 toxicity information. An exposure DB was defined as a collection of data which included at least one of
115 the following information: results of the analysis/monitoring of one or several chemicals in at least one
116 media (e.g., air, urine, food), information on the number of people exposed to at least one chemical,
117 exposure models used to rank or estimate chemical exposure levels or health risk assessment tools (e.g.,
118 software, web application, methods or numerical analysis models) used to qualify or quantify the risk
119 posed to human health resulting from chemical exposure.

120 For this work, only chemical agents, including drugs, natural and manufactured/synthetic chemicals
121 (including unwanted by-products) and air pollutants, were considered as chemical exposome stressors
122 and referred to as chemicals in the remaining of this paper. Biological agents (e.g., mycotoxins, virus),

123 physical agents (e.g., radiation), and active smoking were outside the scope of this inventory. Regarding
124 exposure DBs, only human exposure to chemicals were considered.

125

126 *2.2 Literature search*

127 For the purpose of this study, a review of the literature, conducted by one author, was undertaken using
128 PubMed. All identified papers addressing the subject of toxicological or human exposure DBs regarding
129 chemicals that were published in peer-reviewed journals were considered for inclusion. Edited books
130 and book chapters were also considered. Gray literature was also examined. No records that were
131 deemed relevant to the scope of this study were purposely excluded from this review. Using PubMed,
132 the following search strategy/query was used: *(toxicological OR toxicology OR exposure OR exposome)*
133 *AND (database OR dataset OR databank)*. The search was carried out on February 7th 2022 and was
134 limited by language: 'English' and 'French' and by date: from 1950 to the end of January 2022.

135 When possible, the main information (e.g., DB name, topic, country, number of data observations,
136 references) and other meta-data regarding the identified DBs were retrieved. The meta-data of interest
137 to be retrieved during the literature search are listed in Table 1. The characteristics/meta-data of each
138 DB were collected using only information that was available from the screened records. Descriptors of
139 interest that were never available or almost never reported in the screened records were excluded.
140 Because of time and resource constraints, only an inventory was performed, and no effort was made to
141 rate, assess or rank DBs on quality criteria. In addition, no effort was made to reach out directly to the
142 owners of the DBs to request additional details.

143 **Table 1:** List of meta-data of interest to collect from the literature search

| Database type | Database descriptor | Example |
|--------------------------------------|-----------------------------|--|
| Toxicological and exposure databases | Name | ACROPOLIS |
| | Creator | Institution |
| | Owner | Institution |
| | Contact person | Name and email address |
| | Status | Still active |
| | Data availability | Open |
| | Country/Scale | The Netherlands |
| | Time period | 2001-2010 |
| | Number of agents | 97 |
| | Topic/Agents | Chemicals |
| | Full list of chemicals | Arsenic, cadmium, lead |
| | Reference | Bopp et al. 2018 |
| | Toxicological database only | Data type |
| Exposure database only | Data origin | Original study |
| | Exposure setting | Environmental exposure |
| | Media | Air |
| | Monitoring strategy | Airborne monitoring (area) from 8,405 air stations located in 38 countries |
| | Number of data | 166 |
| | Data type | Samples |

144

145 2.3 Creation of an online tool – TEDI

146 To make the inventory of toxicological and exposure DBs easily findable, accessible, interoperable, and
 147 re-useable, a user friendly and interactive web application/tool, named TEDI (Toxicological and
 148 Exposure Database Inventory), was created. It was built using the Shiny package in the R programming
 149 language (Chang et al. 2018). TEDI can be accessed on any device with an internet browser and requires
 150 no programming knowledge to use. TEDI is freely available to anyone who wants to use it and is
 151 available at: <https://exporisk-timc.imag.fr/TEDI/>.

152 Data analysis, all calculation and the creation of TEDI were performed with R software 4.1.2® (R Core
 153 Team, Vienna, Austria) for Windows 10©.

