KWR 2020.125 | October 2020

Assessment of requirements for the NORMAN Bioactivity database

KWR

Bridging Science to Practice

Collaborating Partners

Report

Assessment of requirements for the NORMAN Bioactivity Database

KWR 2020.125 | October 2020

Project number
403214/001/002

Project manager
Ton van Leerdam

Client
Norman Network

Author(s)
T.E. Pronk, T. Hamers, M.M.L. Dingemans, B.I. Escher, P.A. Neale

Quality Assurance
Marja Lamoree (KWR)

Sent to
Valeria Dulio (NORMAN)

Keywords

Bioassay, Norman, Database, Bioactivity

Year of publishing
2020

More information
dr. ir. T.E. Pronk
T +31 629337806
E tessa.pronk@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl

Contents

1 Introduction

The work described in this report, to assess the requirements for a new NORMAN Bioactivity database, was funded by Norman following a proposal from VU and KWR. A project team was formed in 2020. This team was a collaboration between VU, KWR, UFZ, Griffith. The people participating in this project team are the authors of this report. In this document an overview of considerations on the functionality and requirements for the NORMAN Bioactivity database is given. Some of the considerations are from the project team, others come from the questionnaire that was sent out to potential users.

Goal and scope of the database

In the envisioned NORMAN Bioactivity database, activity/toxicity data for individual water relevant chemicals will be collected from bioassays that measure endpoints relevant for water quality monitoring, or close analogues of those bioassays (e.g. ToxCast assays). This information can be used for different water quality related activities, such as mixture modelling, selection of a relevant bioassay battery for water quality monitoring, prioritization of chemicals for inclusion in monitoring programs, and prioritization of suspects for confirmation.

2 Context: Similarity to other databases

Norman currently hosts several databases in the Norman Database System for the collection & evaluation of data / information on emerging substances in the environment, see https://www.norman-network.com/nds/. There are two existing databases in the Norman database collection that contain useful functionality relevant for the envisioned NORMAN Bioactivity database:

2.1 Ecotoxicology database

This database is substance based. The goal of the database is to provide PNEC values which are derived using a harmonized approach and agreed upon by a large group of experts. The Predicted No Effect Concentration (PNEC) is the concentration of a chemical which marks the limit at which below no adverse effects of exposure in an ecosystem are measured. This is experimental or predicted by QSAR.

The Ecotoxicology database consists of the following modules:

- SEARCH ecotox data – Selection of reference studies for PNEC derivation;
- CRED evaluation – Quality check of selected reference studies;
- PNEC derivation – Automated derivation of PNEC values;
- Lowest PNEC – Voting system for selection of the Lowest PNEC value for each substance contained in the database;
- Statistics – Final list of Lowest PNECs for all substances in the database

The following fields are mandatory at submission:

7	Editor
8	Date
11	Title
12	Author(s)
13	Year
14	Bibliographic source
16	Compartment
17	Test type
20	Substance name
21	CAS No.
24	Purity [%] of test substance
36	Effect measurement
37	Endpoint
38	Test duration
42	Scientific name
63	Measured or nominal concentrations used?
121	Effect concentration qualifier
122	Effect concentration
134	Existing reliabilty score
135	Reliability score system used
140	Affiliation issuing the reliabilty score
136	Existing rationale for reliability

Example output:

Ecotox data for: **Atraton**

Freshwater chronic

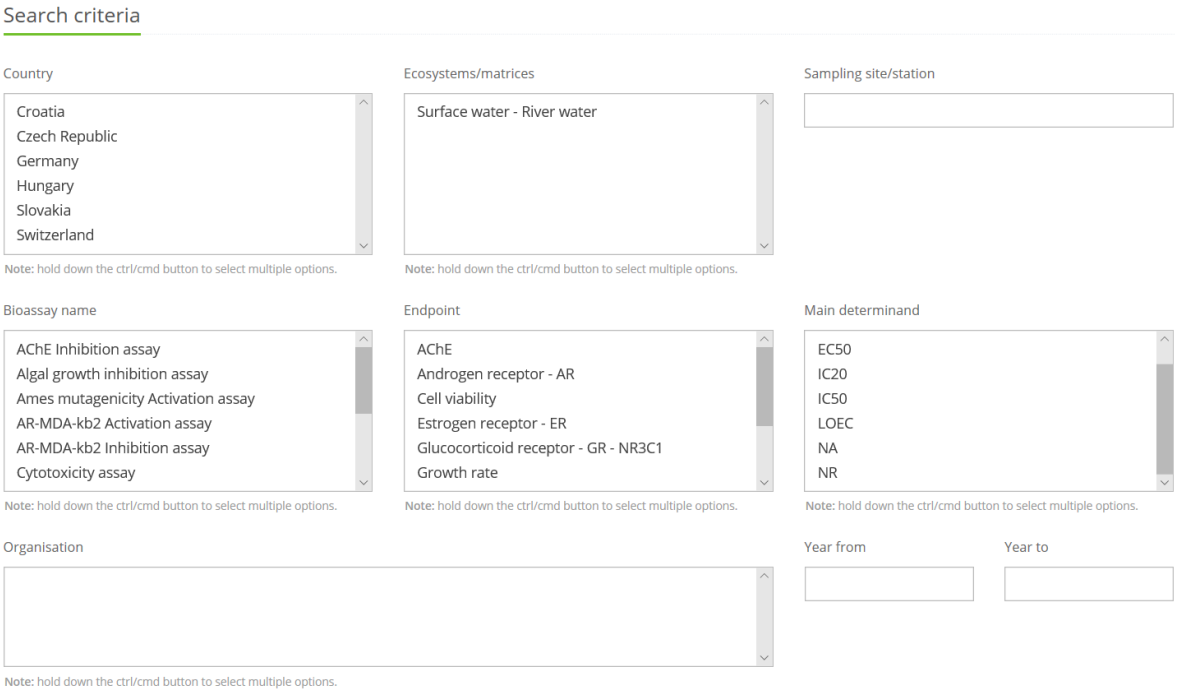
Search:

Biotest ID	Taxonomic group	Scientific name	Endpoint	Duration	Effect measurement	Test type	Standard test	Effect based on	Purity [%]		Effect value [µg/L]	Reference	Measured or nominal
	All	All	All	All	All	All	All	All	All		All	All	All
EPA2357853	Algae	*Chlorella fusca var. vacuolata*	EC50	24 h	populationgrowthrate	experimental result	no	active ingredient	= 95	=	79.9	EPA62304	measured (not specified)
EPA2357854	Algae	*Chlorella fusca var. vacuolata*	EC01	24 h	populationgrowthrate	experimental result	no	active ingredient	= 95	=	1.5	EPA62304	measured (not specified)
EPA2357855	Algae	*Chlorella fusca var. vacuolata*	NOEC	24 h	populationgrowthrate	experimental result	no	active ingredient	= 95	=	12	EPA62304	measured (not specified)

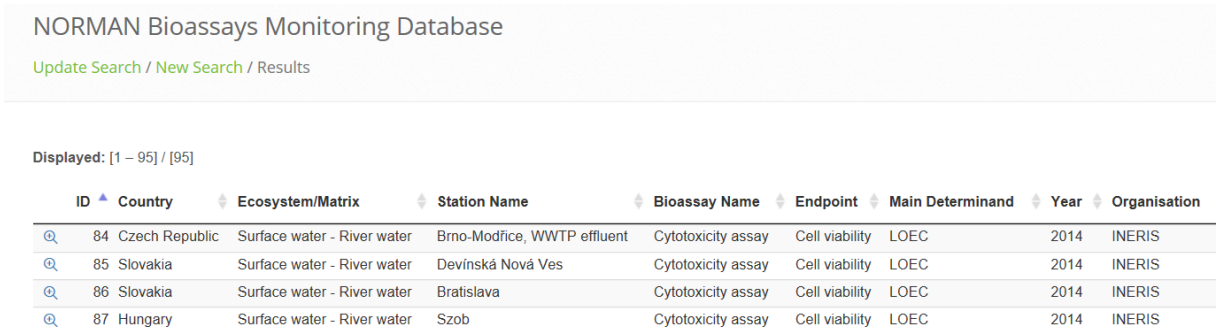
2.2 Bioassay database

This database is bioassay based. The goal of this database is to provide data obtained by analysis of environmental samples (monitoring) with bioassays. Metadata on the datasets is included.

