

DATABÁZOVÁ PODPORA EVIDENCE KONTAMINOVANÝCH MÍST NA ÚZEMÍ ČESKÉ REPUBLIKY

Jaroslav Ráček¹, Tomáš Ludík¹, Jan Ministr²

¹ Fakulta informatiky, Masarykova univerzita

² Ekonomická fakulta, VŠB – Technická univerzita Ostrava

ABSTRAKT

Příspěvek popisuje současný stav národního informačního systému pro evidenci kontaminovaných míst v České republice. Na základě výsledků analýzy národních datových zdrojů byl v roce 2008 navržen nový datový model pro jednotné ukládání dat kontaminovaných míst. V závěrečné části příspěvku jsou popsány způsoby reprezentace geografických dat v nové databázi.

ABSTRACT

This paper presents the real state of the national information system for evidence of contaminated areas in the Czech Republic. On the base of the results of analysis of national data sources the new unified data model for contaminated sites data has been developed in 2008. There are also described formats of representation of geographical data in the new database 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

Tento článek navazuje na loňský příspěvek autorů [5], v němž byl představen projekt Ministerstva životního prostředí (MŽP) č. 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“ a v němž byly popsány výzkumné cíle autorů vytčené pro rok 2008. V letošním příspěvku autoři popisují své výsledky z roku 2008 a z prvního čtvrtletí roku 2009. Jedná se zejména o zmapování struktur jednotlivých databází, které v Česku evidují ekologické zátěže, analýzu jejich nedostatků a potřeb rozšíření a prvotní návrh nové jednotné datové struktury, která by měla nahradit (sjednotit) dosavadní systémy.

2. SOUČASNÝ STAV DATABÁZE KONTAMINOVANÝCH MÍST

Evidence kontaminovaných míst České republiky je v současnosti složitě strukturovaný systém. Nejčastěji uváděné dělení systému je na část označovanou jako SEKM (Systém evidence kontaminovaných míst) a PKM (Priority kontaminovaných míst) [1].

Z pohledu implementace systému však existuje jiné rozdělení, které vyplývá ze způsobu, jak jsou data uvnitř systému zpracována. Opět se jedná o rozdělení na dvě části, z nichž první je Manipulační datový sklad (MDS) chápáný jako dynamická část systému a druhou částí je Centrální datový sklad (CDS), který představuje statickou část. Oba tyto celky jsou umístěny ve vzájemně nezávislých sítích a komunikují pomocí jednorázové předávky standardně strukturovaných a ověřených dat, a to vždy s vědomím příslušných správců.

Manipulační datový sklad je struktura samostatně provozovaných databází a databázových aplikací sloužící pro příjem, aktualizaci a kontrolu vstupních dat v systému. Naprostá většina operací s tady vedoucí k jejich změně se v systému provádí právě zde. Poté, co jsou data zpracována do podoby splňující požadovaná kritéria na kvalitu a strukturu, dochází k jejich finálnímu zápisu do systému. Manipulační sklad tedy slouží jako brána pro data ucházející se o začlenění do hlavní databáze. Již dříve zavedená cenná data jsou tak bezpečně oddělena od jakýchkoliv manipulací, které by mohly znamenat poškození či jiné nechtěné změny v centrální databázi. Hlavní operace které se provádí s daty v manipulačním skladu jsou:

- kontrola formální správnosti (struktura a úplnost) přijímaných dat ze strany anotátorů a případné drobné opravy či zaslání dat zpět s žádostí o opravu nebo doplnění,
- unifikace a standardizace dat z různých zdrojů a různých formátů,
- integrace dat různých typů (tabulkové a GIS formy dat),
- aktualizace zastaralých a importování nových dat,
- aktualizace centrálního databázového skladu jednorázovými datovými předávkami,
- evidence přijatých a zpracovaných dat,
- archivace (i neaktuálních) dat a evidence změn dat i strukturálního vývoje databáze.

