

METODIKA HARMONIZACE, AGREGACE A ANONYMIZACE DAT KRIMINALITY

Jiří HORÁK

Institut geoinformatiky, Hornicko-geologická fakulta, Vysoká škola báňská – Technická univerzita Ostrava,
17. listopadu 15, 708 33 Ostrava, Česká republika
jiri.horak@vsb.cz

Abstrakt

Metodika vznikla v rámci projektu „Geoinformatika jako nástroj pro podporu integrované činnosti bezpečnostních a záchranných složek státu“ a byla certifikována PP PČR v roce 2015. Metodika navrhuje postupy pro úpravu a sjednocení dat, souhrnně označené jako harmonizace dat, pro geokódování, pro agregaci dat a jejich anonymizaci. Tomu odpovídají 3 části metodiky, které jsou ale současně provázané – agregaci dat je potřebné provádět nad harmonizovanými daty a stejně tak anonymizovaná data zpravidla využívají agregaci. Metodika rovněž umožňuje nastavení vhodných mechanismů pro automatizaci některých činností, zejména pro strategické a administrativní analýzy kriminality. Metodika je určena pro všechny, kteří pracují s daty kriminality, připravují je pro analýzy, pro předání jiným subjektům a pro publikaci. Obecné principy (např. harmonizace dat, realizace geografické a časové dimenze) jsou ale uplatnitelné pro kohokoliv, kdo pracuje s časoprostorovými událostmi.

Abstract

Methodology of Harmonisation, Aggregation and Anonymization of Crime Data: The methodology was developed in the frame of the project „Geoinformatics as a tool to support integrated activities of safety and emergency units“ and certified by Police of the Czech Republic in 2015. The methodology proposes processes for data modification and unification (data harmonisation), for geocoding, data aggregation and anonymisation. Three parts of the methodology are linked together – it is necessary to provide data aggregation from harmonised data, and data anonymisation requires data aggregation. The methodology enables setting of suitable mechanics for activity automation especially for strategic and administrative crime analysis. The methodology is intended for everybody dealing with crime data, preparing it for analysis, for passing it to other subjects and for publication. Common principles (i.e. data harmonisation, implementation of geographical and temporal dimensions) are valuable also for other domains where spatiotemporal events are utilized.

Klíčová slova: kriminalita, harmonizace, agregace, GIS

Keywords: crime, harmonisation, aggregation, GIS

1. ÚVOD

Efektivní zpracovávání dat, zajištění srovnatelnosti a jisté objektivnosti výsledků analýz kriminality využívaných např. při přípravě programů prevence kriminality vyžaduje zavedení standardních postupů nejlépe ve formě certifikovaných metodik.

Data kriminality představují evidenci jednotlivých deliktů v informačních systémech bezpečnostních složek, která vytváří základní zdroj dat pro statistické a analytické vyhodnocení kriminality a predikci kriminality. Operativní záznamy nemohou zajistit dokonalé sjednocení a harmonizaci vkládaných údajů. Zavedení přísných kontrol konzistence údajů by zpomalovalo práci uživatelů a komplikovalo plnohodnotný a věrný zápis skutečnosti, kterou není vždy možné dokonale popsat prostřednictvím předem definovaných číselníků.

Pro potřeby statistického a analytického zpracování dat je však nezbytné provést základní systémové kontroly dat, vyjasnit nejednoznačné údaje, vhodně interpretovat jednotlivé situace, sjednotit popis a klasifikaci událostí, vytěžit doprovodné či odvozené informace.

Další běžným krokem při analýze dat je jejich agregace. Stávající agregace podle organizačního členění bezpečnostních složek často nevyhovuje některým požadavkům veřejné správy a rovněž neumožňuje údaje snadno kombinovat s dalšími datovými zdroji z daného území a časového intervalu.

Údaje o stavu, struktuře, intenzitě a vývoji kriminalitě patří k množině dat, o které má zájem řada dalších orgánů veřejné správy i veřejnost. Podmínkou jejich poskytnutí je neohrožení probíhajícího plnění úkolů podle trestního řádu, plnění dalších úkolů na úseku vnitřního pořádku a bezpečnosti, ochrany osobních práv a ochrany citlivých údajů. Metodika proto obsahuje i návrh vhodného postupu anonymizace dat.

