

Übersicht über maschinenlesbare Dateiformate

Klassische Textdateien (*.txt), Open Document Formats (*.odt, *.ods) und Tagged Image File Format (*.tif, *.tiff) sind »offen« im Sinne offener Standards (kursiv gesetzte Formate), stellen aber aufgrund ihrer geringen Strukturierung oft eine große Herausforderung für die maschinelle Verarbeitung dar. Sie sollten nur verwendet werden, wenn dies von den Nutzern explizit gewünscht wird oder der Aufwand für die Umwandlung unverhältnismäßig ist. Dies trifft insbesondere auf Rasterbildformate zu. Ist die fehlerfreie maschinelle Extraktion der in einem Schriftstück enthaltenen Aussagen bereits aus einem Text-File (*.txt) schwierig, so kann aus einem im »offenen« PNG-Format reproduzierten Dokument schon der Text oft nicht fehlerfrei extrahiert werden. Andererseits kann es Anwendungsfälle geben, bei denen gerade die Nachnutzung der Rasterbilddaten gewünscht wird (z. B. Luftbilder).

Zum gegenwärtigen Stand der Technik kommen als »maschinenlesbar« u. a. folgende Dateiformate in Betracht:

Formatbezeichnung	Beschreibung	Dateiformat
»Texte«		
<i>Klassische Textdateien¹</i>	<i>bestehen aus darstellbaren Zeichen</i>	<i>.txt</i>
<i>Open Document Formats</i>	<i>XML-basiertes Dateiformat für Tabellen, Grafiken Präsentationen und Textverarbeitung</i>	<i>.odt, .ods</i>
Resource Description Framework	Sprache zur Beschreibung von Informationen über Ressourcen, die durch eindeutige Bezeichner (URIs) identifiziert werden	.rdf
»Tabellen«		
Comma Separated Value	Tabellen mit einfach strukturierten Daten	.csv
Open Document Formats	XML-basiertes Dateiformat für Tabellen, Grafiken und Präsentationen	.odt, .ods
JSON (JavaScript Object Notation)	Datenaustauschformat in einfach lesbarer Textform	.json
»XML-Formate«		
Extensible Markup Language	Beschreibungssprache zur Darstellung von strukturierten Daten in lesbarer Form	.xml
Newsfeed / Web-feed Syndication	XML-basiertes Schema zur Auslieferung von häufig geänderten Inhalten	.rss
»Geodatenformate«		
Geography Markup Language	Auszeichnungssprache zum Austausch und zum Speichern von geografischen Informationen im Vektorformat	.gml
GPS Exchange Format	XML-basiertes Datenformat zur Speicherung von Geodaten	.gpx

¹ Bei Nutzung des TXT-Formats muss sichergestellt sein, dass die konkret verwendete Zeichenkodierung aus dem Kontext ersichtlich ist. Gemäß SAGA Version de.bund 5.0.0 (Modul Technische Spezifikationen, Abschnitt 7.2) und gleichlautend Anlage §9(1)_SAGA_Sachsen.pdf Abschnitt 2.6 dieses Handlungsleitfadens soll UTF-8 verwendet werden. ISO 8859-1, IOS 8859-15 sowie UTF-16 genießen lediglich Bestandsschutz.

Formatbezeichnung	Beschreibung	Dateiformat
Keyhole Markup Language	XML-basierte Sprache für geografische Informationen zur Darstellung in 3D-Betrachtern	.kml, .kmz
GeoJason (statt ESRI Shapefile)	JSON basierter Standard zum Speichern von zwei dimensional Geographischen Daten (Punkte, Linien und Polygone)	JSON
ESRI Shapefile	Einfaches Geodatenformat im Desktop-GIS-Umfeld	.shp, .shx, .dbf, .prj
»Metadaten Formate«		
Web Catalogue Service	Dienst, der selbst keine Geodaten enthält, sondern lediglich beschreibende Metadaten	WCAS
»Schnittstellen Formate«		
Web Map Service	Dienst, um Kartensichten auf geografische Informationen in Bildformaten anzuzeigen	WMS
Web Feature Service	Downloaddienst bzw. optional zur Manipulation von Geodaten im Format GML	WFS
Web Map Tile Service	Geodienst der einen Webservice definiert, um digitale Karten kachelbasiert anzubieten und abzurufen (spezifiziert durch das Open Geospatial Consortium)	WMTS
OpenGIS® Catalogue Services Specification; ISO Metadata Application Profile (ISO 19115/ISO 19119)	Spezifikation von Diensten zum Suchen von Metadaten über Geodaten, Geodienste und Geoanwendungen	CSW
»Bilder«		
<i>Scalable Vector Graphics</i>	<i>Spezifikation zur Beschreibung von zweidimensionalen Vektorgrafiken</i>	<i>.svg</i>
<i>Graphics Interchange Format</i>	<i>CompuServe Rasterbild-Format (offen durch Patentablauf)</i>	<i>.gif</i>
<i>Portable Network Graphics</i>	<i>Grafikformat, das 16 Mio. Farben, verlustfreie Kompression und eine inkrementelle Anzeige unterstützt</i>	<i>.png</i>
<i>Joint Photographic Experts Group</i>	<i>Format für die Speicherung und den Austausch von Fotos und Grafiken mit Farbverläufen</i>	<i>.jpg .jpeg .jp2</i>
<i>Tagged Image File Format</i>	<i>Format zur Speicherung gerasteter Bilder</i>	<i>.tif .tiff</i>

Quellen: [Open-Government-Vorgehensmodell – Vorschläge zur Umsetzung von Open Government in Österreich \(Version 1.1\)](#) und [Open-Data-Rahmenkonzept der Stadt Leipzig vom 15.4.2014](#)