

Geographische Daten und Datenformate

In der [vorhergehenden Lektion](#) haben wir gelernt, dass **Informationen aus interpretierten Daten** bestehen und das Daten wiederum Messwerte, Signale oder Umfrageergebnisse sein können. Fügen wir diesen Daten nun noch einen Raumbezug hinzu (direkt oder indirekt), so können wir bereits von **geographischen Daten oder Geodaten** sprechen ('Geo' = Erde, 'Graph' = Zeichnen/Beschreiben). Der Raumbezug (oder die [Georeferenz](#)) also **macht den entscheidenden Unterschied unserer Geodaten zu „herkömmlichen Daten“** aus:

„Unter Geodaten versteht man digitale Daten, welchen auf der Erdoberfläche bestimmte räumliche Lagen zugewiesen werden können (Geoinformationen, Geobezug)“ (GISWIKI <http://giswiki.org/wiki/Geodaten>)

Geodaten = Daten + Raumbezug + Metadaten

Der Raumbezug

Als **Raumbezug** verstehen wir die **Verortung in einem räumlichen Referenzsystem** welches i.d.R. die Erdoberfläche oder der erdnahe Weltraum sein können. Dabei bezieht sich die Verortung auf ein geometrisches Formprimitiv wie Punkt, Linie oder Fläche. Wir unterscheiden den **direkten Raumbezug** und den **indirekten Raumbezug**.

Indirekter Raumbezug

Ein **indirekter Raumbezug** liegt dann vor, wenn die Verortung **uneindeutig oder beweglich** ist. So stellt eine **Adresse** einen indirekten Raumbezug dar, da sich die räumliche Verortung verschieben kann. Auch Gemeinden sind nicht eindeutig und beweglich verortet (z.B. verschiebt sich das geographische Zentrum einer Gemeinde durch Ausdehnung oder Eingemeindung).

Direkter Raumbezug

Ein direkter Raumbezug wird üblicherweise durch Koordinatenpaare (Koordinatentupel) beschrieben. Diese sind **eindeutig und unbeweglich**. Ein **Punkt** besitzt logischer Weise genau **ein Koordinatenpaar**, eine **Linie mindestens zwei** und eine **Fläche mindestens drei** Koordinatenpaare (sonst wäre es eine Linie 😊).



Es müssen nicht immer Koordinaten sein! Jeder Ort der Erde lässt sich auch mit **3 Worten** beschreiben! Glauben Sie nicht? Versuchen Sie es: *Knaben.Pfeife.Tagebuch* → <https://what3words.com/de/>

Die Metadaten

Ein weiterer Bestandteil unserer Geodaten sind die **Metadaten** oder **Metainformationen**. Sie beschreiben unsere Geodaten hinsichtlich ihrer zeitlichen und räumlichen Ausdehnung, das verwendete Koordinatenreferenzsystem, die Maßeinheiten, die Art und Weise der Speicherung, die Zeichenkodierung etc. und können je nach Speichertyp auch sonstige Informationen enthalten (Urheber, Lizenz, Sprache, Kontaktinformationen, Farbvorschläge etc).

Unterscheidung nach Geobasisdaten und Geofachdaten

Unterscheiden wir Geodaten nach Ihrem Zweck, so differenzieren wir - zumindest in Deutschland - amtliche **Geobasis-** und **Geofachdaten**. Sie können als Raster (z.B. Digitale Topographische Karte) oder als Vektor (z.B. Flurstücke) vorliegen.

Geobasisdaten

Bei **Geobasisdaten** handelt es sich laut der [Arbeitsgemeinschaft der Vermessungsverwaltungen](#) (AdV) um Daten des amtlichen Vermessungswesens, die einen Raumbezug herstellen und die Landschaft sowie Gebäude und Grundstücke nachweisen und beschreiben, ohne einen Anwendungsbezug herzustellen. Die Geobasisdaten stellen damit, wie der Name schon sagt - die **Grundlage** oder eben die **Basis** (den „Hintergrund“) dar, **worauf die jeweiligen Fachdaten verortet werden** können.

| Abkürzung | Bezeichnung |
|-----------|---|
| ALKIS | Amtliches Liegenschaftskataster Informationssystem |
| ATKIS | Amtliches Topographisch-Kartographisches Informationssystem |
| AFIS | Amtliches Festpunktinformationssystem |
| DOP | Digitales Orthophoto |
| DLM | Digitales Landschaftsmodell |

