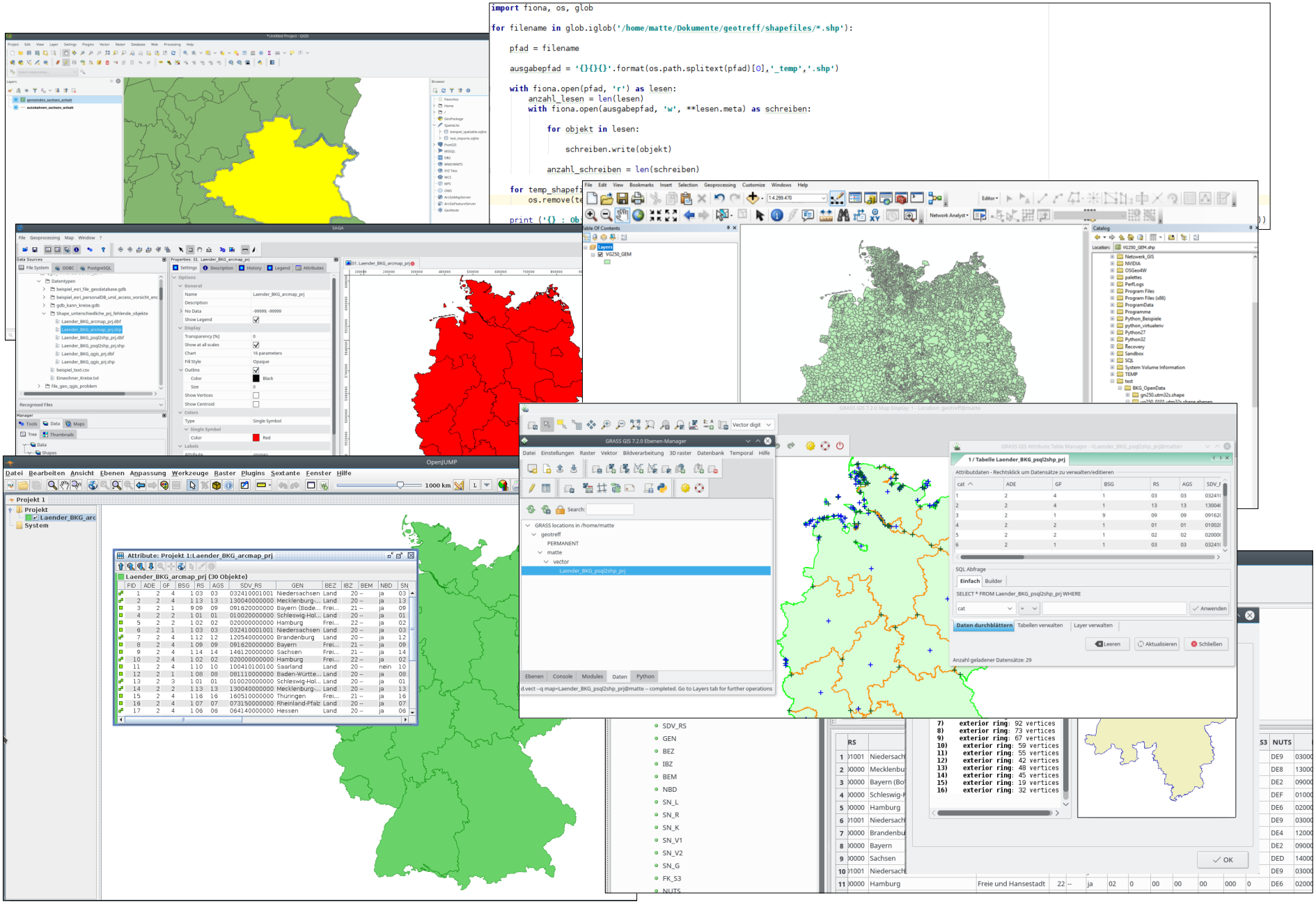
Herausforderungen beim Wechsel zwischen Datenformaten und Programmen



Herausforderungen beim Wechsel zwischen Datenformaten und Programmen

(1) Die Ausgangssituation

(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

(3) Wechsel des Koordinatenbezugssystems

(4) Unterschiedliche Arbeitsabläufe und Geoverarbeitungswerkzeuge

(5) Hinweise zu gemischten Umgebungen

(6) Fazit

(1) Die Ausgangssituation

Es existieren die unterschiedlichsten GIS, CAD-Systeme, Bibliotheken (und deren Implementierung) sowie eine Vielzahl an Datei- und Datenformaten

Unterschiedlichste Gründe können zu einem Wechsel oder zur Nutzung von gemischten System führen → der Workshop ist so neutral wie möglich hinsichtlich der Softwaresysteme gehalten, eine Bewertung ist nicht beabsichtigt

(1) Die Ausgangssituation

Es existieren die unterschiedlichsten GIS, CAD-Systeme, Bibliotheken (und deren Implementierung) sowie eine Vielzahl an Datei- und Datenformaten

Umfrage

- CAD in Verwendung?
- Kommerzielle GIS in Verwendung?
- Außer ESRI?
- Opensource außer QGIS?
- kein Windows?
- zentrale Datenbank?
- häufige Nutzung eines Datenformates <u>außer</u> Shapefile?