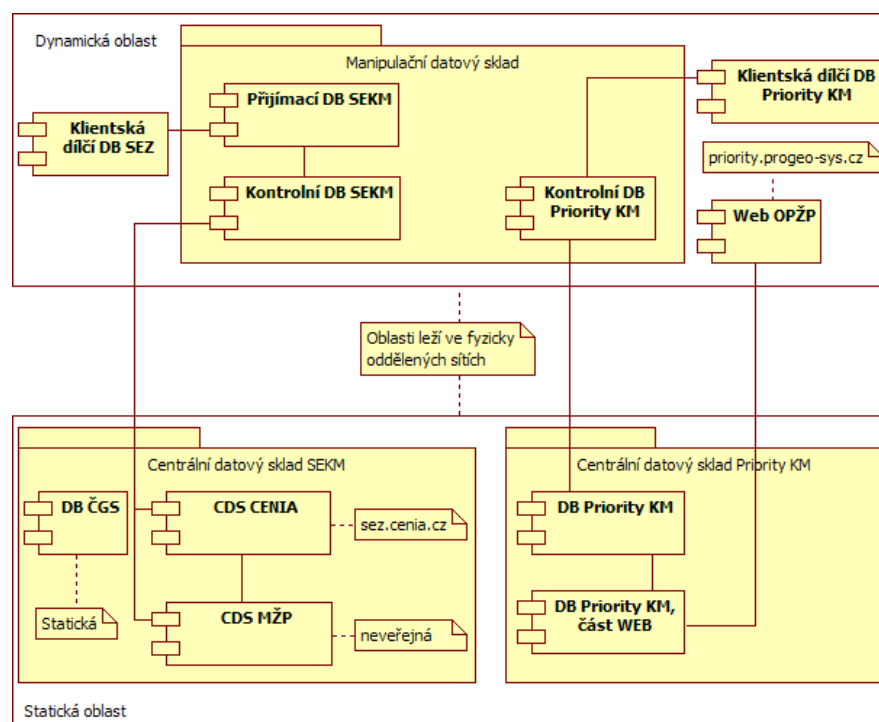
MDS se dále člení na přijímací databázi SEKM, kontrolní databázi SEKM a kontrolní databáze Priority KM. Přijímací databáze SEKM pouze přijímá data od klientských aplikací. Kontrolní databáze SEKM naopak slouží pouze pro konečnou komunikaci s centrálním datovým skladem, jsou zde prováděny pouze kontrolní rutiny odhalující nekonzistenci nebo nesrovnalosti v datech, případné opravy jsou prováděny v přijímací databázi SEKM. Do kontrolní databáze Priority KM se ukládají data pořízená klientskými aplikacemi Priority KM a po následné kontrole jsou zaslána do centrálního datového skladu.

Centrální datový sklad je úložiště, do něhož se importují už jen data, jež prošla kontrolou kvality a struktury v manipulačním datovém skladu. Import dat do centrálního datového skladu se provádí jednorázovým dávkovým přepisem celé databáze, tedy kompletním nahrazením novými daty. Struktura CDS kopíruje logické dělení celého systému, obsahuje tedy části SEKM a Priority KM.

Centrální databáze SEKM je fyzicky umístěna na serverech České informační agentury životního prostředí (CENIA) a zároveň v místní síti MŽP. Databáze SEKM v počítačové síti MŽP je neveřejná (přístupná jen ze sítě MŽP) a obsahuje i další data, která se v jiných částech systému nevyskytují (např. ekologické audity, reporting, detailní grafické vrstvy). Databáze SEKM na serverech agentury CENIA je veřejná a slouží zároveň jako jeden ze zdrojů dat mapové aplikace na Portálu veřejné správy.

Centrální část databáze Priority KM je rovněž umístěna v počítačové síti MŽP. Skládá se ze dvou dílčích databází. Část „Priority KM“ slouží pro integraci dat z klientských aplikací a druhá databáze „Priority KM část WEB“ je určena pro příjem dat z dotazníku na internetových stránkách. Data z obou databází se slučují do databáze Priority KM a v této jsou

pak data podrobena samotnému hodnocení (klasifikaci) v procesu určování priorit odstraňování ekologických zátěží. Celkové schéma systému ukazuje následující obrázek.



Obrázek 1: Struktura systému evidence kontaminovaných míst v ČR.

3. NÁVRH JEDNOTNÉ DATOVÉ STRUKTURY

Z analýzy [3] struktury a funkcionality systému evidence kontaminovaných míst vyplynulo, že se jedná o velmi složitou a nevhodně strukturovanou architekturu. Datové modely jednotlivých částí systému sice v minulosti vyšly ze společného základu, v současnosti však již díky vlastnímu vývoji nejsou totožné, což způsobuje problémy s kompatibilitou. Identifikace duplicit, které se v datech vyskytují, je pracná a vyžaduje manuální zpracování dat.