Cílem uplatnění metodiky je poskytovat pro analýzy sjednocené a srovnatelné (případně také garantované) údaje, které zajistí objektivitu analýz, srovnatelnost výsledků a jejich správnou interpretaci. Pro publikaci a předávání dat zajistí jednotný postup vytváření sekundárních dat upravených takovým způsobem, aby se zvýšila jejich informační hodnota, snadnost interpretace a současně umožnilo i vytváření ošetřených dat, neobsahující citlivé údaje, které dosud brání širšímu využití dat o kriminalitě.

Metodika navrhuje postupy pro úpravu a sjednocení dat, souhrnně označené jako harmonizace dat, pro geokódování, pro agregaci dat a jejich anonymizaci. Tomu odpovídají 3 části metodiky, které jsou ale současně provázané – agregaci dat je potřebné provádět nad harmonizovanými daty a stejně tak anonymizovaná data zpravidla využívají agregaci. Metodika rovněž umožňuje nastavení vhodných mechanismů pro automatizaci některých činností, zejména pro strategické a administrativní analýzy kriminality.

Metodika vznikla v rámci projektu „Geoinformatika jako nástroj pro podporu integrované činnosti bezpečnostních a záchranných složek státu“ a byla certifikována PP PČR v roce 2015 (Horák, 2015).

Metodika je určena pro všechny, kteří pracují s daty kriminality, připravují je pro analýzy, pro předání jiným subjektům a pro publikaci. Obecné principy (např. harmonizace dat, realizace geografické a časové dimenze) jsou ale uplatnitelné pro kohokoliv, kdo pracuje s časoprostorovými událostmi.

2. HARMONIZACE DAT

Harmonizací dat se rozumí sjednocení dat, odstranění různých typických situací nekonzistence a vnitřních rozporů v datech.

Při zpracování se musí dodržovat princip nedotknutelnosti primárních zápisů. Harmonizovaná (tj. upravená) data se proto zapisují do nových atributů.

Cílem harmonizace je dosažení co nejkvalitnějšího výchozího stavu dat pro jejich další zpracování, hodnocení a interpretaci. Důležitou podmínkou harmonizace je také sjednocení používaných dat a číselníků při integraci dat, tj. spojování a společné zpracování dat buď z různých zdrojů nebo z téhož zdroje s různou časovou platností.

2.1 Kontroly

Postup začíná zásadami tvorby číselníků (např. povinné vymezení období platnosti) a kontrolami dostupnosti a sémantiky číselníků. Následuje ověření referenční integrity (Pokorný, Halaška 2003, Teorey et al. 2006). Z praktických důvodů jsou kontroly rozděleny na obecně platné (standardizované) číselníky a specifické tematické číselníky.

Ověření referenční integrity u základních standardizovaných územních číselníků kontroluje zejména způsob zápisu adres. Zápis adresy sjednocuje a standardizuje Vyhláška č. 359/2011 Sb., o základním registru územní identifikace, adres a nemovitostí. Popis variant formátování zápisu adresy včetně příslušných údajů je v příloze certifikované metodiky „Využití databáze RÚIAN pomocí webových služeb“ (Augustýn et al., 2013).

Standardizace adres se doporučuje do úrovně částí obcí, další části adresy (resp. popisu lokalizace) je vhodné ponechat bez standardizace. Nižší adresní pole (název ulice a veřejného prostranství, domovní čísla atd.) se v řadě případů používají pro zápis lokalizace mimo adresní místa v závislosti na struktuře evidence.

V případě využití **tematických standardizovaných číselníků** se provádí kontrola zadaných hodnot vůči obsahu těchto číselníků. Přitom se využívá zejména číselníků a klasifikací zavedených sdělením ve Sbírce zákonů.

V případně ostatních **nestandardizovaných číselníků** se provádí kontrola obsahu proti aktuální verzi číselníku. Týká se to např. klasifikace trestných činů a přestupků, identifikace útvarů bezpečnostních složek u např. evidence útvarů, který přijal oznámení nebo provádí šetření. V případě kontroly útvarů se provádí rovněž kontrola, zda událost řeší pouze útvar správného typu. Na základě rozlišení jednotlivých typů útvarů by delikt neměly řešit útvary typu kancelář, auditor apod.

