PollutantDB

Contaminant transport and risk parameters – in-browser app • Hydrogeological Repository

What it does

PollutantDB is the parameter lookup a contaminant transport or risk calculation starts with. Pick up to six constituents and up to six aquifer materials and it puts the published values side by side: **solubility, log K_{ow}, Henry's law constants, vapour pressure, density and viscosity**; **drinking water standards** and **degradation half-lives**; **sorption coefficients** for the constituent–material pairs that have been measured; and the **hydraulic conductivity, porosity, specific yield and specific storage** of the materials themselves, with **field dispersivities** against the scale they were measured at.

Nothing is reduced to a single accepted number. Where fifteen references report a solubility, the app shows the median, the range and all fifteen values with the paper each came from – because the spread between them is the uncertainty the calculation inherits, and a property backed by one 1979 measurement should not look like one backed by fifteen.

What the database holds

174 constituents – the EPA priority pollutants, the regulated drinking water contaminants and the common radionuclides – and 85 aquifer materials, from glaciofluvial sand to fractured granite. Every number in it is a published value traceable to one of 457 references.

Table	Records	What it holds
Physicochemical properties	1 447	Solubility (407), log K_{ow} (299), vapour pressure (264), Henry's law constant (210), density (208) and viscosity (59), for 132 constituents.
Drinking water standards	74	71 maximum contaminant levels and 3 maximum residual disinfectant levels, from the US EPA national primary drinking water standards.
Half-life	174	One degradation half-life estimate per constituent, most with the range the reference gave around it.
Sorption	193	Distribution coefficients K_d for 138 measured constituent–material pairs, covering 45 constituents on 31 materials.
Flow properties	367	Hydraulic conductivity (170), porosity (108), specific yield (60) and specific storage (29), for 73 of the materials.
Dispersivity	90	Longitudinal dispersivity against the travel distance of the test, for 32 materials, from a 0.09 m laboratory column to a 43 km regional study.

Units the source database does not record

Three of the property columns are stored as bare numbers with no unit anywhere in the database. Rather than guess, each was established from the values themselves against compounds whose properties are not in dispute, and the app labels them accordingly.

Property	Unit used	How it was established
Vapour pressure	mmHg	Benzene at 76 and toluene at 22, both at 20 °C, are the mmHg values.

Property	Unit used	How it was established
Henry's law constant	atm·m³/mol	Benzene at 0.00548 and TCE at 0.0091 match the published constants in these units.
Viscosity	cP	Benzene at 0.634 cP at 22 °C. The values are *dynamic* viscosity, despite the source column being named for kinematic viscosity.

The units that are recorded are kept as published and converted alongside: mL/g and L/kg are the same number and are pooled as L/kg; ft/s becomes m/s and 1/ft becomes 1/m; years and months become days; and g/L, ppm and ppb become mg/L. Every table and export carries both the published figure with its own unit and the converted one, so the original is never lost.

Read the temperature with the number. Most properties here were measured at 20 or 25 °C and the reference's note says which. Henry's law constant and vapour pressure both roughly double over that 5 °C, so part of the spread within a constituent is temperature rather than disagreement between the references.

Sorption and retardation

K_d is shown per constituent–material pair, as the geometric mean of the measurements of that pair with their range and count. The app turns it into a retardation factor with the linear equilibrium isotherm, $\mathbf{R = 1 + \rho_b K_d / n_e}$, for a bulk density and effective porosity you set – optionally taking each material's own median published porosity where the database has one. R = 5 means the constituent's centre of mass travels at a fifth of the groundwater velocity, and that flushing it out again takes five times as long.

K_d belongs to the pair, not to the chemical. The same compound sorbs an order of magnitude more strongly to a silt with organic carbon in it than to a clean gravel, and each measurement carries the pH, ionic strength and organic carbon content of the column it was made in. Only 138 of the 14 790 possible pairs have been measured; the app shows a dash for the rest rather than a value borrowed from a similar material.

The isotherm also assumes sorption is instantaneous, reversible and independent of concentration. Where it is not – strongly sorbing metals, high concentrations, rate-limited diffusion into a matrix – R understates the tail the plume leaves behind.

Dispersivity, and the scale it belongs to

Dispersivity rises with the distance travelled, because a longer path samples more of the heterogeneity that produces the spreading in the first place. Across these 89 measurements a least-squares fit in log–log space gives $\alpha_L = 0.091\ L^{0.74}$ ($R^2 = 0.59$), so a 100 m problem sits near 2.7 m – but the scatter at any one scale runs to two or three orders of magnitude, and that matters as much as the trend through it.