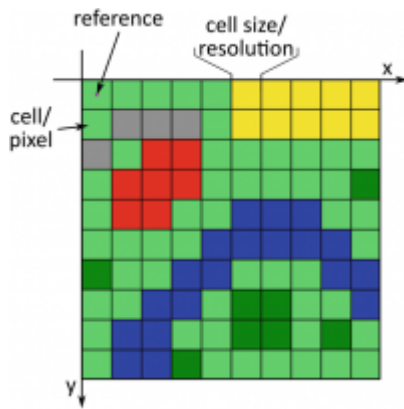
Geofachdaten

Geofachdaten stellen fachspezifische Informationen bereit, welche verortet auf den Basisdaten dargestellt werden können. Die Kombination aus Basis- und Fachdaten kommuniziert ein Fachthema auf einer Basiskarte, anhand welcher wir uns orientieren können. Geofachdaten können sehr unterschiedlich und vielfältig sein.

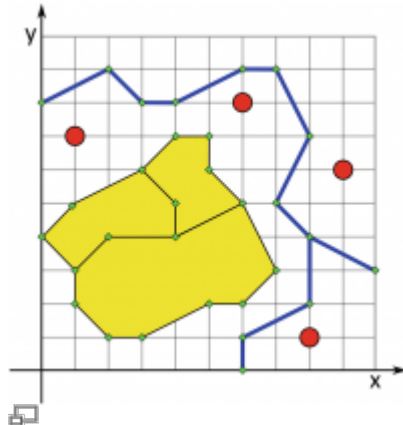
| Abkürzung | Bezeichnung |
|------------|---|
| BORIS | Bodenrichtwert Informationssystem |
| NIBIS | Niedersächsisches Bodeninformationssystem |
| CLC | CORINE Land Cover |
| Natura2000 | Natura2000 Schutzgebiete |

Raster- und Vektor-Geodaten

Betrachten wir die **Speicherung** von Geodaten, so unterscheiden wir **Raster-Geodaten** und **Vektor-Geodaten**.



Aufbau von **Raster-Geodaten**¹⁾



Aufbau von **Vektor-Geodaten**²⁾

Raster entstehen z.B. aus einem Scan, einem Satellitenfoto, einer Radar- oder Lasermessung. Die einzelnen Pixelwerte repräsentieren Messergebnisse. Diese können die an einem Fotosensor oder Scanner gemessenen Reflexionseigenschaften des aufgenommenen Objekts sein oder die Ergebnisse der in einem Gebietsraster verteilten Messstationen zur Lärm- oder CO²-Messung. Es handelt sich in diesen Fällen um **Primärdaten**. Weiterverarbeitet, z.B. durch eine Reklassifikation, durch Abdigitalisieren oder Vektorisieren erzeugen wir daraus **Sekundärdaten** - häufig in Form von **Geovektoren**.



Stammen die Daten „berührungsfrei“ von Satelliten, Flugzeugen oder Drohnen, so sprechen wir auch von **Fernerkundungsdaten**. (Satellitenbild, Höhenmodell, Multispektralaufnahme)

Geovektoren entstehen in den häufigsten Fällen aus der Weiterverarbeitung, Klassifikation, Analyse oder Modellierung etc. von Rasterdaten. Es handelt sich dann um **Sekundärdaten**. Sie besitzen neben Ihrer Verortung in einem Bezugssystem, anders als bei Rasterdaten, eindeutige **Geometrien**. Sie können im einfachsten Sinne diskrete **Punkte** sein aber auch **Strecken** oder **Polygone**.

