

Volume

1

ERSTELLEN VON GPKG AUS HARMONISIERTEN GML

SKRIPTBASIERTER ERSTELLUNG VON GEOPACKAGE AUS HARMONISIERTEN GML DATEIEN MITTELS GDAL

INSPIRE THEMENSPEZIFISCHE QGIS-KONFIGURATION

HANDBOOK

UMWELTBUNDESAMT WIEN

Erstellen von GPKG aus harmonisierten GML

INHALTSVERZEICHNIS

KAPITEL I

Geopackage	2
------------------	---

KAPITEL II

GDAL/GMLAS	4
GMLAS-Treiber	5

KAPITEL III

Skript für die Erstellung von GPKG aus GML.....	7
Workflow für die Erstellung des GPKG Ausführen des Skriptes....	8

Geopackage

Darstellung der Eigenschaften des Geopackage Formates und Begründung warum dieses speziell für das Encoding von harmonisierten INSPIRE Geodaten geeignet ist

GeoPackage (GPKG) ist ein offener, nicht proprietärer, plattformunabhängiger, auf bereits existierenden Standards aufbauender Standard, um Geodaten (Vektor- und Rasterdaten) in einer Datei zu speichern.

„Das GeoPackage vereint alle Vorteile einer räumlichen Datenbank mit dem Komfort einer einzelnen, kleinen und leicht austauschbaren Datei. In einer einzigen Geopackage Datei lassen sich sowohl (multi-)Linien, (multi-)Punkte, (multi-)Polygone sowie Rasterdaten und geometrieloze Tabellen und Texte abspeichern. Es gibt dabei kaum Einschränkungen“ (Quelle: <https://lernplattform.map-site.de/doku.php?id=qgis:advanced:lernpfad-a:lektion-3>). Daraus lassen sich zahlreiche Vorteile des GPKG gegenüber dem noch immer für den Datenaustausch weit verbreiteten und beliebten ESRI Shapefile Format ableiten (Quelle: <https://www.teuderun.de/workshops/qgis/geopackage/>):

- Unicode-Support
- Keine Begrenzung bei der Länge von Feldnamen oder Zeichenfolgen (bei Feldinhalten)
- Feldtypen für ganzzahlige oder Dezimalzahlen, Boolean, BLOB etc.
- NULL-Werte
- Maximale Dateigröße liegt bei 140 TB
- Es handelt sich um ein topologisches Austauschformat (=topologische Informationen werden gespeichert)
- Es können mehrere Geometrietypen pro Datei gespeichert werden, d.h. man kann ein gesamtes GIS-Projekt mit allen Features in einer einzigen Datei speichern. Das erleichtert die Weitergabe der Daten enorm.

Weitere Eigenschaften des GPKG sind:

- Neben Vektordaten können auch Rasterdaten in der GeoPackage-Datei gespeichert werden. Das Erstellen von Rasterkatalogen ist daher kein Problem.
- Es können nicht-lineare Geometrietypen (Kreisbögen, Splines) gespeichert werden.
- Metadaten und Informationen zur Projektion sind integriert.
- Attributdomänen (Stichwort "Gültigkeitsbereiche") und Subtypen können definiert werden.
- Die Erstellung von Tabellenbeziehungen ist möglich.

- Zahlreiche zusätzliche Informationen können in der Geopackage-Datei hinterlegt werden: Drucklayouts, Style-Dateien, SVGs oder Bilddateien (Logos), sogar eine oder mehrere QGIS-Projektdateien!
- Da das Format auf SQLite basiert, unterstützt Geopackage die Datenbanksprache SQL
- Gegenüber dem Shape-Format gibt es einen Dateigröße-Vorteil bei der Speicherung derselben Geodaten.

Durch die oben angeführten Eigenschaften stellt das GPKG eine gute und vor allem sehr nutzerfreundliche Alternative dar um harmonisierte INSPIRE Geodaten in einem alternativen Encoding anzubieten. Da GPKG auf SQLite aufbaut und somit die wesentlichen Eigenschaften einer räumlichen Datenbank aufweist, ist es möglich die vollständige Modellsemantik der INSPIRE Datenmodelle abzubilden. Darüber hinaus können für die einzelnen INSPIRE Themen spezifische View erstellt werden, die eine unmittelbare Nutzung von Informationen die in verschiedenen Tabellen gespeichert sind in GIS-Clients wie z.B. QGIS ermöglichen. Zusätzlich ermöglicht das GPKG die direkte Ablage von Styles welche eine themenspezifische kartographische Aufbereitung der Daten ermöglicht.