Das Shapefile – Der Quasi-Standard – ein geeignetes Austauschformat?

wichtige Vorteile

Nahezu jede Software kann Shapefiles importieren oder exportieren

Offen dokumentierter Standard

Vergleichsweise unkomplizierte Geometriespeicherung

Das Shapefile – Der Quasi-Standard – ein geeignetes Austauschformat?

wichtige Nachteile

Eingeschränkte Attributdatenspeicherung (10 Zeichen Attributname, Texte mit max. 254 Zeichen, kein echtes NULL)

Nur Geometrie und Sachdaten sind einheitlich standardisiert. Raumbezug, Indizes, Kodierung, Metadaten sind je nach Software verschieden

Eingeschränkte Geometrien (keine Kreisbögen und Kurven), keine Topologie und ungenaue Koordinatenspeicherung (Verarbeitung Nachkommastellen je Software verschieden)

keine Speicherung von komplexen Datenmodellen möglich

(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

Das Shapefile – Der Quasi-Standard – ein geeignetes Austauschformat?

- Attribut-Namen sind auf 10 Zeichen beschränkt
- Texte werden nach 254 Zeichen abgeschnitten (kein Hinweis der Software)
- Datenmodell wird nicht gespeichert (Semantik, Aktualität, Alias etc.)

bev_beispiel :: Objekte gesamt:4, gefiltert: 4, gewählt: 0

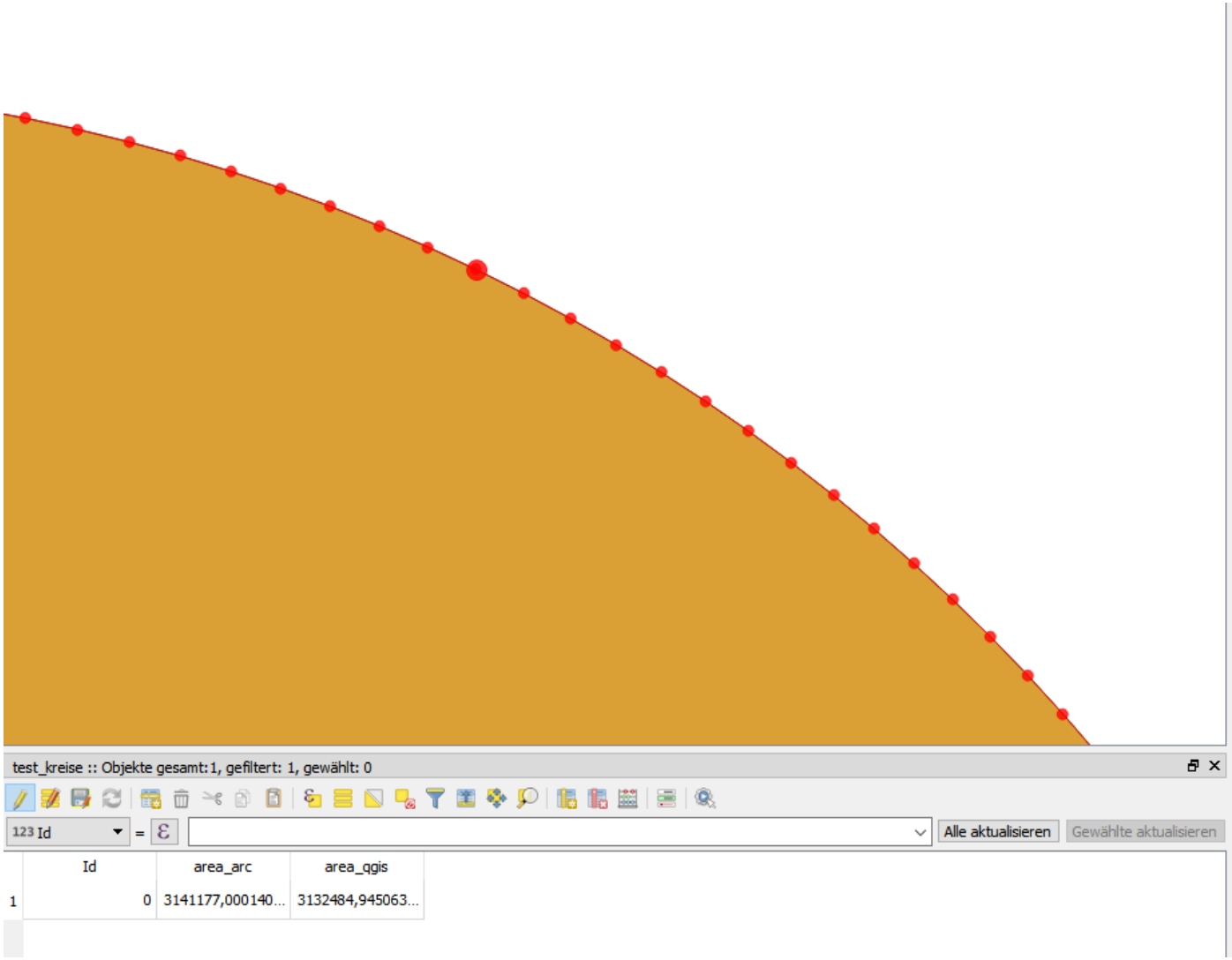
	fid	Einwohner_Maenner	Einwohner_Frauen	Beschreibung	len_vorher
1	4	16579	11830	creating shapefil... formats. This doc...	509
2	2	5356	1223		NULL
3	3	16410	26272		NULL
4	1	18022	17223		NULL

textlaengen_beispiel :: Objekte gesamt:4, gefiltert: 4, gewählt: 0

	fid	Einwohner_	Einwohne_1	Beschreibu	len_vorher	len_nach
1	3	16410	26272			
2	4	16579	11830	describes why sh... Environmental Sy...	509	254
3	1	18022	17223			
4	2	5356	1223			

(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

Das Shapefile – Der Quasi-Standard – ein geeignetes Austauschformat?



