

Reporting Chemical Data in the Environmental Sciences

Sivani Baskaran,* Parviel Chirsir, Shira Joudan, Raoul Wolf, Evan E. Bolton, Paul A. Thiessen, and Emma L. Schymanski



Cite This: <https://doi.org/10.1021/acsenvironau.5c00034>



Read Online

ACCESS |



Metrics & More



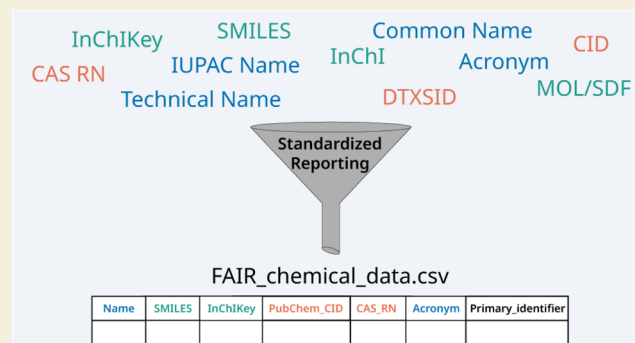
Article Recommendations



Supporting Information

ABSTRACT: Environmental sciences, including environmental chemistry and toxicology, are highly interdisciplinary fields that integrate researchers with various backgrounds and expertise. This interdisciplinary aspect is critical to addressing issues of chemical pollution, environmental sustainability, and health. However, a standardized method for reporting chemical data is needed to address these issues effectively. This becomes increasingly important as both the number of chemical structures and our reliance on and use of computational analysis and cheminformatics tools grow. This paper provides background, examples, and recommendations on how to report chemical data in a findable, accessible, interoperable, and reproducible (FAIR) manner within environmental science disciplines. Ultimately, the goal is to broaden the scope and applicability of environmental research to help the entire community tackle the issues of chemical pollution and sustainability in a comprehensive manner.

KEYWORDS: chemical information, data, FAIR, open access, accessibility, reproducibility, environmental science



to help the entire community tackle the issues of chemical pollution and sustainability in a comprehensive manner.

1. INTRODUCTION

Environmental science is an interdisciplinary and collaborative field that brings together expertise and knowledge from a wide range of disciplines, including analytical, environmental, organic, and physical chemistry, toxicology, and biology, plus atmospheric, health, social, and data sciences. One key aim of this scientific intersection is to understand the behavior of chemicals in the environment and their interactions with humans and other organisms. However, this objective is sometimes impeded by data accessibility issues. Although the largest chemical databases contain hundreds of millions of chemicals, they are known to be incomplete, as many analytical signals detected in the environment cannot be matched to any chemicals in these databases.^{1–3} While notable progress has been made to improve the accessibility of scientific knowledge via the rise of open access publishing options, less attention has been given to making the chemical data associated with environmental studies findable, accessible, interoperable, and reusable (FAIR).⁴ Building on recent recommendations on FAIR chemical data,⁵ this perspective aims to provide background, examples, and finally, guiding principles for reporting chemical data in environmental sciences.

A chemical's identity is crucial information in the chemical and environmental sciences. This information is used to distinguish chemicals listed in global chemical registries, determine their properties, apply for patents, describe use applications, and so on. In 2003, Glüge et al.⁶ reviewed 8590 substances from the European Chemicals Agency (ECHA)

database and cross-referenced them with multiple chemical databases. They identified several issues in the reported data due to inconsistent identifiers and missing stereochemical information.⁶ An earlier 2008⁷ analysis of multiple databases found errors in the structure of chemicals reported relative to other chemical identifiers.⁷ As informatics approaches become more accessible, errors such as these will be increasingly exacerbated due to “cross-pollination” between different databases as data are copied and shared across platforms. However, errors are not necessarily always errors per se, but can arise due to discrepancies in how the data was collected and processed. Whether the data are obtained from the field, lab, or through computational methodologies, the data presented in scientific publications holds immense value to other researchers. Publications are most useful to the chemical community when the chemicals described are well-defined using standard and precise identifiers.

For the purposes of this article, it is important to first define the concepts “chemical compound” and “chemical substance”. This is illustrated with examples in Figure 1. The International

Received: February 26, 2025

Revised: July 9, 2025

Accepted: July 11, 2025

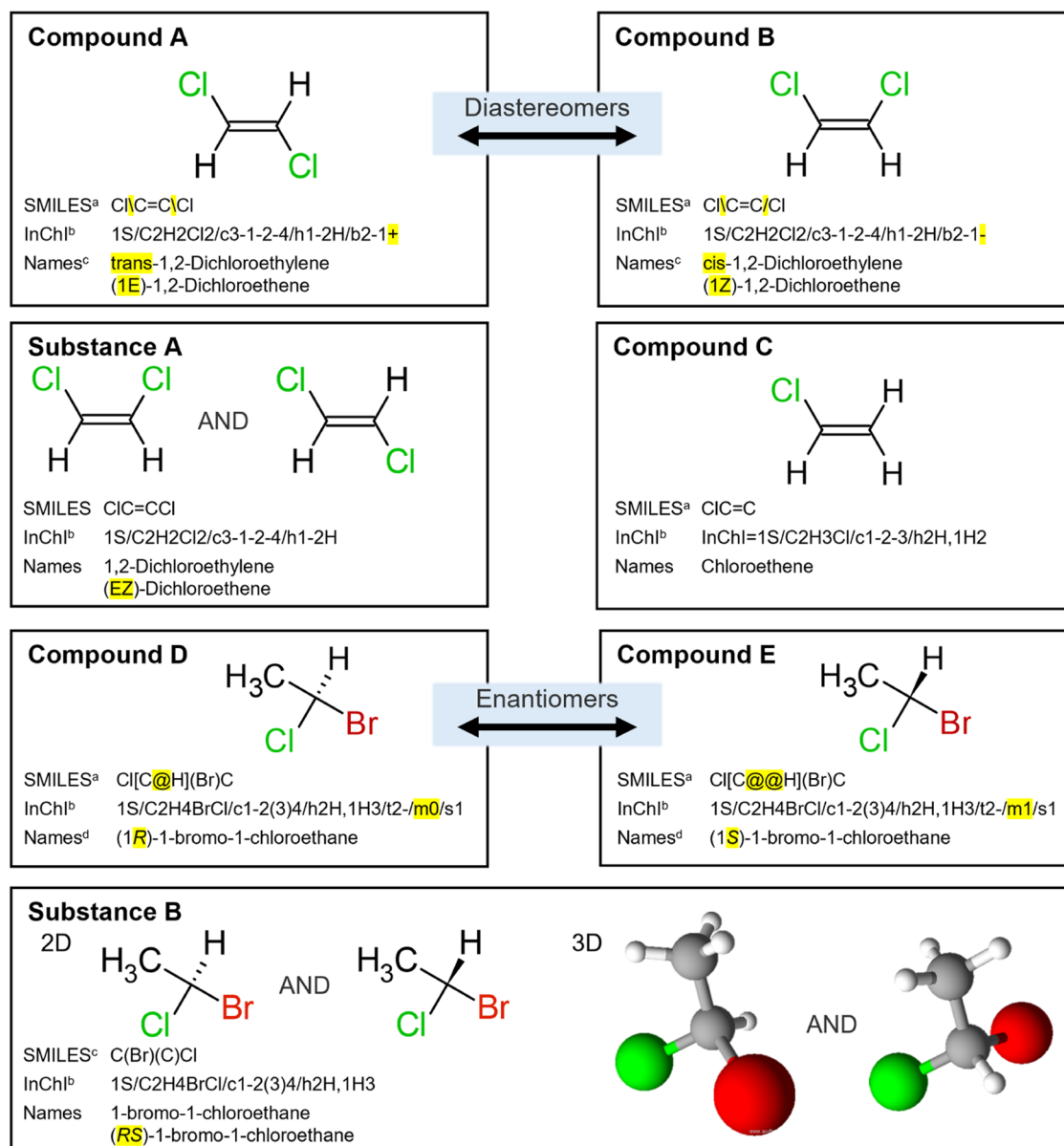


Figure 1. Examples distinguishing the differences between compounds and substances as defined in this work. SMILES, InChIs, and some names are provided for each chemical. Distinct differences in these identifiers between similar substances are highlighted in yellow. Compounds A and B are diastereomers, stereoisomers that are not mirror images of each other. Compounds D and E are enantiomers, stereoisomers that are mirror images of each other. Substance B can also be depicted as a single 2D structure using a single bond instead of hashed or wedged bonds. Identifiers were generated manually or taken from ^aACD/Labs ChemSketch, ¹⁰ ^bOpen Babel, ¹¹ ^cCAS Common Chemistry, ^{12–14} ^dPubChem. ^{15,16} Structures were generated using ACD/Labs ChemSketch. ¹⁰

Union of Pure and Applied Chemistry (IUPAC) definition of a molecule is a neutral entity, made up of at least two atoms.⁸ In this article, a chemical compound is composed of a single entity, made up of at least two atoms, which collectively has either a neutral or charged state, and has specific physical–chemical properties (see Figure 1). IUPAC's definition of a chemical substance is a unit of matter with a constant composition that can be characterized by the molecules, formula units, and atoms it is made up of.⁹ By these definitions distinct structural and stereoisomers are considered different chemical compounds (see Figure 1, compounds A–E); while a nonstereochemistry specific entity with stereocenters (see Figure 1, substance A, also Figure SI 1), mixtures, Unknown or Variable composition, Complex reaction products and/or Biological materials (UVCBs), or polymers are considered to

be substances. Thus, with these definitions, (1) both substances and compounds have distinct physical–chemical properties, (2) chemical compounds can also be chemical substances, but (3) not all substances are chemical compounds. Importantly, the definitions of chemical compounds and substances used by published databases and systems can differ from the definitions presented here and from each other.

