

Datu savietojamība :: Failu formāti

Kārlis Kalviškis

12.02.2021



Telpisko datu digitālā apstrāde

Datu apmaiņas un savietojamības ierobežojumi

- Juridiskie.
- Administratīvie.
- Subjektīvie

- **Failu formāti un versijas.**
- Datu struktūra.
- Datu formāts.
- Datu saturs (kvalitāte, piesaiste telpai un laikam utt.).
- Datu attēlošana.

Tehniskie standarti

Ja savām vajadzībām īslaicīgos projektos var naivi uzticēties, ka programma zinās labāk, kādā formātā datus glabāt, tad ilglaicīgai datu lietošanai un, vēl jo vairāk, datu apmaiņai ar citiem lietotājiem, ir jāapzinās dažādu formātu priekšrocības un trūkumi.

Lekcijā tiks apskatīti ierobežojumi, kuri saistīti ar datu struktūru, formātu un saturu. Šie ierobežojumi ir svarīgi ne tikai pie datu apmaiņas, bet arī, piemēram, lai būtu pieejami pirms desmit gadiem paša savāktie dati un veidotās kartes.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Failu formātu pieejamība

- Atvērtie datu formāti:
 - ir pieejams apraksts, kā dati tiek glabāti failā.
- Slēgtie datu formāti:
 - datu īpašnieks kļūst atkarīgs no programmas izstrādātāja;
 - pēc iespējas vajadzētu izvairīties no slēgtajiem datu formātiem;
 - ja tiek izmantoti dati slēgtajos datu formātos, jāveido rezerves kopijas atvērtos datu formātos, lai palielinātu datu lasāmības iespējamību arī nākotnē.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Failu formāti

- Teksta faili (ASCII, ANSI, UNICODE).
 - Iespējas apskatīt un labot ar parastu teksta redaktoru.
 - Piemēroti konfigurācijas failiem, tai skaitā arī projekta failiem.
 - Kā trūkums minami lieli apjomi, kā arī tas, ka šo failu apstrādē programmatūra patērē vairāk laika.
 - Teksta fails vēl nenozīmē, ka to viegli var lasīt/labot ar parastu teksta redaktoru.

Teksta faili nenozīmē, ka viņi ir viegli saprotami, piemēram, pat ļoti laba iztēle varētu nepalīdzēt saprast rastra attēlu, kas pierakstīts teksta failā.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Failu formāti

- Binārie faili.
 - Nepieciešama speciāla programmatūra failu apskatei un labošanai.
 - Parasti optimizēti ielasīšanai un labošanai.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Failu formāti

- Izmēra ierobežojumi:
 - formāta ierobežojumi (TIFF 4 Gb, BigTIFF teorētiski bez ierobežojuma);
 - failu sistēmas un OS ierobežojumi (FAT – 2 Gb, FAT32 – 4 Gb).



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Failu formāti

- Failu nosaukumu ierobežojumi:
 - nosaukuma garums;
 - lielie/mazie burti;
 - ne-ASCII simboli;
 - pilna ceļa garums (ceļš + faila vārds).



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Telpiskie datu failu formāti*

- Rastrattēlu un karšu failu formāti: tiff, geotiff, jpeg, jp2, png, gif, bmp, hmr, ecw, img, raw, mrsid, ...
- Vektorattēlu un karšu failu formāti: dxf, dwg, dgn, svg, eps, wmf, emf, pdf, swf, ...

* Paplašinājums pats par sevi nenosaka faila formātu! Piemēram, paplašinājums „DRW” var būt failam, kurš veidots ar „Micrografx Draw” vai „Micrografx Designer”, tikpat labi ar „Autocad”, „MacDraw”, „CADS Planner” u.c. – tie visi ir formāta ziņā pilnīgi atšķirīgi faili.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Atribūtdatu failu formāti*

- Atsevišķi faili: txt, tab, csv, xls, ods, mdb, dbf, sqlite, db, ...

* Paplašinājums pats par sevi nenosaka faila formātu!



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Apvienotie telpisko un atribūtdatu failu formāti

- *ESRI shape faili, ESRI coverage, ARC/INFO export fails, ESRI GeoDatabase, MapInfo (mid/mif un tab), Idrisi, GRASS, ILWIS, ...*
- *Vietējās datubāzes: GeoPackage, SpatiaLite (SQLite), MS Access.*

Datubāzēs glabājas gan vektor-, gan rastrkartes.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Objektorientēti formāti

- HDF – (*Hierarchical Data Format*), var saturēt rastra attēlu, paletes, anotācijas, daudzdimensionālus masīvus utt. Objektī var būt sagrupēti;
- SDTS – *the Spatial Data Transfer Standard (USGS)*, paredzēts telpisku datu apmaiņai starp atšķirīgām sistēmām.
- ...

USGS – U.S. Geological Survey.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Karšu projektu faili

- *.map – *Idrisi (Map Composition File)*.
- *.mxd – *ESRI ArcMap (ArcGIS Map Document)*.
- *.pmf – veido ar *ESRI ArcGIS Publisher*, lasa ar *ESRI ArcReader (ESRI Published Map File)*.
- *.qgs – *Quantum GIS (Quantum GIS Project)*.
- *.qgz – *Quantum GIS (Quantum GIS Project)*
- ...

Nesatur pašus telpiskos un atribūtdatus. Parasti nesavietojami starp dažādām versijām. Idrisi un ESRI veido binārā formā – ar „roku” nevar izlabot kļūdas. QGS ir XML formātā – iespējams labot un atgūt informāciju.



© Kārlis Kalviškis, 2020.

Teksta faili



Telpisko datu digitālā apstrāde

Teksta faila parametri

- Atdalītāji var būt:
 - tabulators, atstarpe;
 - semikols, komats, | ;
 - pāreja uz jaunu rindu;
 - vai cits simbols vai simboli.
- Fiksēta teksta vieta (atdalītāji nav vajadzīgi).
- *Extensible Markup Language (XML)* – paplašināmās iezīmēšanas valoda – saimes faili.
- *Well-Known Text (WKT)* – labi atpazīstams teksts.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Teksta faila parametri

- Skaitļu decimāldaļas atdalītāji:
 - punkts;
 - komats.
- Decimālskaitļi izteikti ar mantisu un kāpinātāju, piemēram, $3,8E-5$ (= 0,000038).