Ke klasifikaci deliktů se využívá číselník takticko-statistické klasifikace (TSK), který je rozšířen o podrobný popis přestupků evidovaných zejména obecními policiemi. Správcem číselníku TSK je PČR.

Příkladem dalšího tematického číselníku je označení dálnic, silnic a místních komunikací, které je doporučeno provádět podle vyhlášky č. 104/1997 Sb., §2.

Dále se provádí **kontrola konzistence časových údajů**. Ta zahrnuje kontrolu případů evidovaných mimo adekvátní časový rozsah. Zjednodušeně řečeno, jde o kontrolu záznamů, které vykazují čas událostí v budoucnosti nebo v příliš vzdálené minulosti. K tomu slouží automatizovaná kontrola, že uložená kalendářní data jsou v odpovídajícím časovém rozsahu. Za standardní časový rozsah je považován interval (aktuální datum - 6 měsíců) až (aktuální datum). Pro vybrané trestné činy a přestupky, které jsou uvedeny v příloze metodiky, se používá odlišný interval (až 5 let).

Dále se provádí automatizovaná kontrola návaznosti sledovaných časů.

Významnou je také **kontrola konzistence souřadnic**. Tato kontrola v současnosti zahrnuje kontrolu nekonzistentních souřadnic v ČR, zejména uváděných v S-JTSK. Provádí se automatizovaná kontrola, zda se zadané souřadnice nachází v ČR. Využívá se selekce záznamů, jejichž souřadnice leží mimo území ČR. Kontrola se provádí pomocí rozsahu území ČR (EXTENT) nebo mimo polygon ČR (kontrola pomocí WITHIN). Analogicky lze provádět kontrolu polohy na rozsah (extent) či polygon sledovaného území.

Konkrétní limitní hodnoty souřadnic v systému S-JTSK pro ČR jsou v tab. 1, limitní zeměpisné souřadnice (používané např. při měření GPS) v tab. 2. Limity jsou proti jiným autorům (např. Bečvář 2015) zúžené na pouze potřebný rozsah. V případě souřadnic S-JTSK/05 jsou hodnoty o 0,5 mil. vyšší, tj. např. pro minimum X bude 5905000 a maximum 6228000.

Tab. 1 Rozsah souřadnic S-JTSK pro území ČR

	X v S-JTSK	Y v S-JTSK
Minimum pro ČR	905 000	431 000
Maximum pro ČR	1 228 000	935 000

Tab. 2 Rozsah zeměpisných souřadnic pro území ČR

	Zeměpisná šířka (LAT)	Zeměpisná délka (LON)
Minimum pro ČR	48,55°	12,09°
Maximum pro ČR	51,06°	18,85°

Jednoduchým kontrolním kritériem je rovněž selekce záznamů, pro jejichž souřadnice S-JTSK platí, že souřadnice X je menší nebo rovna souřadnici Y. Jde o účinnou a jednoduchou kontrolou, která odhaluje zejména případy přehození X a Y souřadnic.

Další je kontrola situací možného rozporu mezi polohou a typem deliktu. Provádí se kontrola místa deliktu a detekují se 3 možné problémové situace:

- a) Delikt je lokalizován na služebně PČR a typ lokalizace je místo spáchání nebo místo následku, ale klasifikace deliktu tomu neodpovídá (např. porušení domovní svobody, vloupání do rodinného domu apod.). Výběr se provede na základě vzdálenosti místa deliktů a polohy služebny, která je menší než 10 m.
- b) Delikt, který pravděpodobně vzhledem k typu události nemá přesnou lokalizaci. Např. místo deliktu je daleko od známého adresního bodu a jde o t.č. krádeže vloupáním kódy D.01 až D.17 dle TSK nebo Krádeže věcí v objektech (bez vloupání) (E.12 dle TSK).
- c) Delikt, který se stal na veřejné komunikaci (silnice, dálnice, železniční těleso apod.), ale souřadnice odpovídají adresnímu bodu. Nejdříve je nutné vybrat záznamy deliktů podle identifikace objektu veřejné komunikace a z nich vybrat záznamy, kde je vzdálenost místa spáchání deliktu od libovolného adresního bodu menší než 5 m.

Takto vybrané záznamy je nutné dodatečně geokódovat, protože je pravděpodobné, že nemají správně zadané souřadnice.