This article begins by reviewing commonly used chemical identifiers in environmental sciences: chemical names, database identifiers such as CAS Registry Numbers (CAS RN), and single-line text-based structural chemical representations such as Simplified Molecular Input Line Entry System (SMILES). The pros and cons of the various identifiers are discussed, along with how and where to obtain this information and the

Table 1. Description of Some Common Chemical Identifiers

chemical identifier	description
CAS RN	CAS Registry Number assigned by the American Chemical Society; CAS RNs have a standardized three-block format containing 2 to 7 digits, 2 digits, and 1 digit, respectively; the final digit is used as a check digit that can be used to verify valid CAS RNs ¹⁷
CID	Compound IDentifier used in PubChem
common name	colloquially used term/acronym to refer to the chemical; often does not describe the structure
DTSID	DSSTox Substance IDentifier used in the Environmental Protection Agency's CompTox Dashboard
InChI	International Chemical Identifier—a standardized text-based identifier for chemical structures that encodes molecular information
InChIKey	a three-block hash key of the InChI (skeleton-stereochemistry-charge) with 27 characters (14-10-1); the second block also includes isotopic information, the version used, and whether the original InChI is standardized ¹⁸
IUPAC Name	a name derived using International Union of Pure and Applied Chemistry (IUPAC) naming conventions
MOL/SD file	an MDL Molfile (MOL) contains information about the atoms, bonds, connectivity and coordinates of a single molecule. Multiple entries can be stored in a structure data (SD) file (SDF), separated by “\$\$\$”
SMILES	Simplified Molecular Input Line Entry System—a line notation for describing the structure of chemical species using short ASCII strings
synonyms	other names, variations, spellings, simplifications, or abbreviations of the other names
technical name	name used and registered by companies in patents or products

role that names, structures, and numerical identifiers play in the reporting of data. Finally, recommendations are provided on what information should be given when reporting chemical data and associated results in the environmental sciences, based on the FAIR chemical guidance.⁵

2. CHEMICAL DATA

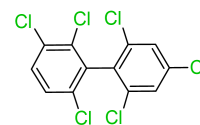
Regardless of the type of research undertaken, the reported data are often underpinned by the chemical information on the substances studied, i.e., information used to identify the chemical. Typical chemical identifiers reported include a common name (e.g., DDT) or CAS RNs (e.g., 50-29-3), but these reporting conventions can be prone to errors and challenges. Table 1 includes an overview of various ways to exchange chemical information.

2.1. Chemical Names

Variations in naming conventions can create challenges in matching and correctly identifying chemicals. Starting with DDT, domain-specific knowledge would lead to the assumption that the acronym is referring to *p,p'*-DDT or 4,4'-DDT, but without knowledge of common or discipline-specific naming practices, it could also be misinterpreted as *o,p'*-DDT or 2,4'-DDT (Table SI 1, rows 8 and 9). Next, polychlorinated biphenyls (PCBs) are a well-established group of chemicals, with many different naming systems used to describe single congeners. A congener is used to describe chemicals with similar structure and/or properties. For PCBs, there are 209 unique congeners.¹⁹ Box 1 includes some examples of names used for one hexachlorobiphenyl congener, PCB 150, revealing issues of order, separators (“-” vs “ ”), and various naming conventions. Although none of these names are incorrect, an exact text match will work only if the names in the database and search query match perfectly.

Box 1: A non-exhaustive list of “common” names for PCB 150.

- PCB 150
- PCB-150
- 2,2',3',4,6,6'-Hexachlorobiphenyl
- 2,2',3,4',6,6'-Hexachlorobiphenyl
- 2,4,6,2',3',6'-Hexachlorobiphenyl
- 2,3,6,2',4',6'-Hexachlorobiphenyl
- 2,2',3',4,6,6'-Hexachloro-1,1-biphenyl
- 2,2',3,4',6,6'-Hexachloro-1,1-biphenyl
- 1,1'-Biphenyl, 2,2',3,4',6,6'-hexachloro-
- 1,1'-Biphenyl, 2,2',3',4,6,6'-hexachloro-
- 1,2,4-trichloro-3-(2,4,6-trichlorophenyl)benzene



The stereochemistry specification is extremely important in some environmental chemistry and toxicology applications, where 3D conformations of chemicals are relevant. Different stereoisomers can have different physical–chemical properties, and thus, their environmental and biological implications can differ. A study testing the toxicity of 6PPD-quinone found that the *S* enantiomer was more toxic than the racemic mixture or the *R* enantiomer.²⁰ Hence, the naming and identification of specific isomers have chemical, environmental, and regulatory importance. More details regarding differences in isomerization in chemical structures are included in Section SI 2 and Brunning.²¹

Another example is hexabromocyclododecane, also commonly referred to as HBCD or HBCDD. The name hexabromocyclododecane indicates a cyclic chain of 12 carbons with 6 bromine substitutions. Under the Stockholm Convention, HBCD is listed as a persistent organic pollutant and supporting documents primarily refer to 6 of the most common isomers of 1,2,5,6,9,10-hexabromocyclododecane,^{22,23} which has 16 different stereoisomers (see Section SI 2). When multiple stereoisomers are possible, each isomer is assigned a Greek letter; enantiomer pairs are assigned the same letter and sometimes differentiated using + or – based on their optical activity.²⁴ While useful as a quick reference, these letters pose many challenges and can produce errors when working with computational software that may not recognize these characters. It is also important to note that the locations of the bromine substitutions are not specified in the original common name “hexabromocyclododecane”; thus, outside the regulatory context of the Stockholm Convention, HBCD can represent 77 possible structural isomers, assuming 6 single

bromine substitutions. This count does not include possible stereoisomers for each of the structural isomers.

In some cases, there are preferred English spellings for chemicals depending on the region, for example, endosulphan²⁵ vs endosulfan.²⁶ The use of *f* over *ph* is now recommended by IUPAC as the proper name for sulfur compounds.²⁷ The distinction between phosphorus (noun) and phosphorous (adjective) in English has also led to much confusion and has been used interchangeably in chemical naming systems.^{28–30} However, not all publications and documents are in English and additional identifiers are needed to complement a chemical name. The problems with using names to identify chemicals are exacerbated for large, novel chemicals, making matching and identifying papers referencing the same chemical increasingly difficult.

In chemistry, there are standardized naming processes defined by IUPAC, which are described extensively in the Blue Book^{31,32} for organic chemistry, the Red Book for inorganic chemistry,³³ and the Purple Book³⁴ for polymers. Using these guidelines, only one IUPAC name should exist per substance. IUPAC has also consolidated shorter guidelines that are available online;^{35–40} however, the use of such guidelines can be cumbersome and time-consuming. There are tools for generating the IUPAC names using the structure of a compound (e.g., ChemDoodle⁴¹) and parsers that can provide a structure of a chemical given the IUPAC name (e.g., Open Parser for Systematic IUPAC nomenclature^{42,43}). Despite IUPAC's best intentions, there are issues in reading and writing IUPAC names. Taking *p,p'*-DDT again, the IUPAC name for this molecule is reported to be 1,2,4-trichloro-3-(2,4,6-trichlorophenyl)benzene by PubChem⁴⁴ and CAC-TUS⁴⁵ and 1,1'-(2,2,2-trichloroethane-1,1-diyl)bis(4-chlorobenzene) by CompTox⁴⁶ and ChemSpider.⁴⁷ Complicated structures have further issues, such as the superscripts in the IUPAC name 1,3,5,7-tetrazatricyclo[3.3.1.1^{3,7}]decane for methenamine.^{48–50} While these standardized names can help determine the exact molecular structure of a chemical, they are not easy to generate or interpret for large molecules and, as demonstrated, are not completely interoperable.

2.2. Database Identifiers

Different databases and registries have introduced many numerical identification systems to index their entries and reduce these types of errors, including, CAS RNs (American Chemical Society),^{17,51} Compound and Substance IDentifiers (CIDs and SIDs respectively, PubChem, United States National Institute of Health),^{52,53} DTXSIDs and DTXCIDs (CompTox Chemistry Dashboard, United States Environmental Protection Agency),⁵⁴ ChEMBL IDs (ChEMBL Database, European Molecular Biology Laboratory),⁵⁵ ChEBI IDs (Chemical Entities of Biological Interest Database, European Molecular Biology Laboratory),⁵⁶ and ChemSpider IDs (CSID, Royal Society of Chemistry).⁵⁷ Because these identifiers are specific to a database, they typically map to a single entry within the database; however, only entries that are already within these respective resources are assigned an identifier. In many cases, the databases cross-reference each other, and almost all attempt to include alternative identifiers in their list of chemical identifiers. Some identifiers, such as CAS RNs, are not readily available unless they have been previously published or made available via CAS Common Chemistry⁵⁸ and can only be authoritatively identified using proprietary tools such as SciFinder.⁵⁹ Additionally, database

identifiers may be deleted, replaced, or deprecated over time. As a result, a chemical structure may have two or more identifiers. For example, *p,p'*-DDT has a current (50-29-3) and deleted (1081843-15-3) CAS RN that can be found in different databases and publications. While one database may consider one name or structure to be unique, another may split the substance into multiple entries to account for stereochemistry, isomerization, or tautomerization. This can often depend on the curation and standardization algorithms used in the databases. This is discussed further in Section 2.4.