Kreisbögen und Kurven werden generalisiert (je nach Software),
Flächenberechnungen weichen stark voneinander ab (ca. 3% bei 1000m Radius)
auch Abweichungen bei Kreisflächen innerhalb einer Software je Format

(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

Das Shapefile – Der Quasi-Standard – ein geeignetes Austauschformat?

Räumlicher Index ist nicht offen dokumentiert → Z.B. Dateien aus ArcMap werden in anderer Software unvollständig angezeigt, je nach Zoomstufe

(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

Das Shapefile – Der Quasi-Standard – ein geeignetes Austauschformat?

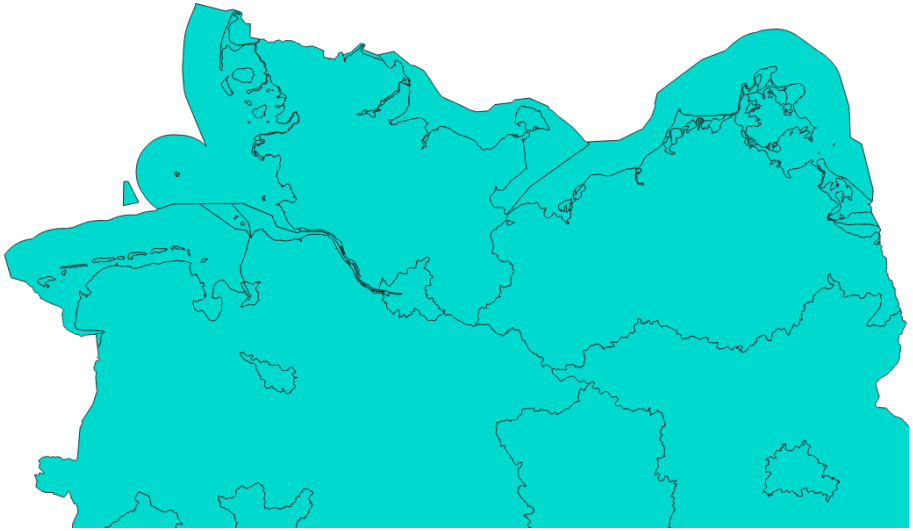
PROJCS["ETRS_1989_UTM_Zone_32N",GEOGCS["GCS_ETRS_1989",DATUM["D_ETRS_1989",SPHEROID...

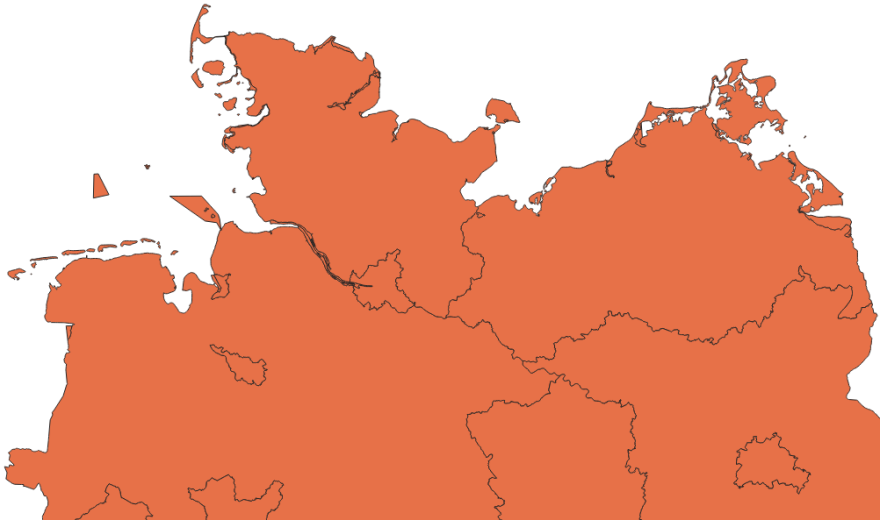
PROJCS["ETRS89 / UTM zone 32N",GEOGCS["ETRS89",DATUM["European_Terrestrial_Reference_System_1989",SPHEROID...

Projektionsdatei nicht standardisiert. Unterschiedliche Projektionsdatei wird zwar inhaltlich korrekt erkannt aber verwirrt Nutzer durch falsche Bezeichnungen und verhindert korrekte Transformation

(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

Das Shapefile – Der Quasi-Standard – ein geeignetes Austauschformat?





<u>Seltene</u> Fehler beim Laden der Geometrien -> einzelne Flächen verschwinden in vielen Implementierungen und werden beim Speichern unwiederbringlich gelöscht

(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

Das Shapefile – Der Quasi-Standard – ein geeignetes Austauschformat?

Hinweise für die Nutzung wenn keine Alternative

- nur standardisierte Teile übergeben (.shp, .dxf, .shx)
- Raumbezug, Kodierung, Datenmodell, Metadaten separat dokumentieren, explizit darauf hinweisen und im Zielsystem neu setzen
- Indizes und Bounding-Box im Zielsystem neu erstellen
- Test der Textlängen, Topologien etc.
- Sachdaten im Zweifelfall abtrennen und über Extra-Datei (kein Spreadsheet wie EXCEL) und gemeinsamen Schlüssel übergeben

(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

Alternative Austausch- und Arbeitsformate

Comma Separated Values - .csv Dateien

- in praktisch jedem System verfügbar
- Zusatzinformationen müssen extern weitergeben werden (Raumbezug, Semantik, etc.)
- keine eigene Geometrie-Logik, Speicherung als Koordinaten oder Well-Known-Text
- Austausch zwischen Betriebssystemen kann zu nicht erkannten Zeilenumbrüchen führen (je nach verwendeten Programm)
- Speicherung als Unicode-Text mit „Byte Order Mark“ sinnvoll
- unkomprimiert große Dateien (als Archiv in Ordnung)

```
ID, WKT
1, "MULTIPOLYGON (((479853.067298559
5969913.68631609,479931.250047623
5969903.27387711,480008.926912039 usw.
```

