

# Ve veřejné databázi najdete už 50 milionů údajů

Západoevropské a skandinávské statistické úřady nabízejí **on-line přístup k výsledkům statistických zjišťování** přes internet. V ČSÚ plní tuto úlohu veřejné databáze.

Myšlenka vytvořit databázi statistických údajů určených pro veřejnost vznikla v období 2001–2002. V lednu 2006 se veřejná databáze objevuje na našich webových stránkách (viz <http://wdb.czso.cz>). Uživatelům jsme zpočátku nabízeli údaje Sčítání lidu, domů a bytů 2001, volební výsledky nebo atraktivní čísla z demografické statistiky. Přestože databáze dnes nabízí desítky milionů údajů ze všech oblastí statistiky, stále se nám nedaří ukládat všechny údaje určené pro veřejnost. Důvodem jsou výrazně odlišné způsoby zpracování sociálních a ekonomických statistik, které nevyužívají jednotný popis vypočítaných údajů.

Zásadní změnu by měl přinést připravovaný projekt redesignu statistického systému plánovaný do roku 2013 (pozn. red.: viz článek „Už za několik měsíců přinesou projekty výsledky“, s. 20).

Veřejná databáze má dvě části, primární představuje obrovské úložiště dat s detailním popisem každého údaje, v sekundární jsou uloženy předdefinované statistické tabulky, grafy, mapy, které



Veřejná databáze umí rozpoznat **i nejpreciznější údaje**.

umožňují uživatelům statistické údaje prohlížet. Zájemci o data si mohou prohlížet několik tisíc

předdefinovaných statistických tabulek. Vybírat si mohou v navigačním stromu kapitol a podkapitol, který se používá na webových stránkách a v katalogu našich produktů. Využít mohou i fulltextový vyhledávač.

## Dynamická databáze

Většina tabulek je navržena tak, aby uživatel mohl interaktivně nastavit požadované období, území

nebo ukazatel, případně přepínat formu zobrazení mezi tabulkou – grafem – mapou. Důležitou vlastností veřejné databáze je její dynamičnost. Teprve v okamžiku, kdy si uživatel vybere požadovanou tabulku, dojde k jejímu naplnění daty z primární části databáze. Obsah prezentovaných tabulek, grafů a kartogramů se tak přidává do prázdných šablon výstupních

## Milióny dat

Veřejná databáze obsahovala koncem října 2011 zhruba **50 milionů údajů**. Za poslední 3 roky do ní přibýlo **každý měsíc v průměru 900 tisíc nových dat**. U každého jsou zapsány vedle věcného, územního nebo časového vymezení také informace o důvěrnosti údaje, kvalitě, zdroji nebo intervalu zveřejnění.

objektů až ve chvíli požadavku na jejich zobrazení.

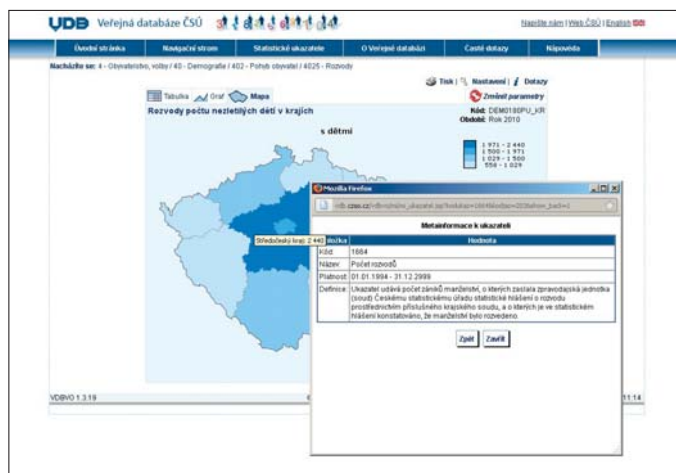
Zvládnutí je možnost automaticky zobrazovat nejnovější údaje o ukazateli. Každou půlnoc totiž dochází k analýze uložených dat a k označení nejčerstvějších nebo nejpreciznějších údajů pro každý statistický ukazatel. Této možnosti využívá řada předdefinovaných tabulek – například Vybrané statistické údaje za obec (obr. 1), kde jsou k dispozici údaje za různé

## Specialita naší databáze

Veřejnou databázi jsme založili **na detailním popisu každého zveřejněného údaje** a ne na popisu statistických tabulek, jak bylo běžné v ostatních statistických úřadech. Toto řešení nám umožňuje hlídat konzistenci údajů z různých šetření a respektovat základní pravidlo statistiků – **zajištění jednoty čísla**.



Obrázek 1: Ukázka předdefinované tabulky s využitím neaktuálnějších údajů.



Obrázek 2: Interaktivní zobrazení metainformace ke kartogramu.