In other instances, the name or chemical referenced in a publication is not included in any chemical databases, or there may not be a numerical identifier assigned to a particular structure or available on an open-access platform. Property data for 6-OH BDE-68, a polybrominated diphenyl ether congener (PBDE 68) with an additional hydroxy group, was published by Liu et al.,⁶⁰ while this chemical (at the time of writing) appears only on PubChem (CID: 101542767) the name used by Liu et al. is not used or included under PubChem's synonyms list. Shorthand names such as 6-OH BDE-68 must be identified and manually added to any database and registries, which can take time or may never occur unless specifically requested.

It is not uncommon for environmental scientists to work with ionizable substances. One excellent example is perfluorooctanesulfonate, which is a charged ionic form of perfluorooctanesulfonic acid (PFOS). With a charge of -1 , perfluorooctanesulfonate exists as a salt with lithium and potassium as common counterions (Figure SI 5). Each salt is often assigned a unique identifier within databases and has distinct structural identifiers; however, these entries do not account for all possible counterions. If the salt is purchased directly from a supplier, it can be useful to specify which salt is being used and specifically identify the analyte or ion of interest. The importance of the salt is relative to the methodology; for example, in toxicology, it is important to use the correct molecular weight when weighing a salt and converting it to concentration units of PFOS in a toxicology study, whereas in environmental water samples, PFOS exists in a dissociated state.

When working with spreadsheets, CAS RNs pose a particular problem: they can be misinterpreted as a date by Microsoft Excel. A good example of this is dichloromethane; its CAS RN 75-09-2 can, for example, be automatically converted to September 2, 1975. This error occurs whenever the first set of digits in the CAS RN contains 2 digits or contains 4 digits which can be read as a year (e.g., 75 or 1975) and the second set of digits contains two digits between 01 and 12. The genomics field experiences a similar problem with some gene symbols (e.g., MARCH1); this has since been addressed by updated nomenclature guidelines and renaming symbols that were affected during data handling.⁶¹ Analyses of genomic papers indicate that approximately 20% of gene names were erroneous in papers that contained supplementary Excel files⁶² and approximately 30% of all publications in the field contained these types of errors.⁶³ No comparable study examining CAS RN errors within environmental science or chemistry publications was found. In 2023, Microsoft Excel introduced an option to remove the automatic date format conversion process; however, preventing these errors relies on researchers working with this type of data to toggle on this functionality before using or opening files that contain CAS RNs and it is not guaranteed to work.⁶⁴ When adding CAS

RNs to a Microsoft Excel document, it is possible to avoid automatic conversion of the CAS RN to a date by introducing an apostrophe before the CAS RN (e.g., '75-09-2). However, this applies only to XLS/XLSX files and such additions will be lost if the file is saved as a CSV and reopened as an Excel document and conversely may be retained as part of the CAS RN when opened by other programs or software.

2.3. Structural Representations

There are multiple methods for describing chemical structures that provide varying levels of detail. Single-line text-based notations that are used to describe chemical compounds include SMILES, InChI, and InChIKeys.

SMILES is the most human-readable format; with a bit of practice, it is possible to generate and read the SMILES of simpler and smaller compounds without the use of computational tools.⁶⁵ SMILES can come in a variety of flavors, including standardized, canonical, QSAR-ready, MS-ready, universal, kekulized, etc. The terms are not always used consistently between toolkits and databases, and in the case of MS- and QSAR-ready SMILES, there can be different or reduced levels of specificity depending on the purpose and use case.^{66,67} By definition, SMILES can be created from multiple starting points, so the order of the atoms may change, and a single structure can have many equivalent and valid SMILES. In some forms, the specificity of the stereochemistry (e.g., relative positions or aromaticity) or dative bond pairs are excluded from the structure, an issue sometimes undetected by those unfamiliar with SMILES, which has led to much confusion in reporting structures (see Section SI 3 for additional information and examples).

Returning to the 1,2,5,6,9,10-HBCDD example from Section 2.1, the nonstereospecific SMILES should be associated only with 1,2,5,6,9,10-HBCDD without defined stereochemistry, but SMILES notation can be used to differentiate between all 16 isomers (see Table SI 1 and Figures SI 1 and SI 2). As a rule of thumb, it is good to check if the SMILES produces the same structure with the same level of specificity one would expect by using free chemical drawing software such as ACD/Labs ChemSketch,¹⁰ CDK Depict,⁶⁸ Smi2Depict,⁶⁹ or MolView.⁷⁰ Other widespread tools such as Marvin/Marvin-Sketch^{71,72} and ChemDraw⁷³ can also be used for the same purpose. SMILES is one of the most common input formats for QSAR models in environmental chemistry and toxicology (in part due to its ability to be included within a spreadsheet); however, there are some issues to keep in mind. For some models, different SMILES for the same chemical can produce different results (e.g., EPISuite⁷⁴), while others perform an internal standardization process (e.g., OPERA⁷⁵). Thus, it is better to have prepared all SMILES data in the same way before using them as input for any QSAR or predictive tool to ensure consistent results.

The International Chemical Identifier or InChI was developed by IUPAC.^{18,76} Unlike SMILES, InChI is calculated using a standardized algorithm, and thus one InChI exists per structure. This enables InChIs to be used as identifiers. However, InChIs are generally much longer than SMILES and contain several special characters (commas and semicolons) that can cause parsing issues. For large molecules, the length of InChIs can also lead to truncation errors if they exceed database limits.^{18,77} To alleviate these issues, the hashed form InChIKey was introduced to serve as a machine-readable identifier. An InChIKey can be generated from an InChI, but

an InChI can not be computed from an InChIKey.^{18,76} If an InChIKey is the only information available, a database or look-up table containing the InChIKey is required to retrieve the chemical's structure. As such, providing an InChIKey without its corresponding InChI or SMILES can be problematic, especially if one describes a novel structure.

InChIs have different layers containing structural information. The standard form, denoted with "1S/", is calculated with standard settings and thus always comparable (see Heller et al.¹⁸ for more details). The use of a nonstandard InChI, denoted with "1/", indicates that nonstandard (and therefore not always comparable) settings were used and are sometimes needed to capture advanced stereochemistry or tautomers. Each layer provides different information about the chemical structure, such as atom connectivity, stereochemistry, and charges. The standardized InChI produces a unique standard InChIKey, denoted by the "SA" at the end of the second block. Nonstandard InChIKeys are denoted by "NA" at the end of the second block. The standard InChI does not always differentiate different tautomers; in these cases, nonstandard InChIs can be used to define specific tautomerization. Nonstandard InChIs may also be necessary to define nonstereospecific structures. In the case of a structure with relative stereochemistry, the SMILES and standard InChI may reflect the structure of one of the enantiomers; the use of a nonstandard InChI and InChIKey allows for the specificity of relative-stereoisomerization (Table SI 1, see Yerin⁷⁸ for additional examples). Since standard InChIs are produced using fixed rules, SMILES generated from standardized InChIs may not always provide the most environmentally relevant structure. For example, atrazine has five tautomers (Table SI 2), which are all described by the same standardized InChI string. Thus, if only one single-line notation is used, SMILES is often preferred to define the structure with the necessary stereochemistry, while InChI and InChIKey can be used to verify and identify the SMILES.

The structure of a chemical can also be expressed visually with lines and atom symbols. Drawing the structure can help reduce the ambiguity of other chemical identifiers, and there are various ways a structure can be drawn and annotated (see Figure SI 3). For simple structures, general drawing tools can be used to create and share the structure as an image file. However, using chemistry-specific tools for drawing substances is generally much easier, reducing the potential for errors, increasing interoperability—as the shared file can be easily modified and saved in recognized exchange formats—and they are generally better equipped to handle large and more complex substances. These structures can be exported as an image file or shared using specific file types, such as SK2, SKC, CDX, or CML, which can only be read by specific tools, as discussed later. MOL (MOLfile) contains a connection table with information about the bonds and can include 2D or 3D coordinates of a chemical and its physical–chemical property data.⁷⁹ An Structure Data File (SDF) can contain multiple MOL files⁷⁹ and thus can contain 2D or 3D chemical data for any number of chemicals. Both MOL and SDF files are machine-readable and, to a certain extent, human readable. There are two versions of MOL and SDF files, which are used interchangeably but look notably different.⁸⁰ The original format was developed by Dalby et al.⁷⁹ and current standards for v2000 and v3000 are set by BIOVIA.⁸¹ v2000 is more widely used, however v3000 can handle more complex chemistry.^{80,81}

Table 2. Selection of Free Tools Available for Finding or Calculating Different Chemical Identifiers^{a,94–97}

	Unique Database/Registry Identifier	SMILES	InChI	InChIKey	2D/3D Structure	IUPAC Name	Synonyms	CAS RN	DTXSID	CID	2D/3D MOL
Databases											
PubChem ⁵²	CID	X	X	X	X	X	X	X	X	X	X
CompTox Chemical Dashboard ⁵⁴	DTXSID	X	X	X	X	X	X	X	X		X
Chemical Identifier Resolver ⁹⁴		X	X	X	X	X	X	X			X
ChEBI ⁵⁶	CHEBI ID	X	X	X	X	X	X	X			X
CAS Common Chemistry ^{58,95}	CAS RN	X	X	X	X	X	X	X			X
ChemSpider ⁵⁷	CSID	X	X	X	X	X	X	X			X
NIST Chemistry Webbook ⁹⁶		X		X	X		X	X			X
ChEMBL ⁵⁵	ChEMBL ID	X	X	X	X						X
Transforming Identifiers											
OpenBabel ¹¹ *		X	X	X	X	X	X				X
ACD/Labs ChemsSketch ¹⁰ *		X	X	X	X	X					X
MolView ⁷⁰					X						
Chemical Translation Service ⁹⁷		X		X			X	X			
Smi2Depict by ChemDB ⁶⁹		X			X						
CDK depict ⁶⁸					X						

^a“X” indicates that the tool can provide or produce these data; shaded cells indicate what input values or formats are accepted by the tool. * indicates that the tool is a downloadable program.

