

Zdroje údajov a zdravotnícke informačné systémy

Ciele kapitoly

Zahraničné a domáce zdroje

Štatistické štandardy

Domáca rutinná štatistika

Medzinárodné zdroje štatistických údajov

Súhrn kapitoly

Literatúra

Ciele kapitoly

V tejto kapitole chceme predstaviť zdroje údajov, z ktorých sme čerpali v príkladoch v knihe alebo ktoré sú častým zdrojom dát pri spoznávaní zdravia verejnosti. Zastavme sa na chvíľu pri výraze *zdroj údajov*. Čo sa vlastne za tým skrýva? V knihe o biomedicínskej štatistike [1] uvádzame zdroje údajov pre výskum a sústreďujeme sa najmä na tie, ktoré sa získavajú experimentálne, meraním. V tejto knihe sú hlavným zdrojom čísla, ktoré už boli spracované a následne boli publikované. Teda nebudeme sa zaoberať detailmi vzniku údajov, ich získaním či postupom zberu, hoci sa tomu nedá v niektorých prípadoch vyhnúť, ale zameriame sa na tie zdroje, kde je možné získať súbory dát.

Zamýšľame predstaviť dostupné zdroje údajov o demografii, o ochoreniach, o zdrojoch pre zdravie verejnosti z domácich i zahraničných inštitúcií. Pokúsime sa uviesť najdôležitejšie z tých, ktoré pôsobia v tejto oblasti a ktorých štatistické údaje používame v príkladoch uvedených v nasledujúcich kapitolách.

Tam, kde je to potrebné, uvádzame aj spôsoby prístupu k poskytovaným údajom. Ukážeme príklady použitia databázových systémov, tabuľkových editorov a najmä funkcií prostredia {R}. Nie je, prirodzene, v našich silách ukázať všetky zdroje, museli sme ich vyberať zo širokej palety. Zohľadnili sme viaceré kritériá. Z domácich zdrojov sme sa zamerali na oficiálne štatistiky pochádzajúce najmä zo *Štatistického úradu Slovenskej republiky*, a tiež z publikácií *Národného centra zdravotníckych informácií*. Z medzinárodných sme vybrali v prvom rade tie, kam Slovenská republika

pravidelne prispieva, napr. medzinárodné organizácie združujúce štáty sveta, akou je *Svetová zdravotnícka organizácia*. Tiež sme bližšie predstavili tie inštitúcie, ktoré poskytujú údaje o regionálnych združeniach, ktorých je Slovenská republika členom, ako sú zdroje Európskej únie, reprezentované najmä organizáciou *Eurostat*. Uvádzame aj niektoré dohovory a štandardy, ktorými sa zber, spracovanie či distribúcia údajov na národnej či medzinárodnej úrovni riadi, napríklad ISO norma pre narábanie s údajmi o čase.

Jedným z cieľov tejto kapitoly je poukázať na zdroje, priniesť ich základnú mapu a ozrejmiť princípy prístupu k údajom. V nasledujúcich kapitolách, kde budeme s tými údajmi narábať, už potom nebude potrebné sa k týmto základom vracáť.

Zahraničné a domáce zdroje

Oficiálne štatistiky

Základným prameňom sú *oficiálne štatistiky*. Tento výraz sa používa pre dáta zozbierané a rozširované množinou národných alebo medzinárodných agentúr. [2] Dáta sa zbierajú s cieľom poskytovať kvantitatívne podklady pre rozhodovanie o stratégiách, politikách a podpore výskumu. To je aj témou tejto publikácie. Za výrazom agentúry sa skrývajú rôzne inštitúcie. V niektorých krajinách tieto funkcie plnia národné štatistické úrady, na Slovensku je ním Štatistický úrad Slovenskej republiky. Inde je to súbor vládnych organizácií, z ktorých každá zbiera a spracúva určitú časť národnej štatistiky. Ekonomické údaje spracúva centrálna banka. Okrem toho k štatistikám prispievajú aj jednotlivé rezorty (zdravotníctvo, školstvo, poľnohospodárstvo, sociálne veci a iné) dátami, ktoré zhromažďujú pre svoju potrebu. Národné organizácie majú obvykle povinnosť (väčšinou danú zákonom) poskytovať dohodnuté údaje medzinárodným organizáciám. Tieto sú buď na regionálnej úrovni, ako Eurostat, ktorý sa zaoberá najmä členskými krajinami Európskej únie. Na globálnej úrovni pôsobia *Spojené národy* (United Nations), ktoré zbierajú a poskytujú štatistiky z celého sveta. Okrem UN globálne pôsobia aj špecializované medzinárodné inštitúcie, napríklad pre oblasť zdravia je to *Svetová zdravotnícka organizácia* (World Health Organization), pre oblasť potravinárstva a výživy je ňou *Organizácia pre potraviny a poľnohospodárstvo* (Food and Agriculture Organization) a *Svetová banka* (World Bank) pre financie a ekonomiku.

Spôsob činnosti

Všetky spomínané organizácie si vymieňajú údaje, na ich základe spracúvajú správy a rozširujú ich. Preto ak navštívite stránky Svetovej zdravotníckej organizácie na internete alebo priamo jej úradovne na Slovensku, alebo niektorú z centrálnych úradovní, všade nájdete publikácie, ktoré približujú určitý aspekt záujmu SZO.

Rovnako to platí aj o ďalších medzinárodných organizáciách, a v neposlednom rade aj o inštitúciách, ktoré sa zaoberajú štatistikou na národnej úrovni. Môžeme zovšeobecniť, že národné a medzinárodné zdroje so štatistickými údajmi na národnej úrovni sú všeobecne dostupné či už z internetu alebo v tlačenej forme. Minimálne raz za rok sa obnovujú vo forme súhrnných štatistických prehľadov. Na nižších úrovniach, v prípade Slovenskej republiky teda na úrovni kraja či okresu alebo mesta, táto situácia až taká jednoduchá nie je. Získať niektoré z údajov je často možné až dopytovaním sa zodpovedných úradov.

Štatistické štandardy

Národná ani medzinárodná štatistika sa nezaobíde bez dohôd na nadnárodnej úrovni o spôsobe výmeny štatistických údajov. Zdravotníkom dobre známa Medzinárodná klasifikácia chorôb [3] je príkladom šandardizácie príčin smrti. Zaoberali sme sa ňou a aj niektorými ďalšími klasifikáciami v samostatnej kapitole.

ISO 8601

Tu si však bližšie všimneme štandard, ktorý patrí medzi takzvané ISO normy a ktorý pod číslom 8601 špecifikuje spôsob narábania s niektorými zo štatistických elementov a ich formátov. [4] V úvode sa píše, že cieľom tohto medzinárodného štandardu je eliminovať riziko nesprávnej interpretácie a predísť neistotám a ich dôsledkom. Táto norma obsahuje číselnú reprezentáciu informácií, ktoré sa týkajú času, ako sú dátumy, týždne, časové intervaly, kombinácie dátumu a času, a samotné informácie o čase v priebehu dňa. Tieto špecifikácie sa týkajú výhradne gregoriánskeho kalendára a 24-hodinového dňa. Prehľad položiek, ktoré sú určené touto normou, je v tabuľke 1.

-
- Kalendárne dátumy, vyjadrené výrazmi kalendárneho roku, kalendárneho mesiaca a kalendárneho dňa v mesiaci;
 - Ordinálne dátumy vyjadrené výrazmi ako kalendárny rok a kalendárny deň v roku;
 - Dátumy v týždni vyjadrené výrazmi kalendárneho roku, čísla kalendárneho týždňa a kalendárny deň v týždni;
 - Miestny čas, založený na 24-hodinovom systéme sledovania času;
 - Coordinated Universal Time (koordinovaný univerzálny čas) v dni;
 - Miestny čas a rozdiel voči Coordinated Universal Time;
 - Kombinácia dátumu a času v dni;
 - Časové intervaly;
 - Opakované časové intervaly.
-

Tabuľka 1 Prehľad položiek, ktoré určuje norma ISO 8601 [4]

Ako príklad normy si uvedieme vyjadrenie kalendárneho dátumu. Kalendárny rok je predstavený štyrmi číslicami. Kalendárne roky sú číslované vo vzostupnom poradí podľa gregoriánskeho kalendára podľa hodnôt v rozsahu (0000) do (9999). Hodnoty v rozsahu (0000) po (1582) by sa mali používať na základe vzájomného dohovoru partnerov pri výmene informácií. Kalendárny mesiac je predstavený dvoma číslicami. Január je predstavený ako (01) a nasledujúce kalendárne mesiace sú číslované vzostupne. Kalendárny deň v mesiaci je reprezentovaný dvoma číslicami. Prvý kalendárny deň ľubovoľného kalendárneho mesiaca je predstavený ako (01) a nasledujúce kalendárne dni toho istého kalendárneho mesiaca sú číslované vzostupnou postupnosťou.

Ak je potrebné kompletne reprezentovať kalendárny deň, malo by to byť v jednom z numerických výrazov, kde (YYYY) predstavuje kalendárny rok, (MM) poradové číslo kalendárneho mesiaca v kalendárnom roku a (DD) poradové číslo kalendárneho dňa v kalendárnom mesiaci.

