



## Tutorial: Datenformate



### Datenformate und deren Nutzung: Tipps und Tricks

In dieser Lerneinheit sollen gängige Datenformate beschrieben und deren Nutzung bei der Verarbeitung offener Geodaten erläutert werden.

Zudem werden Tipps und Tricks gesammelt, um die gemeinsame Verarbeitung der Daten zu erleichtern.



Bildquelle: <sup>1</sup> <sup>2</sup>

### Einführung

Für die Verarbeitung offener Geodaten kommen die unterschiedlichsten Formate zum Einsatz. Dies hängt zum einen davon ab, mit welcher Software gearbeitet werden soll, zum anderen aber auch davon, in welchem Format der Anbieter offener Daten diese bereitstellt. In Abbildung 1 (s. Abb. 1) wird diese Formatvielfalt am Beispiel von GovData ([www.govdata.de/](http://www.govdata.de/)), dem offenen Datenportal des Bundes und der Länder, illustriert.

Man erkennt deutlich, dass knapp die Hälfte textbasierte Formate wie csv, html oder xml sind. Gut 10 % werden im Excelformat, also dem proprietären Format von Microsoft, bereitgestellt. Zusammen mit der textbasierten Form csv lässt sich erahnen, dass es sich bei etwa einem Drittel der angebotenen Daten um Sachdaten/Attribute im Sinne einer Weiterverarbeitung in Geo-Informationssystemen (GIS) handelt. Knapp 10 % sind pdf-Dokumente (z.B. Gesetzestexte, Berichte etc.), die einer Weiterverarbeitung eher schlecht zugänglich sind.

Bei nahezu einem weiteren Drittel der offen bereitgestellten Daten handelt es sich im weitesten Sinne um Geoinformationen, die in unterschiedlichster Verarbeitungsreife herunterladbar sind, mit etwa 25% meist nur als Kartendienst, also geeignet zur Betrachtung oder als Hintergrundkarte. Nur zu 7% stehen wohl direkt in GIS verwertbare Geodatenformate zur Verfügung.

<sup>1</sup> Bildquelle: [Icons made by <a href="http://www.freepik.com" title="Freepik">Freepik</a> from <a href="https://www.flaticon.com/" title="Flaticon">www.flaticon.com</a> is licensed by <a href="http://creativecommons.org/licenses/by/3.0/" title="Creative Commons BY 3.0" target="\\_blank">CC 3.0 BY</a></div>](http://www.freepik.com)

<sup>2</sup> [Icons made by <a href="https://www.flaticon.com/authors/smashicons" title="Smashicons">Smashicons</a> from <a href="https://www.flaticon.com/" title="Flaticon">www.flaticon.com</a> is licensed by <a href="http://creativecommons.org/licenses/by/3.0/" title="Creative Commons BY 3.0" target="\\_blank">CC 3.0 BY</a></div>](https://www.flaticon.com/authors/smashicons)