(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

Alternative Austausch- und Arbeitsformate

Markup Language und GeoJSON – XML/GML/KML/Json Dateien

- wenn als Im- und Export verfügbar gut geeignetes Transportformat
- Text basiert, daher unkomprimiert große Dateien
- transportiert Geometrien, Logik und Zusatzfunktionen
- nicht als Arbeitsformat geeignet
- Nachkommastellen der Koordinaten können variieren

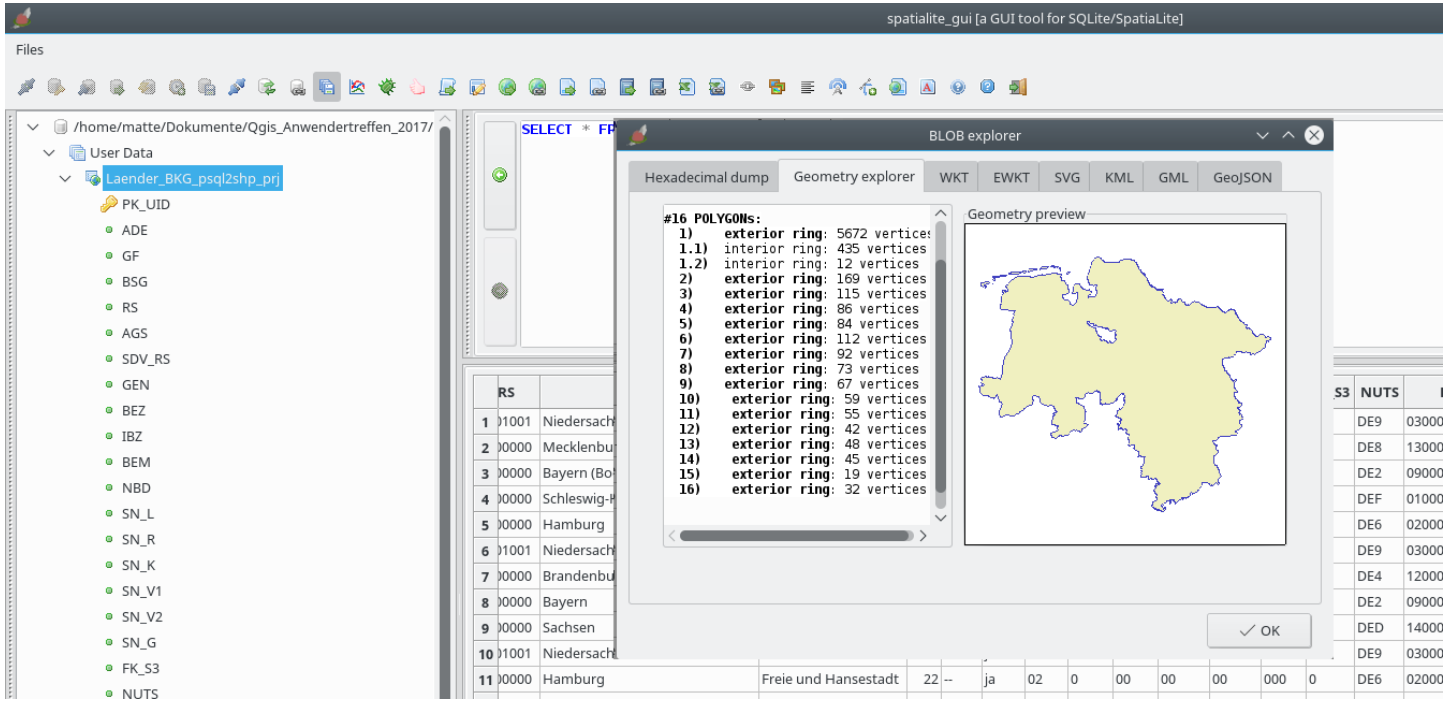

```
<?xml version="1.0" encoding="utf-8" ?>
<ogr:FeatureCollection
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://ogr.maptools.org/ gml_export.xsd"
    xmlns:ogr="http://ogr.maptools.org/"
    xmlns:gml="http://www.opengis.net/gml">
  <gml:boundedBy>
    <gml:Box>
      <gml:coord><gml:X>280353.0870717983</gml:X><gml:Y>5235877.555857686</gml:Y></gml:coord>
      <gml:coord><gml:X>921261.5513513577</gml:X><gml:Y>6106243.849926789</gml:Y></gml:coord>
    </gml:Box>
  </gml:boundedBy>

  <gml:featureMember>
    <ogr:gml_export fid="gml_export.0">
      <ogr:geometryProperty><gml:MultiPolygon
srsName="EPSG:25832"><gml:polygonMember><gml:Polygon><gml:outerBoundaryIs><gml:LinearRing><gml:coordinates
>479853.067298559,5969913.68631609 479931.250047623,5969903.27387711 480008.926912039,5969917.84223009
480084.160098617,5969954.35320206 usw.
```

(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

Alternative Austausch- und Arbeitsformate

Spatialite und Geopackage

- wenn verfügbar sehr gut geeignete Transport- und Arbeitsformate
- kann Geometrien, Logik und Zusatzfunktionen transportieren
- auf Datei-basierter relationaler Sqlite-Datenbank aufbauend
- Vorteile RDBMS (ACID, SQL, Views, Beziehungen, Trigger etc.)
- Geopackage ist OGC-Standard (hoffentlich zunehmende Verbreitung bei kommerziellen Systemen, aktuelle ArcMap-Versionen gute Umsetzung)
- Spatialite nutzt eigene Treiber und bietet weitere Funktionen



