

# **Geosciences, Environmental Sciences/PANGAEA**

Robert Huber  
MARUM – Universität Bremen

**FAIRification STEP 4: domain specific metadata**

# PANGAEA - Data Publisher for Earth & Environmental Science



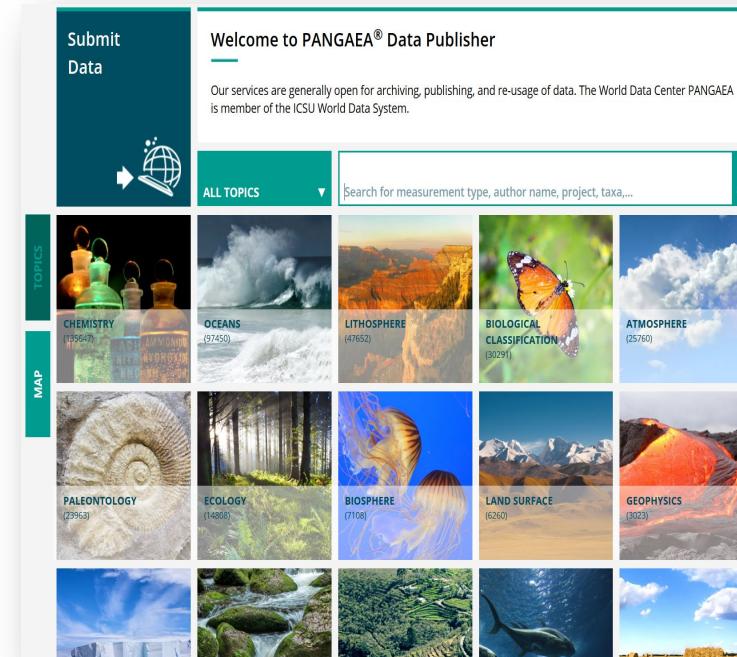
**PANGAEA.**  
Data Publisher for Earth & Environmental Science

- Founded in 1993
- Jointly managed by AWI and MARUM.
- Datasets from researchers, projects, research centres and infrastructures (national & international)
- Focus on marine & environmental data
- Certified long term data archive (WDS, WMO, DSA).

 ALFRED-WEGENER-INSTITUT  
HELMHOLTZ-ZENTRUM FÜR POLAR-  
UND MEERESFORSCHUNG

**marum**  
Center for Marine  
Environmental Sciences

> 400.000 scientific datasets (05.10.2021)



Welcome to PANGAEA® Data Publisher

Our services are generally open for archiving, publishing, and re-use of data. The World Data Center PANGAEA is member of the ICSU World Data System.

Submit Data

ALL TOPICS ▾

Search for measurement type, author name, project, taxa,...