Po provedení nezbytných kontrol a příslušných oprav se přistupuje k dalšímu kroku harmonizace dat a tím je doplnění dat.

2.2 Doplnění dat

První z nich je **výpočet referenčního času spáchání deliktu**. Důvodem je skutečnost, že zpravidla se pracuje s 1 datem události, i když v informačních systémech je v případě kvalitnější evidence uváděn časový interval. Pro časoprostorové analýzy se však běžně předpokládá 1 datum události.

Situaci výpočtu referenčního času spáchání deliktu z intervalu komplikují 2 okolnosti:

- udávaný interval je příliš dlouhý pro jednoznačné určení referenčního kalendářního data či hodiny spáchání deliktu (např. delší než 1.5 dne pro určení referenčního dne). Často je to spojeno s problémem neurčitosti v určení jednoho či obou konců intervalu (např. pro interval 1.1.2015-31.12.2015 je ve většině případů nesmyslné uvádět střední datum v červenci).
- Některé případy (zde typy trestné či přestupkové činnosti) jsou svou podstatou dlouhodobé a nemá smysl uvádět 1 datum či hodinu (typicky zanedbávání povinné výživy, omezování a zbavení osobní svobody, nedovolené pěstování rostlin obsahující omamnou látku (§ 285), podpora a propagace hnutí směřujícího k potlačení práv a svobod občanů apod.). Seznam kategorií deliktů, u kterých nelze určit hodinu spáchání, je v tab. 3.

Tyto případy jednotně řeší metodika.

V případě nejednoznačného určení konkrétního dne je možné alespoň určit měsíc, případně rok, a k zápisu použít textový datový typ (např. RRRRMM00, kde 00 ukazuje neurčený konkrétní den). Konkrétní výpočty pro určení jednotlivých případů jsou uvedeny v metodice.

Tab. 3 Seznam deliktů, u kterých nelze určit hodinu spáchání

Kód TSK	Název skupiny deliktů
G	Ostatní majetková trestná činnost
H	Trestné činy proti hospodářské soustavě
I	Trestné činy proti hospodářské kázní
J	Trestná činnost proti měně
K	Daňová trestná činnost
M	Porušování nehmotných práv a nekalá soutěž
P.02	Zanedbání povinné výživy

V dalším kroku se provádí částečně automatizované a částečně manuální doplnění identifikace objektů, ke kterým se vztahuje místo deliktu, a rovněž událostí na trasách. Zápis se provádí do nových atributů TYPOBJ, OBJEKT, PODOBJEKT, OVEROBJ, ODKUD, KAM, LINKA, SPOJ, KILOMETRAZ.

Dále se doporučuje provést doplnění klasifikace deliktů. Při zpracování se jeden delikt klasifikuje více kódy typu deliktu. Pro zajištění lepších výběrů a sjednocení popisu zpracování se provádí částečně automatizované a částečně manuální doplnění klasifikace deliktů.

K potřebným krokům patří rovněž harmonizace zápisu adresních atributů. Harmonizace je určena k úpravě zápisu adresy tak, aby odpovídající nové atributy následně mohly lépe sloužit pro geokódování, výběry deliktů podle územního členění či jejich klasifikaci. Provádí se automatizované naplnění nových atributů HNAZOB, HNAZCOB, HNAZULICE, HCOR a HCP, případně i automatizované doplnění kódů obcí, částí obcí, ulic pro zajištění jedinečnosti identifikace. V zásadě se provádí vypuštění textových řetězců, které se nemají v příslušném poli vyskytovat, a uplatnění oprav známých chyb, evidovaných ve speciálním seznamu.

V případě chybějících či nedůvěryhodných souřadnic se provádí geokódování dat. Možnosti a výsledky geokódování uvádí např. (Horák et al., 2015), Havlík (2010).

Geokódování míst se přednostně provádí na základě harmonizovaných adresních atributů. Současně se zapisuje přesnost geokódování (kvalita). Zjištěné souřadnice se zapisují do nových polí GJTSTKX a GJTSTKY, přesnost do pole GKVALITA. Je vhodné rovněž evidovat v samostatné struktuře podrobný zápis průběhu geokódování, např. datum a čas zpracování, zápis požadavku, vrácená odpověď, typ lokalizace, typ adresy.