154

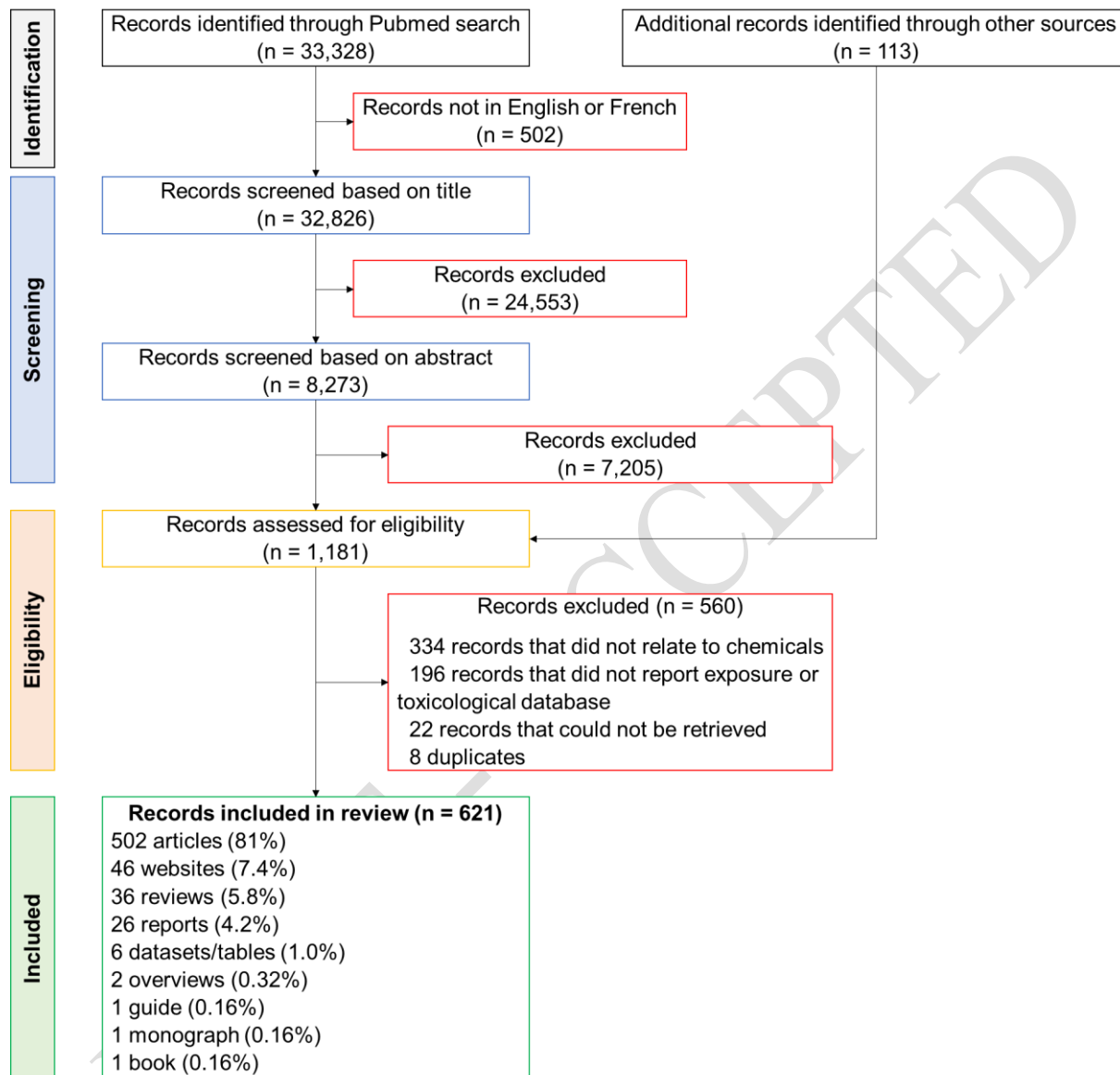
155 3. Results

156

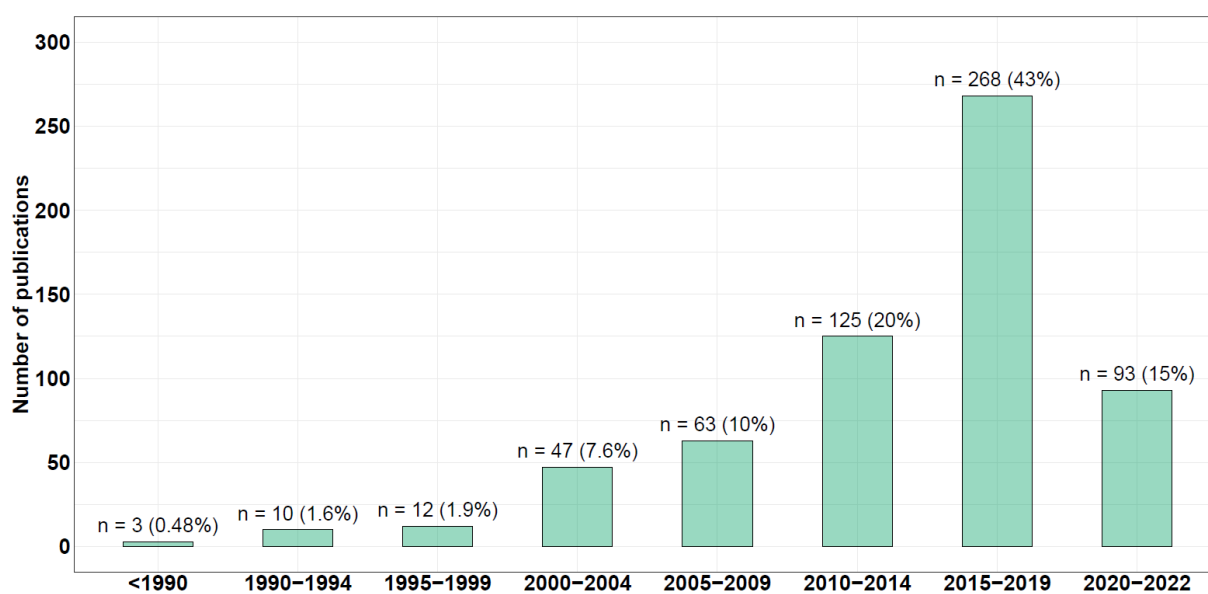
157 3.1 Literature search

158 The literature search yielded 33,328 records that had been published before the end of January 2022. A
 159 total of 113 additional records were identified through other sources (gray literature). Irrelevant and
 160 duplicate records (32,820) were excluded based on the examination of titles, abstracts, and full-texts,
 161 after which 621 records were kept (Fig.1). Most of the records screened were articles (81%), followed

162 by websites (7.4%), reviews (5.8%), reports (4.2%) and other resources (1.7%). Most references were
 163 published after 2005 (88%), and almost two-third of them (58%) were published since 2015 (Fig.2).
 164



165
 166 **Fig.1** Flow chart depicting the literature search and the evaluation process for finding relevant records



168

169 **Fig.2** Number of records per 5-year period

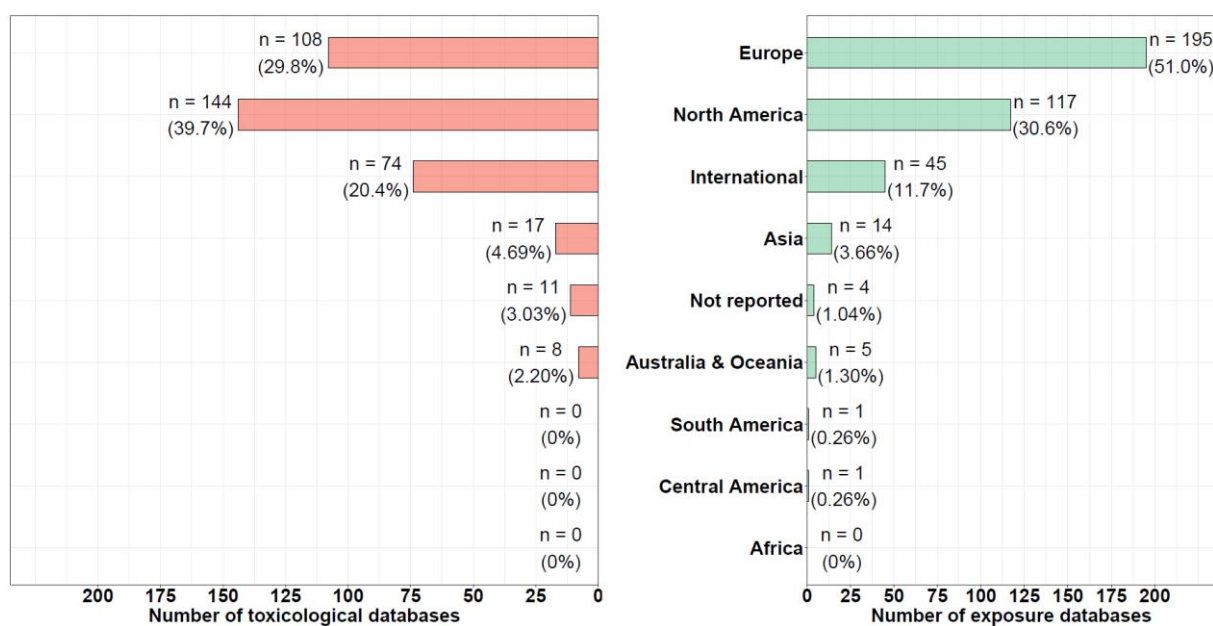
170 Regarding the bar on the right side of the graph (2020-2022), the number of records corresponds only
 171 to the total number of records from 2020, 2021 and January 2022

172

173

3.2 Database inventory

174 A total of 715 toxicological and exposure DBs were identified through the literature search (Fig.1).
 175 Some records described more than one DB, which explains why there were more DBs identified than
 176 records screened (715 vs. 621). There were 333 DBs (46.6%) solely classified as toxicological DBs, 353
 177 DBs (49.4%) solely classified as exposure DBs and 29 DBs (4.1%) that were classified in both
 178 categories. The geographical coverage showed that most DBs were developed in North America and
 179 Europe (Fig.3). No DB from Africa was found while only a few DBs from South America, Central
 180 America, Asia, Australia and Oceania were identified.



182

183 **Fig.3** Number of databases by continent and database type

184

185 A total of seven descriptors of interest were excluded and not recorded on TEDI because they were
 186 either not available or almost never reported in the screened records. These descriptors were the owner,
 187 the creator, the contact person, the data availability, the full list of chemicals, the time period and the
 188 status of the DB. As a results, to address this shortcoming, a proof of concept was undertaken using the
 189 screened records to investigate the possibility of collecting more detailed information about the missing
 190 descriptors of interest. This proof of concept focused only on one descriptor (the full list of chemicals)
 191 and was restricted to exposure DBs reporting results of airborne and/or biological monitoring to
 192 polycyclic aromatic hydrocarbons (PAHs), which is one of the most studied family of carcinogens
 193 (Mallah et al. 2022; Yang et al. 2021). The results of the proof of concept are detailed below in
 194 subsection 3.2.3.