Základný formát:	YYYYMMDD
Príklad:	19850412
Rozšírený formát:	YYYY-MM-DD
Príklad:	1985-04-12

Ďalšie podrobnosti si nájde záujemca v samotnej norme, nebudeme sa preto zaoberať jej úplným znením, je voľne prístupná v anglickom jazyku. Upozorňujeme na ňu preto, lebo sa stále viac využíva pri komunikácii údajov a je právne záväzná pre všetky krajiny Európskej únie. Zároveň je názorným príkladom štandardizácie, ktorá sa týka aj štatistiky.

IS verejnej správy

Ministerstvo dopravy, pôšt a telekomunikácií SR, Sekcia informatizácie spoločnosti, Odbor pre informačnú bezpečnosť a štandardy vydalo *Metodický pokyn k štandardom pre informačné systémy verejnej správy a ich rozdeleniu a náležitostiam, spôsobu ich aktualizácie a oblastiam ich uplatnenia pre informačné systémy verejnej správy*. [5] Metodický pokyn dopĺňa rozdelenie a náležitosti štandardov, spôsob ich aktualizácie a oblasti ich uplatnenia pre informačné systémy verejnej správy, ako nástrojov výkonu informačných činností pri utváraní a prevádzkovaní informačných systémov verejnej správy alebo ich častí. Medzi oblasti štandardizácie, ktoré upravuje tento pokyn, patria technické štandardy, štandardy prístupnosti, elektronická výmena dát, štandardy pre názvoslovie elektronických služieb, bezpečnostné štandardy. Tento zoznam nie je uzavretý a do ďalších výnosov sa pripravujú oblasti, ako napríklad dátové štandardy, štandardy pre priestorovú identifikáciu atď. Odporúčanou požiadavkou pri výmene štatistických dát a metadát je používanie štandardu *Statistical Data and Metadata Exchange, verzia 2.0 (SDMX 2.0)*. [2]

SDMX

Spoločným úsilím a s podporou Bank for International Settlements (BIS), European Central Bank (ECB), Eurostat, International Monetary Fund (IMF), OECD, OSN, a Svetovej banky sa tieto zamerali na vývoj efektívnejších postupov pre výmenu a poskytovanie dát a metadát v rámci ich pôsobnosti. Spoločne vytvorili SDMX [2], za čím sa skrývajú technické a štatistické štandardy a pokyny, spolu s IT architektúrou a IT nástrojmi, ktoré majú byť použité na účinnú výmenu a poskytovanie štatistických údajov a metaúdajov. Štandardizované formáty súborov na spracovanie dát a metadát a štandardizovaný obsah týchto súborov sú predpokladom na automatizovanú produkciu, spracovanie a výmenu dát SDMX a metadát súborov medzi národnými a medzinárodnými štatistickými organizáciami. Čo sa týka domén, kde sa tieto štandardy hlavne uplatňujú, radí sa tam okrem financií a bankovníctva aj demografia, vzdelávanie a epidemiológia. Samotné štandardy sú prísne technické a nie je cieľom tejto publikácie ich podrobnejšie rozoberať. Sú totiž určené najmä pre technikov, ktorí vytvárajú informačné systémy.

Formáty pre výmenu údajov

Nárast výmeny údajov medzi rôznymi inštitúciami viedol k potrebe štandardizácie popisu údajov tak, aby ich bolo možno vymieňať medzi rôznymi platformami (napr. databázami alebo tabuľkovými editormi bez rozdielu v pôvode). Jedným z nich je jazyk XML, čo znamená eXtensible Markup Language, v preklade rozšíriteľný značkovací jazyk, ktorý vyvinulo a štandardizovalo konzorcium W3C (World Wide Web Consortium) ako pokračovanie jazyka SGML a zovšeobecnenie jazyka HTML. [6] Umožňuje jednoduché vytváranie konkrétnych jazykov na rôzne účely a široké spektrum rôznych typov údajov. Jazyk je určený predovšetkým na výmenu údajov medzi aplikáciami a na zverejňovanie dokumentov. Umožňuje opísať štruktúru dokumentu z hľadiska vecného obsahu jednotlivých častí a nezaobera sa sám osebe vzhľadom dokumentu alebo jeho časti. Vzhľad sa potom definuje pripojeným štýlom. Ďalšou možnosťou je pomocou rôznych štýlov vykonať transformáciu do iného typu dokumentu alebo do inej štruktúry XML. Microsoft v balíku Microsoft Office 2007 a novších ukladá všetky súbory v tomto formáte a na označenie pridáva písmeno *x* do rozšírenia názvu súboru, ako napríklad *.docx*, *.xlsx* a *.pptx*. Tieto formáty a prípony súborov sa používajú v programoch Microsoft Word, Microsoft Excel a Microsoft PowerPoint.

Častým formátom je textový súbor označovaný ako CSV, teda *Comma-separated values* (hodnoty oddelené čiarkami). [7] Ide o súbor, ktorý obsahuje čistý text (angl. plain text), teda bez formátovacích znakov, aké používajú textové editory či iné programy. Je určený na ukladanie tabuľkových dát. Súbor v tomto formáte sa skladá z ľubovoľného počtu záznamov (riadkov), oddelených znakom nového riadku. Každý záznam obsahuje stĺpce, ktoré sú oddelené iným znakom, prevažne čiarkou (,) alebo

tabulátorom. Zvyčajne majú všetky záznamy rovnaký počet stĺpcov. Tento formát sa bežne používa pri výmene aj väčšieho množstva údajov a týmto spôsobom možno prenášať informáciu napríklad medzi tabuľkovým editorom a databázovým programom alebo prostredím {R}. Prvý riadok zvyčajne obsahuje názvy stĺpcov, ktoré prijímajúci program, napríklad {R} interpretuje ako názvy premenných.

Mnohé organizácie ponúkajú dáta priamo vo formáte určenom pre tabuľkové editory, či niektorý z formátov pre databázy.

Slovníky

Na záver tejto časti by sme chceli upozorniť, že štandardov je viac, ako sme mohli uviesť na tomto mieste. Medzi štandardy môžeme zaradiť aj názvoslovie, ktoré v anglickej verzii zastupujú napríklad slovníky epidemiológie [8, 9]. Prvý z nich bol aj preložený do slovenčiny. [10] V slovenčine vyšiel aj slovník demografie. [11] Veľmi dobrú službu poskytuje EUROSTAT prostredníctvom web stránky *Statistics Explained*. [12] Je to portál s doplnkovými informáciami pre príležitostných a pravidelných používateľov, je tiež encyklopédiou o štatistike Európskej únie a v neposlednom rade aj štatistický slovník. Obsahuje štatistické údaje a vysvetlenia, ale tiež priame odkazy na najnovšie údaje a na všetky druhy príslušných informácií, ktoré môžu byť užitočné pre pochopenie štatistík. Jadrom portálu sú štatistické články. Každý súbor štatistík Eurostatu je popísaný jedným alebo niekoľkými článkami. Typický článok začína uvedením aktuálnosti použitých údajov, pokračuje krátkym úvodom do článku, a jeho podrobným obsahom.

Domáca rutinná štatistika

Zákonné normy

Právnym základom je *Zákon o štátnej štatistike* [13], ktorý upravuje podmienky získavania údajov a informácií potrebných na posudzovanie sociálno-ekonomického vývoja, postavenie a pôsobnosť orgánov, ktoré vykonávajú štátnu štatistiku, úlohy orgánov verejnej moci v oblasti štátnej štatistiky, práva a povinnosti spravodajských jednotiek, ochranu dôverných štatistických údajov pred zneužitím, poskytovanie a zverejňovanie štatistických údajov, zabezpečovanie porovnateľnosti štatistických informácií a plnenie záväzkov vyplývajúcich z medzinárodných zmlúv v oblasti štátnej štatistiky, ktorými je Slovenská republika viazaná. Zákon zároveň vymedzuje základné pojmy.

ŠÚSR

Štatistický úrad SR vykonáva úlohy štátnej štatistiky. Za podmienok a v rozsahu ustanovenom uvedeným zákonom môžu úlohy štátnej štatistiky vykonávať aj ministerstvá a štátne organizácie. Ako zdroj štatistických údajov je potrebné spomenúť

tiež informačné systémy verejnej správy, ktoré zabezpečujú získavanie, poskytovanie a sprístupňovanie údajov, zhromažďovanie, spracúvanie, prenos, ukladanie, archiváciu a likvidáciu údajov v rámci zákona. [14]

Štatistika zdravia

Štatistika v oblasti zdravia zahŕňa dve oblasti – štatistiku zdravotníctva a štatistiku pracovnej neschopnosti pre chorobu a úraz. Údaje v oblasti štatistiky zdravotníctva sa zisťujú najmä v rámci štátnych štatistických zisťovaní s ročnou periodicitou. Tie vykonáva v pôsobnosti Ministerstva zdravotníctva SR Národné centrum zdravotníckych informácií. [15] Ide o údaje z oblasti ambulantnej a lôžkovej starostlivosti, spoločných liečebných a vyšetrovacích zložiek, údaje o sieti zdravotníckych zariadení v SR a o zdravotnom stave. NCZI má k dispozícii aj ďalšie informácie o zdravotnom stave obyvateľstva zo štatistických zisťovaní realizovaných na základe legislatívy Ministerstva zdravotníctva SR. Ďalšími zdrojmi údajov zo zdravotníctva sú administratívne zdroje Úradu verejného zdravotníctva SR, Štátneho ústavu pre kontrolu liečiv, zdravotných poisťovní a Sociálnej poisťovne. Veľa užitočných informácií poskytnú popisy metodiky, ktoré obsahujú základné charakteristiky každého z ukazovateľov, jednak pre oblasť zdravotníctva a pre pracovnú neschopnosť.