(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

Alternative Austausch- und Arbeitsformate

Personal- / Filegeodatabase

- gute Datenformate (nicht nur) für den Transport
- sehr viel (Geometrie)-Logik beim Import (Geometrie säubern, Grid basierte Koordinaten, Attribute/Datentypen etc.)
- sehr umfangreiche Zusatzfunktionen
- proprietär, aber Treiber für einfaches Lesen und Schreiben sind freigegeben, daher auch in Opensourcesoftware geeignetes Austauschformat
- In Linux/Macos teils problematisch (Treiberintegration Gdal je nach Distribution)
- offener reverse engineered Treiber fürs lesen verfügbar (funktioniert recht zuverlässig, allerdings nicht mit Kreisen und Kurven, diese gehen ohne Hinweis verloren) → in Umgebungen sinnvoll ohne prop. Treiber
- Personalgeodatabase in Verbindung mit Opensource 64-bit Anwendung problematisch

(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

Alternative Austausch- und Arbeitsformate

Drawing Interchange Format - DXF

- Standard CAD-Austauschformat
- Datenspeicherung / Export in CAD oft nach GIS untypischer Philosophie
- stark Software- und Versionsabhängig
- ganzer Zoo an „Konvertierungstools“
- etwas geht meist verloren, im Besonderen bei komplexen Zeichnungen
- nach Möglichkeit anderes Exportformat nutzen, viele CAD Systeme setzen auf interne relationale Datenbanken und bessere Austauschformate → der CAD eigene Export in ein „Geodatenformat“ ist dann oft besser als der GIS-Import einer .dxf

(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

Alternative Austausch- und Arbeitsformate

Spreadsheet Formate - .xls(x) , .ods

- Excel und Co sind <u>keine</u> Datenbanken und haben einen anderen Zweck
- Programme interpretieren zu viel (was teilweise nicht abschaltbar ist)
- Datenverlust kann beim Export durch den Treiber schon geschehen
- Import geschieht oftmals durch Interpretation der Datenformate
- vermeiden wenn möglich...

One in Five Genetics Papers Have Errors Caused by Microsoft Excel

Spreadsheet programs like Excel are creating issues when they reformat scientific data.

Artikel von Avery Thompson von www.popularmechanics.com basierend auf der Studie:
Ziemann, Mark, Yotam Eren, und Assam El-Osta. 2016.
„Gene Name Errors Are Widespread in the Scientific Literature“.
Genome Biology 17 (1). https://doi.org/10.1186/s13059-016-1044-7.

(2) Auf der Suche nach dem geeigneten Austausch- und Arbeitsformat

Kandidaten für ein geeignetes Austauschformat – Diskussion

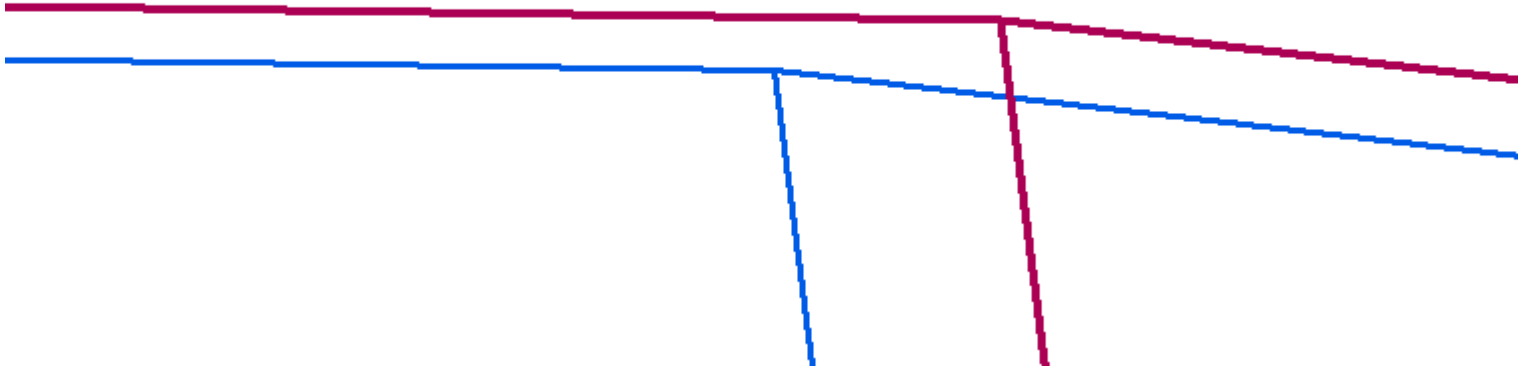
Geopackage	OGC-Standard, Dateibasierte Datenbank mit allen Vorteilen von DB, sehr gutes Geodatenformat, auch Darstellung und Raster integrierbar	
Spatialite	Ähnlich Geopackage, jedoch kein Standard, zusätzlicher Funktionsumfang zur Geoverarbeitung, Routing etc.	
File-Geodatabase	Proprietäres ESRI-Format, sehr gutes Geodatenformat mit sehr vielen Funktionen, Treiber für lesen uns schreiben wird von ESRI zur Verfügung gestellt	
Personal-Geodatabase	ESRI-Format auf der Basis von MS ACCESS, gutes Geodatenformat, viele Funktionen, außerhalb der ESRI-Welt nur Geodatenspeicherung möglich	
GML / KML / GJSON	Textbasierte Formate, dokumentiert und standardisiert, sehr gutes Transportformat, aber teils Versions- und Implementierungschaos	
CSV	Texte; alle Logik und Geometriespeicherung muss selbst eingerichtet und implementiert werden (Geometrie als Koordinaten und Well-Known-Text) – überall verfügbar	
DWG / DXF	Oft einzige Möglichkeit des Transports aus CAD-Systemen heraus, stark Software- und Versionsabhängig, oft Verluste	

