

TESTOVÁNÍ PROSTŘEDÍ ORACLE SPATIAL NA PŘÍKLADU EPIDEMIOLOGICKÝCH DATDaniela SZTURCOVÁ¹, Pavel ŠVEC¹, Václav HÖNIG^{2,3}

¹ Institut geoinformatiky, hornicko-geologická fakulta, VŠB-TUO, 17. Listopadu 15, 708 33, Ostrava- Poruba, Česká republika, *daniela.szturcova@vsb.cz*, *pavel.svec1@vsb.cz*

² Přírodovědecká fakulta Jihočeské univerzity, Branišovská 31, 370 05 České Budějovice

³ Biologické centrum AV ČR, Parazitologický ústav, Branišovská 31, 370 05 České Budějovice
honigva@gmail.com

Abstrakt

V příspěvku nabízíme ukázky zpracování a vizualizace prostorových dat v prostředí prostorové databáze Oracle. Vycházeli jsme z úloh nad sadou epidemiologických dat získaných v rámci projektu "Klíšťata a jimi přenášená onemocnění v Jihočeském kraji a Bavorsku". Ty již byly dříve řešeny v prostředí ArcGIS. Stejnou sadu dat jsme převedli do prostředí Oracle Spatial a poté formulovali obdobné úlohy pomocí Spatial SQL. Použitelnost prezentovaného prostředí předvádíme na několika příkladech spolu s ukázkami vizuálních výstupů. V závěru jsme se pokusili stručně shrnout základní hodnocení obou prostředí, které jsme použili pro analýzy nad prostorovými daty.

Abstract

In this paper we show examples of spatial data processing and visualization in the Oracle spatial database. We used analyses over a set of epidemiological data obtained in the framework of the project "Ticks and tick-borne infectious diseases in the conditions of South Bohemia and Bavaria". The said analyses have been previously solved in the ArcGIS environment. The project data set was converted to Oracle Spatial, and then we formulated a similar tasks using Spatial SQL.

Klíčová slova: Spatial SQL; prostorová databáze; klíšťata; epidemiologická data; klimatická data

Keywords: Spatial SQL; spatial database; ticks; epidemiological data; climatic data

1. ÚVOD

V posledních letech se na trhu objevují nové technologie pro zpracování prostorových dat. Ať už z důvodu tlaku uživatelů, kteří objevují pohodlí při práci s daty s prostorovým kontextem, nebo z důvodu tlaku na výrobce software, aby své produkty rozšířili o práci s prostorovými daty.

Jedním z výrobců, kteří nejen že tento trend zachytili, ale můžeme i říct, že dnes určují standardy, je firma Oracle. Už od svého počátku koncipovala svou filosofii na novinkách. V 70. letech minulého století patřila k prvním, kteří dokázali implementovat novou myšlenku ohledně zpracování dat. Jednalo se tehdy o relační model, který zcela odlišným způsobem přistupoval k ukládání a následné manipulaci s daty. Na rozdíl od tehdy rozšířeného hierarchického modelu byl princip, poprvé popsany E. Coddem [1], založen na využití matematických základů – pojmu relace, jako formálního popisu vztahů mezi prvky.

Tato technologie po desítkách let ukázala svoji životaschopnost a naprosto převládla v databázových technologiích. V dnešní době se rozšiřuje ještě o objektové principy, které v případě propojení obou vede k databázím označovaným jako objektově-relační.

Jak databázové technologie souvisí s prostorovými daty? Výrobci systémů řízení báze dat, kteří zachytili trend zpracování prostorových dat, dnes uživatelům nabízejí možnost pracovat nikoli jen s klasickými

atributovými daty, ale také nabízejí zpracování dat svázaných jak s geometrickou složkou, tak i topologickými vztahy. Ukážeme si takovou nabídku, která případného uživatele bude stát čas na

- stažení aplikací,
- jejich instalaci,
- a seznámení se, jak je používat.