SLOVSTAT

Štatistický úrad SR poskytuje údaje prostredníctvom databáz, ktoré sú prístupné na internete. Na účely špeciálnych štatistických analýz je možno požiadať o údaje, ktoré bežne prístupné nie sú. SLOVSTAT [16] je databáza určená pre odbornú i laickú verejnosť. Obsahuje aktuálne informácie a časové rady ukazovateľov sociálno-ekonomického vývoja Slovenskej republiky. Databáza sprístupňuje základné údaje zo všetkých štatistických okruhov v mesačných, štvrtročných alebo ročných časových radoch. Mnohé ukazovatele sú podrobnejšie členené. Údaje sú pravidelne aktualizované a sprístupňované v krátkom časovom intervale po ich oficiálnom zverejnení. Pri revíziách sú upravované celé časové rady údajov. Databáza obsahuje rozsiahle množstvo tabuliek z jednotlivých štatistík, preto prístup sa dá realizovať cez navigačný strom, ktorý sa dynamicky otvára a zatvára kliknutím na hyperlinku tematického a následne štatistického okruhu až po najnižšiu úroveň navigácie so zoznamom tabuliek zvolenej štatistiky. Základnú os navigačného stromu tvoria štatistické témy, ako napríklad demografická a sociálna štatistika, podnikateľské štatistiky alebo viacstranné štatistiky. Každá téma sa ďalej člení cez vetvy na štatistické okruhy, v ktorých sú v častiach navigačného stromu rozvetvené podrobnejšie štatistiky daného okruhu a príslušné časové rady. Názov tabuliek bližšie určuje oblasť zobrazovaných ukazovateľov a časový rad. Vytvárajú sa dynamicky na základe voľby ukazovateľov a obdobia. Údaje sú v nich zobrazené v riadkoch ako časový vývoj hodnôt ukazovateľa, pričom stĺpce určujú obdobie danej hodnoty. K dispozícii je

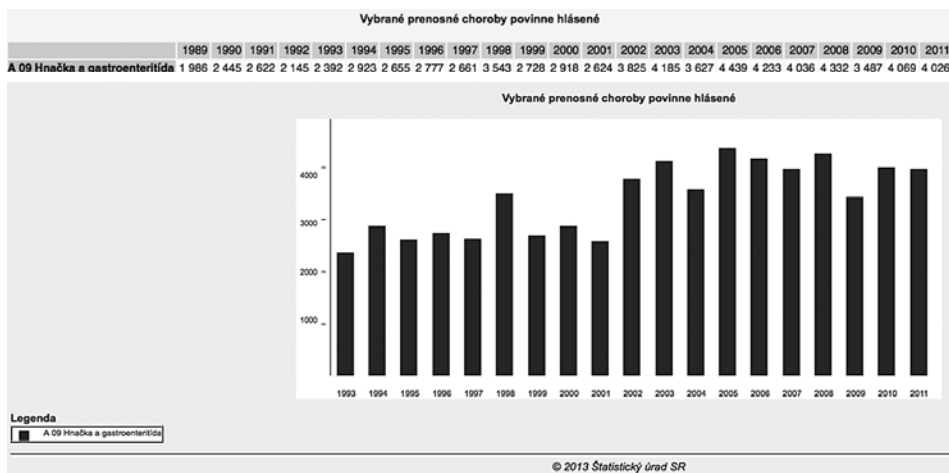
taktiež voľba „textové menu“ (ponuka v hornej lište), ktorá sprístupňuje rozbalený celý navigačný strom s prehľadom poskytovaných štatistík. V aplikácii sú uvedené ako pomôcky pre používateľov skratky a používané značky, návod na prácu a rýchle vyhľadávanie podľa rôznych kritérií.

-
- Štatistická infraštruktúra
 - Demografia a sociálne štatistiky
 - o Demografia a sociálne štatistiky
 - Demografická štatistika, metodika
 - Charakteristiky pohybu obyvateľstva
 - Demografické charakteristiky
 - Štruktúra obyvateľstva podľa veku
 - Štruktúra obyvateľstva podľa rodinného stavu
 - Zahraničná migrácia
 - o Trh práce
 - o Príjmy, výdavky a spotreba domácností
 - o Školstvo
 - o Zdravie
 - Vybrané ukazovatele o zdravotníctve
 - Pracovné miesta a postele v zdravotníckych zariadeniach (2005 – 2010)
 - Lekári v zdravotníckych zariadeniach a postelový fond v zariadeniach ústavnej zdravotnej starostlivosti podľa oddelení k 31. 12. (1995 – 2010)
 - Lekárne, spotreba liekov (1995 – 2011)
 - Vybrané prenosné choroby povinne hlásené (1989 – 2011)
 - Pracovná neschopnosť
 - Finančné ukazovatele v zdravotníctve
 - Historické dáta
 - o Sociálne zabezpečenie
 - Dôchodkové zabezpečenie a nemocenské poistenie
 - Štátne dávky
 - Sociálna pomoc
 - Sociálna ochrana
 - Historické dáta
 - o Kultúra
 - o Kriminalita a požiare
 - Makroekonomické štatistiky
 - Podnikateľské štatistiky
 - Peňažníctvo, financie, zahraničný obchod
 - Poľnohospodárstvo, lesníctvo, rybníctvo
 - Viacstranné štatistiky
-

Tabuľka 2 Štruktúra databázy SLOVSTAT [16]

Štatistika zameraná na problematiku zdravia prirodzene netvorí hlavnú časť ponúkaných štatistík. Obsahuje najmä agregované údaje, obmedzene použiteľné na podrobnejšie štúdium. Údaje o zdravotníckej starostlivosti sa publikujú súhrnne za

rezorty Ministerstva zdravotníctva SR, Ministerstva obrany SR, Ministerstva vnútra SR, Ministerstva spravodlivosti SR, Ministerstva dopravy, pôšt a telekomunikácií SR. Ukazovatele obsahujú krátku charakteristiku. Príklad poskytovaných údajov je na obrázku 1. Vybrané dáta možno exportovať do Excelu alebo vo forme textu oddeleného čiarkou na ďalšie spracovanie, prípadne vytlačiť. Výber premenných je jednoduchý, intuitívny.



Obrázok 1 Údaje o incidencii hnačky a gastroenteritídy v SR od roku 1989 do roku 2011 v tabuľkovej forme a vo forme grafu (Zdroj SLOVSTAT)

NUTS

Štatistický úrad SR poskytuje tiež údaje o regiónoch Slovenskej republiky, ako aj o obciach a mestách. Základom na vymedzenie regiónov je ich klasifikácia nazývaná „Nomenklatúra územných štatistických jednotiek“ („Nomenclature des Unités Territoriales Statistiques“ – NUTS). Zaviedol ju Štatistický úrad Európskej komisie (Eurostat) v spolupráci s národnými inštitútmi pre štatistiku, v prípade Slovenska so Štatistickým úradom Slovenskej republiky. Delenie NUTS nemusí nevyhnutne zodpovedať administratívnejmu deleniu štátu. V SR sú určené úrovne NUTS (ako štatistické jednotky) na základe opatrenia Štatistického úradu SR a jeho akceptácie Eurostatom nasledujúco: NUTS I – Slovenská republika; NUTS II – 4 štatistické územné jednotky: Bratislavský kraj, západné Slovensko (Trnavský, Trenčiansky a Nitriansky kraj), stredné Slovensko (Žilinský a Banskobystrický kraj), východné Slovensko (Prešovský a Košický kraj)/; NUTS III – osem VÚC (samosprávnych krajov); NUTS IV – 79 okresov; NUTS V – 2883 obcí. V minulosti používané úrovne NUTS 4 a NUTS 5 sa premenovali na LAU 1 resp. LAU 2 (z angl. „Local Area Units“) (Tabuľka 3).

