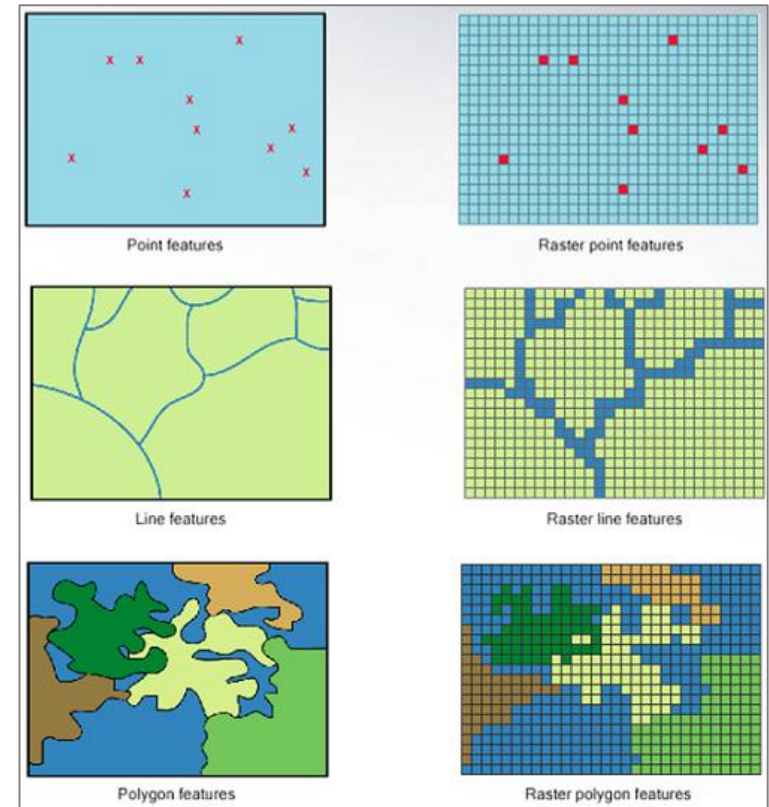




Geospatial Data in Environmental Science

Geospatial Data

- „where“ your data is located (with non-spatial components of „what“ it describes)
- types:
 - vector data** (points, lines, polygons, information stored in attribute tables) & **raster data** (pixel grid cells, information stored in numerical values)
- Geospatial data has definition of coordinates (in geographic, e.g. WGS84, and projected, e.g. UTM32, coordinate systems)
- various file formats exist to store and share geospatial data
- Different metadata standards are suitable for different topics of (geospatial) data
- (Geospatial) data can be accessed and published via topic-related data repositories



from <https://gsp.humboldt.edu/olm/Lessons/GIS/08%20Rasters/RasterToVector.html>

Geospatial Data in Environmental Science

Which data from the 4 pictures would be represented best by which spatial data type (vector points, lines, polygons, or raster)?



Flight paths of migratory birds
(Picture from S. Uryadnikov/Shutterstock.com)



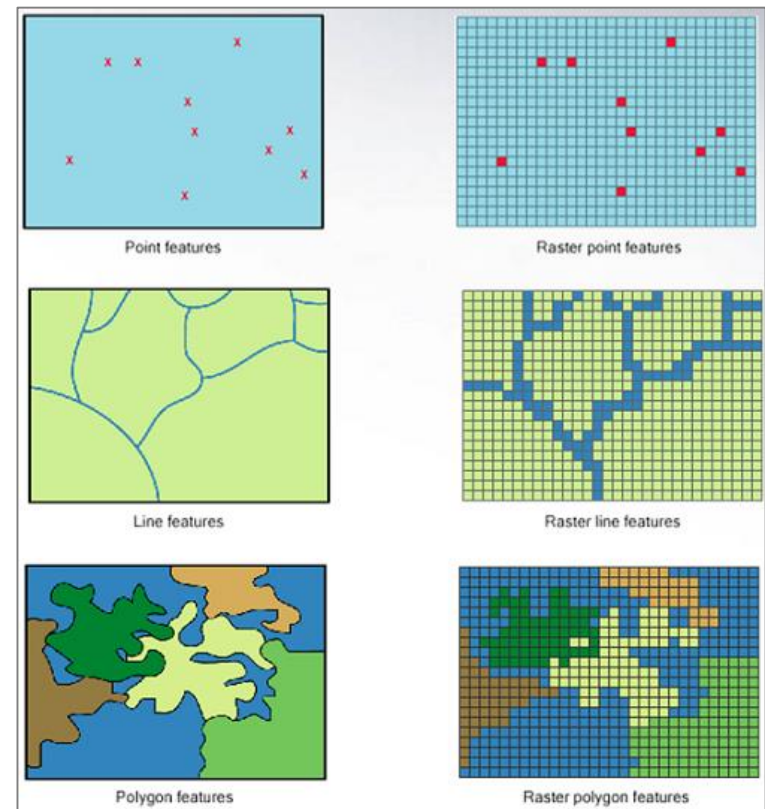
Habitat area mapping
(Picture from Westhäuser, E., & Richert, E. (2017):
Biotoptypenausstattung der Spülhalde
Münzbachtal bei Halsbrücke.)