GPKG stellt somit ein Format zur Verfügung um die Daten welche teilweise in sehr komplexen INSPIRE Datenmodelle gespeichert vorliegen in nutzerfreundliche Art und Weise bereit zu stellen.

GDAL/GMLAS

GDAL/OGR und GMLAS Treiber zur Erstellung von GPKG aus GML

GDAL/OGR stellt ein kommandozeilenbasierte Hilfsprogramme zur Verfügung. Über diese können eine Vielzahl von Raster- und Vektorgeodatenformate umgewandelt und prozessiert werden.

Für die Prozessierung der harmonisierten INSPIRE Datensätzen welche in GML Encoding vorliegen wird des Weiteren der von der Europäischen Umweltagentur entwickelte Treiber GMLAS benötigt. GMLAS liest XML/GML-Dokumente auf Grundlage beliebiger, zugehöriger XML-Schemas (XSD). Aus dem Applikationsschema (=INSPIRE Datenmodell) auf welchen das jeweilige GML Dokuments (=harmonisierter Datensatz) basiert, wird unmittelbar das relationale Datenbankmodell abgeleitet und die Geodaten in diesem abgelegt. Diese Art des Datentransfers wird auch als schemabasierter Datentransfer bezeichnet. Anderes als beim anderen Austauschformaten wird neben den Daten auch die entsprechende Modellsemantik der Daten übermittelt bzw. dem eigentlichen Datentransfer vorangestellt (Abbildung 1)

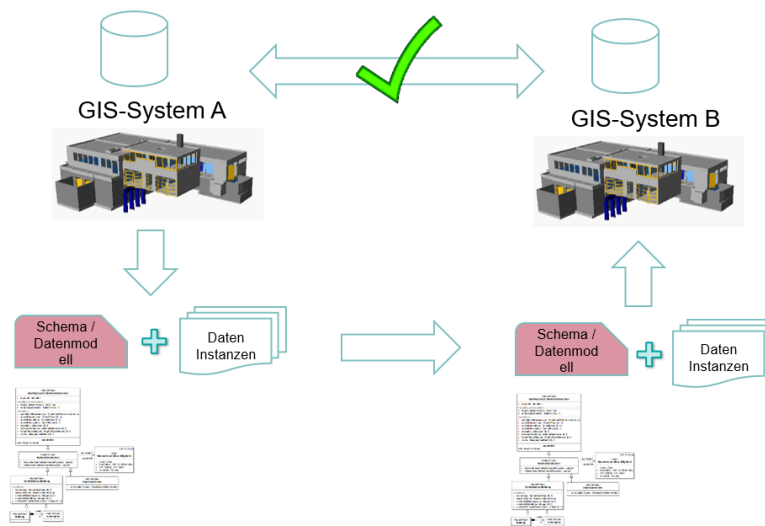


ABBILDUNG 1 – Schemabezogener Datentransfer. Dem eigentlichen Datenaustausch wird das Applikationsschema vorangestellt, welches eine Übermittlung der gesamten Modellsemantik des Systems A und dessen exakte Reproduktion in System B ermöglicht. Somit können komplexe Datenbankinhalte verlustfrei übermittelt werden.

GMLAS-Treiber

In diesem Kapitel wird nur auf jene Aspekte des GMLAS Treibers eingegangen, welche unmittelbar für die Nutzung der zur Verfügung gestellten Skripte erforderlich sind. Eine Vollständige Dokumentation aller Eigenschaften bzw. auch der Limitierung des Treibers ist unter: <https://gdal.org/drivers/vector/gmlas.html> verfügbar.

Konfigurationsdatei des Treibers - gmlasconf.xsd

Für die zur Verfügung gestellten Skripte wurden die Grundeinstellungen welche in der Datei gmlasconf.xsd definiert sind **NICHT GEÄNDERT**. Prinzipiell können anhand dieser Konfigurationsdatei folgende Einstellungen die das Verhalten des Treibers wesentlich beeinflussen vorgenommen werden:

- ob entfernte Schemata heruntergeladen werden sollen. Standardmäßig aktiviert.
- ob der lokale Cache von Schemata aktiviert ist. Standardmäßig aktiviert.
- Pfad des lokalen Cache. Standardmäßig: \$HOME/.gdal/gmlas_xsd_cache
- ob die Validierung des Dokuments anhand der Schemas aktiviert werden soll. In der Voreinstellung deaktiviert.
- ob ein Fehler bei der Validierung dazu führen soll, dass das Öffnen des Datensatzes fehlschlägt. Standardmäßig deaktiviert.
- ob die Metadaten Schichten standardmäßig angezeigt werden sollen. Standardmäßig deaktiviert.
- ob immer ein 'ogr_pkid'-Feld erzeugt werden soll. Standardmäßig deaktiviert. Das Einschalten dieser Option kann bei Ebenen nützlich sein, die ein ID-Attribut haben, dessen Eindeutigkeit in verschiedenen Dokumenten nicht gewährleistet ist. Dies kann zu Problemen führen, wenn mehrere Dokumente in eine Zieldatenbanktabelle eingefügt werden.
- ob Ebenen und Felder, die im XML-Dokument nicht verwendet werden, entfernt werden sollen. Standardmäßig deaktiviert.
- ob OGR-Array-Datentypen verwendet werden können. Standardmäßig aktiviert.
- ob die XML-Definition der GML-Geometrie als OGR-Stringfeld gemeldet werden soll. Standardmäßig deaktiviert.
- ob nur XML-Elemente, die von gml:_Feature oder gml:AbstractFeature abgeleitet sind, im ersten Durchgang der Schemaerstellung berücksichtigt werden sollen, wenn mindestens ein Element in den Schemata von ihnen abgeleitet ist. Standardmäßig aktiviert.
- mehrere Regeln um zu konfigurieren ob und wie xlink:href aufgelöst werden soll.
- eine Definition von XPath's von Elementen und Attributen, die ignoriert werden müssen, um die Anzahl der OGR-Ebenen und -Felder zu verringern.

Folgende Parameter wurden im Skript verwendet:

EXPOSE_METADATA_LAYERS=YES/NO: ob die Metadatenlayer "_ogr_fields_metadata", "_ogr_layers_metadata", "_ogr_layer_relationships" und "ogr_other_metadata" standardmäßig eingetragen werden sollen. Die Voreinstellung ist NO.

VALIDATE=YES/NO: ob das Dokument gegen die Schemata validiert werden soll. Die Validierung erfolgt beim Öffnen des Datasets. Voreinstellung ist NO.

SWAP_COORDINATES=AUTO/YES/NO: Ob die Reihenfolge der x/y- oder long/lat-Koordinaten vertauscht werden soll. Im AUTO-Modus bestimmt der Treiber anhand des srsName, ob eine Vertauschung vorgenommen werden muss. Wenn der srsName urn:ogc:def:crs:EPSG::XXXX lautet und die Reihenfolge der Koordinaten in der EPSG-Datenbank für diese SRS lat,long oder northing,easting ist, dann vertauscht der Treiber sie in die GIS-freundliche Reihenfolge (long,lat oder easting,northing). Bei anderen Formen von SRS (z. B. EPSG:XXXX) wird die GIS-freundliche Reihenfolge angenommen und daher keine Vertauschung vorgenommen. Wenn SWAP_COORDINATES auf YES gesetzt ist, werden die Koordinaten immer in der Reihenfolge vertauscht, in der sie im GML erscheinen. Wenn es auf NO gesetzt ist, bleiben sie in der gleichen Reihenfolge. Die Voreinstellung ist AUTO.

REMOVE_UNUSED_LAYERS=YES/NO: Legt fest, ob nicht verwendete Feature Types / Datentypen usw. bei der Erstellung des Datenbankschemas entfernt werden sollen. Die Voreinstellung ist NO.

REMOVE_UNUSED_FIELDS=YES/NO: Legt fest ob unbenutzte Felder aus dem Datenbankschema entfernt werden sollen. Standardwert: NO

Skript für die Erstellung von GPKG aus GML

Anhand des folgenden Skriptes kann unmittelbar aus einer beliebigen GML Datei ein Geopackage erstellt werden

Das vorliegende Skript ermöglicht die Erstellung eines GPKG, welches alle Inhalte eines harmonisierten INSPIRE Datensatzes entsprechend der INSPIRE Datenspezifikation in einem relationalen Datenmodell ablegt. Um die Nutzerfreundlichkeit des GPKG zu erhöhen, wurde des Weiteren spezifische Views erstellt, welche die unmittelbare Nutzung der unterschiedlichen Informationen des Datensatzes die oft in unterschiedlichen Tabellen gespeichert sind ermöglicht.

Zum Ausführen des Skriptes muss GDAL am Rechner installiert sein. Dieses kann anhand des OSGeoW4 – Installer vorgenommen werden.