K automatizovanému geokódování je možné využít aplikace GeoCoder (VŠB-TU Ostrava 2015 v rámci projektu GISBS) nebo API rozhraní vybraných mapových serverů.

Riziko chybějících geokódovaných událostí spočívá zejména v možnosti vzniku prostorové systematické chyby (Andersen, Malleson 2013), pokud existuje problém s lokalizacemi jistého typu či v jistém území. Podle Ratcliffe (2004 in Andersen, Malleson 2013) je minimální hranicí úspěšnosti geokódování, která by ještě neměla způsobovat prostorovou systematickou chybu, 85%.

3. AGREGACE DAT

Agregace dat kriminality je využívána pro tvorbu policejních statistik, pro analytické účely, hodnocení vývoje, trendů atd. Je také základem některých technik predikce.

Cílem agregace je především potlačení vlivu náhodných efektů u dat s vysokým rozlišením, které by znesnadnily analýzy a interpretace, snadnější identifikace trendů a vazeb (zejména závislostí, asociací) (Harries, 1999, Chainey, Ratcliffe 2013).

Agregace je rovněž významným nástrojem anonymizace dat.

Pokud se vyskytuje více míst deliktu, provádí se agregace nezávisle na každé místo deliktu.

Agregace dat kriminality je prováděna podle základních kritérií:

- Území
- Čas
- Klasifikace podstaty deliktu
- Charakteristika delikventa a okolností deliktu

Agregace je rovněž základem k definování OLAP (Teorey et al. 2006, Loshin 2012, Horák, Horáková 2007).

Specifickou technikou prostorové agregace je výpočet jádrového vyhlazení.

3.1 Geografická agregace

Agregaci je potřebné provádět trojím způsobem:

- Podle geometrického vymezení
- Podle administrativního uspořádání územně správních jednotek
- Podle organizačního členění bezpečnostních útvarů

Některé geometricky vymezené jednotky mohou spadnout do více územních jednotek. Důvodem může být:

- Poloha na hranici 2 či více jednotek
- Velikost buňky, která obsahuje více jednotek
- Čas – zatímco jednotky geometrického vymezení jsou stálé, administrativní či organizační jednotky mohou podléhat změnám.

Doporučená nejnižší agregační jednotka **podle geometrického vymezení** je buňka 100x100m, orientovaná podle os souřadnic S-JTSK (hranice buněk dle souřadnic S-JTSK v celých stovkách metrů). Vyššími agregačními jednotkami jsou buňky 500x500 m, 1 x 1 km a 5 x 5 km.

Backer (2014) uvádí 6 typů měřítek pro grid. Pro 4. úroveň (Communes and urbanized districts) doporučuje grid 100 x 100 m, pro 3. úroveň pro národní regiony doporučuje 1 km x1 km. I když jsou v literatuře doporučovány větší agregační jednotky (150m, 200 m, 250 m v systému PRECOBS (Hruška et al., 2015) či CriPa (Steiner, Glasner, 2015), je vhodné základní nastavit na 100m z důvodu skladebnosti. Pokud by např. byla použita základní buňka 150 m, nelze z takto agregovaných dat vytvořit následující agregace v kilometrové síti.

Výpočet identifikátoru základní stometrové čtvercové buňky se provádí následovně:

$$\text{Fix}([\text{SJTSK_Y}]/100)*(100000)+\text{Fix}([\text{SJTSK_X}]/100)$$

V budoucnu pro 3D buňky se počítá s nejmenší buňkou 100m x 100m x 1 patro.

Agregace podle administrativního uspořádání musí zajistit skladebnost vyšších územních jednotek.

K identifikaci územních jednotek se používají číselníky ČSÚ (viz kap. 3.1.2).

Nejnižší agregační jednotka **podle administrativního uspořádání** je část obce. Vyššími agregačními jednotkami jsou obec, obec s rozšířenou působností, okres, kraj a stát.

Další územní členění je **podle organizačního členění** bezpečnostních složek. Tyto jednotky jsou nezávislé na administrativním členění území, resp. nemusí ho zcela kopírovat. Kromě toho bezpečnostní složky řeší i případy, jejichž místo je mimo jejich obvod působnosti.