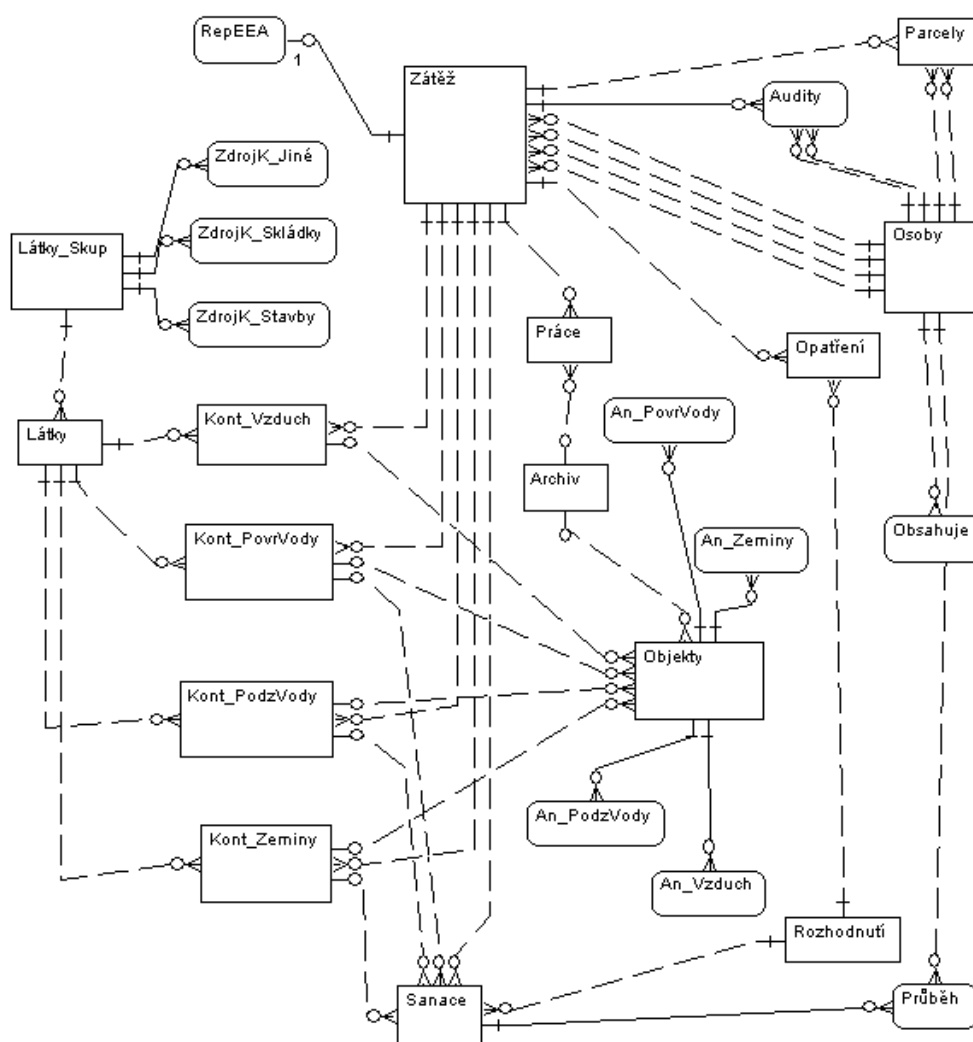
V datových modelech se vyskytují porušení normálních forem. Implementace jednotlivých částí systému je na velmi rozdílné a místy zastaralé úrovni. To vše jsou argumenty, které podporují záměr zásadního reinženýringu stávající struktury systému [2], jehož cílem bude vytvoření jednotného datového modelu a následná implementace odpovídající současným trendům a standardům v oblasti environmentální informatiky.

Dalším aspektem, který bylo třeba při návrhu nové datové struktury zohlednit, byla možnost automatické nebo poloautomatické konverze dat z dalších datových zdrojů, které v ČR existují. Z toho důvodu byla při návrhu zohledněna i některá partikulární řešení z jiných informačních systémů. Konkrétně byly v rámci pilotního projektu probíhajícího v roce 2008 vzaty v úvahu datové struktury dalších patnácti informačních systémů (např. Generel geologie Magistrátu města Brna, Informační systém odpadového hospodářství Výzkumného ústavu vodohospodářského v Praze nebo LandUse vrstva GIS Integrovaného záchranného systému Jihomoravského kraje).