(3) Wechsel des Koordinatenbezugssystems

Potsdam-Datum und Pulkovo-Datum nach ETRS89

NTV2/Beta2007 Ansatz des BKG für Potsdam-Datum unproblematisch, keine einfache Transformation von Pulkovo-Datum

- Umwandlung Potsdam-Datum mit hoher Genauigkeit möglich
- einfache Pulkovo-Datum-Transformation mit 3-4m Abweichung in Sachsen-Anhalt
- Transformationsdienste des LVermGeo verfügbar
- LSA-Trans bietet als freies Programm die Möglichkeit der genaueren Umwandlung (Linien und Flächen müssen allerdings in Stützpunkte zerlegt, umgewandelt, und korrekt wieder zusammengesetzt werden) <u>Aber: momentan keine Weiterentwicklung, keine 64-bit-Version</u>



(4) Unterschiedliche Arbeitsabläufe und Geoverarbeitungswerkzeuge

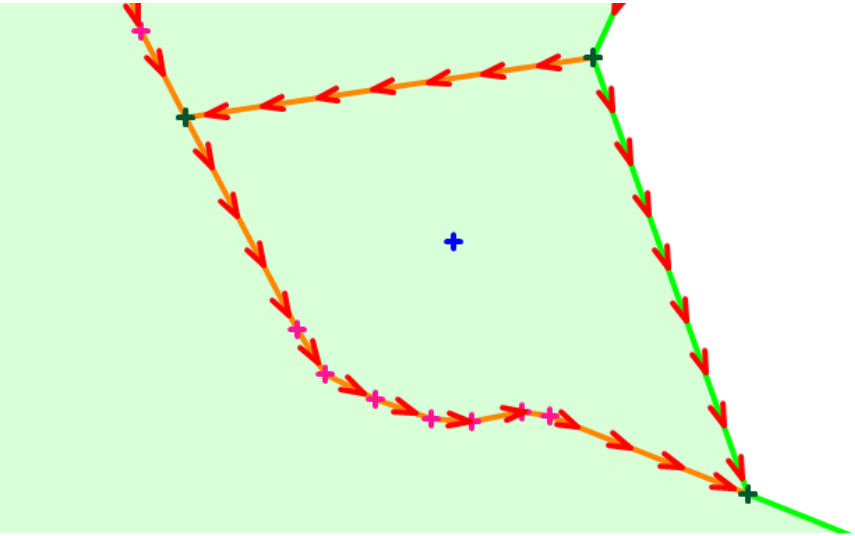
Unterschiedliche Ergebnisse und Arbeitsschritte je nach Software

Jede Software verfolgt andere Philosophie

- z.B. behebt ArcMap direkt beim Digitalisieren Geometriefehler ohne den Nutzer darauf hinzuweisen und säubert und reprojiziert vor jeder Geoverarbeitung
- viele Programme gehen mit ungültigen Geometrien anders um → von Fehlermeldung mit Abbruch über Hinweis das einzelne übergangen werden bis hin zu unerkannt fehlerhaften Ergebnissen
- Kreise und Kurvenzüge meist problematisch beim Wechsel
- echte topologische Bearbeitung

  erlauben nur wenige

  Programme und Formate

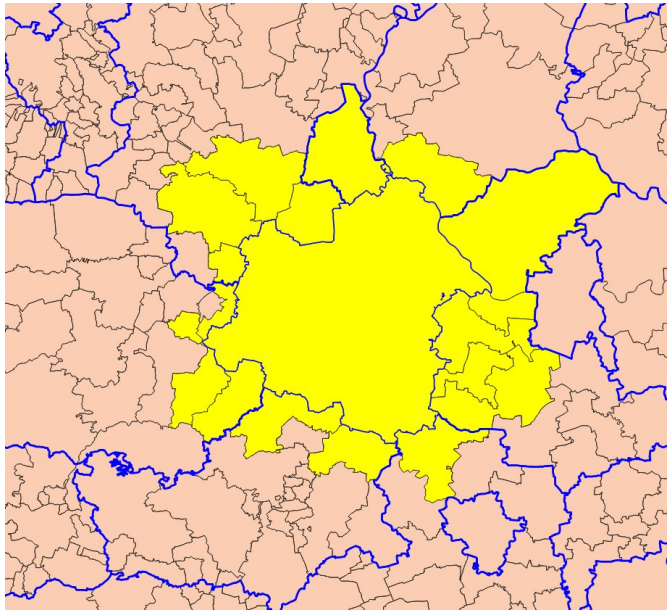


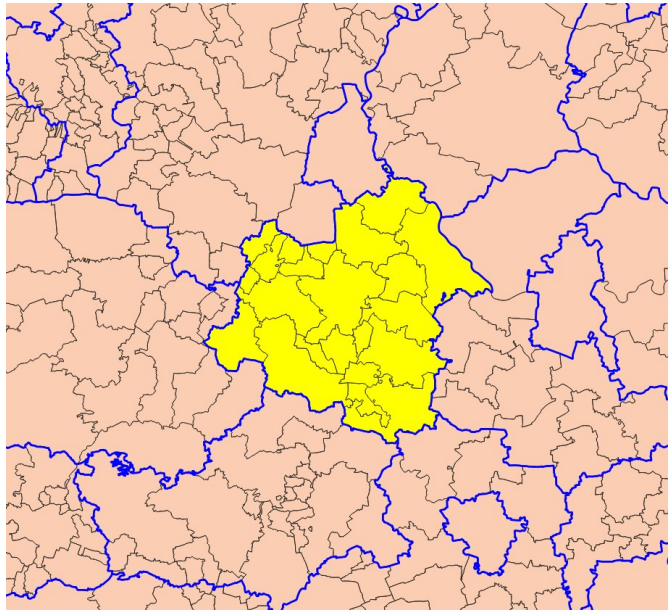
(4) Unterschiedliche Arbeitsabläufe und Geoverarbeitungswerkzeuge

