



**HAL**  
open science

## Accumulation-depuration data collection in support of toxicokinetic modelling

Aude Ratier, Sandrine Charles

► **To cite this version:**

Aude Ratier, Sandrine Charles. Accumulation-depuration data collection in support of toxicokinetic modelling. 2021. hal-03362531

**HAL Id: hal-03362531**

**<https://hal.science/hal-03362531v1>**

Preprint submitted on 1 Oct 2021

**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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# 1 Accumulation-depuration data collection in support 2 of toxicokinetic modelling

3 Aude Ratier<sup>1</sup> and Sandrine Charles<sup>1,\*</sup>

4 <sup>1</sup>Université de Lyon, Université Lyon 1, CNRS UMR5558, Laboratoire de Biométrie et Biologie Evolutive, 69100  
5 Villeurbanne, France

6 \*corresponding author: Sandrine Charles ([sandrine.charles@univ-lyon1.fr](mailto:sandrine.charles@univ-lyon1.fr))

## 7 ABSTRACT

Regulatory bodies requires evaluation of bioaccumulation of chemicals within organisms with the objective to better assess risks linked to the toxicity of active substances. To this end, toxicokinetic (TK) data are particularly useful to relate the chemical exposure concentration to the accumulation and depuration processes happening within organisms. The bioaccumulative property of substances is quantified by bioaccumulation metrics obtained by fitting TK models to data collected from bioaccumulation tests. The internal concentrations of the studied substances are measured within organisms  
8 at regular time points during both accumulation and depuration phases, and their time course is captured by TK models thus providing bioaccumulation metrics. Still today, raw TK data remain difficult to access, as most of the time provided within papers in plots only. To increase accessibility to TK data, we present in this paper a wide collection of raw data sets extracted from the scientific literature in support of TK modelling to be performed with the MOSAIC<sub>bioacc</sub> web application (<https://mosaic.univ-lyon1.fr/bioacc/>).

## 9 Background & Summary

10 The Environmental Risk Assessment (ERA) workflow for chemicals of interest, as described in some European regulations (*e.g.*,  
11 for plant protection products in marketing authorisation applications (EU regulation No 283/2013)), requires a bioaccumulation  
12 test, as for example on fish according to the OECD Test guideline 305<sup>1</sup>. Such a test consists in an accumulation phase followed  
13 by a depuration one and the time course of the internal concentration within fish is measured at regular time points during  
14 both phases. The resulting data allow to model the time course of the exposure within organisms (denoted toxicokinetic, TK),  
15 summarized in the end via bioaccumulation metrics (BCF/BSAF/BMF for water, sediment and food exposure, respectively).  
16 From a regulatory point of view, these bioaccumulation metrics are key decision criteria to determine the bioaccumulative  
17 property of active substances<sup>2</sup>, and to further assess potential risks that are associated to them according to the exposure sources.

18 Kinetic bioaccumulation metrics are always defined as ratios between uptake and elimination rates, these latter being

19 estimated from TK models<sup>1,3</sup>. More precisely, TK models relate the exposure concentration to a given substance to the internal  
20 concentration within organisms, considering various processes such as accumulation, depuration, metabolism and excretion  
21 (ADME)<sup>4</sup>. Different types of TK models have been proposed, all being compartment models<sup>5,6</sup>. Organisms are usually  
22 described as unique compartments with input and output fluxes described by the TK models. Data collected from standard  
23 bioaccumulation tests are used to fit the TK models leading to uptake and elimination rate estimates. Finally, bioaccumulation  
24 metrics are calculated in support of regulatory decision making.

25 Some years ago, the United States Environmental Protection Agency (US EPA) has created a database for ecotoxicology  
26 data providing such bioaccumulation metrics according to species-compound combinations<sup>7</sup>. However, raw data are rarely  
27 provided in this database as this would require to collect them directly from the corresponding scientific papers. Moreover, when  
28 available, raw data are often provided as plots only. Nevertheless, to make updates or develop new TK modelling frameworks,  
29 it is crucial for researchers to benefit of a collection of raw TK data in order to check the robustness of their new approaches.  
30 While developing the MOSAIC<sub>bioacc</sub> web application<sup>8</sup> (<https://mosaic.univ-lyon1.fr/bioacc/>), we dealt with  
31 such a lack of raw data to fully test our innovative approach. Indeed, MOSAIC<sub>bioacc</sub> provides estimates of TK model parameters  
32 and bioaccumulation metrics with their uncertainties for a large range of species-compound combinations (*e.g.*, aquatic or  
33 terrestrial organisms exposed to metals, hydrocarbons, active substances, etc.), encompassing different exposure routes and  
34 elimination processes. For example, it was particularly difficult to collect raw TK data with biotransformation processes being  
35 involved<sup>9</sup>, preventing us to fully test the robustness of this tool for the widest as possible diversity of input data types.