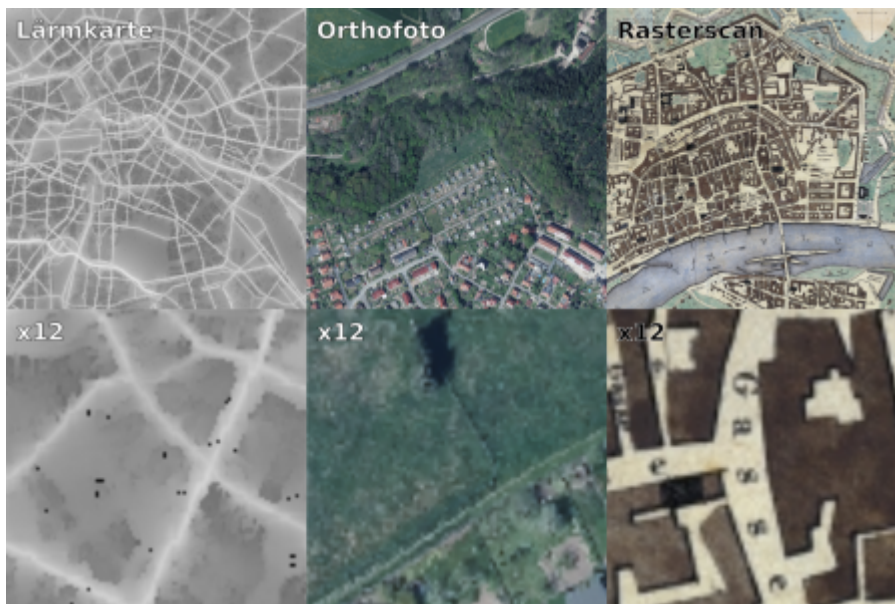
Während **Punkte** keine Ausdehnung besitzen und deshalb nur **ein einzelnes Koordinatenpaar** haben, bestehen **Linien aus mindestens 2 Koordinatenpaaren** und **Polygone aus mindestens 3** wobei hier eine Koordinate den Anfang sowie das Ende einer geschlossenen Linie markiert (Polygone sind eigentlich „Ringe mit Füllung“). All diesen Geometrien sind Informationen hinterlegt oder auch **Fachinformationen** oder **Attributwerte**. Alle Fachinformationen/Attributwerte finden wir zusammengefasst in der **Attributtabelle** wieder. Für **jede Geometrie, finden wir genau eine Zeile** in der Attributspalte wieder. Nehmen wir die drei Geometriebeispiele aus Abb. 3, so

erwarten wir für die drei Punkte auch 3 Attributzeilen und für die Linie und das Polygon jeweils nur eine Zeile in der Attributtabelle. Ein Geovektor besteht also immer aus:

- einer räumlichen Ausdehnung und Lage (Koordinaten, Ausmaße, Länge)
- einer Geometrie (Punkt/Linie/Fläche);
- Sachinformationen/Attributwerte (z.B. Naturschutzgebiet, Bahnlinie oder Mülleimer)

Bei Geoinformationssystemem kommen sowohl **Raster-** und **Vektordaten** als auch **Datenbanken** und einfache **Text-** oder **Tabellen-Dateien** zum Einsatz. Eine Übersicht der wichtigsten Geo-Dateiformate mit Entwickler und Kompatibilität finden Sie auf Wikipedia: [W GIS-Dateiformat](#) Im Folgenden sind die wichtigsten Speicher-Formate in Kürze erläutert:

Raster-Geodaten



Beispiele für Rasterdaten

Ein **GeoRaster** besteht aus Zeilen und Spalten wodurch sich **Zellen** bzw. **Pixeln** ergeben. Jede dieser Zellen/Pixel besitzt **genau einen Wert**. Im Unterschied zu einem herkömmlichen Rasterbild (z.B. ein Digitalfoto) besitzt das Georaster zudem einen **Raumbezug** und eine **geographische Ausdehnung**.

Vergrößern wir eine Rastergrafik, so sehen wir die Pixel aus welchen sie besteht. Bei digitalen Fotos würde man dann sagen: „das Bild ist verpixelt“. Aber tatsächlich sind alle Fotos *verpixelt*. Die Auflösung, also die **Dichte bzw. Größe** der Pixel, entscheidet darüber, ob wir es natürlich wahr nehmen oder eben *verpixelt* sehen.

Bei einem Georaster haben die Pixel eine **tatsächliche räumliche Ausdehnung**. z.B. eine Landsat³⁾-Bild hat eine räumliche Pixelauflösung von ca. 30×30 Meter, eine Orthofoto durch Beflug⁴⁾ kann 1×1 Meter oder noch höher aufgelöst sein.

[Weitere Informationen zu Rasterdaten im QGIS Handbuch](#)

Beispiele für Raster-Geodaten

| Abkürzung | Bezeichnung | Bedeutung |
|-----------|-----------------------------|-------------------------------------|
| DOP | Digitales Orthophoto | Georeferenziertes Luftbild |
| NIR | Nahes Infrarot | Georeferenziertes Infrarot Luftbild |
| DGM | Digitales Geländemodell | Höhen der Reliefoberfläche |
| DOM | Digitales Oberflächenmodell | Höhen der tatsächlichen Oberfläche |

Typische und häufige Raster-Speicherformate

Eines der häufigsten GeoRaster-Formate ist das [GeoTIFF](#) (*.tiff oder *.tif). Der Unterschied zum **herkömmlichen TIFF-Raster** ist nicht aus dessen Erweiterung ersichtlich - beide heißen gleich - doch ein GeoTiff besitzt eingebaute **Informationen zum Raumbezug** ([Koordinatenbezugssystem](#)). Dabei ist keine weitere Datei notwendig ⁵⁾.