Unterschiedliche Ergebnisse und Arbeitsschritte je nach Software

Werkzeuge liefern andere Ergebnisse je nach Software

- z.B. bei Puffern unterschiedliche Generalisierung von Kreisen und Linienabschlüssen
- Unterschiedliche Toleranzen bei räumlichen Verschneidungen und unterschiedliche Methoden





(4) Unterschiedliche Arbeitsabläufe und Geoverarbeitungswerkzeuge

Unterschiedliche Ergebnisse und Arbeitsschritte je nach Software

Werkzeuge liefern andere Ergebnisse je nach Software

- die häufig verwendeten Prüftools liefern unterschiedliche Ergebnisse (neben fehlerhafter Berechnungen und unterschiedlichen Toleranzen auch unterschiedliche Fehlerdefinitionen)
- Beispiel von der Fossgis 2018 empfehlenswert (https://media.ccc.de/v/2018-5330-geometrie-_und_topologiefehler_finden_und_korrigieren)
- z.B. „Berühren sich in einem Punkt“ ist im SimpleFeature-Modell korrekt, wird aber von manchen Prüftools als „Selbstüberschneidung“ erkannt

(5) Hinweise zu gemischten Umgebungen

Strategien beim parallel Betrieb verschiedener Software

Herausforderung

- im Gegensatz zu einmaliger Migration ständiger Wechsel zwischen Softwaresystemen und eventuell Formaten
- Arbeitsschritte müssen gleichwertige Ergebnisse liefern
- es darf kein Informationsverlust auftreten
- Schulungs- und Administrationsaufwand minimieren
- unterschiedliche, teils wechselnde, Lizenzmodelle je Software

(5) Hinweise zu gemischten Umgebungen

Strategien beim parallel Betrieb verschiedener Software

Vorüberlegungen

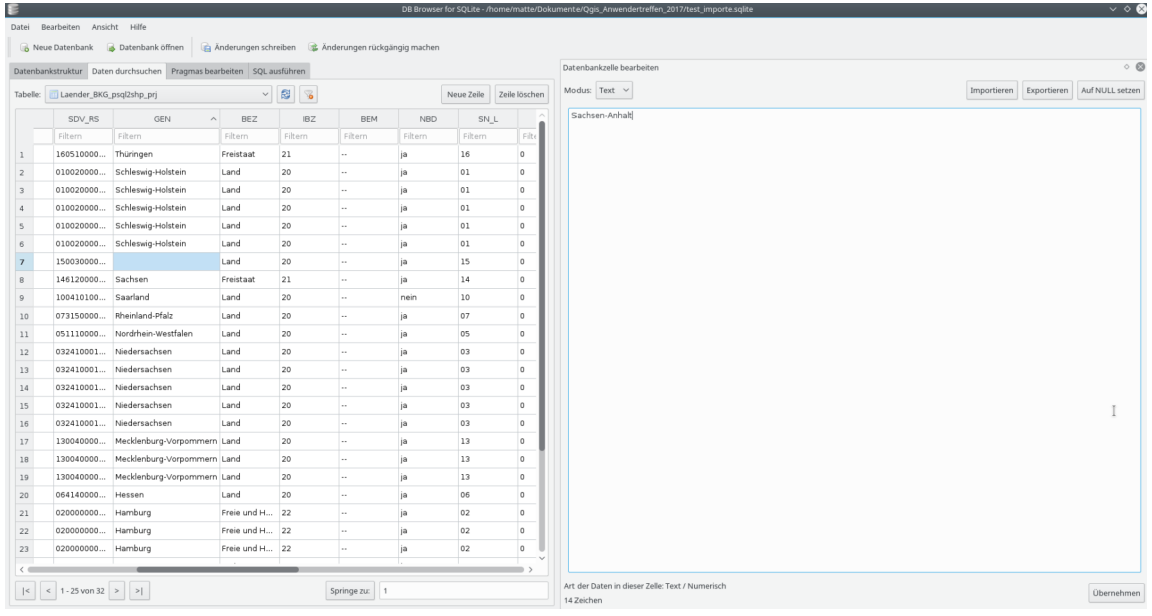
- wann immer möglich zentrale Datenbank verwenden → häufig mit Zusatzkosten verbunden (z.B. ArcGIS Enterprise + ArcMap Standard Lizenz zum Bearbeiten notwendig) oder nicht in jeder Software gleichwertig verfügbar
- prüfen welche Funktionen benötigt werden, oft werden keine Geometrien bearbeitet sondern nur Sachdaten ergänzt → Trennung Geodaten und Sachdaten (Verknüpfungs-ID)
- Opensource-Programme eine Option?!
- Größe des Nutzerkreises → bei wenigen Nutzern Dateien auf Netzlaufwerken oft der pragmatische Weg
- Browserbasierte Dateneingabe? Generalanbieter (ArcGIS Online, Boundless, Mapbox etc.)
- spielen CAD-Systeme eine Rolle?