36 This motivated us in creating a new public database presented in this paper. This new database gathers together more  
37 than 90 data sets of published bioaccumulation tests, concerning more than 20 genus and more than 70 substances. Some  
38 data sets concern several exposure routes (water, soil or sediment, and/or food) and several possible elimination processes  
39 (excretion, growth dilution and/or biotransformation). All the collected data were standardized in the same units and uploaded  
40 in MOSAIC<sub>bioacc</sub>, ensuring the use of the same methodology to get bioaccumulation metrics for all the data sets. This database  
41 should allow to overcome the current lack of raw TK data in ecotoxicology. Indeed, our purpose with this first version of the  
42 database is to motivate other researchers to share their data based on the FAIR (Findability, Accessibility, Interoperability and  
43 Reuse) principles which became today almost a duty. Our database can be considered as a proof of concept of the added-value  
44 of sharing raw TK data. It allows anyone to reuse data, for example to test new modelling frameworks. This database should  
45 also facilitate the design of new bioaccumulation experiments, as well as the comparison of results between several studies,  
46 benefiting of the unified calculation method for bioaccumulation metrics as provided by MOSAIC<sub>bioacc</sub>. Finally, we hope that  
47 this database could help changing of paradigm in ecotoxicology, through an increase in a more wide sharing of raw data that  
48 would certainly lead to better knowledge in the perspective of ERA.

## 49 **Methods**

50 The conceptual workflow we followed to collect the raw TK data is summarized on Figure 1 and detailed below.

## 51 **Data and literature sources**

52 One of our aims was to test the robustness of the generic feature of the MOSAIC<sub>bioacc</sub> methodological framework from a wide  
53 collection of TK data sets, encompassing a variety of genus, substances, exposure routes and elimination processes. For this  
54 purpose, the US EPA ecotoxicology database<sup>7</sup> could not be used as raw observations are usually absent together with a lack of  
55 information (*e.g.*, biotransformation processes are usually not informed). Consequently, an in-depth search in literature with  
56 Scopus was performed with specific keywords according the expected type of TK data sets. These keywords were for example:  
57 "TK model aquatic", "TK model terrestrial", "TK model biotransformation", "TK biotransformation rate", "TK model food  
58 exposure", "TK model sediment exposure", "TK model water exposure", etc. For each article retained, we manually checked if  
59 TK data were available in a table or in a plot, in the manuscript or in the supporting information. A total of 28 studies was  
60 finally selected for a full analysis with a TK model.

## 61 **Toxicokinetic data**

### 62 ***Literature source data extraction***

63 Only few publications provided raw data as table of internal concentrations of a chemical (and its potential metabolite(s))  
64 within organisms versus time. For most of the studies, data were only available in plots. Thus, plots were converted in JPEG  
65 files from screenshots (native screenshot tool in Windows 10). Then images were imported in WebPlotDigitizer<sup>10</sup>, which assists  
66 the users to extract the underlying numerical data of a plot. The calibration of axes and the selection of points for extraction is  
67 performed by the user. Raw data are exported from WebPlotDigitizer to CSV files. Then, CSV files data were formatted for  
68 MOSAIC<sub>bioacc</sub> requirements, manually adding information on replicates and exposure concentration according to the Methods  
69 section of each article. For each collected TK data set, the corresponding file was named by the genus, the substance, the  
70 duration of the accumulation phase, the first author and the year of publication, possibly adding comments (*e.g.*, size of the  
71 tested organism) when available.

### 72 ***Data standardization***

73 Data were standardized to always express exposure concentrations in  $\mu\text{g.mL}^{-1}$  (for water exposures) or in  $\mu\text{g.g}^{-1}$  (for sediment  
74 or food exposures). The internal chemical concentrations in organisms were translated in  $\mu\text{g.g}^{-1}$ , as required for a proper use  
75 with MOSAIC<sub>bioacc</sub>.