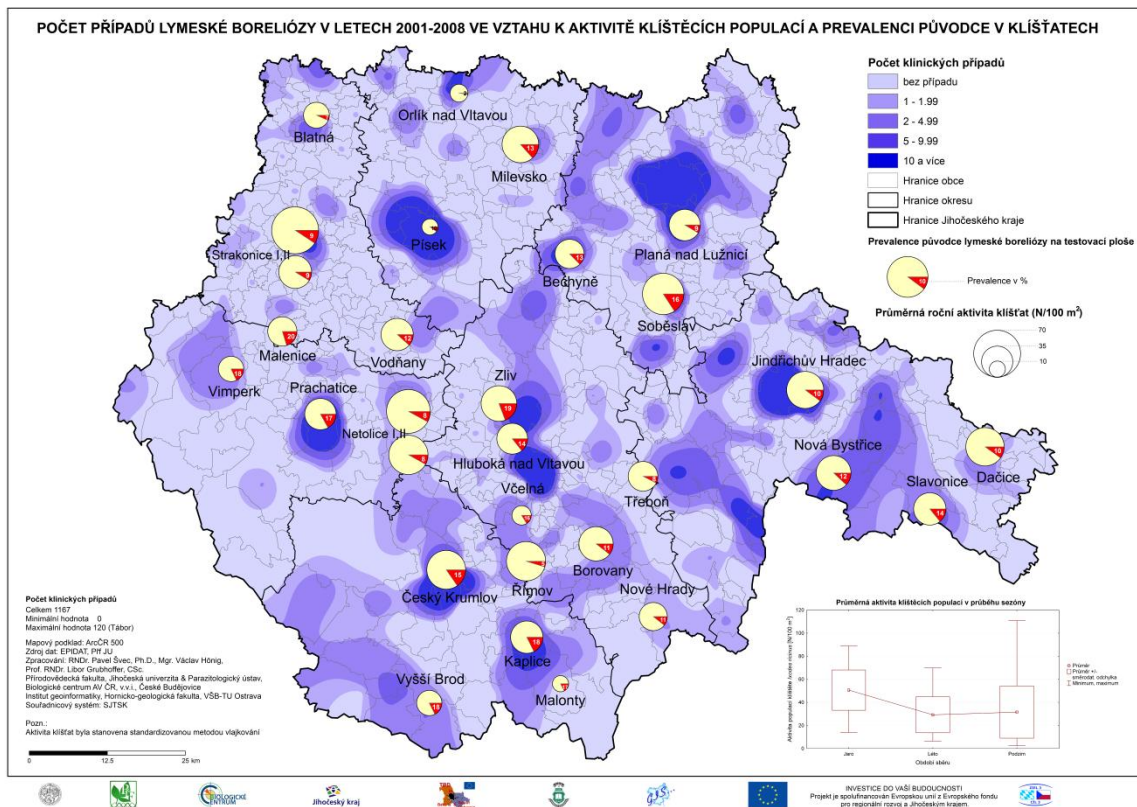
Předvedeme propojení databáze Oracle s vizualizační extenzí GeoRaptor. Firma Oracle nabízí instalaci svého produktu Oracle ve verzi Express Edition, označované jako Oracle XE, zcela zdarma. Také prostředí pro ovládání databáze je možno ze stránek firmy stáhnout bez poplatku - jedná se o klientskou stranu označovanou SQL Developer. K vizualizaci prostorových dat jsme pak použili prostředí GeoRaptor, který je dnes již integrován právě do klientského prostředí SQL Developer.

V našem příspěvku chceme ukázat některé možnosti takové technologie, která ukazuje na prolínání dvou dříve oddělených skupin software. Následný text je rozdělen do kapitol, kde popíšeme data, se kterými jsme prostorovou databázi zkusili a posléze ukážeme několik příkladů použití.

2. DATA A METODY

Data, se kterými jsme pracovali, jsme získali v rámci tříletého projektu „Klíšťata a jimi přenášená onemocnění v Jihočeském kraji a Bavorsku“, který probíhal v letech 2009-2011. Projekt byl zaměřen na geografické zmapování výskytu klinických případů klíšťové encefalitidy (KE) a lymeské boreliózy (LB) v Jihočeském kraji a v Bavorsku, stanovení aktivity klíšťat, modelování predikce rizika infekcemi klíšťaty přenášených onemocnění a dalších aktivit v tomto území. V rámci projektu došlo ke sběru klíšťat na 30 lokalitách v Jihočeském kraji a 20 lokalitách v částech Bavorska. Klíšťata byla sbírána v Jihočeském kraji v roce 2008 a v Bavorsku v roce 2010. Sběry proběhly ve třech vlnách reprezentujících aktivitu klíšťat v rámci sezóny - jarní sběr (květen), letní sběr (červenec), podzimní sběr (září). Celkem bylo v Jihočeském kraji nasbíráno 20 057 jedinců, kteří byli dále laboratorně testováni na přítomnost klíšťaty přenášených nákaz (Hönig et al. 2011). V rámci tohoto příspěvku jsme pracovali pouze s daty Jihočeského kraje.

Po celou dobu projektu byly využity geografické informační systémy (GIS), konkrétně software ArcGIS verze 9.X od firmy ESRI a dále program Statistica a MS Excel. GIS byl použit pro nalezení vhodných lokalit pro sběr klíšťat, tvorbu mapových podkladů např. o rozšíření počtu onemocnění KE a LB či vizualizaci počtu nasbíraných klíšťat a údajích o prevalenci klíšťat (Obr. 1). V prostředí GIS byly také zjišťovány různé fyzicko-geografické, environmentální a klimatologické charakteristiky (NDVI, teplota povrchu v době sběru klíšťat). Zjištěné hodnoty byly vždy vztaženy k jednotlivé lokalitě či konkrétnímu sběru klíšťat. V kombinaci GIS a statistického software Statistica bylo také provedeno modelování a predikce rizika klíšťaty přenášených onemocnění. Výsledky jsou pro veřejnost dostupné prostředím mapového portálu na gis.vsb.cz/klisjata.



Obr. 1. Ilustrace mapového výstupu. Ukázka vizualizace výstupů v rámci projektu v prostředí ArcGIS.