V případě PČR je nejnižší agregační jednotka podle organizačního členění útvar PČR (OOP). Vyššími agregačními jednotkami jsou Okres (Územní odbor), Kraj (krajské ředitelství) a ČR.

Základní agregační jednotkou u obecní policie je okrsek. Vyšší jednotky jsou označovány podle konkrétní organizace činnosti obecní policie, nejvyšší pak celá územní působnost příslušné obecní policie.

Další možností provádění prostorové agregace dat je jádrové vyhlazení.

Jádrové vyhlazení představuje specifickou techniku, kterou lze využít pro dosažení agregace dat v území, i když její metodické principy jsou odlišné od běžných agregací (Horák et al. 2010, Inspektor et al. 2014, Carlos et al., 2010, Levine 2007). Popis metodiky uplatnění jádrového vyhlazení je předmětem certifikované metodiky „Metodika identifikace anomálních lokalit kriminality pomocí jádrových odhadů“ (Ivan et al., 2015).

3.2 Agregace podle času

Jako součást popisu deliktů se eviduje datum a čas výskytu události (např. spáchání přestupku). Podobně významnou součástí popisu událostí je datum a čas např. pro záznamy IZS. Tyto informace umožňují vytváření časových řad, na jejichž základě lze hodnotit časový vývoj kriminality či například provádět predikce kriminality.

Z praktického hlediska je vhodné oddělit od sebe datum události, které lze použít pro časové hodnocení kriminality, a čas události. Kombinace kalendářního data a určení času události není vhodná, protože by vznikla nedostatečná agregace.

Doporučená nejnižší agregační jednotka podle kalendářního data spáchání deliktu je 1 den. Vyššími agregačními jednotkami jsou měsíc, kvartál a rok. Při agregaci se využívá i nepřesně určených dat, které jsou uloženy jako RRRRMM00 nebo RRRR0000. Pro agregaci na úrovni dne se časově nepřesně určené

události evidují jako „datum nezařazeno“, pro agregaci na úrovni měsíce se časově nepřesně určené události evidují jako „měsíc nezařazeno“.

3.3 Agregace podle klasifikace podstaty deliktu

Ke klasifikaci deliktů se využívá číselník takticko-statistické klasifikace (TSK), který je rozšířen o podrobný popis přestupků evidovaných zejména obecními policiemi.

Při agregaci se postupuje tak, aby v dané úrovni klasifikace byl delikt zaznamenán nanejvýš 1x v příslušném kódu klasifikace. Tedy každý delikt je zaznamenán maximálně 1x v každé třídě klasifikace, maximálně 1x v každé nadtřídě klasifikace a maximálně 1x v každé kategorii klasifikace. Tím je dokumentován celý rozsah skutkové podstaty delikty a současně se uměle nezvyšuje počet deliktů.

4. ANONYMIZACE DAT

Cílem anonymizace je umožnit poskytování dat o kriminalitě v takové formě, aby byla zajištěna ochrana osobních (případně firemních) údajů tak, aby nedošlo k ohrožení základních lidských práv dle Listiny základních práv a svobod, zejména (ale nejen) nedotknutelnost osoby, soukromí a lidské důstojnosti, a její majetková práva.

Anonymizace dat může být prováděna kombinací následujících postupů:

- Agregace dat – především územní agregace
- Selekcce pro nadlimitní počet případů
- Náhodné roztřesení
- Náhodné přičítání a odečítání
- Projekce na vybranou sadu atributů

Při volbě metody či jejich kombinace se musí především vzít v úvahu typ (klasifikace) deliktů, jejich četnost a charakter území, ve kterém má být anonymizace provedena.

Součástí anonymizace dat je rovněž omezení předávaných atributů (zpravidla klasifikace deliktu, referenční datum a čas).

1) V **bodové formě s přesnou polohou** místa spáchání lze zveřejnit nebo předat následující události (s vybranými atributy):

Ukončené trestné činy dle TSK:

- v kategorii S (Dopravní nehody) bez vnitřního rozlišení.

Ukončené přestupky dle TSK:

- v kategorii PR01 (Na úseku BESIP a silničního hospodářství) všechny,
- v kategorii PR02 (Na úseku ochrany před alkoholismem a toxikomanií) všechny,
- v kategorii PR03 (Proti občanskému soužití) pouze třída PR030101 (Proti veřejnému pořádku dle § 47) a PR030201 (Proti veřejnému pořádku dle § 48).