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Teksta faila parametri

- Pāreja uz jaunu rindu:
 - LF (Unix, Linux, FreeBSD, MacOS X)
 - CR (Mac OS)
 - CR + LF (DOS, Windows, Symbian OS)
 - UNICODE tabulā atrodami 7 simboli (ieskaitot LF un CR), kuriem būtu jāveido jauna rinda.

Universālā kodu sarakstā iekļauti vēl divi atdalītāji – līniju atdalītājs (U+2028; HTML:   LSEP) un rindkopu atdalītājs (U+2029; HTML:   PSEP). Šie simboli ir tikai teksta formatēšanai un tie nav komandsimboli.

Ne visi redaktori/sistēmas saprot visus simbolus.



© Kārlis Kalviškis, 2020.

Līniju beigu simbolu saraksts

- **LF:** pāreja uz jaunu rindu, *Line Feed*, U+000A, 10
- **VT:** stateniskā tabulācija, *Vertical Tab*, U+000B, 11
- **FF:** jauna lapa, *Form Feed*, U+000C, 12
- **CR:** atgriezt galviņu, *Carriage Return*, U+000D, 13
- **NEL:** nākām rinda, *Next Line*, U+0085, 133
- **LS:** līniju atdalītājs, *Line Separator*, U+2028, 8232
- **PS:** rindkopu atdalītājs, *Paragraph Separator*, U+2029, 8233

Ar simbolu nosaukumiem angļu valodā, unikoda apzīmējumu un koda kārtas numuru decimālā skaitīšanas sistēmā.

Telpisko datu digitālā apstrāde

Nesapraستی rindu nobeigumi

```

Izmanto_LS_teksts.txt x
1 <p>The Conference working language is English.
Scientific topics:
<li>Systematics, Morphology & Phylogeny;
<li>Baltic Fauna & chorology;<li>Baltic Flora & chorology;
<li>Genetics and Biotechnology;
<li>Conservation Biology;<li>Ecology and Ecosystem Management;
<li>Forest Management & Biological Diversity;
File Edit Search Help
<p>The Conference working language is English. □□Scientific topics:□<li>Systematics, Morphology & Phylogeny;□<li>Baltic Fauna & chorology;
Izmanto_LS_teksts.txt x
1 <p>The Conference working language is English. chorology; Mar
2
line: 1 / 2 col: 0 sel: 0 INS TAB mode: Unix (LF) encoding: UTF-8 (with BOM) filetype: None scope: unknown

```

Pirmajā redaktorā redzams, ka ieraksts tiek uzvertts kā viena rindkopa sadalīta pa vairākām rindām.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Teksta failu piemēri

- Datu bāze (atdalīta teksta fails).

"Vieta "	→ "Garums"	→ "dg"	→ "Platums"	→ "dp"
"Aizkraukle"	→ "25 14"	→ 25,23	→ "56 36"	→ 56,6
"Aizpute"	→ "21 36"	→ 21,6	→ "56 43"	→ 56,72
"Alūksne"	→ "27 03"	→ 27,05	→ "57 25"	→ 57,42
"Auce"	→ "22 54"	→ 22,9	→ "56 28"	→ 56,47
"Baldone"	→ "24 25"	→ 24,42	→ "56 46"	→ 56,77
"Baloži"	→ "24 03"	→ 24,05	→ "56 46"	→ 56,77
"Balvi"	→ "27 16"	→ 27,27	→ "57 50"	→ 57,83
"Bauska"	→ "24 12"	→ 24,2	→ "56 25"	→ 56,42
"Brocēni"	→ "22 34"	→ 22,57	→ "56 42"	→ 56,7
"Cēsis"	→ "25 16"	→ 25,27	→ "57 19"	→ 57,32
"Dagda"	→ "27 32"	→ 27,53	→ "56 06"	→ 56,1
"Daugavpils"	→ "26 31"	→ 26,52	→ "55 52"	→ 55,87
"Dobele"	→ "23 17"	→ 23,28	→ "56 38"	→ 56,63
"Ērgļi"	→ "25 38"	→ 25,63	→ "56 54"	→ 56,9



© Kārlis Kalviškis, 2020.

Teksta failu piemēri

- Telpiska datu bāze (vektorkarte)

MapInfo Data Interchange Format (MIF)

*.MIF

*.MID

```
Version 300
Charset "WindowsBalticRim"
Delimiter ","
CoordSys Earth Projection 1, 104
Columns 3
  id Smallint
  Apraksts Char(32)
  Vertiba Decimal(7, 3)
Data

Pline 20
24.432854 57.811151
24.432854 57.617937
..
```

MIF ir piemērots formāts, kad karti „jāveido ar roku”.

Telpisko datu digitālā apstrāde

Teksta failu piemēri

- Telpiska datu bāze (vektorkarte)

The Drawing eXchange Format (DXF)

0	0	8	30
SECTION	SECTION	Valsts	0.000000
2	2	62	0
HEADER	TABLES	4	VERTEX
9	0	66	8
\$EXTMIN	ENDSEC	1	Valsts
10	0	40	10
300000.000000	SECTION	1.000000	471133.399000
20	2	41	20
154000.000000	BLOCKS	1.000000	316873.736000
9	0	0	30
\$EXTMAX	ENDSEC	VERTEX	0.000000
10	0	8	0
780000.000000	SECTION	Valsts	VERTEX
20	2	10	8
454000.000000	ENTITIES	472833.677522	Valsts
0	0	20	10
ENDSEC	POLYLINE	315185.656896	470250.676214
			..

© Kārlis Kalviškis, 2020.

Nāk no CAD lietotnēm. Ir dažādas versijas. Izstrādāja un uztur Autodesk. Nejaukt ar AutoCAD Binary DXF.



Telpisko datu digitālā apstrāde

Teksta failu piemēri

- *WKT* formāta teksta fails.

```
id|wkt|apraksts
```

```
1|LINESTRING(172.07 -43.60, 172.15 -43.57)|cels1
```

```
2|LINESTRING(172.07 -43.60, 171.92 -43.54, 171.88 -43.58)|cels2
```

- Ģeometrisko objektu veidi kartogrāfiskām vajadzībām:
 - POINT;
 - LINESTRING;
 - POLYGON;
 - MULTI- (POINT, LINESTRING, POLYGON);
 - GEOMETRYCOLLECTION.

WKT var tik izmantots datu apmaiņai kā arī glabāšanai datubāzēs, kas uztur telpiskus datus. Ir virknē vēl citu ģeometrisko objektu veidi, bet tos neatbalsta ne *PostGIS*, ne *Spatialite*.

