

OPTIMALIZACE DATABÁZE KONTAMINOVANÝCH MÍST

Jaroslav Ráček¹, Tomáš Ludík¹, Jana Sedláčková²

¹ Fakulta informatiky, Masarykova univerzita

² Fakulta informačních technologií, Vysoké učení technické v Brně

Abstrakt

Příspěvek popisuje současný stav národní databáze pro evidenci kontaminovaných míst v České republice a metodiku jejího redesignu, kterou je modifikovaná metoda logického modelování zaměřující se na normalizaci datových struktur. V závěrečné části příspěvku je zmíněn pilotní projekt na území Jihomoravského kraje.

Abstract

This paper presents the current state of the national information system for the reporting of contaminated areas in the Czech Republic and the methodology of its redesign. The modified method of logical modeling focused to the data model normalization is used as the redesign methodology. There is also described the pilot project of South Moravian Region in the paper.

Klíčová slova

Databáze kontaminovaných míst; environmentální informatika; redesign informačního systému.

Key words

Database of Contaminated Areas; Environmental Informatics; Information System Redesign.

1. ÚVOD

Systém evidence kontaminovaných míst (SEKM) je integrovaný databázový systém sloužící k evidenci lokalit postižených kontaminací zemín, stavebních konstrukcí nebo půdního vzduchu a podzemních či povrchových vod. Databáze eviduje i místa s kontaminací dosud neověřenou. Datový obsah současné databáze je ve finální podobě uložen na serveru České informační agentury životního prostředí (CENIA), kde je integrován do celého komplexního systému geoportálu životního prostředí ČR, a dále na serveru Ministerstva životního prostředí České republiky (MŽP), kde je dostupný v rámci LAN. Všechny jeho součásti mohou využívat pracovníci MŽP a CENIA na základě udělení oprávnění administrátorem či správcem systému. Odborná veřejnost má k většině údajů databáze přístup na webových portálech MŽP a CENIA prostřednictvím aplikace MapServeru CENIA. Vlastníkem tohoto databázového systému je MŽP.

2. SOUČASNÝ STAV DATABÁZE KONTAMINOVANÝCH MÍST

Integrovaná databáze kontaminovaných míst je v dnešní podobě poměrně složitě strukturována a rozdělena do tří datových účelových skupin. Jsou to vlastní SEKM, dále dříve samostatná databáze skládek pocházející od České geologické služby (ČGS) a databáze pro klasifikaci priorit sloužící pro rozhodování při přidělování finančních prostředků na sanace ekologických zátěží.

Provádění změn, jako je například přidání či odstranění záznamu, je možné povolit pouze oprávněným, odborně fundovaným a dobře zaškoleným uživatelům, protože hodnota dat je vskutku vysoká a cena sanace některých lokalit se pohybuje v řádech miliard korun. Na základě provozních zkušeností bylo rozhodnuto o zabezpečeném provozu databáze tak, aby

byla zajištěna její bezproblémová informační funkčnost nelimitovaná vnějšími zásahy a současně byla garantována i vysoká kvalita vstupních dat. Dalším důvodem pro vytvoření samostatných skupin (buď vzájemně kompatibilních a tím v hlavních rysech snadno integrovatelných), byla potřeba kontinuálního rozvoje a uzpůsobování databáze momentálním požadavkům a potřebám bez omezení její dosavadní funkčnosti.

Koncepce SEKM [1] předpokládá kontinuální sběr dat o kontaminovaných místech v České republice a jejich jednotné shromáždění v jediné univerzální geodatabázi. Takto formulovaná koncepce klade požadavky na funkce databáze, jimiž jsou především:

- unifikace a standardizace dat z různých datových zdrojů,
- integrace dat různých typů,
- aktuálnost dat,
- archivace dat,
- publikace a prezentace obsahu,
- výběr a export dat pro další interpretaci.

Pro zajištění výše uvedených funkcí je koncipováno řešení, které dělí celou problematiku správy integrované databáze SEKM na dvě logické oblasti, jimiž jsou:

- statická oblast (centrální datový sklad),
- dynamická oblast (manipulační datový sklad).

Statická oblast je prezentační částí databáze, která navenek představuje aktuální stav databáze po provedených kontrolách.

Dynamická oblast (oblast změn) je provozní částí databáze, zachycující její vývoj a zajišťující archivaci dat (i neaktuálních). Fyzicky je tato část tvořena samostatně provozovanými databázemi, které nemohou být ovlivněny zásahy neoprávněných uživatelů, a je spravována v oddělené počítačové síti od prezentační oblasti.