### 76 ***MOSAIC<sub>bioacc</sub> running***

77 Once standardized, each collected data set was uploaded in MOSAIC<sub>bioacc</sub> ([https://mosaic.univ-lyon1.fr/  
78 bioacc](https://mosaic.univ-lyon1.fr/bioacc)). The generic modelling framework of MOSAIC<sub>bioacc</sub> automatically provided both kinetic and steady-state bioac-  
79 cumulation metrics, unifying their calculation among data sets allowing relevant comparisons or a correct classification of  
80 substances according to their bioaccumulative property for example. Kinetic bioaccumulation metrics (given as medians  
81 and 95% credible intervals) were saved in a local repository, and the report with all fitting outputs downloaded in the same  
82 repository, for each data set.

## 83 Storage

84 An R project<sup>11</sup> was created in a local repository of the MOSAIC<sub>bioacc</sub> Shiny application, containing all collected data sets,  
85 the bibtex file with all references, all reports and all kinetic bioaccumulation metric estimates. Then, an RMarkdown file<sup>12,13</sup>  
86 allows to create the final table based on information collected from the name of the data set and from the data set itself (*e.g.*,  
87 column headers, number of data, number of replicates), as well as from the bibtex file. The R package 'DT'<sup>14</sup> was used to  
88 combine all collected information in a user-friendly table, and the RMarkdown file was knitted in HTML format for display to  
89 the user. In the future, each new data set added into the repository will make the RMarkdown automatically knitted<sup>13</sup> for update.

## 90 Data Records

91 The database can be accessed directly on-line at [http://lbbe-shiny.univ-lyon1.fr/mosaic-bioacc/data/](http://lbbe-shiny.univ-lyon1.fr/mosaic-bioacc/data/database/index_readme.html)  
92 [database/index\\_readme.html](http://lbbe-shiny.univ-lyon1.fr/mosaic-bioacc/data/database/index_readme.html) or via MOSAIC<sub>bioacc</sub> clicking the "More scientific TK data" button. An example of the  
93 output of the database is shown in Figure 2. The collected raw TK data of the database consist in time courses of several types  
94 of chemicals bioaccumulated in various species via different exposure routes. Each data set is summarized by:

- 95 • the file name (raw data directly downloadable by clicking on the file name, in TXT or CSV format),
- 96 • the genus of the tested organism,
- 97 • the category of the organism (*e.g.*, aquatic, terrestrial, etc.),
- 98 • the tested substance,
- 99 • the duration of the accumulation phase,
- 100 • the tested exposure routes and for which data are available (*e.g.*, water, sediment, food, pore water),
- 101 • the total number of data in the data set (and in brackets the number of replicate(s)),
- 102 • the kinetic bioaccumulation metric median value with its 95% credible interval,
- 103 • the report which contains all the results from MOSAIC<sub>bioacc</sub> (in PDF file),
- 104 • the link to the reference or the source of data,
- 105 • some additional comments (*e.g.*, lipid fraction, growth, biotransformation).

106 A summary of all data sets is presented in Table 1. Genus were separated in nine categories: aquatic invertebrates ( $n = 51$ ),  
107 fish ( $n = 21$ ), insects ( $n = 5$ ), aquatic worms ( $n = 5$ ), terrestrial worms ( $n = 3$ ), seawater sponges ( $n = 2$ ), seawater plants ( $n =$   
108 1), aquatic algae ( $n = 1$ ) and terrestrial invertebrates ( $n = 1$ ). The most represented genus in the database are *Oncorhynchus*  
109 (fish) and *Gammarus* (aquatic invertebrate), followed by *Daphnia* (aquatic invertebrate), classically used in ecotoxicological  
110 experiments. Recommended genus by OECD guidelines for bioaccumulation tests are *Eisenia* and *Enchytraeus* for terrestrial

111 organisms (OECD 317)<sup>15</sup>, and *Tubifex* or *Lumbriculus* for aquatic invertebrates exposed to sediment (OECD 315)<sup>16</sup>; data sets  
112 for these species are available in the database ( $n = 8$ ).

113 Substances were divided in nine classes: pesticides ( $n = 23$ ), hydrocarbons ( $n = 14$ ), metals ( $n = 9$ ), nanoparticles ( $n =$   
114  $5$ ), polychlorobiphenyls (PCB,  $n = 5$ ), flame retardants (brominated or chlorinated,  $n = 3$ ), pharmaceutical products ( $n = 2$ ),  
115 polyesters ( $n = 2$ ) and octyphenol ( $n = 1$ ). Among all data sets, the majority of bioaccumulation tests were performed via spiked  
116 water. Besides, 18 data sets account for biotransformation processes, considering from 1 to 8 metabolites.