While there are many challenges with using structural representations, their use in combination with classical chemical identifiers, including names and database identifiers, can help reduce errors in chemical identification and identify erroneous entries. As previously noted, databases can have erroneous data where names and structures do not match.^{6,7} These errors are often due to a lack of available information regarding the exact structure of a compound, and if sufficient chemical structure information is provided with chemical names (and vice versa), errors can often be detected and corrected.

2.4. Structural Complexities

The structural representations discussed above work well to describe many compounds and substances. However, chemical information on ambiguous substances, mixtures, UVCB substances, and polymers is more challenging to manage in databases and for use with cheminformatics tools.

A substance that is ambiguously described may be lacking information regarding specificity of bond locations, e.g., trichloropyrene (CAS RN: 83690-29-3) or HBCD (CAS RN: 25637-99-4⁸² and DTXSID8025383⁸³) or have ambiguous stereochemistry (e.g., 6PPD-quinone or 1,2,5,6,9,10-HBCDD; see also Figure 1, substances A and B). Gobbi and Lee⁸⁴ note that when the stereochemistry of a substance is ambiguous, the substance should be considered to be a mixture of the stereoisomers. These ambiguous substances are often treated as singular compounds, and the need for specificity in the stereochemistry is not always apparent; it is therefore beneficial to be as specific as possible when reporting data. Stereoisomers can sometimes be identified by searching databases and registries for the substructure or similar substances (see Section SI 2); alternatively, some chemistry drawing tools such as ACD/Labs ChemSketch have methods

for generating all possible stereoisomers for a given structure.¹⁰ In the case of mass spectrometry, stereochemistry information is rarely available, and if there is no way of determining it otherwise, reporting structures without stereochemistry is sometimes the best approach to match the given information.

Some mixtures, such as racemic mixtures, are composed of equal quantities of two enantiomers (see Figure 1, substance B). In other instances, the ratio of constituents or the composition of the mixture is unknown and the substance could be classified as a UVCB. Take for example the UVCB, “Technical chlordane”, which is a commercial mixture that includes chlordane isomers and related compounds including isomers of nonachlor, *cis*-, and *trans*-nonachlor.⁸⁵ Previous analysis identified 120 components of the technical mixture, which includes 13 nonachlor C₁₀Cl₉ isomers.⁸⁶ The most common isomers, *cis*-nonachlor and *trans*-nonachlor, are found in some databases, individually or combined with a generic structure (Table SI 1). It is difficult to combine the details for all of these “technical chlordane” substituents into one identifier, and in many cases, not all substituents of the UVCB are known. Lai et al. provide specific recommendations for reporting and describing UVCBs.⁸⁷ Notably, they suggest providing as much detail as possible regarding the substance and suggest the use of extended SMILES, i.e., SMILES with additional layers of notation capturing structural unspecificity, and future use of Mixture InChIs (MInChIs), which is an IUPAC project in development.^{87,88} While no system for describing UVCBs is touted to be perfect, providing as much information as possible increases the interoperability of the chemical data and could allow future researchers and developments to retroactively improve the identification of these UVCBs in the future.

Polymers can also be difficult to describe or model.^{89,90} Audus and De Pablo note that polymers are often not made up of single entities but can be composed of different branches and contain chiral or multiple monomers.⁹⁰ Some databases also assign identifiers (e.g., CAS RN), to specific monomers, however, this does not take into account the complexity of synthetic polymers which can be composed of various combinations of unique monomers.⁹⁰ Japan's National Institute for Materials Science has published a database of polymers (PoLyInfo) with its own numbering system and more recently development of an ontology system for polymers.^{91,92}

3. FINDING AND GENERATING CHEMICAL DATA

It is difficult to obtain chemical identifiers without prior knowledge of the available chemical resources. Information can be obtained from a wide range of databases, journal articles, reference books, encyclopedias, safety data sheets, chemical registries, and chemical data tools. Depending on the preference and research scope of the user, some sources may be considered more reliable than others, yet no data source is comprehensive or complete. Access to data sources may be limited due to commercial licensing or intellectual property rights, while the accuracy of the data is dependent on the level of curation and reliability of the reported data or tool. These resources contribute immense value to the environmental sciences. Table 2 provides a nonexhaustive list of open resources available for obtaining chemical identifiers; while some of these resources provide additional information about the chemical structures, this table focuses on the availability of chemical identifiers (names, database identifiers, and structural representations). The availability of information depends on the compound or substance and the database. There may be instances where the reported CAS RN is a deprecated or deleted number, or different synonyms and spellings are used in different data sets. More resources can be found on the Wikipedia Page for Chemical Databases.⁹³

4. CURRENT REPORTING GUIDELINES

Current author guidelines for reporting chemical data in publications are inconsistent and can differ even between journals from the same publisher. Publishers play a key role in facilitating data sharing and encouraging the application of FAIR principles.⁴

As of early 2025, for the American Chemical Society (ACS) series of journals relating to environmental sciences, the journal guidelines recommended the use of SI units and the inclusion of IUPAC or International Union of Biochemistry and Molecular Biology (IUBMB) substance names. This includes Environmental Science & Technology (ES&T),⁹⁸ Environmental Science & Technology Letters,⁹⁹ ACS ES&T Engineering,¹⁰⁰ ACS ES&T Air,¹⁰¹ ACS ES&T Water,¹⁰² ACS Environmental Au,¹⁰³ and Environment & Health.¹⁰⁴ ES&T, ES&T Letters⁹⁹ and ES&T Air¹⁰¹ further recommend including chemical names or composition in the first mention. Environment & Health¹⁰⁴ and ACS Environmental Au¹⁰³ both recommend including trade names at first mention. These guidelines are in contrast to those of ACS's Journal of Chemical & Engineering Data (J. Chem. Eng. Data) guidelines,^{105,106} which require authors to provide the IUPAC name, CAS RN and 2D structural identifiers (e.g., SMILES, InChIs) for all chemicals studied. For experimental

data, J. Chem. Eng. Data further require the source of the chemicals and details regarding the purity of the substance.

The Society of Toxicology and Chemistry (SETAC) publishes two journals, Environmental Toxicology and Chemistry and Integrated Environmental Assessment and Management. Neither journal has any requirements or suggestions for reporting chemical data and identifiers in their author guidelines as of 2024.^{107,108} However, SETAC has published a Technical Issue Paper on "Recommending Minimum Reporting Information for Environmental Toxicological Studies".¹⁰⁹ While the focus is not only on chemical data, the authors make a clear call for increasing transparency and improving reporting practices.¹⁰⁹ Ågerstrand et al. further suggest that the technical name, CAS RN, purity, source, and physical–chemical properties of the chemicals used in the test be reported.¹⁰⁹ SETAC also has a Data Transparency Policy, which requires authors to include a data availability statement and encourages authors to provide data on open repositories.¹¹⁰

Elsevier provides no suggestions for reporting chemical information in their author guidelines for Chemosphere,¹¹¹ Environment International,¹¹² Environmental Pollution,¹¹³ Environmental Technology & Innovation,¹¹⁴ or Science of the Total Environment¹¹⁵ but do require research data to be provided or inclusion of a statement about its availability. Emerging Contaminants¹¹⁶ and Environmental Chemistry & Ecotoxicology¹¹⁷ are published jointly by Elsevier and China Science Publishing through KeAi and recommend similar reporting practices. However, Environmental Chemistry & Ecotoxicology also requires authors to provide a "stereochemistry abstract," which includes all available stereochemical information for every chiral compound.¹¹⁷ No additional details or requirements for nonchiral substances are mentioned.

The Royal Society of Chemistry (RSC) has taken a distinctive approach that follows the recommendations of Schymanski and Bolton.⁵ For articles published in any of their journals, including Environmental Science: Atmospheres, Environmental Science: Nano, Environmental Science: Processes & Impacts, and Environmental Science: Water Research & Technology, they recommend including a summary file that contains chemical structural information such as the SMILES, InChI or InChIKey, chemical names and synonyms and any relevant metadata.¹¹⁸

These four publishers currently host the most relevant journals in environmental sciences. While all publishers encourage data transparency principles and provide options for authors to submit Supporting Information, most provide little to no guidance on how to share their data. To facilitate data transparency, publishers should provide more guidelines to authors on how to report chemical data to improve data FAIR-ness. The next section discusses how to share chemical information critical to advancing environmental sciences.

5. RECOMMENDATIONS FOR DATA REPORTING

Schymanski and Bolton¹¹⁹ note that CSV (comma-separated value) files are the most interoperable file format available for sharing Supporting information, including chemical information in an interdisciplinary context (i.e., where not all professionals are data scientists or informaticians). These files are easily readable by coding languages and different file systems and through the use of commonly available software such as Microsoft Excel or Google Sheets, and simple free text

editors like the NotePad (Windows) or TextEditor (Apple iOS). Similarly, data can be shared in a TSV (tab-separated value) file.¹¹⁹ Both CSV and TSV file formats are easily read by a large number of free- and subscription-based software tools. The chemistry-specific formats, such as SDF or MOL, require external tools and packages to convert the data into a more user-friendly format and can be daunting for those less familiar with chemistry formats. When working with a large number of chemicals, SDFs can become very large and cumbersome to open and manipulate. However, SDF and MOL files allow for the transfer of 3D structural information of the chemical, which is difficult to do with other chemical identifiers reported in CSV or TSV files. For this reason, the use of CSV or TSV files is encouraged when reporting any chemical data, supplemented with SDF or MOL files as needed or if already available. Table SI 1 contains the chemical identifiers for all substances mentioned in this text and SI (see Section SI 1).