Search window:



Example output:



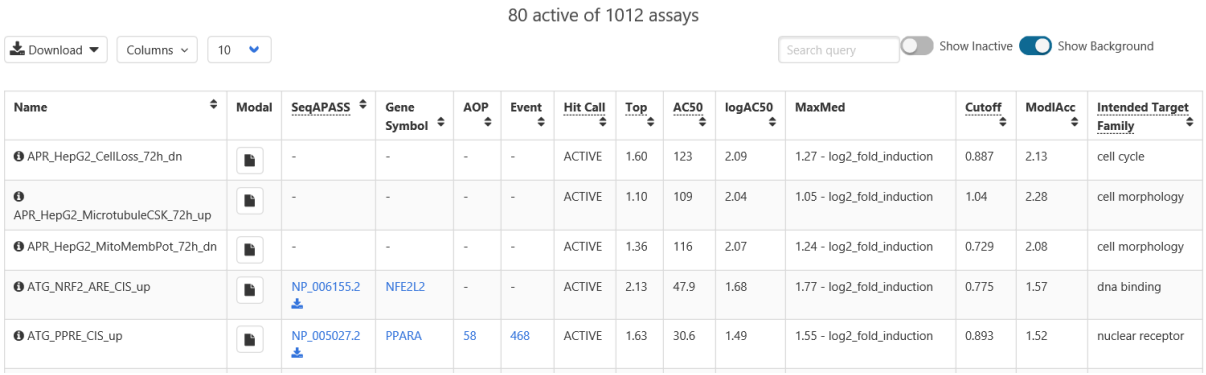
ID	Country	Ecosystem/Matrix	Station Name	Bioassay Name	Endpoint	Main Determinand	Year	Organisation
84	Czech Republic	Surface water - River water	Brno-Modřice, WWTP effluent	Cytotoxicity assay	Cell viability	LOEC	2014	INERIS
85	Slovakia	Surface water - River water	Devínská Nová Ves	Cytotoxicity assay	Cell viability	LOEC	2014	INERIS
86	Slovakia	Surface water - River water	Bratislava	Cytotoxicity assay	Cell viability	LOEC	2014	INERIS
87	Hungary	Surface water - River water	Szob	Cytotoxicity assay	Cell viability	LOEC	2014	INERIS

Per entry, there is descriptive information on the bioassay, the test site, the experimental conditions, and the data source (institute, project).

2.3 Comptox Toxcast EPA database

The Toxcast program was designed to rank and prioritize chemicals. The Toxcast database consists of chemicals having undergone some level of screening in EPA's Toxcast research program from 2007 to the present (last updated 4/11/2017). Toxcast also includes EPA's full, plated contribution of nearly 4000 unique chemicals to the multi-federal agency Tox21 program (TOX21SL). A publication detailing the construction and composition of the Toxcast inventory (Richard et al., Chem. Res. Toxicol. 2016) can be freely downloaded from: http://pubs.acs.org/doi/abs/10.1021/acs.chemrestox.6b00135 (this text is adapted from: https://comptox.epa.gov/dashboard/chemical_lists/TOXCAST)

Example: https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID8031865#invitrodb



80 active of 1012 assays

Download Columns 10 Search query Show Inactive Show Background

Name	Modal	SeqAPASS	Gene Symbol	AOP	Event	Hit Call	Top	AC50	logAC50	MaxMed	Cutoff	ModlAcc	Intended Target Family
APR_HepG2_CellLoss_72h_dn		-	-	-	-	ACTIVE	1.60	123	2.09	1.27 - log2_fold_induction	0.887	2.13	cell cycle
APR_HepG2_MicrotubuleCSK_72h_up		-	-	-	-	ACTIVE	1.10	109	2.04	1.05 - log2_fold_induction	1.04	2.28	cell morphology
APR_HepG2_MitoMembPot_72h_dn		-	-	-	-	ACTIVE	1.36	116	2.07	1.24 - log2_fold_induction	0.729	2.08	cell morphology
ATG_NRF2_ARE_CIS_up		NP_006155.2	NFE2L2	-	-	ACTIVE	2.13	47.9	1.68	1.77 - log2_fold_induction	0.775	1.57	dna binding
ATG_PPRE_CIS_up		NP_005027.2	PPARA	58	468	ACTIVE	1.63	30.6	1.49	1.55 - log2_fold_induction	0.893	1.52	nuclear receptor

Comptox Toxcast has labels for each Toxcast assay endpoint in Adverse Outcome Pathways (AOP) and Event (specific part in AOP). Also, it has a column Hit call that indicates if the compound was active or not. The Target family is also indicated.

3 Requirements for database functionality

3.1 Minimum information for data entries

The envisioned NORMAN Bioactivity database will expand by the submission of measured activity/toxicity data from users. Submission can take place via a customized upload form to upload data, as with other NORMAN databases. The amount of information asked will have to be, as with all databases, a balance between options for extra use and effort of submission. Preferably, automatic filling of cells occurs and drop down lists are provided where possible. We propose to ask submitters to fill in the NORMAN parameters, as indicated in the supplementary table (Supplement Figure S1).

3.2 Search functionality

There is a need for elaborate (faceted) search functionality to select relevant data in the database. A possibility for download of results via spreadsheet would be required. Based on the combined requirements of users (see questionnaire results in Chapter 4) and the team, a concept search functionality is given in Figure 1.