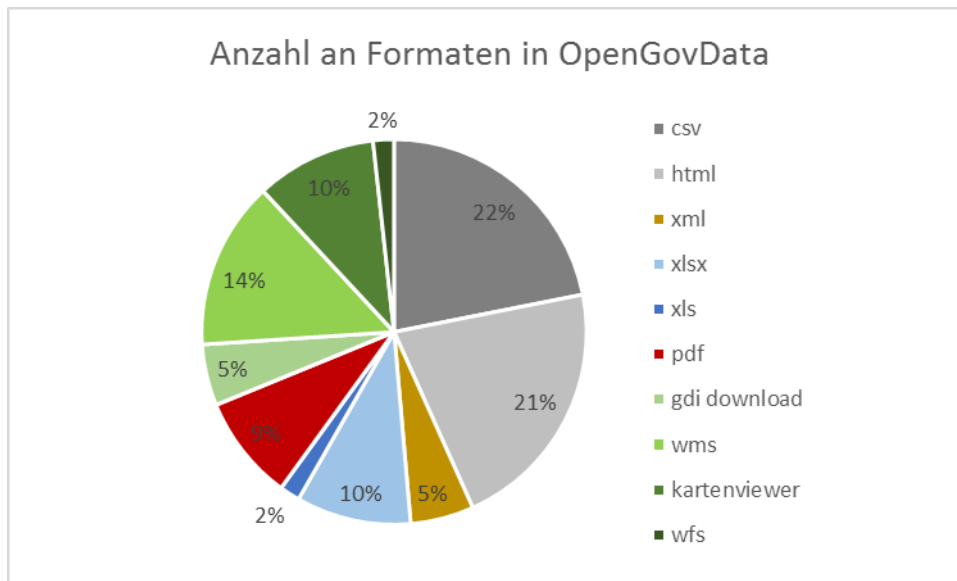


Abb. 1: Formate in OpenGovData (Stand 28.9.2017, 20693 Treffer, 42761 Dokumente aufgrund von Mehrfachformatangeboten).

## Grundlegende Formate

### Binärdateien

Diese Dateien folgen keinem standardisierten Zeichensatz, und haben oft keine Dateiendung. Sie enthalten beliebige Folgen von Bits, ohne eine direkt zugehörige Interpretation. Diese wird erst von dem Programm, mit dem sie ausgelesen werden, mitgebracht. Viele Programme definieren ihre eigenen Dateiendungen, um die Zugehörigkeit zu kennzeichnen. Öffnet man eine Binärdatei in einem Texteditor, sind oft nur unlesbare Zeichen vorzufinden.

### Textdateien (TXT)

Diese Dateien sind unter Windows meistens als .txt geführt. TXT-Textdateien haben keine vorgeschriebene Struktur oder Notation, sondern sind wenig mehr als ein Hinweis für das Betriebssystem, die binären Dateiinhalte als sequentielle Folge von Zeichen aus einem bestimmten Zeichensatz zu interpretieren. Welcher Zeichensatz dabei angenommen wird, hängt von Betriebssystem und gewähltem Editor ab, und kann oft auch im selbigen verändert werden. Die häufigsten Beispiele sind hierbei ASCII oder UTF-8. Wird der Zeichensatz falsch interpretiert kommt es oft zu unlesbaren Fehldarstellungen.

Im Kontext von Datenverarbeitung wird oft eine bestimmte Struktur in .TXT Dateien angenommen oder dynamisch festgestellt, da jedes andere textbasierte Format prinzipiell in einer solchen Datei gespeichert werden kann. Häufig sind die Dateiinhalte in solchen Fällen z.B. wie eine CSV-Datei formatiert.

## Strukturierte Textformate

### CSV-Format

CSV steht für Comma-separated values (seltener Character-separated values) und beschreibt den Aufbau einer Textdatei zur Speicherung oder zum Austausch einfach strukturierter Daten. In CSV-Dateien können Tabellen oder eine Liste unterschiedlich langer Listen abgebildet werden. Die Dateinamenserweiterung lautet .csv.

Ein allgemeiner Standard für das Dateiformat CSV existiert nicht, jedoch wird es im RFC 4180 grundlegend beschrieben. Die zu verwendende Zeichenkodierung ist ebenso wenig festgelegt; 7-Bit-ASCII-Code gilt weithin als der kleinste gemeinsame Nenner.

Im nachfolgenden Auszug ist ein weltweiter Länderdatensatz mit Kürzeln für Nationennamen sowie dem vollen Namen, der Flächengröße, der Bevölkerungszahl, einer regionalen Untergliederung vom Kontinent zur Region sowie Zentrumskoordinaten in geographischer Länge und Breite dargestellt, wie der ersten Zeile im Sinne der Spaltenüberschriften entnehmbar ist. Bezüglich Interpretation muss meist auf einen Begleittext verwiesen werden, der z.B. die Dimensionen, Abkürzungen etc. ausführlich beschreibt.