195

196 **3.2.1 Toxicological databases**

197 Among the 715 DBs identified, 362 (50.6%) were classified as toxicological DBs. Most of the time,
 198 they were made up of data retrieved through literature reviews, data collation and/or data collection.
 199 Data on a wide variety of chemicals were available. The number of chemicals found in the identified

200 toxicological DBs ranged from one single chemical to 90 million. They were 83 (23%) DBs that
201 investigated specific topics or classes of chemicals, in particular nanomaterials (n=20), pesticides
202 (n=19), chemicals in food (n=13), chemical mixtures (n=11), endocrine disrupting chemicals (n=9),
203 carcinogens (n=6) and PAHs (n=4).

204 A broad range of data types were available (Fig.4). A total of 208 (57%) toxicological DBs provided
205 toxicological information while 16 reported ecotoxicological data. There were many types of toxicity
206 information reported such as data on target organ, mode of action, adverse effects, as well as toxicity
207 profile, toxicity thresholds and dose-response relationship of a chemical in different organisms and/or
208 the environment. Some of these DBs also provided information on pharmacokinetic data (n=10), in
209 particular ADME (absorption, distribution, metabolism, and excretion), but also on adverse outcome
210 pathway (n=7).

211 A total of 139 (38%) DBs provided chemical structure and physicochemical properties. Most of these
212 DBs were usually freely accessible and had their own search engine. Information regarding the chemical
213 identity (e.g., name, synonym, formula, structure, identification number) were often provided. A broad
214 range of physicochemical properties could be found such as vapor pressure, solubility, partition
215 coefficients, molar mass, molar volume, pH and critical temperature. Chemotype and activity
216 annotations were sometimes also available

217 A total of 100 (28%) DBs gathered results from experimental studies. Most experimental data came
218 from animal and *in vitro* studies. DBs with experimental data provided results for 1 to 1,563 different
219 species. There were many types of assays that were reported, in particular carcinogenicity, mutagenicity,
220 genotoxicity, hepatotoxicity, acute toxicity, developmental toxicity and skin sensitization tests.

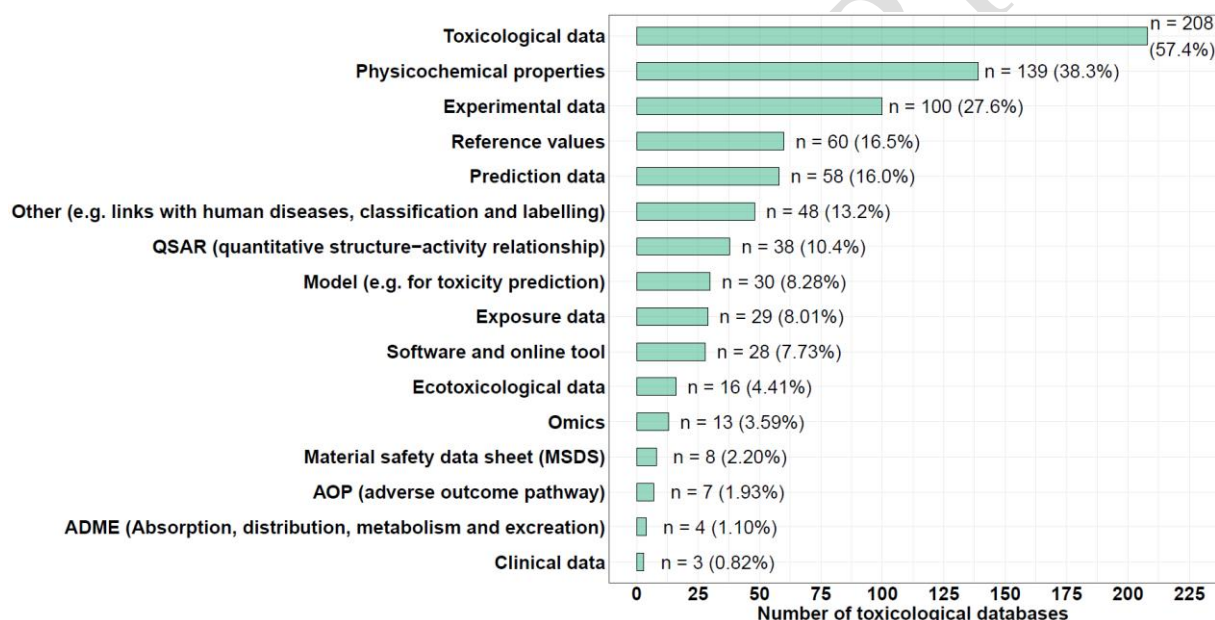
221 Several toxicological DBs provided prediction data (16%) as well as different models (8%) that can
222 mostly be used to predict toxicity. Toxicity prediction was usually performed with quantitative structure-
223 activity relationship (QSAR) (n=38), read-across, structure-activity relationship (SAR), or quantitative
224 structure-property relationship (QSPR) models.

225 A total of 60 (17%) of the DBs gathered reference values. There were also 12 DBs that reported threshold
226 of toxicological concern. A broad range of reference values were recorded, such as cancer slope factors,
227 unit risks or tolerable daily intake. Most of the time, for each reference value, the chemical, the route of

228 exposure, the targeted population, the adverse effect, the target organ, the type of reference value (e.g.,
229 acute, chronic) and the year of establishment was provided.

230 Other information was also available. There were 15 DBs (4%) that recorded safety data, such as
231 material safety data sheets (n=8). A total of 13 (4%) DBs reported omics data, with 9 DBs that gave
232 information on toxicogenomics, 3 on metabolomics and 1 on chemogenomics. A total of 12 DBs
233 recorded classification and labeling information according to the Globally Harmonized System of
234 Classification and Labeling of Chemicals (GHS) and European Union (EU) rules. There were 3 DBs
235 that recorded links between chemicals and human diseases or conditions.

236



237

238 **Fig.4** Number of toxicological databases by data type

239

240 3.2.2 Exposure databases

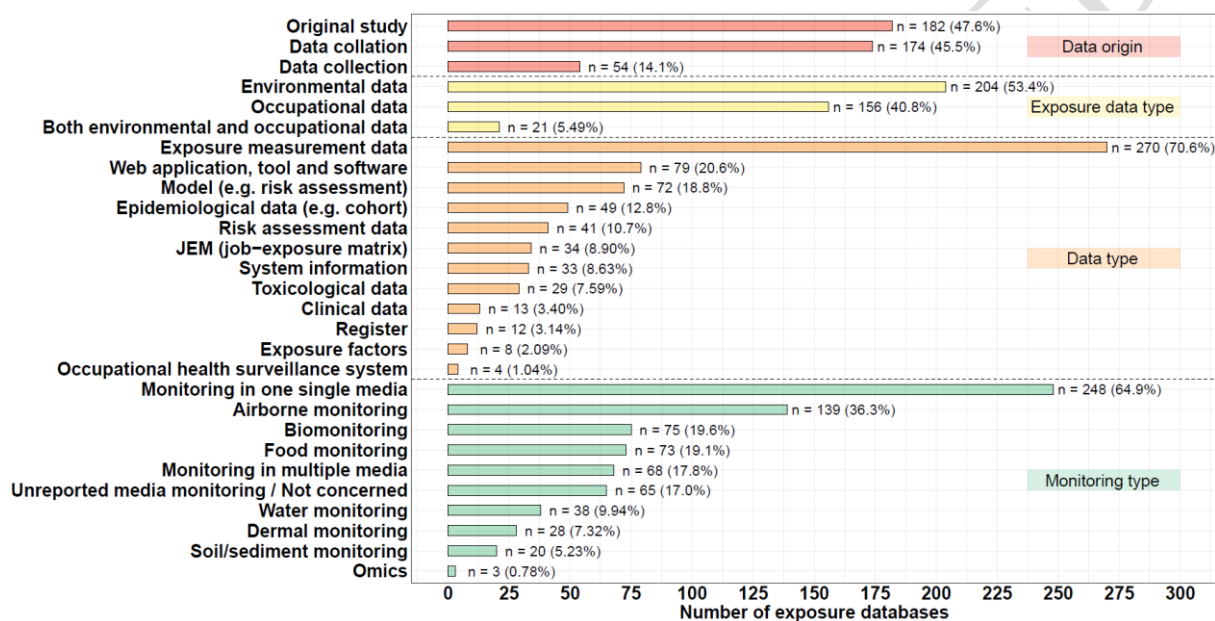
241 Exposure DBs are useful for estimating exposure levels and profiles, estimating the number of
242 individuals exposed, identifying exposure determinants, tracking co-exposures/multiple exposures,
243 triggering and guiding medical surveillance, estimating spatial and temporal exposure traceability, etc.