Vēl cits *WKT* piemērs (lai arī nedaudz sagrozīts) ir *shape* faila *PRJ* fails. Viena no atšķirībām – viss ieraksts atrodas vienā rindā.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Teksta failu piemēri

- Rastra attēls

```
#define count_width 48
#define count_height 16
static char count_bits[] = {
0x00,0x00,0x00,0x00,0x00,0x00,0x00,
0x00,0x00,0x00,0x00,0x00,0x00,0x00,
0x00,0x00,0x00,0x00,0x3c,0x30,0x3c,
0x38,0x38,0x30,0x66,0x38,0x66,0x0c,
0x0c,0x38,0x66,0x30,0x60,0x06,0x06,
0x30,0x66,0x30,0x60,0x06,0x06,0x30,
0x66,0x30,0x38,0x3e,0x3e,0x30,0x66,
0x30,0x60,0x66,0x66,0x30,0x66,0x30,
0x60,0x66,0x66,0x30,0x66,0x30,0x60,
0x66,0x66,0x30,0x66,0x30,0x66,0x66,
0x66,0x30,0x3c,0x30,0x3c,0x3c,0x3c,
0x30,0x00,0x00,0x00,0x00,0x00,0x00,
0x00,0x00,0x00,0x00,0x00,0x00,0x00,
0x00,0x00,0x00,0x00,0x00};
```

X BitMap

0 1 3 6 6 1

(ļoti palielināts,
redzami atsevišķi pikseli)



© Kārlis Kalviškis, 2020.

X BitMap fails ir tāds kā C valodas izpildāms kods. Senatnīgs formāts.

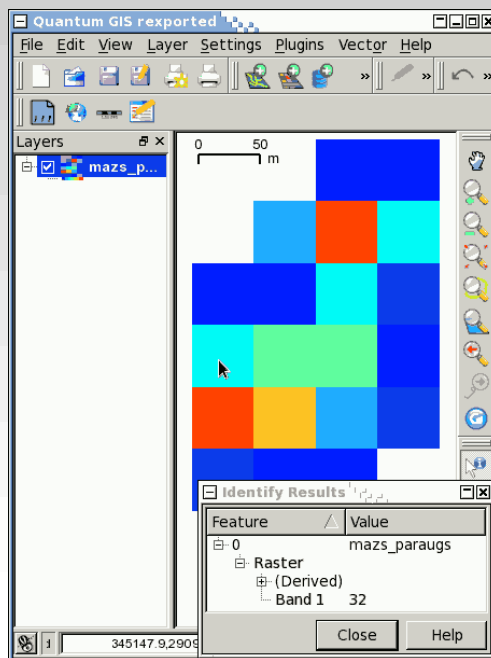
Telpisko datu digitālā apstrāde

Teksta failu piemēri

- Rastrkarte

ARC/INFO ASCII Grid Files

```
ncols      4
nrows     6
xllcorner 345125.8
yllcorner 290860.6
cellsize  50.0
NODATA_value -9999
-9999 -9999 5 2
-9999 20 100 36
3 8 35 10
32 42 50 6
88 75 27 9
13 5 1 -9999
```



Telpisko datu digitālā apstrāde

XML saimes formāti

- GML (*Geography Markup Language*)
The Open Geospatial Consortium, Inc (OGC)
ISO 19136:2007
- ArcXML (*Arc eXtensible Markup Language*)
ESRI
- KML (*Keyhole Markup Language*)
KMZ – kompresēts KML + pievienotie faili
Google un OGC
- SVG (*Scalable Vector Graphics*)
W3C

Diezgan neērti lasīšanai/labošānai
parastā teksta redaktorā.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Karšu apkopojumi

- Rastrkaršu apkopojumi – virtuālās kartes (*.vrt), *Idrisi* attēlu grupas (*.rgf), ... Veido mozaīkas vai daudzkanālu kartes. Parasti ir parasti teksta vai *xml* faili, kuros ir atsauces uz rastrkartēm.



© Kārlis Kalviškis, 2020.

Binārie faili



Telpisko datu digitālā apstrāde

Vektorkartes

- *ESRI:*
 - *ArcInfo Coverage;*
 - *Shape file;*
 - *GeoDatabase.*
- *Bentley Microstation DGN* formāts.
 - Vecā versija nesagādā problēmas.
 - Jaunā versiju (saukts par 8.), izmanto atvērtu maksas faila formātu. Licences nosacījumu dēļ nav iekļaujama brīvajā atvērta koda programmatūrā.

ArcInfo Coverage – piemērs, kādēļ ir slikti slēgtie failu formāti. Kādreiz tas bija ESRI ĢIS programmatūras pamatformāts. Mūsdienās ArcGIS ar pamatllicenci to nemāk atvērt.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Vektorkartes

- *DWG – DXF* binārā versija.
- *MapInfo TAB* fails (fails.dat, fails.id, fails.map, fails.tab).
- *Spatialite* datubāzes fails.
- *GeoPackage*.
- ...

Sākot no „AutoCAD” 10. versijas pastāv AutoCAD Binary DXF, kurā skaitliskās vērtības ir pierakstītās binārā formā. DWG ir pilnībā binārs fails.

Spatialite fails ir binārā formā. Pašā ģeometrijas laukā telpiskie dati var glabāties vai nu binārā, vai teksta formātā.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Rastrkartes

- Rastattēls + *world* fails vai metadatu palīgfails. Ne vairāk kā trīs (krāsu) kanāli, izņemot *TIFF*, kurā iespējams saglabāt vairākus kanālus.
- Koordinātas pašā failā – *GeoTIFF*, *MrSID*, *JPEG2000*, *ECW*, Erdas Imagine (.img), ... Parasti iespējams saglabāt vairākus kanālus.
- Datubāzes failā – *Spatialite*, *GeoPackage*.

MrSID un *ECW* ir slēgti formāti, kuru lietošanu ierobežo šo formātu licences. Labāk tā vietā izmantot *JPEG2000*, kas daļēji ir brīvi lietojams.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Rastrkartes

- Vērtību masīvs (*raw*) – *Idrisi Raster Format* (*.rst – vērtības, *.rdc – apraksts, *.smp – palete, *.ref – koordinātu sistēma), *BSQ*, *BIL*, *BIP*, ... Vērtību virkne. Atsevišķā failā tiek norādīts cik biti atvēlēti vienai vērtībai, cik aiļu un rindu dotajā masīvā, kā arī kādā secībā ierakstīti dati. Vienā failā var būt vairāki kanāli.