Da das Skript auch themenspezifische Views erstellt, muss für die jeweilige INSPIRE Datenspezifikation das entsprechende Skript heruntergeladen werden. Zurzeit werden Skripte für die vier INSPIRE Datenspezifikationen

- Habitats and Biotops:
- Existing and Use
- Planned Land Use
- Protected Sites

themenspezifische Skripte angeboten

Ein weiteres Skript welches keine themenspezifischen Views sowie auch keine QGIS Konfigurationsdatei anbietet und für alle INSPIRE Themen anwendbar ist auf der österreichischen INSPIRE Assistenzseite verfügbar.

Die ZIP Datei zu den jeweiligen Skripten weisen die folgende Ordern Struktur und Inhalte auf:

/Input: Ordner in welchen die GML Datei aus der das GPKG abgeleitet werden soll gespeichert werden muss

/Skript: Im Ordern /Skript befindet sich das eigentliche Skript welches bevor dieses ausgeführt werden kann noch adaptiert werden muss

/Output: Verzeichnis in dem das Fertiggestellte GPKG gespeichert wird

QGIS-Konfigurationsdatei (Hauptverzeichnis) : Diese ist nur für die themenspezifischen Skripte vorhanden. In dieser Konfigurationsdatei sind

- die Beziehungen der Datenbanktabellen (in den Projekteigenschaften) definiert
- Spezifische Formulare die eine Anzeige der 1:n Relationen ermöglichen
- Joins von Tabellen auf die Tabelle des Feature Types des harmonisierten Datensatzes
- Styles für die Darstellung der Views und des Feature Types des harmonisierten Datensatzes

Workflow für die Erstellung des GPKG | Ausführen des Skriptes

In weiterer Folge wird der Workflow für die Erstellung eines GPKG dargestellt

- 1.) Download der GML Datei aus einem der INSPIRE Datenportale.
- 2.) Download des Files welches die Skript Datei beinhaltet → Entpacken der Zip-Datei
- 3.) Öffnen des Hauptordners der Zip-Datei
- 4.) Kopieren der GML Datei für welche das GPKG erzeugt werde soll in den Unterordner **/Input**
- 5.) Öffnen des eigentlichen Skriptes in einem Texteditor (z.B. Notepad ++) und Adaptierung der folgenden Parameter

ogr2ogr

-lco ASPATIAL_VARIANT=GPKG_ATTRIBUTES -f GPKG

Keine Adaptierung notwendig. Definition des Fileformats GPKG und das nicht räumliche Tabellen angelegt werden dürfen

-dsco VERSION=1.2

Keine Adaptierung notwendig. Definition der verwendeten GPKG Version

-oo SWAP_COORDINATES=**YES/NO**

Eventuell SWAP_COORDINATES auf YES setzen falls der Datensatz nicht lagerichtig angezeigt wird.
Default Value ist AUTO. Im AUTO-Modus bestimmt der Treiber anhand des srsName *urn:ogc:def:srs:EPSG::XXXX* lautet und die Reihenfolge der Koordinaten in der EPSG-Datenbank für diese SRS lat,long oder northing,easting ist, dann tauscht der Treiber sie in die GIS-freundliche Reihenfolge (long,lat oder easting,northing) aus. Bei anderen Formen von SRS (z. B. EPSG:XXXX) wird die GIS-freundliche Reihenfolge vorausgesetzt und daher keine Vertauschung vorgenommen.

-oo EXPOSE_METADATA_LAYERS=**YES/NO**

Falls mehrere GML Files importiert werden muss dieser Parameter für den letzten Import auf **YES**, für alle anderen auf **NO** gesetzt werden.