(5) Hinweise zu gemischten Umgebungen

Strategien beim parallel Betrieb verschiedener Software

Trennung von Geometrie und Sachdaten, kann zentrales Netzlaufwerk pragmatische Lösung sein ?

- Geodaten enthalten nur einen Verknüpfungsschlüssel, Sachdaten werden in anderer Software bearbeitet
- wird häufig mit Excel oder Access genutzt → wann immer möglich Excel vermeiden (Formatierungschaos, Zellen/Zeilen können aus versehen verschoben werden etc.)
- wenn kein Access zur Verfügung über Opensource Varianten nachdenken wie LibreOffice-Base, DB Browser für Sqlite (siehe unten)



(5) Hinweise zu gemischten Umgebungen

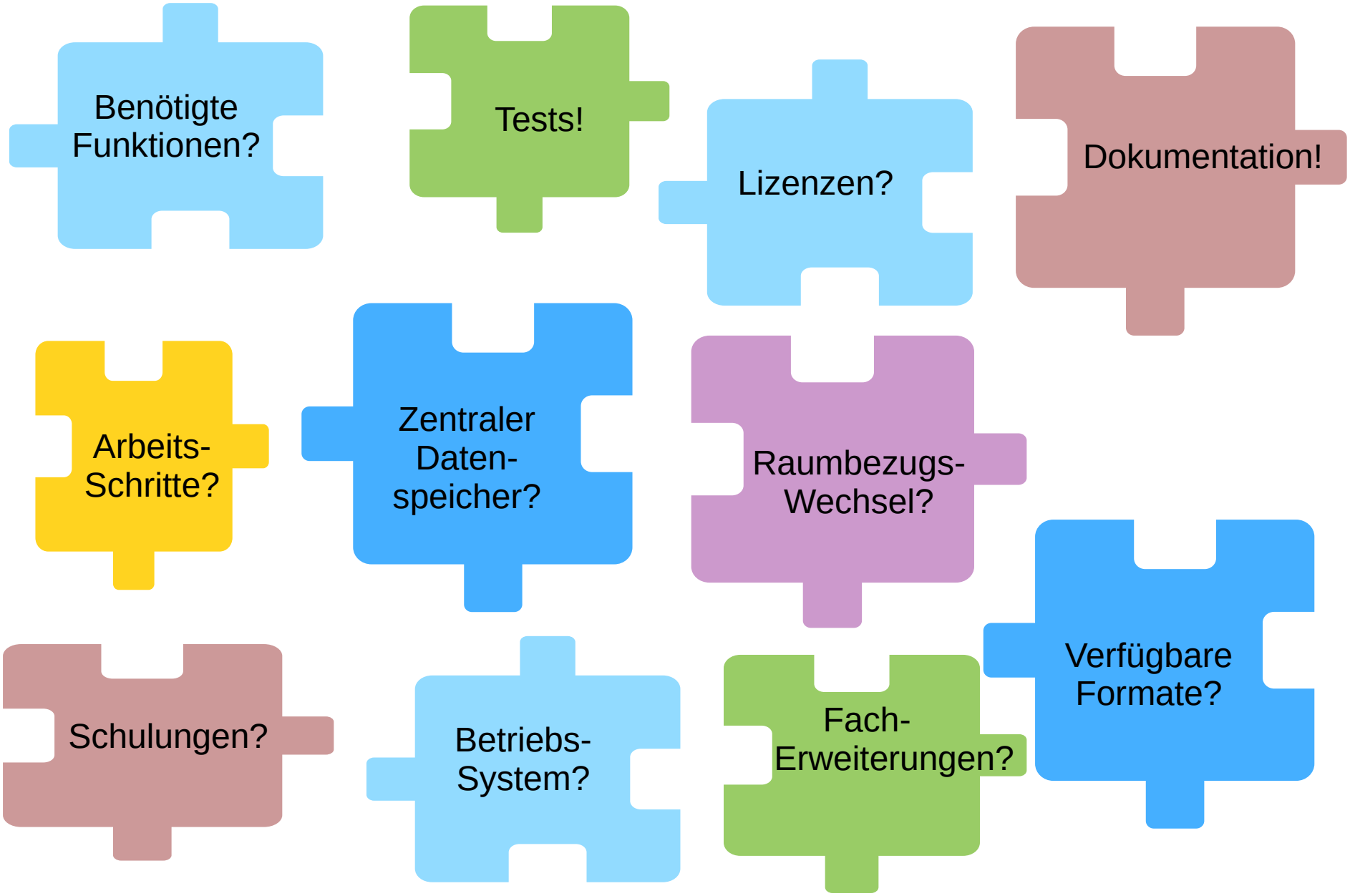
Strategien beim parallel Betrieb verschiedener Software - Diskussion

Hinweise

- Arbeitsabläufe je Software prüfen und definieren
- bestimmte Arbeitsschritte nur mit einer Anwendung / Arbeitsvorschrift umsetzen (Flächenberechnungen, Koordinatentransformationen etc.)
- keine Arche für Tools zulassen (im Besonderen Webseiten ala Koordinatenfinder, Transformationstools, Konverter etc.)
- z.B. Nutzung von GDAL/OGR, JTS/GEOS und deren Implementierungen oder FME
- Automatisierung von Prozessen mit den oben genannten Systemen über zentrale Scripte oder eigene getestete Tools (viele GIS mit Ablaufmodelierung)

(6) Fazit

Einzelne Prüfschritte beim Softwarewechsel - Diskussion



(6) Fazit

Einzelne Prüfschritte beim Softwarewechsel - Diskussion

Erfahrungen? Beispiele?

-

Vielen Dank für ihre Aufmerksamkeit
und Mitarbeit !