**Auszug 1:** Aussehen einer csv-Datei im Editor am Beispiel eines weltweiten Länderdatensatzes

```
FID_;FIPS;ISO2;ISO3;UN;NAME;AREA;POP2005;REGION;SUBREGION;LON;LAT
;AC;AG;ATG;28;Antigua and Barbuda;44;83039.000000000000000;19;29;-61.7830009;17.0779991
;AG;DZ;DZA;12;Algeria;238174;32854159.000000000000000;2;15;2.6320000;28.1630001
;AJ;AZ;AZE;31;Azerbaijan;8260;8352021.000000000000000;142;145;47.3950005;40.4300003
;AL;AL;ALB;8;Albania;2740;3153731.000000000000000;150;39;20.0680008;41.1430016
```

**Tipps und Tricks**

csv-Dateien können mit den unterschiedlichsten Texteditoren (Editor, Wordpad) geöffnet und darin auch einfach modifiziert werden (Suche nach bestimmten Zeichen, Veränderung der Trennzeichen, Änderung von Sonderzeichen und deutschen Umlauten), falls verarbeitende Programme damit Probleme haben.

**XML-Format**

Die Extensible Markup Language (XML) ist eine Auszeichnungssprache zum Strukturieren von Dokumenten und Daten. Auszeichnungssprachen (engl. Markup Languages) geben Regeln vor, mit denen Teile eines Textes in bestimmter Weise markiert (ausgezeichnet) werden können, um ihnen zusätzliche Semantik und Eigenschaften hinzuzufügen, meist mit dem Ziel einen Text maschinenlesbar zu machen.

XML ist eine vereinfachte, erweiterte Version der SGML (Standard Generalized Markup Language), und ist damit genau wie diese eine Metasprache: Eine Sprache, mit der je nach Anwendungsfall eine spezifischere Sprache definiert werden kann.

Aufgebaut ist XML hauptsächlich aus Paaren von Tags, die sich wie Klammern verhalten. Wird ein Tag A in Tag B geöffnet, muss Tag auch wieder vor Tag B geschlossen werden. Einige Tags stehen für sich alleine und öffnen keine „Klammer“. Attribute können direkt in den Tags spezifiziert werden, die restlichen Inhalte stehen zwischen den Paaren.

**Auszug 2:** Beispielhafte XML-Datei

```
<person>
  <name> John </name>
  <isAlive> true </isAlive>
  <age> 25 </age>
  <address>
    <cityStreet> New York, 21 2nd Street </cityStreet>
    <postalCode> 10021-3100 </postalCode>
  </address>
```

```
<children> </children>
<spouse> </spouse>
</person>
```

## JSON-Format

Die Programmiersprache JavaScript wurde mit einem Fokus auf Webanwendungen und Objektorientiertheit entworfen. Da das Web zu großen Teilen auf textbasierten Protokollen basieren sollte, wurde ein allgemeingültiges, strukturiertes Datenformat für Objekte benötigt. Aus dieser Idee entstand die JavaScript Object Notation, kurz JSON. Als Austauschformat ist JSON heute der größte Konkurrent zum XML-Format.

Es ist nicht nur textbasiert, sondern auch gut von Menschen lesbar. Objekte bestehen aus Mengen von *Attribut-Wert-Paaren*. Die Objekte selbst haben dabei standardmäßig nach der Tradition von JavaScript keine festgelegte Struktur, weitere Attribute können nach Belieben benannt und hinzugefügt werden. Die Attribute sind textuell repräsentiert, für die Werte gibt es folgende Typen:

- Number: Wobei kein Unterschied zwischen Integer und Float gemacht wird.
- String: Eine Folge von Zeichen (Chars), durch Anführungszeichen markiert.
- Boolean: Die möglichen Werte sind True oder False
- Array: Eine geordnete Liste von Werten, durch eckige Klammern markiert und mit Kommas getrennt.
- Objekt: Eine ungeordnete Menge von Attribut/Wert paaren, durch geschweifte Klammern markiert und mit Kommas separiert, die Paare selbst werden durch einen Doppelpunkt geteilt. Gleiche Notation wie die übergeordneten Objekte.
- null: Ein leerer Wert, durch das Wort ‚null‘ ausgezeichnet

### Tipps und Tricks

JSON Dateien sind gut menschenlesbar, oft übersichtlicher als XML-Dateien mit den gleichen Inhalten. Fehler sind jedoch teilweise schwerer festzustellen als in XML, da die Inhalte und Notation weniger strikt sind.