U některých typů deliktů nemusí být anonymizace prováděna a mohou být poskytnuty s přesnou polohou. Jde zejména o delikty, kde z polohy deliktu nelze žádným způsobem usuzovat na bydliště pachatele či oběti deliktu, zejména v kombinaci s provedenou agregací podle charakteristiky deliktu na obecnou skupinu, ze které nelze usuzovat na konkrétní typ činnosti.

2) V **bodové formě s náhodně roztřesenou polohou** místa spáchání lze zveřejnit nebo předat následující události (s vybranými atributy):

Ukončené trestné činy dle TSK:

- v kategorii T (Toxikománie) v rozlišení na nadtřídy (zejména odlišení T.05, která se vztahuje k alkoholu).
- Přestupky v jiné kategorii než výše uvedené.

Náhodné roztřesení může být definováno jako přidání náhodně vygenerované (za předpokladu rovnoměrné distribuce) hodnoty z intervalu $<-100\text{ m}, +100\text{ m}>$ k souřadnici X a přidání náhodně vygenerované (za předpokladu rovnoměrné distribuce) hodnoty z intervalu $<-100\text{ m}, +100\text{ m}>$ k souřadnici Y.

Je nutné upozornit, že ponechání bodové reprezentace je pro řadu čtenářů mapy zavádějící, protože i přes důkladná upozornění mají tendenci spojovat delikt právě s místem zobrazení a následně provádět mylné interpretace.

3) Ve **formě linie** lze zveřejnit nebo předat následující události (s vybranými atributy):

Ukončené trestné činy dle TSK:

- Krádeže na osobách - v dopravním prostředku (E.0707).

4) Ve **formě polygonu** (územní jednotka, buňka) lze zveřejnit nebo předat počty všech typů deliktů, pokud polygon zahrnuje více než 10 adresních bodů. Adresní body se přebírají z aktuálního RÚIAN.

Selekce pro nadlimitní počet případů je jednoduchou a spolehlivou metodou, která zajišťuje kontrolovanou úroveň agregace tak, aby nedošlo ke ztotožnění se skutečnými nositeli. Stanoví se limitní počet deliktů, při kterém je již spolehlivě zajištěna nemožnost identifikace skutečného nositele údaje. Následně se vyberou pouze ty územní jednotky, které mají vyšší než limitní počet deliktů uvnitř jednotky.

5. ZÁVĚR

Metodika je určena pro všechny, kteří pracují s daty kriminality, připravují je pro analýzy, pro předání jiným subjektům a pro publikaci. Metodika slouží pro zajištění jednotného postupu předzpracování dat kriminality a jejich interpretace, kontrole zapsaných dat, vytváření harmonizovaných údajů pro analytické potřeby, zejména hodnocení stavu a vývoje kriminality a návrhy preventivních opatření apod.

Je uplatnitelná jak pro data Policie ČR, tak pro další bezpečnostní složky, které evidují případy porušení zákonů (delikty), tj. jak pro trestné činy, tak i přestupky.

Metodika je připravena tak, aby byl možný co nejvyšší rozsah automatizace jednotlivých úkonů a zajistilo se vyloučení subjektivního faktoru při kontrole a zpracování dat. Lze ji tedy uplatnit při realizaci integritních omezení na úrovni databází a přípravě programových kódů pro zpracování dat.

Metodika má 3 základní části (harmonizace dat, agregace dat a anonymizace dat), kde nezbytnou prerekvizitou pro realizaci agregace dat a/nebo anonymizace dat je harmonizace dat, kterou není možné z procesu vypustit.

PODĚKOVÁNÍ

Metodika byla vytvořena v rámci projektu VF20142015034 „Geoinformatika jako nástroj pro podporu integrované činnosti bezpečnostních a záchranných složek státu“. Metodika byla certifikována Policejním prezidiem PČR.

LITERATURA

Andresen M.A., Malleson N. 2013: Spatial heterogeneity in crime analysis. In Leitner M. (Ed.) Crime Modeling and Mapping Using Geospatial Technologies. Geotechnologie and Environment 8, DOI 10.1007/978-94-007-4997-9_1. Springer, Dordrecht.

Augustýn et al. (2013): Využití databáze RÚIAN pomocí webových služeb. Certifikovaná metodika ČUZK.