Climate Data
(Picture from DWD)



Water sampling
(Picture from Tulane University)



Geospatial Data in Environmental Science

In environmental science:

- Spatiality is a core component in environmental data
 - Understanding environmental phenomena, their distributions, dynamics, changes, interdependencies, influence
 - Research needs data from various sources, in various formats, depending on data topics
- Research data management and Good scientific practice in working with (spatial) data are fundamentally important in environmental research

Geospatial data formats (selection)

		Geospatial Format	Description	Application examples
Vector		Shapefile (.shp+.shx+.dbf+...)	for vector data storage and sharing, developed by ESRI (but also works in other GIS), widely known format, 2GB file limit	Research data mapping, analysis and visualization (points, lines, polygons) in GIS
		Keyhole Markup Language (.kml/.kmz)	Developed for Google Earth, can display various features (points, images, 3D models, text) using KML encoding (XML notation for geospatial data)	Visualization of data (e.g. sampling points saved in Google) in Google Earth and other 2D/3D GIS software
		GeoJSON (.geojson)	Open format for simple vector data with their attributes, based on JSON	Areas of interest (state and country outlines)
		GPS Exchange Format (.gpx)	Open format for vector data (waypoints, tracks, routes), based on XML schema for attribute annotation, info stored in tags	In software and smartphone applications to tag information to location (e.g. Image and name to a place)
Raster		GeoPackage (.gpkg)	Applicable for vector and raster data, non-proprietary format developed by Open Geospatial Consortium, data stored in a SQLite library	Format widely used in QGIS projects, encoding format for INSPIRE data, suitable for desktop and mobile applications
		GeoTIFF (.tif/.geotiff)	GeoTIFF follows a Open Geospatial Consortium standard to georeference TIFF raster images, allows loss-free storage of images, can contain multiple bands/images in one file	Satellite and aerial imagery (Sentinel-2 satellite specialized for vegetation), modelled and inter/extrapolated datasets (examples: global landcover maps, national weather prediction maps)
		NetCDF (.nc)	Self-describing file format for array-oriented scientific data, applicable to grid data, defines dimensions (spatial & temporal), variables and attributes	Meteorological forecasting, atmospheric and oceanographic modelling, geophysical modelling and data storage and sharing

Metadata Standards for Geospatial Data

What is metadata?

- Metadata describes data
- Who: collectors, owners, contact person/institution
- What: topic, parameters, quality, project, funder
- Where: location of collection, of processing, of storage
- When: date and time of collection, of processing, of publication
- How: Methods for collection/analysis/etc., access rights (license), cost

Metadata Standards:

- Guidelines, conventions and protocols describing data by providing metadata that is relevant for the dataset
- Standardized metadata schemas support understanding, managing, and sharing of data
- supports FAIR data management
- For research with geospatial data:
metadata with standardized annotation for the spatial aspects of described data

from <https://www.ncei.noaa.gov/resources/metadata>

Metadata Standards for Geospatial Data

Metadata Standards define what metadata needs to be stored, topics cover:

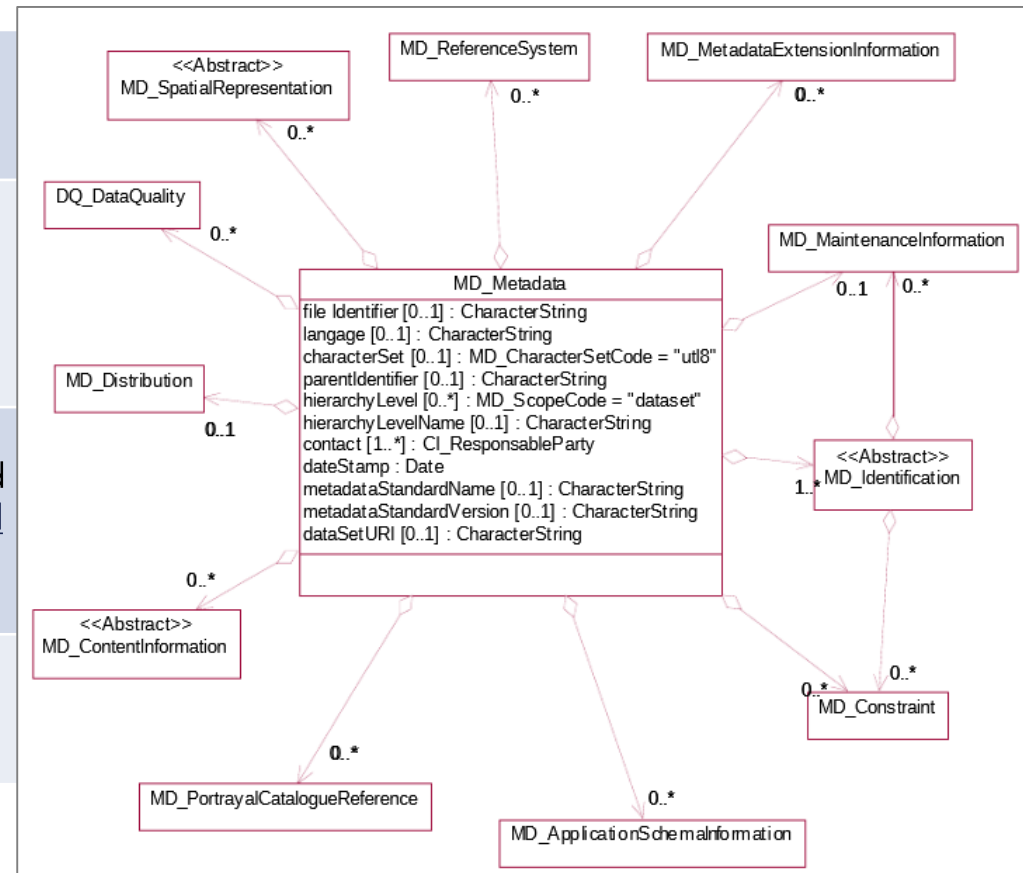
- Title
- Creator/Collector/Author
- Description/Abstract
- Keywords
- Geographic Location/Coverage
- Temporal Coverage
- License
- ...

Further, detailed topics can be:

- Taxonomic information
- Habitat description
- Study design
- Sampling Protocol
- Sampling Methods
- Analysis Methods
- Data Quality
- Data Repository
- ...

Metadata Standards for Geospatial Data: Examples

Metadata Standard	ISO 19115
Purpose/ Background	Metadata standard in general for geographic information Highly adaptable for a wide range of spatial data
Selected example for standard application:	Metadata structure for datasets in PANGAEA and for the INSPIRE Geoportal of the EU follow the ISO 19115
Source for Standard	published current version DIN EN ISO 19115-1:2014-07 , not free