The file should contain clearly defined headers (CSV/TSV) or tags (MOL/SDF) and at least one chemical identifier: SMILES, InChI, a name, or InChIKey according to Schymanski and Bolton.⁵ It is best practice to avoid using spaces in the header names. Spaces can be replaced with “_” to maintain readability and reduce errors when reading files into external programs. Standardizing and using the FAIR Chemical Structures Template,⁵ with consistent column or header names helps cheminformaticians and data scientists to consolidate information to sync with databases and registries, without the need for additional data mapping.

To report novel chemicals accurately, at least two chemical identifiers should be reported, including at least one of SMILES or InChI. Any number of additional chemical identifiers that indicate common or trademarked names and database numbers (CAS RNs, CIDs, etc.) can also be included at the discretion of the author. It is also important to define the primary identifier (i.e., the most reliable of the information given) to help resolve occurrences of mismatching information (see Table SI 1). Having multiple sources of information about the chemical identity reduces the possibility of misidentification; defining the primary piece of information helps resolve potential clashes that remain. Since, for instance, CAS RNs can only be reliably obtained from a closed database (i.e., SciFinder⁵⁹) and publicly available mappings may be absent or prone to errors, it is useful to know whether this is the primary identifier during curation. A thorough description of the substance studied has the potential to improve the usability of the research when working with more complex scenarios, such as mixtures and polymers. Authors of all environmentally related disciplines are encouraged to adopt these best practices for reporting chemical information, and environmental science journals are encouraged to adopt minimum reporting standards for chemical identification (Box 2). These recommendations align with those made in an IUPAC Technical Report for good reporting practices of physical–chemical property measurements.¹²⁰

Obtaining the necessary identifiers depends on the certainty of the primary identifier. If the structure of the compound is available, the SMILES can be generated using drawing programs, which can then be used to search for the CAS RN on SciFinder⁵⁹ or CAS Common Chemistry,⁵⁸ or to search directly in CompTox¹²¹ or PubChem.¹²² Once a single database entry is found, verify that the record matches the specificity of the substance being characterized before extracting other chemical identifiers. If there is already a

Box 2

Recommendations for Reporting Chemical Information

File Format:

- CSV or TSV file
- MOL or SDF if the 3D conformation of the substance is important

Chemical Identifiers (minimum 2), indicating primary identifier:

- Structural identifier (minimum 1):
 - SMILES
 - InChI
- Other Identifier (minimum 1):
 - Chemical name
 - Database identifier
 - Optionally:
 - * InChIKey
 - * Additional names (IUPAC, common, technical) or synonyms
 - * Additional database identifiers (CAS RN, CID, DTXSID)
 - * Drawing of structures

primary identifier associated with the substance, e.g., CAS RN, check first that the entry matches the chemical you expect on the database platform, in this case, CAS Common Chemistry⁵⁸ or SciFinder.⁵⁹ If there is no database or registry data available about the chemical, or the available information is not as ambiguous or as specific as required, the structural identifier can be generated directly, using tools such as Open Babel¹¹ or using ACD/Labs ChemSketch.¹⁰ Erroneous or mismatched data found within a registry can be reported by email¹²² or through built-in error functions (e.g., ChemSpider¹²³ and CompTox “Submit Comment”¹²¹). When working with multiple chemicals at once, many tools, including CompTox,¹²¹ PubChem,¹²² and Open Babel,¹¹ are capable of running batch searches or conversions using chemical identifiers. Most registries also provide an Application Programming Interface (API) access to their data.

This paper focuses on guidelines for reporting chemical information; however, to improve FAIR practices and increase the applicability and rigor of scientific research, authors should aim to provide additional information regarding experimental and modeling data that they generate and use. Previous publications have provided guidance on reporting data related to environmental exposure,¹²⁴ plant bioaccumulation test results,¹²⁵ mass spectrometry¹²⁶ and property measurements.¹²⁰ If something was considered during the research design phase, it is likely something that others in the field or related fields would be interested in.

Ideally, all research articles related to chemistry in the environmental sciences should have at least one Supporting Information text or document containing a summary of the chemicals discussed. Almost all journal publications are entirely digital and provide options to include one or more SI documents. Additionally, the SI can be included as part of a preprint copy on ChemRxiv or on another free repository such as re3data¹²⁷ (recommended by SETAC Journals¹¹⁰), Zenodo,¹²⁸ or FigShare.¹²⁹ In some instances, it is not possible to submit CSV or TSV files as supplementary documentation. As an example, Table SI 1 is available on Zenodo at [10.5281/zenodo.14931110](https://doi.org/10.5281/zenodo.14931110) as a CSV file and as part of the Supporting

Information as an *XLSX* file. Sharing data and providing detailed SI are extremely useful to other researchers and allow the data to have better reach and a larger impact on future research.

The inclusion of Supporting Information with chemical data makes the data within the research article more findable and accessible to the reader. It prevents and reduces errors associated with the misidentification of chemicals, allowing data presented in the articles to be more readily used in future research, whether that involves building new models, designing experiments, or developing methods. Presenting data in consistent and clear formats helps make the integration of data into larger data sets easier for researchers and data managers, exemplified in initiatives such as PubChem,^{52,53} CompTox,⁵⁴ and the NORMAN Suspect List Exchange¹³⁰ (NORMAN-SLE). A further example of this includes data-driven approaches such as machine learning (ML) and artificial intelligence (AI), which require readily (re)usable data sets. If data is not made readily reusable, it is logical to assert that it will slow the rate of discovery and innovation within the fields of environmental science, chemistry, and toxicology.

The ultimate goal of scientific publishing is to disseminate findings and research. By doing so, we influence and impact future research and facilitate further developments in science and society. Therefore, the usefulness of results relating to chemical data is often dependent on the meta information that is made available and how well the chemical is identified. Results from exposure experiments investigating the toxicity or bioaccumulation of a chemical must be interpreted within the context of the identity, purity, concentration, and properties of the chemicals used. Environmental observations and measurements of chemicals can be used by other researchers to verify their own work and prioritize and deprioritize chemicals of emerging concern. Physical–chemical and toxicity data are useful in exposure and risk assessment, monitoring, experimental design, and the development and verification of quantitative structure–property–activity relationships (QSPR/QSAR).

During the days of print-only journals, the length of a journal article was far more restricted, and thus detailed reporting on the data and methodologies was often very limited. However, with the move to digital publishing, open access platforms, and data sharing, these restrictions no longer exist. Instead, we are confined only by journal and copyright limitations and our capacity and willingness to share data.

■ ASSOCIATED CONTENT

Data Availability Statement

Table SI 1 contains chemical information for all chemicals mentioned or referenced in this text and the SI as a CSV file on Zenodo at [10.5281/zenodo.14931110](https://doi.org/10.5281/zenodo.14931110).

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsenvironau.5c00034>.

Additional details, figures, and tables (citing refs 10,18,24, and 131) (PDF)

Chemical information for all chemicals mentioned or referenced in this text (Table SI 1) (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Sivani Baskaran – Norwegian Geotechnical Institute (NGI), 0484 Oslo, Norway; orcid.org/0000-0001-6192-5072; Email: sivanibaskaran1@gmail.com

Authors

Parviel Chirsir – Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 4367 Belvaux, Luxembourg

Shira Joudan – Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada; orcid.org/0000-0002-2736-8329

Raoul Wolf – Norwegian Geotechnical Institute (NGI), 0484 Oslo, Norway; orcid.org/0000-0001-5971-8525

Evan E. Bolton – National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), National Institutes of Health (NIH), Bethesda, Maryland 20894, United States; orcid.org/0000-0002-5959-6190

Paul A. Thiessen – National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), National Institutes of Health (NIH), Bethesda, Maryland 20894, United States; orcid.org/0000-0002-1992-2086

Emma L. Schymanski – Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 4367 Belvaux, Luxembourg; orcid.org/0000-0001-6868-8145

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsenvironau.5c00034>

Author Contributions

CRediT: **Sivani Baskaran** conceptualization, visualization, writing - original draft, writing - review & editing; **Parviel Chirsir** writing - original draft, writing - review & editing; **Shira Joudan** conceptualization, writing - review & editing; **Raoul Wolf** writing - review & editing; **Evan E. Bolton** writing - review & editing; **Paul A. Thiessen** writing - review & editing; **Emma L. Schymanski** conceptualization, supervision, writing - original draft, writing - review & editing.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

S.B., P.C., R.W., and E.L.S. acknowledge funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 101036756, project ZeroPM: Zero pollution of persistent, mobile substances. E.L.S. acknowledges funding by the Luxembourg National Research Fund (FNR), grant reference A18/BM/12341006. For the purpose of open access, and in fulfillment of the obligations arising from the grant agreement, the author has applied a Creative Commons Attribution 4.0 International (CC BY 4.0) license to any Author Accepted Manuscript version arising from this submission. The Natural Sciences and Engineering Research Council of Canada is acknowledged for funding S.B. via a Postdoctoral Fellowship and S.J. via a Discovery Grant (NSERC RGPIN-2023-04369). The work of P.A.T. and E.E.B. was supported by the National Center for Biotechnology Information of the National Library of Medicine (NLM), National Institutes of Health. We also acknowledge Emma Palm and Alessandro Sangion for fruitful discussions regarding chemical identifiers.