Pierakstīti var secīgi kanālus (*BSQ* – band sequential) (vispirms pirmais kanāls, tad otrais u.t.t.), pikselus (*BIP* – band interleaved by pixel) (vispirms visi pirmās rindas pirmās ailes pikseļi, tad pirmās rindas otrās ailes u.t.t.), vai rindas (*BIL* – band interleaved by line) (vispirms visu kanālu pirmās rindas, tad otrās rindas u.t.t.).



© Kārlis Kalviškis, 2020.

Kartēšanā biežāk izmantotie rastrattēlu failu formāti



Telpisko datu digitālā apstrāde

Biežāk lietotie rastrattēlu failu formāti

Piesaistei koordinātām vajag *world* vai metadatu failu!

.....

- *Joint Photographic Experts Group* izstrādātais formāts fotogrāfijām (*.jpg, *.jpeg, *.jpe).
- *CompuServe's Graphics Interchange Format* (*.gif).
- *World Wide Web Consortium* izstrādātais *Portable Network Graphics* (*.png).
- *Microsoft Windows Bitmap formatted image* (*.bmp).
Nekompresēts attēls.

Failu paplašinājumi nenosaka faila formātu un otrādi.

GeoTIFF != JPG ar geotag.

Pēdējais ir fotografēšanas punkta koordinātas, turpretī GeoTIFF ir karte (koordinātām piesaistīts attēls).



© Kārlis Kalviškis, 2020.

Rastrattēlu failu formāti (turpinājums)

- *Tagged Image File Format* (*.tif, *.tiff). Vairākas versijas. Iecienījuši tipogrāfijās. Ir versija telpiski piesaistītu karšu glabāšanai (*GeoTiff*).
- *JPEG 2000* (*.jp2, *.j2k), izstrādāja *Joint Photographic Experts Group*, izmanto citus (piedevām, dažādus) saspiešanas algoritmus nekā *JPG*. Pie vienādiem failu izmēriem *JPEG 2000* formāta attēls ir ar mazāk kropļojumiem kā *JPG*. Ir paredzēta vieta koordinātām.

JPEG 2000 tiek izmantots samērā sarežģīts un resursietilpīgs saspiešanas / atspiešanas algoritms. *JPG 2000* standartam ir vairākas daļas. Tikai pirmā daļa ir bezmaksas lietošanai.

GDAL atbalsts (2017.11.16):

JP2ECW vajag ECW SDK

JP2KAK Kakadu bibliotēku

JP2Lura vajag Lurawave bibliotēku

JP2MrSID vajag MrSID SDK

JP2OpenJPEG vajag OpenJPEG bibliotēku (v2)

Telpisko datu digitālā apstrāde

Rastrattēlu failu formātu salīdzinājums

Formāts	Krāsu dziļums, bitos	Cik krāsas var indeksēt, bitos	Iespējama caurspīdība	Faila saturs saspiests	Attēla bojājumi	Iespējami vairāki attēli vienā	Animācijas iespējas	CMYK atbalsts
BMP	1, 4, 8, 16, 24, 32	1, 4, 8	—	—	—	—	—	—
TIFF	1, 4, 8, 16, 24, 32, 40, ...	1, 4, 8	maska	— / +	— / +	+	—	+
GIF	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 6, 7, 8	indekss	+	—	+	+	—
JPG	8, 24, 32	—	—	+	+	priekšapskatei	—	— / +
PNG	1, 4, 8, 16, 24, 32, 48, 64	1, 4, 8	indekss / maska	+	—	—	—	—



© Kārlis Kalviškis, 2020.

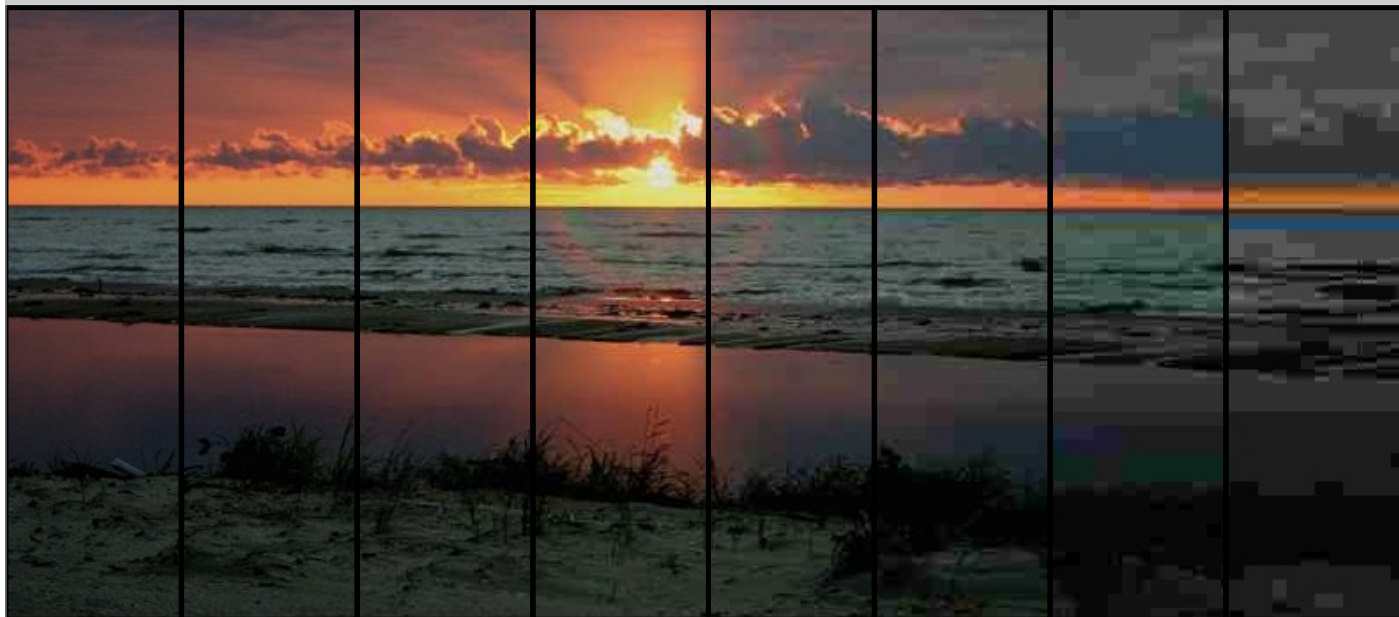
Telpisko datu digitālā apstrāde

JPG kompresija

mazāk

← *saspiešana* →

vairāk



labāka

← *kvalitāte* →

sliktāka



© Kārlis Kalviškis, 2020.