### Auszug 3: Beispielhafte JSON-Datei

```
"Person": {
  "name": "John Smith",
  "isAlive": true,
  "age": 25,
  "address": {
    "cityStreet": "New York, 21 2nd Street",
    "postalCode": "10021-3100"
  },
  "children": [],
  "spouse": null
}
```

## Dokumentformate

### HTML-Format

Eine auf der SGML (Standard Generalized Markup Language) basierende Auszeichnungssprache für Hypertext-Dokumente. Hypertext-Dokumente sind strukturierte, digitale Dokumente, die untereinander verlinkt werden können. HTML enthält standardmäßig Unterstützung für Hyperlinks, Bilder, Metainformationen und anderen Inhalten. Dokumente sind in HTML-Kopf (Header) und HTML-Körper (Body) unterteilt. Der Header enthält allgemeine Informationen wie Titel und Metadaten, der Body die eigentlichen, dargestellten Inhalte.

HTML legt dabei hauptsächlich die Semantik fest, enthält aber von sich aus keine Hinweise zur Formatierung. Die Formatierung wird oft in separaten CSS-Dateien (Cascading Style Sheets) festgelegt, und kann auch je nach Browser variieren. Werden dynamisches Verhalten und komplexere Funktionen benötigt, kann außerdem JavaScript-Code hinzugefügt werden.

Aufgrund ihrer Wichtigkeit im heutigen Internet, wird die Sprachspezifikation stetig erweitert. So wurden Versionen der Sprache entworfen, die nicht auf SGML, sondern auf XML basieren. Die neueste Erweiterung ist HTML5, welche nicht mehr strikt SGML oder XML folgt, und unter Anderem stark erweiterte Unterstützung dynamischer und multimedialer Inhalte enthält.

#### Auszug 4: Beispielhafte HTML-Datei

```
<!DOCTYPE HTML>
<html>
  <head>
    <title> Webkarte </title>
    <meta charset="utf-8">
  </head>
  <body>
    <div id="map"> Load map here </div>
  </body>
</html>
```

### PDF-Format

Das Portable Document Format (PDF) wurde von Adobe System entwickelt, um es zu ermöglichen, Dokumente zwischen verschiedenen Plattformen transportieren zu können, so dass diese ihre originalgetreue Formatierung beibehalten. Typische Konvertierungs-Fallstricke werden dabei umgangen, in dem nicht implizit Strukturen wie „Überschrift“ oder „Absatz hier“ vorgegeben werden, sondern Positionen von Texten und Bildern explizit in einem Vektorformat festgeschrieben werden. Sogar verwendete Schriftarten können direkt in den .pdf-Dateien mitgeliefert werden. Das Anhängen weiterer Daten wie z.B. von Inhaltsverzeichnisse und Kommentaren ist auch möglich.

Dadurch das PDF als Austauschformat für fertiggestellte Dokumente entwickelt wurde, ist es nur schwer möglich diese zu verändern. Einige Programme erlauben zwar kleinere Veränderungen wie das Ausbessern von Tippfehlern, jedoch sollte das Arbeiten an PDF-Dateien immer über die Ursprungsdatei und erneutes Exportieren ins PDF Format erfolgen.

## Tabellen- und Datenbankformate

### XLS/XLSX/XLSM-Format

Für die Verarbeitung tabellarisch organisierter Daten eignet sich u.a. das Tabellenkalkulationsprogramm Microsoft Excel. Bei der Dateierdung .xls handelt es sich um das proprietäre Dateiformat der Versionen 97 bis einschließlich 2003, danach lautet die Endung .xlsx (bzw. .xlsm, sofern Makros eingebaut sind).