117 According to ECHA (2017)<sup>2</sup>, bioaccumulation metrics below 1,000 mean that the substance is not bioaccumulative, whereas  
118 one ranging between 2,000 and 5,000 corresponds to a bioaccumulative substance. If the bioaccumulation metric is higher  
119 than 5,000, the substance is classified as very bioaccumulative. These ranges are reported in Table 1, where bioaccumulation  
120 metrics are higher than 5,000 for eight data sets, indicating the very bioaccumulative property of the substances for the studied  
121 genus. On an ecotoxicological point of view, the most higher bioaccumulation metrics were obtained for *Daphnia* exposed to  
122 alpha-cypermethrin, triphenyltin hydroxide (TPT), diazinon and imidacloprid; *Hydropsyche*, *Chaetopteryx*, *Tubifex* and *Asellus*  
123 exposed to cypermethrin, highlighting the potential toxicity of cypermethrin for aquatic and terrestrial organisms.

## 124 Technical Validation

125 The validity of the collected data was checked by considering several aspects in the corresponding papers: (i) limitations in the  
126 method and/or source material, (ii) unreported replicates (experimental variability missing), (iii) inaccuracy in the collection  
127 techniques used to extract data from source content, and (iv) the quality of the results we obtained (based goodness-of-fit and  
128 precision criteria) when reusing the data in our TK modelling framework.

129 For some studies, raw data were not always consistent with the associated plot (*e.g.*, differences in time points or measured  
130 concentrations in organisms) as provided within the original paper. When a possible inaccuracy collection of data was noticed,  
131 the data set was removed from the database. Besides, most of the studies did not provide all data, for example the measurement  
132 concentration in each replicate were missing. Indeed, most of the time, plots only provide average measured data with error  
133 bars representing standard deviation, standard error or other uncertainty measures, without giving access to the raw observed  
134 concentrations. In such cases, only the mean values were collected.

135 The uncertainty due to data extraction from plots with the WebPlotDigitizer tool<sup>10</sup> is legitimate to consider in order to see  
136 if a bias was introduced in collected data sets. To quantify this bias, we performed ten times the collection of TK data for  
137 a same data set (.jpeg file) and calculated the extraction errors possibly due to the shape of points, the axis calibration, the  
138 image resolution or some other factors. For example, a variation of  $0.12 \pm 0.13\%$  was observed for the data set of Chen *et al.*,  
139 (2017)<sup>17</sup> (Figure 2-a in their paper), which is relatively low ( $< 0.5\%$  as usually considered for coefficients of variation). This is  
140 representative of what we had for most of the plots.

141 In order to support the quality of each data set provided in our database, MOSAIC<sub>bioacc</sub> was used to fit TK models for all data  
142 sets under the same conceptual framework. Hence, the bioaccumulation metrics and TK model fitting plots are systematically

143 provided (Figure 3), as well as a collection of goodness-of-fit (GOF) criteria allowing to check the relevance of the results. All  
144 this information is gathered together in a PDF report downloadable from the MOSAIC<sub>bioacc</sub> database.

145 More than 95% of data sets of the MOSAIC<sub>bioacc</sub> database are already published. The reuse with MOSAIC<sub>bioacc</sub> allows  
146 to compare fitting results to those of the original study and to confirm the relevance of including the data in the database. In  
147 addition, most of the publications (25 among 28) do not provide uncertainty around predicted concentrations (at least not  
148 visualized in the original plots), while we provide, for each data set, the 95% credible intervals around median predicted  
149 concentrations, as well as uncertainties for all bioaccumulation metrics. It is an undeniable added-value we bring to the  
150 ecotoxicological community thanks to the revisit of all these data sets.