Veidots kā mozaīka no 8 attēliem, kuri saglabāti atsevišķi ar dažādu kompresijas pakāpi. Kreisā puse ir veidota no attēla ar vismazāko kompresiju, labā – ar visaugstāko.

Telpisko datu digitālā apstrāde

JPG kompresija

- Formāts paredzēt fotogrāfiju glabāšanai.
- Raibāki attēli ir mazāk jutīgi pret *JPG* kompresijas kropļojumiem.
- Fotogrāfijās ar vienmērīgām krāsu pārejām jāizmanto maza kompresijas pakāpe (ne lielāka par 5, tas ir, kvalitātei jābūt ne sliktākai par 95 %).
- *JPG* formāts nav piemērots grafisku attēlu glabāšanai.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

JPG kompresijas radītie kropļojumi



Attēla izmērs: 400×600, kvalitāte: 90, nokrāsu samazināšana: nav.

© Kārlis Kalviškis, 2020.

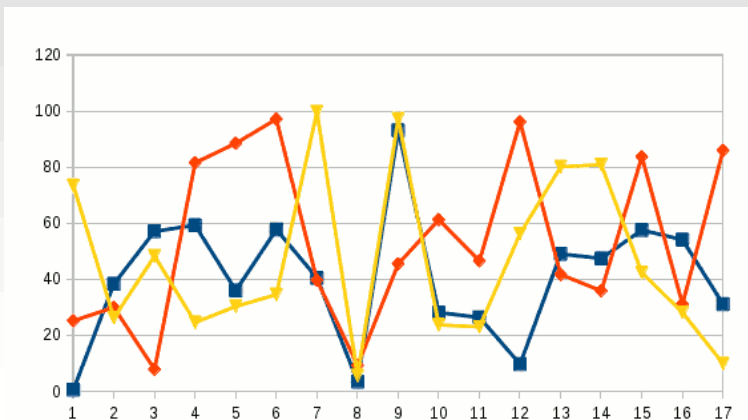


JPEG „klupšanas akmens” ir lēzenas toņu pārejas un smalki zīmējumi.

Telpisko datu digitālā apstrāde

JPG kompresijas radītie kropļojumi

- *PNG* attēls bez kropļojumiem. 24 bitu faila izmērs ~39,3kb. Ar optimizētu krāsu dziļumu(bitu skaitu pikselī) – ~10,8 kb.
- *JPG* – kompresija: 90, faila izmērs: ~48,8 kb.



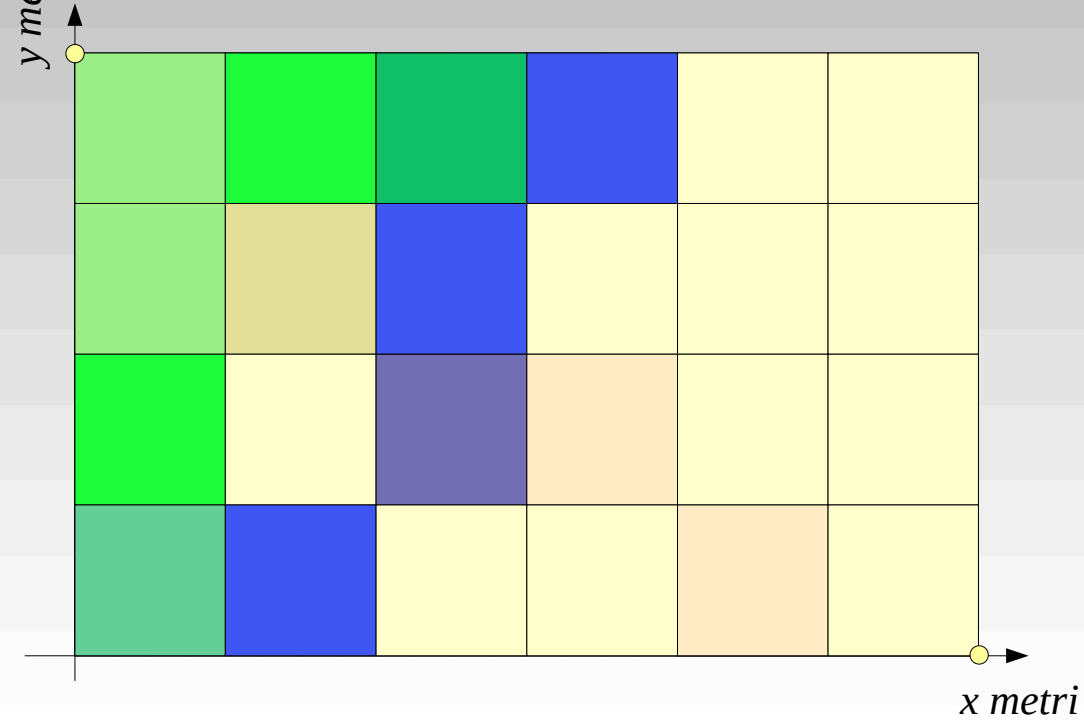
© Kārlis Kalviškis, 2020.