Excel bietet umfangreiche Berechnungen mit Formeln und Funktionen, eine Organisation in Arbeitsmappen, die Dateien entsprechen, aus Blättern, die in Registern angezeigt werden, und aus Zellen, die die Daten enthalten. Pro Arbeitsmappe sind bis zu 65.536 Tabellenblätter möglich. Die Zellen eines Tabellenblatts sind in Zeilen und Spalten eingeteilt und können über ein Zellbezugssystem angesprochen werden. Jede Zelle kann durch eine Kombination aus Buchstabe und Zahl eindeutig identifiziert werden, dem sogenannten Zellbezug, der aus Zeilen- und Spaltenangabe besteht. Die erste Zelle in der oberen linken Ecke heißt A1, wobei A die erste Spalte und 1 die erste Zeile bezeichnet und der Bezug genau genommen auch noch den Tabellenblattnamen einschließt, da Formeln in verschiedenen Blättern und Mappen den gleichen Bezug aufweisen können, etwa `Tabelle1!C4 + Tabelle3!C4`.

Für die Anzeige von Werten steht eine Reihe von Formaten zur Verfügung. Neben vorgefertigten Formaten, wie Datum und Uhrzeit und Sonderformaten für Postleitzahlen, können benutzerdefinierte Formate angegeben werden. Zellen können in Excel auch Text enthalten.

#### *Tipps und Tricks*

Beim Nutzen von Excel sollte auf sinnvolle Formate der spaltenweise angegebenen Merkmale/Attribute geachtet werden (statt Standard z.B. eine Zahl, bei der Nachkommastellen definiert werden können, Text für Beschreibungen, Datumsformate bei Zeitangaben usw.).

Hat man die Excel-Datei geöffnet, sollte geschaut werden, ob alle Zellen mit Werten belegt sind. Datenlücken sollten mit eindeutigen Werten versehen werden, die nicht im Wertebereich des Merkmals liegen. Hat man z.B. Luftaußentemperaturen in °C gemessen, so wird der Wertebereich i.d.R. so zwischen -25° und + 40° liegen. Ist kein Wert vorhanden, kann dies durch den Wert -99 klar kenntlich gemacht werden. Dies lässt sich dann in einer späteren Auswertung leichter herausfiltern.

Beim Einlesen von Excel-Dateien in GIS-Software kommt es oftmals zu Problemen in der Übertragung der Dateiformate in den Zellen. In der GIS-Software wird dann z.B. <NULL> angezeigt. Oftmals hilft hier ein einfaches Exportieren aus Excel heraus ins csv-Format. Die csv-Datei kann dann in das GIS-Programm eingelesen werden.

### DBF-Format

Das Dateiformat ist für dBase-Datenbanken definiert, einem der ältesten Datenbankmanagementsysteme. Auch wenn dBase selbst kaum noch in Benutzung ist, finden sich seine .dbf-Dateien immer noch in einigen langlebigen Anwendungen. Sie werden in ArcGIS-Shapefiles genutzt, und können nach wie vor sowohl von MS Excel, als auch von LibreOffice und Open-Office Calc geöffnet werden.

Es handelt sich um ein relativ komplexes Binärformat, im Grunde erlaubt es jedoch das Speichern einer Tabelle, mit Headern für die Gesamttabelle und für die einzelnen Spalten. Der Inhalt der Felder wird in ASCII (einem simplen Textformat) gespeichert. Im Spaltenheader finden sich Informationen

dazu, was für einen Datentyp der ASCII-Code in den Feldern der Spalte hat, wie er also interpretiert werden sollte.

#### *Tipps und Tricks*

Viele der Beschränkungen, auf die man bei der Nutzung von ArcGIS Shapefiles stößt, sind auf das Alter des DBF-Formats zurückzuführen. Für Spaltennamen ist nur eine sehr begrenzte Anzahl an Bytes vorgesehen, außerdem gibt es schnell Probleme mit Umlauten. Diese sind im ASCII-Standard nicht vorgesehen, und müssen mittels einiger ASCII-Codes, welche für Internationale Sonderzeichen freigelassen wurden, nachträglich interpretiert werden. Geschieht diese Interpretation nicht korrekt, kommt es zu Problemen.