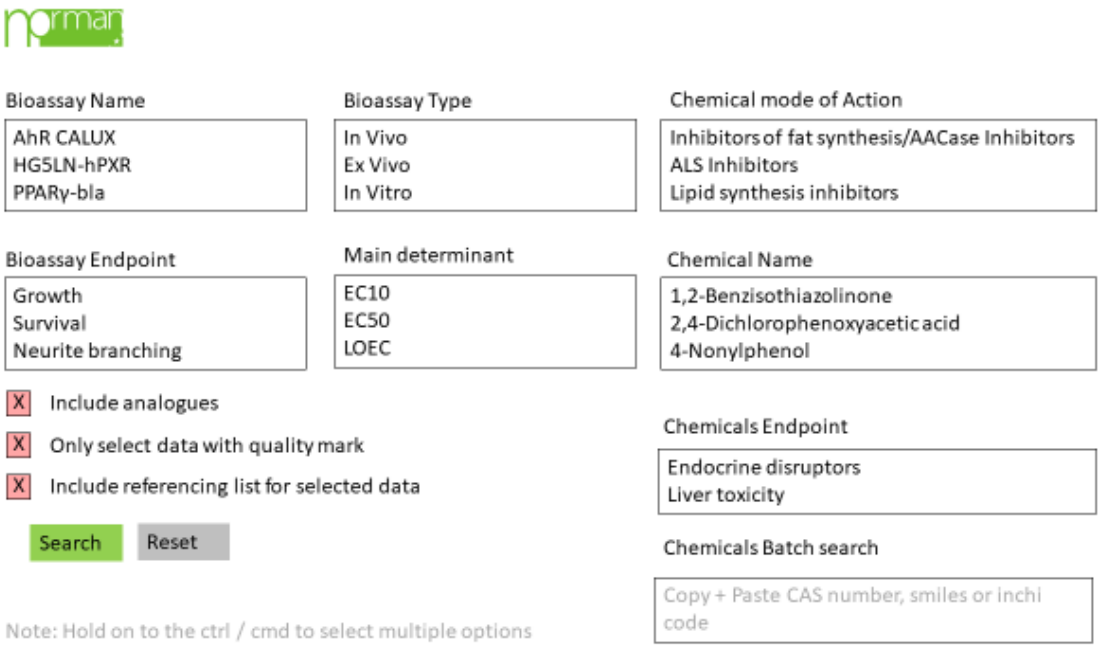


Figure 1. A concept for search and select functionality for the Bioactivity database

For a concept submission form, see Table S1 in the Excel supplement: Table S1. Supplement. Excel sheet with fields (mandatory, optional, automatic) to fill when submitting data. Fields are based on the NORMAN Ecotox DB. A distinction is made between fields for in vivo (ex vivo) or in vitro bioassays. In red are newly proposed fields.

3.3 Due credit

The project team would like to see the contributors to the database to get due credit for their work. When downloading data, a suggestion for referencing can be included in the output. Or, a window showing: your selection contains data from: - list of institutes and research groups - .

Referencing can be done to a publication that is relevant, an institute that uploaded the data or to a database where the data is deposited (with a persistent identifier / url).

3.4 Compliance of elements of the other databases to requirements of the NORMAN Bioactivity database

The goal of the Ecotoxicity database is to obtain PNECs for chemicals in ecological systems. The data is logically limited to in vivo systems. This is not the goal of the Bioactivity database, in which on overview of the activity of substances per bioassay (both in vivo, ex vivo and in vitro) is one of the goals. The Ecotoxicity database however has information on individual substances, which is also a requirement for the Bioactivity database. For this reason the information of the Ecotoxicity database could be fit *in* the Bioactivity database.

The search functionality as it is presented in the Ecotoxicity database is however incomplete for the purposes of the Bioactivity database. The search functionality limits itself to data per compound while the Bioactivity database needs also data per bioassay, or endpoint.

The Bioassay database has a search functionality that resembles more the optimal requirements of the Bioactivity database. However, search for data on compounds is not possible because this database only includes monitoring data.

The Comptox database has labels for the bioassays in Adverse Outcome Pathways (AOP) and Event (specific part in AOP). Also, it has a column Hit call that indicates if the compound was active or not. The Target family is also indicated. It would be recommendable to have these items also in the Bioactivity database, for easy integration of results between the two databases.

In short, for the NORMAN existing databases the data structure should resemble the Ecotoxicity database, the search functionality should resemble the Bioassay database, with some metadata elements (and possibly even data points) of the Comptox database.

4 Requirements of prospective users: Summary of the Bioactivity database questionnaire

In order to get additional user requirements for the Bioactivity database, a questionnaire was made consisting of six questions on the use and requirements of the database. The questionnaire was initially send to the e-mail list of NORMAN. It was requested in the questionnaire to forward the e-mail or link to other interested persons. The questionnaire was open from 13 May 2020 to 10 June 2020. In total, 45 persons replied. A total of 33 persons left their e-mail address for further contact.

Some important points from the project team on the questionnaire outcome:

- A lot of respondents generate data that could be included
- Most respondents are respondents that think they can use the NORMAN Bioactivity database in their work
- Possibility for integration / aligning with the Comptox Toxcast data would be good
- Ability to restrict data is not often required
- All search functionalities mentioned are welcomed and more are suggested
- All information mentioned is welcomed and more are suggested
- Quality of the data is important
- Some respondents offer help to contribute

Guiding and closing text of the NORMAN Bioactivity database Questionnaire

To give more context on what was asked of respondents in the questionnaire, we give the guiding and closing tekst that was used in the questionnaire here:

“With this questionnaire we would like to collect your needs and opinions to consider in the construction of a new NORMAN Bioactivity database. For other, existing databases in the Norman Database System, please visit https://www.norman-network.com/nds/.

In the envisioned NORMAN Bioactivity database, activity/toxicity data for individual water relevant chemicals will be collected from bioassays used in water quality monitoring, or close analogues thereof. This information can be used for different water quality related activities, such as mixture modelling,

selection of a relevant bioassay battery and prioritization of chemicals for inclusion in monitoring programs, and of suspects for confirmation.

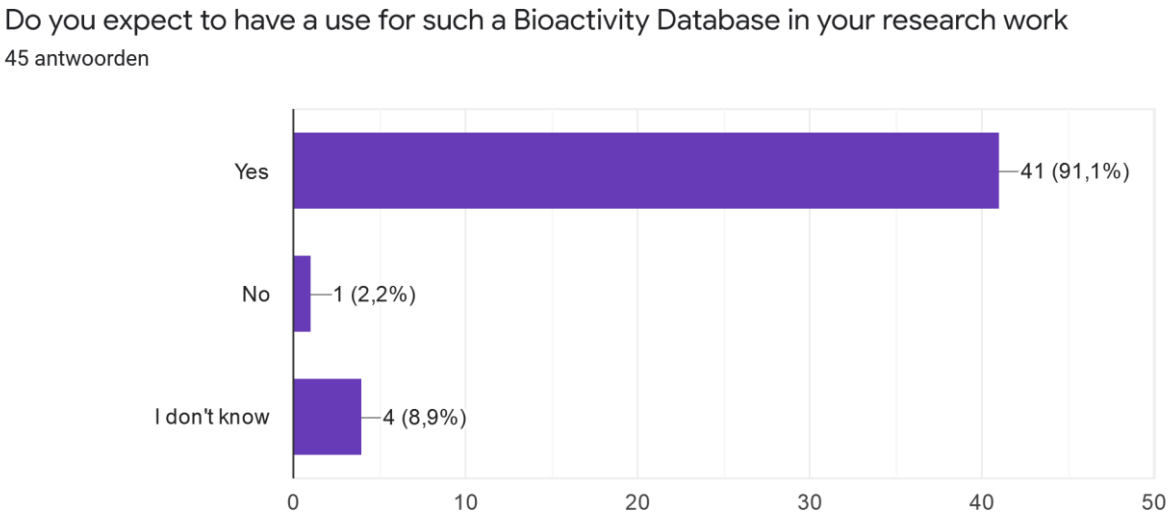
It is foreseen that the envisioned NORMAN Bioactivity database will expand by the submission of newly developed activity/toxicity data from users.

You can leave your e-mail address below if you allow us to reach out to you for follow-up questions."