Außerdem finden wir immer häufiger das relativ neue [JPEG2000 oder JP2](#), ein **für Geodaten besser geeignetes JPEG**, welche komprimiert gespeichert werden und einen **Raumbezug ohne World-Datei** haben kann. Sie sind relativ klein, aber **ungeeignet zum Speichern diskreter Werte** (z.B. wie sie in Rohdaten vorkommen; Digitale Geländemodelle, Lärmkarten etc.).

Bei Geländemodellen, Höhenmodellen und auch Punktwolken werden häufig einfache Text-Formate wie das [XYZ-Format](#) verwendet.

Auch das [GeoPackage](#) ist in der Lage Raster-Daten verlustfrei zu speichern. Noch finden sich GeoPackages relativ selten, trotz ihrer vielen Vorteile.

Eine vollständige Liste aller GeoRaster-Formate findet man hier: [GDAL Raster Format List](#)

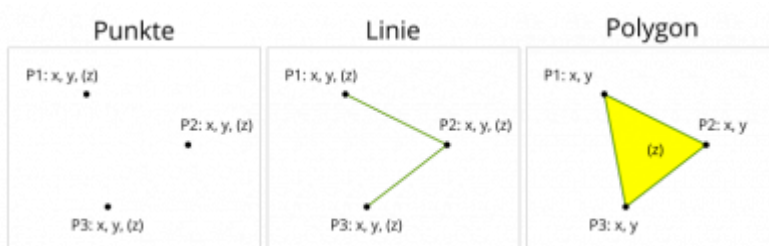


Ein Raster-Datensatz zum Ausprobieren und Üben finden Sie hier:

Harz-Raster-DEM

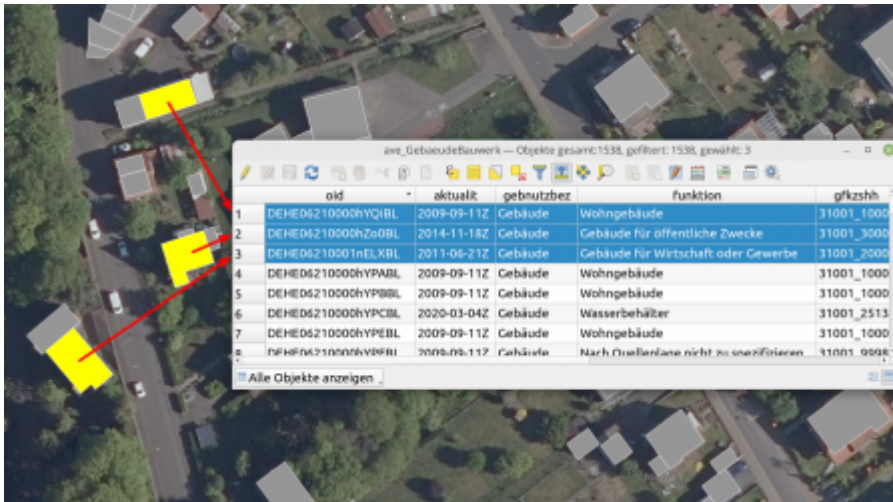
. Es handelt sich dabei um zwei SRTM-Szenen (Digitales Höhenmodell).

Vektor-Geodaten



Die drei Vektor-Grundformen: Punkte, Linien und Flächen

Eine **Vektorgrafik** besteht aus **Punkten, Kreisen, Linien** oder **Polygonen**. Die Lage eines Vektor-Punktes (**Node** oder **Stützpunkt**) wird durch **Koordinaten** beschrieben. Bei **GeoVektoren** sind dies **geographische Koordinaten**. Verbindet man diese diskreten Punkte miteinander, so erhält man Linien oder - wenn die Linie dort aufhört, wo sie begonnen hat und damit eine Fläche einschließt - Polygone. Die *Strecken* zwischen den Stützpunkten können als **Segment** bezeichnet werden. Da die Segmente immer Strecken zwischen zwei Punkten sind und Punkte theoretisch unendlich klein sind, sind Vektordaten „frei skalierbar“ - sie *verpixeln* nie!