Backer L. (2014): EFGS and the integration of Geography and Statistics. Proceedings of EFGS 2014 Krakow.

Bečvář L. (2015): Program UNITRANS – transformace souřadnic. verze 11.02. Uživatelská příručka. http://www.archeologickyatlas.cz/downloads/Unitrans_manual.pdf

Carlos, H.A., Shi, X., Sargent, J., Tanski, S., Berke, E.M. (2010): Density estimation and adaptive bandwidths: A primer for public health practitioners. *International Journal of Health Geographics*, 9:39.

Harries, K., *Mapping Crime: Principles and Practice*. Washington, DC: The National Institute of Justice, 1999. 206 s

Havlík M. (2010): Průvodce geokódováním zdravotnických dat databáze EPIDAT. BP UP Olomouc. http://www.geoinformatics.upol.cz/dprace/bakalarske/havlik10/images/havlik_bakalarska_prace.pdf

Horák J.: *Metodika harmonizace, agregace a anonymizace dat kriminality*. Certifikovaná metodika. PP PČR 2015.

Horák J., Horáková B. (2007): Datové sklady a využití datové struktury typu hvězda pro prostorová data. In *Sborník mezinárodního symposia GIS Ostrava 2007*. Ostrava 28.1.-31.1.2007. 26 stran. ISSN 1213-2454.

Horák J., Inspektor T., Čaha J., Kukuliač P.: Geokódování objektů podle adresy. In *sborník „GIS Ostrava 2015 - Současné výzvy geoinformatiky“*, Ostrava, 26-28.1.2015. 11 stran. http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_2015/sbornik/papers/gis2015541c4ad4ce8fc.pdf

Horák J., Ivan I., Inspektor T., Hruška-Tvrđý L. (2010): Monitoring of Socially Excluded Localities of Ostrava City. In *Sborník mezinárodního symposia GIS Ostrava 2010*. Ostrava 24.1.-27.1.2010. 14 stran. ISBN 978-80-248-2171-9.

Hruška L., Foldynová I., Šotkovský I., Zapletalová L., Václavík T., Žurovec I., Fújak R., Ševčík J. (2015): *Mapy budoucnosti - moderní nástroj ke zvýšení efektivity a kvality veřejné správy v oblasti prevence kriminality založený na analýze a predikci kriminality*. Ostrava, 2015. s. 317.

Chainey, S., Ratcliffe, J., (2013): *GIS and Crime Mapping*. *Mastering GIS: Technol, Applications & Mgmt.* John Wiley & Sons, 448 p. ISBN 9781118685198.

Inspektor, T., Ivan, I., Horák, J. (2014): Mapping and Monitoring Unemployment Hot Spots towards Identification of Socially Excluded Localities: case study of Ostrava. *Journal of Maps*, vol. 10, No. 1, 02 Jan 2014. P. 35-46. ISSN 1744-5647. <http://dx.doi.org/10.1080/17445647.2013.847806>. DOI: 10.1080/17445647.2013.847806.

Ivan I., Horák J.: *Metodika identifikace anomálních lokalit kriminality pomocí jádrových odhadů*. Certifikovaná metodika. MV ČR 2015.

Levine, N. (2007): *CrimeStat: A spatial statistics program for the analysis of crime incident locations (v3.1)*. N Levine & Associates, Houston, TX, and the National Inst. of Justice, Washington, DC.

Loshin D. (2012): *Business Intelligence: The Savvy Managers Guide*. MORGAN Kaufmann, Newnes, USA. Pp.370. ISBN 978-0-12-385889-4.

Pokorný J., Halaška I. (2003): *Databázové systémy*. Vysokoškolská skripta ČVUT, Praha 2003.

Steiner, F., Glasner, P. (2015). *Zkušenosti v oblasti mapování, analýz a predikce kriminality u policie v Rakousku*. *Sborník příspěvků – Mapy budoucnosti*, Praha, 2015. http://www.prevencekriminality.cz/evt_file.php?file=838

Teorey T., Lightstone S., Nadeau T. (2006): *Database modelling and design: Logical design*. Elsevier. ISBN 978-0-12-685352-0

Vyhláška č. 359/2011 Sb., o základním registru územní identifikace, adres a nemovitostí.