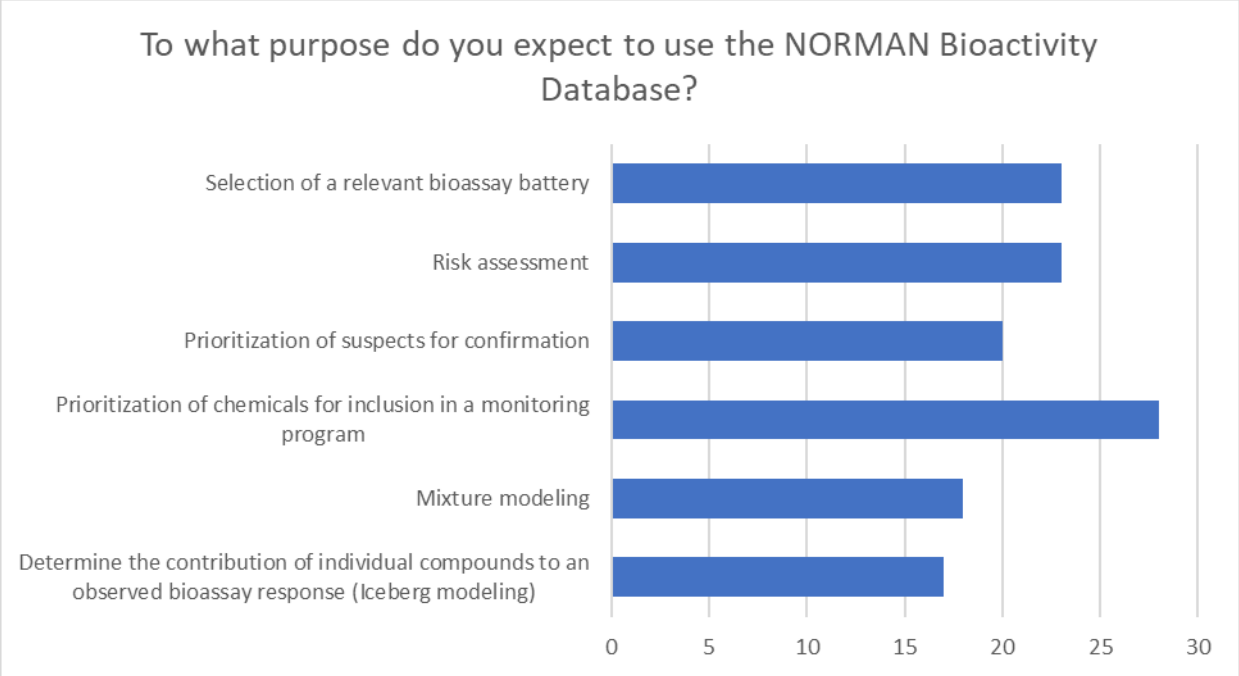
"Please note that we will present an anonymized overview of responses in a brief report to NORMAN. Your contact details will not be shared with any other persons, aside from the persons below who are responsible for setting up a concept version of the NORMAN Bioactivity database.

Thank you very much for filling in this questionnaire. Please share this link with anybody that may be interested. With high regards, Timo Hamers (VU, Amsterdam), Tessa Pronk (KWR Water Research Institute), Milou Dingemans (KWR Water Research Institute), Beate Escher (UFZ), Peta Neale (Griffith University)"

Question 1.



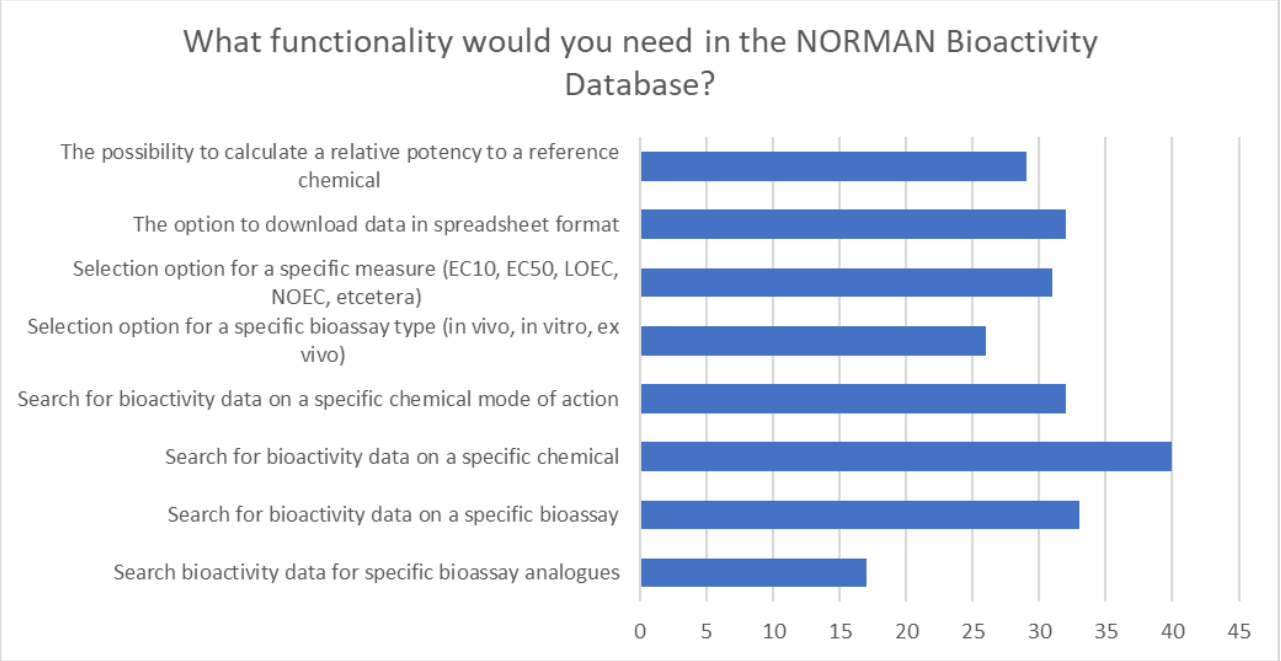
Question 2.



Extra answers:

- Teaching all of the above
- AOP development
- Calibration of prospective risk to bioassay-derived field effects and biomonitoring patterns
- Compounds co-exposure effects
- Dissimination of knowledge
- Explanation of an observed bioactivity in environmental samples
- Finding activity of chemicals in bioassays
- integration, analysis, data mining