Návrh nového datového modelu si proto kladl za cíl:

- sjednotit informační domény databází SEKM i Priority KM,
- podpořit kompatibilitu s dalšími informačními systémy,
- normalizovat datový model alespoň podle třetí normální formy,
- sjednotit způsoby geografické reprezentace ploch,
- zavést nové číselníky tam, kde je to potřeba,
- zlepšit popis sémantik jednotlivých entit a jejich atributů a definovat kardinality vazeb,
- rozšířit a zefektivnit vypovídací schopnost datového modelu tak, aby lépe odrážel skutečný svět,
- zachovat co největší kompatibilitu s původními datovými modely (SEKM a Priority KM) při co nejnižší nutnosti přepracování záznamů.

Jelikož část SEKM je vývojově starší a část Priority KM vznikla vypuštěním nebo agregací některých datových položek původního modelu SEKM a přidáním nových atributů pro stanovení priorit, vycházel návrh z datového modelu SEKM a jeho úpravy byly vedeny tak, aby byl nový model kompatibilní i s Priority KM. Základní členění nově navrženého datového modelu je na následujícím obrázku.



Obrázek 2: Návrh nového datového modelu systému evidence kontaminovaných míst v ČR.

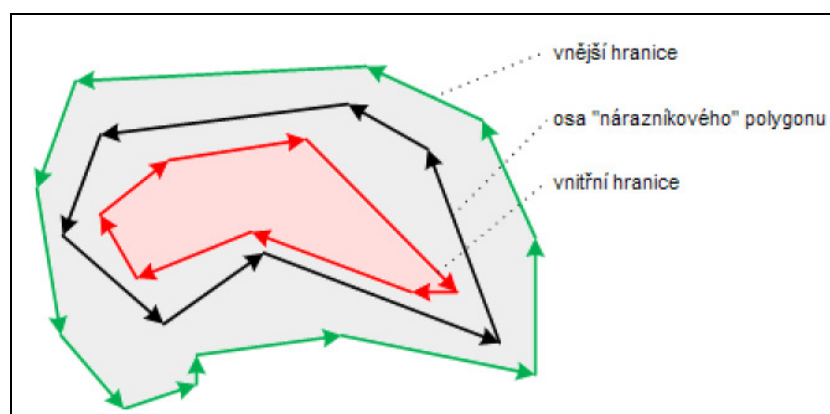
Datový model na obrázku 2 ukazuje, jakým způsobem jsou vzájemně uloženy informace o ekologických zátěžích (lokality) a objektech na těchto lokalitách, kde dochází k odběru vzorků půdy, vody a v některých případech i vzduchu. Odebrané vzorky jsou analyzovány v laboratořích a naměřené hodnoty chemických látek jsou rovněž uloženy do databáze. Systém dále eviduje i jednotlivé sanace na lokalitách včetně administrativních opatření, která k nim vedla.

4. GEOGRAFICKÁ REPREZENTACE LOKALIT

Jedním z významných úkolů při návrhu nové datové struktury bylo určit, jakým způsobem má být uložena geografická reprezentace kontaminované lokality. Zatímco v původní databázi SEKM bylo částečně podporováno ukládání lokalit formou ohraničujících polynomů, ve většině dalších analyzovaných databází byly ukládány pouze souřadnice jednotlivých bodů, kde se odebíraly vzorky, takže lokalita nebyla reprezentována plochou, ale byla tvořena shlukem bodů. Nová datová struktura podporuje ukládání obou způsobů geografické reprezentace.

Celou situaci dále komplikuje fakt, že pro řadu míst nelze jednoznačně říct, zda jsou kontaminována (tj. jsou na nich překročeny limity některé ze sledovaných chemických látek), nebo nikoli. Míra kontaminace jednotlivých míst se zároveň může v čase měnit.

