

**METODIKA TVORBY MULTIDIMENZIONÁLNÍ DATABÁZE PRO PREVENCI 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 MV ČR v roce 2015. Metodika prakticky uplatňuje principy Business Intelligence pro oblast kriminality. Zabývá se vytvářením sjednocené multidimenzionální databáze sekundárních agregovaných dat, využitím specifických analytických a reportovacích nástrojů a zajištěním přenosu anonymizovaných agregovaných údajů z různých informačních systémů veřejné správy pro podporu rozhodování v oblasti prevence kriminality. Obecné principy jsou uplatnitelné pro kohokoliv, kdo má zájem vybudovat multidimenzionální databáze socioekonomických dat, zejména ve veřejné správě.

**Abstract**

**Methodology of Creation of a Multidimensional Database for Crime Prevention:** 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 Ministry of Interior in 2015. The methodology practically applied principles of Business Intelligence for crime domain. It deals with design and creation of unified multidimensional database of secondary aggregated data, with utilisation of specific analytical and reporting tools and implementation of transfer of the anonymized aggregated data from various information systems of the public sector for decision support in the crime prevention. Common principles can be adapted for everybody interested to build multidimensional databases of socioeconomic data namely those in the public sector.

**Klíčová slova:** kriminalita, multidimenzionální databáze, OLAP, business intelligence, GIS

**Keywords:** crime, multidimensional database, OLAP, business intelligence, GIS

**1. ÚVOD**

Prevence kriminality se snaží působit na různorodé příčiny kriminality (Strategie 2012), proto musí využívat široké datové základny, aby dokázala monitorovat a adresovat různé problémy spojené s rizikem výskytu trestných činů.

Objem dostupných dat v socioekonomické sféře prudce narůstá. Kromě standardních dat šetření ČSÚ vč. SLDB je stále rostoucí výzvou využití dalších zdrojů dat v informačních systémech veřejné správy. Data vznikající operativní činností, často evidenční a kontrolní činností jednotlivých orgánů, mohou být efektivně využita pro řadu jiných účelů, zejména přípravy různých analýz a strategií.

Jednou z možností, jak zužitkovat informační hodnotu operativních dat představuje koncept multidimenzionálního datového skladu a OLAP technologií, které pracují s konsolidovanými a agregovanými daty.

Metodika (Horák, 2015b) popisuje logický koncept multidimenzionální databáze, popisuje její strukturu, definuje jednotlivé dimenze sledování reality, definuje vhodnou granularitu dat, popisuje jejich hierarchické vazby. Popisuje rovněž vhodné zdroje dat a postupy transformace dat do multidimenzionální databáze.

Metodika je určena pro všechny subjekty, které se zabývají prevencí kriminality a chtějí při analýzách stavu kriminality využít koncept multidimenzionální databáze, která umožní harmonizované a agregované údaje o kriminalitě spojovat a analyzovat s vybranou sadou především socioekonomických dat (ze stejného území) a

usnadnit tak včasné odhalení stavu a změn lokálních environmentálních faktorů, které mohou mít vliv na vývoj kriminality a/nebo přímo identifikaci významných změn ve vývoji kriminality (z hlediska intenzity i struktury) s cílem včasného přijímání cílených opatření na snížení kriminality.

Cílem metodiky je standardizace tvorby multidimenzionálních databází pro prevenci kriminality, zajištění jejich srovnatelnosti, interoperability, usnadnění kombinace dat v analýzách (vč. integrace statistických dat PČR a městské a obecní policie), možnost snadného rozvoje takové poznatkové základny, tvorba dlouhých a standardizovaných časových řad, zlepšení objektivnosti posuzování stavu a vývoje kriminality mezi jednotlivými územími a termíny, vytvoření standardní databáze pro snazší tvorbu analytických či reportovacích aplikačních nadstaveb, zajištění snadné prostorové, časové a tematické skladebnosti údajů vznikajících v různých prostředích, zajištění rychlého přenosu anonymizovaných agregovaných údajů z různých informačních systémů veřejné správy pro zlepšení informovanosti o aktuálním vývoji socioekonomické situace v daném území a posílení podpory rozhodování v oblasti prevence kriminality.