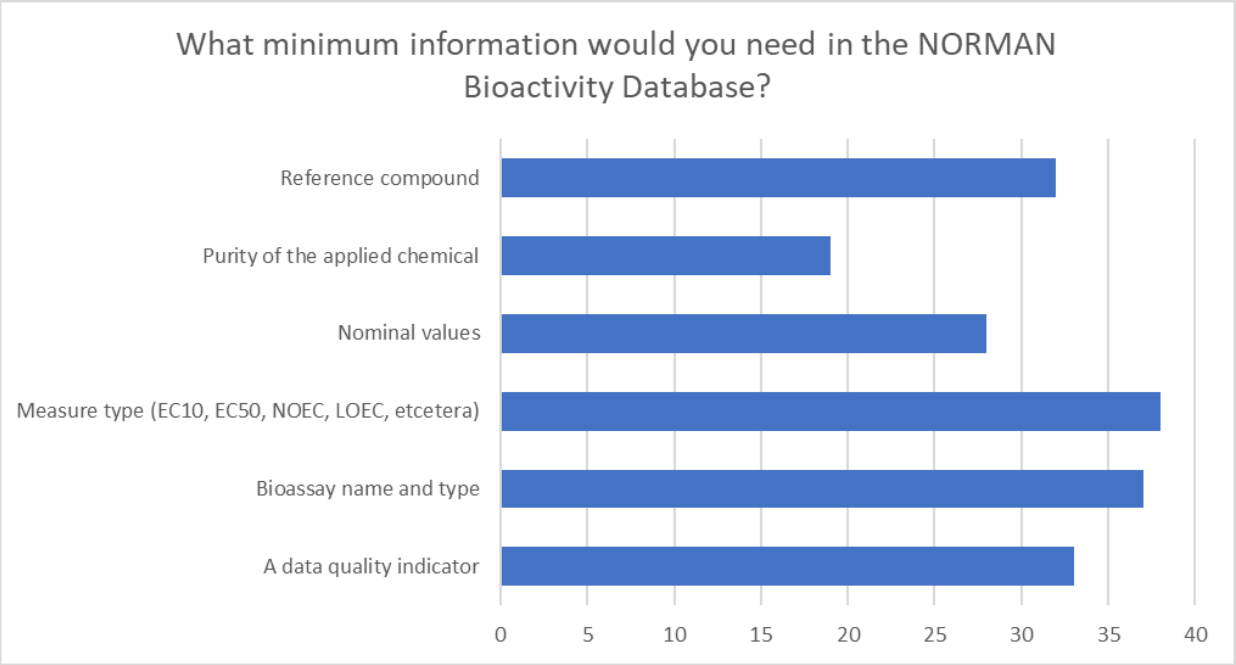
Question 3.



Extra answers:

- Access to info on culture conditions, exposure, dose response and analysis methods used
- Application programming interface to retrieve data with own workflows (e.g. written in R or Python)
- Categories based on applicability relating to EU regulations (e.g. Watch List parameters & WFD Priority substances)
- Comparison to other bioassays (as is possible)
- Extract bioactivity information (LOEC) per substance per bioassay
- search for multiple chemicals => and/or ability to integrate this into other NORMAN efforts (e.g. NORMAN-SLE) so I could work with PubChem on integrating this data
- Search, sort by
- Selection on chemicals

Question 4.



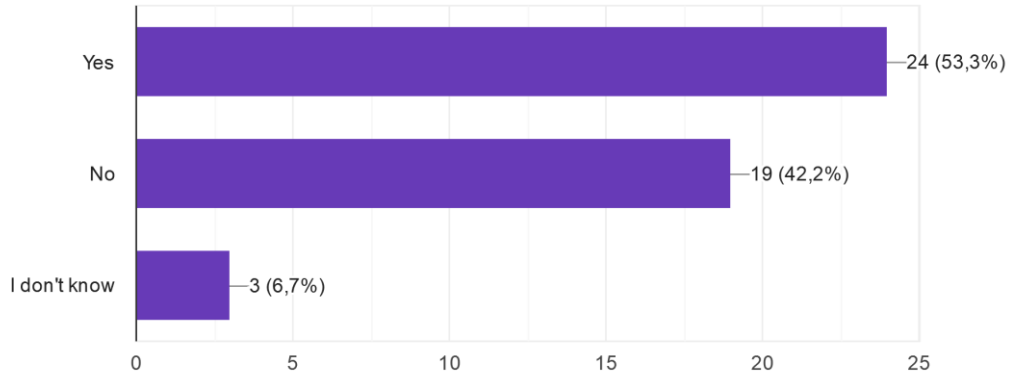
Extra answers:

- Availability of full concentration-response data rather than single endpoints only. It should be possible to recalculate with own methods
- Open Access Data under FAIR conditions
- Bioactivity of individual waterrelevant chemicals in bioassays from SIMONI and DEMEAU
- Chemical name (IUPAC), CAS number, molecular weight, log Kow, water solubility, type of bioassay, quantitative data (e.g. LOEC, EC50) if available, otherwise qualitative data (+,-,+/-)
- Detailed method description
- Exposure duration and name of species tested
- If available, a link to published report where more detailed information is provided (e.g. dose-response profile and modeling, biological model specific information, etc)
- Important to know the quality criteria used for the data entered in the database. Access to source and ideally raw data would also be really helpfull!
- Measured concentrations, if any
- Number of replicates
- Real measured concentrations in case of single compound testing rather than nominal concentrations
- Reference to dataset, publication, lab, responsible researcher (traceability)
- Sufficient chemical information for chemists to use (happy to contribute to help provide this)

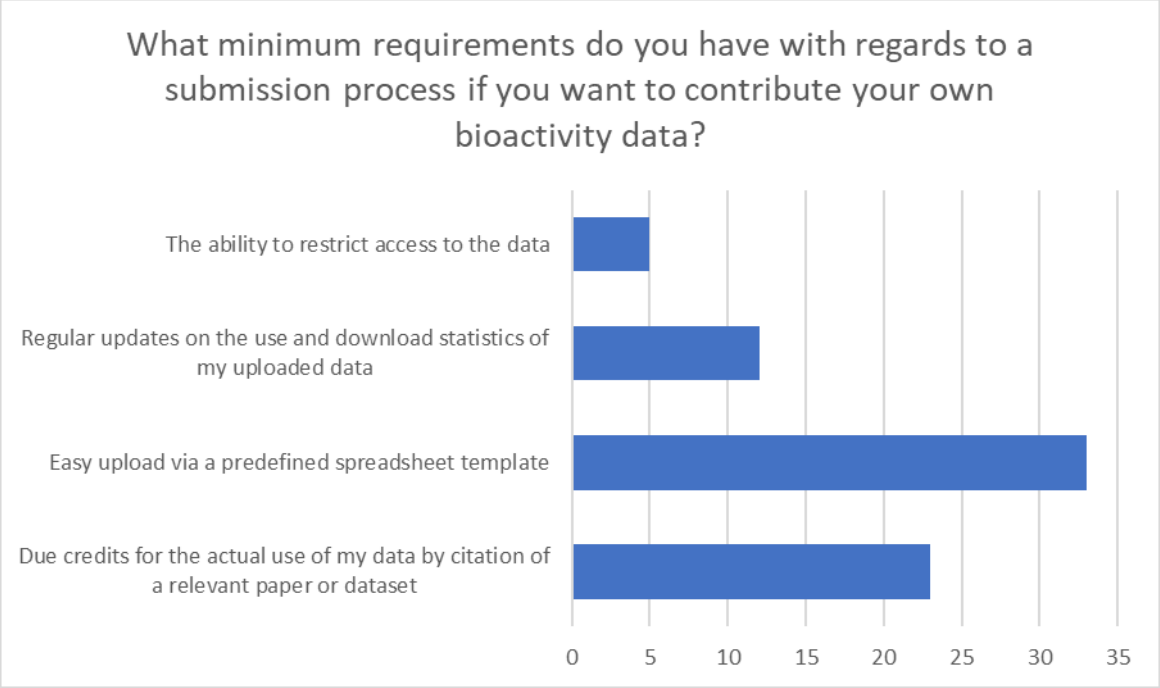
Question 5.

Do you generate activity/bioactivity data for individual chemicals in your research that would be fit to include in the NORMAN Bioactivity Database?

45 antwoorden



Question 6.