Es sollte bei der Nutzung von Shapefiles also am besten auf lange Spaltennamen und Umlaute verzichtet werden.

## Vektordatenformate

### SVG-Format (basiert auf XML)

Das Scalable Vector Graphics (SVG) Format ist der Web-Standard für die Beschreibung von Vektorgrafiken, also Grafiken, die nicht über ihre Pixel definiert sind, sondern aus Primitiven wie Linien, Polygonen und Kurven bestehen, und dynamisch je nach Bildschirmgröße, Zoomstufe und Auflösung neu und in passender Auflösung gezeichnet (gerendert) werden können.

Für die Speicherung von Fotos sind sie damit nicht geeignet, jedoch sehr wohl für die vielen Illustrationen und Grafiken, die sich auf fast jeder Webseite finden. Da SVG auf XML basiert, handelt es sich hier also um ein textuelles Bildformat, welches auch von Menschen les- und veränderbar ist. Auch Animationen sind im Standard vorgesehen, ohne dass die XML-Datei dynamisch verändert werden muss. Dafür gibt es verschiedene Zeitskalen, auf denen Veränderungen festgeschrieben werden können, z.B. „Rotiere für 5 Sekunden“ oder „Ändere deine Farbe bei einem Mausklick“.

Das Format ist dabei hauptsächlich auf Darstellung ausgerichtet, und speichert keine zusätzlichen Informationen wie z.B. Raumbezug. Punkte können hier also z.B. nicht in einem übergeordneten Koordinatenreferenzsystem (CRS, siehe Tutorial Koordinatenreferenzsysteme) definiert werden, diese müssen relativ zur derzeit gegebenen „Leinwand“ (Canvas) mittels eines gefüllten Kreises gezeichnet werden.

#### Auszug 5: Beispielhafte SVG-Datei (Header Unvollständig)

```
<svg version="1.0" ...>
  <circle cx="10" cy="20" r="5">
  <polyline points="5,20 10,25 15,20 10,15">
</svg>
```

### DXF-Format

Das Drawing Interchange File Format (DXF) stammt aus dem Kontext von CAD-Programmen und wurde von der Firma Autodesk entwickelt. Es ist heutzutage der Industriestandard zum Datenaustausch zwischen verschiedenen CAD-Lösungen. Auf .dxf-Dateien wird dabei normalerweise nicht gearbeitet, sondern es wird in die Formate des genutzten Programmes im- und exportiert. Das Format kann sowohl textuell als auch binär gespeichert und verwendet werden.

## Rasterdatenformate

### JPEG-Format

Das JPEG-Dateiformat (eigentlich JFIF, für „JPEG File Interchange Format“) ist ein Standard zur Speicherung von Rasterbildern nach der JPEG-Norm. Die JPEG-Norm geht auf die Joint Photographic Experts Group der ISO zurück. Es existieren weitere derart genormte Dateiformate, .jpg ist jedoch das am häufigsten verwendete, auch wenn es nicht die vollständige Norm umsetzt.

Die Norm, und damit auch das Dateiformat, fokussiert sich auf das Anbieten verschiedener Kompressionsverfahren, je nach gewünschtem Anwendungszweck. Klassischerweise werden JPEG-Dateien heute verwendet um Fotos verlustbehaftet zu komprimieren und abzuspeichern. Die häufig zu sehenden blockigen Artefakte von im Internet verbreiteten Fotos sind meist auf zu häufige oder zu drastische JPEG-Kompressionen zurückzuführen. Das heutzutage seltenere, aber bei schwachem Internetempfang immer noch sichtbare Schritt-für-Schritt-Verbessern der Bildauflösung ist auch eine JPEG-Funktion.

Da die Speicherung oft ohne weitere Konfigurationsangebote vom genutzten Bildbearbeitungsprogramm vorgenommen wird, sind viele Anwender sich gar nicht bewusst, dass JPEG auch verlustfreie Kompression, die Auswahl verschiedener Farbräume und unterschiedliche Bildaufbauverfahren unterstützt.