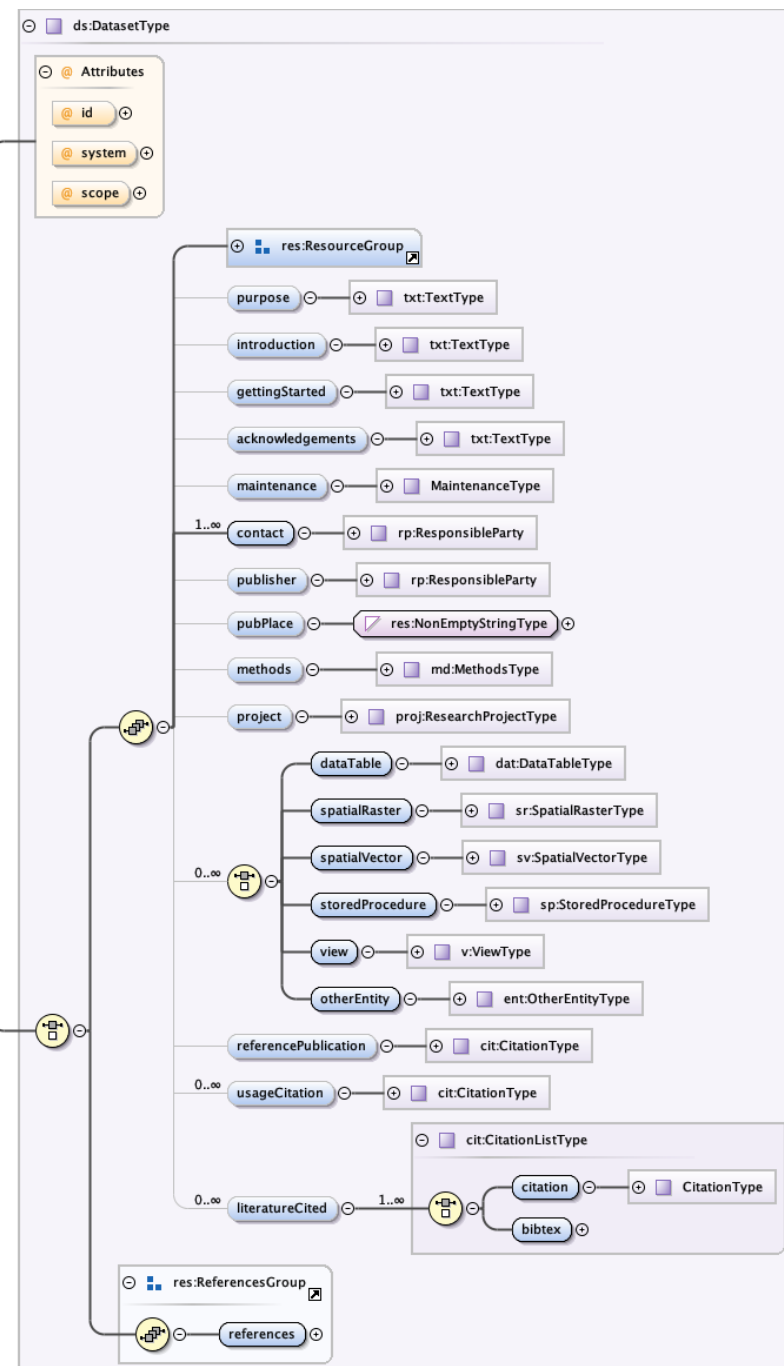
From Pierkot et al. (2005): Modelling heterogeneous and distributed spatial datasets to support updates management. ([Link](#))

Metadata Standards for Geospatial Data: Examples

Metadata Standard	Ecological Metadata Language (EML)
Purpose/ Background	designed to specifically describe ecological research data
Selected example for standard application:	The EDI and the KNB work with the EML standard
Source for Standard:	Jones et al. (2019): Ecological Metadata Language version 2.2.0. KNB Data Repository. doi:10.5063/F11834T2