Metodika navazuje na Metodiku harmonizace, agregace a anonymizace dat kriminality (Horák, 2015a).

## 2. STRUČNÉ PŘÍPOMENUTÍ PRINCIPŮ MULTIDIMENZIONÁLNÍ DATABÁZE A OLAP

OLAP je často chápán za součást Business Intelligence. Za základní funkce technologií Business Intelligence jsou považovány (Badard et al. 2012) - reporting, dashboarding, scoreboarding, OLAP, data mining, proces mining, analytika řízení obchodní výkonnosti (business performance management analytics) a prediktivní analytika (predictive analytics). OLAP je prezentován jako kategorie DSS nástrojů, která zajišťuje intuitivní a efektivní přístup do datového skladu (data warehouse) nebo datového tržiště (datamart).

Základem řešení OLAP je multidimenzionální datová struktura datového skladu. V multidimenzionální databázi se ukládají konsolidovaná, agregovaná data.

Lacko (2003) vymezuje v rámci multidimenzionální kostky dvě složky: fakta a dimenze. Dimeze jsou logicky nebo hierarchicky uspořádané údaje. Gála et. al (2006) popisují, že dimenzi lze vnímat jako číselník. Horák (2013) uvádí, že se jedná o aspekty, dle nichž jsou organizována fakta a podle kterých se provádí analýza agregovaných dat.

Fakta jsou potom jednotlivé, zpravidla numerické a opakovaně zjišťované hodnoty (Horák, 2013). Tabulka faktů je potom středobodem každé OLAP databáze, s níž jsou propojeny dimenze, které poskytují tabulce faktů interpretaci. U faktů je důležitá jejich granularita (Horák, 2013). Vysoká granularita znamená uložení dat na nízkém stupni agregace a tedy velice detailní. Vysoká granularita je dat je vhodná zejména proto, že generalizovat data lze vždy, ale opačný postup možný není. Vytvořit data o nízké granularitě tedy problém není, ale přidat data do OLAP kostky již problémem být může.

Business Intelligence (BI) ve spojení s OLAP, SOLAP (spatial OLAP) a role ETL v BI jsou vysvětleny např. v OGC White Paper (Badard et al. 2012). Principy datových skladů, exekutivní informační systémy, OLAP (on-line analytical processing) jsou vysvětleny např. v Dohnal, Pour (1997), Pokorný (2004), Pirkli (2004), Loshin (2012), Teorey et al. (2006) nebo KHEOPS (2005).

## 3. OBSAH METODIKY

Základem metodiky je návrh struktury multidimenzionální databáze společně se základními procesy jejího budování a využití pro analýzy.

Struktura byla navržena na základě rozsáhlé analýzy potenciálně významných zdrojů dat (část interních, řada externích). Umožňuje sjednotit pohledy na data, data vhodně integrovat, agregovat a připravit pro analýzy a reporting.

Pro výstavbu multidimenzionální databáze se doporučuje dodržovat zejména následující principy:

- a) datový sklad by měl být fyzicky i logicky oddělen od provozních systémů. Data z provozních systémů se převádějí do datového skladu, kde se po transformaci ukládají způsobem, který vyhovuje analytickému a prezentačnímu zpracování výstupů.

- b) Fyzická realizace multidimenzionální databáze může být provedena buď relačním databázovým modelem s odpovídající multidimenzionální strukturou (systémy ROLAP) nebo prostřednictvím speciální multidimenzionální databáze, které multidimenzionální struktury podporují nativně (místo uložení dat v relačních tabulkách aplikují zpravidla vícerozměrná pole, známá z programovacích technik) a mají k dispozici specializovaný multidimenzionální SŘBD (Pokorný 2004).
- c) Data by měla být ukládána s vysokou granularitou, tj. v nízkém stupni agregace, protože se předpokládá jejich využití na místní úrovni.
- d) Fakta by měla být co nejvíce aditivní, aby je bylo možné sumarizovat podle více dimenzí.
- e) V dimenzionálních tabulkách se doporučuje upřednostnit implicitní hierarchii, tj. zařazení potřebných atributů přímo v tabulce dimenze, která není normalizovaná.

