

MOŽNOSTI VYUŽITÍ INFORMAČNÍCH SYSTÉMŮ NEMOCNIC K HLÁŠENÍ POPULAČNÍCH DAT – ZKUŠENOSTI MASARYKOVA ONKOLOGICKÉHO ÚSTAVU

ROLE OF HOSPITAL INFORMATION SYSTEMS IN POPULATION-BASED CANCER REGISTRIES – THE EXPERIENCE OF MASARYK MEMORIAL CANCER INSTITUTE

BRABEC P.^{1,2}, KLIMEŠ D.², ANDRES P.¹, DUŠEK L.², VYZULA R.¹

¹ MASARYKŮV ONKOLOGICKÝ ÚSTAV, BRNO

² LÉKAŘSKÁ FAKULTA MASARYKOVY UNIVERZITY, BRNO

Souhrn

Nemocniční informační systémy (NIS) procházejí neustálým vývojem, který ale mnohdy nereaguje na aktuální potřeby nemocnic. To je způsobeno především složitostí problematiky a dlouhodobým sběrem neparаметrických dat, nedostupných pro retrospektivní audity a analýzy. Svůj díl viny má i častá implementace dodatečných modulů, které nakonec z NIS vytvářejí heterogenní prostředí a komplikují jednoduchou analýzu a reporting. Těmto problémům je nutné předcházet. Pokud nebudou data jednotlivých pracovišť alespoň částečně parametrizovaná, nebude v budoucnu možné nad nimi provádět analýzu. Nebude tak možné rozhodovat o efektivním řízení léčby, vyhodnocovat ekonomickou situaci a provádět jakýkoli benchmarking. V současné době je v ČR velice málo nemocnic, které by skutečně měly svá data „pod kontrolou“ tak, aby dokázaly pohotově argumentovat přímo čísly. Tato práce krátce představuje jedno z efektivních řešení dané situace. Jde o datový sklad (Data Ware House) a na něj navazující reportovací systém vyvíjený na Masarykově onkologickém ústavu v Brně. Jedním z viditelných benefitů tohoto vývoje je možnost automatizace hlášení do populačních registrů, včetně Národního onkologického registru.

Klíčová slova: nemocniční informační systém, datový sklad, onkologie.

Summary

Hospital information systems represent very dynamic segment of IT market, however not always coupled with demands of concrete health care facility. It is evidently a consequence of complexity of the situation and general lack of parametric data retrospectively accessible for analytical processing. The processing and reporting is also complicated by rather partial solution of IT problems including supplementation with separated software modules. We should precede these problems as parametric data is indispensable for any optimization of therapy, cost-effectiveness evaluation and benchmarking as well. There are only a few hospitals in the Czech Republic that fully control their data and are able to argue with parametric and rational reports. This paper introduces one of effective solutions for the situation, i.e. data warehouse and associated reporting that has been implemented in the Masaryk Memorial Cancer Institute in Brno. One of straightforward and visible benefit of the solution is automated reporting to population – based registries, including Czech National Cancer Registry.

Key words: hospital-based information system, data warehouse, oncology.

Úvod

Z běžné praxe víme, že moderní nemocnice se dnes neobejdou bez elektronického informačního systému (NIS). V devadesátých letech nastal rozmach „komputerizace“, který zasáhl snad všechna odvětví lidské činnosti. Nemocnice nebyly výjimkou a budovaly se různé informační systémy, které měly usnadnit orientaci ve sbíraných parametrech. Cílem těchto systémů je usnadnit běžnému lékaři přehled o pacientovi a jeho onemocnění, hodnocení nákladnosti a efektivnosti léčby a další analýzy. Při zamyšlení nad těmito cíly si každý z nás musí položit otázku „Povedlo se nám to?“. A odpovědi nejsou jednoznačné. V běžné klinické praxi je dnes samozřejmostí, že lékař zná základní údaje o pacientovi poté, co zadá do svého nemocničního informačního systému rodné číslo