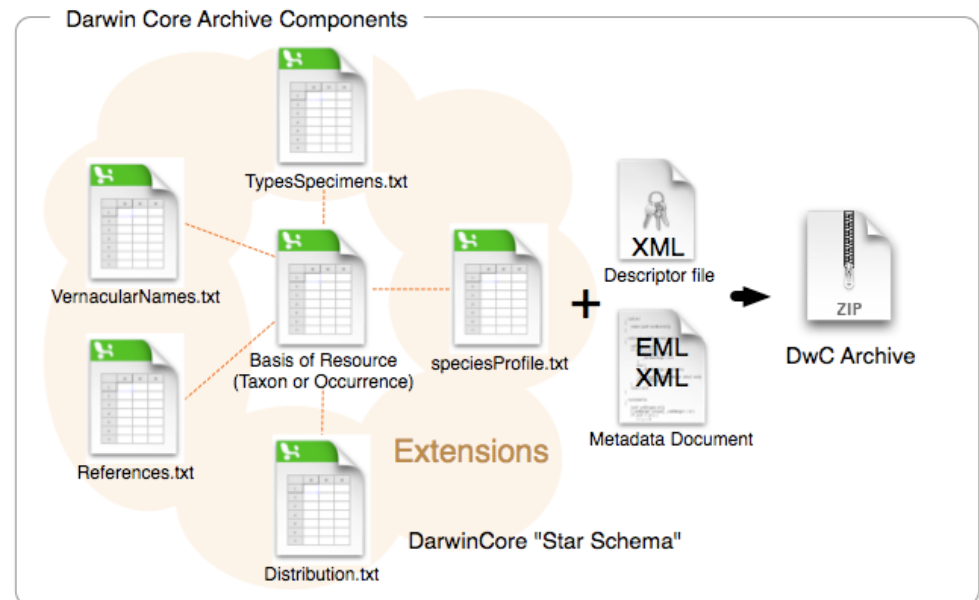
Schematic graphical representation of the EML structure for the dataset and other entities:
https://eml.ecoinformatics.org/schema/eml-dataset_xsd

dataset
Type ds:DatasetType



Metadata Standards for Geospatial Data: Examples

Metadata Standard	Darwin Core (DwC)
Purpose/ Background	applicable for biodiversity research with taxonomic information, particularly the occurrence and distribution of species, is an extension of the Dublin Core Metadata Element Set
Selected example for standard application:	OBIS and GBIF portals implement and promote applying the DwC
Source for Standard:	Documentation and further links: https://dwc.tdwg.org/



GBIF IPT User Manual, Darwin Core:
<https://ipt.gbif.org/manual/en/ipt/latest/darwin-core>

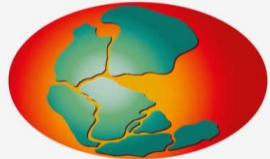
TRUST Principles for digital repositories



Transparency	To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.
Responsibility	To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.
User Focus	To ensure that the data management norms and expectations of target user communities are met.
Sustainability	To sustain services and preserve data holdings for the long-term.
Technology	To provide infrastructure and capabilities to support secure, persistent, and reliable services.

From Australian Research Data Commons: [Trust Principles](#) and Lin et al. (2020): The TRUST Principles for digital repositories. <https://doi.org/10.1038/s41597-020-0486-7>

Repositories and Access Points for Ecological Data



PANGAEA
Data Publisher for Earth &
Environmental Science.

Data Publisher for Earth & Environmental Sciences, operated by Helmholtz's AWI and MARUM (Uni Bremen), review and curation through editors, publishing with DOI, no cost, CoreTrustSeal



open non-profit data publishing platform committed to availability and re-use of all research data. Data is published in any format and any discipline. Data is curated and published with CC0 license and DOI, Data Publishing Charges, CoreTrustSeal



Easy publication and sharing without curation! For research items that are not suitable for publishing in subject-related repositories, but have a preservation value. Variety of metadata options (Dublin Core, DataCite Metadata Schema), DOI, no cost. Own responsibility for publication quality.

Platform to collaborate on **software projects** and share code with co-workers and scientific community. No DOI, but many platforms (e.g. Zenodo) integrate the option to connect a publication to GitHub projects.



international organization, goal to make the world's biodiversity data accessible, makes data from different publishers available via GBIF portal



Foundation providing a platform that allows access to data from multiple member repositories, puts emphasis on education on data & data management



Re3data is a global registry of research data repositories that covers research data repositories from different academic disciplines. It is funded by DFG and sustained by a working group from multiple research institutions.