244 Among the 715 DBs identified, 382 (53.4%) were classified as chemical exposure DBs. There were
245 many types of exposure DBs, such as exposure measurement DBs (71%), epidemiological DBs (13%),
246 job-exposure matrices (JEM) (9%), exposure information systems (9%), national registers (3%), clinical

247 DBs (3%) and occupational health surveillance systems (1%) (Fig.5). In addition, around 20% of the
248 exposure DBs were web application/tool or provided different statistical models that were used, for
249 example, to estimate and/or rank exposure levels or to conduct health risk assessment. A wide variety
250 of exposures to chemicals were covered, with between 1 to 103,817 chemicals reported. Most data
251 originated from original studies (48%) and data collation – i.e., compilation of data directly from the
252 source (e.g., companies, physicians, hygienists, agencies) - (46%). Some exposure DBs pertained solely
253 to environmental exposures (53%) or to occupational exposures (41%) while 6% of them related to both.
254 Most exposure DBs had a national stature. Occupational exposure DBs usually represented multiple
255 industries, occupations, and workplaces over multiple years and sometimes decades.
256 National registers provided information on the number of exposed people (mostly workers) and their
257 exposure to given chemicals (usually carcinogens). In these registers, only a few chemicals were
258 recorded. Exposure information systems also provided information about the number of people exposed,
259 the levels of exposure, the places where the exposure occurred and sometimes emission inventories.
260 Clinical DBs gathered cases, exposure information and factors seeking to explain disease occurrence or
261 prognosis, while epidemiological DBs recorded data to study exposure determinants that play a role in
262 the occurrence of a given disease in a particular human population. Regarding JEMs, they were
263 commonly used to assign exposure estimates, within a given industry, department or for a specific job
264 title, to subjects in epidemiological studies. JEMs were usually created for occupational exposure but
265 also existed for the general population.

266 Regarding exposure measurement DBs, some of them had administrative purposes and were made for
267 insurance, control, compliance or regulatory purposes. This was in particular true for occupational
268 exposure DBs. Within industry, chemical exposure measurements were typically carried out for the
269 purpose of compliance with occupational safety and health regulations or under specific circumstances
270 (e.g., incidents, turnarounds). When reported, there were between 50 to more than 2 million samples
271 recorded in exposure measurement DBs. DBs with a high number of samples were usually made from
272 data collation and/or literature search (data collection). Most samples were measured into one single
273 media (65%). The monitoring strategies the most used were airborne monitoring (37%), biomonitoring
274 (19%) (e.g., urine, blood) and contaminant monitoring in food (19%) (Fig.5). Airborne monitoring

275 consisted of measuring the concentrations of chemicals in the air. Biomonitoring consisted of measuring
 276 the concentrations of chemicals and/or their biomarkers in body fluids (e.g., blood, urine, and breast
 277 milk) or tissues (e.g., hair, nails, fat, and bone). In contrast to targeted analyses of single compounds,
 278 exposomic biomonitoring, which are untargeted analyses of many analytes using omics technologies
 279 (e.g., metabolomics), were still rare (only 3 DBs identified). Some exposure measurement DBs
 280 contained measurements of numerous chemicals, while others focused on one family of compounds.
 281



282
 283 **Fig.5** Number of exposure databases by data type

284
 285 3.2.3 Proof of concept – PAH exposure measurement databases

286 Among the 270 exposure measurement DBs identified, a total of 71 (26%) contained information about
 287 exposure to PAHs. Occupational exposures to PAHs were investigated alone in 49% of the DBs and in
 288 parallel to environmental exposures in 10% of them. For most DBs (73%), it was possible to collect the
 289 full list of PAHs and/or PAH biomarkers. A total of 28% DBs reported measurements of PAH substitutes
 290 (products containing or emitting PAHs) and 18% did not report which PAH congeners were measured.
 291 Most DBs recorded airborne samples (73%), with between 1 to 39 PAHs measured. Only 8% of the DBs
 292 reported both airborne monitoring and biomonitoring. When biomonitoring was conducted (31% of the

293 time), between 1 to 38 biomarkers were measured, with 1-hydroxypyrene levels reported most of the
294 time (82%).

295

296 3.3 Online inventory – TEDI

297 The results of the inventory were implemented online. Fig.6 presents a screenshot of the TEDI tool. On
298 top of the web page (on the left), there are 8 tabs, written in orange fonts, that can be clicked on. The
299 active tab has an orange background and a bold black font. The first three tabs (starting from the left)
300 display as interactive tables the toxicological, exposure and PAH exposure measurement DBs,
301 respectively. The “*Reference*” tab provides a table listing all of the references mentioned in the reference
302 column of the three aforementioned tables. The “*News & Articles*” tabs will report articles describing,
303 using or citing TEDI. The “*Help/Info*” tab provides general information regarding TEDI, such as the
304 goal, strengths and limitations of the tool as well as information regarding table interactivity, information
305 pertaining to condition of use, data privacy, data protection and legal mentions and how to report bugs
306 or mistakes. The “*Contact*” tab provides contact information details while the “*Mentions légales*” tab
307 stipulates the legal mentions in French.

308 All of the tables available on TEDI are interactives. There is a help button below each table, named
309 ‘*How to interact with this table?*’, which explains how to interact with all the tables. On top of each
310 table (on the left), there are 7 action buttons. The “*Copy*” button allows the user to copy the table to
311 clipboard. The “*CSV*”, “*Excel*” and “*PDF*” buttons enable the user to export the table to CSV, XLS or
312 PDF format, respectively. The “*Print*” button allows the user to print the table. The user can choose
313 which columns to visualize using the “*Column visibility*” button. The last button, “*Show 10 rows*” allows
314 the user to choose the number of rows to display. When the former button is not set on the “*Show all*
315 *rows*” option, the user will have to click on the buttons under the table (on the right) to navigate through
316 the table. Beware, for exporting or printing the entire table, the user needs to first click on the number
317 of rows to display button (named “*Show 10 rows*” by default), then choose the *All* option and finally to
318 click on one of the three export/download buttons.

319 The interactive tables allow the user to perform queries, either on a specific column or on the entire
320 table. The global search is performed across all searchable columns and is starting when typing text on

321 the “*Search*” rectangle/bar located on top of the table (on the right). Searching on individual columns
 322 can be performed by typing text on the white rectangle/bar below each column header. Queries are not
 323 case sensitive and provide on-the-fly filtering with immediate feedback to the user. For the global search,
 324 the order of the words in the query does not matter. Clicking on the cross symbol that appears on the
 325 right of the query will cancel the search. Columns containing numbers (e.g., number of data) can be
 326 filtered either using the rule that appears when clicking on the column search rectangle under the column
 327 header or by typing the following command in the search rectangle cell: '*LB ... UB*', where *LB* and *UB*
 328 are the lower and upper bounds, respectively.