pacienta. Současně je schopen zjistit, kdy a jak byl pacient léčen, jaká mu byla předepsána medikace, kdy má přijít na kontrolu. Před jistou dobou, kdy jsme byli všichni zvyklí na papírové karty, by něco podobného vypadalo jako nemožné. Z tohoto pohledu byl jistě hlavní cíl naplněn. Přesto stále zůstává otevřena otázka sledovanosti různých parametrů a jejich využitelnosti pro analýzy dat. Řada parametrů není adekvátně zpřístupněna a zaniká v textových poznámkách a především klinický popis stavu pacienta tak mizí z následných analýz. Jinými slovy zjišťujeme, že nestačí pouze data sbírat, ale je nutné sledovat také jejich budoucí možný informační přínos. Tento článek není diskuse o tom, zda je tak či onak správně, nebo zda některá nemocnice dělá něco lépe než jiná. Problém využitelnosti dat a jejich ceny byl již v minulosti

mnohokrát popsán v různých publikacích. Cílem tohoto článku je poukázat na jednu z možných cest, jak získat kvalitní využitelné informace o pacientech a chodu nemocnice, a přitom ještě při relativně nízkých nákladech. Je logické, že dosáhneme-li tohoto cíle, můžeme automatizovat veškerá následná hlášení, která z takových dat vycházejí. Pro českou onkologii to představuje především velmi lákavou a efektivní automatizaci hlášení do Národního onkologického registru.

Popis současného stavu

Problémem dnešních nemocnic není nedostatek informací, ale jejich **využitelnost, nákladovost** jejich pořizování a možnost **integrace**. Jedním z typických příkladů je i hlášení NOR, které je dáno zákonem a nařizováno ministerstvem zdravotnictví. ÚZIS v letošním roce zavedl pilotní projekt, ve kterém je hlavním cílem integrovat data z hlášení NOR do jednotných datových struktur v NIS u onkologických pracovišť. To je jistě rozumná myšlenka, která ale otevřela řadu problémů souvisejících více se stavem NIS než se stavem NOR. Vyjdeme-li z předpokladu, že hlášení NOR má danou datovou strukturu a unifikovanou metodiku, pak hlavní problém spočívá v otázce, kde získat data a jak je napojit do centrální databáze. U parametrických NIS by vlastně žádné problémy nebyly – řešila by se pouze technologická otázka, jak data pro NOR vyexportovat a z jakých zdrojů. To ale bohužel není případ české onkologie, která musí nejprve řešit, jak exportovatelná data do NIS dostat, kde je sbírat a také kolik celý tento projekt bude vlastně stát peněz.

Dalším podobným problémem může být otázka provázanosti parametrů mezi systémy, např. „Kolik máme v každém měsíci unikátních pacientů ve třetím klinickém stadiu, s jakou diagnózou, jakou mají nejčastější léčbu, jaké jsou s nimi spojené náklady a jaký je průměrný počet lůžkodnů u těchto pacientů?“ Z pohledu hodnocení péče jde o zcela základní a vlastně ještě jednoduchou otázku. Z pohledu IT je to ale velice komplikovaný dotaz, na který dokáže málokdo odpovědět během krátké chvíle. Z praxe víme, že pokud systémy, a to nejen NIS, nejsou provázané, nejsme schopni adekvátně odpovědět na dotaz kombinující různé datové zdroje. Dotaz skončí jako analytické zadání a čeká se týden – dva, než ho někdo zodpoví na základě manuálně vytažených a zkombinovaných čísel. Přitom můžeme narazit i na problém, že některý potřebný parametr není v NIS či jiném systému vůbec parametricky veden a jeho vyhodnocení tedy vyžaduje procházet textové záznamy. Ze dvou týdnů se tak stává úkol i na měsíc nebo delší období.