REFERENCES

- (1) Schollée, J. E.; Schymanski, E. L.; Avak, S. E.; Loos, M.; Hollender, J. Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. *Anal. Chem.* **2015**, *87*, 12121–12129.
- (2) Chao, A.; Al-Ghoul, H.; McEachran, A. D.; Balabin, I.; Transue, T.; Cathey, T.; Grossman, J. N.; Singh, R. R.; Ulrich, E. M.; Williams, A. J.; Sobus, J. R. In Silico MS/MS Spectra for Identifying Unknowns: A Critical Examination Using CFM-ID Algorithms and ENTACT Mixture Samples. *Anal. Bioanal. Chem.* **2020**, *412*, 1303–1315.
- (3) Zweigle, J.; Bugsel, B.; Fabregat-Palau, J.; Zwiener, C. PFAScreen – an Open-Source Tool for Automated PFAS Feature Prioritization in Non-Target HRMS Data. *Anal. Bioanal. Chem.* **2024**, *416*, 349–362.
- (4) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, No. 160018.
- (5) Schymanski, E. L.; Bolton, E. E. FAIR Chemical Structures in the Journal of Cheminformatics. *J. Cheminf.* **2021**, *13*, No. 50.
- (6) Glüge, J.; McNeill, K.; Scherlinger, M. Getting the SMILES Right: Identifying Inconsistent Chemical Identities in the ECHA Database, PubChem and the CompTox Chemicals Dashboard. *Environ. Sci.: Adv.* **2023**, *2*, 612–621.
- (7) Young, D.; Martin, T.; Venkatapathy, R.; Harten, P. Are the Chemical Structures in Your QSAR Correct? *QSAR Comb. Sci.* **2008**, *27*, 1337–1345.
- (8) International Union of Pure and Applied Chemistry Molecule. *IUPAC Compendium of Chemical Terminology*; IUPAC, 2006.
- (9) International Union of Pure and Applied Chemistry (IUPAC). Chemical Substance. *IUPAC Compendium of Chemical Terminology*; IUPAC, 2006.
- (10) ChemSketch. ACD/Labs, 2023. <https://www.acdlabs.com/resources/free-chemistry-software-apps/chemsketch-freeware/> (accessed Jan 15, 2025).
- (11) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011**, *3*, No. 33.
- (12) CAS Common Chemistry 1,2-Dichloroethylene. https://commonchemistry.cas.org/detail?cas_rn=540-59-0 (accessed Feb 14, 2025).
- (13) CAS Common Chemistry cis-1,2-Dichloroethene. https://commonchemistry.cas.org/detail?cas_rn=156-59-2 (accessed Feb 14, 2025).
- (14) CAS Common Chemistry trans-1,2-Dichloroethylene. https://commonchemistry.cas.org/detail?cas_rn=156-60-5 (accessed Feb 14, 2025).
- (15) PubChem (1R)-1-Bromo-1-chloroethane. <https://pubchem.ncbi.nlm.nih.gov/compound/96570687https://pubchem.ncbi.nlm.nih.gov/compound/96570687> (accessed Feb 14, 2025).
- (16) PubChem (1S)-1-Bromo-1-chloroethane. <https://pubchem.ncbi.nlm.nih.gov/compound/23616462> (accessed Feb 14, 2025).
- (17) CAS, a division of the American Chemical Society Check Digit Verification of CAS Registry Numbers. <https://www.cas.org/training/documentation/chemical-substances/checkdig> (accessed Jan 15, 2025).
- (18) Heller, S. R.; McNaught, A.; Pletnev, I.; Stein, S.; Tchekhovskoi, D. InChI, the IUPAC International Chemical Identifier. *J. Cheminf.* **2015**, *7*, No. 23.
- (19) United States Environmental Protection Agency Table of PCB Species by Congener Number, 2003. <https://www.epa.gov/sites/default/files/2015-09/documents/congenertable.pdfhttps://www.epa.gov/sites/default/files/2015-09/documents/congenertable.pdf> (accessed Feb 16, 2025).
- (20) Di, S.; Xu, H.; Yu, Y.; Qi, P.; Wang, Z.; Liu, Z.; Zhao, H.; Jin, Y.; Wang, X. Environmentally Relevant Concentrations of S-6PPD-quinone Caused More Serious Hepatotoxicity than R-enantiomer and Racemate in *Oncorhynchus Mykiss*. *Environ. Sci. Technol.* **2024**, *58*, 17617–17628.
- (21) Brunning, A. A Brief Guide to Types of Isomerism in Organic Chemistry, 2014. <https://www.compoundchem.com/2014/05/22/typesofisomerism/> (accessed Feb 05, 2025).
- (22) United Nations Environment Program. The Stockholm Convention on Persistent Organic Pollutants, 2023. <https://chm.pops.int/TheConvention/Overview/TextoftheConvention/tabid/2232/Default.aspx> (accessed June 11, 2025).
- (23) Report of the Conference of the Parties to the Stockholm Convention on Persistent Organic Pollutants on the Work of Its Sixth Meeting, 2013. <https://chm.pops.int/TheConvention/ConferenceoftheParties/ReportsandDecisions/tabid/208/Default.aspx> (accessed June 11, 2025).
- (24) International Union of Pure and Applied Chemistry (IUPAC). Optical Activity. *IUPAC Compendium of Chemical Terminology*; IUPAC, 2006.
- (25) Mallah, S. A.; Shaikh, H.; Memon, N.; Qazi, S. Fabrication of 1-Octane Sulphonic Acid Modified Nanoporous Graphene with Tuned Hydrophilicity for Decontamination of Industrial Wastewater from Organic and Inorganic Contaminants. *RSC Adv.* **2023**, *13*, 21926–21944.
- (26) Alarcon, P. C.; Kitanovski, Z.; Padervand, M.; Pöschl, U.; Lammel, G.; Zetzsch, C. Atmospheric Hydroxyl Radical Reaction Rate Coefficient and Total Environmental Lifetime of α -Endosulfan. *Environ. Sci. Technol.* **2023**, *57*, 15999–16005.
- (27) So Long Sulphur. *Nat. Chem.* **2009**, *1*, 333.
- (28) Hairston, N. Phosphorus: Time for Us to Oust Bad Spelling. *Nature* **2003**, *426*, 119–119.
- (29) Iheagwara, O. S.; Ing, T. S.; Kjellstrand, C. M.; Lew, S. Q. Phosphorus, Phosphorous, and Phosphate. *Hemodialysis Int.* **2013**, *17*, 479–482.
- (30) Government of Canada Public Services and Procurement Canada. Phosphorus, Phosphorous—Writing Tips plus—Writing Tools – Resources of the Language Portal of Canada—Canada.ca. <https://www.noslangues-ourlangues.gc.ca/writing-tips-plus/phosphorus-phosphorous> (accessed Feb 19, 2025).
- (31) Favre, H. A.; Powell, W. H. *Nomenclature of Organic Chemistry: IUPAC Recommendations and Preferred Names 2013*; Royal Society of Chemistry: Cambridge, 2014.
- (32) Moss, G. Blue Book, 2023. <https://iupac.qmul.ac.uk/BlueBook/PDF/> (accessed Oct 30, 2024).
- (33) Nomenclature of Inorganic Chemistry—IUPAC Recommendations 2005. *Chemistry International*; Royal Society of Chemistry, 2005; Vol. 27.
- (34) Jones, R. G.; Kahovec, J.; Stepto, R. F. T.; Wilks, E. S.; Hess, M.; Kitayama, T.; Metanowski, W. V. *Compendium of Polymer Terminology and Nomenclature: IUPAC Recommendations*, 2008; RSC Publications: Cambridge, 2009.
- (35) Hiorns, R. C.; Boucher, R. J.; Duhlev, R.; Hellwich, K.-H.; Hodge, P.; Jenkins, A. D.; Jones, R. G.; Kahovec, J.; Moad, G.; Ober, C. K.; Smith, D. W.; Stepto, R. F. T.; Vairon, J.-P.; Vohlidal, J. A Brief Guide to Polymer Nomenclature (IUPAC Technical Report). *Pure Appl. Chem.* **2012**, *84*, 2167–2169.
- (36) Hiorns, R. C.; Boucher, R. J.; Duhlev, R.; Hellwich, K.-H.; Hodge, P.; Jenkins, A. D.; Jones, R. G.; Kahovec, J.; Moad, G.; Ober, C. K.; Smith, D. W.; Stepto, R. F. T.; Vairon, J.-P.; Vohlidal, J. A Brief Guide to Polymer Nomenclature, 2012. <https://iupac.org/wp-content/uploads/2019/07/140-Brief-Guide-to-Polymer-Nomenclature-Web-Final-d.pdf> (accessed Oct 30, 2024).
- (37) Hartshorn, R. M.; Hellwich, K.-H.; Yerin, A.; Damhus, T.; Hutton, A. T. Brief Guide to the Nomenclature of Inorganic Chemistry. *Pure Appl. Chem.* **2015**, *87*, 1039–1049.
- (38) Hartshorn, R. M.; Hellwich, K.-H.; Yerin, A.; Damhus, T.; Hutton, A. T. Brief Guide to the Nomenclature of Inorganic Chemistry, 2017. <https://iupac.org/cms/wp-content/uploads/2016/07/Inorganic-Brief-Guide-VI-1.pdf> (accessed Oct 30, 2024).
- (39) Hellwich, K.-H.; Hartshorn, R. M.; Yerin, A.; Damhus, T.; Hutton, A. T. A Brief Guide to the Nomenclature of Organic Chemistry, 2021. <https://iupac.org/wp-content/uploads/2021/06/>

Organic-Brief-Guide-brochure_v1.1_June2021.pdf (accessed Oct 30, 2024).