Jádrem řešení multidimenzionální datové struktury je 6 tabulek faktů (ve schématech označeny s předponou FTAB), určených pro evidenci odlišných typů událostí a objektů v území a čase. Tato sestava je rozšiřitelná, současně pro většinu funkcí není nutné, aby byla realizována v celém navrženém rozsahu.

Dvě dimenze jsou pro všechna data společná - jedná se o geografickou dimenzi a dimenzi časovou. Budou se vyskytovat u všech dat a jejich existence je přirozeně předpokládána. U konkrétních tabulek faktů se pak tyto dimenze vyskytují ve více variantách současně.

U většiny dimenzí nelze vyloučit, že budou existovat záznamy událostí, pro které není určena odpovídající nejnižší jednotka použitá v dané dimenzi (např. nejmenší územní jednotka, nejmenší časový interval, nejpodrobnější klasifikace deliktu apod.). Proto u nich musí existovat jednotka „nezařazeno“, která umožní ukládat (agregovat) údaje s neurčenou základní jednotkou. Kromě toho struktury by měly umožnit i ukládání údajů známých pouze za hierarchicky vyšší jednotky (např. není určena část obce, ale pouze obec; trestný čin byl páchan v období vymezeném časovým intervalem, který překračuje základní použitou časovou jednotku; není známá podrobná klasifikace deliktu, ale pouze jeho zařazení do nadřídky či kategorie), aby bylo maximálně využito dostupné informace a příslušné delikty byly započítány při agregaci na vyšší jednotky.

Součástí přípravy dat pro multidimenzionální databáze je potřebné předzpracování dat. To zahrnuje kontroly dat, jejich harmonizaci a doplnění. Vzhledem k principiálnímu významu geografické dimenze je nezbytné zajistit kvalitní geokódování dat, aby získaná data byla přiřazena ke správným územním celkům a to i při velkém detailu (malých územních celcích). Postupy vhodného předzpracování dat popisuje Metodika harmonizace, agregace a anonymizace dat kriminality (Horák, 2015a).

Pro potřeby OLAP doporučujeme nerozlišovat úroveň rozpracovanosti trestního řízení u osob a za pachatele označovat i osoby nejen odsouzené, ale již i podezřelé ze spáchání deliktu (a samozřejmě všechny další stupně). Analogicky se postupuje v případě přestupků. Je zřejmé, že toto opatření povede ke zvýšení množství informací o osobách, které jsou původci deliktů, ale na druhou stranu může výsledky analýzy negativně ovlivnit, protože účast osob na deliktu není prokázána.

Základní principy agregace dat jsou vysvětleny v Metodice harmonizace, agregace a anonymizace dat kriminality (Horák, 2015a).

V metodice jsou podrobně popsány kroky přípravy jednotlivých tabulek faktů a dimenzionálních tabulek. Část z nich je dále popsána.

### 3.1 Příprava tabulky faktů o obyvatelích

Tabulka FTAB\_OBYVATELSTVO eviduje počet obyvatel v závislosti na vybraných charakteristikách. Tabulka umožňuje sledovat vývoj základních demografických charakteristik v čase a prostoru a současně poskytuje datovou základnu pro výpočty indikátorů, které mají charakter intenzity jevů přepočtených na stejný počet obyvatel. Takovým je např. index kriminality (Zapletal, Novotný 2004).

