Supplementary Material

Table 1. A description of the data categories used in the curation of an interdisciplinary database of chemical novelty and bioactivity in the deep sea by 2024.

Data Category	Description
Domain	Taxonomic Domain identified by the publication and defined by WoRMs[89] if available.
Kingdom	Taxonomic Kingdom identified by the publication defined by WoRMs[89] if available.
Phylum	Taxonomic Phylum identified by the publication defined by WoRMs[89] if available.
Class	Taxonomic Class identified by the publication defined by WoRMs[89] if available.
Order	Taxonomic Order identified by the publication defined by WoRMs[89] if available.
Family	Taxonomic Family identified by the publication defined by WoRMs[89] if available.
Genus	Taxonomic Genus identified by the publication defined by WoRMs[89] if available.
Species	Taxonomic Species identified by the publication defined by WoRMs[89] if available.
Strain ID/Accession number	The unique identifier assigned to the specific strain or sequence of the recorded organism, if available.
Database code	The code used by the database the original record was derived from, if available.
Compound name	The chemical name given to the compound in the publication, if available.
SMILES	The Simplified Molecular Input Line Entry System (SMILES) line notation that represents chemical structures using short ASCII strings, if available.
Pathway	Chemical pathway defines by ClassiFire[88] if available.
Superclass	Chemical superclass defines by ClassiFire[88] if available.
Class	Chemical class defines by ClassiFire[88] if available.
Bioactivity	General bioactivity type and detailed bioactivity results inclding the IC50 or MIC values where possible.
Location	General descriptive location.
Latitude	Latitude in decimal degrees if available.
Longitude	Longitude in decimal degrees if available.
Depth (m)	Depth of sample collection in meters if available.
DOI	The Digital Object Identifier (DOI) of the publication from which the record was derived.
Chemical Novelty	If the compound was deemed chemically novel (N) or not (-) at the time of publication.
Additional information	Any additional information, for example other DOIs that have recorded the compound previously.
Environment	A description of if the sample was collected from the pelagic or benthos, and any details of the habitat or geomorphological feature from which the sample was derived if available.
WoRMS	If the recorded taxonomic classification is (Y) or is not (N) featured in the WoRMS catalogue[89].

Table 2. A summary of the curated interdisciplinary database that describes the abundance, distribution, chemistry, taxonomy, phylogeny, and environment of chemical novelty and bioactivity in the deep sea by 2024.

Data Collection	Count	Description
Total Data Entries	2 907	A database of all deep-sea bioactivity and chemical novelty records (compounds and extracts).
Novel Compounds	2 195	Recorded as "novel" or "new" at the time of publication (77 % of compounds recorded).
Bioactive Compounds	1 870	Compounds that have shown any form of bioactivity (65 % of records).
Bioactive Extracts	30	Compounds are not defined.
Chemistry		
Chemical Compounds	2 877	Entries defining chemical compounds, extracts excluded.
Chemical Compound Classes	216	445 compounds could not be defined to a class using the NP classifier platform[88] and open access literature.
Chemical Compound Super Classes	67	337 compounds could not be defined to a superclass using the NP classifier platform[88] and open access literature.
Chemical Compound Structure Pathways	8	Alkaloids (474), Amino acids and peptides (259), Carbohydrates (12), Cyanocobalamin (1), Fatty acids (148), Polyketides (903), Shikimates and phenylpropanoids (131) Terpenoids (733). 180 compounds could not be defined to a pathway using the NP classifier platform[88] and open access literature.
Bioactivity *single compounds/extracts may exhibit multiple activities; thus 1900 bioactive compounds/extracts generate 2290 recorded bioactive results.		
Antibacterial Compounds	504	
Antiviral Compounds	98	
Antifungal Compounds	136	
Cytotoxic Compounds	849	
Anti-inflammatory Compounds	205	
Antioxidant Compounds	88	
Enzyme activity	175	
Antiparasitic Compounds	34	
Other Bioactivity	201	
Taxonomy		
Kingdoms	4	Animalia, Bacteria, Chromista, Fungi. 10 records identified to kingdom only.
Phyla	16	Actinobacteria, Aquificae, Arthropoda, Firmicutes, Proteobacteria, Chordata, Cnidaria, Echinodermata, Mollusca, Nematoda, Porifera, Foraminifera, Ochrophyta, Ascomycota, Basidiomycota, Zygomycota. 9 records identified to phyla only.
Classes	32	2 records identified to class only.

Orders	69	1 record identified to order only.
Families	142	14 records identified to family only.
Genera	192	1 161 records identified to genus only (40 %).
Species	223	1 710 records identified to species level (59 %).
Literature		
Databases	7	MarinLit[4], CMNPD[74], WoRMS[89], NP Classifier[88], Reaxys[86], Chem Spider[87], Time Tree[90].