## 151 Usage Notes

152 Raw TK data are particularly useful to better understand the relationship between exposure chemical concentrations in the  
153 environment and their potential bioaccumulation within organisms. They are also used by researchers or regulators to evaluate  
154 the bioaccumulative property of chemicals then to establish the link with the life-history traits of effects on organisms. Our  
155 database also gives an easy access to formatted raw TK data that can be used to improve current TK models or to develop new  
156 conceptual modelling frameworks. This database is not an overview of all available TK data in the literature, but it encompasses  
157 several types of genus, substances, routes of exposure and elimination processes, giving a proof of concept for raw TK data  
158 storage and sharing. Our database aimed at developing and validating our generic TK modelling framework, as illustrated  
159 in this paper where the evaluation of the robustness of MOSAIC<sub>bioacc</sub> through the whole collection of data sets proved to be  
160 successful. In the past, evaluation and improvement of TK modelling frameworks have been limited due to the lack of raw  
161 TK data for a large variety of species and chemicals, especially for data considering biotransformation<sup>9</sup>, often more scarce  
162 than classic TK data. From our database, such evaluation and improvement should be facilitated and researchers are invited to  
163 contribute to this database by sharing their own data sets.

164 If bioaccumulation testing is required, it may be difficult to plan the experimental design since it depends on the substance,  
165 the species and the exposure route under consideration. Usually, a search in literature provides a possible range of exposure  
166 concentrations for the substance, but the amount likely to be bioaccumulated in organisms of the studied species remains  
167 unknown, as well as how long it will require to reach the steady-state at the end of the accumulation phase. Hence, the  
168 MOSAIC<sub>bioacc</sub> database provides a quick search option to identify similar genus, or to choose a substance with a similar mode  
169 of action to consider in support of new experiments to plan.

170 We sincerely hope this database will support further research to increase our knowledge on bioaccumulation of substances  
171 for which there are today clear societal challenges of environmental protection. Indeed, in ERA, bioaccumulation metrics are  
172 key decision criteria to classify a substance as non-bioaccumulative, bioaccumulative or very bioaccumulative. Most often,  
173 these metrics are calculated from two or three different modelling approaches, which makes their comparison difficult, in  
174 addition to be given only as a point mean value (without uncertainty). Our database should then improve this classification

175 thanks to the kinetic bioaccumulation metrics we provide from the same methodology for all data sets, that clearly quantify the  
176 uncertainty summarized by the 95% credible intervals around median values. More globally, our database should strongly  
177 encourage researchers to share their data with the scientific community by directly providing the raw data in their publications.

## 178 Code Availability

179 Data sets are available at MOSAIC<sub>bioacc</sub> (<https://mosaic.univ-lyon1.fr/bioacc>) or directly at [http://lbbe-shiny.univ-lyon1.fr/mosaic-bioacc/data/database/index\\_readme.html](http://lbbe-shiny.univ-lyon1.fr/mosaic-bioacc/data/database/index_readme.html). To reproduce all the outputs pro-  
180 vided in the reports for all data sets of the database, data can be downloaded from the database then uploaded in MOSAIC<sub>bioacc</sub>.  
181 In addition, for each data set, the corresponding R code can be downloaded if using the R software directly is the preferred  
182 option of the user.  
183

## 184 References

- 185 **1.** OCDE. Test no. 305: Bioaccumulation in fish: Aqueous and dietary exposure, <https://doi.org/10.1787/9789264185296-en> (2012).  
186
- 187 **2.** ECHA. Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.11: PBT/vPvB Assessment,  
188 [10.2823/128621](https://doi.org/10.2823/128621) (2017).
- 189 **3.** Schlechtriem, C., Fliedner, A. & Schäfers, C. Determination of lipid content in fish samples from bioaccumulation studies:  
190 contributions to the revision of guideline OECD 305. *Environ. Sci. Eur.* **24**, 13, [10.1186/2190-4715-24-13](https://doi.org/10.1186/2190-4715-24-13) (2012).
- 191 **4.** MacKay, D. & Fraser, A. Bioaccumulation of persistent organic chemicals: Mechanisms and models. *Environ. Pollut.* **110**,  
192 375–391, [10.1016/S0269-7491\(00\)00162-7](https://doi.org/10.1016/S0269-7491(00)00162-7) (2000).
- 193 **5.** Grech, A. *et al.* Toxicokinetic models and related tools in environmental risk assessment of chemicals. *Sci. The Total.*  
194 *Environ.* **578**, 1–15, <https://doi.org/10.1016/j.scitotenv.2016.10.146> (2017).
- 195 **6.** Landrum, P. F., Lydy, M. J. & Lee II, H. Toxicokinetics in aquatic systems: Model comparisons and use in hazard assessment.  
196 *Environ. Toxicol. Chem.* **11**, 1709–1725, <https://doi.org/10.1002/etc.5620111205> (1992). <https://setac.onlinelibrary.wiley.com/doi/pdf/10.1002/etc.5620111205>.
- 197
- 198 **7.** USEPA. ECOTOX User Guide: ECOTOXicology Knowledgebase System. Version 5.0, <http://www.epa.gov/ecotox/> (2020).
- 199 **8.** Ratier, A. *et al.* Brief communication: new perspectives on the calculation of bioaccumulation factors for active substances  
200 in living organisms. *bioRxiv* [10.1101/2020.07.07.185835](https://doi.org/10.1101/2020.07.07.185835) (2021). <https://www.biorxiv.org/content/early/2020/11/20/2020.07.07.185835.full.pdf>.  
201
- 202 **9.** Ratier, A., Lopes, C., Geffard, O. & Babut, M. The added value of bayesian inference for estimating biotransformation  
203 rates of organic contaminants in aquatic invertebrates. *Aquatic Toxicol.* **234**, 105811, <https://doi.org/10.1016/j.aquatox.2021.105811>  
204 (2021).