-
- Slovenská republika (NUTS 1)
 - Oblasť (NUTS 2)
 - Kraj (NUTS 3)
 - Okres (LAU 1) zodpovedá (NUTS 4)
 - Obec (LAU 2) zodpovedá (NUTS 5)
-

Tabuľka 3 Prehľad štatistických územných jednotiek v SR**RegDat**

Databáza regionálnej štatistiky (RegDat) obsahuje časové rady ukazovateľov hospodárskeho a sociálno-ekonomického vývoja za regióny Slovenskej republiky (okresy, kraje a oblasti podľa dostupnosti údajov). Regionálne údaje z rôznych štatistických okruhov sú vo forme tabuliek v ročných časových radoch. Ich štruktúra je veľmi podobná predchádzajúcej. Prístup k údajom je umožnený databázovým nástrojom PX-WEB. Na ilustráciu možných výstupov sme zvolili výber údajov dát z mesta Bratislava, ktoré je rozdelené do päť okresov. Vybrali sme štatistiku práce a pracovnej neschopnosti pre chorobu a pre pracovný úraz, obe pohlavia spolu a za roky 2001

	Priemerné percento pracovnej neschopnosti		
	2010	2011	2012
Spolu			
Okres Bratislava I			
choroba	1,752	1,647	1,510
pracovné úrazy	0,020	0,010	0,010
Okres Bratislava II			
choroba	2,112	1,964	1,933
pracovné úrazy	0,020	0,017	0,021
Okres Bratislava III			
choroba	1,965	1,821	1,802
pracovné úrazy	0,011	0,023	0,022
Okres Bratislava IV			
choroba	1,714	1,740	1,575
pracovné úrazy	0,023	0,024	0,023
Okres Bratislava V			
choroba	2,151	1,142	2,132
pracovné úrazy	0,022	0,025	0,020

Metodický list. Od roku 2007 sú údaje o pracovnej neschopnosti preberané zo Sociálnej poisťovne. V hodnotách ukazovateľov za SR sú za rok 2007 zahrnuté aj prípady poistencov, ku ktorým nebola zistená územná príslušnosť. Za rok 2007 nie sú ukazovatele v členení podľa pohlavia dostupné. Posledná aktualizácia: 19.8.2013

Obrázok 2 Pracovná neschopnosť v okresoch Bratislavy (Zdroj ŠÚ SR, 2013)

až 2011 (Obrázok 2). Databáza umožňuje pripraviť údaje v rôznych formátoch. Jednoduchý výber dát v kombinácii s možnosťou exportu časových radov dovoľuje používať údaje na podrobnejšie výskumy v oblasti zdravia verejnosti.

MOŠ a MIS

Informačný systém Mestskej a obecnej štatistiky (MOŠ) [17] zabezpečuje potreby ŠÚ SR na centrálnej, krajskej a na úrovni pracoviska ŠÚ SR v oblasti zberu dát, nahrávania hodnôt ukazovateľov za jednotlivé obce SR resp. mestá SR. Informačný systém MOŠ poskytuje štatistické informácie za každú obec a mesto SR s možnosťou definovať pri výstupoch rôzne územné jednotky SR s perspektívnou väzbou na európske štatistiky. Ide o systém, ktorý má prierezový charakter a obsahuje údaje z viacerých úsekových štatistik, doplnený o ďalšie údaje a charakteristiky, ktoré štátna štatistika svojimi zisťovaniami nepokrýva. Paralelne sú definované aj ukazovatele za mestský informačný systém (MIS), s ktorým sa uvažuje pre väčšie mestá v SR. Zatiaľ je tento systém realizovaný len za hlavné mesto SR Bratislava a od roku 2003 aj pre Košice.

Databáza MOŠ [17], ktorá sa každoročne aktualizuje, obsahuje údaje za rok 1993 a od roku 1996 kontinuálne za jednotlivé roky. Zber údajov zabezpečujú v ročnej periodicite pracoviská ŠÚ SR. Báza dát je vedená ako centrálna báza dát, obsahuje údaje za všetky obce SR a tiež ako krajské bázy dát prevádzkované na pracoviskách ŠÚ SR za okresy a obce z okruhu svojej pôsobnosti. Zoznam okruhov a ukazovateľov v databáze MOŠ je v Tabuľke 4.

Všetky údaje sú predkladané buď formátované v tabuľkovej forme, alebo ako text. Ukazovateľ 7, zdravotníctvo a kúpeľníctvo, poskytuje štatistické údaje o polik-

01 – základná charakteristika,	15 – sociálna oblasť,
02 – doprava,	16 – organizačná štruktúra hospodárstva,
03 – technická vybavenosť,	17 – subjekty podľa vybraných ekonomických činností,
04 – demografia,	18 – pôsobnosť orgánov štátnej správy, súdnictva a poistenia,
05 – bytový fond,	19 – vodné toky,
06 – šport,	20 – znečisťovanie ovzdušia,
7a – zdravotníctvo,	21 – klimatické podmienky,
7b – kúpeľníctvo,	22 – geologické podmienky,
08 – obchod a pohostinstvo,	23 – ostatné prvky životného prostredia,
09 – ubytovacie zariadenia,	24 – vybrané výsledky <i>Sčítania ľudu, domov a bytov</i> 1991,
10 – peňažníctvo a poisťovníctvo,	25 – vybrané výsledky <i>Sčítania obyvateľov, domov a bytov</i> 2001.
11 – školstvo,	
12 – kultúra,	
13 – kultúrne pamiatky,	
14 – výmera obce,	

Tabuľka 4 Sústava ukazovateľov sledovaných v štatistickom informačnom systéme MOŠ

linikách, pracoviskách praktických lekárov, zdravotníckych zariadeniach, lekárnach, rýchlej zdravotnej pomoci, vozidlách a kúpeľoch. V ukazovateli 4, demografia, sú zahrnuté základné demografické údaje o narodených, zomretých, o počte obyvateľov, migrácii, o sobášoch a rozvodoch. Definícia každej položky sa zobrazí po zotrvaní kurzora na príslušnom názve ukazovateľa alebo hodnote vo forme popisu ukazovateľa. Podporujú tak jednotný výklad a identifikáciu ukazovateľov sledovaných v MOŠ. Grafické zobrazenia systém neponúka. Ako príklad si zobrazíme údaje o službách pre zdravie v meste Trnava (Obrázok 3).

Tieto údaje sa dajú použiť pre regionálne či miestne analýzy a výskumy.

Trnava		Návrat na okres Trnava
Demografia (31.12.2012)		
Ukazovateľ	Hodnota	
Počet obyvateľov k 31.12. spolu	66073	
muži	31940	
ženy	34133	
Predproduktívny vek (0-14) spolu	8866	
Produktívny vek (15-54) ženy	19193	
Produktívny vek (15-59) muži	21827	
Poproduktívny vek (55+Ž, 60+M) spolu	16187	
Počet sobášov	328	
Počet rozvodov	164	
Počet živonarodených spolu	687	
muži	341	
ženy	346	
Počet zomretých spolu	524	
muži	291	
ženy	233	
Celkový prírastok (úbytok) obyv. spolu	-146	
muži	-116	
ženy	-30	
Zdravotníctvo (31.12.2012)		
Ukazovateľ	Hodnota	
Lekárne a výdajne liekov	nie	
Samostatné ambulancie praktického lekára pre dospelých	nie	
Samostatné ambulancie praktického lekára pre deti a dorast	nie	
Samostatné ambulancie praktického lekára stomatólóga	nie	
Samostatné ambulancie praktického lekára gynekológa	nie	
Životné prostredie (31.12.2012)		
Ukazovateľ	Hodnota	
Skládka komunálneho odpadu	áno	
Komunálny odpad	áno	
Využívaný komunálny odpad	áno	
Zneškodňovaný komunálny odpad	áno	

Obrázok 3 Ukazovatele demografie, zdravotníctva a životného prostredia pre mesto Trnava (Zdroj MOŠ, 2013)

VDC

Výskumné demografické centrum (VDC) [18] bolo založené v roku 2000. Pôsobí v rámci Inštitútu informatiky a štatistiky v Bratislave (INFOSTAT). Činnosť VDC je zameraná hlavne na hodnotenie populačného vývoja a jeho jednotlivých zložiek, spracovanie populačných odhadov a prognóz a vypracovanie analytických a prognostických podkladov na rozhodovanie vrcholových riadiacich orgánov v oblasti hospodárskej, sociálnej a bytovej politiky, zamestnanosti, zdravotníctva a školstva. Plní zároveň úlohy v oblasti metodiky a metodológie, medzi iným sa zaoberá prípravou metód, modelov a nástrojov na demografické analýzy, odhady a prognózy, aktualizáciou a prevádzkou demografického informačného systému a rozvojom demografickej metodiky a metodológie. Zároveň koordinuje demografický výskum na Slovensku. Z hľadiska štúdia zdravia a služieb pre zdravie sú dôležité databázy o počte obyvateľov Slovenska, ako aj o príčinách smrti. VDC sa orientuje v prvom rade na demografiu, a preto aj údaje, ktoré ponúka, majú výhradne demografický charakter. Pre záujemcu o poznávanie zdravia populácie sú dôležité jednak údaje o populácii, pohybe obyvateľstva a tiež príčinách smrti.

-
- Tabuľky základných demografických údajov za SR
 - Pohyb obyvateľstva SR (podrobné demografické údaje)
 - SLOVAKPOPIN web stránka
 - Tabuľky prognóz obyvateľstva SR z roku 2003 (2005-2050)
 - Tabuľky prognóz obyvateľstva oblastí SR z roku 2003 (2005-2025)
 - Tabuľky prognóz obyvateľstva krajov SR z roku 2003 (2005-2025)
 - Tabuľky prognóz obyvateľstva okresov SR z roku 2003 (2005-2025)
 - Tabuľky prognóz obyvateľstva krajov SR z roku 2008 (2008-2025)
 - Tabuľky prognóz obyvateľstva okresov SR z roku 2008 (2008-2025)
 - Tabuľky prognózy obyvateľstva SR z roku 2012 (2012-2060) - nízky, stredný a vysoký variant
 - Úmrtnostné tabuľky SR
 - Stredná dĺžka života pri narodení - kraje SR, Bratislava mesto, Košice mesto
 - Stredné dĺžky života pri narodení - okresy SR
 - Stredné dĺžky života v zdraví na základe EHIS2009
 - Vytriedené príčiny smrti podľa oblasti SR (NUTS2) a podľa krajov SR (NUTS3)
-

Obrázok 4 Položky z databázy údajov prístupných prostredníctvom VDC
(Zdroj: web stránka VDC [18])

Stránka ponúka tabuľky, ale nie transformáciu do formy vhodnej na ďalšie spracovanie. Obsahuje tabuľky prognóz i strednú dĺžku života. Ponúka tiež spojenie na web stránku SLOVAK POPIN [19].