TOPICS

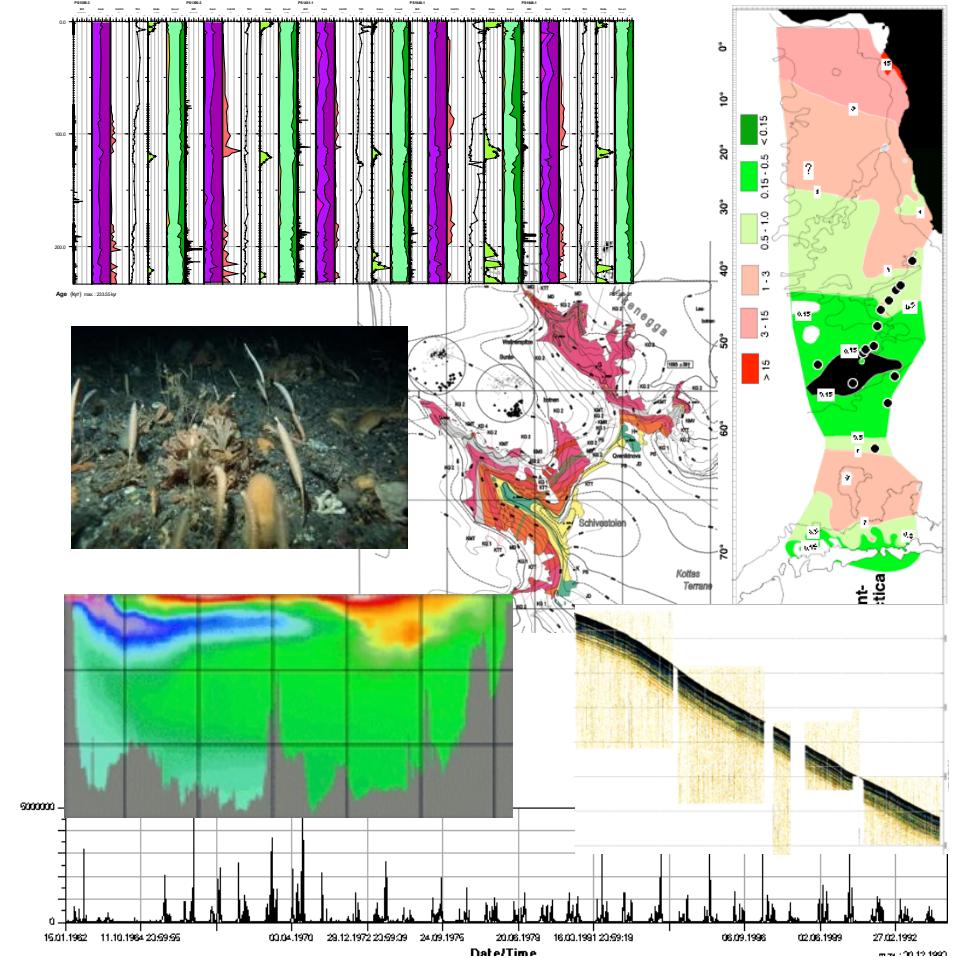
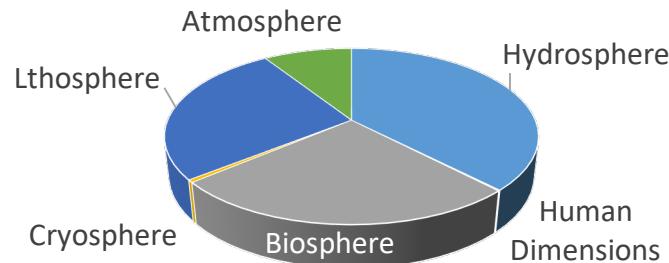
MAP

Chemistry (135657)	Oceans (97450)	Lithosphere (47652)	Biological Classification (30291)	Atmosphere (25760)
Paleontology (23963)	Ecology (14805)	Biosphere (7108)	Land Surface (6260)	Geophysics (3023)

PANGAEA Data Portal (<https://pangaea.de/>)

# PANGAEA Content, Data Types

- Profiles (oce,sed,ice,atm) -> [doi:10.1594/pangaea.103958](https://doi.org/10.1594/pangaea.103958)
- Environmental time series -> [doi:10.1594/pangaea.323487](https://doi.org/10.1594/pangaea.323487)
- Sea bed photos, movies -> [doi:10.1594/PANGAEA.319877](https://doi.org/10.1594/PANGAEA.319877)
- Sediment samples -> [doi:10.1594/pangaea.51749](https://doi.org/10.1594/pangaea.51749)
- Species occurrence data -> [doi:10.1594/PANGAEA.123036](https://doi.org/10.1594/PANGAEA.123036)
- Morphometry -> [doi:10.1594/PANGAEA.704634](https://doi.org/10.1594/PANGAEA.704634)
- Age models, tree rings -> [doi:10.1594/PANGAEA.787572](https://doi.org/10.1594/PANGAEA.787572)
- Stable isotopes -> [doi:10.1594/PANGAEA.52201](https://doi.org/10.1594/PANGAEA.52201)
- Geophysics-> [doi:10.1594/PANGAEA.786480](https://doi.org/10.1594/PANGAEA.786480)



# PANGAEA Content - Dataset



Not logged in

PANGAEA.  
Data Publisher for Earth & Environmental Science

SEARCH SUBMIT ABOUT CONTACT

Citation: **Demske, Dieter; Tarasov, Pavel E (2018):** Absolute counts of pollen and non-pollen palynomorphs from the sediment core XBW1146 from Boston Lake (Boston Hu) covering the last 8540 years. PANGAEA. <https://doi.org/10.1594/PANGAEA.896256>, In supplement to: Tarasov, Pavel E; Demske, Dieter; Leipe, Christian; Long, Tengwen; Müller, Stefanie; Hoelzmann, Philipp; Wagner, Mayke (2018): An 8500-year palynological record of vegetation, climate change and human activity in the Boston Lake region of Northwest China. *Palaeogeography, Palaeoclimatology, Palaeoecology*. <https://doi.org/10.1016/j.palaeo.2018.11.050>