- 205 **10.** Rohatgi, A. Webplotdigitizer: Version 4.4, <https://automeris.io/WebPlotDigitizer> (2020).
- 206 **11.** R Core Team. R: A Language and Environment for Statistical Computing (2021).
- 207 **12.** Allaire, J., Horner, J., Xie, Y., Marti, V. & Porte, N. *markdown: Render Markdown with the C Library 'Sundown'* (2019).
- 208 R package version 1.1.
- 209 **13.** Xie, Y. *knitr: A General-Purpose Package for Dynamic Report Generation in R* (2021). R package version 1.31.
- 210 **14.** Xie, Y., Cheng, J. & Tan, X. *DT: A Wrapper of the JavaScript Library 'DataTables'* (2021). R package version 0.17.
- 211 **15.** OCDE. Test no. 317: Bioaccumulation in terrestrial oligochaetes, <https://doi.org/https://doi.org/10.1787/9789264090934-en>
- 212 (2010).
- 213 **16.** OCDE. Test no. 315: Bioaccumulation in sediment-dwelling benthic oligochaetes, [https://doi.org/https://doi.org/10.1787/](https://doi.org/https://doi.org/10.1787/9789264067516-en)
- 214 [9789264067516-en](https://doi.org/https://doi.org/10.1787/9789264067516-en) (2008).
- 215 **17.** Chen, W., Li, W., Ju, Y., Liao, C. & Liao, V. Life cycle toxicity assessment of earthworms exposed to cadmium-contaminated
- 216 soils. *Ecotoxicology* **26**, 360–369, [10.1007/s10646-017-1769-4](https://doi.org/10.1007/s10646-017-1769-4) (2017).

## 217 **Acknowledgements**

218 The authors are thankful to ANSES for providing the financial support for the development of the MOSAIC<sub>bioacc</sub> web tool.

219 This work was performed using the computing facilities of the CC LBBE/PRABI. This work benefited from the French GDR

220 “Aquatic Ecotoxicology” framework which aims at fostering stimulating scientific discussions and collaborations for more

221 integrative approaches. This work is part of the ANR project APPROVe (ANR-18-CE34-0013) for an integrated approach to

222 propose proteomics for biomonitoring: accumulation, fate and multi-markers (<https://anr.fr/Projet-ANR-18-CE34-0013>). This

223 work was also made with the financial support of the Graduate School H2O’Lyon (ANR-17-EURE-0018) and “Université de

224 Lyon” (UdL), as part of the program “Investissements d’Avenir” run by “Agence Nationale de la Recherche” (ANR). The

225 authors are also thankful to Nina Ceredergreen, Frédéric Gimbert and Roman Ashauer to their help in providing raw TK data.