Obrázek 3: Ukázka grafu s nejstaršími údaji ve VDB.

### Nejžádanější tabulky

Každý měsíc se automaticky aktualizuje žebříček Top 10 nejčastěji zobrazovaných tabulek. Primát drží tabulka Vybrané statistické údaje za obec, u které měsíčně evidujeme kolem sto tisíc zobrazení (jsou zde započítány i přístupy vyhledávacích serverů). V popředí zájmu jsou také tabulky národních účtů. Zejména porovnání ukazatele HDP na obyvatele mezi jednotlivými kraji uživatele láká. Tisíce přístupů každý měsíc zjišťujeme k tabulkám zahraničního obchodu, cenové statistiky, ale zejména demografické statistiky. Rostoucí zájem o výsledky sčítání dokládá i vysoký počet přístupů k nejdelší časové řadě – porovnání historických výsledků sčítání od roku 1869.

ukazatele pro každou ze 6 251 obcí v naší republice. Tabulka nenabízí údaje za předem určené období, ale dohledají se vždy údaje opatřené příznakem „nejnovější údaj“. Stejnou vlastnost využívají i některé tabulky připravené jako přehled ukazatelů pro územně analytické podklady, které slouží zejména krajským a obecním úřadům.

Veřejná databáze umí rozpoznat i nejpřesnější údaje. Časté revize údajů jsou běžné zejména

pro ekonomické statistiky a národní účty. Databáze umožňuje připravovat a prezentovat tabulky i tak, že se v nich zobrazí vždy údaje poslední revize, případně údaje definitivní, tedy takové, u nichž se další revize nepředpokládá. Využíváme toho zejména pro tabulky o HDP a dalších ukazatelích národních účtů.

Možnost rychlého přístupu k tzv. metainformacím je typickou vlastností veřejné databáze.



Obrázek 4: Ukázka nových funkcí pro animovaný graf ve VDB

Ve všech tabulkách má uživatel možnost kliknutím na konkrétní hodnotu zobrazit třeba definici příslušného ukazatele nebo informaci o přesnosti zobrazeného údaje. Totéž platí při kliknutí na sloupec zobrazeného grafu nebo plochu vykresleného kartogramu (obr. 2).

Veřejná databáze nabízí i možnost prezentace statistických dat v animovaných grafech. K tomu využívá volně dostupný nástroj pro vizualizaci dat společnosti Google nazývaný Motion Charts. V nedávné době pod každým animovaným grafem přibyla možnost pro uložení a sdílení odkazu na stav grafu v libovolném okamžiku animace, možnost graf vložit do jiné webové stránky a možnost exportu dat do formátu XLS (obr. 4).

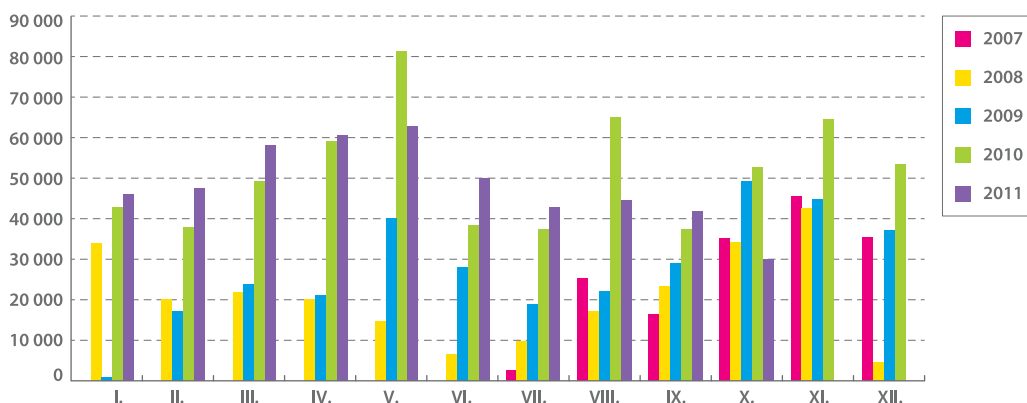
### O co je největší zájem?

Návštěvnost veřejné databáze se pohybuje v řádu statisíců přístupů měsíčně. Po odečtení přístupů vyhledávacích serverů dosáhla průměrná měsíční návštěvnost v prvních devíti měsících roku 2011 50 tisíc zobrazení, což je mírný nárůst proti loňskému roku, ale dvojnásobek v porovnání s rokem 2009. Rapidní růst návštěvnosti v roce 2010 ovlivnil i velký zájem o data s volebními výsledky koncem května a v říjnu (graf na s. 31).