(40) Hellwich, K.-H.; Hartshorn, R. M.; Yerin, A.; Damhus, T.; Hutton, A. T. Brief Guide to the Nomenclature of Organic Chemistry (IUPAC Technical Report). *Pure Appl. Chem.* **2020**, *92*, 527–539.

(41) ChemDoodle 2D. iChemLabs logo, 2025. <https://www.chemdoodle.com/> (accessed Jan 14, 2025).

(42) Lowe, D. Dan2097/Opsin, 2025. <https://github.com/dan2097/opsin> (accessed Jan 14, 2025).

(43) Lowe, D. M.; Corbett, P. T.; Murray-Rust, P.; Glen, R. C. Chemical Name to Structure: OPSIN, an Open Source Solution. *J. Chem. Inf. Model.* **2011**, *51*, 739–753.

(44) PubChem Clofenotane. <https://pubchem.ncbi.nlm.nih.gov/compound/3036> (accessed Jan 14, 2025).

(45) Chemical Identifier Resolver DDT IUPAC Name. https://cactus.nci.nih.gov/chemical/structure/ddt/iupac_name (accessed Jan 14, 2025).

(46) CompTox Chemistry Dashboard DDT. <https://comptox.epa.gov/dashboard/chemical/details/DTXSID4020375> (accessed Jan 14, 2025).

(47) ChemSpider DDT | C14H9Cl5. <https://www.chemspider.com/Chemical-Structure.2928.html> (accessed Jan 15, 2025).

(48) PubChem Methenamine. <https://pubchem.ncbi.nlm.nih.gov/compound/4101> (accessed Feb 16, 2025).

(49) Methenamine. <https://comptox.epa.gov/dashboard/chemical/details/DTXSID6020692> (accessed Feb 16, 2025).

(50) CAS Common Chemistry Methenamine. https://commonchemistry.cas.org/detail?cas_rn=100-97-0 (accessed Feb 16, 2025).

(51) CAS, a division of the American Chemical Society CAS Registry. <https://www.cas.org/cas-data/cas-registry> (accessed Jan 15, 2025).

(52) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem in 2021: New Data Content and Improved Web Interfaces. *Nucleic Acids Res.* **2021**, *49*, D1388–D1395.

(53) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2025 Update. *Nucleic Acids Res.* **2025**, *53*, D1516–D1525.

(54) Williams, A. J.; Grulke, C. M.; Edwards, J.; McEachran, A. D.; Mansouri, K.; Baker, N. C.; Patlewicz, G.; Shah, I.; Wambaugh, J. F.; Judson, R. S.; Richard, A. M. The CompTox Chemistry Dashboard: A Community Data Resource for Environmental Chemistry. *J. Cheminf.* **2017**, *9*, No. 61.

(55) Zdravil, B.; Felix, E.; Hunter, F.; et al. The ChEMBL Database in 2023: A Drug Discovery Platform Spanning Multiple Bioactivity Data Types and Time Periods. *Nucleic Acids Res.* **2024**, *52*, D1180–D1192.

(56) Hastings, J.; Owen, G.; Dekker, A.; Ennis, M.; Kale, N.; Muthukrishnan, V.; Turner, S.; Swainston, N.; Mendes, P.; Steinbeck, C. ChEBI in 2016: Improved Services and an Expanding Collection of Metabolites. *Nucleic Acids Res.* **2016**, *44*, D1214–D1219.

(57) Pence, H. E.; Williams, A. ChemSpider: An Online Chemical Information Resource. *J. Chem. Educ.* **2010**, *87*, 1123–1124.

(58) CAS Common Chemistry. <https://commonchemistry.cas.org/> (accessed Jan 13, 2023).

(59) CAS Common Chemistry. CAS, a division of the American Chemical Society CAS SciFinder. <https://www.cas.org/solutions/cas-scifinder-discovery-platform/cas-scifinder> (accessed Feb 16, 2025).

(60) Liu, H.; Shi, J.; Liu, H.; Wang, Z. Improved 3D-QSPR Analysis of the Predictive Octanol–Air Partition Coefficients of Hydroxylated and Methoxylated Polybrominated Diphenyl Ethers. *Atmos. Environ.* **2013**, *77*, 840–845.

(61) Bruford, E. A.; Braschi, B.; Denny, P.; Jones, T. E. M.; Seal, R. L.; Tweedie, S. Guidelines for Human Gene Nomenclature. *Nat. Genet.* **2020**, *52*, 754–758.

(62) Ziemann, M.; Eren, Y.; El-Osta, A. Gene Name Errors Are Widespread in the Scientific Literature. *Genome Biol.* **2016**, *17*, No. 177.

(63) Abeysooriya, M.; Soria, M.; Kasu, M. S.; Ziemann, M. Gene Name Errors: Lessons Not Learned. *PLOS Comput. Biol.* **2021**, *17*, No. e1008984.

(64) Fifadra, C. Control Data Conversions in Excel for Windows and Mac, 2023. <https://insider.microsoft365.com/he-il/blog/control-data-conversions-in-excel-for-windows-and-mac> (accessed July 01, 2024).

(65) Daylight Chemical Information Systems Inc.. Daylight Theory: SMILES. <https://www.daylight.com/dayhtml/doc/theory/theory.smiles.html> (accessed Jan 14, 2025).

(66) Mansouri, K.; Moreira-Filho, J. T.; Lowe, C. N.; Charest, N.; Martin, T.; Tkachenko, V.; Judson, R.; Conway, M.; Kleinstreuer, N. C.; Williams, A. J. Free and Open-Source QSAR-ready Workflow for Automated Standardization of Chemical Structures in Support of QSAR Modeling. *J. Cheminf.* **2024**, *16*, No. 19.

(67) McEachran, A. D.; Mansouri, K.; Grulke, C.; Schymanski, E. L.; Ruttkies, C.; Williams, A. J. “MS-ready” Structures for Non-Targeted High-Resolution Mass Spectrometry Screening Studies. *J. Cheminf.* **2018**, *10*, No. 45.

(68) Cdk/Depict. The Chemistry Development Kit, 2025. <https://github.com/cdk/depict> (accessed Jan 14, 2025).

(69) Chen, J.; Swamidass, S. J.; Dou, Y.; Bruand, J.; Baldi, P. ChemDB: A Public Database of Small Molecules and Related Chemoinformat Ics Resources. *Bioinformatics* **2005**, *21*, 4133–4139.

(70) MolView. <https://app.molview.com> (accessed Jan 15, 2025).

(71) ChemAxon. Introduction to MarvinSketch | Chemaxon Docs. <https://docs.chemaxon.com/display/lts-europium/introduction-to-marvinsketch.md> (accessed Jan 15, 2025).

(72) ChemAxon. Marvin—Chemical Drawing Software. <https://chemaxon.com/marvin> (accessed Jan 15, 2025).

(73) ChemDraw. Revvity Signals. <https://revvitysignals.com/products/research/chemdraw> (accessed Jan 15, 2025).

(74) United States Environmental Protection Agency Estimation Programs Interface Suite. United States Environmental Protection Agency, 2012. <https://www.epa.gov/tsca-screening-tools/download-epi-suite-estimation-program-interface-v411> (accessed June 28, 2017).

(75) Mansouri, K.; Grulke, C. M.; Judson, R. S.; Williams, A. J. OPERA Models for Predicting Physicochemical Properties and Environmental Fate Endpoints. *J. Cheminf.* **2018**, *10*, No. 10.

(76) InChI Trust. InChI Homepage. <https://www.inchi-trust.org/> (accessed Jan 14, 2025).

(77) Heller, S.; McNaught, A.; Stein, S.; Tchekhovskoi, D.; Pletnev, I. InChI—the Worldwide Chemical Structure Identifier Standard. *J. Cheminf.* **2013**, *5*, No. 7.

(78) Yerin, A. Improvements in InChI Treatment of Stereo-configuration, 2017. <https://www.inchi-trust.org/wp/wp-content/uploads/2017/11/11-InChI-Stereo-Yerin-201708.pdf> (accessed Feb 25, 2025).

(79) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244–255.

(80) Apodaca, R. L. Ten Reasons to Adopt the V3000 Molfile Format, 2021. <http://depth-first.com/articles/2021/11/17/ten-reasons-to-adopt-the-v3000-molfile-format/> (accessed Feb 17, 2025).

(81) BIOVIA Databases 2020 CTFile Formats, 2020. https://discover.3ds.com/sites/default/files/2020-08/biovia_ctfileformats_2020.pdf (accessed Feb 17, 2025).

(82) CAS Common Chemistry Hexabromocyclododecane. https://commonchemistry.cas.org/detail?cas_rn=25637-99-4 (accessed Feb 17, 2025).

(83) CompTox Chemistry Dashboard Hexabromocyclododecane. <https://comptox.epa.gov/dashboard/chemical/details/DTXSID8025383> (accessed Feb 17, 2025).