Data se kterými jsme pracovali, můžeme obecně rozdělit do dvou skupin:

Data vztahující se k lokalitě sběru - data se vztahují k lokalitě, ve které byl sběr klíštět prováděn. Jedná se o údaje o lokalitě sběru, které jsou v průběhu sledovaného období v podstatě neměnné (např. nadmořská výška, expozice vůči světovým stranám, sklon svahu, vegetační pokryv) a data průměrovaná za celou sezónu (např. průměrná roční teplota povrchu, vzduchu).

Data vztahující se ke sběru - v tomto případě se jedná o data, vztažená na jednotlivé sběry. Ke každému sběru jsou pak uvedeny údaje vztažené pro daný časový okamžik sběru. Příkladem dat z této kategorie je například hodnota NDVI, jež je závislá na ročním období, počet klíštět nasbíraných za jednotlivý sběr a výsledky laboratorního testování, výsledky fytoocenologického snímkování, klimatická data z ČHMÚ pro jednotlivé sběry (min. a max. teplota, srážky) atd. Celkem jsme měli k dispozici sadu 120 atributů (charakteristik) pro každou lokalitu. Data byla uložena v souborové geodatabázi.

3. POUŽITÍ TECHNOLOGIE ORACLE LOCATOR

Zkusme se nyní podívat na možnost použít databázové technologie, jak získat odpovědi na otázky, které v projektu "Klíštěta a jimi přenášená onemocnění v Jihočeském kraji a Bavorsku" vyvstaly. Jednalo se zejména o ověřování souvislostí mezi polohou lokalit sběru a různými vlivy, u kterých se předpokládalo, že mohou mít vliv na velikost populace klíštět. Jak již bylo zmíněno výše, v projektu se používala technologie firmy ESRI a statistika byla tvořena výpočty v MS Excel či programu Statistica. Ovšem ukázalo se, že v prostorové databázi můžeme pomocí dotazů vytvořit statistické výstupy i vyhledat geoobjekty vyhovující prostorovému filtru. Některé takové úlohy jsme ověřili v prostředí databáze Oracle.

3.1 Práce s prostorovými daty v prostředí Oracle

Jak již bylo zmíněno v úvodu, firma Oracle je vedoucí společností na trhu s databázovými technologiemi. První náznaky práce s prostorovými daty lze najít už ve verzi 7 z devadesátých let minulého století, kdy se objevila možnost ukládat bodová data, později se k bodům přidávají i linie a polygony a první prostorové operátory. Od verze 8i je možné pracovat s georeferencováním, transformací souřadnic apod. Od verze 9i jsou to už prostorové agregované funkce, ve verzi 10g práci s rastrovými daty, topologií a síťovými analýzami. V Oracle 11g přibývá 3D support, WMS, modelovací nástroje.

Rozšířením svého produktu o prostorová data nabídla firma Oracle svým zákazníkům nejen systém řízení báze dat (laicky označované jako databáze Oracle), ale také možnost pracovat s prostorem. Uživatelé mohou – jak jsou klienti znalí GIS aplikací zvyklí – ukládat prostorové objekty jako geoprvky a mohou s nimi provádět i některé analytické úlohy nad prostorovými daty. Do doby změny koncepce se prostorová data do běžné relační databáze ukládala jako dva či tři sloupce s hodnotami souřadnic, nyní lze opravdu pracovat s prostorovými prvky včetně funkcí, které jsou postaveny na geometrii geoprvku. U Oracle databáze bylo toto rozšíření dáno i změnou na úrovni paradigmatu – technologie zpracování dat se posunula od relačního modelu k objektově-relačnímu.

Zde si popíšeme, co nám vlastně objektově-relační databázový model nabízí. Jako důležité pro uživatele vidíme princip prolínání atributových dat a reprezentace prostorové složky. Každý záznam v databázové tabulce lze chápat jako řádek obsahující popisná data spolu s objektově pojatým popisem geometrické a metadatové složky objektu. *V čem vidíme zásadní rozdíl od pojetí GIS software je možnost ukládat do jedné tabulky bodové, liniové i polygonové prvky.* Je to umožněno způsobem implementace standardu OGC.

Geoobjekt je popsán formou objektu, kde atributové složky umožňují jak popis prvku, tak také objektový přístup k operacím nad prostorovou složkou dat. U každého geoprvku je pak vyplňováno pět atributů:

SDO_GTYPE, SDO_SRID, SDO_POINT, SDO_ELEM_INFO a SDO_ORDINATES. První dvě složky představují metainformace o typu geoobjektu a souřadnicové soustavě. Další atribut SDO_POINT má v případě bodových dat vyplněny koordináty bodu, nebo – v případě jiného typu dat – hodnotu NULL. V takovém případě se jedná o geoobjekt jiného typu (linie, polygon, sada prvků) a pak jsou naplněny hodnoty SDO_ELEM_INFO a SDO_ORDINATES. Ty naopak u atributu SDO_POINT mají uvedenu hodnotu NULL. Tím je zajištěno, že se do jedné tabulky mohou ukládat data o geoprvcích s různými geometriemi, jen musí zůstat zachována stejná dimenze pro všechny prvky.