#### *Tipps und Tricks*

JPEG sollte, wenn möglich, nur zum Speichern von Fotos genutzt werden, nicht von Grafiken mit vielen einfarbigen, flächenhaften Bildteilen. Die Kompressionsmethoden sind auf Fotos und die menschliche Wahrnehmung komplexer Bilder zugeschnitten, und sorgen auf niedrigeren Stufen für drastisch verringerte Bildgrößen bei kaum wahrnehmbarer Qualitätsverschlechterung. Die Kompressionsmethoden sind für Grafiken weniger effektiv, und beeinträchtigen die Qualität dieser schnell massiv. Besser ist hier PNG.

### PNG-Format

Bei PNG (Portable Network Graphics) handelt es sich um ein weiteres weit verbreitetes Datenformat für Bilder. Im Gegensatz zu JPEG ist es jedoch standardmäßig verlustfrei. Es fokussiert sich vor allem auf Grafiken, nicht auf Fotos, und unterstützt Funktionen wie Transparenz, unterschiedliche Farbmodi und Metadaten.

PNG wurde als freier und offener Ersatz für das GIF-Format entworfen, da dieses Anfangs patentierte Algorithmen benutzte und wenig Farbpräzision erlaubte. GIFs erfreuen sich jedoch nach wie vor einiger Beliebtheit, da sie im Gegensatz zu PNG Animationen erlauben.

### TIFF-Format

Weniger häufig im Internet vorzufinden als JPEG und PNG, ist TIFF, kurz für Tagged Image File Format. Der Grund hierfür ist, dass es sich hauptsächlich auf professionelle Anwender wie Grafikdesigner, Fotografen und Verlage konzentriert.

TIFF erlaubt ein präzises Anpassen der Bilddatei auf den gewünschten Verwendungszweck, durch die Verwendung verschiedener Tags, die im Bild hinzugefügt und von Programmen interpretiert werden können. Einige dieser Tags sind vorgegeben, wie z.B. das Definieren von genutzten Farbräumen, Kompressionsmethoden und vielen weiteren Schritten, die in anderen Bildformaten vorgeschrieben



sind. Tags können jedoch auch nach Belieben für jede Spezialanwendung definiert werden, und müssen dann nur von der entsprechenden Fachsoftware richtig interpretiert werden.

## Geodatenformate

### GeoTIFF-Format (basiert auf TIFF)

Hierbei handelt es sich weniger um ein eigenes Dateiformat, als um eine Sammlung von zusätzlichen Tags für das TIF-Format. Diese fügen Koordinatenreferenzsysteme und Projektionsinformationen zu der normalen TIFF-Datei hinzu. Kombiniert mit der Möglichkeit zur hohen Farbgenauigkeit, dem Speichern mehrere Bilder pro Datei und dem verlustlosen Speichern im TIF-Format, hat sich GeoTIFF zu einem Standard in der Raster-Geodatenverarbeitung entwickelt. Es wird z.B. häufig in der Verarbeitung von Satelliten- und Luftaufnahmen genutzt, bei denen es stark auf Genauigkeit ankommt.

### GeoJSON-Format (basiert auf JSON)

GeoJSON definiert das strukturierte Speichern räumlicher Informationen im Rahmen des JSON-Formats. Dafür müssen den räumlichen Objekten bestimmte Attribut/Wert-Paare zugeordnet werden, z.B. um welche Art von Geometrie es sich bei dem Objekt handelt als String, und die Geometriedaten selbst als Objekt.

#### Auszug 6: Beispielhafte GeoJSON-Datei

```
{
  "type": "Feature",
  "geometry": {
    "type": "Point",
    "coordinates": [125.6, 10.1]
  },
  "properties": {
    "name": "Dinagat Islands"
  }
}
```

### SHP-Format

Ein von der Firma ESRI entwickeltes und offen verfügbares Binärformat für räumliche Geometrien. Es ist eines der am weitesten verbreiteten GIS-Datenformate, und ist zumindest außerhalb des Webs der de-facto-Standard.