## 226 **Author contributions statement**

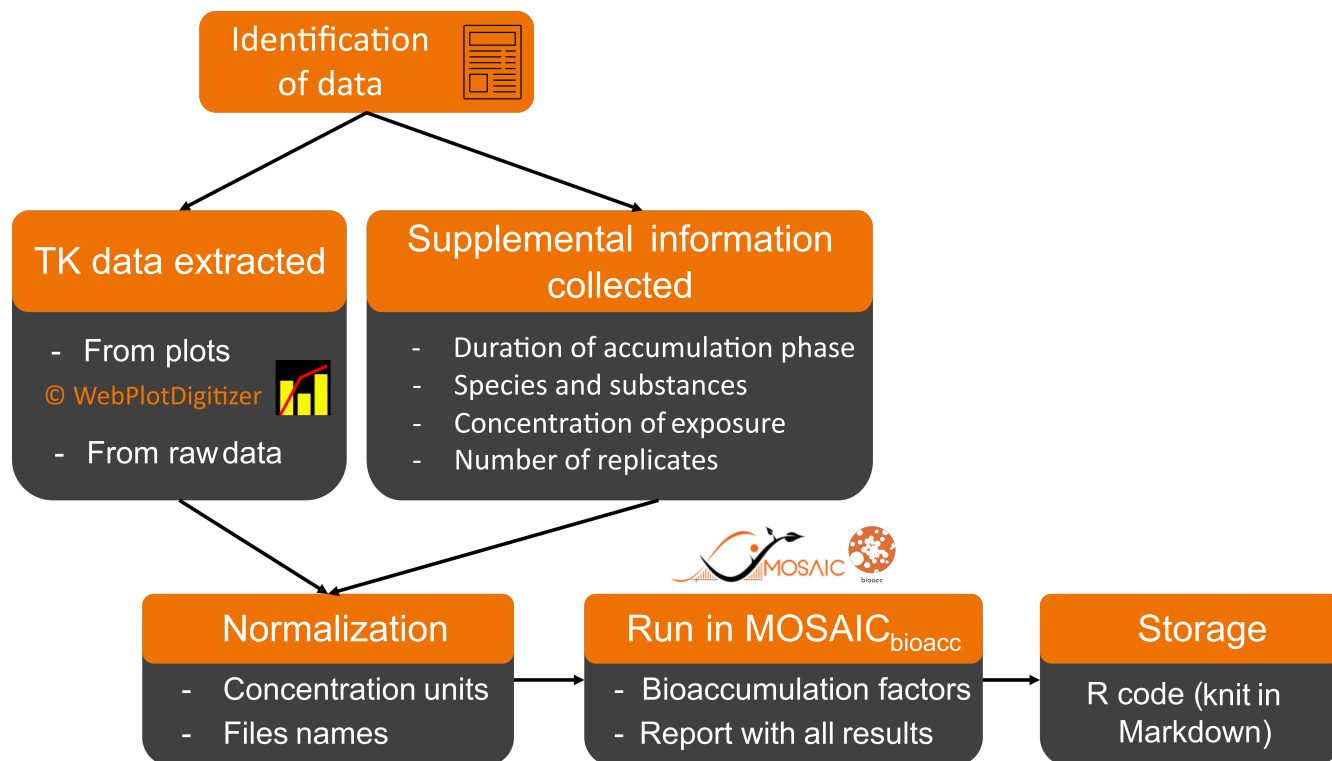
227 A.R. and S.C. participated to the collection of data sets, analysed the results in MOSAIC<sub>bioacc</sub> and elaborated the database. All

228 authors reviewed the manuscript.