Jede Geometrie besitzt genau eine Zeile in der Attributtabelle (Auswahl: Gelb in der Karte, Blau in Attributtabelle)

Von ganz besonderer Bedeutung ist die **Verknüpfungen zu Attributdaten (auch Sachdaten genannt) in der Attributtabelle bzw. Sachdatentabelle**. Jede **Vektorgeometrie besitzt demnach eine Zeile in der Attribut-/Sachdatentabelle** und eine theoretisch beliebige Anzahl von Attributen. Neben der freien Skalierbarkeit von Vektordaten ist das die wichtigste Eigenschaft und bedeutendster Unterschied zu Rasterdaten!

Betrachtet man den nebenstehenden Screenshot, erkennt man den Zusammenhang zwischen Geometrie (Gebäudeumriss) und Zeile der Attributtabelle. Jede Geometrie hat **exakt einen Eintrag in der Attributtabelle**.

Eine Ausnahme machen hier die sog. **Multi-Geometrien⁶⁾**. Diese können aus mehreren Geometrieteilen bestehen welche sich eine Zeile der Attributtabelle „teilen“ → *Beispiel: Zwei Gebäudeteile (Garage + Wohnhaus) haben die gleichen Attributeigenschaften da es sich thematisch auch um ein Gebäudekomplex handelt - also auch nur eine Attributzeile*

Ob es sich um einen Single oder Poly-Typ handelt, wird beim Erzeugen der Daten entschieden. Die Entscheidung dazu muss wohl überdacht werden!

[Weitere Informationen zu Vektordaten im QGIS Handbuch](#)

Typische und häufige Vektor-Speicherformate

Derzeit findet das von ESRI™ aus den 90ern stammende [W Shapefile](#) noch am häufigsten Verwendung. Allein der Begriff *Shapedatei* oder einfach nur *Shape* hat sich in der GIS-Branche als **Art Standard** etabliert. Dabei gibt es viele anderer GeoVektor-Formate, die durchaus Vorteile und Vorzüge haben (SpatialLite, GeoPackage...). Dennoch: Das Shapefile ist nach wie vor das **Top Austausch-Format** und wird es noch einige Zeit bleiben.

Das [W Shapefile](#) ist aber **nicht nur eine** Datei, sondern besteht aus **mindestens drei voneinander abhängigen Dateien**:

| Erweiterung | Funktion | Erforderlich? |
|---------------|--|---------------|
| *.shp | Speicherung der Geometriedaten | Ja |
| *.dbf | Sachdatentabelle (Attributtabelle) | Ja |
| *.shx | Indexdatei zur Verknüpfung der Geometrien mit den Attributwerten (Sachdaten) | Ja |
| *.prj | Projektion der Daten (nicht vorausgesetzt aber wichtig!) | Ja/In |
| *.cpg | Zeichensatz der Attributtabelle (z.B. UTF-8 für Sonderzeichen und Umlaute) | Ja/In |
| *.atx | Attributindex | Nein |
| *.sbx & *.sbn | Räumlicher Index | Nein |
| *.qix | GDAL-Index | Nein |
| *.aih & *.ain | Index für Tabellenverknüpfungen | Nein |
| *.shp.xml | Metadaten zum Shapefile | Nein |
| *.qml | QGIS-Stildatei (damit lassen sich QGIS-Stilinformationen mit austauschen) | Nein |



Shapefiles haben diverse Einschränkungen: Die **maximale Größe** eines Shapefile beträgt **2GB**; die **Länge der Feldnamen** (Attributnamen) ist auf **10 Zeichen** begrenzt, es können maximal **255 Felder** (Attribute) gesetzt werden, **Fließkommazahlen werden als Text gespeichert** und können **Rundungsfehler** enthalten...

Alternativen zum Shapefile sind [W SpatialLite](#), [W GeoJSON](#), [W KML/KMZ](#) oder ganz besonders das [GeoPackage](#). Alle haben ihre Vor- und Nachteile. Arbeitet man viel mit QGIS, rate ich zu SpatialLite oder GeoPackage - *Ein-Datei-SQLite-Datenbanken*. Möchte man mit **Raster** und **Vektoren** in einer Datei arbeiten, empfiehlt sich GeoPackage.