Lokalizace obyvatelstva vychází z evidence trvalého bydliště. Pro některé účely by byla vhodná okamžitá lokalizace obyvatel, ke které však chybí datové zdroje. Geografická agregace je zajištěna prostřednictvím identifikace jednotek podle geometrického vymezení a podle administrativního uspořádání územně

správních jednotek. V souladu s dalšími typy evidence se doporučuje pro geometrické vymezení ukládat identifikátor buňky 100x100m (ID\_100m), orientované podle os souřadnic S-JTSK, který je identifikátorem v dimenzionální tabulce DT\_CTVEREC. Identifikace čtverce vychází z polohy jeho aktuálního bydliště. Pro každou osobu se předpokládá max. 1 bydliště (v daný čas). Podle administrativního uspořádání se ukládá identifikátor části obce v příslušné verzi, který je identifikátorem v dimenzionální tabulce DT\_UZEM\_JEDNOTKY.

Stejně jako v případě pachatelů a obětí se používá kategorizace věku do pětiletých skupin s některými výjimkami (např. 15-17 let) a většinou spojených kategorií (např. nad 30 let). Příslušnost k věkové kategorii (ID\_VEK) se eviduje pouze 1x.

K rozlišení počtu obyvatel podle pohlaví slouží identifikátor pohlaví osoby (ID\_POHLAVI), který navazuje na dimenzionální tabulku DT\_POHLAVI\_PO.

Dalšími charakteristikami osob jsou vzdělání, národnost, rodinný stav a ekonomická aktivita. Ve všech případech se eviduje pouze 1 klasifikace v dané charakteristice, aktuální a nejvyšší platná.

Základním zdrojem dat je v ideálním případě Registr obyvatel (využívají se atributy věk, pohlaví, rodinný stav). Další charakteristiky obyvatel jako vzdělání, národnost či ekonomická aktivita jsou v současnosti dostupné pouze ze SLDB, je však možné, že ČSÚ bude v budoucnu schopen poskytnout takové údaje z jiného zdroje dat.

**Tab. 1** Struktura tabulky FTAB\_OBYVATELSTVO

Název	Datový typ	Integritní omezení
ID_JEDNOTKA	Char(10)	FK
ID_100M	Char(10)	FK
ID_DATUM	Char(8)	FK
ID_POHLAVI	Char(10)	FK
ID_VEKOVE_KATEGORIE	Char(10)	FK
ID_VZDELANI	Char(10)	FK
ID_NAROD	Char(3)	FK
ID_STAV	Char(1)	FK
ID_EKON_A	Char(2)	FK
P_OBYVATEL	Integer	NN

Pokud by to zdroje dat umožnily, bylo by vhodné evidovat také počet přistěhovalých, vystěhovalých a zahraničních imigrantů, počet dojíždějících, případně další významné charakteristiky ve vztahu k obyvatelstvu.

### 3.2 Příprava tabulky faktů o objektech

Je zřejmé, že výskyt některých typů kriminality souvisí s výskytem jistého typu objektů v území. Např. velká obchodní střediska přitahují drobnou majetkovou kriminalitu, noční kluby a bary mohou souviset s různorodou trestnou a přestupkovou činností. Vývoj počtu objektů jistého typu může tedy ovlivňovat vývoj počtu páchaných deliktů či jejich další charakteristiky jako je objasněnost, charakteristiky obětí či pachatelů atd.

Jednoduchým způsobem lze stav a vývoj počtu zájmových objektů v území sledovat prostřednictvím k tomu určené tabulky faktů. Pak je možné v každé územní jednotce či čtverci zjistit např.:

- Počet heren, resp. počet výherních automatů
- Počet zastaváren
- Počet sběren druhotných surovin
- Počet obchodů
- Počet restauračních podniků

- Počet zastávek veřejné dopravy (s rozlišením)
- Počet čerpacích stanic
- Počet charit, azylových domů, nocleháren, zařízení pro matky s dětmi apod.
- Počet ubytoven
- Počet mateřských škol
- Počet základních škol
- Počet středních škol
- Počet firem

Poměrně jednoduchá struktura tabulky faktů FTAB\_OBJEKTY zahrnuje 2 geografické dimenze, dimenzi kalendářního data a dimenzi typu objektu a samozřejmě této kombinaci odpovídající počet objektů.