Dle zkušeností různých pracovišť v ČR patří mezi nejčastější problémy:

- ☑ výrazná heterogenita informačních systémů,

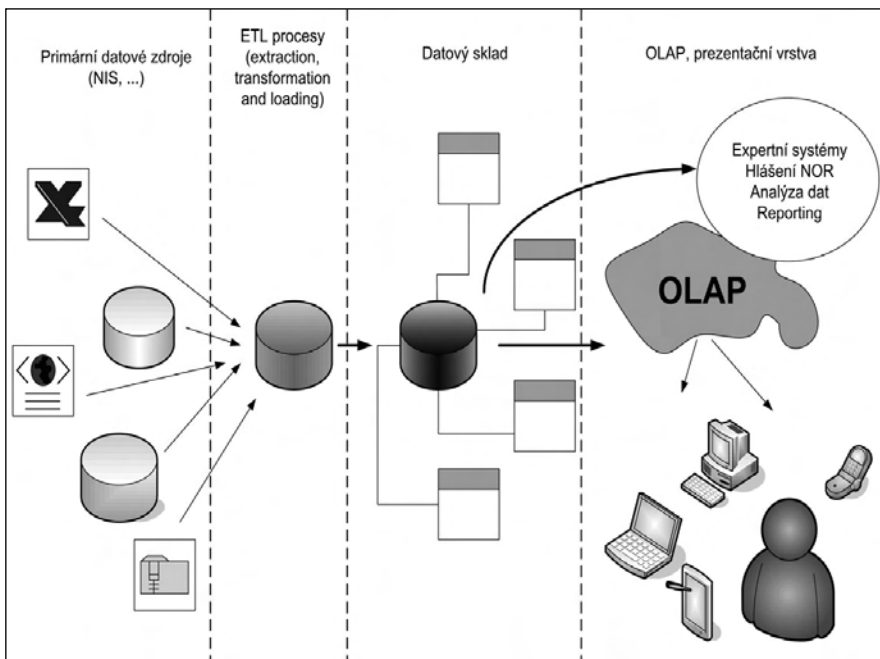
- ☑ nejednoznačná identifikace provázanosti parametrů mezi systémy,
- ☑ chybějící dokumentace,
- ☑ nedostatečná definice finálních výstupů,
- ☑ záznamy parametrů v textové podobě,
- ☑ roztržitost modulů,
- ☑ špatná implementace,
- ☑ systém nerespektuje procesní chod nemocnice,
- ☑ složitá integrace.

Možná řešení současného stavu

Z výše popsaných problémů by se mohlo zdát, že splnit předpoklad automatizované využitelnosti dat je prostě nemožné nebo příliš komplikované. Není tomu tak. Optimismus posílí již jen uvědomění, kolik informací v současné době již máme plně parametricky v různých systémech téměř v každé nemocnici (minimálně hlášení plátcům zdravotní péče a laboratorní výsledky). Bylo by ale chybou snažit se obsáhnout všechny informační systémy v jednotné ideální struktuře, která bude 100% flexibilní a přitom bude mít i akceptovatelné provozní náklady. Takový jednotný systém ani neexistuje. Dnešní nemocnice se totiž neobejdou bez IT či ICT oddělení, neustálého doplňování parametrů a rozšiřování datových struktur. Vznikají nové přístroje, diagnostické metody, mění se legislativa atd., a při všech těchto změnách je neustále nutné data integrovat a analyzovat.

Jedním z možných řešení výše popsané situace je datový sklad (Data Ware House DWH). Datový sklad si můžeme pro jednoduchost představit jako velkou databázi, která obsahuje kromě dat samotných i vazby a definice vazeb mezi jednotlivými systémy. Jakmile jsou data uložena, můžeme na ně nahlížet různými zobrazovacími nástroji a dále je zpracovávat do struktur, jako jsou například OLAP kostky. Dále je možné přímo vytvářet reporty, napojovat další aplikace apod. Příklad, jak vypadá proces získávání dat do DWH, ilustruje Obrázek 1.

Obrázek 1: Proces získávání dat do datového skladu (Data Ware House DWH)



Implementace datového skladu (DWH) na Masarykově onkologickém ústavu v Brně

Masarykův onkologický ústav se jako jedno z progresivních pracovišť rozhodl technologie DWH využít a v roce 2004 se rozběhla implementace zahrnující významné informační systémy nemocnice. Zvolené řešení datového skladu bylo vybráno s předpokladem, že bude dostatečně robustní nejen z hlediska spolehlivosti a bezpečnosti, ale zajistí také analytické funkce manažerského informačního systému s požadovanými výkonovými a ekonomickými parametry. Řešení je charakterizováno zejména následujícími vlastnostmi:

1. Použitá metodologie relačního on-line analytického zpracování umožňuje provádět **komplexní analýzu dat** (včetně ad-hoc analýzy, analýzy jednotlivých položek, implementace data mining metod, atd.) a **minimalizuje dodatečné investice** při dalším rozvoji systému (rozšiřování počtu uživatelů, objemu analyzovaných dat, jejich granularity, počtu dimenzí a aplikace komplexnějších analytických metod).
2. Použitá technologie tzv. ETL procesů (zajišťujících přenos dat z provozních systémů do datového skladu a kontrolu jejich kvality) využívá distribuovaného zpracování pro zajištění **vysoce spolehlivosti** procesu a nesymetrického šifrování a systému certifikátů pro zajištění **bezpečnosti a autenticity dat**.
3. Systém využívá „open source“ technologie, což kromě **nízkých celkových nákladů** znamená i vyšší bezpečnost, úroveň **ochrany investice** a transparentnost celého řešení (v důsledku přístupu ke zdrojovým kódům všech použitých komponent).
4. Technologie obsahuje prezentační vrstvu, která naplní předpoklad jednoduchosti a využitelnosti informací v reálném čase

Cílem prezentační vrstvy je zpřístupnit analytickou informaci uživateli tak, aby pro něj byla **snadno vyhodnotitelná, snadno dostupná** (tj. „ve správný čas na správném místě“) a **důvěryhodná**. Samozřejmým požadavkem je potřeba volit takové technologie, které jsou optimální z hlediska **celkových nákladů na vlastnictví** systému (tedy nízké náklady na údržbu a administraci, na provoz hardware a síťové infrastruktury, nízké náklady na rozšiřování počtu uživatelů).

Splnění výše uvedených požadavků je zajištěno vhodným výběrem použitých technologií a jejich promyšlenou implementací:

- **Snadné a jednoznačné vyhodnocení informace.** Kromě standardních výstupů ve formě vhodně formátovaných a parametrizovatelných tabulek je možné analytickou informaci prezentovat ve formě 2D grafů, map nebo přímo exportovat do souborů tabulkového procesoru. Je také možné vytvářet 3D prezentace dat s možností interaktivní manipulace a tak umožnit přehled o struktuře dat a trendech jejich vývoje.
- **Snadná dostupnost informace.** Analytická informace musí být snadno dostupná jak uvnitř sítě, tak i pro mobilní uživatele. Kromě interaktivního přístupu (např. přes web portál) je vhodné doručit potřebnou informaci uživateli i proaktivně pomocí nejrůznějších komunikačních kanálů (např. e-mail, SMS, fax).
- **Nízké celkové náklady na vlastnictví systému.** Kritické části systému jsou implementovány na platformě operačního systému GNU/Linux, díky jehož vysoké stabi-

lity a snadné údržbě je možné udržet velmi nízké náklady na provoz řešení.

Konkrétní požadavky a návrh datové struktury je součástí detailní projektové dokumentace, která jasně specifikuje strukturu, do kterých mají dodavatelé systémů exportovat nemocniční data, včetně popisu jednotlivých parametrů. Obecně lze říci, že řešené požadavky se týkají následujících oblastí:

- klinická data (pacient, hodnocení pacienta v čase, vývoj nemoci),
- ekonomická data (náklady, výnosy, hospodaření nemocnice...),
- multilaterální data (náklady na pacienta, obecně data agregovaná z více systémů).

Přístup řešení pro uživatele

Přístup prostřednictvím www portálu je vhodný pro široké spektrum uživatelů, kterým může poskytnout různou míru funkcionality a interaktivity, vždy optimalizovanou podle jejich konkrétních potřeb. Vývojáři (tzv. „power-user“) mohou prostřednictvím portálu vyvíjet nové reporty a analytické objekty (metriky, filtry, atd.), které pak mohou využít ostatní uživatelé na různých pozicích. Předem připravené reporty jsou snadno začlenitelné např. do portálu instituce, kde pak mohou poskytovat potřebné informace pro širokou komunitu uživatelů, a to i těch, kteří nejsou pro práci s OLAP technologií nijak speciálně školeni.