Map Satellite

Data

Download dataset as tab-delimited text (use the following character encoding: UTF-8: Unicode (PANGAEA default))

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
Sample ID	Depth [m]	Age [ka BP]	Lyc(ad) [#]	Pinus s/g Diplo [#]	Pinus s/g Hapl. [#]	Pic [#]	Abi [#]	Jun-T [#]	Bet.ud [#]	Bet.n-T [#]	RosaeAP.ud [#]	Sor [#]	Ctt [#]	Pir [#]	Pop [#]	Ulm [#]	Rha [#]	Sal [#]	Hpp [#]	Ele [#]	Core (CORE)	
101	0.140	0.126	119		1	0	4	1	2	2	0	0	0	1	0	2	1	0	4	2		
103	0.400	0.453	64	1		0	4	0	1	0	0	0	0	0	0	0	1	1	0	0	4	
105	0.690	0.818	39	0		0	3	0	0	0	0	1	0	1	0	1	3	0	0	5		
107	0.890	1.070	57	2	1	3	0	0	0	0	0	0	0	0	0	2	4	1	2	4		
241	1.300	1.501	121	1	0	6	0	1	1	0	0	0	0	1	0	2	10	0	0	0		
231	1.500	1.707	58	2	0	10	1	0	2	0	1	0	1	1	2	4	1	2	2			
221	1.700	1.913	39	2	0	4	0	1	0	0	2	0	0	0	3	3	0	4	3			
211	1.900	2.119	83	0	1	3	0	1	0	0	0	0	0	0	2	0	0	0	1	0		
201	2.100	2.325	36	2	0	3	0	0	0	0	1	0	0	0	0	0	0	0	0	14	3	
347	2.260	2.490	63	2	0	9	0	0	3	1	0	0	0	0	0	0	0	0	0	3	2	
337	2.460	2.696	21	1	0	4	0	0	1	0	0	0	0	0	0	0	0	1	0	1		
332	2.560	2.799	37	1	0	4	0	0	0	0	0	0	0	0	0	1	0	0	0	1		
321	2.760	3.005	28	1	0	5	0	0	0	0	2	0	0	0	2	0	0	0	0	0		
311	2.960	3.211	59	1	0	6	0	0	0	0	1	0	2	0	2	0	2	0	0	4		
301	3.160	3.417	33	2	0	9	0	6	1	0	0	0	0	1	2	0	0	0	0	4		
451	3.205	3.463	36	0	1	3	0	5	2	0	0	0	1	0	1	0	0	0	1	0		
440	3.420	3.685	53	0	0	7	0	2	0	0	0	0	0	0	2	0	0	0	0	0		
431	3.620	3.891	55	1	0	5	0	1	0	0	0	0	0	0	0	0	0	1	2	0		
421	3.820	4.097	34	0	1	6	0	1	1	0	0	0	0	0	3	0	0	4	1			
444	4.020	4.292	26	0	4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0		