### Chystané změny

Veřejná databáze na přelomu roku 2011–2012 nabídne první předběžné výsledky Sčítání lidu, domů a bytů 2011. Během příštího roku

Návštěvnost veřejných databází (k 20. 10. 2011)



Zdroj: ČSÚ

se budou postupně prezentovat další výsledky tohoto nejrozsáhlejšího cenzu. Vše už v inovované grafické podobě a s rozšířenou nabídkou možností pro uživatele.

Pro novou verzi veřejné databáze připravujeme novou navigační strukturu. Přehlednost nabízených tabulek bude zvýšena presenta-

cí základních tabulek s možností zobrazit podrobnější doplňkové tabulky. Časové řady s jednotným vzhledem pro všechny statistické oblasti budou rovněž dostupné prostřednictvím novější verze veřejné databáze.

Kromě toho přinese nová verze také možnosti tvorby plovoucích

časových období i jiných složitějších struktur v tabulkách. Uživatelé budou mít další možnosti úpravy nabízených tabulek, např. záměnu hlavičky a legendy tabulek, seřazení sloupců či řádků, skrytí či zobrazení některých částí tabulky, výběr z více grafů a kartogramů k jedné tabulce, možnost zobra-

zení složitějších grafických výstupů včetně úpravy jejich nastavení, možnost exportu jednoho či více datových objektů do \*.pdf souboru a v neposlední řadě i možnost vložení objektů z databáze na jiné webové stránky.

### Exportujeme do \*.xml

Využívání webových služeb je perspektivou pro veřejné databáze všude ve světě. Naše současná databáze podporuje export obsahu každé předdefinované tabulky ve formátu \*.xml, díky kterému mohou být data dále automaticky zpracována aplikacemi uživatelů. Pro potřeby automatické aktualizace statistických údajů na Portálu veřejné správy byla ve veřejné databázi zprovozněna webová služba, která na určené adrese generuje aktuální data ve formátu \*.xml pro vybrané ukazatele za všechny obce.

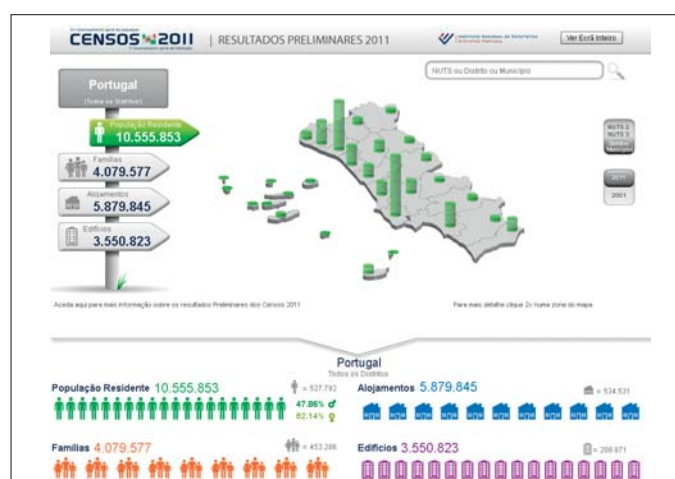
## Uživatelé přestávají být pasivní

Inspirací pro veřejnou databázi jsou každoroční mezinárodní konference o marketingu a výstupních databázích, kde prezentujeme **naše novinky**.

Trendem uplynulých let byly nové možnosti vizualizace statistických dat. Letos zaujaly produkty a služby vyvinuté pro mobilní zařízení, zejména pro iPhone a iPad.

### Aktivní producenti

Velkou diskusi vyvolala nová myšlenka o tom, že uživatelé statistických dat už nejsou pouhými konzumenty produktů statistických úřadů, ale stávají se aktivními producenty nových aplika-



Ukázka vizualizace statistických dat z Portugalska.

cí, služeb a výstupů využívajících statistická data.

Hlavním úkolem statistických úřadů na poli diseminace podle zastánců této teze není přitáhnout návštěvníky na svoje webové strán-

ky, ale naopak vytvořit nástroje a služby pro automatické sdílení dat, díky nimž by další informační producenti a koncoví uživatelé mohli statistické údaje využívat ve svých vizualizacích, interaktivních

aplikacích, produktech a zařízeních. Výstupy statistických údajů by tímto způsobem mohly být mnohem více šířeny prostřednictvím oblíbených serverů, portálů a mobilních zařízení a využívány v takové formě a podobě, která koncovým zákazníkům vyhovuje a na kterou jsou zvyklí.

V tomto směru již vykročil například statistický úřad Katalánska ve Španělsku, který vyvinul a zveřejnil popis uceleného aplikačního rozhraní pro své statistické ukazatele a nabídl tak veřejně možnost využití svých dat v jakýchkoliv produktech či aplikacích na internetu.

Eduard Durník  
Martin Černý  
oddělení výstupních databází