Extra answers:

- Embargo options should be possible
- You should consider to submit single datasets to zenodo or so (like Suspect List Exchange) to provide a smart way for dataset citation.
- A template file and quality check by independent reviewer
- an independent quality check
- Data input should be checked by a colleague, if experimental data is included permission for use should be asked by the owner
- FAIR and Open Access
- Make submission as easy as possible by automating as much as possible, this could e.g. be done for chemical information as long as certain fields are provided in the template
- minimal information standards, define limits of assay, provide other metadata about bioassay
- Review of the data before upload
- Sufficient meta-data should be added (upload date, data generating lab, etc.)

Extra Remarks from the questionnaire:

- Compiling such a dataset requires very strict quality control and a check for plausibility of test results
- Good initiative!
- This initiative is great! It would be really helpful to us. It looks like there is some similarity with Comptox (https://comptox.epa.gov/dashboard/) ToxCast 21, although this initiative will definitely have added value (e.g. other bioassays included). It would be handy if the use of both databases would be made easy, e.g. that you can download data from the Norman bioactivity database in a similar table format as provided by Comptox, so that you can easily paste data for one compound from both databases together.
- Great initiative! I really hope it will be realized and useful. It would be great to have such a database also for bioactivity data for actual environmental samples. We have one in house for Swedish soil samples.
- Preferably, the possibility to calculate relative potency towards a reference compound should not be restricted to a predefined reference compound. It would be good if each user could select his/her own reference chemical, maybe with an asterisk indicating which compounds have been suggested as reference chemicals in previous studies.
- Regarding the minimum information requirements, this is really a minimum. All the other suggested types of information (purity, data quality, etc) are highly appreciated, of course.
- Is it possible to expand an existing NORMAN database?
- Open access is important
- Excellent initiative. I presently use the USEPA ECOTOX database and ToxCast, but spend a lot of time on ensuring data is of proper quality. Hence, quality assurance or at least explicite measures on how quality is assessed would be really helpful.

5 Next steps to take

This report and the table in the supplementary material, with proposed variables are a concrete starting point for development of the Bioactivity database. In order to take these ideas further. We list further steps here.

- Present the results in the working group WP2
- Based on the current report and feedback of the working group construct a proposal (November 2020) before the general assembly to obtain resources for next steps
- Discuss the proposal in the NORMAN General Assembly

5.1 Further development of the NORMAN Bioactivity database

- Decide on the properties of a prototype and discuss the building of it with a database expert (contact Jaroslav Slobodnik)
- Fill the prototype with case study data (this could be the data from an experiment of Beate Escher, UFZ, 100 compounds)
- Contact respondents of the questionnaire for contributing data for the prototype

- Test round of the prototype
- Further, detailed requirement analysis:
    - Users indicate that the Comptox Toxcast database is already used. It would be good not to have two different go-to places:
      Investigate how to link or include Comptox Toxcast database data
    - One of the functionalities of the Comptox Toxcast database is to view the effect size with concentration:
      Investigate possibilities for raw data inclusion and dashboard visualizations of these data
    - In the Ecotoxicity database there is not a possibility for users to autonomously upload data:
      Investigate self-upload possibilities for contributing persons, and do a user experience study to find how the filling in is in practice.
    - In different experiments, different values are found:
      Investigate options for majority vote to settle on value for range of results
    - A requirement could be to have reference compound information if this is not given in a particular experiment:
      Investigate the option to link reference compounds between tests
    - In the user inquiry there was a request for an indication of quality of the data:
      Investigate what fields and values can represent 'quality' in the reliability score for the Bioactivity database
    - Many people have in vitro data on environmental samples, and no place to put results*:*
      Investigate if a separate database in in vitro environmental samples would be worthwhile to complement the Bioactivity database
    - Some institutes may have the opportunity to provide in-kind contribution:
      Ask for in-kind contribution of institutes for building and filling in the prototype
    - The reference compound may have a different cytotoxicity endpoint:
      Investigate if this should be described in a separate uploaded experiment

I Supplement: Screenshot example Excel upload form

In red are additions to the Norman Ecotoxicity database. The colums 'in/ex vivo' and 'in vivo' are suggested fields for the NORMAN Bioactivity database.

Table 1 #	Table 1 Header	Table 1 / 2 Description	Norman Ecotoxity DB	in/ex vivo	in vitro
147	Ecotox DS ID	A system-generated ID to identify upload dataset (DS) with ecotox data.	Autom.	Autom.	Autom.
1	Biotest ID	A system-generated ID to identifiy each ecotox endpoint.	Autom .	Autom.	Autom.
2	Data source	Name of the source where the data was taken from, e.g. the "UBA Etox" or "EPA Ecotox" databases, or "Experimental results" if new.	-	Mandatory	Mandatory
3	Data source ID	ID to identify the data entry at the source.	-	Optional	Optional
4	Data source reference ID	ID to identify the reference at the source.	-	Optional	Optional
5	Data protection status	Indicates whether the study is confidential or publically available.	NORMAN	NORMAN	NORMAN
6	Data source link	URL or website, DOI, pdf-link to the original study, entry in database, etc.	Mandatory	Mandatory	Mandatory
7	Editor	Editor who created or modified the record.	Mandatory	Mandatory	Mandatory
8	Date	Date representing the day the record was created or modified.	Mandatory	Mandatory	Mandatory
xx	Laboratory performing	Institute or laboratory name that performed the study		Optional	Optional
xx	Laboratory publishing	Institute or laboratory name that performed the study		Mandatory	Mandatory
9	Reference type	Type of the reference of the source document: e.g. literature study, GLP study report, grey literature, Risk Assessment Report, etc.	NORMAN	Mandatory	Mandatory
10	Reference ID	A system-generated ID to identify the reference of the study.	Autom.	Autom.	Autom.
11	Title	Title of the publication	Mandatory	Mandatory	Mandatory
12	Author(s)	All authors names of the publication	Mandatory	Mandatory	Mandatory
13	Year	Year of the publication	Mandatory	Mandatory	Mandatory
14	Bibliographic source	Journal or publisher name + city of publisher + Issue and pages	Mandatory	Mandatory	Mandatory
16	Compartment	Tested compartment (matrix), e.g. freshwater, marine water, freshwater sediments, marine sediments, etc.	-	Optional	Not applicable
17	Test type	Indicates whether the reported value is experimental or estimated.	-	Not applicable	Not applicable
18	Acute/Chronic	Indicates whether the test is considered to be short term or long term.	NORMAN	NORMAN	Not applicable
19	NORMAN Substance ID	NORMAN IDY used in NORMAN databases (e.g. SUSDAT).	Autom.	Autom.	Autom.
xx	DSS ID	Comptox ID of compound		Autom.	Autom.
20	Substance name	Substance name used in the study.	Mandatory	Mandatory	Mandatory
21	CAS Number	Substance CAS number used in the study.	Mandatory	Mandatory	Mandatory
22	EC Number	EU number used in the study, in cases where no CAS exist.	NORMAN	NORMAN	NORMAN
xx	Molecular weight test substance	Molecular weight in g/mol		Autom.	Autom.
23	Purity qualifier	Qualifier for the % purity value, e.g. >, =, <.	Mandatory	Mandatory	Mandatory
24	Purity [%] of test substance	Reported purity of test material in %.	Mandatory	Mandatory	Mandatory
25	Supplier of test item	Source / Merchant, where the test material was purchased.	-	Optional	Optional