# PANGAEA - Growth

Integral part of science

More than 240 European to international projects since 1995

<https://www.pangaea.de/projects>

highly heterogenous & dynamic multidisciplinary

Number of data sets

~406.000

Number of data items

>19 Billion

Data volume

<3 PB

Increase

~5% per y

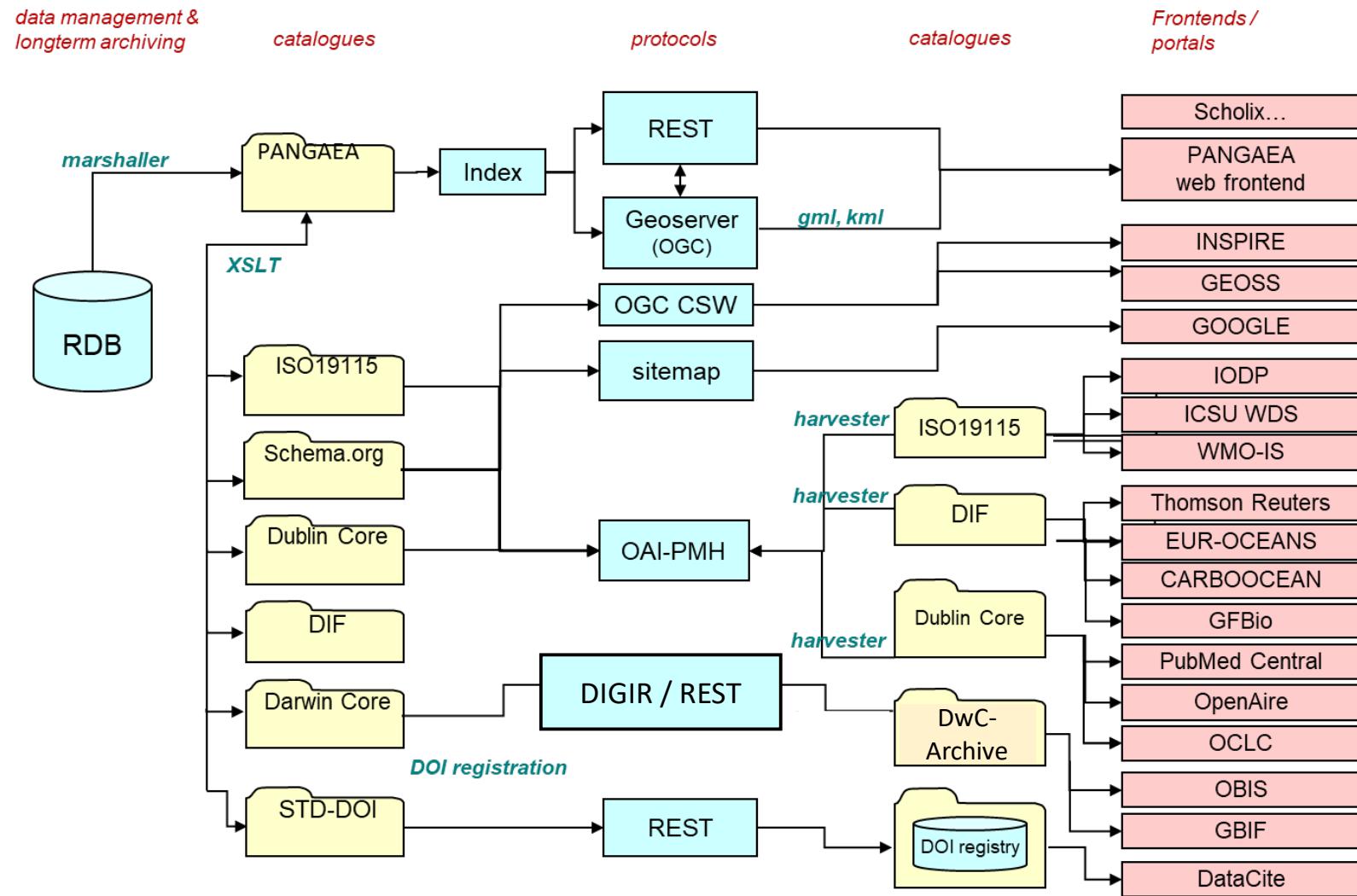
>10.000 registered users

12 – 15.000 unique users / month

# PANGAEA - Dissemination

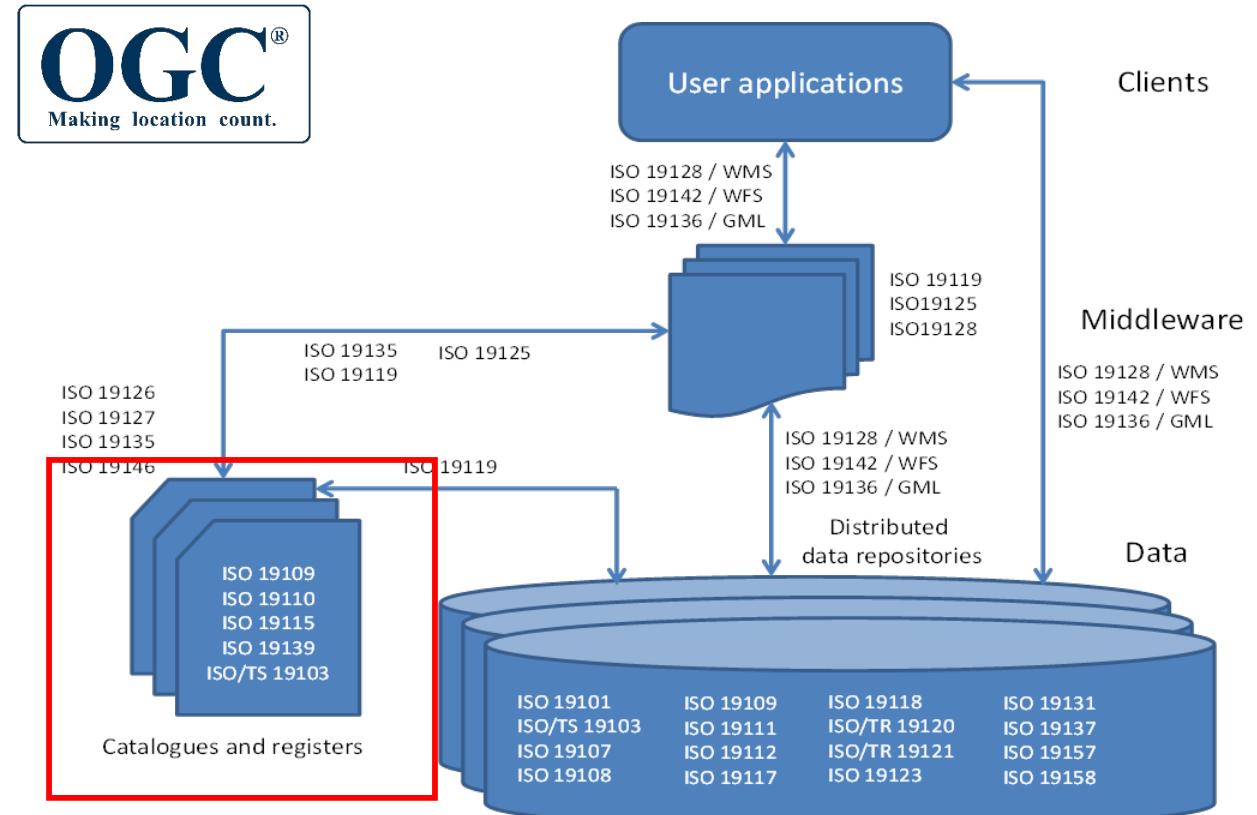


# PANGAEA - Interfaces



# ISO19115/19139 – part of OGC SDI

- Geographic Information – Metadata
- Good for: e.g. maps, remote sensing, gridded data etc.
- Very complex!
- Several versions for XML: e.g. 19139 , 19115-3
- Community profiles
  - INSPIRE



Exchange: as e.g. OAI-PMH or CSW payload

Image source: [https://eo4geo.sbg.ac.at/KULeuven/Technical\\_Introduction\\_SDI/Images\\_Module01/Standards\\_for\\_SDI.png](https://eo4geo.sbg.ac.at/KULeuven/Technical_Introduction_SDI/Images_Module01/Standards_for_SDI.png)