329 It is possible to sort the table based on one column by clicking on a column double arrow head located
 330 on the right of each column header. Clicking on the arrowhead facing the top of the web page will sort
 331 the chosen column either alphabetically or in ascending numerical order, depending on the type of data
 332 the column contains (numerical values or texts). It is also possible to change the order of the columns
 333 of a table by clicking and dragging the column header to the left or to the right.

334 In each column, cells with terms underlined and written in orange font refer to special terms and
 335 abbreviations that can be clicked on to display the full definition/details. References and access links
 336 (last column of each table) are also underlined and written in orange font and can be clicked on to open
 337 the website where the paper/record can be viewed/obtained. Note that empty cells correspond to non-
 338 reported values.

339

TEDI - Toxicological and Exposure Database Inventory

[Toxicological databases](#)
[Exposure databases](#)
[PAH databases](#)
[References](#)
[News & Articles](#)
[Help/Info](#)
[Contact](#)
[Mentions légales](#)
[Tabs](#)

There are currently a total of 362 toxicological databases (out of 715) listed in TEDI.

Tab. 1: Inventory of toxicological databases

Action buttons Search:

| Name of the database | Country/Scale | Number of agents | Topic/Agents | Data type | Reference |
|----------------------|-----------------|------------------|---|--|--|
| All | All | All | All | Column search | All |
| ACROPOLIS | The Netherlands | | Pesticide residues | Monte Carlo Risk Assessment software embedded in a web-based environment that allows for the assessment of dietary exposure to pesticide residues, with information on experimental data (results from in vivo and in vitro studies) | Bopp et al. 2018 ; van Klaveren et al. 2015 |
| ACToR | United States | 500000 | Environmental chemicals | Physicochemical properties, experimental data , exposure data, toxicological data , health risk assessment tool | Bond and Garny. 2019 ; Judson. 2010 ; Judson et al. 2012 ; Mahammedi et al. 2020 |

Horizontal slider **Table navigation**

Showing 1 to 2 of 362 entries

[How to interact with this table?](#)

Previous ... Next

340

341 **Fig.6** Screenshot of the TEDI app

342 Orange fonts that are underlined refer to hyperlinks providing either the definitions and descriptions of
343 terms or direct access to the mentioned references. Buttons allow user actions such as copying the table
344 to clipboard, saving the table as CSV/XLS/PDF, printing the table and selecting columns to visualize.
345 It is possible to make global (entire table) and column specific queries. Beware, when opening TEDI on
346 a web browser there are several horizontal sliders that will be available, with one on the web page itself,
347 one inside the toxicological DB table (at the bottom) and one inside the exposure DB table (at the
348 bottom). To access all of the available information reported in these two tables, the user has to use the
349 different horizontal sliders (moving it on the right or left) or to download the table.

350

351 **4. Discussion**

352 Toxicological and exposure DBs are complementary tools for conducting risk assessment that can help
353 linking exposures to adverse outcomes, establish reference values, develop primary prevention and thus
354 decrease disease incidence (Atwood et al. 2019). An online inventory collecting meta-data on various
355 toxicological and exposure DBs was created. A total of 715 DBs were catalogued, described and are
356 available on TEDI.

357

358 *4.1 Remarks about databases identified*

359

360 *4.1.1 Data availability and geographical coverage*

361 The understanding of, access to and acquisition of relevant information was often difficult. For many
362 DBs, especially exposure DBs, the summary information reported was very limited. Several available
363 resources (papers, reports, and websites) lacked necessary details to provide adequate descriptions of
364 the DBs. In addition, some DBs had no name (7.2%) and it was difficult if not impossible to identify the
365 proper contact person for potential correspondence and access to additional information. Based on the
366 current inventory, DBs developed for Eastern Europe, Southern Europe, Africa, South America, Latin
367 America and Asia were lacking.

368

369 4.1.2 Data accuracy and quality

370 The identified DBs contained a wealth of information that was relevant to risk assessment, but the value
371 of these sources could not be considered equal. According to some authors, the quality and accuracy of
372 available information greatly varies and is inconsistent from one DB to another, which has a direct
373 impact on the quality and reproducibility of the resulting risk assessment that may be conducted (Bopp
374 et al. 2018).

375 DBs were heterogeneous in nature, not only in terms of objectives and strategies but also in the accuracy
376 and completeness of information available as well as in the general architecture and the semantics used.
377 This heterogeneity mainly came from different economic, technical, practical and human choices that
378 had been made by the founder(s)/creator(s) of each DB as well as the disparate purposes for which DBs
379 were developed.

380

381 4.1.3 Toxicological databases

382 For most existing chemicals, there are no comprehensive toxicological data or some form of human
383 health risk assessment has not been undertaken (Wignall et al. 2018). In addition, determining which
384 available toxicological DB is the most suitable for addressing the goal(s) of a study depending on the
385 desired aim might be potentially difficult (Laamanen et al. 2008). Indeed, the reported health effects
386 could be different from one DB to another, depending on the methods, tests, experimental conditions,
387 assumptions made, chemicals and species used. In addition, the degree of reliability and validity of the
388 available data could highly vary, making it difficult to determine which relevant information to use
389 (Beck et al. 2016).

390

391 4.1.4 Exposure databases

392 Most exposure DBs provided low temporal resolution information due to the complexity of gathering
393 complete, rich long-term time-series data. Only a sample of the whole population of interest was usually
394 studied, which may only represent a narrow picture of the development of a disease. In addition, most
395 exposure DBs (52%) reported results from monitoring of chemicals in various media/matrices (e.g.,
396 food, water, sediments). While these types of monitoring are easier to undertake than monitoring

397 chemicals in human body fluids or tissues, they do not provide information on individuals but only on
398 media contamination, which makes the results sometimes difficult to interpret and not necessarily useful
399 for the establishment of prevention and protection measures.

400 Regarding occupational measurements DBs, chemical exposure measurements were typically carried
401 out for the purpose of compliance with occupational safety and health regulations or under specific
402 circumstances (e.g., incidents, turnarounds), which explained why the focus was most often on
403 carcinogens and high exposure tasks/events.

404

405 *4.2 Limitations of this work*

406 While a comprehensive literature search was conducted, the search was not exhaustive and not
407 completely systematic. The search sentence may have not yielded every possible existing toxicological
408 and exposure DBs, therefore results presented in this work only reflects this query. Hence, the
409 identification of some type of DBs (e.g., JEMs) was not optimal. For instance, when screening records
410 related to exposure data, it was found that some exposure models were either part of exposure DBs or
411 provided exposure measurement data. Therefore, they were classified/considered as a category of
412 exposure DBs in this work. If a literature search had been conducted using specific keywords related to
413 exposure models, it would have likely allowed for the identification of more exposure models than what
414 was found in this work. The same is true for other categories of toxicological and exposure DBs. This
415 might also explain why some DB categories, in particular exposure measurement DBs were more
416 identified than other exposure DB categories.

417 Only one author conducted the inventory and created the online inventory. Therefore, it is possible that
418 some DBs may have been missed and the presence of some mistakes (e.g., typos) and bias cannot be
419 excluded. Two independent reviewers may have decreased the chance of systemic error and bias. In
420 addition, the literature search was conducted using only one source (PubMed). To conduct a more
421 comprehensive and systematic work, other bibliographic DBs such as Web Of Science, Scopus or
422 Dimensions could have been used. The information reported on TEDI came from screened records
423 (mostly papers) that may not report the most up-to-date information regarding the DB(s). Thus, some
424 information on TEDI is likely to be incorrect and is subjected to evolve if new data become available.

