

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. 21

Specifikace bloku pro transformaci struktury dat CENTRAL – IN/OUT

Název souboru: RSIS_ZD001P21_CENRAL_IN_OUT.pdf

Počet stran přílohy (bez tohoto krycího listu): 4

Administrace přílohy: Ing. Pavel Skotnica

Verze ke zveřejnění

Zadání subsystému CENTRAL – IN/OUT (C-I/O)

1. Základní charakteristika

- 1 | Nedílnou součástí procesu zpracování je nahrání dat ze standardních formátů typu .xls, .dbf, .xml, .csv, .txt do centrálního úložiště nebo také export opačným směrem do zvoleného formátu. Vedle uvedeného importování/ exportování je aplikace C-I/O určena zejména pro tvorbu zadání a posléze samotnou realizaci transformace uložených externích a administrativních dat v centrálním úložišti do standardního datového modelu pro centrální zpracování (DMCZ) nebo na druhou stranu pro opačný převod z DMCZ do zvolené struktury a zvoleného formátu. Pro převod do DMCZ je nutný metadatový popis dat a následné namapování na metadata. Převod do DMCZ je nutný pro využití standardních nástrojů pro centrální zpracování, jehož součástí je nástroj pro převod do UF a k přenosu do DWH, VDB.
- 2 | Subsystém je součástí SIS CENTRAL a je spravován a aplikován nad centrální databází. Využívá metadata ze subsystémů SMS.
- 3 | Důvodem provádění importů a exportů, resp. převodu dat není jen poskytování dat externím uživatelům, ale ČSÚ je mnohdy nuceně používat nástroje, které nejsou přímo aplikovatelné na oracleovskou databázi, a proto se musejí převádět do formátů pro ně přijatelných. Mezi činnostmi, které musí úřad zabezpečovat patří:
- sezónní očišťování (nutný převod do .xls v dané struktuře pro DEMETRu);
 - identifikace důvěrných dat (převod do .txt v dané struktuře pro tau/mí Argus);
 - kontrola a verifikace dat SASovskými makry (dodávané ze strany Eurostatu);
 - spouštění specifických maker v SASu nebo v Excelu pro zpracování odhadů vynucených ze strany Eurostatu;
 - výstupy pro Eurostat ve formátu GESMES, SDMX.
- 4 | Po využití jiných aplikací ve zpracování je nutné výstupy z externích aplikací dostat do centrální databáze a doplnit je o příslušné metainformace a převést do vhodných struktur.

2. Popis současného stavu

- 5 | V současné době neexistuje souhrnná aplikace pro import/export dat do centrální databáze (kromě importu z formátu UF) a už vůbec neexistuje nástroj pro převod do standardního datového modelu, který je využíván pro centrální zpracování.
- 6 | Existuje program pro import dat DBF tabulek, který by měl být zahrnut do aplikace C-I/O, resp. vylepšen. Odborní pracovníci využívají PL/SQL Developer pro kopírování excelovských tabulek do oracleovských databází, resp. se využívá ODBC Driverů z Excelu, Accessu apod. V případě potřeby nahrání dat z externích zdrojů (mezi ně patří AdZd) se programují aplikace na míru (což se může ukázat jako nejefektivnější cesta). Navíc současná transformace externích zdrojů neumožňuje namapování na a rozšíření dat o standardní metadatový popis a tím pádem ani import do DWH apod.

3. Cílový stav

- 7 | Cílovým stavem je mít k dispozici souhrnnou aplikaci, která v sobě integruje specifikaci (zadávání), realizaci a evidenci procesů C-I/O
- a) nahrání dat do centrálního úložiště (myšleno centrální databáze v širším slova smyslu, nejen prostory pro zpracování, ale i pro nahrání, importy...);
 - b) exportu dat z centrálního úložiště;
 - c) převod dat z centrálního úložiště do DMCZ;
 - d) převod dat z DMCZ z centrální databáze pro potřeby tvorby externích výstupu (pro bod b)) .
- 8 | Aplikace umožní vedle zadávání, realizace a jejich evidence, také tvorbu metainformací k procesům souvisejícím se správou. Dále umožní zahrnutí externích dat do Centrálního zpracování a výstupy pro Kvalitu, DWH a tvorbu souvisejících výstupů. Dalším přínosem je standardizování procesů exportů a následných importů pro využití v jiných aplikacích a SW pro statistickou činnost, jako je

např. sezónní očišťování, ochrana důvěrnosti dat a bezpečnosti, specifické analýzy a procesy požadované uživateli (ECB, Eurostat, ...).

- 9 | Aplikace poskytne pracovníkům vedle řízení, spouštění, parametrizování, sledování i plnou kontrolu nad průběhem zpracování externích dat v rámci centrálního zpracování a možnost sledovat jeho průběh, případně analyzovat proces zpracování po jeho ukončení.
- 10 | Navrhovaný systém by měl být otevřený pro případné další doplňování funkčnosti a měl by poskytovat takové rozhraní vůči svému okolí, které bude jeho schopnosti dále rozvíjet a zvyšovat komfort pro uživatele zabývající se správou zpracování a řízení úloh.