SLOVAK POPIN

Informačnú sieť o populácii (Population Information Network, POPIN) ustanovila v roku 1979 Organizácia spojených národov, jej Hospodárska a sociálna rada. Cieľom organizácie je sprístupniť medzinárodné, regionálne a národné informácie o populácii. Slovenská republika je členom tohto zoskupenia a aktívne prispieva k naplneniu jej cieľov poskytovaním údajov. SLOVAK POPIN je v anglickom jazyku a aj všetky dokumenty, ktoré je možné získať z tejto stránky, sú v angličtine. Údaje, ktoré možno z tejto stránky sťahovať, sú rozdelené do typicky demografických kategórií (Obrázok 5).

Štatistika populácie	Rozvodovosť
Všeobecné informácie	Potraty
Hlavné demografické ukazovatele	Pôrodnosť
Veková štruktúra	Úmrtnosť
Sobášnosť	Príčiny smrti
	Migrácia

Obrázok 5 Obsah databázy SLOVAK POPIN (Zdroj [19])

Na použitie v štúdiu zdravia verejnosti sú určite všetky vhodné, aj keď štatistika úmrtí podľa príčin smrti je pravdepodobne najčastejšou témou. Štruktúra voľne prístupných údajov je voliteľná z troch foriem: súbory *XML*, *XLS* (MS Office Excel 2003) alebo ako *HTML*. Dáta v prvom formáte sú primárne určené na uloženie v databázovom prostredí. *XLS* sú údaje prednostne čitateľné programom Excel. Posledný z formátov je určený na prehliadanie tabuliek priamo na obrazovke a prípadný prenos z nej pomocou *COPY-PASTE*.

Pre prostredie {R} je najjednoduchšie čítať súbory *CSV*. Tvorcovia {R} prostredia čítanie súborov *XLS* neodporúčajú pre príliš komplikované špecifikácie argumentov funkcie, aj keď to je možné. Lepšie je údaje exportovať do formátu *CSV* a potom importovať do {R}. [20] Dá sa tiež využiť prenos údajov prostredníctvom schránky (clipboard). Ukážeme si najprv postup na základe vytvorenia súboru *CSV* zo súboru *XLS* (Príklad 1) vo výbere z databázy POPIN. Údaje získame z *Main Demographic Indicators* vo formáte *XLS*, čo by malo automaticky otvoriť Excel stránku. Z nich vyberieme tie, ktoré nás zaujímajú a preniesieme ich do prázdneho listu v Excel. Potom ich zapamätáme ako textový súbor s rozšírením *.csv*. Ten už vieme prečítať do prostredia {R}. V prostredí R Commander je potrebné zadať názov premennej, do ktorej sa majú údaje uložiť. Bude vytvorená vo forme poľa údajov. Tiež musíme potvrdiť, že názvy premenných sú v prvom riadku. Keďže čítame súbor na disku počítača, tak to musíme uviesť. Vidíme, že je použitá bodkočiarka ako oddeľovač medzi údajmi. My síce pracujeme s údajmi v celých číslach, ale v prípade desatinných je potrebné uviesť, či na vyznačenie desatinných miest použijeme čiarku

	A	B	C	D	E	F
2						
3	Tab.	Main demographic indicators				
4		1950	1951	1952	1953	1954
5	Population (as of 1st January)	3447085	3485530	3533282	3576852	3629425
6	- males	1671836	1688740	1712071	1732792	1761164
7	- females	1775249	1796790	1821211	1844060	1868261
8	Population (as of 1st July)	3463446	3508698	3558137	3598761	3661437
9	- males	1678970	1700135	1724275	1744174	1779490
10	- females	1784476	1808563	1833862	1854587	1881947
11	Main age groups (as of 1st January) - males					
12	0-14	499540	509544	521450	536009	552504
13	15-59	1025525	1031702	1041093	1046007	1055347
14	60+	146771	147494	149528	150776	153313
15	0-19	653259	659781	669282	681206	697911
16	20-64	920090	929954	942083	949918	960151
17	65+	98487	99005	100706	101668	103102

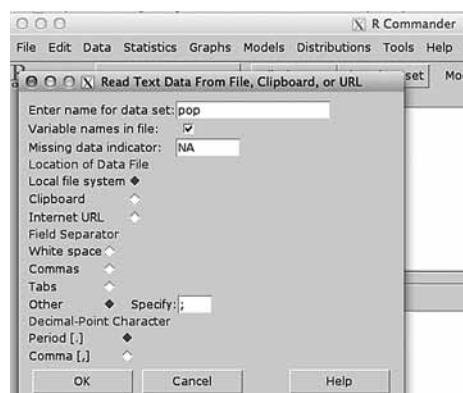
Vybrané údaje a vložené ako hodnoty (ináč by sa preniesli nuly)

1994	1995	1996	1997	1998	1999	2000
5336455	5356207	5367790	5378932	5387650	5393382	5398657

Vstupné údaje po načítaní do Excel

1994; 1995; 1996; 1997; 1998; 1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2010; 2011
 5336455; 5356207; 5367790; 5378932; 5387650; 5393382; 5398657; 5378783; 5378951; 5379161; 5380053; 53
 84822; 5389180; 5393637; 5400998; 5412254; 5424925; 5392446

Súbor vo formáte .CSV, v prvom riadku sú nadpisy stĺpcov a v druhom samotné hodnoty



```
> pop <- read.table("/Users/martinrusnak/Desktop/pop.csv", header=TRUE, sep=";", na.strings="NA", dec=".", strip.white=TRUE)
> pop
  X1994 X1995 X1996 X1997 X1998 X1999 X2000 X2001 X2002 X2003 X2004 X2005 X2006 X2007 X2008
1 5336455 5356207 5367790 5378932 5387650 5393382 5398657 5378783 5378951 5379161 5380053 5384822 5389180 5393637 5400998
  X2009 X2010 X2011
1 5412254 5424925 5392446
```

Výsledok, údaje prečítané do prostredia R

Príklad 1 Prečítanie údajov z POPIN do prostredia {R}

alebo bodku. Odošleme požiadavku a určíme súbor, z ktorého sa bude čítať. Pokiaľ sme všetko zadali správne, potom by sme mali získať premennú s obsahom čísel z databázy.

Podobným postupom načítame údaje z tabuľky, ktorá sa zobrazí prehliadačom po zvolení formátu HTML. Požadované údaje vyberieme do schránky príkazom

T 2.8.2 PACIENTI V PRAVIDELNEJ DIALYZAČNEJ LIEČBE (PDL) PODĽA ZÁKLADNEJ DIAGNÓZY A ÚZEMIA ZDRAVOTNÍCKYCH ZARIADENÍ

PATIENTS IN REGULAR DIALYSIS TREATMENT (RDT) BY DIAGNOSIS AND TERRITORY OF HEALTH ESTABLISHMENT

Diagnóza Diagnosis	Počet pacientov v PDL / Number of patients in RDT									
	SR	BL	TA	TC	NI	ZI	BC	PV	KI	
Primárne glomerulonefritídy Primary glomerulonephritis	576	94	57	51	89	76	57	69	83	
Pyelonefritída Pyelonephritis	615	67	77	64	66	81	65	117	78	
Polycystická choroba obličiek u dospelých (dominantný typ) / Polycystic kidney disease by adults (dominant typ)	224	21	19	24	36	36	35	28	25	
Poškodenie obličiek hypertenziou Injury of kidney by hypertension	327	56	30	46	27	37	63	10	58	
Renálne vaskulárne ochorenia Renal vascular diseases	207	24	16	18	38	34	27	9	41	
Poškodenie obličiek spôsobené diabetes mellitus Injury of kidney due to diabetes mellitus	1351	121	171	170	154	148	175	207	205	
Neznáma Unknown	115	26	12	25	10	14	1	25	2	
Iná Other	637	49	44	78	75	73	91	117	110	

Pomocou COPY-PASTE sme ich preniesli do Wordu a vymazali texty pred riadkami, preniesli sme tiež nadpisy stĺpcov

S údajmi z tejto tabuľky by sme chceli ďalej pracovať

```
> kidney <- read.table("clipboard", header=TRUE, sep=" ", na.strings="NA", dec=".", strip.white=TRUE)
```

Schránku (clipboard) vybranú zo stránky vo Word sme prečítali do prostredia R

```
> kidney
```

```
SR BL TA TC NI ZI BC PV KI
1 583 78 78 48 64 70 68 95 82
2 704 70 82 84 57 92 107 127 85
3 237 21 19 30 38 41 39 26 23
4 294 48 27 30 23 30 57 15 64
5 206 21 14 30 30 34 29 11 37
6 1297 100 169 187 142 134 167 199 199
7 86 23 8 15 7 15 3 5 10
8 594 79 34 43 101 73 65 100 99
```

Príklad 2 Prečítanie údajov zo Štatistickej ročenky do premenných v prostredí {R}

COPY alebo Ctrl-C a potom je postup rovnaký ako v predchádzajúcom príklade. Jediným rozdielom je označenie možnosti *Clipboard* v okienku *R Commander*. Prirodzene odpadá ukladanie na disk a výber súboru, všetko ostatné je rovnaké.