Vložení jedné ze stanic ČHMÚ jako bodového prvku vypadá takto:

```
INSERT INTO stanice_geo VALUES (2, 'Temelín', 'C1TEME01',
SDO_GEOMETRY(2001,
NULL,
SDO_POINT_TYPE(-762023.5,-1140243.6, NULL),
NULL,
NULL));
```

Po vytvoření prostorového indexu pak prostorové funkce využívají charakteristiku danou „umístěním“ prvku ve virtuální prostorové struktuře. Vyhodnocování dotazu probíhá ve dvou fázích. V prvním filtru se hledají kandidáti vyhovující vyhledávací podmínce, čímž se omezí množina prvků, nad kterými probíhají náročné prostorové výpočty. Teprve v druhé fázi je ověřováno, zda kandidát splňuje prostorový filtr či ne.

Co uživatel potřebuje znát k tomu, aby mohl námi nabízené řešení použít? Je vhodná znalost jazyka SQL. Jedná se o silný nástroj, který dovoluje uživateli přistupovat k datům z mnoha hledisek. Díky dlouholetému vývoji a všeobecně uznávanému standardu je dotazovací jazyk SQL (Structured Query Language) používán uživateli, kteří chtějí nad daty zjistit více, než kolik jim nabízí „klikací“ aplikace. Tento jazyk můžeme chápat jako sadu příkazů, které uživateli dovolují ve vlastní režii tvořit novou strukturu databáze, manipulovat s daty, konstruovat pohledy na data z různých úhlů, apod.

Pro prostorová data implementovala firma Oracle v rámci standardu ISO [4] rozšíření jazyka SQL o několik důležitých prvků, které je označováno Spatial SQL. Jednak – v souladu s definováním databázových entit –

Ize nadefinovat geoprvek jako entitu společně s jeho prostorovou složkou. K ní se dají uložit i metadata jako je souřadnicový systém, prostorové okno – souřadnice osy X (Longitude) a Y (Latitude) (případně i Z), včetně přesnosti, s jakou budeme moci daná data v souřadnicovém systému používat.

Dále je nutné pro použití některých prostorových funkcí data zaindexovat. To řeší tvorba prostorového indexu založeném na R-stromu, který zrychlí práci s daty při operacích vyhledávání a umožní také využít topologických vazeb, vyplývajících z prostorových vazeb uložených prvků. V samotném jazyce se pak objevila nová klíčová slova, která při jejich použití umožňují zjistit například vzdálenost geometrií prvků, či nejbližšího souseda apod.

Co je však podstatné, uživateli databáze je nabízeno pracovat s daty v rámci celého souboru dat najednou během jedné operace. Oproti GIS prostředí lze nadefinovat filtr nad několika různými typy geoprvků současně. V GIS technologiích se komplikovanější výběr řeší postupným vypracováním nad jednotlivými vrstvami, vytváří se vrstvy s mezivýsledky a ty se poté „skládají“ do konečné podoby.

3.2 Datová struktura

Ukažme si princip použití jazyka SQL rozšířeného o prostorovou složku. Nejprve si stručně popišme, jaká data máme k dispozici. Tabulky v databázi můžeme rozdělit na dva typy. Jeden typ obsahoval pouze hodnoty datového typu text, číslo, datum (atributové položky), u druhého typu byly pro každý záznam ještě navíc uložena data popisující prostorovou složku. Omezili jsme se na data vektorová, pro která lze využít topologických vztahů vzniklých po prostorovém zaindexování. Tabulky pak měly tento význam:

sber_jaro, sber_let, sber_podzim – data o sběrech v lokalitách, kde probíhaly sběry klíšťat ve třech obdobích roku 2008 (jaro, léto, podzim)

lokality_jk – popis lokalit v Jihočeském kraji,

KE – data o počtu případů klíšťové encefalitidy (KE) v obcích, kde se sledované lokality nacházejí

LB – data o počtu případů lymeské boreliózy (LB) pro sledované lokality

stanice_geo – seznam stanic ČHMÚ, které obsahují polohu stanic a jejich označení

hranice_okresy – hranice okresů Jihočeského kraje

Veškeré zde uvedené ukázky byly zpracovány v prostředí Oracle XE 10g¹ s přístupem přes klienta SQL Developer. Aby výsledky práce s prostorovými daty byly viditelné pro lepší pochopení souvislostí, budeme je vizualizovat v extenzi GeoRaptor².

Nyní přistupme k ukázkám, jak nám rozšíření Spatial SQL může pomoci ověřovat hypotézy spojené s výskytem klíšťat v různých oblastech.