Obě části (statická a dynamická) komunikují pomocí jednorázové dávkové předávky standardně strukturovaných dat. Touto dávkou je po následných kontrolních rutinách aktualizována centrální databáze celkovým nahrazením všech údajů. Jedná se prakticky vždy o kompletní výměnu dat centrální databáze v prezentační oblasti.

Dle této koncepce s přihlédnutím k následnému vývoji a potřebám je SEKM rozdělen do několika fyzicky oddělených funkčních celků se shodnými hlavními klíčovými poli a položkami. Tyto celky se navzájem doplňují, avšak nemohou změnit svůj obsah vlivem změn v jiném celku. Data mezi těmito celky se mění vždy s vědomím příslušných správců pomocí exportních a importních rutin. Hlavními funkčními provozními celky jsou:

- centrální databáze SEKM,
- databázový systém pro klasifikaci priorit kontaminovaných míst,
- manipulační datový sklad.

3. PROJEKT REINŽENÝRINGU SEKM

Z výše uvedeného popisu vyplývá, že současný stav databáze není z informatického hlediska nijak ideální. V důsledku historického vývoje byly do celého systému integrovány nehomogenní (dříve samostatné) části veřejných informačních systémů a nad nimi vytvořeny různé aplikace, které vycházely spíše z administrativních rozhodnutí úředníků než z principů softwarového inženýrství. Toto se projevilo v nedostatečné propojenosti jednotlivých částí, kdy je řada dat přenášena poloautomaticky nebo téměř manuálně, ačkoli by toto šlo plně automatizovat nebo data sdílet. Přestože se jednotlivé záznamy v různých částech systému podařilo uvést do vzájemně kompatibilní podoby, z pohledu kompatibility vůči ostatním informačním systémům soukromého ale i veřejného sektoru, je současný stav nevyhovující.

Záznamy SEKM se mezi sebou rovněž výrazně liší z pohledu kvality a důvěryhodnosti dat. Jsou zde patrné značné rozdíly mezi záznamy pořizovanými v různých časových obdobích, ale i na různých územních celcích.

Z těchto důvodů zahájilo MŽP v roce 2007 řešení výzkumného projektu s názvem „Zhodnocení struktury stávající databáze starých ekologických zátěží, definování kritérií pro hodnocení jejich vlivu na ŽP a pro stanovení priorit jejich odstraňování s důrazem na brownfields“. Řešitelem projektu je pražská společnost DHV ČR a spoluřešitelem zodpovědným za infromatickou stránku projektu je Fakulta informatiky MU v Brně.

Celý projekt nabývá na svém významu zejména v souvislosti s plánovanou inventarizací kontaminovaných míst, kterou připravuje MŽP na rok 2009. V rámci této inventarizace budou do stávající databáze SEKM doplňovány nové kontaminované lokality a následně stanoveny priority jejich sanace. Vedle dat, která budou pořízena během terénního průzkumu podezřelých míst, budou do databáze importovány i údaje z řady informačních systémů různých organizací, z nichž nejvýznamnější jsou Česká geologická služba, Ministerstvo průmyslu a obchodu, Český hydrometeorologický ústav, správy povodí nebo armáda ČR. Z podniků, které do systému budou rovněž vkládat data, lze uvést například České dráhy, České aerolinie, Czechinvest, správy jednotlivých letišť nebo sítě čerpacích stanic.

4. METODIKA OPTIMALIZACE SEKM

Pro posouzení a zlepšení stávající struktury systému [2] byly stanoveny tři základní oblasti, kterými se řešitelský tým zabývá. Jsou to:

- optimalizace vnitřních datových struktur a nad nimi pracujících funkcí,
- posouzení stávajícího obsahu dat a návrh validačních kritérií nových dat,
- vytvoření konverzních můstků pro zajištění kompatibility s ostatními informačními systémy.

Prvním krokem při modifikaci systému je úprava jeho datových struktur a následný návrh a implementace funkcí, které nad těmito strukturami pracují. Pravděpodobně nejdůležitějším úkolem je upravit datový model tak, aby z něj byly odstraněny duplicity datových atributů a odvoditelné údaje, které se v průběhu vývoje v modelu objevily. Z tohoto pohledu je klíčovou činností normalizace datového modelu vedoucí alespoň k dosažení 3. normální formy. Vzhledem ke skutečnosti, že se jedná o systém postavený již v devadesátých letech na principech třívrstvé architektury a na relačním databázovém modelu, přistoupili řešitelé k využití modifikované formy již poměrně staré metody softwarového inženýrství, kterou je *logické modelování* [2], [4]. Kroky modifikované metodiky spočívají v tom, že stávající stav SEKM slouží jako výchozí model systému, čímž odpadají úvodní kroky původní metodiky, jimiž je tvorba prvotního funkčního modelu a následné odvození základních datových elementů. Tyto modely byly získány reverzním inženýrstvím SEKM pomocí nástroje CASE Studio 2. Další kroky jsou již v souladu s originální metodikou logického modelování a spočívají v normalizaci datového modelu a dosažení třetí normální formy, následné vytvoření logického procesní modelu a jeho implementace.