NCZI

Národné centrum zdravotníckych informácií (NCZI) je štátnou príspevkovou organizáciou, ktorá sa zaoberá okrem iného zberom, spracovaním a poskytovaním zdravotníckych štatistických údajov. [15] Predstavuje teda organizáciu, ktorá je hlavným poskytovateľom údajov o zdraví a zdravotníckych službách v Slovenskej republike. Základným produktom, ktorý pozná dúfajme každý, kto sa zaoberá zdravím verejnosti, je Edícia ročeniek, kde sa vydávajú každoročne *Zdravotnícke ročenky SR*. [21] Na web stránke sú ročníky od roku 1996 vo formáte Adobe *pdf*. Tým je zároveň dané aj obmedzenie pri používaní údajov. Transformácia z tohto formátu do takého, ktorý možno použiť na ďalšie spracovanie, je dosť náročná na čas. Mnohé z údajov sú len vo forme grafov. Jednou z možností, ako tieto údaje využiť, je ich prepísať do listu tabuľkového editora. Určite to nie je príliš elegantné riešenie. Iným riešením je využiť schopnosť funkcie R *read.table()* čítať obsah schránky (clipboard) (Príklad 2). Pozor, v tabuľkách sa oddelujú tisíce medzerou, pre správne prečítanie tieto medzery je nutné vymazať, pretože ináč to {R} rozpozná ako dve hodnoty a vyhlási chybu, pretože mu nebude sedieť počet stĺpcov. Výsledok načítania vloží do premennej triedy dátové pole. Narábanie s ním uľahčuje funkcia *attach()*, ktorá umožní pracovať s jednotlivými premennými.

Tvorivý čitateľ určite nájde aj iné spôsoby prenesenia údajov z ročeniek do prostredia štatistických programov, tabuľkových editorov alebo databáz.

Ďalšie zaujímavé údaje (tiež vo formáte *pdf*) sú dostupné v časti *Edícia analytických publikácií*. [21] Tu nájde záujemca údaje o incidencii onkologických ochorení či vývoj úrazov a ďalšie publikácie.

Zaujímavým zdrojom údajov sú Národné zdravotné registre. NCZI ich definuje ako *zdravotnícke informačné systémy, ktorých úlohou je predovšetkým zhromažďovať a spracovávať údaje o závažných, hromadne sa vyskytujúcich ochoreniach a poskytovať odbornej verejnosti základné epidemiologické a klinické charakteristiky súvisiace s výskytom a príčinami týchto ochorení*. [22] Nasledujúca tabuľka 5 podáva prehľad registrov, ktoré sú v správe NCZI. Niektoré poskytujú dáta len na základe zdôvodnenej požiadavky. Z iných, napríklad onkologického registra, sú verejne prístupné publikácie s tabuľkami údajov vo formáte *pdf*.

Národný register základných zdravotných údajov
Národný register pacientov s onkologickým ochorením
Národný register pacientov s diabetes mellitus 1. typu
Národný register pacientov s vrodenou chybou srdca
Národný register pacientov so srdcovo-cievnyim ochorením
Národný register pacientov s cievnyim ochorením mozgu
Národný register pacientov s chronickým ochorením pľúc
Národný register pacientov s vrodenou vývojovou chybou

Tabuľka 5 Národné zdravotné registre v správe NCZI

Mimo správy NCZI je Národný transplantačný register, Národný register pacientov s tuberkulózou, Národný register pacientov s prenosnými ochoreniami a Národný artroplastický register.

Medzinárodné zdroje štatistických údajov

Dostupnosť validných, spoľahlivých a porovnateľných informácií o zdraví, ktoré umožňujú formovať miestne, regionálne, národné a globálne stratégie, možno naplňať prostredníctvom štyroch vzájomne prepojených iniciatív: zlepšenie technológie a metód merania zdravotného stavu obyvateľstva, posilnením národných kapacít a motiváciou vlád zhromažďovať a analyzovať užitočné údaje o zdravotnom stave, o stanovení globálnych noriem a štandardov pre základné miery zdravia a ich meranie. Zároveň o poskytovaní svetovo platných, spoľahlivých a porovnateľných správ, ktoré umožnia hodnotenie vstupov, a výsledkov poskytovania služieb v oblasti zdravia. [23] Medzinárodné zdroje údajov v podstate kopírujú systém národných štatistík s tým rozdielom, že poskytujú prehľad o údajoch z rôznych krajín. Ich cieľom je hlavne poskytnúť materiál na porovnanie medzi krajinami. Často sú menej podrobné ako národné štatistiky.

MacroDataGuide

MacroDataGuide [24] je používateľská príručka pre výskumných pracovníkov a študentov, používaná globálne pri hľadaní kvalitných údajov pre spoločenskovedný výskum. Informuje nielen o dostupnosti údajov – kde ich možno nájsť, ale tiež poskytuje kvalitatívne hodnotenie týchto zdrojov. Vytvorila a udržiava ju *Norwegian Social Science Data Services* (NSD), čo je spoločnosť s ručením obmedzeným, ktorú vlastní Ministerstvo vzdelávania a vedy v Nórsku. Portál vychádza zo správy vypracovanej pre *European Social Survey* (ESS), ktorá skúma dostupnosť a porovnateľnosť existujúcich zdrojov kontextových štatistických údajov. Tie môžu byť zaujímavé pre používateľa ESS dát. Web portálom je webová stránka, ktorá prináša informácie spoločne z rôznych zdrojov jednotným spôsobom. Obvykle každý zdroj informácií je na stránke jedného odboru (portlet), a používateľ môže vybrať, ktoré z nich chcete zobrazit'. [25]

Zdroje *MacroDataGuide* obsahujú posúdenie aktuálnej dostupnosti a pokrytia kontextových zdrojov dát. Kým pôvodne bol určený pre používateľov európskych mikroúdajov (ESS používateľov), postupne rozširuje svoj záber a dnes poskytuje makroúdaje pre mnohé krajiny. Zdroje dát sa značne líšia v zameraní a rozsahu a zahŕňajú všeobecné databázy, ktoré obsahujú sociálne a ekonomické dáta, veľké dátové množiny s radom ukazovateľov politických premenných, a indexy produkované mimovládnyimi a medzivládnyimi organizáciami. Dátové súbory zostavujú nezávislí

výskumní pracovníci. Informácie poskytnuté na tomto portáli pochádzajú z rôznych inštitúcií a výskumných projektov. Okrem toho sú tam zaradené aj nezávislé recenzie databáz uverejnené v časopisoch a relevantné štúdie. Pre záujemcu, ktorý hľadá zdroje údajov z celého sveta, poskytuje portál účinný nástroj, ktorý ušetrí čas a náklady.

Ďalej sa pozrieme na tri najvýznamnejšie zdroje z medzinárodných organizácií. Prvým sú dve databázy Svetovej zdravotníckej organizácie a to *Health for All* a *Mortality database*. Druhým sú databázy Eurostatu, ktoré sa týkajú zdravia.

Svetová zdravotnícka organizácia

Svetová zdravotnícka organizácia (SZO) zostavuje a zverejňuje štatistiky o širokej škále tém spojených so zdravím, ktoré ponúka záujemcom či výskumným pracovníkom vo forme viacerých databáz a tlačенých publikácií. Prednostne sa poskytujú informácie o úmrtnosti, chorobnosti, zdravotnom stave, pokrytí službami a o rizikových faktoroch.

European Health for All, Európska databáza, Zdravie pre všetkých

Európska databáza Zdravie pre všetkých (HFA) [26] poskytuje prístup k širokému spektru základných zdravotníckych štatistík pre 52 členských štátov európskeho regiónu SZO. Údaje pochádzajú z úradných registrov a prieskumov. Zhromažďujú sa z rôznych zdrojov. Časť z nich sa každoročne zbiera priamo v členských krajinách. Ďalšia časť dát pochádza od technických jednotiek SZO, ktoré získavajú príslušné štatistické informácie v rámci ich vlastného odboru. Druhotné zdroje, akými sú údaje poskytované medzinárodnými organizáciami a agentúrami, tvoria rovnako dôležitý zdroj dát. Databáza obsahuje asi 600 ukazovateľov, ktoré sú usporiadané v skupinách: demografia a sociálno-ekonomická štatistika, ukazovatele úmrtnosti, chorobnosť, invalidita a hospitalizácie, rizikové faktory životného štýlu a prostredia, zdroje zdravotnej starostlivosti, využívanie služieb zdravotnej starostlivosti a náklady, zdravie matiek a detí. On-line verzia databázy je založená na pop-up oknách, okrem nej je možné stiahnuť aj offline verziu. Databáza umožňuje používateľom vybrať si krajinu, premenné a roky pri vytváraní tabuliek. Výsledky výberu sa zobrazia buď vo forme tabuliek, máp, alebo grafov. Všetky výstupy možno vytlačiť alebo v offline verzii exportovať na nasledujúce spracovanie inými nástrojmi. Údaje boli zostavené, overené a spracované jednotným spôsobom s cieľom zlepšiť medzinárodnú porovnateľnosť štatistík.