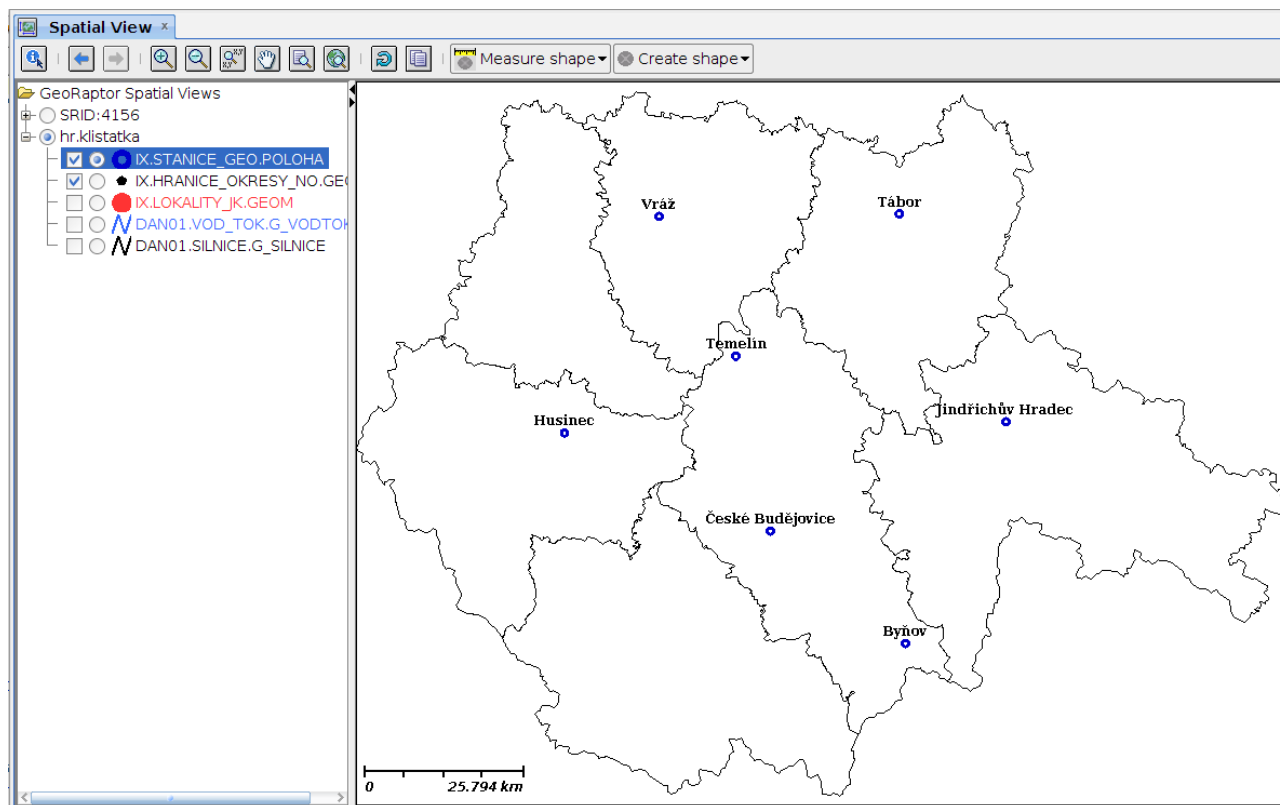
3.3 Ukázky práce v prostředí Oracle Locator

Úloha 1:

Nejprve ukážeme, co nám výše popsané prostředí nabízí. Chceme-li se „podívat“ na geobjekty v databázi uložené, postačí využít funkce pro vizualizaci prostorové složky objektů z databáze a použít zobrazení v prostředí GeoRaptor, jak je vidět na obrázku 2.

1 Zdarma poskytovaná licence firmy Oracle včetně možnosti zpracování prostorových dat v modulu Locator. Placené verze Oracle – Standard Edition a Enterprise Edition – obsahují rozšíření Spatial s větším množstvím prostorových funkcí, včetně možnosti práce s rastrovými daty.

2 GeoRaptor je volně šiřitelná aplikace, publikovaná na webových stránkách sourceforge.net/projects/georaptor/. Uživateli je tak dán k dispozici nástroj pro zviditelnění prostorových prvků, nadefinovaných v databázi či ve výsledku dotazu.



Obr. 2. Ukázka vizualizace prostorových dat v prostředí GeoRaptor. Zobrazeny jsou dvě tabulky: stanice ČHMÚ modrým symbolem a hranice okresů pomáhají vymezit prostorový kontext sledovaných geobjektů.