Klasifikace typů objektu je řešena prostřednictvím identifikátoru ID\_TYPOBJ, který má vazbu na dimenzionální tabulku (DT\_TYPOBJEKTU). V případě, že jeden fyzický objekt slouží více účelům, je možné uvést více odpovídajících identifikátorů typů objektu.

Doporučení ke zdrojům objektů vybraných typů:

- Čerpací stanice – veřejně dostupný seznam MPO (<http://www.mpo.cz/dokument153307.html>), aktualizace 1x ročně v září, nebo získávání seznamů přímo z MPO na základě smluvní spolupráce.
- Zastávky veřejné dopravy – využití centrální databáze zastávek VLD a MHD vč. jejich souřadnic společnosti CHAPS s.r.o. po zajištění smluvní spolupráce.
- Herny – veřejně dostupný seznam všech technických herních zařízení v České republice, kterým vydalo Ministerstvo financí povolení (<http://www.mfcr.cz/cs/soukromy-sektor/regulace/loterie-a-sazkove-hry/prehled-povolenych-zarizeni>). Aktualizace 1x měsíčně.
- Základní, střední a vyšší odborné školy - veřejně dostupný seznam na <http://stistko.uiv.cz/register/vybskolrn.asp>.
- Vysoké školy - veřejně dostupný seznam na <http://stistko.uiv.cz/proavs/provsass.asp>
- Zdravotnická zařízení – seznam poskytovatelů zdravotních služeb s kontaktními údaji z Národního registru poskytovatelů zdravotních služeb poskytuje ÚZIS (Koordinační středisko rezortních zdravotnických informačních systémů) prostřednictvím aplikace na <https://eregpublicsecure.ksrzis.cz/Registr/NRPZS/ZdravotnickeZarizeni> nebo poskytování seznamu ÚZIS na základě smluvního ujednání. Aktualizace zdroje průběžná, vzhledem k využití pro různé účely (mj. krizové řízení).
- Služby s živnostenským oprávněním (obchodní centra, obchody, restaurační zařízení, ubytovny, sběrný druhotných surovin) - automatizované zpracování podnikatelských subjektů MPO, které provozuje na adrese [http://www.rzp.cz/cgi-bin/aps\\_cacheWEB.sh?VSS\\_SERV=ZVWSBJFND](http://www.rzp.cz/cgi-bin/aps_cacheWEB.sh?VSS_SERV=ZVWSBJFND) službu vyhledání podnikatelského subjektu nebo poskytování seznamu provozovatelů vybraných živností ze strany MPO.

Pokud by bezpečnostní složky využívaly pro svou činnost databázi referenčních objektů, bude možné z ní pravidelně plnit potřebnou tabulku faktů k objektům.

**Tab. 2** Struktura tabulky FTAB\_OBJEKTY

Název	Datový typ	Integritní omezení
ID_JEDNOTKA	Char(10)	FK
ID_100M	Char(10)	FK
ID_TYPOBJ	Char(10)	FK
ID_DATUM	Char(8)	FK
P_OBJEKTY	Integer	NN

### 3.3 Příprava tabulky faktů o zdraví

Tvorbu sociálního prostředí také ovlivňuje výskyt sociálně patologických jevů a zdravotních okolností. Současně také takové agregované informace mohou být důležité pro další úkoly bezpečnostních složek v souvislosti se zajištěním pořádku, ochranou obyvatelstva před zdravotními riziky apod.

Pomocí této struktury bude možné určit v každé územní jednotce či čtverci např.:

- Počet léčených uživatelů drog
- Počet nových léčených uživatelů drog
- Počet vydaných injekčních setů v L/K centrech
- Počet sebraných injekčních stříkaček nebo jehel v terénu
- Počet případů vybraných infekčních nemocí (např. virová hep. B a C, syfilis, svrab)
- Počet hospitalizovaných ve zdravotnických zařízeních dle bydliště dle vybrané klasifikace (zejména na odděleních AT – návykové nemoci, příp. infekční choroby apod.)
- Počet zemřelých dle vybrané klasifikace (např. známky týrání, sebevraždy, otrava narkotiky, otrava jinými léčivými, náhodná otrava alkoholem a expozice)

Zdrojem dat mohou být informační systémy resortu zdravotnictví, spravované ÚZIS nebo Ministerstvem zdravotnictví.