xx	Reference compound name	Reference compound used in the study		Not applicable	Mandatory
xx	CAS Number Reference Compound	Reference Compound CAS number used in the study.		Not applicable	Optional
xx	EC Number Reference Compound	EU number Reference Compound used in the study, in cases where no CAS exist.		Not applicable	Optional
xx	Molecular weight reference compound	Molecular weight in g/mol		Autom.	Autom.
xx	Purity qualifier Reference Compound	Qualifier for the % purity value of Reference Compound, e.g. >, =, <.		Not applicable	Optional
xx	Purity [%] of Reference Compound	Reported purity of Reference Compound in %.		Not applicable	Optional
xx	Supplier of Reference Compound	Source / Merchant, where the Reference Compound was purchased.		Optional	Optional
26	Vehicle substance	Name of vehicle (i.e. solvent) used.	-	Optional	Optional
27	Concentrations of vehicle or impurities	Was the concentration of the vehicle or other impurities reported? The vehicle should not exceed 0.01 %, i.e. to exert no effect.	-	Optional	Optional
28	Radio labeled substance	Was the test substance radiolabeled? Which isotope was used? How was it quantified?	-	Optional	Optional
29	Preparation of stock solutions	Was the method for preparing the stock solution reported? Is the amount of test substance in the formulation indicated?	-	Optional	Optional
30	Standard qualifier	Qualifier for the test standard used, e.g. "according, equivalent or similar to" the test guideline used.	-	Optional	Optional
31	Standard used	Standard used, e.g. OECD Guideline 210 (Fish, Early-Life Stage Toxicity Test), or reference to a certain paper.	-	Optional	Optional
33	Principles of method if other than guideline	Description of the test protocol if other than the guideline, giving a justification for using this method if appropriate.	-	Optional	Optional
34	GLP Certificate	Indicates whether the study is GLP compliant, e.g. yes, no, not reported.	-	Optional	Optional
35	Effect	Observed effect category, e.g. growth, reproduction, mortality.	Autom.	Optional	Optional
36	Effect measurement	Effect measurement, e.g. biomass vs growth rate, number of offspring or time to first breed, mortality, survival or immobilisation.	Mandatory	Mandatory	Mandatory
37	Endpoint	Endpoint type to which the effect concentration refers, e.g. LC50, EC10, EC50, IC50, AC50	Mandatory	Mandatory	Mandatory
143	Response site	Molecular site within the test species which is affected by the toxicant.	-	Optional	Optional
xx	Bioassay name	Name of the bioassay		Not applicable	Mandatory
xx	Bioassay type	Type of bioassay in vitro, in vivo, ex vivo		Not applicable	Mandatory
xx	Bioassay endpoint AOP	Type of toxicological endpoint that the bioassay tests (e.g. liver steatosis)		Not applicable	Mandatory
xx	Bioassay Target family	cell morphology, cell cycle, dna binding, etc.		Not applicable	Mandatory
xx	Bioassay Event	Point within AOP		Not applicable	Mandatory
38	Test duration	Numeric value of the time until the effect is reported.	Mandatory	Mandatory	Optional
39	Test duration Unit	Unit of the test duration	Mandatory	Mandatory	Optional
40	Total test duration	Duration the test was conducted for.	-	Optional	???
41	Recovery considered?	Was recovery of test organism considered in the study design?	-	Optional	Not applicable
42	Scientific name	Scientific name of the test species.	-	Mandatory	Autom.
43	Common name	Common name of the test species.	-	Mandatory	Autom.
44	Taxonomic group	Taxonomic group or trophic level of test species : e.g. Algae and plants, Fish, Cyanobacteria, etc.	-	Autom.	Autom.

45	Initial body length of control group	Body length of test species at the beginning of the experiment.	-	Optional	Not applicable
144	Final body length of control group	Body length of test species at the end of the experiment.	-	Optional	Not applicable
46	Body length Unit	Unit of body length	-	Optional	Not applicable
47	Initial body weight of control group	Body weight of the controll group.	-	Optional	Not applicable
48	Body weight Unit	Unit of body weight	-	Optional	Not applicable
49	Initial cell density of control group (cells/mL)	Determination of cell concentrations or chlorophyll measurement at the begining of the experiment.	-	Optional	Optional
138	Final cell density of control group (cells/mL)	Determination of cell concentrations or chlorophyll measurement at the end of the experiment (for the control).	-	Optional	Optional
141	Deformed or abnormal cells / organism	Indication whether any deformed or abnormal cells or organism are observed.	-	Optional	Optional
50	Reproductive condition of the control	Reproductive condition of test species (number of offspring, mean coefficient of variation for section-by-section specific growth rate) according to minimum requirements.	-	Optional	Not applicable
52	Lipid %	Lipid content of the test species.	-	Optional	Not applicable
145	Lipid method	Method used to determine the lipid content in test species.	-	Optional	Not applicable
53	Age	Age of the test species or inoculum.	-	Optional	Not applicable
54	Age unit	Unit of age	-	Optional	Not applicable
55	Life stage	Life stage of test species, e.g. Juvenile, adult, nymph, etc.	-	Optional	Not applicable
56	Gender	Gender of the test species.	-	Optional	Not applicable
57	Strain, clone	Strain/clone of the test species.	-	Optional	Optional
58	Source (laboratory, culture collection)	Source, where test species were purchased. Is there information about the history of the control species? Have the test species been collected from a polluted site?	-	Optional	Optional
59	Culture handling	Information, how the test species culture was handeled (conditions).	-	Optional	Optional
60	Acclimation	Did the test species have enough time to acclimate to the test conditions prior to the test begin?	-	Optional	Not applicable
61	Nominal concentrations	Nominal concentrations or doses tested.	Mandatory	Mandatory	Mandatory
62	Measured (initial) concentrations	Measured concentrations or doses at the beginning of the test.	-	Optional	Optional
63	Measured or nominal concentrations used?	If the deviation from the nominal concentration is greater than 20 per cent, are results based on the measured concentration?	-	Optional	Optional
64	Limit test	Indicator whether the experiment was a limit test.	-	Optional	???
65	Range finding study	Indicator whether the study was performed for range finding.	-	Optional	Optional
66	Analytical matrix	Indication whether a specific tissue or the whole organism is analyzed.	-	Optional	Optional
67	Analytical schedule	How often and when was the concentration determined?	-	Optional	Optional
68	Analytical method	Analytical method: description of method.	-	Optional	Optional
69	Analytical recovery	How was the recovery of the compound?	-	Optional	Optional
70	Limit of quantification	How was the limit of quantification?	-	Optional	Optional
71	Exposure regime	Exposure regime (static, semistatic, flow through system, change of water, repeated doses).	Mandatory	Mandatory	Optional
76	Exposure route	Exposure route, e.g. aqueous, dietary, injection.	-	Optional	Not applicable
72	Exposure duration	Duration the test species were exposed to the test substance.	Mandatory	Mandatory	Mandatory
73	Exposure duration Unit	Unit of exposure duration	Mandatory	Mandatory	Mandatory
74	Application frequency	How often was the test material applied? Was the test solution renewed?	-	Optional	Optional
75	Application frequency unit	Unit of application frequency (if applicable)	-	Optional	Optional
142	Negative control used?	Indicator whether a negative control was used?	-	Optional	Optional
77	Positive control used?	Indicator whether a positive control was used?	-	Optional	Optional
78	Positive control substance	Which substance was used as a positive control?	-	Optional	Optional
79	Effects in positive control	Did the positive control show the expected effect?	-	Optional	Optional
80	Vehicle control used?	Indicator whether a vehicle control was used?	-	Optional	Optional
81	Effects in vehicle control	Was the solvent control significant different from control?	-	Optional	Optional
xx	Reference compound used?	Indicator whether reference compound was used?	-	Optional	Mandatory
xx	Effects for reference compound	Did the reference compound show the expected effect?	-	Optional	Optional
xx	Check for cyotoxicity	In case of non-cytotoxic endpoint, was cytotoxicity checked? Yes, no, not applicable, unknown		Not applicable	Mandatory
xx	Cytotoxicity endpoint (IC50)	What is the endpoint			Optional