Rastrkaršu piramīdas



Telpisko datu digitālā apstrāde

Rastrkaršu piramīdas



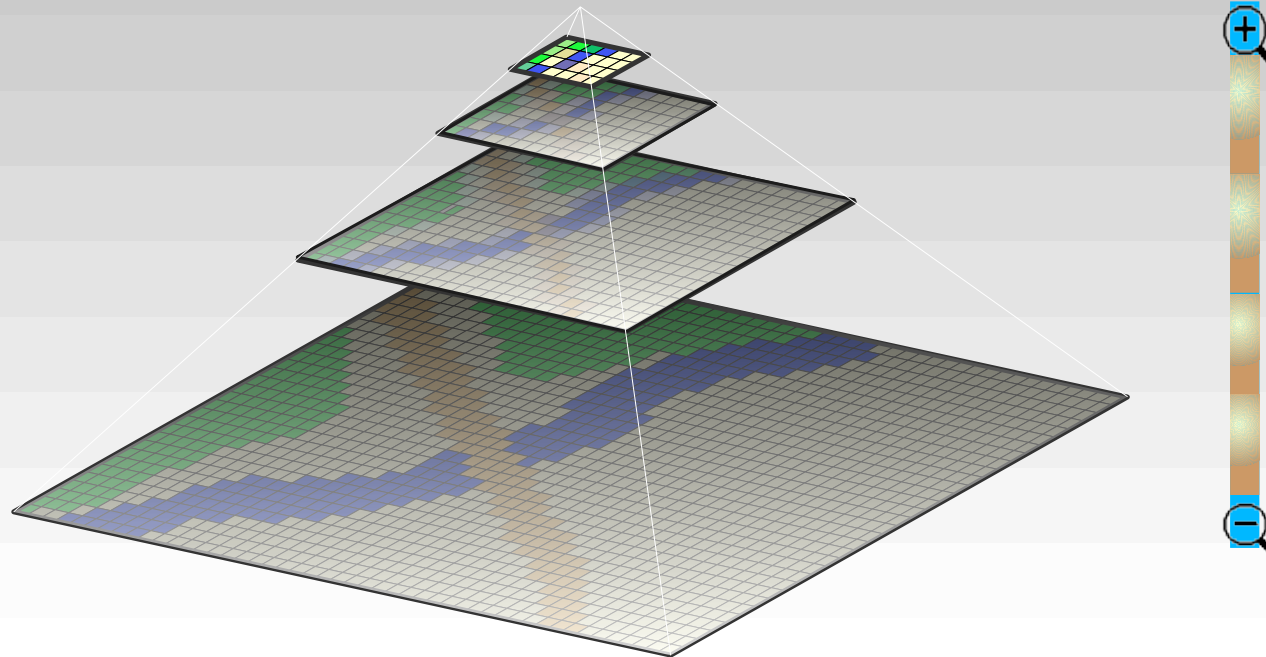
Pikseļi: $48 \times 32 \Rightarrow 24 \times 16 \Rightarrow 12 \times 8 \Rightarrow 6 \times 4$

© Kārlis Kalviškis, 2020.



Telpisko datu digitālā apstrāde

Rastrkaršu piramīdas



Atkarībā no faila formāta,
piramīdas var atrasties tai pašā failā,
kur karte, vai arī atsevišķā failā.



© Kārlis Kalviškis, 2020.

Datu servisi



Attālinātie dati

- Datubāzes „lokālajā tīklā” (nepieciešams serveris):
 - PostGIS (PostgreSQL), MySQL/MariaDB, Oracle, MS SQL ...
- Tīmekļa serveri balstīti uz atvērtiem standartiem:
 - *Web Map Service (WMS)* (attēls);
 - *Web Feature Service (WFS)* (aptaujājama karte, var veidot savu simboloģiju).
 - *Web Coverage Service (WCS)* (pārklājumi – laikā un telpā saistīti dati).

Lokālais tīkls var būt tikpat labi tas pats dators, pie kura strādā, kā dators blakus istabā vai citā kontinentā.

WMS – lietotājs saņem serverī sagatavotu rastrkarti.

WFS – lietotājs saņem vektorkarti. Tajā var veikt aptaujas gan telpiskajos, gan atribūtdatos. Attēlošanai pielieto savu simboloģiju.

WCS – dati izmantojami lietotāja veidotos modeļos. Dati var būt daudzdimensionāli.

QGIS „iedzimtās” vektorkartes



Telpisko datu digitālā apstrāde

Vektorkartes kā vietējie faili

- *Shape* faili;
- *Spatialite*;
- *GeoPackage*.



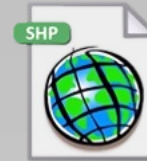
© Kārlis Kalviškis, 2020.



ESRI Shape faili



Telpisko datu digitālā apstrāde



Shape faila vēsture

- Sens formāts.
- Formātu izveidoja *Environmental Systems Research Institute (ESRI)* pagājušā gadsimta 90-os gados.
- Formāts ir publiski pieejams:
*„ESRI Shapefile Technical Description
An ESRI White Paper—July 1998”*.



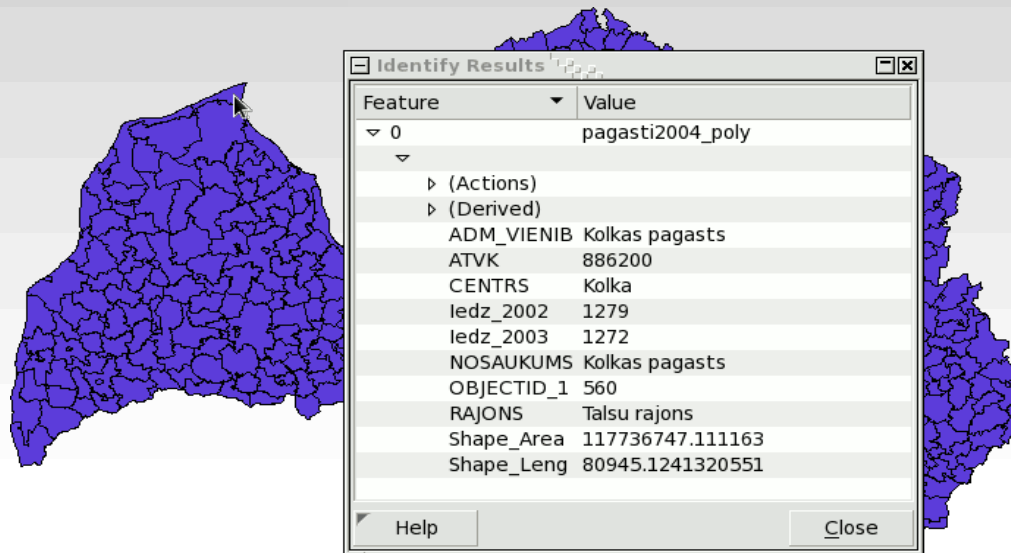
© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde



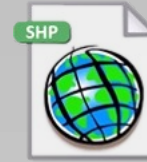
Telpiskie dati un atribūtdati

- Telpiskie dati var būt vai nu 2- vai 3-dimensionāli (x, y, z).
- Var saturēt noietā ceļa vērtību (m).