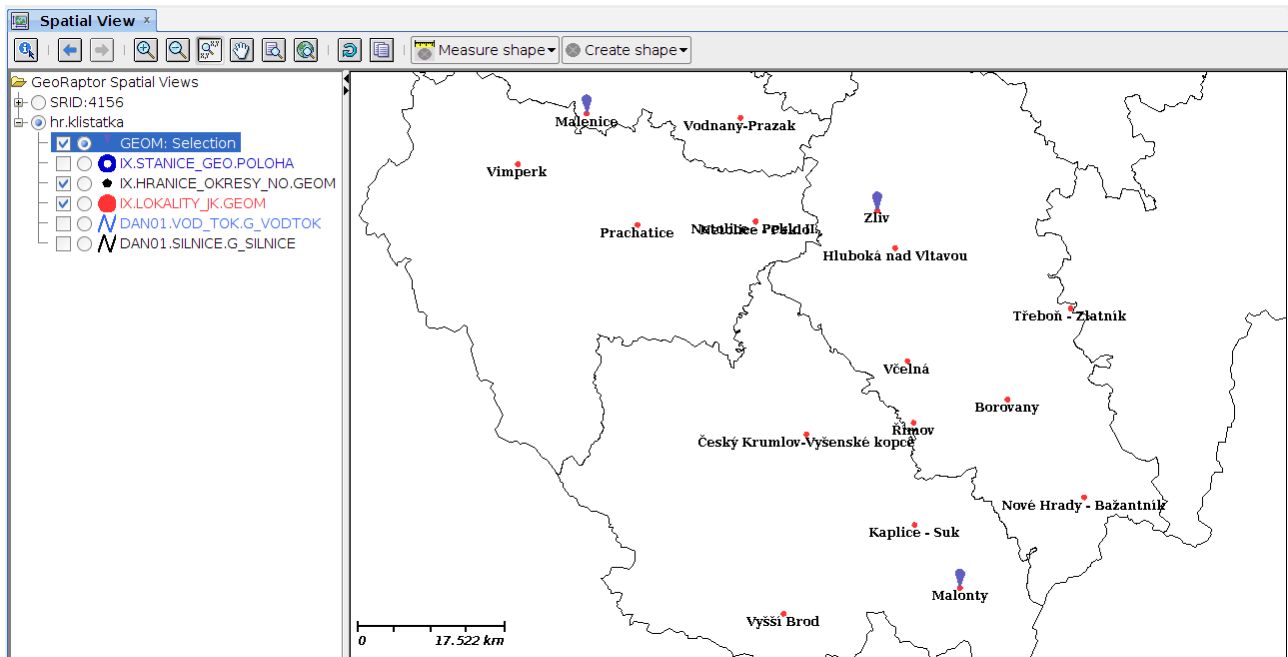
Chceme-li vytvořit základní statistiku o datech, která používáme, většinou máme k dispozici funkce v tabulkových procesorech nebo použijeme některý ze statisticky orientovaných software. V databázi jsou nám k dispozici agregované funkce, jejichž použití pomáhá zjistit jednoduché základní charakteristiky datového souboru.

Úloha 2:

Zkusme si nyní ukázat, jak pomocí dotazu s použitím agregované funkce zjistíme, ve které sledované lokalitě bylo hlášeno maximální množství případů lymeské boreliózy (LB)? Maximum patří mezi agregované funkce, které jsou v jazyce SQL definovány. Protože máme taková data propojena i na prostorovou složku, lze z výsledku dotazu ihned zjistit polohu místa s maximálním výskytem LB.

Dotaz pak může vypadat takto:

```
SELECT l.cislo, l.geom FROM lokality_jk l
WHERE l.cislo in (SELECT lb.idlokality FROM lb WHERE lb.vyskyt =
(SELECT min(lb2008) FROM lb));
```



Obr. 3. Výsledek vyhledání lokality s minimálním počtem případů lymeské boreliózy. Stejnou hodnotu minima mají tři lokality: Malenice, Malonty a Zliv. Tyto lokality s minimem případů LB jsou na obrázku zvýrazněny modrým symbolem.

Úloha 3:

Budou-li nás zajímat souhrnné hodnoty, například:

Kolik bylo nasbíráno všech jedinců klíšťat (nymf, samic i samců) na jednotlivých lokalitách při všech sběrech?

K výstupu tohoto typu můžeme využít možnost nadefinovat pohled (View), který nám umožní uchovat agregovaná data v požadované podobě. Pohled nám pak dále poslouží ke zjištění celkové sumy nasbíraných jedinců, kde je opětovně využita agregovaná funkce sum. Vytvoření pohledu a výpočet počtu jedinců na jednotlivých lokalitách jsou patrné na obrázku 4. V prostředí relační databáze je použita jedna agregační funkce, v prostředí ArcGIS to znamená kombinaci několika kroků.

```

create view sumacni_v as
select j.nazev, j.lokalita, (sum(j.nymfy) + sum(j.samice) + sum(j.samci)) as "Suma jedinců za jaro",
(sum(l.nymfy) + sum(l.samice) + sum(l.samci)) as "Suma jedinců za léto",
(sum(p.nymfy) + sum(p.samice) + sum(p.samci)) as "Suma jedinců za podzim"
from sber_jaro j, sber_let l, sber_podzim p
where j.lokalita=l.idlokality and j.lokalita=p.idlokality
group by j.nazev, j.lokalita
order by j.lokalita;

```

	LOKALITA	Suma jedinců za jaro	Suma jedinců za léto	Suma jedinců za podzim
1	1	414	315	209
2	2	533	419	172
3	3	257	111	259
4	4	250	78	101
5	5	337	39	188
6	6	252	182	247
7	7	420	187	665
8	8	246	124	81
9	9	84	60	60
10	10	137	80	53
11	11	361	384	193
12	12	405	104	234

Obr. 4. Vytvoření pohledu na agregovaná data v jednotlivých lokalitách využitím funkce sum. Sumy byly vypočteny pro jednotlivá období sběru a byl zjištěn celkový počet nasbíraných jedinců pro každou lokalitu zvlášť.