Databáza úmrtnosti, Mortality Database

Databáza úmrtnosti SZO [27] obsahuje viac ako dva milióny záznamov príčin úmrtí v jednom dátovom súbore. Tie zahŕňajú úmrtia registrované v národných re-

gistračných systémoch úmrtí podľa príčin smrti tak, ako boli kódované príslušným vnútroštatným orgánom. Tieto údaje predstavujú oficiálne národné štatistiky v tom zmysle, že Svetovej zdravotníckej organizácii ich odovzdali príslušné orgány zainteresovaných krajín. Pri práci s údajmi je potrebné zvážiť národné rozdiely v registrácii úmrtí a kódovaní postupov, najmä v použití kódov pre zle definované a neznáme príčiny. SZO poskytuje tieto údaje s podmienkou, že nebudú použité na komerčné účely. Prehľad súborov s popisom obsahu je v tabuľke 6.

Samotné údaje sú rozdelené podľa rokov do viacerých skupín. Súbor s dokumentáciou obsahuje všetky kódy a vysvetlenia k nim, bez ktorých sa s údajmi nedá pracovať. Súborov sú formátované ako textové, s rozšírením .csv a nie sú určené pre nezaškoleného užívateľa. Dátové súbory, spolu s potrebnými pokynmi, štruktúrami súborov, referenčnými tabuľkami sú primárne určené najmä na výskumné účely. Osoby, ktoré ich používajú, potrebujú databázové nástroje na prácu s nimi, keďže pre bežné kancelárske programy sú príliš veľké a komplikované.

Názov súboru	Obsah
Documentation	Súbor formátovaný v MS Word, ktorý obsahuje informácie o WHO Mortality Database, špecifikácie súborov a zoznam príčin smrti
Availability	Súbor formátovaný v MS Excel, ktorý obsahuje zoznam krajín a rokov, pre ktoré údaje o úmrtnosti a populácii sú dostupné
Country codes	Kódy a názvy krajín
Notes	Poznámky k niektorým údajom z určitých rokov a krajín
Populations and live birth	Referenčné populácie a štatistika živo narodených
Mortality, ICD-7	Súbor s údajmi, ktoré obsahujú podrobné údaje o úmrtnosti pre siedmu revíziu MKCH (International Classification of Diseases).
Mortality, ICD-8	Súbor s údajmi, ktoré obsahujú podrobné údaje o úmrtnosti pre ôsmu revíziu MKCH (International Classification of Diseases).
Mortality, ICD-9	Súbor s údajmi, ktoré obsahujú podrobné údaje o úmrtnosti pre deviatu revíziu MKCH (International Classification of Diseases).
Mortality, ICD-10	Súbor s údajmi, ktoré obsahujú podrobné údaje o úmrtnosti pre desiatu revíziu MKCH (International Classification of Diseases).

Tabuľka 6 Popis obsahu súborov, ktoré obsahuje databáza SZO Mortality Database

Register indikátorov a meraní

Register indikátorov a meraní (IMR) je ústredným zdrojom definícií indikátorov, vo forme textov a formátoch čitateľných počítačom. [28] Poskytuje kompletne a dobre štruktúrované metadáta o indikátoroch, harmonizáciu riadenia definícií indikátorov a číselníkov, prístup k definíciám indikátorov prostredníctvom internetu. Takto zabezpečuje súlad medzi rôznymi oblasťami štatistických dát. Je podporovaný prostredníctvom formátu SDMX-HD. V rámci IMR sa nachádza úložisko

dát Global Health Observatory Data Repository, ktoré obsahuje rozsiahly zoznam ukazovateľov, ktoré sa môžu vyberať na základe témy alebo prostredníctvom viacnásobnej otázky. Predstavuje hlavné úložisko údajov, indikátorov a informácií zo zdravotnej štatistiky Svetovej zdravotníckej organizácie. Základom úložiska je 50 súborov dát o prioritne zdravotných témach vrátane mortality a závažnosti ochorenia, rozvojových cieľoch tisícročia (detská výživa, zdravie detí, matiek a reprodukčné zdravie, očkovanie, HIV/AIDS, tuberkulóza, malária, zanedbávané ochorenia, voda a kanalizácia), neprenosných chorôb a rizikových faktoroch, choroby epidemického významu, zdravotné systémy, environmentálne zdravie, násilie a zranenia, rovnosť a iné. Okrem toho poskytuje on-line prístup k ročnému zhrnutiu údajov, ktoré sa týkajú zdravia pre 194 členských štátov SZO, pod názvom Svetová zdravotnícka štatistika.

World Health Statistics

- ▶ Mortality and burden of disease
- ▶ Cause-specific mortality and morbidity
- ▶ Selected infectious diseases
- ▶ Health service coverage
- ▶ Risk factors
- ▶ Health workforce, infrastructure, essential medicines
 - Health workforce
 - Infrastructure
 - Essential medicines
- ▶ Health expenditure
- ▶ Health inequities
- ▶ Demographic and socioeconomic statistics
- ▶ Health information systems and data availability
- ▶ Environmental health
- ▶ Epidemic prone diseases
- ▶ Equity
- ▶ Global Information System on Alcohol and Health
- ▶ Health systems
- ▶ Health-related Millennium Development Goals
- ▶ Health workforce
- ▶ HIV/AIDS and other STIs

Health workforce, infrastructure, essential medicines: Health workforce

Showing

Show filters

Download this data as
 CSV (codes only) | CSV (text only) | CSV (text and codes) | Excel (SpreadsheetML) | HTML (flat table) | GHO XML

Details: off

		Number of nursing and midwifery personnel ^f	Number of pharmaceutical personnel ^f	Number of physicians ^f	Number of psychiatrists ^f	Nursing and midwifery personnel density (per 10 000 population) ^f
Singapore	2009	20 712	1 030	5 323	130	59
Slovakia	2009	1 789				3.28
	2008	3 5539			621	65.8
	2007		2 517	16 201		
Slovenia	2009	16 480	1 086	4 915	143	83.94
	2008	16 009	972	4 854		81.6
Solomon Islands	2009	1 080		118		20.53
	2008		53			
Somalia	2005	694	53	89		14.5
	2009				4	
	2006	965	50	300		1.1
South Africa	2005					
	1997					
	2009				136	
Spain	2004	18 4459	12 521	34 829		40.8
	2010	224 800	37 000	174 100	3 894	51.1

Obrázok 6 Príklad výberu z Global Health Observatory Data Repository

Na obrázku 6 je vidieť, že údaje okrem tabuľky sa dajú stiahnuť v rôznych formátoch, napríklad text oddelený čiarkami, teda .csv, tiež vo formáte čitateľnom Excel alebo HTML. To poskytuje skutočne široké možnosti spracovania údajov na potreby štúdia rôznych parametrov zdravia na globálnej – celosvetovej úrovni.

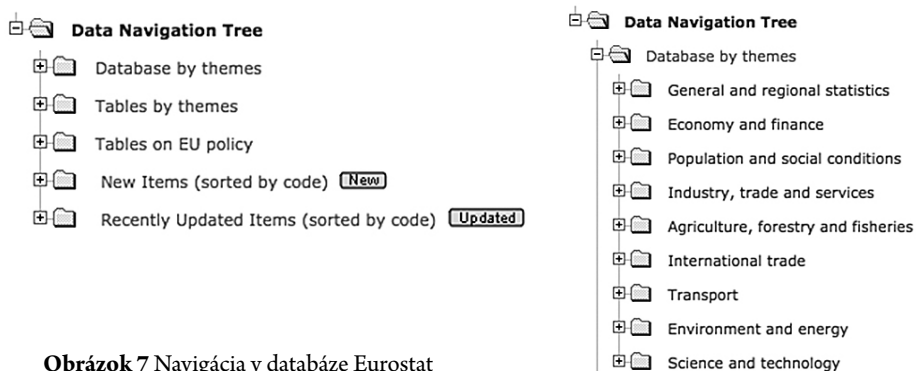
Eurostat

Eurostat [29] je štatistický úrad Európskej únie, ktorý sa nachádza v Luxemburgu. Jeho úlohou je poskytovať Európskej únii štatistiku, ktorá umožňuje porovnanie krajín a regiónov na európskej úrovni. Eurostat bol založený v roku 1953 na podporu Spoločenstva uhlia a ocele. V priebehu rokov sa jeho úloha rozšírila a po založení Európskeho spoločenstva v roku 1958 sa Eurostat stal jedným z generálnych riaditeľstiev (DG) Európskej komisie. Kľúčovou úlohou je poskytovať štatistiky pre iné generálne riaditeľstvá a ostatným európskym inštitúciám dáta tak, aby mohli definovať, realizovať a analyzovať politiky Spoločenstva. Eurostat takto ponúka celý rad dôležitých a zaujímavých údajov, ktoré vlády, podniky, vzdelávací sektor, novinári a verejnosť môžu použiť na svoju prácu a v každodennom živote. Vzhľadom na uvedené, Eurostat nie je klasickou štatistickou ustanovitzňou, s akými sa stretávame na národnej úrovni. Jeho úlohou nie je zhromažďovať údaje, tie sa zbierajú v členských štátoch. Hlavnou úlohou je spracovávať a zverejňovať porovnateľné štatistické informácie na európskej úrovni. Aj z toho dôvodu sa snaží dospieť k spoločnému štatistickému „jazyku“, ktorý zahŕňa pojmy, metódy, štruktúry a technické normy. Inými slovami, ide mu o zabezpečenie porovnateľnosti pomocou harmonizovanej metodiky.