Eine Zusammenfassung aller OGR-GeoVektor Formate findet man über die [OGR Vector Format List](#)

(Kommagetrennte) Texte und Tabellen

Sofern Text- oder Tabellendateien geographische Koordinaten besitzen, können diese auch als Geodaten bezeichnet und entsprechend in einem GIS angewendet werden. Die Einfachheit der Text- oder Tabellen-Dateien ist dabei deren Vorteil:

Beispiel: Der folgende Inhalt einer *.txt oder *.csv Datei kann in QGIS geöffnet werden:

```
name;x;y
Punkt1;464391.784;5662142.072
Punkt2;464391.784;5671642.072
Punkt3;478241.784;5662142.072
Punkt4;478241.784;5671642.072
```

Da Text-Dateien unterschiedlich aufgebaut sein können und beispielsweise unterschiedliche Trenn- oder Dezimalzeichen verwendet werden können, sollte der Import von Texten über die Datenquellenverwaltung → Getrennte Texte erfolgen. Hier hat man die Möglichkeit QGIS mitzuteilen, wie die zu importierende Datei aufgebaut und zu interpretieren ist. Mehr dazu in der Lektion **Import zeichengetrennter Texte und Tabellen**



Hin und wieder stößt man auf ***.xyz-Dateien**. Hierbei handelt es sich ebenfalls um einfache getrennte Texte ohne Header oder Metabeschreibung. Häufig werden darin **Punktwolken** (ungridded ASCII) oder **DGMs** (gridded ASCII) ⁷⁾ abgespeichert. QGIS kann in letzterem Fall das *.xyz direkt als DGM einladen ⁸⁾.

GeoDatenbanken



GeoDatenbanken oder räumliche Datenbanken sind **objektrelationale Datenbanken** in welchen **Geoobjekte** mit **Geometrie** und **Raumlage** relational gespeichert werden.

Geodatenbanken ermöglichen eine **effiziente** und **schnelle Verwaltung** von Geodaten, eine einfache und **reccourcenschonende Speicherung** sowie gute **Backupfunktionen**. Mit Hilfe einer spezifischen Syntaxsprache können komplexe **Filter** gesetzt, **Berechnungen** durchgeführt und **Abfragen** erstellt werden. Datenbanken sind die **erste Wahl bei umfangreicheren GIS-Projekten**.

Bekannte räumliche Datenbanken sind [PostgreSQL-PostGIS](#) und [SQLite-Spatialite](#). Auch **Oracle** und **Microsoft Access** Datenbanken können zu räumlichen Datenbanken erweitert werden.

Zusammenfassung



- Geodaten sind Daten mit Raumbezug
- Der Raumbezug kann direkt oder indirekt sein
- Geodaten können Raster oder Vektoren sein
- Rasterdaten sind i.d.R. Primärdaten
- Sekundärdaten sind abgeleitete Daten (z.B. Digitalisierung eines Orthofotos)
- Geodaten besitzen neben dem Raumbezug auch Metainformationen
- Metainformationen beschreiben die Geodaten näher
- Vektordaten bestehen aus Geometrien und dazugehörigen Attributen
- Wir unterscheiden in Deutschland Geobasis und Geofachdaten

1) , 2)

Pászto V., Redecker A., Macků K., Jürgens C., Moos N. (2020) Data Sources. In: Pászto V., Jürgens C., Tominc P., Burian J. (eds) Spationomy. Springer, Cham. https://doi.org/10.1007/978-3-030-26626-4_1

3)

Serie ziviler Erdbeobachtungssatelliten der NASA zur Fernerkundung der kontinentalen Erdoberfläche sowie der Küstenregionen.

4)

mit Flugzeugen, Hubschraubern oder UAV's

5)

viele geographischen Raster-Formate benötigen zusätzlich eine sogenannte **World-Datei** welche den Raumbezug wieder gibt

6)

Multipunkt, Multilinie, Multipolygon

7)

z.B. LiDAR-Modelle aus Laserscan-Daten

8)

wenn nicht, dann:

<https://gis.stackexchange.com/questions/323117/rasterizing-ungridded-xyz-lidar-files>

From:

<https://lernplattform.map-site.de/> - **Lernplattform für OpenSource GIS**

Permanent link:

https://lernplattform.map-site.de/doku.php/qgis/basic/a_grundlagen/al01_geographische_datenformate

Last update: **2022/05/16 10:39**