Datová struktura tabulky faktů FTAB\_ZDRAVOTNI zahrnuje 2 geografické dimenze a dimenzi kalendářního data, které jsou řešeny stejným způsobem jako u tabulky faktů o obyvatelích.

V souvislosti s vazbou na obyvatelstvo jsou použity 2 charakteristiky, které mohou pomoci lépe vymezit podskupinu obyvatel, kterých se registrovaný problém týká – věk a pohlaví. Věk a pohlaví používají stejné dimenze jako v případě tabulky faktů o obyvatelích. Pro evidenci platí stejná doporučení.

Novými specifickými tematickými dimenzemi jsou evidence vybraných lidských chorob, kde identifikátor choroby se váže na dimenzionální tabulku DT\_ZDRAV\_OBYV, a evidence typu případu, která rozlišuje nové případy od stávajících (incidence versus prevalence), sebrané versus vydané injekční sety či jehly, hospitalizované, zemřelé, u kterých se vyskytla identifikovaná choroba. Zde je identifikátor napojen na dimenzionální tabulku TYP\_PRIPADU. V případě evidence chorob se evidují všechny zájmové choroby (tj. počet případů může být vyšší než počet nemocných osob).

Pro výpočty intenzit se použije poměr s počty obyvatel (případně příslušných specifických kategorií zejména dle věku a pohlaví) evidovaných v tabulce faktů o obyvatelích.

**Tab. 3** Struktura tabulky FTAB\_ZDRAVOTNI

Název	Datový typ	Integritní omezení
ID_JEDNOTKA	Char(10)	FK
ID_100M	Char(10)	FK
ID_DATUM	Char(8)	FK
ID_TYP_ZDRAV	Char(10)	FK
ID_TYP_PRIPAD	Char(10)	FK
ID_VEKOVE_KATEGORIE	Char(10)	FK
ID_POHLAVI	Char(10)	FK
P_OBYVATEL	Integer	NN

### 3.4 Příprava tabulky faktů o nezaměstnanosti

Významný vliv na zejména majetkovou kriminalitu může mít majetková situace obyvatel, jejíž zhoršení může nastat v souvislosti s nezaměstnaností. Celkově však nezaměstnanost může vytvářet podmínky či ovlivňovat i jiné druhy deliktů a je významnou součástí tvorby sociálního prostředí.

Informace o situaci v nezaměstnanosti zprostředkovává tabulka faktů FTAB\_NEZAMEST. Tvorba tabulky vyžaduje spolupráci s resortem MPSV ČR, resp. s Úřadem práce.

Pomocí této struktury bude možné určit v každé územní jednotce či čtverci:

- Počet nezaměstnaných
- Počet nezaměstnaných s max. základním vzděláním (VABC)
- Počet dlouhodobě nezaměstnaných (délka evidence 12 měsíců a více)
- Počet nezaměstnaných vyžadujících primárně CZ-ISCO9 (nekvalifikované práce)

**Tab. 4** Struktura tabulky FTAB\_NEZAMEST

Název	Datový typ	Integritní omezení
ID_JEDNOTKA	Char(10)	FK
ID_100M	Char(10)	FK
ID_DATUM	Char(8)	FK
ID_VZDELANI	Char(10)	FK
ID_DELKA_EVID	Char(10)	FK
ID_CZISCO	Char(10)	FK
ID_POHLAVI	Char(10)	FK
ID_VEKOVE_KATEGORIE	Char(10)	FK
P_NEZAM	Integer	NN

Datová struktura tabulky faktů zahrnuje 2 geografické dimenze a dimenzi kalendářního data, které jsou řešeny stejným způsobem, jako u tabulky faktů o obyvatelích či zdraví. Předpokládá se agregace dat v pravidelném časovém intervalu 1 měsíce, protože vyšší frekvence by nepřinesla vyšší přidanou hodnotu. V souvislosti s vazbou na obyvatelstvo jsou použity 2 charakteristiky, které mohou pomoci lépe vymezit podskupinu obyvatel, kterých se registrovaný problém týká – věk a pohlaví. Věk a pohlaví používají stejné dimenze jako v případě tabulky faktů o obyvatelích. Pro evidenci platí stejná doporučení.