## 229 **Competing interests**

230 The authors declare no competing interests.

## 231 **Figures & Tables**



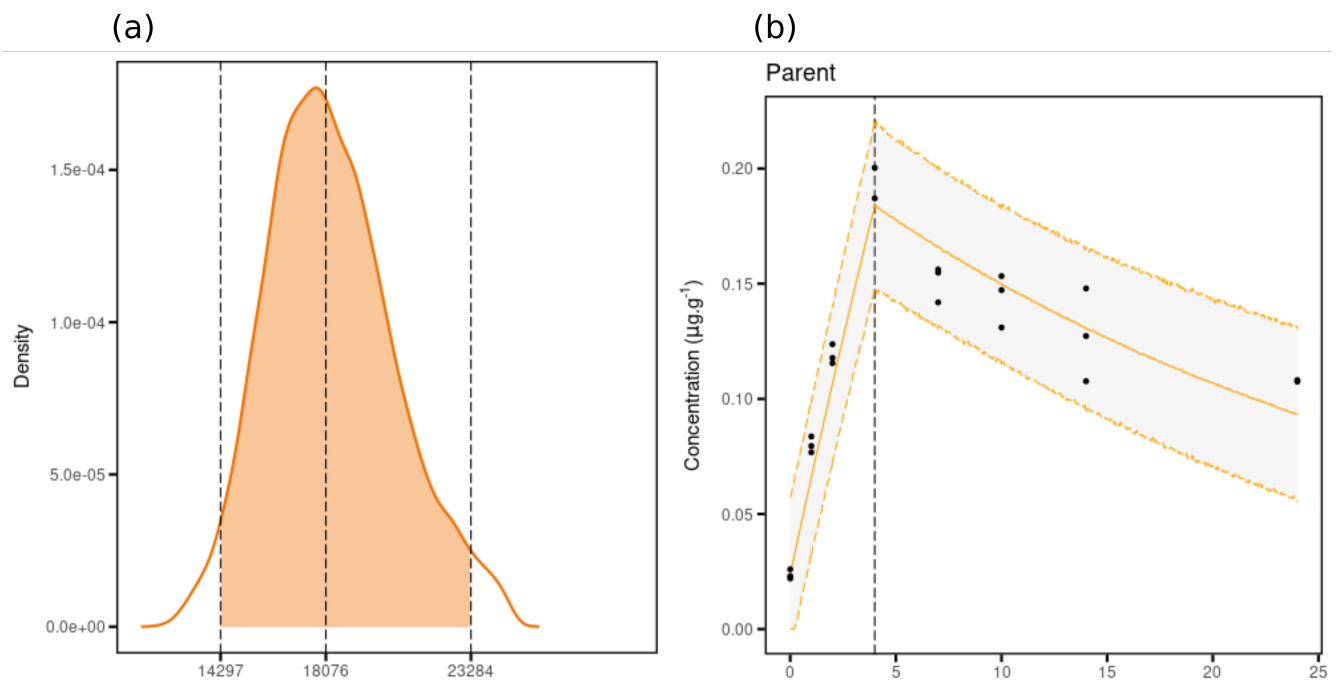
**Figure 1.** Conceptual framework of the collection of TK data and their storage in the database of MOSAIC<sub>bioacc</sub>.

dataset	genus	category	substance	duration of accumulation	exposure route	n (replicate)	Kinetic bioaccumulation factor	report	reference	comments
<a href="#">Asellus_cypermethrine_24h_Dalhoff2020.txt</a>	<i>Asellus</i>	aquatic invertebrate	cypermethrine	24 hours	water	16 (2)	10109 [9283 - 10988]	<a href="#">Link</a>	<a href="#">Dalhoff et al. (2020)</a>	
<a href="#">Axinella_Ag_7.04d_Genta2012.txt</a>	<i>Axinella</i>	seawater sponge	Ag	7.04 days	water	66 (3)	193 [118 - 281]	<a href="#">Link</a>	<a href="#">Genta-Jouve et al. (2012)</a>	
<a href="#">Axinella_Hg_7.04d_Genta2012.txt</a>	<i>Axinella</i>	seawater sponge	Hg	7.04 days	water	66 (3)	159 [35 - 212402]	<a href="#">Link</a>	<a href="#">Genta-Jouve et al. (2012)</a>	

**Figure 2.** Screenshot of a sample from the database available from MOSAIC<sub>bioacc</sub>.

**Table 1.** Summary of collected data. The classification of bioaccumulative properties are the same as in ECHA (2017)<sup>2</sup>.

Genus (n = 25)	Substances (n = 64)	Exposure route	Elimination process	Kinetic bioaccumulation metric (median)	Number of studies
aquatic invertebrate (n = 51) fish (n = 21) insect (n = 5) aquatic worm (n = 5) terrestrial worm (n = 3) seawater sponge (n = 2) seawater plant (n = 1) aquatic algae (n = 1) terrestrial invertebrate (n = 1)	metals (n = 9) PCB (n = 5) pesticides (n = 23) flame retardants (n = 3) pharmaceutical products (n = 2) hydrocarbons (n = 14) octyphenols (n = 1) polyesters (n = 2) nanoparticles (n = 5)	water (n = 57) food (n = 21) sediment (n = 12)	biotransformation (n = 18) growth dilution (n = 0)	< 1000 (n = 75) 1000 < BAF < 2000 (n = 4) 2000 < BAF < 5000 (n = 3) > 5000 (n = 8)	28



**Figure 3.** Example of results given by MOSAIC<sub>bioacc</sub>: (a) the bioaccumulation metric and (b) the fitted TK model predictions with observed data. This example comes from the data set 'Male\_Gammarus\_Single.txt' available from MOSAIC<sub>bioacc</sub>.