xx	Cytotoxicity method	In case of non-cytotoxic endpoint, which method was used to determine cytotoxicity?		Not applicable	Mandatory
xx	Protein or lipid content	What is the protein or lipid content of the ..(?)			Optional
82	Intervals of water quality measurements	How often was the water quality measured?	-	Optional	Optional
83	Intervals of water quality measurements Unit	Time unit of water quality measurements	-	Optional	Optional
84	pH	Test pH applicable to the measured test value or mean or range.	-	Optional	Optional
85	Adjustment of pH	Was the test pH adjusted / buffered?	-	Optional	Optional
86	Temperature	Temperature used in the experimental determination of the test value.	-	Optional	Optional
87	Temperature Unit	Temperature Unit	-	Optional	Optional
88	Conductivity	Conductivity used in the experimental determination of the test value.	-	Optional	Optional
89	Conductivity Unit	Conductivity Unit	-	Optional	Optional
90	Light intensity	Light intensity used in the experimental determination of the test value.	-	Optional	Optional
91	Light intensity Unit	Light intensity Unit	-	Optional	Optional
92	Light quality (source and homogeneity)	Light quality used in the experimental determination of the test value.	-	Optional	Optional
93	Photo period	Photo period used in the experimental determination of the test value.	-	Optional	Optional
94	Hardness	The concentration of calcium and magnesium ions in water, measured as mg/L calcium carbonate equivalent.	-	Optional	Optional
95	Hardness Unit	Hardness Unit	-	Optional	Optional
96	Chlorine	Chlorine concentration determined in the test.	-	Optional	Optional
97	Chlorine Unit	Chlorine Unit	-	Optional	Optional
98	Alkalinity	Alkalinity used in the experimental determination of the test value.	-	Optional	Optional
99	Alkalinity Unit	Alkalinity Unit	-	Optional	Optional
100	Salinity	Salinity used in the experimental determination of the test value. Especially important for marine or brackish tests.	-	Optional	Optional
101	Salinity Unit	Salinity Unit (e.g. given in ‰)	-	Optional	Optional
102	Total Organic Carbon	Total carbon content used in the experimental determination of the test value.	-	Optional	Optional
103	Total Organic Carbon Unit	Organic Carbon Unit	-	Optional	Optional
104	Dissolved oxygen	Dissolved oxygen concentration should be greater than or equal to 3 mg/l or > 60% in control and test vessels.	-	Optional	Optional
105	Dissolved oxygen Unit	Dissolved oxygen Unit	-	Optional	Optional
106	Use of sand or sediment, and its characteristics	Description of the sand or sediment, and its characteristics used (total organic carbon, particle size, etc.).	-	Optional	Optional
xx	Type of test vessel	Beaker glass, Greiner Tube, 6-well plate, 96-well plate, etc.		Optional	Optional
107	Material of test vessel	Material of aquarium/container, considering the lipophilicity of the test substance.	-	Optional	Optional
108	Volume of test vessel	Volume (ml) of test test vessels used.	-	Optional	Optional
109	Open or closed system	Is the system closed or open, considering the volatility of the test substance.	-	Optional	Optional
110	Aeration	Was the test vessel aerated?	-	Optional	Optional
xx	Description of growth medium	Was the growth medium described?			Optional
111	Description of test medium	Was the test medium described? ~~Conditions suitable for test species?~~	-	Optional	Optional
112	Culture medium different from test medium?	Is the culture medium different from test medium?	-	Optional	Optional
113	Feeding protocols	Are feeding protocols (for long-term tests) reported?	-	Optional	Optional
114	Type and amount of food	Food composition reported?	-	Optional	Optional
115	Number of organisms per replicate	Number of organisms, or algal cell concentration that were used per replicate.	-	Optional	Not applicable
xx	Number of test concentrations?	How many concentrations of the test substance were tested?		Optional	Mandatory
116	Number of replicates per concentration	Was the number of replicates per treatment reported?	Mandatory	Optional	optional

125	Concentration based on	The effect is based on test material, active ingredient, free ion, etc.	NORMAN	NORMAN	NORMAN
117	Statistical method used	Which statistical method was used? Is it appropriate?	-	Optional	Optional
118	Trend	Is there an increasing or decreasing trend in the dose response curve?	-	Optional	Optional
119	Significance of result	Was the test significant?	-	Optional	Optional
120	Significance level	Significance level for NOEC and LOEC data (e.g. p = 0.05).	-	Optional	Optional
xx	Substance Hit call	Active / No response		Optional	Optional
xx	Cytotoxicity excluded	In case of non-cytotoxic endpoint, can interference of cytotoxicity with reported result be excluded?		Not applicable	Mandatory
121	Effect concentration qualifier	Qualifier for the effect value, e.g. >, =, <.	Mandatory	Mandatory	Mandatory
122	Effect concentration	Effect value (e.g. LC50) in µg/L or kg/L.	Mandatory	Mandatory	Mandatory
123	Effect concentration unit	Concentration unit, e.g. µg/L, µg/Kg	Mandatory	Mandatory	Mandatory
xx	Effect concentration derivation	linear, log, Hill, etc? curve			Mandatory
xx	Reference compound effect concentration qualifier	Qualifier for the effect value, e.g. >, =, <.		Optional	Mandatory
xx	Reference compound effect concentration	Effect value (e.g. LC50) in µg/L or kg/L.		Optional	Mandatory
xx	Reference compound effect	e.g. LC50, LC10, etc			Mandatory
xx	Reference compound effect derivation	linear, log, Hill, etc? curve			Mandatory
124	Estimate of variability for LC and EC data	Was an estimate of variability for LC and EC data given?	-	Optional	Optional
127	Result comment	Any comments regarding the test results.	-	Optional	Optional
129	Dose-response reported in figure/text/table	Dose-response reported in figure/text/table?	-	Mandatory	Mandatory
131	Availability of raw data	Is the raw data available (e.g. Doc III = Effect data)? Link to the data.	-	Mandatory	Mandatory
133	General comment	General comments, e.g. QSAR Software used	-	Optional	Optional
134	Existing reliabilty score	Reliability score: e.g. 1 to 4 from CRED, 5 = not evaluated.	Mandatory	Mandatory	Optional
135	Reliability score system used	Reliability according to: e.g. Klimisch, CRED, etc.	Mandatory	Mandatory	Optional
140	Affiliation issuing the reliabilty score	Institute, Organisation or Authority that released the reliability score.	Mandatory	Mandatory	Optional
136	Existing rational reliability	Rationale for the reliablilty classification.	Mandatory	Mandatory	Optional
137	Regulatory purpose	Indication of the regulatory or other programme for which the data are used.	-		
139	Purpose status	Indication if the study was used for regulatory purposes, i.e. key study, supporting study.	-		

Figure S1. Screenshot of the proposed upload form. In red are additions to the fields of the Ecotoxicity database upload form. Green texts are most recently added fields. Green cells indicate the Mandatory or automatic fields.