425 Potential misclassification of DBs is possible since the content of several DBs spanned multiple
426 categories so that DBs often do not sit comfortably in a single category. This is not surprising because
427 risk is a function of hazard and exposure and therefore some degree of exposure information can come
428 into play during hazard identification and the establishment of dose-response relationship. This
429 explained why some DBs (4.1%) were classified as both toxicological and exposure DBs. Another
430 definition regarding toxicological and exposure DBs could have resulted in another DB classification.
431 TEDI provides access to 6 descriptors for toxicological DBs and to 12 descriptors for exposure DBs.
432 The level of description of each DB reported in TEDI is heterogeneous and remains fairly general due
433 to the broad scope of the literature search, which lead to a high number of data sources that differed in
434 nature and content. Indeed, information reported in the records screened were found to be highly diverse
435 and inconsistent across studies resulting in challenges summarizing relevant information from the
436 reviewed records. Most of the time, many interesting and important information were not available, such
437 as the database availability (e.g., public), the terms of use, the full list of chemicals, the number of
438 variables (columns) and observations (rows) available or the database format (e.g., MS Excel sheet).
439 Some information, such as the full list of chemicals are not available in TEDI, because this information
440 was often unavailable and because it was not possible and feasible to list every chemical available in
441 each DB due to the high number of chemicals sometimes available (up to several millions). Therefore,
442 it is unfortunately not currently possible to pull up information on a specific chemical, with the exception
443 of PAHs.

444

445 *4.3 Perspectives from this work*

446 As pointed out before, this work is far from being exhaustive, both in terms of DBs listed and in terms
447 of information and meta-data retrieved. The inventory and its website remain open for new entries and
448 a more global coverage would certainly be preferable to support the broader objectives to support human
449 health risk assessment, which could help promote collaborative and harmonized research in this area.
450 A possible evolution of TEDI would be to integrate new core descriptors and meta-data from each DB,
451 in particular the creator, the owner, the contact person, the data availability, the full list of chemicals,
452 the coding system(s) used, the time period considered (e.g., data from 1995 to 2010) and the status of

453 the DB (e.g., still maintained). These descriptors were initially considered but were most of the time
454 missing from the records screened, and were therefore not collected and not recorded. The proof of
455 concept showed that this was possible, at least for a specific type of DB, for a specific family of
456 chemicals (PAHs) and for one of the missing descriptors. Another possible evolution of TEDI could be
457 to focus on specific chemical families and to perform the same work than the one that was done for
458 PAHs.

459 To collect the missing aforementioned meta-data and descriptors, another perspective from this work
460 could be to send a questionnaire to each contact person of the identified DBs. Such initiative, which was
461 beyond the scope of this work, would allow for the correction of potential mistakes and bias currently
462 present in TEDI. It would also allow for the improvement of the level of description of each DB reported
463 on TEDI.

464

465 *4.4 Comparison with other international efforts to compile databases*

466 To the best of the author's knowledge, to date, there is no review providing a comprehensive compilation
467 of existing toxicological and exposure DBs, consolidated in a single resource that can support chemical
468 risk assessment. Previous reviews on the subject in the past 15 years have only focused either on DBs
469 that can assist computational, drug discovery and system toxicology (usually in the public domain) or
470 on exposure DBs regarding carcinogens in the workplace. In addition, with the exception of one study
471 (Pawar et al. 2019), existing reviews only listed a few DBs (no more than 41). TEDI, which listed 715
472 DBs focusing solely on chemicals, was complementary to other existing reviews and provided more
473 information about existing toxicological and exposure DBs and their characteristics related to the
474 chemical risk assessment. As a result, most DBs mentioned in TEDI (81%) were not listed in any other
475 reviews.

476 Regarding toxicological DBs, one study listed 15 toxicological (in vitro, in vivo toxicity and ontology)
477 public DBs that can be used for assisting computational toxicology (Judson 2010). Another study
478 provided a comprehensive inventory of available data resources to support computational toxicology,
479 which is complementary to TEDI (Pawar et al. 2019). In their work, they listed more than 900 chemistry,
480 drug discovery, ADME, toxicological, clinical trials, biological, omics, protein-protein interactions,

481 patents, pathways, animal alternative methods and nanomaterials DBs. However, 88% of the DBs listed
482 in TEDI are not listed in their review. Another study reported 18 DBs on carcinogenicity and
483 mutagenicity that are available in the public domain (Benigni et al. 2013). A review from 2014 described
484 development of various bioinformatics tools that may be applied in the field of biodegradation (Arora
485 and Bae 2014). This review also provided the list of 32 chemical DBs that enable classification
486 identification and risk assessment of chemicals or that describe their environmental properties, toxicity
487 and distribution. Another study published in 2014 described 14 toxicogenomics and toxicology DBs that
488 can address question within systems toxicology (Fostel et al. 2014). A more recent study reported a list
489 of 41 publicly available DBs that provides information on hazards and risks of industrial chemicals, and
490 assessed the quality of each DB described (Bond and Garny 2019). An inventory of around 100 in silico
491 resources (DBs, software, models and applications) that can be used for physiologically-based kinetic
492 modeling and to predict physicochemical properties and chemical similarity was also published that year
493 (Madden et al. 2019).

494 Regarding exposure DBs, a report from the European Risk Observatory provided the list of 26
495 carcinogen exposure DBs and evaluated these existing sources of information, identified major
496 knowledge gaps and described some new approaches needed to assess and prevent occupational cancer
497 risks (Lißner et al. 2014). A recent overview reported an online inventory on exposure assessment tools
498 (<https://occupationalexposuretools.net>) (Peters et al. 2021). Meta-data on 36 general population JEMs,
499 11 exposure DBs, and 29 occupational coding systems from more than 10 countries were available. This
500 inventory was created within the Network on the Coordination and Harmonization of European
501 Occupational Cohorts (OMEGA-NET) framework, which aims to enable optimization of the use of
502 industrial and general population cohorts across Europe to advance etiological research. This inventory
503 reported less exposure DBs than TEDI, however, information recorded were more descriptive and
504 accurate as they were obtained directly from DB owners.

505 Regarding health risk assessment tools, one study introduced and compared seven air pollution health
506 risk assessment tools (Hassan Bhat et al. 2021). Other authors conducted a systematic review of risk
507 assessment tools for contaminated sites (Mahammedi et al. 2020). For each of the 31 identified tools the
508 relevant risk assessment stages, harm type, hazard category, receptor type and pathways were reported.

509 There are other complementary resources worth mentioning that can provide with apposite and useful
510 information. For instance, the *Nucleic Acids Research* journal has an online molecular biology DB
511 collection that amounts to 1,737 DBs that falls into 15 categories such as metabolic and signaling
512 pathways or microarray data and other gene expression DBs (Rigden and Fernandez 2018). The Global
513 Burden of Disease (GBD), which is an international consortium, provides a tool to quantify health loss
514 from more than 350 diseases, injuries, and risk factors, in 195 countries, by age and sex, from 1990 to
515 the present. The GBD project collected, generated and analyzed over 1 billion data points (Shaffer et al.
516 2019). Finally, TEDI could also be expanded to include open exposure DBs from citizen science
517 projects, which have not been found with the search strategy used in this work.

