

List of fields and definitions used in a data set of fish occurrence records from the Danube River Basin. Terms and definitions are according to Darwin Core standards (dc, <http://www.tdwg.org/standards/450>), Freshwater Core Template at Freshwater Biodiversity Data Portal (fwct, <https://data.freshwaterbiodiversity.eu/>), Global Biodiversity Information Facility (gbif, <https://www.gbif.org>) and Amatulli et al. (2022).

Category	Field	Definition
Identification and Classification	scientificName <sup>dc</sup>	The full scientific name, with authorship and date information if known.
	genus <sup>dc</sup>	The full scientific name of the genus in which the taxon is classified.
	family <sup>dc</sup>	The full scientific name of the family in which the taxon is classified.
	order <sup>dc</sup>	The full scientific name of the order in which the taxon is classified.
	verbatimIdentification <sup>dc</sup>	A string representing the taxonomic identification as it appeared in the original record.
	identificationRemarks <sup>dc</sup>	Comments or notes about the process or conditions of identification, or any other information relevant to the identification of the taxon.
Occurrence Information	occurrenceID <sup>dc</sup>	A globally unique identifier for an individual occurrence record
	materialSampleID <sup>dc</sup>	A globally unique identifier for a physical sample or collection of material related to an occurrence.
	basisOfRecord <sup>dc</sup>	The specific nature of the data record. Categories “PreservedSpecimen”, “MaterialCitation” and “HumanObservation”.
	occurrenceRemarks <sup>dc</sup>	Comments or notes about the occurrence. In this case, whether or not the occurrence could be an environmental outlier detected by 10 outlier detections methods (i.e. kmeans, local outlier factor, isolation forest, Z-score, semi-interquartile range, Hampel filter, reverse jack knifing, One Class support vector machines, adjusted boxplots, and interquartile range).
	individualCount <sup>dc</sup>	The number of individuals present at the time of the occurrence

Category	Field	Definition
	lifeStage <sup>dc</sup>	The age class or life stage of the organism(s) at the time the occurrence was recorded.
Temporal Information	day <sup>dc</sup>	The integer day of the month on which the event occurred.
	month <sup>dc</sup>	The integer month in which the event occurred.
	year <sup>dc</sup>	The four-digit year in which the event occurred, according to the Common Era Calendar.
Collector and Identifier Information	recordedBy <sup>dc</sup>	A person, group, or organization responsible for recording the original occurrence.
	identifiedBy <sup>dc</sup>	A list (concatenated and separated) of names of people, groups, or organizations who assigned the Taxon to the subject.
Geospatial Information	waterBody <sup>fwct</sup>	The name of the water body in which the location occurs.
	subcatchmentID <sup>Amatulli et al. (2022)</sup>	Identification number for each subcatchment in the Danube River Basin <sup>1</sup> . A subcatchment is the land area that drains into a stream reach.
	country <sup>dc</sup>	The name of the country or major geopolitical unit where the occurrence was recorded.
	stateProvince <sup>dc</sup>	The name of the next smaller administrative region than country (state, province, canton, department, region, etc.) in which the location occurs.
	county <sup>dc</sup>	The full, unabbreviated name of the next smaller administrative region than stateProvince (county, shire, department, etc.)
	locality <sup>dc</sup>	The specific description of the place.
	decimalLatitude <sup>dc</sup>	The geographic latitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic center of a location. Positive values are north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive.
	decimalLongitude <sup>dc</sup>	The geographic longitude (in decimal degrees, using the spatial

Category	Field	Definition
		reference system given in geodeticDatum) of the geographic center of a location. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180, inclusive.
	geodeticDatum <sup>dc</sup>	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which the geographic coordinates given in decimalLatitude and decimalLongitude as based.
	coordinatePrecision <sup>dc</sup>	Precision of the coordinates provided for a specific occurrence record. e.g.: coordinatePrecision = 0.0001 (Coordinates are precise to approximately 11 meters at the equator.)
	coordinateUncertaintyInMeters <sup>dc</sup>	The radius of uncertainty around the given geographic coordinates, indicating the potential error in the location of an occurrence.
Georeferencing Information	georeferenceProtocol <sup>dc</sup>	The methods or protocols used to determine the geographic coordinates of an occurrence. Specify the method used for determining the location of the occurrence, such as manual georeferencing from maps, GPS devices, or automated systems.
	georeferenceSources <sup>dc</sup>	A list (concatenated and separated) of maps, gazetteers, or other resources used to georeference the location, described specifically enough to allow anyone in the future to use the same resources.
Collection and Dataset Information	catalogNumber <sup>dc</sup>	An identifier (preferably unique) for the record within the data set or collection.
	collectionCode <sup>dc</sup>	Term used to identify the specific collection within an institution or repository where a specimen, observation, or occurrence record is cataloged.
	institutionCode <sup>dc</sup>	The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record.

Category	Field	Definition
	datasetName <sup>dc</sup>	The name identifying the data set from which the record was derived.
	samplingProtocol <sup>dc</sup>	The names of, references to, or descriptions of the methods or protocols used.
Rights and License	license <sup>dc</sup>	The legal terms under which the occurrence record or dataset is made available, specifying how the data can be used, shared, and redistributed. A standardized string representing a specific license.
	rightsHolder <sup>dc</sup>	The person, organization, or entity that holds the legal rights over the occurrence record or dataset, including the rights to its distribution, use, and licensing.
Bibliographic and Reference Information	reference <sup>dc</sup>	A related resource that is referenced, cited, or otherwise pointed to by the described resource. Link to academic papers, datasets, or authoritative publications that support the information provided in the occurrence record. e.g, a DOI linking to a study, a citation to a paper, a link to the original dataset.
	bibliographicCitation <sup>dc</sup>	A bibliographic reference for the resource. The intended usage of bibliographicCitation, is to provide the preferred way to cite the resource itself.
	gbifId <sup>gbif</sup>	Identification number of the occurrence record at the Global Biodiversity Information Facility (GBIF) database.

Amatulli, G., Garcia Marquez, J., Sethi, T., Kiesel, J., Grigoropoulou, A., Üblacker, M. M., Shen, L. Q., & Domisch, S. (2022). Hydrography90m: a new high-resolution global hydrographic dataset. *Earth Syst. Sci. Data*, *14*, 4525–4550.  
<https://doi.org/10.5194/essd-14-4525-2022>