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

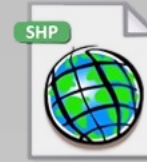


Shape „faila” uzbūve

- Pamatfaili:
 - *fails.shp* – telpiskie dati;
 - *fails.shx* – telpisko datu indeksa fails;
 - *fails.dbf* – atribūtdati.



© Kārlis Kalviškis, 2020.

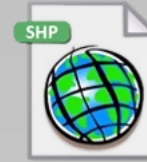


Shape „faila” uzbūve

- Papildus faili:
 - *fails.prj* – projekcijas fails;
 - *fails.cpg* – atribūtdatu tabulas teksta kodējums;
 -
 - *fails.qpj* – „QGIS” veidots projekcijas fails;
 - *fails.shp.xml* – shape faila apraksts (*metafails*);
 - *fails.sbn* – telpiskais indekss binārā formā, izmanto ESRI programmatūra;
 - *fails.?x?* – dažādu indeksu faili;
 - *fails.???* – citi piederīgi faili.

Teorētiski pašā DBF failā var uzdot teksta koda lapu, bet parasti tā izpaliek. Tai vietā var lietot CPG failu.

*.sbn un *.?x? var radīt nepatikšanas, ja vienu un to pašu failu pamīšus labo gan „ArcGIS”, gan kādu ne-ESRI produktu. Programmatūrām var veidoties dažāda izpratne par to, kas ticis pievienots, dzēsts un labots. Tādēļ šos indeksa failus, ja notiek Shape faila pārnese, jādzēš. Tas, protams, neattiecas uz *.shx failu, ka arī tad, ja fails paredzēts tikai skatīšanās režīmā.



Shape failu telpiskie dati

- Vienā *shape* failā var būt tikai:
 - vai nu **punktveida** objekti (*points*),
 - vai nu **līnijveida** objekti (*lines*),
 - vai nu **laukumveida** objekti (*polygons*).
- Failā var glabāties nesakārtoti vektordati (pārklāties laukumi utt.).
- Topoloģijas izveidi un uzturēšanu nodrošina izmantotā programmatūra.

Telpisko datu digitālā apstrāde



Teksta kodējumi (nepilnīgs saraksts)

Kodējuma veids	ID	Valodu grupa
ANSI Windows-1252	1252	Rietumeiropas
ANSI Windows-1257	1257	Baltijas
Unikods UTF-8	65001	

UTF-8 katrs nelatīņu burts rakstās ar diviem simboliem, attiecīgi vajag platāku lauku.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde



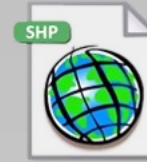
***Shape* faila atribūtdatu labošana**

- *QGIS*
Labot saturu atsevišķi katram objektam.
- *QGIS*
Slāņa īpašības/Lauki (ļauj labot tabulas struktūru).
- *QGIS*
Atribūtdatu tabula (ļauj labot tabulas saturu).



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde



Shape faila atribūtdatu labošana

- *Ar Ooo.Calc (LibreOffice Calc).*
- *Ar MS Office Excel (līdz 2003).*
- *Ar kādu datubāzu vadības sistēmu, kas lasa un raksta atbilstošās versijas (Dbase IV/FoxBase) DBF failus.*



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde



dbf uzbūve

- Pieļaujamo lauku veidi:
 - Teksts: C,254.
 - Vesels skaitlis: N,9,0 (4 baiti).
 - Daļskaitlis: N,19,17 (8 baiti, peldošā komata skaitlis)
(*OpenOffice.org* pieļauj 19,15).
 - Jā/Nē lauks: L (viena simbola lauks, kurš pieņem tikai ieraksta vērtības 'T', 'F', 'Y' un 'N' (piem., *QGIS*) vai '1' un '0' (piem., *ArcGIS*)). Tiešā veidā izveidot šo lauku nepiedāvā ne *QGIS* ne *ArcGIS*.
 - Datums: D (8 simboli, ievadlaukā – 10 simboli).



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde



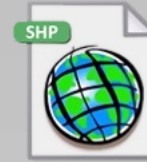
DBF faila labošana ārpus QGIS

- Nedrīkst izmest vai pievienot jaunus ierakstus
- Nedrīkst mainīt ierakstu secību
- Var pievienot jaunus laukus
- Var dzēst laukus
- Var mainīt lauku garumu un secību



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde



Shape failu priekšrocības

- Atvērts formāts.
- Lielākā daļa ĢIS programmatūru to saprot.
- Vienkāršs pēc savas struktūras.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde



Shape failu ierobežojumi

- Pamatā tie ir DBF IV formāta ierobežojumi:
 - neuztur .NULL. (NaN) vērtību (0 != .NULL.);
 - lauku nosaukumā var būt tika 10 simboli (latīņu burti + cipari + „_”);
 - var būt tikai 255 lauki;



© Kārlis Kalviškis, 2020.



Shape failu ierobežojumi

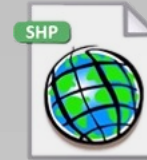
- DBF IV formāta ierobežojumi (turpinājums):
 - uzturētie datu veidi:
 - peldošā komata skaitļi (13 simboli), kurus pieraksta kā tekstu;
 - veseli skaitļi (4 vai 9 simboli);
 - datums (8 simboli), kuru pieraksta kā tekstu;
 - teksts (ne vairāk kā 254 simboli).
 - Tikai viena atribūtdatu tabula.
 - *.dbf un *.shp failu (katra) izmērs nedrīkst pārsniegt ~2 gigabaitus ($2^{31} - 1 = 2\ 147\ 483\ 647$ bitus).

Ja reālos skaitļos pieraksta kā tekstu:

- aizņem daudz vietas;
- ja cipari vairāk kā 12, tad skaitlis tiek noapaļots.

Laikam nav speciāla lauka.

Telpisko datu digitālā apstrāde



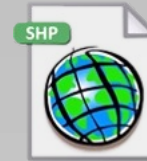
Pielietojums

- Ministru kabineta noteikumi Nr.940
Noteikumi par mikroliegumu izveidošanas un
apsaimniekošanas kārtību, to aizsardzību, kā arī
mikroliegumu un to buferzonu noteikšanu
(spēkā esoši, 29.10.2020)



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde



Pielietojums

- Ministru kabineta noteikumi Nr.940
 - [..]