Obecné pojednání

- 11 | Výstupem projektu by mělo být navržení a později schválení aplikace, která by měla splnit níže vyjmenované funkcionality společně se zavedenými rolemi a s definovanou návazností a vazbami na další subsystémy SMS a subsystémy SIS. Aplikace má být postavena nad databází jejíž datový model pro evidenci a správu je také předmětem zadání. Dále je předmětem zadání specifikování náročnosti na výpočetní techniku a SW požadavky.

4. Funkční specifikace předmětu

4.1. Funkcionalita

- 12 | Funkcionalitu aplikace C-I/O lze rozdělit do následujících věcných bloků:

1. zadání pro IMPORT;
2. zadání pro EXPORT;
3. zadání pro TRANSFER DOVNITŘ datového modelu pro centrální zpracování;
4. zadání pro TRANSFER VEN z datového modelu pro centrální zpracování;
5. realizace zadání z výše uvedených bloků;
6. správa aplikace C-I/O;

Ad 1. IMPORT

- 13 | Aplikace umožní převést data z externích souborů do centrálního úložiště. Pro převod je možné specifikovat
- volbu pro převedení všeho ve shodných/ odpovídajících formátech zdrojového souboru s možností založení tabulky v centrální databázi;
 - nahrání dat z externího zdroje do již existující tabulky s tím, že dle názvu sloupce se provede naplnění příslušných položek/sloupců (nevyskytující se vyloučí), nebo je možné specifikovat konkrétní prolinkování sloupců, resp. provést kombinaci obého;
 - doplnění sloupce/ů o nápočty nebo doplnění sloupce/ů specifikující metainformace s napojením na jiné tabulky nebo doplnění textů, datumových položek apod. informací;
 - pro aktualizaci již existujících tabulek se provede analýza vhodnosti formátů, seznamu vyloučených sloupců.

Ad 2. EXPORT

- 14 | Aplikace umožní převod tabulek, pohledů, resp. výsledku dotazů z centrálního úložiště do zvolených formátů a zvolených souborů. Jedná se o proces v opačném směru s analogickou funkcionalitou jako u zadávání IMPORTu.

Ad 3. TRANSFER DOVNITŘ

- 15 | V Centrální databázi je vhodné importované tabulky z externího formátu/nebo prostředí převést do datových struktur a datového modelu, který je užíván v centrálním zpracování.
- 16 | Je to jiná úloha než předcházející import. Ten je určen pro nahrání dat do centrálního úložiště víceméně ve struktuře nebo logice zdrojového souboru a jeho hlavním účelem je evidence a uložení, aby bylo možné analyzovat i logiku a pojetí zdroje, které by se významnou měrou mohlo po přímém převodu do datových struktur ČSÚ vytratit.

- 17 | Zadání transferu dovnitř zpracovatelského systému si vynucuje naplnění povinných i nepovinných položek pro korektní uložení podle povahy dat. Musí být respektována struktura a model, který je uveden v příloze specifikující datový model a strukturu, to znamená, že každému transferu je nutné specifikovat
- I. pro statistické jednotky
 - a) navázání na již existující statistické jednotky (základní, resp. výběrové soubory – tabulky typu ZZ, VD, VP, VO)
 - b) nebo vytvoření nových tabulek pro statistické jednotky s jejich případnou specifikací atributů (tabulky typu ZZ, VP, VD, VO) a následnou specifikací tabulek meziproměnných, atributů, pomocných proměnných (tabulky typu AP, AZ, AD, MD, MX, případně další typy tabulek, které vzniknou díky zadání pro VYSTUPY, KVALITU, ČASOVÉ ŘADY)
 - c) přiřazení statistické úlohy, verze zpracování, běhu, identifikace statistické jednotky
 - d) + další metadata, která jsou požadována v zadání jiných subsystémů, nebo napojení na ně;
 - II. pro statistické údaje
 - a) způsob získání údaje a případná specifikace kvality údaje
 - b) referenční období, u poměrových ukazatelů případnou specifikaci základního období
 - c) VIP k příslušným položkám a k nim případnou specifikaci doplňkových objektů
 - d) stanovení důvěrnosti a ochrany údaje
 - e) příslušnost k časové řadě
 - f) + další metadata, která jsou požadována v zadání jiných subsystémů, nebo napojení na ně.
- 18 | Pozn. VIP ze subsystému ULOHY v sobě zahrnuje napojení na UKAZ, KLAS, ale i statistický popis – měřicí jednotka, ocenění, funkci, časové určení apod.
- Ad 4. TRANSFER VEN*
- 19 | Aplikace umožní převod tabulek, pohledů, resp. dotazů nad údaji a statistickými jednotkami z centrální databáze uložených v DMCZ. Věcně jde o opačný proces. Tzn. jde o nadefinování propojení statistických jednotek s údaji (VIP, ref. období,...) a o nadefinování pojmenování sloupců a typy formátů s možností nastavení formátů.
- Ad 5. Realizace zadání*
- 20 | Aplikace zajišťuje samotné spouštění procesu zpracování, při kterém lze formou menu vymezit zadání, resp. zadání lze spustit i ze zadávacího formuláře. Procesy lze pouštět v posloupnosti, ale obecně lze provádět zpracování i souběžně (paralelně). Při samotné realizaci je evidován průběh procesu zpracování – čas zahájení a čas ukončení společně s chybovostí procesu, kým byl proces zahájen a s jakým nastavením.
- 21 | Spouštění lze parametrizovat daty zahájení s možností vymezení pro celou realizaci nebo pro jednotlivé skupiny zadání, resp. jednotlivé zadání. Také je možnost spouštění zpracování pomocí batche, tedy dávkově s parametrizací posloupnosti, resp. paralelního zpracování.
- Ad 6. Správa systému*
- 22 | Každá akce uskutečněná pomocí aplikace je evidována (kdy, kde, kým a co). Umožní verzování definovaných zadání pro importy/exporty, resp. pro převody z/do datového modelu pro centrální zpracování. Dále umožní reportování o zadávání změn parametrů, nebo o samotném spouštění zpracování, resp. o aktuálním průběhu zpracování. Také bude možné sledovat naplněnost jednotlivých databázových tablespaces, resp. odhad volného databázového prostoru. Samozřejmě je možnost správy uživatelů, přidávání rolí jednotlivým uživatelům. Je nutná evidence, resp. systém na správu programových modulů, chodů a jejich skupin.
- Tzn.
- zajistit evidenci přihlášení a provedených činností umožnit dělat reporty z této evidence;