518

519 **5. Conclusions**

520 Toxicological and exposure DBs play an important role in determining the relationships between
521 exposures and diseases. Access to this type of information is clearly needed and although numerous DBs
522 have been compiled, they are often extremely difficult to find and navigate. This paper presented a
523 searchable web-based inventory (TEDI) of 715 toxicological and exposure DBs that might benefit
524 human health risk assessment resulting from the chemical exposome. Each DB had its own strengths
525 and limitations, depending on the problem formulation, purpose of the assessment, regulatory context,
526 relevance of the exposure, population, and chemical(s) considered in relation to the outcome(s) of
527 interest. The available information regarding each DB was often very limited. In addition, a dearth of
528 information remains, in particular for developing countries, for some life stages of an individual (e.g.,
529 pregnancy, childhood, late adulthood), mixtures of compounds, transformation products and new
530 substances/emerging risks. Only a small fraction of chemicals, their life cycle and the chemical
531 exposome were covered. Many challenges and data gaps remain, as outlined above. New studies to
532 bridge the existing gaps are therefore needed.

533 TEDI is of value to the broader community and could support human health risk assessment to chemicals
534 in various contexts. TEDI can help researchers, risk managers, risks assessors, practitioners, decision
535 makers, stakeholders, sanitary agencies... in finding the available resources useful for assessing health
536 risk. This is particularly true in the context of the launch of the European PARC (Partnership for the

537 Assessment of Risks from Chemicals) project, which aims in particular to promote harmonization of
538 data as well as tools and methods for conducting chemical risk assessment (Zare Jeddi et al. 2021). TEDI
539 remains open for further additions, to enlarge its coverage and include newly developed DBs.

540

541 **Acknowledgments**

542 The author would like to thank Simon Pervier and Tiansi Laranjeira for their help to deploy TEDI online
543 as well as TIMC lab (UMR 5525 - CNRS UGA) and Université Grenoble Alpes for hosting TEDI.

544

545 **Statements & Declarations**

546

547 *Funding*

548 This work was partly supported by the French National Research Agency, within the framework of the
549 “Investissements d’avenir” program [ANR-10-AIRT-0005, 2012; ANR-15-IDEX-0002, 2016]. The
550 funding sources had no role in the study design; in the collection, analysis, and interpretation of data; in
551 the writing of the report; or in the decision to submit the paper for publication.

552

553 *Declarations of interest*

554 None.

555

556 *Author Contributions - CRediT author statement*

557 **Pascal Petit:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation,
558 Data Curation, Writing – Original Draft, Writing - Review & Editing, Visualization, Supervision,
559 Project administration.

560 The work reported in the paper has been performed by the author, unless clearly specified in the text.

561

562 *Data availability statement*

563 The data that support the findings of this study are openly available on TEDI at [https://exporisk-](https://exporisk-timc.imag.fr/TEDI/)
564 [timc.imag.fr/TEDI/](https://exporisk-timc.imag.fr/TEDI/). Further information is available from the corresponding author upon request.

565 **References**

- 566 Andreu-Perez, J., Poon, C.C., Merrifield, R.D., Wong, S.T., Yang, G.Z., 2015. Big data for health. IEEE.
567 J. Biomed. Health. Inform. 19(4), 1193-1208. <https://doi.org/10.1109/JBHI.2015.2450362>.
- 568 Arora, P.K., Bae, H., 2014. Integration of bioinformatics to biodegradation. Biol. Proced. Online. 16, 8.
569 <https://doi.org/10.1186/1480-9222-16-8>.
- 570 Atwood, S.T., Lunn, R.M., Garner, S.C., Jahnke, G.D., 2019. New Perspectives for Cancer Hazard
571 Evaluation by the Report on Carcinogens: A Case Study Using Read-Across Methods in the Evaluation
572 of Haloacetic Acids Found as Water Disinfection By-Products. Environ. Health. Perspect. 127(12),
573 125003. <https://doi.org/10.1289/ehp5672>.
- 574 Beck, N.B., Becker, R.A., Erraguntla, N., Farland, W.H., Grant, R.L., Gray, G., et al. 2016. Approaches
575 for describing and communicating overall uncertainty in toxicity characterizations: U.S. Environmental
576 Protection Agency's Integrated Risk Information System (IRIS) as a case study. Environ. Int. 89-90,
577 110-128. <https://doi.org/10.1016/j.envint.2015.12.031>.
- 578 Benigni, R., Battistelli, C.L., Bossa, C., Tcheremenskaia, O., Crettaz, P., 2013. New perspectives in
579 toxicological information management, and the role of ISSTOX databases in assessing chemical
580 mutagenicity and carcinogenicity. Mutagenesis. 28(4), 401-409. <https://doi.org/10.1093/mutage/get016>.
- 581 Bond, G.G., Garny, V., 2019. Inventory and evaluation of publicly available sources of information on
582 hazards and risks of industrial chemicals. Toxicol. Ind. Health. 35(11-12), 738-751.
583 <https://doi.org/10.1177/0748233719893198>.
- 584 Bopp, S.K., Barouki, R., Brack, W., Dalla Costa, S., Dorne, J.C.M., Drakvik, P.E., et al. 2018. Current
585 EU research activities on combined exposure to multiple chemicals. Environ. Int. 120, 544-562.
586 <https://doi.org/10.1016/j.envint.2018.07.037>.
- 587 Chang, W., Cheng, J., Allaire, J., Xie, Y., McPherson, J., 2018. Shiny: Web Application Framework for
588 R. R Package, version 1.4.0.

589 Dennis, K.K., Marder, E., Balshaw, D.M., Cui, Y., Lynes, M.A., Patti, G.J., et al. 2017. Biomonitoring
590 in the Era of the Exposome. *Environ. Health. Perspect.* 125(4), 502-510.
591 <https://doi.org/10.1289/EHP474>.

592 ECHA. 2015. REACH 2018. ECHA: Helsinki, 2015. [last accessed: 14 May 2022].

593 Fostel, J., van Someren, E., Pronk, T., Pennings, J., Schmeits, P., Shao, J., et al. 2014. Chapter 6.2 -
594 Toxicogenomics and Systems Toxicology Databases and Resources: Chemical Effects in Biological
595 Systems (CEBS) and Data Integration by Applying Models on Design and Safety (DIAMONDS).
596 Toxicogenomics-Based. *Cellular. Models.* 275-290. [https://doi.org/10.1016/B978-0-12-397862-
597 2.00023-1](https://doi.org/10.1016/B978-0-12-397862-2.00023-1).

598 Frelinger, J.A., 2015. Big Data, Big Opportunities, and Big Challenges. *J. Investig. Dermatol. Symp.*
599 *Proc.* 17(2), 33-35. <https://doi.org/10.1038/jidsymp.2015.38>.

600 Gaudet, P., Bairoch, A., Field, D., Sansone, S.A., Taylor, C., Attwood, T.K., et al. 2011. Towards
601 BioDBcore: a community-defined information specification for biological databases. *Database.*
602 (Oxford). 2011, baq027. <https://doi.org/10.1093/database/baq027>.

603 Hassan Bhat, T., Jiawen, G., Farzaneh, H., 2021. Air Pollution Health Risk Assessment (AP-HRA),
604 Principles and Applications. *Int. J. Environ. Res. Public. Health.* 18(4), 1935.
605 <https://doi.org/10.3390/ijerph18041935>.

606 Judson, R., 2010. Public databases supporting computational toxicology. *J. Toxicol. Environ. Health.*
607 *B. Crit. Rev.* 13(2-4), 218-231. <https://doi.org/10.1080/10937404.2010.483937>.

608 Kim, K.N., Hong, Y.C., 2017. The exposome and the future of epidemiology: a vision and prospect.
609 *Environ. Health. Toxicol.* 32, e2017009. <https://doi.org/10.5620/eht.e2017009>.

610 Laamanen, I., Verbeek, J., Franco, G., Lehtola, M., Luotamo, M., 2008. Finding toxicological
611 information: An approach for occupational health professionals. *J. Occup. Med. Toxicol.* 3, 18.
612 <https://doi.org/10.1186/1745-6673-3-18>.

