

Název projektu:

Redesign Statistického informačního systému v návaznosti na zavádění eGovernmentu v ČR

Příjemce: Česká republika – Český statistický úřad

Registrační číslo projektu: CZ.1.06/1.1.00/07.06396

Příloha k zadávací dokumentaci veřejné zakázky „Integrační nástroje, vstupní a výstupní subsystém“

Příloha č. 37

Popis aktuálního stavu veřejné databáze

Název souboru: RSIS_ZD001P37_VDB_popis.pdf

Počet stran přílohy (bez tohoto krycího listu): 2

Administrace přílohy: Ing. Eduard Durník

Verze ke zveřejnění

Popis aktuálního stavu veřejné databáze

Veřejná databáze ČSÚ (VDB) je základním nástrojem procesu diseminace statistických dat.

Je budována jako jednotné úložiště agregovaných statistických dat (ne mikrodat) určených pro on-line prezentaci koncovým uživatelům, zejména externím. Součástí veřejné databáze jsou webové aplikace pro prezentaci statistických dat, a to zejména formou předdefinovaných maket výstupních objektů (tabulek, grafických objektů – grafů, kartogramů). Do těchto maket se odpovídající data dynamicky dotahují vždy až v okamžiku, kdy uživatel chce objekt zobrazit.

Součástí VDB jsou i nástroje na načítání statistických dat (ze zpracovatelských databází nebo datového skladu), ukládání do jednotného úložiště, aplikace pro správu a nástroje pro tvorbu předdefinovaných výstupních objektů (návrhář maket).

Primární úložiště statistických údajů (primární databáze) VDB je řešeno jako databázová tabulka. Základní entitou je statistický údaj, jednoznačně identifikovaný metainformačním popisem (ukazatele, číselníky), obsahuje i další doplňkové informace (atributy, např. důvěrnost dat, interval zveřejnění apod.).

Principem VDB je i to, že VDB sama nic nepočítá – pouze prezentuje dodaná statistická data včetně prezentace odpovídajících metainformací.

VDB 1

Datový model současné VDB je stále ještě orientován na původní popis údajů s využitím metainformačního systému METIS (systém UKAZATELE a CISELNIKY). Vzhledem k tomu, že METIS byl ukončen v roce 2009, byly vytvořeny analogické struktury v systému SMS-KLAS (např. číselník ukazatelů).

Ve VDB je v současné době cca 50 milionů údajů za různé odvětvové nebo průřezové statistiky. Vstup dat do VDB je zajišťován prostřednictvím definované struktury vstupního rozhraní (XML). Zpracovatelské databáze mohou generovat výstupní soubory (obsahující nezbytné metainformace) do vstupního rozhraní VDB, kde speciální aplikace tyto soubory načítá a po kontrole konzistence ukládá do primárního úložiště. Většina zpracovatelských databází ovšem nepodporuje export pro VDB, a tak vstupní data připravují pracovníci oddělení VDB (případně VPO) úpravou dostupných nebo vyžádaných souborů s daty (zpravidla XLS tabulky) a doplňováním popisných metainformací.

Vstupní data se ukládají do databáze interní, data jsou dostupná všem interním uživatelům. Synchronizační aplikace přenáší nová data, která jsou určena externím uživatelům (tedy ne důvěrná), databáze externí.

Data jsou dostupná přes webovou aplikaci, připravené výstupní objekty je možné odkazovat např. z webových stránek ČSÚ. Navigační stromy pro přístup uživatelů vycházejí z analogických stromů na webu ČSÚ případně v Katalogu produktů.

Omezení současné VDB spočívá zejména v nedostatečné aktuálnosti dat. Je to způsobeno neprovázaností na standardní proces zpracování. Řešením je nová verze VDB, která bude vstupní data čerpat z datového skladu (DWH).

Rovněž při prezentaci výstupů existují určitá omezení. Jen velmi obtížně lze současným návrhářem spojit do jednoho výstupu stavové a tokové veličiny, neexistuje možnost z jedné tabulky vytvořit různé grafy apod.