- (84) Gobbi, A.; Lee, M.-L. Handling of Tautomerism and Stereochemistry in Compound Registration. *J. Chem. Inf. Model.* **2012**, *52*, 285–292.
- (85) McGaughy, R. E.; Foureman, G. L.; McClure, P. *Toxicological Review of Chlordane (Technical)*; U.S. EPA Toxicological Review, 1997.
- (86) Dearth, M. A.; Hites, R. A. Complete Analysis of Technical Chlordane Using Negative Ionization Mass Spectrometry. *Environ. Sci. Technol.* **1991**, *25*, 245–254.
- (87) Lai, A.; Clark, A. M.; Escher, B. I.; Fernandez, M.; McEwen, L. R.; Tian, Z.; Wang, V. C.; Schymanski, E. L. The next Frontier of Environmental Unknowns: Substances of Unknown or Variable Composition, Complex Reaction Products, or Biological Materials (UVCBs). *Environ. Sci. Technol.* **2022**, *56*, 7448–7466.
- (88) IUPAC Project Details: InChI Extension for Mixture Composition, 2023. <https://iupac.org/project/2015-025-4-800/> (accessed Oct 18, 2023).
- (89) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative Structure-Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* **2012**, *112*, 2889–2919.
- (90) Audus, D. J.; De Pablo, J. J. Polymer Informatics: Opportunities and Challenges. *ACS Macro Lett.* **2017**, *6*, 1078–1082.
- (91) Ishii, M.; Ito, T.; Sakamoto, K. NIMS Polymer Database PoLyInfo (II): Machine-Readable Standardization of Polymer Knowledge Expression. *Sci. Technol. Adv. Mater.: Methods* **2024**, *4*, No. 2354651.
- (92) Ishii, M.; Ito, T.; Sado, H.; Kuwajima, I. NIMS Polymer Database PoLyInfo (I): An Overarching View of Half a Million Data Points. *Sci. Technol. Adv. Mater.: Methods* **2024**, *4*, No. 2354649.
- (93) List of Chemical Databases, Wikipedia, 2024. https://en.wikipedia.org/w/index.php?title/List_of_chemical_databases (accessed Jan 15, 2025).
- (94) National Cancer Institute Chemical Identifier Resolver. <https://cactus.nci.nih.gov/chemical/structure> (accessed Sept 24 2024).
- (95) Jacobs, A.; Williams, D.; Hickey, K.; Patrick, N.; Williams, A. J.; Chalk, S.; McEwen, L.; Willighagen, E.; Walker, M.; Bolton, E.; Sinclair, G.; Sanford, A. CAS Common Chemistry in 2021: Expanding Access to Trusted Chemical Information for the Scientific Community. *J. Chem. Inf. Model.* **2022**, *62*, 2737–2743.
- (96) Linstrom, P. NIST Chemistry WebBook, NIST Standard Reference Database 69, 2023. <http://webbook.nist.gov/chemistry/> (accessed Jan 16, 2025).
- (97) Wohlgemuth, G.; Haldiya, P. K.; Willighagen, E.; Kind, T.; Fiehn, O. The Chemical Translation Service—a Web-Based Tool to Improve Standardization of Metabolomic Reports. *Bioinformatics* **2010**, *26*, 2647–2648.
- (98) American Chemical Society Author Guidelines—Environmental Science & Technology, 2024. https://researcher-resources.acs.org/publish/author_guidelines?coden=esthag (accessed Jan 15, 2025).
- (99) American Chemical Society Author Guidelines—Environmental Science & Technology Letters, 2024. https://researcher-resources.acs.org/publish/author_guidelines?coden=estlcu (accessed Jan 15, 2025).
- (100) American Chemical Society Author Guidelines—ACS ES&T Engineering, 2024. https://researcher-resources.acs.org/publish/author_guidelines?coden=aeecco (accessed Jan 15, 2025).
- (101) American Chemical Society Author Guidelines—ACS ES&T Air, 2024. https://researcher-resources.acs.org/publish/author_guidelines?coden=aeacd5 (accessed Jan 15, 2025).
- (102) American Chemical Society Author Guidelines—ACS ES&T Water, 2024. https://researcher-resources.acs.org/publish/author_guidelines?coden=aeewca (accessed Jan 15, 2025).
- (103) American Chemical Society Author Guidelines—ACS Environmental Au, 2024. https://researcher-resources.acs.org/publish/author_guidelines?coden=aeacc4 (accessed Jan 15, 2025).
- (104) American Chemical Society Author Guidelines—Environment & Health, 2024. https://researcher-resources.acs.org/publish/author_guidelines?coden=ehnea2 (accessed Jan 15, 2025).
- (105) American Chemical Society Author Guidelines—Journal of Chemical & Engineering Data, 2024. https://researcher-resources.acs.org/publish/author_guidelines?coden=jceaax (accessed Jan 15, 2025).
- (106) American Chemical Society Submission Checklist - Journal of Chemical & Engineering Data, 2022. https://pubs.acs.org/pb-assets/documents/jceaax/JCED_Submission_Checklist.pdf (accessed Jan 15, 2025).
- (107) Oxford Academic General Instructions | Environmental Toxicology and Chemistry. <https://academic.oup.com/etc/pages/general-instructions> (accessed Jan 15, 2025).
- (108) Oxford Academic General Instructions | Integrated Environmental Assessment and Management. <https://academic.oup.com/ieam/pages/general-instructions> (accessed Jan 15, 2025).
- (109) Ågerstrand, M.; Alix, A.; Ajao, C.; Hanson, M.; Hoff, D.; Moermond, C.; Schlekot, T.; Staveley, J. Recommended Minimum Reporting Information for Environmental Toxicity Studies, 2019. <https://www.setac.org/resource/recommended-minimum-reporting-information-for-environmental-toxicity-studies-tip.html> (accessed July 01, 2024).
- (110) SETAC Journals Data Transparency Policy, 2019. https://cdn.ymaws.com/www.setac.org/resource/resmgr/Publications_and_Resources/SETAC-data-transparency-poli.pdf (accessed Jan 15, 2025).
- (111) Elsevier Guide for Authors—Chemosphere. <https://www.sciencedirect.com/journal/chemosphere/publish/guide-for-authors> (accessed Jan 15, 2025).
- (112) Elsevier Guide for Authors—Environment International. <https://www.sciencedirect.com/journal/environment-international/publish/guide-for-authors> (accessed Jan 15, 2025).
- (113) Elsevier Guide for Authors—Environmental Pollution. <https://www.sciencedirect.com/journal/environmental-pollution/publish/guide-for-authors> (accessed Jan 15, 2025).
- (114) Elsevier Guide for Authors—Environmental Technology & Innovation. <https://www.sciencedirect.com/journal/environmental-technology-and-innovation/publish/guide-for-authors> (accessed Jan 15, 2025).
- (115) Elsevier Guide for Authors—Science of The Total Environment. <https://www.sciencedirect.com/journal/science-of-the-total-environment/publish/guide-for-authors> (accessed Jan 15, 2025).
- (116) KeAi Guide for Authors—Emerging Contaminants. <https://www.keaipublishing.com/en/journals/emerging-contaminants/guide-for-authors> (accessed Jan 15, 2025).
- (117) KeAi Guide for Authors—Environmental Chemistry and Ecotoxicology. <https://www.keaipublishing.com/en/journals/environmental-chemistry-and-ecotoxicology/guide-for-authors> (accessed Jan 15, 2025).
- (118) Royal Society of Chemistry. Experimental Reporting Requirements, Characterisation of Compounds and Materials, 2024. <https://www.rsc.org/journals-books-databases/author-and-reviewer-hub/authors-information/prepare-and-format/experimental-reporting-requirements/#characterisation-of-new-compounds> (accessed Dec 18, 2024).
- (119) Schymanski, E. L.; Bolton, E. E. FAIRifying the Exposome Journal: Templates for Chemical Structures and Transformations. *Exposome* **2022**, *2*, No. osab006.
- (120) Bazyleva, A.; Abildskov, J.; Anderko, A.; et al. Good Reporting Practice for Thermophysical and Thermochemical Property Measurements (IUPAC Technical Report). *Pure Appl. Chem.* **2021**, *93*, 253–272.
- (121) United States Environmental Protection Agency. CompTox Chemicals Dashboard Version 2.5.2. <https://comptox.epa.gov/dashboard/> (accessed Feb 17, 2025).
- (122) National Library of Medicine. PubChem, 2022. <https://pubchem.ncbi.nlm.nih.gov/> (accessed Jan 13, 2023).

- (123) Royal Society of Chemistry. ChemSpider: Search and Share Chemistry. <https://www.chemspider.com/> DOI: 10.1016/j.jenv-man.2016.06.065 (accessed Feb 25, 2025).
- (124) Udesky, J. O.; Dodson, R. E.; Perovich, L. J.; Rudel, R. A. Wrangling Environmental Exposure Data: Guidance for Getting the Best Information from Your Laboratory Measurements. *Environ. Health* **2019**, *18*, No. 99.
- (125) Fantke, P.; Arnot, J. A.; Doucette, W. J. Improving Plant Bioaccumulation Science through Consistent Reporting of Experimental Data. *J. Environ. Manage.* **2016**, *181*, 374–384.
- (126) Xi, Y.; Sohn, A. L.; Joignant, A. N.; Cologna, S. M.; Prentice, B. M.; Muddiman, D. C. SMART: A Data Reporting Standard for Mass Spectrometry Imaging. *J. Mass Spectrom.* **2023**, *58*, No. e4904.
- (127) Re3data. <https://www.re3data.org/> (accessed Jan 15, 2025).
- (128) Zenodo. <https://zenodo.org/> (accessed Jan 15, 2025).
- (129) FigShare. Figshare—Credit for All Your Research, <https://figshare.com/> (accessed Jan 15, 2025).
- (130) Mohammed Taha, H.; Aalizadeh, R.; Alygizakis, N.; et al. The NORMAN Suspect List Exchange (NORMAN-SLE): Facilitating European and Worldwide Collaboration on Suspect Screening in High Resolution Mass Spectrometry. *Environ. Sci. Eur.* **2022**, *34*, No. 104.
- (131) Huang, R.; Lin, Z.; Liu, Y.; Wu, X.; Yuan, K. A “Hand-Held” Polarimeter for on-Site Chiral Drug Measurement and Chemical Reaction Monitoring. *Anal. Bioanal. Chem.* **2025**, *417*, 1055–1065.