Výsledná tabulka faktů o nezaměstnanosti tak udává počet nezaměstnaných podle výše uvedených kategorií.

Vedle nezaměstnanosti by bylo vhodné v jiné tabulce faktů evidovat také situaci ve vyplácení nepojistných sociálních dávek. Taková evidence by umožnila mít informace např. o počtu osob pobírajících dávky hmotné nouze, počtu osob pobírajících příspěvek na bydlení nebo počtu osob pobírajících obecně sociální dávky.

### 3.5 Příprava tabulky faktů o bydlení

Sociální charakteristiku území a potenciálně problémová místa (neobydlené prostory) může pomoci indikovat tabulka faktů FTAB\_BYDLENÍ k vybraným charakteristikám bydlení. Soustřeďuje se zejména na znaky problémového bydlení jako je přítomnost neobydlených prostor či bytů, větší stáří obydlí, akumulace osob či domácností v bytě apod.

Pomocí této struktury bude možné určit v každé územní jednotce či čtverci:

- Počet bytů v neobydlených domech
- Počet neobydlených domů s byty
- Průměrné stáří obydlí
- Počet osob v bytech celkem s právním důvodem užívání: nájemní
- Počet neobydlených bytů
- Počet neobydlených bytů v domech s důvodem neobydlenosti nezpůsobilé k bydlení
- Počet obydlí v ostatních budovách se sníženou kvalitou

**Tab. 5** Struktura tabulky FTAB\_BYDLENI

Název	Datový typ	Integritní omezení
ID_JEDNOTKA	Char(10)	FK
ID_100M	Char(10)	FK
ID_DATUM	Char(8)	FK
ID_BYDLENI	Char(10)	FK
ID_OBYDLENOST	Char(10)	FK
ID_VEKOVE_KATEGORIE	Char(10)	FK
ID_PRAV_FORMA	Char(10)	FK
ID_TYP_UZIVANI	Char(10)	FK
P_OBYDLI	Integer	NN

Vedle standardní geografické a časové dimenze jsou využity dimenze věku (sdílený číselník DT\_VEK, který musí nabízet dostatečný věkový rozsah jak pro obyvatele, tak pro stáří bytů a domů) a specifické dimenze jen pro tuto tabulku faktů, kam patří rozlišení bytových a rodinných domů, rozlišení neobydlených objektů podle důvodu, právní forma bydlení a rozlišení typu užívání.

Zdrojem dat pro naplnění tabulky faktů je SLDB. Vedle použití SLDB je možné využít evidence domů a částečně bytů v RÚIAN, která poskytuje mnohem aktuálnější, ale pouze dílčí údaje. K výběru stavebních objektů z RÚIAN, které mohou být předmětem zájmu a agregace, se jeví jako nejvhodnější využití číselníku Způsob využití, kde za hlavní objekty zájmu považujeme kódy 1 až 15, 18, případně i 20. Naopak číselník „Způsob vytápění“ či „Druh svíslé nosné konstrukce“ nepovažujeme za vhodné k implementaci rozlišení budov a jejich výběru. Jinou alternativou je využití RSO (ČSÚ), který obsahuje i evidenci budov, a k dispozici jsou klasifikace budov dle atributu ZPVYBU (způsob využití budovy dle ISKN, atribut 0076), TVYBU (typ využití budovy, atribut 0075), KSD\_CZ\_CC (klasifikace stavebních děl, číselník 80008 a 80026, atribut 5522) a počet bytů (SUM\_BYT).