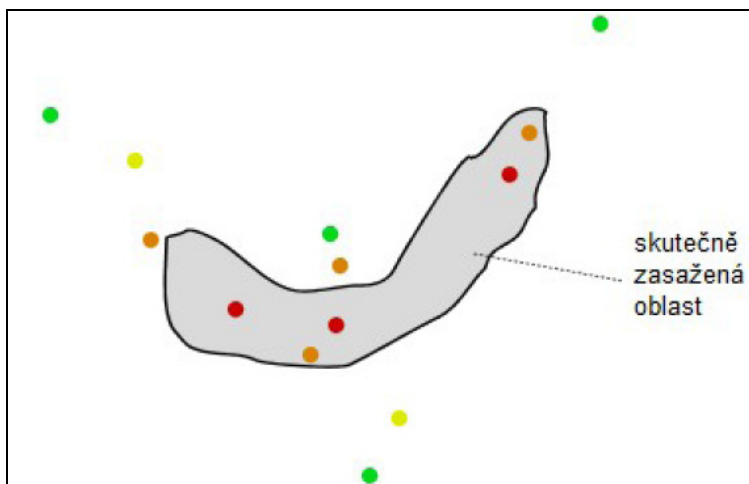
V případě, kdy je lokalita reprezentována polygony, podporuje datový model uložení dvou polygonů, z nichž jeden určuje hranici oblasti, o níž lze říci, že je zcela jistě kontaminovaná (vnitřní hranice), a druhý určuje hranici oblasti, která již s jistotou kontaminována není (vnější hranice). Skutečná hranice kontaminované plochy se pak nachází někde mezi vnitřní a vnější hranicí. Z důvodu kompatibility se staršími daty, kde je zaznamenán pouze jeden ohraničující polygon, podporuje datový model tzv. „nárazníkový“ polygon, který představuje předpokládanou hranici kontaminované plochy, nicméně tuto jedinou hranici je v praxi téměř nemožné s jistotou stanovit. Ke každé lokalitě může v případě, že ekologická zátěž tvořena nesouvislou plochou, existovat více polygonů. Aby bylo možné vytvářet i plochy uvnitř kterých jsou „ostrůvky“, je u každého polygonu uveden údaj, zda je kontaminovaná plocha uvnitř, nebo vně ohraničeného území. Ke každému polygonu je zároveň uvedeno datum jeho platnosti.



Obrázek 3: Příklad lokality reprezentované polygony.

V případě reprezentace lokality pomocí shluku bodů, je u každého bodu uvedena pravděpodobnost kontaminace danou látkou. Existují tak body, kde byla kontaminace

jednoznačně potvrzena nebo vyvrácena (např. provedením kontrolního vrtu) a body, kde se kontaminace předpokládá s určitou pravděpodobností (např. na základě expertního odhadu nebo matematického modelu). Kontaminovanou plochu pak lze rekonstruovat z těchto bodů (např. pomocí Voronoiho diagramů a Delauneyovy triangulace). U každého bodu se opět uchovávají i časové údaje, kdy byl kontaminován a kdy ne.



Obrázek 4: Příklad lokality reprezentované body.

5. ZÁVĚR

Nový datový model, který byl popsán v tomto článku, by měl být v letech 2009 a 2010 implementován a měl by se stát základem nového informačního systému, který bude zabezpečovat připravovanou národní inventarizaci kontaminovaných míst. Vedle vlastní správy dat o kontaminovaných místech se řešitelský tým z Masarykovy univerzity zabývá i vývojem konverzních můstků pro automatickou nebo poloautomatickou konverzi dat ze signifikantních datových zdrojů v ČR [4]. Dále jsou připravovány nástroje pro automatickou identifikaci a odstraňování duplicit v datech. Při spuštění národní inventarizace by proto příslušné pověřené orgány veřejné správy měly mít k dispozici zcela novou, komplexní a technologicky vyspělou informační podporu.

6. LITERATURA

- [1] Blahutová, M., Ráček, J.: Zhodnocení struktury stávající databáze starých ekologických zátěží - zpráva o řešení projektu v roce 2007, DHV, Praha, 2007.
- [2] Král, J.: Informační systémy. Science, Brno, 1998.
- [3] Ráček, J.: Strukturovaná analýza systémů, Masarykova univerzita, Brno, 2006.
- [4] Ráček, J., Ludík, T.: Analýza heterogenních datových zdrojů a jejich konverze do jednotného formátu. 5. Letní škola aplikované informatiky. Masarykova univerzita, Brno, 2008.
- [5] Ráček, J., Ludík, T., Sedláčková, J.: Optimalizace databáze kontaminovaných míst. Tvorba softwaru 2008. VŠB-TU, Ostrava, 2008.

Příspěvek vznikl za podpory projektu MŽP ČR č. SP/4h4/168/07a výzkumného záměru MŠMT ČR č. MSM0021622418.