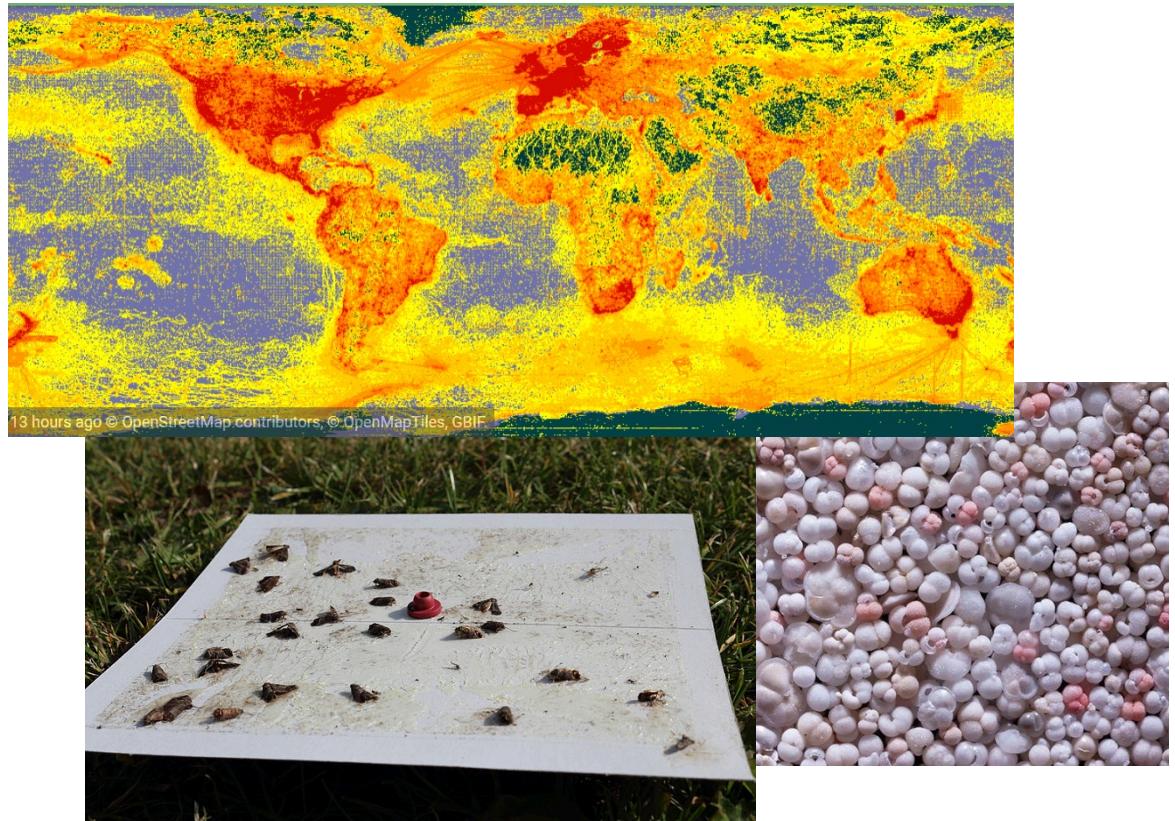
# ISO19139 @PANGAEA

```
<identificationInfo>
  <MD_DataIdentification>
    <citation>
      <CI_Citation>
        <title>
          <gco:CharacterString>Oligocene to Miocene palaeoclimate reconstructions from 14 sites in Serbia</gco:CharacterString>
        </title>
        <date>
          <CI_Date>
            <date>
              <gco:DateTime>2007-06-27T14:03:53</gco:DateTime>
            </date>
            <dateType>
              <CI_DateTypeCode codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml#CI_DateTypeCode" codeListValue="publication" >publication</CI_DateTypeCode>
            </dateType>
          </CI_Date>
        </date>
        <identifier>
          <MD_Identifier>
            <code>
              <gco:CharacterString>https://doi.org/10.1594/PANGAEA.472281</gco:CharacterString>
            </code>
          </MD_Identifier>
        </identifier>
        <citedResponsibleParty>
          <CI_ResponsibleParty id="dataset.author22081" >
            <individualName>
              <gco:CharacterString>Utescher, Torsten</gco:CharacterString>
            <individualName>
            <contactInfo>
              <CI_Contact>
                <address>
                  <CI_Address>
                    <electronicMailAddress>
                      <gco:CharacterString>utescher@geo.uni-bonn.de</gco:CharacterString>
                    </electronicMailAddress>
                  <CI_Address>
                    <address>
                      <CI_Contact>
                        <contactInfo>
                          <role>
                            <CI_RoleCode codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml#CI_RoleCode" codeListValue="author" >author</CI_RoleCode>
                          </role>
                        </CI_Contact>
                      </contactInfo>
                    </address>
                  </CI_Address>
                </address>
              </CI_Contact>
            </contactInfo>
          </CI_ResponsibleParty>
        </citedResponsibleParty>
```

<http://ws.pangaea.de/oai/provider?verb=ListRecords&metadataPrefix=iso19139>

# Environmental & Biodiversity formats

- Occurance data (lat, lon, species)
  - Darwin Core (DIGIR)
  - ABCD (Biocase)
  - Darwin Core Archive (zip EML+DwC data)
- Environmental data
  - Ecological Modelling Languange
  - ISO 19115
- PANGAEA: Darwin Core Archive



Images: GBIF, MARUM, [https://commons.wikimedia.org/wiki/File:Cydia\\_pomonella\\_trap\\_2012-06-05.jpg](https://commons.wikimedia.org/wiki/File:Cydia_pomonella_trap_2012-06-05.jpg)