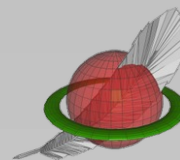
15. Eksperts veic pārbaudi dabā un iesniedz atbildīgajai institūcijai atzinumu, kura saturs atbilst normatīvajiem aktiem sugu un biotopu aizsardzības jomā, kā arī mikrolieguma kartoshēmu digitālā veidā vektordatu formā (.gml, .dwg, .dxf, .dgn vai **.shp** datņu formā) LKS-92-TM koordinātu sistēmā. Eksperts atzinumā ietver šo noteikumu 13.punktā minēto izvērtējumu.

[..]

Izcēlums mans.



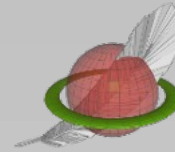
© Kārlis Kalviškis, 2020.



Spatialite datubāze



Telpisko datu digitālā apstrāde



SpatiaLight vēsture

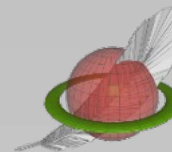
- Salīdzinoši jauns formāts.
- Balstas uz *SQLite* datubāzi (pamatautors *Dwayne Richard Hipp*).
- Formātu un atbilstošās bibliotēkas *Alessandro Furieri* sāka veidot 2008. gadā.
- <http://www.gaia-gis.it/gaia-sins/>



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

SpatiaLite

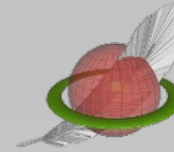


- Viens fails, viens vai vairāki slāņi.
- Gan vektorkartes, gan rastrkartes.
- Ilgu laiku bija viens no diviem *QGIS* „iedzimtajiem” formātiem (otrs ir *Shape* fails).
- Sākot no 2013. gada, *ESRI* programmatūra „*ArcGIS*” (10.2) atbalsta *SpatiaLite*.
- Telpiskos datus *SpatiaLite* formātā var izmantot arī „*AutoCAD Map 3D*” lietotāji.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde



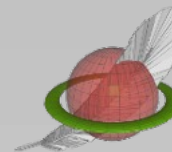
Spatialite vektorkartes

- Katrs ģeometrijas veids savā slānī:
 - punkti (*points*);
 - līnijas (*lines*);
 - daudzstūri (*polygon*);
 - punktu grupas (*multipoints*);
 - līniju grupas (*multilines*);
 - daudzstūru grupas (*multipolygon*);
 - rastrkarte.
- Vienā datubāzē var būt vairāki slāņi.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde



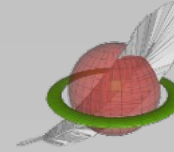
SpatiaLite ēnas puses

- Aizvien vēl strauji attīstās. Pašlaik rit darbs pie *libSpatiaLite* 5. versijas. (5.0.0-beta0, 2018.08.02). Stabilā verija ir 4.3.0 (2015.09.07). Savukārt rastru glabāšanai paralēli izmanto veco *libRasterLite* versiju 1.1g (2013.05.05), kuru pamazām aizstāj ar *libRasterLite2* (pašlaik 1.1.0-beta0, 2018.08.02).



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde



Spatialite ēnas puses

- *GDAL (OGR)* kompilējot bez papildus parametriem neiekļauj *SQLite* atbalstu.
- Komandas *ogr2ogr* noklusētais faila formāts ir *Shape* fails.



© Kārlis Kalviškis, 2020.



Ģeosainis (*GeoPackage*)



Telpisko datu digitālā apstrāde



Īsumā par *GeoPackage*

- *OGC* izstrādāts atvērtais standarts telpiskiem datiem.
- Pirmo reizi publicēta 2014. gadā.
- Balstās uz *SQLite 3*.
- Vektor-, rastr- un atribūdatiem.
- Lielākais izmērs ~140Tb, tā kā pamatā ierobežo failu sistēma nevis ģeosaiņa formāts.

OGC – Open Geospatial Consortium.



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde



Īsumā par *GeoPackage*

- Pieņemtais faila paplašinājums – „gpkg”.
- Standarta apraksts:
<http://www.geopackage.org/spec/>



© Kārlis Kalviškis, 2020.



GeoPackage pielietojums

- Diezgan daudzas programmatūras nodrošina darbu ar *GeoPackage*.
- Vektoru atbalsts jau no *GDAL* 1.11. versijas Lasīt/rakstīt +rastra atbalsts sākot no 2.0. Vēlams lietot vismaz 2.2. versiju.
- *QGIS* lasa/raksta vektordatus sākot no 2.10.1. versijas. Rastra datus sākot no 2.18. versijas.
- *ESRI ArcGIS Desktop* lasa/raksta vektorus sākot no 10.2.2. versijas. Rastra atbalsts sākot 10.3 versijas. *ArcGIS Pro* lasa vektorus sākot no 1.1. versijas.
- ...

Telpisko datu digitālā apstrāde



GeoPackage ēnas puses

- Ne visi atbalsta šo formātu. *GPKG* neatbalsta:
 - *Bentley Systems* produkti, piemēram, *Microstation*.
 - *Autodesk* produkti, piemēram, *AutoCAD*.
 - ...



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde



Pielietojums

- Ģeotelpisko datu apmaiņas formātus nosaka EP direktīva 2007/2/EK (2007. gada 14. marts), ar ko tika izveidota Telpiskās informācijas infrastruktūra Eiropas Kopienā (*INSPIRE*), un tās pavadošās regulas.
 - INSPIRE media-types register
<https://inspire.ec.europa.eu/media-types/application>
 - Media subtype name: geopackage+sqlite3
 - (registered 2018-08-10, last updated 2018-08-10)
 - Intended usage: Common

Balstoties uz šo direktīvu, šim formātam vajadzētu būt vienam no tiem formātiem, kurus valsts iestādēm jāpieņem arī Latvijā.

Papildus uzziņai



Telpisko datu digitālā apstrāde

Izmantotā literatūra

- Programmatūras rokasgrāmatas
- http://www.gdal.org/formats_list.html
- http://www.gdal.org/ogr/ogr_formats.html
- <http://www.opengeospatial.org/standards>
- <https://www.geopackage.org/>



© Kārlis Kalviškis, 2020.

Telpisko datu digitālā apstrāde

Izmantotā literatūra

- ESRI Shapefile Technical Description, An ESRI White Paper — July 1998
<https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>
- <https://docs.bentley.com/LiveContent/content/en/MicroStation%20Help-v13>
- <https://pro.arcgis.com/en/pro-app/help/projects/supported-data-types-and-items.htm>



© Kārlis Kalviškis, 2020.