Přesto stávající VDB nabízí uživatelům stovky výstupních objektů, ve kterých uživatel může vybírat z nabízených parametrů (např. území, čas), přepínat formu zobrazení (tabulka – graf – mapa). Nad daty uloženými ve VDB jsou generovány i speciálním animované grafy (motion-charts).

VDB 2

V souvislosti se zamýšleným Redesignem statistického informačního systému byly od roku 2010 zahájeny práce na vybudování nového systému veřejné databáze.

Změnil se datový model, který je založen na nových subsystémech statistického metainformačního systému SMS (vedle již využívaného SMS-KLAS jde o využívání SMS-UKAZ, SMS-ULOHY, SMS-DISEMINACE a další). Vstupní rozhraní je navrženo tak, aby bylo schopno automatizovaně přebírat statistická data z datového skladu (bez nutnosti dodatečného doplňování informací).

Hlavní idea veřejné databáze se ovšem nemění. I nová VDB (VDB 2) je budována jako databázová tabulka, do které budou uložena agregovaná statistická data jako výsledek zpracování statistických úloh (včetně úloh pro přebírání externích dat).

Pro vstup dat a komunikaci s datovým skladem se připravuje aplikace umožňující vstup dat přes definované univerzální vstupní rozhraní (viz soubor „RSIS_ZD001P89_VDB_VSTUP_XSD“, tj. příloha č. 89, zjednodušený popis je uveden v souboru „RSIS_ZD001P88_VDB_VSTUP“, tj. příloha č. 88). Ta přebírá připravené dávky určené pro VDB z datového skladu (rozhraní DWH-VDB) včetně uložených metapopisů a atributů (primární důvěrnost, kvalita, intervaly zveřejnění apod.). Následují vlastní kontrolní a plnicí procedury včetně synchronizace dat pro externí uživatele (přesun nových dat do externí databáze).

Pro prezentaci statistických dat slouží nový návrhář výstupních objektů (NVO). Umožňuje zadat požadované kombinace statistických proměnných a jejich dimenzí (elementární proměnné) včetně grafických prvků. Tyto objekty budou tvořit páteř prezentačního systému pro on-line prezentaci statistických dat. Jsou v úzké vazbě na webovou prezentaci ČSÚ a na SMS-DISEMINACE. NVO může sloužit i pro vymezení požadavků na statistická data provázáním na SMS-VYSTUPY. Nově se uvažuje s finálními objekty (předdefinované výstupní objekty PVO) založenými na jejich elementech (elementy výstupních objektů – EVO, např. legendy, hlavičky, záhlaví...). Dojde k významnému rozšíření možnosti upravit si finální podobu výstupního objektu uživatelem (vedle výběru parametrů také skrývání řádků/sloupců, změna pořadí, rotace os, apod.).

Zcela novou aplikací se stane vyvíjený návrhář publikací, který umožňuje skládat z předdefinovaných výstupních objektů a dalších objektů výslednou publikaci (ve formátu PDF).

Pro přístup uživatelů je připravovaná zcela nová verze prezentační aplikace. Ta umožní nejen nabízet předdefinované výstupní objekty v novém navigačním stromě, ale bude podporovat související odkazy do dalších systémů DISEMINACE případně na externí objekty. Umožní uživatelům i pokládat uživatelské dotazy (na základě připravených scénářů postupu výběru dimenzí) a počítá se s rozvojem webových služeb (např. stahování větších dávek dat pro registrované uživatele). Bude podporován export dat do standardních formátů (XLS, PDF, CSV, SDMX, XML). Prezentace animovaných grafů se stane čtvrtou standardní formou prezentace.

V cílovém stavu budou aplikace VDB 2 komunikovat zejména se systémy:

- SMS – KLAS, UKAZ, ULOHY, DISEMINACE, VYSTUPY
- DWH
- SIS-DISEMINACE – zejména web a katalog produktů