Uživatel může vytvářet reporty velmi jednoduchým způsobem přenášením (tzv. „**drag and drop**“) **atributů a metrik** do předem připravené mřížky modelu tabulky. Uživatel si tedy sám určí rozměr tabulky metriky, které chce sledovat v návaznosti na vybrané atributy. Součástí definice reportu je i definice filtrace reportu a jeho formátování. Formát zobrazení lze definovat přímo v prostředí tenkého klienta (web browser) interaktivním způsobem využívajícím obvyklých ovládacích prvků pro definici barvy výstupu, typu a velikosti písma, zarovnání atd.

Pro formátování reportu lze také využít tzv. podmíněného formátování. Je možné definovat podmínku (např. náklady větší než...) a na její splnění vázat určitý způsob formátování výstupní buňky reportu (barva, typ písma, buňku je možné doplnit o různé ikony, poznámky atd.).

V průběhu životního cyklu systému roste počet vytvořených objektů (atributů, faktů, metrik, filtrů, reportů, mřížek atd.) a je třeba přesně kontrolovat přístupová práva uživatelů k reportům a k jednotlivým analytickým objektům. Lze si např. představit situaci, kdy primář má být schopen vytvářet reporty zobrazující výkony jednotlivých lékařů jeho oddělení, ale neměl by mít možnost sledovat výkony ostatních oddělení. Pro řízení přístupu k datům je využít systém tzv. „**uživatelských rolí**“. Každý uživatel pak může v systému vystupovat v jedné nebo více rolích a každý objekt (např. report, metrika, atribut) má pak svůj vlastní tzv. „**Access Control List**“. Snadno lze pak definovat, které operace s objektem jsou povoleny a které zakázány pro konkrétní roli.

Pro práci ve vícejazyčném prostředí je výhodná důsledná internacionalizace použité technologie a lokalizace prostředí aplikace podle potřeby konkrétní organizace. Nastavením uživatelských preferencí se **lokalizuje** nejen prostředí www portálu, ale také **názvy** všech vytvořených **objektů** (metrik, filtrů, reportů) a **obsah reportů** (např. názvy diagnóz a formát data).

Alternativou zobrazení dat je grafická reprezentace pomocí **2D grafů** a **3D prezentací**, která umožní uživateli snadno a rychle se zorientovat v rozsáhlých reportech a nacházet různé spojitosti a trendy v datech, které nejsou z tabulkové prezentace na první pohled patrné. Grafická prezentace výsledků je přístupná prostřednictvím tenkého klienta (web browser), a to i včetně možnosti interaktivní manipulace.

Užitečnou pomocnou metodou pro analýzu vícerozměrných dat je **interaktivní zobrazení dat v třírozměrném prostoru** (tyto metody bývají zahrnovány do oblasti vizuálního dolování dat). Možnost interaktivní manipulace s 3D prezentací usnadňuje uživateli pochopení struktury dat, popř. trendů jejich vývoje.

Využití statistického software projektu „R“ umožňuje analytikům použít širokou škálu pokročilých statistických metod a metod pro dolování dat (**data mining**), vyvíjených na nejvýznamnějších světových akademických pracovištích. Grafické knihovny projektu „R“ lze také využít pro generování kvalitní tzv. „statistické“ a „obchodní“ grafiky. Konkrétní implementace na onkologickém pracovišti tak umožňuje analýzu přežití nebo léčebné odpovědi různých skupin pacientů a vícerozměrné hodnocení rizikových nebo potenciálně prognostických faktorů.

Využití kontingenčních tabulek tabulkového procesoru. Po přenosu dat z datového tržiště do tabulkového procesoru je možné velice efektivně provádět tzv. „řezy“ vícerozměrným datovým prostorem, včetně snadného drillování, pivotování a filtrování. Výsledky je možné snadno formátovat, zobrazovat v podobě 2D a 3D grafů a publikovat (ve formě PDF dokumentů nebo XLS souborů). Tento způsob práce s daty je vhodný pro menší objemy zpracovávané informace a menší počty sledovaných atributů a metrik. Data konkrétního reportu je také možno exportovat přímo z rozhraní tenkého klienta do tabulkového procesoru (MS Office, OpenOffice.org, StarOffice) ve formě tabulky včetně formátování. Jako vhodný tabulkový procesor je možné jmenovat produkty OpenOffice.org nebo StarOffice firmy Sun, zejména pro dobře propracovaný a pružný systém definování kontingenčních tabulek a kvalitní grafický výstup (i ve formě 3D grafů). Implementace těchto funkcí znamená, že uživatel získává postupně přístup do datových vrstev různé složitosti a od profilu spotřeby léku v celé nemocnici se tak dostane až k jednotlivým podáním, dávkám apod.