# EML @PANGAEA (DwC Archive)

```
<?xml version="1.0" encoding="UTF-8"?>
<eml:eml xml:lang="eng" packageId="https://doi.org/10.1594/PANGAEA.727338" scope="system" system="http://www.pangaea.de"
xsi:schemaLocation="eml:/ecoinformatics.org/eml-2.1.1 http://rs.gbif.org/schema/eml-gbif-profile/1.1/eml.xsd"
xmlns:md="http://www.pangaea.de/MetaData" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:dc="http://purl.org/dc/terms/"
xmlns:eml="eml:/ecoinformatics.org/eml-2.1.1">
  - <dataset>
    <alternateIdentifier>https://doi.org/10.1594/PANGAEA.727338</alternateIdentifier>
    <alternateIdentifier>dataset727338</alternateIdentifier>
    <title xml:lang="eng">Planktic foraminifera counts of sediment core V14-102</title>
    - <creator>
      - <individualName>
        <surName>CLIMAP Project Members</surName>
      </individualName>
    </creator>
    - <metadataProvider>
      <organizationName>PANGAEA - Data Publisher for Earth & Environmental Science</organizationName>
      <electronicMailAddress>info@pangaea.de</electronicMailAddress>
      <onlineUrl>https://www.pangaea.de</onlineUrl>
    </metadataProvider>
    <pubDate>2009</pubDate>
    <language>eng</language>
    - <abstract>
      <para/>
    </abstract>
    - <intellectualRights>
      - <para>
          This work is licensed under a
          - <ulink url="https://creativecommons.org/licenses/by/3.0/">
            <citetitle>Creative Commons Attribution 3.0 Unported</citetitle>
          </ulink>
        </para>
    </intellectualRights>
    - <distribution>
      - <online>
          <url function="metadata">https://doi.org/10.1594/PANGAEA.727338</url>
        </online>
    </distribution>
    - <coverage>
      - <geographicCoverage>
        - <boundingCoordinates>
          <westBoundingCoordinate>57.183</westBoundingCoordinate>
          <eastBoundingCoordinate>57.183</eastBoundingCoordinate>
          <northBoundingCoordinate>10.25</northBoundingCoordinate>
          <southBoundingCoordinate>10.25</southBoundingCoordinate>
        </boundingCoordinates>
      </geographicCoverage>
    </coverage>
  </dataset>
</eml:eml>
```

# Domain specific schema.org

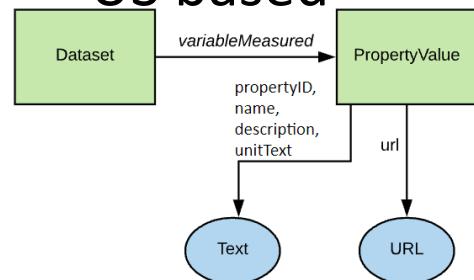
- **Bioschemas**
  - Life sciences Schema.org types
  - ELIXIR
- **ESIP (Earth Science Information partners)**
  - Generic & domain specific schema.org advice

The screenshot shows the Bioschemas schema.org page for the **Taxon** type. The page has a red header with links for Schema.org, Documentation, Schemas, and About. A search bar is at the top right. The main content area starts with a heading "Taxon" and a sub-heading "A Schema.org Type". It includes a note: "This term is proposed for full integration into Schema.org, pending implementation feedback and adoption from applications and websites." Below this, there's a breadcrumb navigation "Thing > Taxon" and a link "[more...]".

A set of organisms asserted to represent a natural cohesive biological unit.

Property	Expected Type	Description
Properties from Taxon		
<code>childTaxon</code>	Taxon or Text or URL	Closest child taxa of the taxon in question. Inverse property: <code>parentTaxon</code>
<code>hasDefinedTerm</code>	DefinedTerm	A Defined Term contained in this term set. Closest parent taxon of the taxon in question. Inverse property: <code>childTaxon</code>
<code>parentTaxon</code>	Taxon or Text or URL	
<code>taxonRank</code>	PropertyValue or Text or URL	The taxonomic rank of this taxon given preferably as a URI from a controlled vocabulary – (typically the ranks from TDWG TaxonRank ontology or equivalent Wikidata URIs).
Properties from Thing		

- **US based**



In its most basic form, the variable as a `schema:PropertyValue` can be published as:

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@type": "Dataset",
  "name": "Removal of organic carbon by natural bacterioplankton communities as a function of pCO2 from laboratory experiment",
  ...
  "variableMeasured": [
    {
      "@type": "PropertyValue",
      "name": "Bottle identifier",
      "description": "The bottle number for each associated measurement."
    }
  ]
}
```



**Thank you**