Úloha 4:

Dále díky vypočteným sumám můžeme zjišťovat, kolik jedinců máme nasbíráno celkem, kolik jich tedy máme vyšetřovat? Toho dosáhneme jednak celkovou sumací za každé období, přičemž součet těchto sum nám ukáže počet všech nasbíraných jedinců za všechny lokality a všechna období sběru. Konstrukce dotazu a výsledné hodnoty jsou předvedeny na obrázku 5.

```

select sum("Suma jedinců za jaro") as "Nasbíráno na jaře",
sum("Suma jedinců za léto") as "Nasbíráno v létě",
sum("Suma jedinců za podzim") as "Nasbíráno na podzim",
(sum("Suma jedinců za jaro") + sum("Suma jedinců za léto") + sum("Suma jedinců za podzim")) as "Celkem nasbíráno"
from sumacni_v;

```

	Nasbíráno na jaře	Nasbíráno v létě	Nasbíráno na podzim	Celkem nasbíráno
1	9115	5262	5680	20057

Obr. 5. Znovu použití funkce sum a předvedení výstupních hodnot jako vstupů do aritmetické operace v témže dotaze.

Úloha 5:

Budeme-li chtít zjišťovat vliv klimatických poměrů na počet nasbíraných jedinců klíčící populace, pak je dobré vědět, které lokality spadají do určité vzdálenosti od měřících stanic. Díky prostorové funkci *sdo_distance* jsme schopni zjistit vzdálenost mezi všemi bodovými geoobjekty navzájem. My ovšem výpis omezíme na vyhledávání pěti nejbližších lokalit sběru u každé stanice ČHMÚ. Pro lepší přehlednost

seřídíme výsledná data podle abecedního pořadí v názvech stanic a vzdálenosti mezi stanicí a lokalitou. Část výsledku je vidět na obrázku 6, včetně zápisu dotazu v klientském prostředí SQL Developeru.

```
select l.nazev as "Lokalita sběru", s.name as "Stanice ČHMÚ",
mdsys.sdo_geom.sdo_distance(l.geom,s.poloha,1) as "Vzdálenost"
from lokality_jk l, stanice_geo s
where (sdo_nn(l.geom, s.poloha, 'sdo_num_res=5',1)='TRUE')
order by 2, 3;
```

Lokalita sběru	Stanice ČHMÚ	Vzdálenost
1 Nové Hrady - Bažantník	Byňov	3177.04554494558
2 Borovany	Byňov	15364.793032563
3 Třeboň - Zlatník	Byňov	21108.6486330959
4 Malonty	Byňov	21504.5535172084
5 Kaplice - Suk	Byňov	23428.4927430429
6 Prachatic	Husinec	3607.85869867659
7 Malenice	Husinec	11743.5436986249
8 Netolice - Peklo II.	Husinec	13878.2891236105
9 Netolice - Peklo	Husinec	13892.5689630503
10 Vodnany-Prazak	Husinec	14883.5254550026
11 Jindřichův Hradec	Jindřichův Hradec	5557.91199210984
12 Nová Bystřice	Jindřichův Hradec	17796.2291951444
13 Soběslav	Jindřichův Hradec	22237.2840576468
14 Třeboň - Zlatník	Jindřichův Hradec	25663.5557997
15 Planá nad Lužnicí	Jindřichův Hradec	28549.4184385789
16 Zliv	Temelín	14301.5650816808
17 Bechyně	Temelín	15012.5715404436
18 Vodnany-Prazak	Temelín	16425.1072788214

Obr. 6. Vyhledávání nejbližších lokalit. Seznam pěti nejbližších lokalit pro každou stanici ČHMÚ včetně uvedení vzdálenosti mezi geobjekty v metrech.

Úloha 6:

Zkusme zjistit, kolik bylo v každém okrese nasbíráno jedinců při jarním sběru.

Postupujeme tak, aby na základě topologických vztahů byly lokality přiřazeny do patřičného okresu (využití funkce *sdo_join*) a teprve potom je provedeno sečtení hodnot pro nymfy, samice, samce a všechny jedince na lokalitě (obrázek 7).

```
select o.nazev as "Okres", sum(j.nymfy) as "Počet nymf",
sum(j.samice) as "Počet samic", sum(j.samci) as "Počet samců",
(sum(j.nymfy) + sum(j.samice) + sum(j.samci)) as "Suma jedinců za jaro"
from table(sdo_join(
'lokality_jk', 'geom',
'hranice_okresy_no', 'geom',
'mask=ANYINTERACT')) c,
lokality_jk l, hranice_okresy_no o, sber_jaro j
where c.rowid1 = l.rowid and c.rowid2 = o.rowid
and j.lokalita = l.objectid
group by o.nazev
ORDER BY o.nazev;
```

Okres	Počet nymf	Počet samic	Počet samců	Suma jedinců za jaro
1 Jindřichův Hradec	1245	73	68	1386
2 Prachatic	1286	85	83	1454
3 Písek	566	22	20	608
4 Strakonice	1326	85	102	1513
5 Tábor	873	37	34	944
6 České Budějovice	2014	80	86	2180
7 Český Krumlov	965	32	33	1030

Obr. 7. Výpočet nasbíraných klíšťat. Pro jednotlivé okresy jsou na základě polohy lokalit spočteny sumy jedinců v jarním sběru.