Při posuzování stávajícího obsahu databáze z pohledu vyplnění jednotlivých atributů a celkové kvality dat se řešitelé opírají o statistické metody. Data jsou analyzována pomocí vícedimenzionálních statistických metod a porovnávána s vybranými referenčními územími, u nichž se předpokládá, že jejich data jsou v pořádku. Následně se stanoví období a regiony, z nichž data nedosahují potřebnou kvalitu nebo zcela chybí. Poté jsou příslušné orgány místní samosprávy žádány, aby data doplnily.

Časově nejnáročnějším úkolem je zajištění kompatibility s okolními systémy. MŽP vyvinulo pro hlášení nově zjištěných ekologických hlášení souhrnný formulář, který se

v praxi používá buď v tištěné podobě nebo jako tabulka aplikace MS Excel. V rámci projektu je vyvíjeno webové rozhraní, které odpovídá tomuto formuláři a umožňuje přímý zápis do SEKM. Dále jsou vytvářeny konverzní můstky mezi SEKM a informačními systémy, které budou sloužit jako zdroje dat. V úvodní fázi se jedná o konverzní můstky mezi SEKM a systémem magistrátu města Brna, dále systémy provozovanými CENIA a daty společnosti Czechinvest.

5. PILOTNÍ PROJEKT NA ÚZEMÍ JIHOMORAVSKÉHO KRAJE

Výše popsané kroky nejsou aplikovány na celé země České republiky. K tomu dojde až v roce 2009 při celoplošné oficiální inventarizaci kontaminovaných ČR. Pro rok 2008 si řešitelé v rámci přípravy na tuto inventarizaci stanovili za cíl provést pilotní inventarizaci ekologických zátěží a reinženýring informační infrastruktury na území Jihomoravského kraje. Jedná se konkrétně o okresy Blansko, Brno-město, Brno-venkov a Znojmo. Tyto lokality nabízejí všechny základní typy krajiny od ryze městské aglomerace na území města Brna až po zachovalou přírodní krajinu v okolí Národního parku Podyjí. Dále se jedná o území, na němž jsou zastoupeny všechny typy ekologických zátěží, což nabízí možnost statistického srovnávání kvality obsahu a množství dat s ostatními regiony v ČR.

6. ZÁVĚR

Příspěvek popisuje část řešení výzkumného projektu jehož cílem je zhodnocení a optimalizace existujícího systému veřejné správy. Na základě svých dřívějších zkušeností si dovoluují autoři konstatovat, že se jedná o názorný příklad rezortního IS provozovaného některým ministerstvem v uplynulých deseti letech. Na celkové architektuře systému se výrazně projevíly dřívější požadavky úředníků, které vedly k nesčetným modifikacím a slučováním s dalšími částmi. Vznikl tak velmi složitě strukturovaný celek, jehož správa je navíc omezoována řadou legislativních nařízení. Má-li být systém po svém redesignu nasazen zpět do praxe, bude muset být výsledné řešení kompromisem mezi ideálním stavem z pohledu informačních struktur a mezi stavem, který požadují legislativní předpisy. Právě z tohoto důvodu probíhá nejdříve redesign datových struktur, který má za úkol připravit kvalitní půdu pro následné budování aplikační a prezentační vrstvy systému.

7. LITERATURA

- [1] Blahutová, M., Ráček, J.: Zhodnocení struktury stávající databáze starých ekologických zátěží - zpráva o řešení projektu v roce 2007, DHV, Praha, 2007.
- [2] Gane, C., Sarson, T.: Structured system analysis: Tools and techniques. Prentice-Hall, Englewood Cliffs, 1979.
- [3] Král, J.: Informační systémy. Science, Brno, 1998.
- [4] Ráček, J.: Strukturovaná analýza systémů, Masarykova univerzita, Brno, 2006.

Příspěvek vznikl za podpory projektu SP/4h4/168/07 „Zhodnocení struktury stávající databáze starých ekologických zátěží, definování kritérií pro hodnocení jejich vlivu na ŽP a pro stanovení priorit jejich odstraňování s důrazem na brownfields“.