Shapefiles existieren dabei im Gegensatz zu anderen Formaten nicht nur als eine .shp Datei. Mehrere weitere Dateien mit gleichem Namen, aber anderen Endungen müssen zur korrekten Nutzung vorhanden sein, wie z.B. .dbf (Attributtabelle), .shx (Indextabelle) und .prj (Projektionsdaten). .shp selbst deckt nur die einzelnen Geometrien ab, das SHP-Format unterstützt jedoch auch Attributwerte, Indizierungen, Projektionen und vieles mehr. Diese werden jeweils in ihren eigenen Dateien abgelegt. Aufgrund der binären Formatierung sind Shapefiles nicht menschenlesbar, können allerdings in nahezu allen Programmen geöffnet werden, die den Import von Geodaten erlauben.

#### Tipps und Tricks

Viele Programme werden beim Importieren von Shapefiles nur die .shp-Datei im Dateibrowser anzeigen. Die restlichen benötigten Dateien werden automatisch nachgeladen. Gibt es ein Problem

mit einer Shapefile, sollte zuerst geprüft werden, ob die anderen Dateien auch vorhanden und im selben Ort gespeichert sind.

### **GML-Format (basiert auf XML)**

Die Geographic Markup Language (GML) ist sowohl als Norm definiert als auch durch das OGC spezifiziert. Hierbei handelt es sich um eine Spezialisierung von XML, mit einem Fokus auf raumbezogene Objekte. Sie erlaubt die Spezifikation von Objekten, Geometrien und Koordinatenreferenzsystemen, aber auch von weiteren Konzepten wie z.B. Kartenstilen. Der hauptsächliche Anwendungszweck von GML ist es, als Austauschformat für webbasierte Geoinfrastrukturen zu fungieren.

### **KML-Format (basiert auf XML)**

Ähnlich wie GML ist die Keyhole Markup Language (KML) eine geographische Spezialisierung der XML, ursprünglich jedoch fokussiert auf die Benutzung in Google Earth. Sie enthält daher nicht nur geometrische Informationen, sondern auch erweiterte Informationen zur Visualisierung der Daten. GML kann sogar mittels KML um Stilinformatoren erweitert werden: Beide sind Teil des OGC-Standards.

### **GeoRSS-Format (basiert auf XML)**

RSS-Feeds werden genutzt, um Änderungen im Inhalt eines Webdienstes zu veröffentlichen. Diese werden serverseitig festgestellt und gespeichert, und dann vom Empfänger aktiv abgerufen. Es müssen hier also keine serverseitigen Abonnentenlisten geführt werden. Nutzer benötigen jedoch ein Programm zum Abfragen und Abonnieren von Feeds. Die Veröffentlichungen sind dabei in XML formatiert, können also sowohl von Menschen direkt gelesen, als auch maschinell weiterverarbeitet werden.

Aufbauend auf RSS erlaubt GeoRSS dann, Georeferenzierungen standardisiert mitzupublizieren. Dies wird unter anderem durch die Nutzung von GML anstatt XML verwirklicht.

### **WKT-Format**

Ein OGC (Open Geospatial Consortium) Dateistandard für das textbasierte Speichern von Geodaten wie Vektoren und Koordinatenreferenzsystemen.

#### **Auszug 7: Beispiele für Objektrepräsentation im WKT-Format**

```
POINT X (4468298 5333781)
LINESTRING(1 12, 10 20, 5 6, 1 12)
GEOMETRYCOLLECTION( POINT(1 12), LINESTRING(1 12, 10 20))
```

### **Literatur**

Bill, R. (2016): Grundlagen der Geo-Informationssysteme. 6. Auflage. Wichmann Verlag, Offenbach-Berlin. 867 Seiten. Kapitel 8.6.

Seip, C., Korduan, P., Zehner, M.L. (2017): Web-GIS. Grundlagen, Anwendungen und Implementierungsbeispiele, Wichmann Verlag, Heidelberg, 552 Seiten.