- každý objekt a atribut má evidenci kdy a kdo prováděl změnu a také má své datумы platnosti;
- verzování objektů a specifikace platnosti objektů je v souladu s pravidly systému GA SMS.

4.2. Uživatelské role

23 | Každá z výše uvedených funkcionalit má možnost nadefinování role s následujícími vlastnostmi:

- role prohlížení – bez možností provádění změn, resp. spouštění;
- role preeditační – provést změny, aktualizace, nové nastavení, ale zatím neschválené, tudíž neplatné, fakticky se jedná o návrh;
- role editační se schválením – provést změny, aktualizace, nové nastavení, které jsou platné a zároveň se mohou editovat jiné návrhy a provést případné schválení návrhu;
- role operátorská – umožní provádět spouštění zpracování s možností vymezení spouštěné úlohy;
- role správcovská – přidělení rolí uživatelům, ale i doplnění aplikací, atributů (parametrů) pro sledování, resp. nastavování metainformací souvisejících s centrálním zpracováním.

4.3. Vazby

- SMS-KLAS, SMS-UKAZ – zdroj číselníků a popisů pro naplnění /čtení metainformací;
- SMS-ULOHY – zdroj pro popis úloh, harmonogramy společně s verzemi a běhy zpracování, definování VIP;
- SMS-CENTRAL – umožnění spouštění aplikace C-I/O, resp. jejích programů;
- SIS-CENTRAL – samotná realizace nad centrální databází, ať už import/export s externími zdroji, tak samotný převod do vhodného datového modelu

Přeneseně přes SMS-CENTRAL, resp. SIS-CENTRAL jsou tvořeny výstupy pro SMS-KVALITA, VDB, DWH, DISEMINACE.

4.4. Očekávaná datová architektura

24 | Aplikace bude mít veškerá data uložena v centrální oraclovské databázi uživatelem příslušného zpracování (CSUrrrr). Veškerá data budou ukládána se všemi potřebnými metadaty a bude zajištěna kontrola jejich konzistence.

25 | Metadaty budou zadávat interní zaměstnanci ČSÚ z centra i z krajských pracovišť. Každý uživatel bude mít vlastní konto a bude mít právo editovat jen objekty, ke kterým má právo. Všichni uživatelé budou mít právo vidět (jen pro čtení - roli prohlížení) všechny objekty.

4.5. Rozhraní subsystému

26 | Vstupním rozhraním je rozhraní vytvořené v subsystému SMS-ULOHY a SMS-KLAS, popř. i SMS-UKAZ (view v Oracle DB SMS), SMS-KVALITA.

27 | Z aplikace CENTRAL bude vytvořeno výstupní rozhraní (view) pro subsystém SMS-KVALITA, SIS VSTUP, SIS CENTRAL, VDB, DISEMINACE, REGISTRY.

28 | Struktura a obsah rozhraní budou definovány dodatečně při schvalování datového modelu.

29 | Aplikace bude využívána pouze interními zaměstnanci, nebude mít definováno výstupní rozhraní pro externí uživatele.