Databáza EUROSTAT

Eurostat poskytuje údaje prostredníctvom web stránky s prehľadným používateľským prostredím.

Z obrázku 7 je vidieť, že prístup k údajom je prehľadný a účelný. Väčšinu údajov je možno získať vo viacerých formátoch – *XLS*, *CSV*, *HTML*, *PDF*, *PC-AXIS*, *SPSS*, a *TSV*. S prvými štyrmi sme sa už oboznámili. *PC-Axis* [30] je rodina programových produktov, ktorá sa skladá z niekoľkých programov pre prostredia Windows a Internet s cieľom prezentovať štatistické informácie. Je to výsledok celosvetového projektu spolupráce na šírenie štatistík. *SPSS* [31] je komerčný balík štatistických



Obrázok 7 Navigácia v databáze Eurostat

programov, široko využívaný vo vede a výskume. TSV formát je rovnako ako CSV textový súbor, kde však jednotlivé položky oddelujú tabulátory namiesto čiarok. Okrem samotných údajov Eurostat ponúka metadáta, ktoré popisujú jednotlivé položky. Je to veľmi užitočný zdroj informácií o štatistických údajoch.

CODED je databáza pojmov a definícií a ďalších on-line slovníkov, ktoré sa týkajú štatistík získaných prieskumom. Ďalej sa tu nachádzajú právne predpisy EÚ a metodické príručky, ktoré súvisia so štatistikou. Tiež sú tu medzinárodné štatistické klasifikácie a číselníky, ako aj slovník Medzinárodného štatistického inštitútu ISI (*International Statistical Institute*) [32] a ďalšie on-line a synonymické slovníky, ktoré sa týkajú štatistík. V neposlednom rade sa tu nachádza prehľad metodík prieskumov, ktoré sa používajú na národných úrovniach v príprave štatistiky pre EÚ, správ o kvalite atď. Je tu tiež už spomínaná Euro-SDMX Metadata Štruktúra (ESMS), vo forme súboru medzinárodných štandardov na výmenu štatistických informácií medzi organizáciami. Nakoniec sa tu dajú nájsť štandardné číselníky odporúčané pre tvorbu databáz a prenos dát.

Geografický informačný systém (GIS) integruje hardvér, softvér a dáta na vytváranie, správu, analýzu a zobrazenie všetkých foriem geograficky prezentovanej informácie. Služba Eurostatu s názvom GISCO podporuje a stimuluje využitie GIS v rámci Európskeho štatistického systému a Európskej komisie. Využíva sa na riadenie a šírenie geografickej referenčnej databázy Komisie. Produkuje mapy, priestorové analýzy, podporuje geo-referencovanie štatistík a poskytuje používateľskú podporu pre používateľov GIS.

Súhrn

Pokúsili sme sa zhrnúť najdôležitejšie zdroje údajov z rutínnej štatistiky z domova i zo zahraničia. Nie je pri tom možné predstaviť všetky zdroje a ani by to nebolo účelné. Napokon záujemca s konkrétnym problémom na riešenie si vhodný zdroj údajov nájde. To, čo sme popísali v tejto kapitole slúži najmä na orientáciu a v nasledujúcich kapitolách sa budeme na niektoré z uvedených zdrojov odvolávať. Čitateľovi tejto kapitoly odporúčame, aby popri čítaní popisu jednotlivého zdroja si otvoril web stránku a pokúšal sa sám nájsť, prehliadať položky, ktoré produkt ponúka. Len tak získa komplexnejší obraz a predstavu o prezentovaných zdrojoch a ich možnom využití.

Literatúra

1. Rusnák, M., Rusnáková, V., Majdan, M. Bioštatistika pre študentov verejného zdravotníctva. I. ed, 2010, Trnava: Typi Universitatis Tyrnaviensis, Vydavateľstvo Trnavskej univerzity. 216 s.

2. Nelson, C., Gager, J., Gregory, A. SDMX 2.1 User Guide. 2012, Statistical Data and Metadata Exchange SMDX.
3. WHO. International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Version for 2007. 2007 [cited 2010 30.12.]; Available from: <http://apps.who.int/classifications/apps/icd/icd10online/>.
4. ISO, INTERNATIONAL STANDARD ISO 8601. Data elements and interchange formats — Information interchange —Representation of dates and times. 2004, ISO copyright office: Case postale 56, CH-1211 Geneva 20.
5. Ministerstvo dopravy pôšt a telekomunikácií SR, Metodický pokyn k štandardom pre informačné systémy verejnej správy a ich rozdeleniu a náležitostiam, spôsobu ich aktualizácie a oblastiam ich uplatnenia pre informačné systémy verejnej správy, Ministerstvo dopravy pôšt a telekomunikácií SR, Editor. 2006: Bratislava.
6. Harold, E. R. XML™ Bible. 2004, Foster City, CA, Chicago, IL, Indianapolis, IN, New York, NY: IDG Books Worldwide, Inc. An International Data Group Company.
7. Nassau, C. CSV File Format Specification. 2012, Curzon Nassau: London.
8. Last, J. M. A Dictionary of Epidemiology. 1995, Oxford: Oxford University Press. 196.
9. Porta, M., Greenland, S., Last, J. M. Editors. A Dictionary of Epidemiology. 2008, Oxford University Press, New York: Oxford. p. 289.
10. Last, J. M. Slovník epidemiológie. 1999, Bratislava: USAID.
11. Jurčová, D. Slovník demografických pojmov. 2005, INFOSTAT – Inštitút informatiky a štatistiky. Výskumné demografické centrum: Bratislava.
12. EUROSTAT. Statistics Explained. 2013 [cited 2013 2013-01-17]; Available from: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Statistics_Explained.
13. Zákon o štátnej štatistike, in: Zbierka zákonov č. 298/2010, 2010: Bratislava.
14. Zákon o informačných systémoch verejnej správy a o zmene a doplnení niektorých zákonov, in: 275, N.R.S. republiky, Editor. 2006: Bratislava.
15. Národné centrum zdravotníckych informácií (NCZI). [cited 2013 27.1.]; Available from: <http://www.nczisk.sk/O-nas/Pages/default.aspx>.
16. SLOVSTAT. 2012 [cited 2012 10.1.2012]; Available from: <http://www.statistics.sk/pls/elisw/vbd>.
17. Štatistický informačný systém MOŠ. [cited 2013 27.01]; Available from: <http://app.statistics.sk/mosmis/sk/run.html>.
18. Výskumné demografické centrum (VDC). [cited 2013 27.01.]; Available from: http://www.infostat.sk/vdc/sk/index.php?option=com_content&task=view&id=5&Itemid=4.
19. Slovak POPIN. [cited 2013 28.01.]; Available from: <http://www.infostat.sk/slovakpopin/>.
20. Ripley, B. R. Data Import/Export. 2012: R Core Team
21. NCZI. Zdravotnícke ročenky SR. Edícia analytických publikácií [cited 2013 31.01.]; Available from: <http://www.nczisk.sk/Publikacie/Pages/Edicia-analytickych-publikacii.aspx>.
22. NCZI. Národné zdravotné registre. [cited 2013 02.02.]; Available from: <http://www.nczisk.sk/Registre/Narodne-zdravotne-registre/Pages/default.aspx>.
23. Murray, C. J., Lopez, A. D., Wibulpolprasert, S. Monitoring global health: time for new solutions. BMJ, 2004. 329(7474): p. 1096 – 1100.
24. Norwegian Social Science Data Services. MacroDataGuide. 2013 [cited 2013 16.02.]; Available from: <http://www.nsd.uib.no/macrodataloguide/>.
25. Business dictionary. Web portal. [cited 2013 16.02.]; Available from: <http://www.businessdictionary.com/definition/portal.htm>
26. WHO EURO. European Health for All Database. [cited 2013 16.02.]; Available from: <http://data.euro.who.int/hfad/>.

27. WHO. WHO Mortality Database. 2012; Available from: <http://www.who.int/whosis/mort/download/en/index.html>.
28. WHO. Indicator definitions and metadata. 2012 16.12.2012]; Available from: <http://www.who.int/whosis/indicators/en/>.
29. EUROSTAT. EUROSTAT-DATABASE. 2013 [cited 2013 19.02.]; Available from: http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database.
30. Statistics Sweden. PC-Axis web site. [cited 2013 11.03.]; Available from: http://www.scb.se/Pages/StandardNoLeftMeny____314045.aspx.
31. IBM. IBM SPSS Statistics [cited 2013 11.03.]; Available from: <http://www-01.ibm.com/software/analytics/spss/>
32. ISI. International Statistical Institute. [cited 2013 20.08.]; Available from: <http://www.isi-web.org>.