Pokud bychom chtěli v prostředí ArcGIS výsledky stejné aritmetické operace museli bychom postupovat jinak. Možnosti práce s atributy jsou v prostředí ArcGIS poněkud omezené. Výběry se provádějí také pomocí jazyka SQL, ovšem jedná se o stavebnici dotazů (Building a query). Je možné provádět výběry současně s více atributy v jedné vrstvě, nelze však definovat například minimální či maximální hodnotu. Často jsme však nuceni provádět sekvence jednotlivých výběrů na již provedeném výběru. Pro pokročilejší práci a výpočty v atributové tabulce se často neobjedeme bez exportu a následného zpracování např. v software Statistica či MS Office a zpětném importu do ArcGIS (tak tomu bylo i při přípravě modelů rizika nákazy klíšťaty přenášených patogenů).

ZÁVĚR

V našem příspěvku jsme se pokusili ukázat, že na trhu se objevují technologie, které uživatelům pro práci s prostorovými daty nabízejí nová řešení. Jednu z takových možností, zajímavou nejen prakticky nulovými finančními nároky, ale také možností aktivní práce s daty, jsme aplikovali nad daty týkající se aktuální epidemiologické problematiky. V příkladech jsme předvedli několik ukázek, které demonstrují propojení prostorového kontextu a jednoduchých statistických funkcí.

Za silné stránky použitého řešení považujeme možnost ukládání geoprvků různé geometrie (body, linie, polygony) do jedné tabulky, kde jsou uloženy v jednom sloupci. Důležitá je možnost vyhledávat informace o geoprvcích bez toho, že bychom museli vytvářet řadu po sobě jdoucích dotazů nad již vybranými prvky. S pomocí agregovaných funkcí můžeme nad daty provádět jednoduchou statistiku bez exportu do externích programů.

Co je dále podstatné – nad stejnými daty může pracovat více uživatelů současně bez nutnosti databázi replikovat pro každého zvlášť. Každý z uživatelů může samostatně pracovat nad stejnými daty, pouze je nutno vnímat omezení plynoucí například z editačních operací, kdy po zamknutí záznamu je aktuální hodnota dat výhradně známa vlastníkově této operace. To jsou však situace, které má databáze Oracle již vyřešeny transakčním zpracováním.

Představené řešení je vhodné zejména pro ty, kteří jsou zvyklí pracovat s databázemi obsahující velké množství dat. Nepotřebují provádět rozsáhlou manipulaci a analýzu dat, data potřebují jednoduše vizualizovat v prostoru. Příkladem takovýchto řešení mohou být např. data ze senzorických sítí, kdy ze senzorů proudí (ať již kontinuálně či diskrétně) velké množství dat a informací a uživatel je potřebuje zobrazit v prostoru nad mapou.

Slabší stránku funkčnosti použitého řešení lze považovat oblast vizuálních výstupů. Například je nemožné přemístit popisky jednotlivých prvků – pro všechny z dané tabulky platí stejné pravidlo pro umístění (viz obrázek 3). Vnímáme také celkově menší komfort a možnosti, na které jsme zvyklí v GIS prostředí, což je dáno primárním určením databázové technologie – práce s daty ve formě tabulek. Pro laika se může jevit import dat a jeho nastavení jako komplikovaná záležitost, kdy při přenosu jednoduché shapefilové vrstvy je nutno nastavit několik parametrů (souřadnicový systém, klíčové atributy, kontrolu datových typů apod.). Zpracování rastrových podkladů v této variantě není možné, neboť je nabízena pouze v plné verzi databáze Oracle. Pro mnoho uživatelů může být překážkou alespoň základní znalost jazyka SQL.

Nabízené řešení považujeme za vhodné pro uživatele, kteří potřebují pracovat současně s databází i prostorovými daty, nemají možnost investovat do komerčních produktů a jsou ochotni přistoupit na některá omezení – například velikost databáze do 4GB, nemožnost pracovat plnohodnotně s rastrovými daty a smíří se s jednodušší vizualizací výstupů.

LITERATURA

- [1] Codd, E. F. (1970) A Relational Model of Data for Large Shared Data Banks. Communications of the ACM, Volume13 (6),1970. 377–387.
- [2] Hönig V., Zubriková D., Vögerl D., Švec P., Pfister K., Grubhoffer L. (2011) Klíšťata a jimi přenášená onemocnění v Jihočeském kraji a Bavorsku/ Zecken und zecken-übertragene Krankheiten in Südböhmen und Bayern. Jihočeská univerzita, České Budějovice. 80 s. ISBN: 878-80-7394-309-7.
- [3] Oracle Sun server:
http://www.xbitlabs.com/news/cpu/display/20111004222532_Oracle_We_Will_Continue_to_Sell_Systems_with_x86_Microprocessors.html [cit. 5. 8. 2012].
- [4] http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=53698 [cit. 5. 8. 2012].