### 3.6 Příprava dimenzionálních tabulek

V metodice je popsána specifikace a příprava jednotlivých dimenzionálních tabulek. Pro příklad je uvedeno několik zajímavostí.

Zajímavá může být třeba hierarchizace času, která umožňuje vytvářet snadno agregace hodinového rozpětí, které označuje např. špičku či večerní hodiny.

Důležitá je také klasifikace objektů a podobjektů, která rozlišuje základní typy objektů a pak typické prostory přidružené či včleněné do objektů (parkoviště, šatna, garáž, schodiště apod.).

V dimenzi věku se používají základní kategorie po 5 letech, ale rovněž jakékoliv další, i neuzavřené věkové intervaly. To umožňuje korektně pracovat i s nepřesným určením věku.

## 4. ZÁVĚR

Metodika je určena pro všechny, kteří pracují s daty kriminality a chtějí je využívat s uplatněním principů OLAP a Business Intelligence, vytvářet sjednocené multidimenzionální databáze sekundárních agregovaných dat, usnadnit provádění strategických a administrativních analýz kriminality a návrhy preventivních opatření, umožnit využití specifických analytických a reportovacích nástrojů a zajistit rychlý přenos anonymizovaných agregovaných údajů z různých informačních systémů veřejné správy pro podporu rozhodování v oblasti prevence kriminality.

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. Obecné principy jsou ale uplatnitelné pro kohokoliv, kdo má zájem vybudovat multidimenzionální databáze socioekonomických dat, zejména ve veřejné správě.

Metodika popisuje přípravu tabulek faktů i jednotlivých dimenzionálních tabulek. Poskytuje podrobný popis potřebných struktur a poskytuje návody pro jejich tvorbu i z hlediska procesního, vč. definice základních integritních omezení.

Navrhuje se rovněž agregovat údaje o objektech a jevech z řady zdrojů, které mohou sloužit k vytváření vhodné univerzální údajové základny pro další socioekonomické analýzy.

## PODĚKOVÁNÍ

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

## LITERATURA

Badard T., Kadillak M., Percivall G., Ramage S., Reed C., Sanderson M., Singh R., Sharma J., Vaillancourt L. (Editors) (2012): Geospatial Business Intelligence (GeoBI). OGC White paper. Ref. Number OGC 09-044r3

Dohnal J., Pour J.: Architektury informačních systémů v průmyslových a obchodních aplikacích. Ekopress 1997. ISBN 80-86119-02-5.

Horák J. (2015a): Metodika harmonizace, agregace a anonymizace dat kriminality. Certifikovaná metodika. Ostrava, 2015.

Horák J. (2015b): Metodika tvorby multidimenzionální databáze pro prevenci kriminality. Certifikovaná metodika. MV ČR 2015.

KHEOPS (2005): JMAP Spatial OLAP – On-Line Analytical Processing for Spatial Databases – Innovative technology to support intuitive and interactive exploration and analysis of spatio-temporal multidimensional data, KHEOPS Technologies, 2005, [http://www.oifii.org/do-org/dod/geosoa-scg-ulaval-ca\\_kheops-tech-com\\_jmap-solap\\_WP\\_JMap\\_SOLAP.pdf](http://www.oifii.org/do-org/dod/geosoa-scg-ulaval-ca_kheops-tech-com_jmap-solap_WP_JMap_SOLAP.pdf)

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

Pirkl D. Tvorba datových skladů – pohled zevnitř. Konference Datakon. Brno 2004. [http://www.datakon.cz/datakon04/d04\\_it\\_pirkl.pdf](http://www.datakon.cz/datakon04/d04_it_pirkl.pdf)

Pokorný J. (2004) Konstrukce databázových systémů. Vysokoškolská skripta ČVUT, Praha 2004.

Strategie (2012): Strategie prevence kriminality ČR 2012-2015. MV ČR. 33 stran.

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.

Zapletal J., Novotný O. (2004): Kriminologie. ASPI Publishing, 2. přeprac. vydání, 451 s., ISBN 80-7357-026-2