613 Lißner, L., Kuhl, K., Kauppinen, T., Uuksulainen, S., 2014. Exposure to carcinogens and work-related
614 cancer: A review of assessment methods. European Risk Observatory Report. European Risk
615 Observatory, Luxembourg.

616 Madden, J.C., Pawar, G., Cronin, M.T.D., Webb, S., Tan, Y.M., Paini, A., 2019. In silico resources to
617 assist in the development and evaluation of physiologically-based kinetic models. *Comput. Toxicol.* 11,
618 33-49. <https://doi.org/10.1016/j.comtox.2019.03.001>.

619 Mahammed, C., Mahdjoubi, L., Booth, C.A., Akram, H., Butt, T.E., 2020. A systematic review of risk
620 assessment tools for contaminated sites - Current perspectives and future prospects. *Environ. Res.* 191,
621 110180. <https://doi.org/10.1016/j.envres.2020.110180>.

622 Mallah, M.A., Changxing, L., Mallah, M.A., Noreen, S., Liu, Y., et al. 2022. Polycyclic aromatic
623 hydrocarbon and its effects on human health: An overview. *Chemosphere.* 296, 133948.
624 <https://doi.org/10.1016/j.chemosphere.2022.133948>.

625 Pawar, G., Madden, J.C., Ebbrell, D., Firman, J.W., Cronin, M.T.D., 2019. In Silico Toxicology Data
626 Resources to Support Read-Across and (Q)SAR. *Front. Pharmacol.* 10, 561.
627 <https://doi.org/10.3389/fphar.2019.00561>.

628 Persad, A.S., Bobst, S., 2020. Chapter 48 - Risk assessment. *Information Resources in Toxicology (Fifth*
629 *Edition). Volume 1: Background, Resources, and Tools.* 509-512. [https://doi.org/10.1016/B978-0-12-](https://doi.org/10.1016/B978-0-12-813724-6.00048-7)
630 [813724-6.00048-7](https://doi.org/10.1016/B978-0-12-813724-6.00048-7).

631 Peters, S., Vienneau, D., Sampri, A., Turner, M.C., Castano-Vinyals, G., Bugge, M., et al. 2021.
632 Occupational Exposure Assessment Tools in Europe: A Comprehensive Inventory Overview. *Ann.*
633 *Work. Expo. Health.* wxab110. <https://doi.org/10.1093/annweh/wxab110>.

634 Rigden, D.J., Fernandez, X.M., 2018. The 2018 Nucleic Acids Research database issue and the online
635 molecular biology database collection. *Nucleic. Acids. Res.* 46(D1), D1-D7.
636 <https://doi.org/10.1093/nar/gkx1235>.

637 Shaffer, M., Sellers, S.P., Baker, M.G., de Buen Kalman, R., Frostad, J., Suter, M.K., et al. 2019.
638 Improving and Expanding Estimates of the Global Burden of Disease Due to Environmental Health Risk
639 Factors. *Environ. Health. Perspect.* 127(10), 105001. <https://doi.org/10.1289/EHP5496>.

640 Siroux, V., Agier, L., Slama, R., 2016. The exposome concept: a challenge and a potential driver for
641 environmental health research. *Eur. Respir. Rev.* 25(140), 124-129.
642 <https://doi.org/10.1183/16000617.0034-2016>.

643 Sowe, S.K., Zettsu, K., 2014. Curating Big Data Made Simple: Perspectives from Scientific
644 Communities. *Big. Data.* 2(1), 23-33. <https://doi.org/10.1089/big.2013.0046>.

645 Steckling, N., Gotti, A., Bose-O'Reilly, S., Chapizanis, D., Costopoulou, D., De Vocht, F., et al. 2018.
646 Biomarkers of exposure in environment-wide association studies - Opportunities to decode the
647 exposome using human biomonitoring data. *Environ. Res.* 164, 597-624.
648 <https://doi.org/10.1016/j.envres.2018.02.041>.

649 Vineis, P., Chadeau-Hyam, M., Gmuender, H., Gulliver, J., Herceg, Z., Kleinjans, J., et al. 2017. The
650 exposome in practice: Design of the EXPOsOMICS project. *Int. J. Hyg. Environ. Health.* 220(2 Pt A),
651 142-151. <https://doi.org/10.1016/j.ijheh.2016.08.001>.

652 Wignall, J.A., Muratov, E., Sedykh, A., Guyton, K.Z., Tropsha, A., Rusyn, I., et al. 2018. Conditional
653 Toxicity Value (CTV) Predictor: An In Silico Approach for Generating Quantitative Risk Estimates for
654 Chemicals. *Environ. Health. Perspect.* 126(5), 057008. <https://doi.org/10.1289/ehp2998>.

655 Wild, C.P., 2005. Complementing the genome with an "exposome": the outstanding challenge of
656 environmental exposure measurement in molecular epidemiology. *Cancer. Epidemiol. Biomarkers.*
657 *Prev.* 14(8), 1847-1850. <https://doi.org/10.1158/1055-9965.EPI-05-0456>.

658 Wild, C.P., 2012. The exposome: from concept to utility. *Int. J. Epidemiol.* 41(1), 24-32.
659 <https://doi.org/10.1093/ije/dyr236>.

660 Yang, L., Zhang, H., Zhang, X., Xing, W., Wang, Y., et al. 2021. Exposure to Atmospheric Particulate
661 Matter-Bound Polycyclic Aromatic Hydrocarbons and Their Health Effects: A Review. *Int. J. Environ.*
662 *Res. Public. Health.* 18(4), 2177. <https://doi.org/10.3390/ijerph18042177>.

663 Zare Jeddi, M., Virgolino, A., Frantke, P., Hopf, N.B., Galea, K.S., et al. 2021. A human biomonitoring
664 (HBM) Global Registry Framework: Further advancement of HBM research following the FAIR
665 principles. *Int. J. Hyg. Environ. Health.* 238, 113826. <https://doi.org/10.1016/j.ijheh.2021.113826>.

666 Zheng, H., Porebski, P.J., Grabowski, M., Cooper, D.R., Minor, W., 2017. Databases, Repositories, and
667 Other Data Resources in Structural Biology. *Methods. Mol. Biol.* 1607, 643-665.
668 https://doi.org/10.1007/978-1-4939-7000-1_27.

PROOF - ACCEPTED

669 **Figure Captions**

670

671 **Fig.1** Flow chart depicting the literature search and the evaluation process for finding relevant records

672

673 **Fig.2** Number of records per 5-year period

674 Regarding the bar on the right side of the graph (2020-2022), the number of records corresponds only
675 to the total number of records from 2020, 2021 and January 2022

676

677 **Fig.3** Number of databases by continent and database type

678

679 **Fig.4** Number of toxicological databases by data type

680

681 **Fig.5** Number of exposure databases by data type

682

683 **Fig.6** Screenshot of the TEDI app

684 Orange fonts that are underlined refer to hyperlinks providing either the definitions and descriptions of
685 terms or direct access to the mentioned references. Buttons allow user actions such as copying the table
686 to clipboard, saving the table as CSV/XLS/PDF, printing the table and selecting columns to visualize.
687 It is possible to make global (entire table) and column specific queries. Beware, when opening TEDI on
688 a web browser there are several horizontal sliders that will be available, with one on the web page itself,
689 one inside the toxicological DB table (at the bottom) and one inside the exposure DB table (at the
690 bottom). To access all of the available information reported in these two tables, the user has to use the
691 different horizontal sliders (moving it on the right or left) or to download the table.