Použité technologie

Navrhované řešení manažerského informačního systému na Masarykově onkologickém ústavu je založeno na koncepci relační on-line analýzy dat, (tzv. **ROLAP** - Relational On-Line Analytical Processing), která poskytuje vynikající možnosti analýzy jak agregovaných ukazatelů, tak i atomických dat (např. dat jednotlivých obchodních transakcí). Využitím těchto prostředků je možné efektivně připravovat širokou škálu výstupů, od jednoduchých reportů až po komplexní analýzy, jinými technologiemi nerealizo-

vatelnými. Systém umožňuje zpřístupnit výstupy uživatelům jak prostřednictvím tenkých klientů (WEB technologie), tak i formou přímé analýzy prostřednictvím kontingenčních tabulek v tabulkovém procesoru nebo je automaticky vhodně zformátované distribuovat uživatelům prostřednictvím faxu, elektronické pošty nebo SMS zpráv. Podobně jako obecný systém pro podporu rozhodování, tak i řešení využívající technologie projektu MOÚ je rozděleno do tří základních vrstev: ETL procesy, datový sklad a prezentační vrstva. Tzv. ETL procesy zajišťují extrakci dat z primárních datových zdrojů, jejich transformaci a zavedení do datového skladu (z angl. „Extraction, Transformation and Loading“). Podle určené struktury tato vrstva může být reprezentována datovým tržištěm nebo datovým skladem.

Vlastní úložiště datového skladu (ev. datového tržiště) může být zabezpečeno obecným relačním systémem řízení báze dat. V současné době se jeví jako výhodné pro uvažované objemy využít technologie databáze **MySQL** (<http://www.mysql.com/>), která disponuje vynikajícími vlastnostmi pro využití v režimu tzv. RO (read-only) databáze při jedné z nejlepších propustností na trhu. Pro využití v oblasti datových skladů stojí za zmínku zejména následující vlastnosti:

- možnost uzamčení tabulek v paměti a využití hash indexů,
- možnost komprimace tabulek a indexů,
- možnost uložení a rychlého čtení dat přímo z indexových B-stromů,
- tzv. „Merge“ tabulky umožňující partikulární komprimaci virtuální datové tabulky,
- tzv. stripping tabulek umožňující jejich rozložení přes více diskových svazků a tím zvýšení I/O propustnosti,
- možnost replikace databáze na více fyzických systémech jak pro zvýšení spolehlivosti, tak pro rozložení zátěže při čtecích operacích.

Systém řízení báze dat MySQL je produktem švédské firmy MySQL AB. Jedná se o Open Source technologii publikovanou jak pod GPL, tak pod komerční licenci. Se svými více než 4 miliony aktivních instalací se řadí mezi celosvětově nejrozšířenější databázové platformy. Díky tak širokému nasazení a dostupnosti zdrojových kódů se vyznačuje výbornou stabilitou. V oblasti datových skladů firma MySQL AB disponuje referencí databáze o velikosti tabulek faktů 50 GB, celkové velikosti 200 GB (7 mil. záznamů/měsíc) udržující historii až 10 let.

Závěr

Řešení datového skladu bylo představeno jako efektivní cesta k získání informací z vnitřních databází nemocnice. Masarykův onkologický ústav se pro tuto cestu rozhodl již před 2 lety a v současné době disponuje robustním nástrojem, díky kterému není problém reportovat a analyzovat data z hlášení plátcům zdravotní péče nebo pracovat s daty hlášení NOR. Dostupná data lze rovněž automaticky exportovat do jiných systémů. V současnosti na MOÚ probíhá parametrizace klinického záznamu onkologického pacienta.