The 0.1L rule of thumb is conservative, not central. Two thirds of these measurements fall below it and the median ratio of dispersivity to distance is 0.035. The app draws both the rule and the fit over the data so the choice is made with the scatter in view. Read the value off the scale of the problem being modelled; it is not a property of the material.

Aquifer flow properties

Hydraulic conductivity across these materials spans nearly twelve orders of magnitude, from 3×10^{-13} m/s for an unweathered marine clay to 0.15 m/s for a clean gravel. That and specific storage are log-normally distributed, so both are summarised by their **geometric mean**; porosity and specific yield carry the arithmetic mean. Each is shown with its

range and the number of values behind it, and the chart can draw every other material in the database behind the selection for context.

What you get

Output	Detail
Property summary	Median, range and count of every property for up to six constituents side by side, with CAS number, chemical class and the overall K_d across all materials.
Published values	Every individual value behind a property, with its unit as published and the full reference.
Standards and half-life	Maximum contaminant levels, and half-lives in days with the range the reference gave.
Sorption	K_d by material for the selection, charted on a log axis, and the retardation factor for parameters you set.
Flow properties	Hydraulic conductivity, porosity, specific yield and specific storage per material, with every published value.
Dispersivity	All measurements against scale on log–log axes, the fitted power law and the 0.1L rule, with the selected materials picked out.
References	All 457, searchable, filterable by what cites them, and restrictable to the references behind the current selection.
Exports	CSV of any block, for the selection or the whole database, one row per published value with its unit, reference and note; PNG of every chart; and a printable report of the whole selection.

How a run works

Step	What happens
1. Constituents	Search by name or CAS number, filter by chemical class, select up to six.
2. Properties	The six physicochemical properties, summarised and then value by value.
3. Standards and persistence	Drinking water standards and half-lives.
4. Aquifer materials	Select up to six; their flow properties, summarised and value by value.
5. Sorption and retardation	K_d for the measured pairs, and R for your bulk density and porosity.
6. Dispersivity	The scale relationship, with the selection highlighted.
7. References	Everything the selection rests on.
8. Report	A standalone parameter sheet, ready to print or save as PDF.
9. Export	CSV of any part of the database.

Where the numbers come from, and what was repaired

The source is a compilation of the contaminant transport literature held as nine tables. It stores everything as text, with no declared units on three columns and no foreign keys, so the dataset the app reads is built from it once by a script

that parses the numbers, records the unit each was published in and adds an SI-consistent copy. Four repairs were needed, all of them documented in that script:

- Six misspelled chemical classes – "chlorinate alkane", "aromaitc" and the like – were folded into the spelling already present in the same column, so a filter on a class returns all of its compounds.
- Two CAS numbers appeared in two spellings between references. The check digit identifies which of each pair is the real one.
- One viscosity recorded as 12345 was dropped as a placeholder; the next largest value in that column is 17.2.
- One dispersivity record pairs a travel distance of 10 µm with a dispersivity of 22.8 km, from a paper on regional systems. The dispersivity is kept and the distance dropped, which leaves the record in the table with the reason attached and off the chart. Left in, that one point moved the fitted exponent from 0.74 to 0.39 and R^2 from 0.59 to 0.20.

Limitations worth knowing

- These are literature values compiled from many settings, not measurements from your site. They are the right basis for a screening calculation, a sensitivity range or a first estimate where nothing has been measured, and the wrong basis for a design or a compliance decision.
- A half-life is one literature estimate. Degradation depends on the redox state, the temperature and the microbial community of the specific setting, and the ranges in the reference notes – often a factor of ten either side – are the honest measure of that.
- Coverage is uneven. 132 of the 174 constituents carry physicochemical properties, 74 carry a drinking water standard, 45 appear in the sorption table, and 73 of the 85 materials carry flow properties. A dash means nothing has been measured, not that the value is zero.
- The drinking water standards are the US EPA's. They are not South African standards and should not be read as a local compliance limit.
- One standard is recorded in µg/m³, an air value in a water table. It is shown as published and left out of the mg/L comparison rather than converted.
- Six constituents and six materials is the limit for one selection. More makes the charts unreadable before it makes them slow.

Everything runs in your browser. The whole database is 55 kB compressed and is downloaded once when the page opens. Every summary, chart, retardation factor and export is computed on your machine; nothing about what you select is sent anywhere or stored.

Credit

The values are the work of the researchers listed in the 457 references and of the compilers who assembled them into the source database. This app presents their published numbers, with the reference attached to each, and claims no authorship of them.

Beta. PollutantDB is published for evaluation and feedback. Check any figure you rely on against the reference shown beside it before using it in an assessment.