	Default Value YES findet Anwendung wenn der Parameter nicht angegeben wird.
-oo REMOVE_UNUSED_FIELDS= YES /NO	YES wenn Tabellenspalten die keine Einträge im GML aufweisen auch nicht in den Tabellen der Datenbank angelegt werden sollen. NO – es werden alle Tabellenspalten die im Applikationsschema definiert sind in der Datenbank angelegt. Dadurch entstehen große und zum Teil sehr unübersichtliche Tabellen mit zahlreichen Spalten die den Wert NULL aufweisen.
-oo REMOVE_UNUSED_LAYERS= YES	YES wenn für alle Feature Types die nicht in der GML Datei verwendet wurden auch keine Tabellen im Datenbankschema erstellt werden sollen.
-s_srs EPSG:3045	Definition (bzw. Überschreiben) des räumlichen Referenzsystems des Ausgangsdatensatzes (in dem Fall EPSG: 3045). Notwendig falls dieses beim Einlesen der GML Datei nicht erkannt wird
-t_srs EPSG:3035	Umprojektion auf dieses SRS und Definition des räumlichen Referenzsystems des Zielsystems (in dem Fall EPSG: 3035)
- -a_srs EPSG:3035	Nur notwendig falls das SRS im GPKG nicht erkannt wird. Im Unterschied zu -t_srs wird der Datensatz nicht umprojiziert
-nlt CONVERT_TO_LINEAR	Keine Adaptierung notwendig. Komplexe Geometrien (Kurven, Kreise...) werden in lineare Geometrien approximiert.
-forceNullable	Keine Adaptierung notwendig. Not NULL Constraints der Quellebene werden nicht auf die Zielebene übertragen.
../Output/ biotopkartierungen.gpkg	Filename des GPKG das erstellt wird
GMLAS:../Input/ filename.gml	Filename der GML-Datei die in GPKG umgewandelt werden soll.
-APPEND /OVERWRITE	Default wird ein CREATE ausgeführt (=Erstellung des DB Schema und einfügen der Daten). Beim Import einer einzelnen Datei muss somit kein der Parameter nicht gesetzt werden. Als weitere Option steht -OVERWRITE zur Verfügung. Durch den Parameter wird die existierende Datenbank mit allen Inhalten überschrieben. Die bereits in der Datenbank eingefügten Daten gehen verloren! APPEND: Bei Verwendung des Parameters -APPEND werden die Datensätze in der bestehenden Datenbankstruktur angehängt. Für die Erzeugung eines Datensatzes bei den mehrere GML-Files in ein GPKG geschrieben werden muss dieser Parameter gesetzt werden. ...

6.) Abspeichern der Skript Datei

7.) Ausführen der Skript Datei

7.1.) Öffnen der OSGeo4W Shell (Abbildung 2)



ABBILDUNG 2 – Öffnen der OSGeo4W Shell zum Ausführen des Skriptes zur Erstellung des Geopackage.

7.2.) Navigation in den Ordner in dem das Skript abgespeichert ist

7.3.) Ausführen des Skriptes – Aufrufen der *.bat Datei in der OSGeo4W Shell

7.4.) Nach Durchlauf des Skriptes kann das GPKG mittels dem QGIS Projekt im Hauptordner direkt geöffnet werden (Abbildung 3)

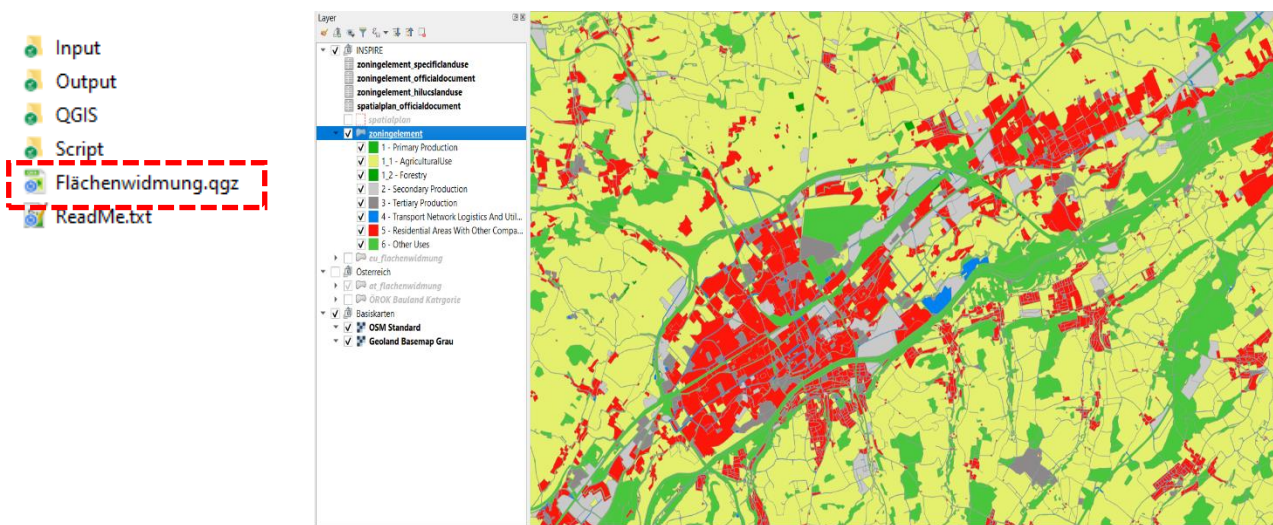


ABBILDUNG 3: Öffnen des erstellten GPKG mit den im QGIS-Projekt konfigurierten Datenschichten, Formularen und Verknüpfungen