

Statistický metainformační systém a Sčítání lidu, domů a bytů 2011

Neoddělitelnou součástí předběžných výsledků ze sčítání jsou i doprovodné informace, tzv. **metadata**. Proč jsou potřebná nejen pro uživatele statistických výstupů, vysvětlují odborníci na Sčítání 2011.

Pro zpracování a prezentaci výsledků SLDB 2011 byl využit statistický metainformační systém ČSÚ. Koncepce vybudování statistického metainformačního systému (SMS) vznikla v ČSÚ v polovině minulého desetiletí, vlastně v době zahájení příprav SLDB 2011. Projekt SLDB 2011 na začátku počítal s využitím existujících částí SMS – číselníků a klasifikací. Postupující zavádění SMS a získávání praktických zkušeností v letech 2008 a 2009 však nakonec vedlo k rozhodnutí ČSÚ o širším využití SMS ve zpracování a prezentaci SLDB 2011.

Využití SMS

Před zahájením realizační části SLDB 2011 existovaly v ČSÚ dílčí nástroje SMS, např. aplikace pro správu klasifikací a číselníků, statistických proměnných a některých metadat statistických úloh. V oblasti metodické a obsahové byly získávány zkušenosti se systémem metainformačním popisem statistických údajů v oblasti ekonomických statistik.

Vzhledem k tomuto výchozímu stavu SMS a charakteru SLDB 2011 byla navržena koncepce

maximálního využití SMS v procesu SLDB 2011 od fáze zpracování vstupních údajů ve zpracovatelské databázi SLDB až po prezentaci výstupních údajů nástroji veřejné databáze ČSÚ.

Konkrétní využití SMS v rámci SLDB 2011 lze rozdělit do 4 oblastí: **popis údajů SLDB, kontroly a autokorekce vstupních údajů, odvozování mikroúdatů a agregace.**

Výpočet agregovaných údajů je plně řízen metadatovým zadáním, který přesně vymezuje: co má být počítáno (jaké agregáty jsou požadovány); jak to má být počítáno (z kterých mikroúdatů, pomocí jaké agregační funkce); jak má vypadat metadatový popis výsledných agregátů (tj. jaké jsou hodnoty třídících charakteristik a dalších dimenzionálních vymezení výsledných agregátů).

Přínosy a zkušenosti

I když aktivity SLDB 2011 budou dále intenzivně probíhat, je možné již nyní udělat krátké vyhodnocení některých přínosů a získaných zkušeností.

Z hlediska přínosů SMS pro SLDB 2011 stojí na prvním místě zavedení metodického přístupu a poskytnutí nástrojů pro systematické definování přesného zadání ze strany metodiků SLDB. Toto je určitě jeden z důležitých faktorů včasného dodání a kvality předběžných výsledků SLDB 2011.

Výše zmiňovaná koncepce plně automatizovaných modulů zpra-

cování dat, jejichž chování lze při opakovaném spuštění měnit pouze na základě změny metadatových popisů algoritmů, se ukázala jako klíčová během metodického ladění algoritmů kontrol a odvození, a při předpokládaném budoucím rozšiřování množiny výstupních mikroúdatů a agregátů v závislosti na dodatečných požadavcích na výstupy.

Strukturovaný metainformační popis statistických údajů SLDB 2011 v datovém skladu a veřejné databázi budou oceněny především v budoucnosti při další práci s těmito daty.

Projekt SLDB 2011 má ovšem také nesporný přínos pro zavádění SMS v ČSÚ. Řada praktických zkušeností byla získána v metodické a obsahové oblasti např. obsahové konsolidaci číselníků, popisu agregovaných údajů. Hlavním přínosem je však ověření aktivního využití metainformačních popisů při vlastním zpracování (např. pro kontroly, odvozování, agregování) a ověření procesu od definování požadavků na výstupy, přes zpracování výstupů a poskytování dat a metadat veřejné databázi.

Co je SMS?

Metadata, tj. doprovodné informace vztahující se ke sledovaným statistickým údajům, jsou potřebná nejenom pro uživatele statistických výstupů, ale pro všechny účastníky procesu tvorby statistických údajů – respondenty, správce administrativních zdrojů dat, statistiky, metodiky, IT specialisty i management. Každý z účastníků tohoto procesu je v některé fázi **tvůrcem i uživatelem metadat**.

Podstatou SMS je poskytnout nástroje pro vytváření a poskytování statistických metadat. Pod touto poměrně obecnou definicí si však nelze představovat pouze aplikace informačního systému, sloužící k zadávání či čerpání metadat. Ty tvoří pouze třetinu celé problematiky. Druhou část SMS tvoří metodická a obsahová stránka SMS, tj. detailní vymezení potřebných metadat a způsobu obsahového naplňování pro konkrétní potřeby statistiky. Praktické zkušenosti dále ukázaly, že nelze nasadit nástroje a metodiky SMS a zajistit jejich životaschopnost bez důsledného aktivního využívání metadat v procesech zpracování statistických údajů. Za třetí část SMS je proto třeba považovat oblast využívání metadat. SMS tedy není pouze nástrojem nebo částí statistického informačního systému (SIS). Jedná se o koncepci, na které je budoucí SIS podle **Globální architektury statistického informačního systému (GA-SIS)** ČSÚ založen. Využití SMS v rámci SLDB 2011 je jedním z logických kroků naplňování této koncepce.

Pavla Trendová
odbor obecné metodiky
Štěpánka Morávková
oddělení metodiky, analýz
a diseminace sčítání
Martin Bubeník
NESS Czech