



Západočeská univerzita v Plzni
Katedra informatiky a výpočetní techniky
Univerzitní 8
30614 Plzeň
Česká republika

Rozsáhlé informační systémy – standardizace metadat

Odborná práce ke státní doktorské zkoušce

Petr Včelák

Technická zpráva č. DCSE/TR-2011-05
Září, 2011

Distribuce: veřejná

Rozsáhlé informační systémy – standardizace metadat

Petr Včelák

Abstrakt

Rozsáhlé informační systémy musí komunikovat a spolupracovat s okolím. Pro zajištění interoperability systémů je potřeba zajistit, aby spolupracující systémy používaly shodné slovníky i význam informací, které si mezi sebou mají vyměňovat. Práce se zabývá návrhem metod a principů pro zajištění interoperability při zpracování metadat s návazností na jejich další použití. Cíl výzkumného informačního systému pro podporu rozsáhlých medicínských a zdravotnických heterogenních dat reálných pacientů a především navrhovaného zpracování metadat je v zajištění jednotného prostředí a rozhraní pro podporu medicínského výzkumu primárně zaměřeného na problematiku cévní mozkové příhody.

Kopie zprávy jsou dostupné na
<http://www.kiv.zcu.cz/publications/>
nebo na žádost poslanou na adresu:

Západočeská univerzita v Plzni
Katedra informatiky a výpočetní techniky
Univerzitní 8
30614 Plzeň
Česká republika

Obsah

1	Úvod	1
2	Teoretické pozadí	2
2.1	Metadata	2
2.1.1	Typy metadat	2
2.1.2	Struktura a zápis	3
2.1.3	Standardy	4
2.1.4	Sémantický web	4
2.2	Resource Description Framework	4
2.3	Medicínská data	4
2.3.1	DASTA	5
2.3.2	HL7	6
2.3.3	DICOM	7
2.3.4	SITS	7
2.3.5	Ostatní	7
2.4	Interoperabilita	8
3	Aktuální stav a související práce	9
3.1	Stav DASTA vs. HL7	9
3.2	Formát SITS-XML	9
3.3	Vyšetření a léčba cévní mozkové příhody	11
3.4	Struktura dat pro výzkum CMP	11
3.5	Získání výzkumných dat	12
3.6	Ochrana osobních údajů	14
3.6.1	Osobní údaje pacienta	14
3.6.2	Proces anonymizace	15
3.6.3	Identifikace pacienta v obrazových datech	17
3.6.4	K-anonymita	17
3.7	Navrhovaný výzkumný informační systém	18
3.7.1	Motivace a princip	18
3.7.2	Stručný popis a vymezení hranice	19
3.7.3	Specifikace požadavků pro podporu výzkumu	19
4	Návrh zpracování metadat	22
4.1	Získání metadat	22
4.2	Uchování metadat	22
4.3	Interoperabilita a použití metadat	25
4.4	Schéma provedení experimentu	26

5	Ověření navrženého zpracování metadat	28
5.1	Anonymní medicínská data	28
5.2	Získání základních metadat při importu	28
5.3	Tvorba metadat experimenty	29
5.4	Interoperabilita metadat	29
5.5	Uchování metadat a rychlost odezvy	29
6	Závěr	30
7	Návrh cílů dizertační práce	30

1 Úvod

V rigorózní práci se primárně zabývám standardizací metadat a zajištěním interoperability v oblasti rozsáhlých výzkumných medicínských a zdravotnických informačních systémů. Výzkum je založen na velkém množství heterogenních medicínských dat reálných pacientů, která jsou v různých formátech a verzích. Cílem výzkumného informačního systému je zajistit jednotné prostředí a rozhraní pro podporu medicínského výzkumu obecně i přesto, že je nyní výzkum zaměřen na oblast mozku a problematiku cévní mozkové příhody. Cévní mozkové příhody tvoří jsou celosvětově nejčastější příčinou smrti. V České republice se jedná a druhou nejčastější příčinou smrti. [4, 16] Cévní mozkové příhody jsou také druhou nejčastější příčinou demence, nejčastější příčinou epilepsie v pokročilém věku a častou příčinou depresí. [1, 20, 22]

Navržený výzkumný informační systém podporuje všechny základní cíle informačního systému – získávání a přenos informací, zpracování informací (evidence, organizace, kategorizace, třídění, vyhledávání, agregaci), uložení informací, zpřístupnění informací (tisk, vizualizace) a v neposlední řadě odvozování nových informací. Pro poslední jmenované odvozování je navržena také metodika pro provádění tzv. experimentů, kterými lze z existujících informací i zdrojových dat vytvářet nové informace. Ve výzkumném informačním systému dochází k požadovanému zpracování metadat a uložené informace mohou sloužit pro navazující znalostní, expertní ale i výukové systémy.

Osobní údaje a znalost identity pacienta nejsou pro medicínský výzkum významné. Proto musí být veškerá data před přenosem anonymizována. Musí být odstraněny nebo nahrazeny identifikační a osobní údaje pacienta s důrazem na zachování vnitřních vazeb mezi daty. Proto klinické zprávy, laboratorní výsledky, obrazová dokumentace i další dokumenty zůstanou i pro výzkumné účely propojeny. Zachování vazeb a tedy zdravotnické historie pacienta ve zdrojových datech přináší další možnosti jejich sekundárního využití. Původní identifikátory (rodná čísla) jsou jednotně nahrazována novým anonymním identifikátorem.

Partnery výzkumu jsou Fakultní nemocnice v Plzni (FNPL), Lékařská fakulta v Plzni Karlovy univerzity v Praze (LFP) a Fakulta aplikovaných věd (FAV) na Západočeské univerzitě v Plzni (ZČU). Přímo spolupracují neurologická klinika, radio-diagnostická klinika, oddělení neurochirurgie, oddělení chirurgie, Katedra informatiky a výpočetní techniky a Katedra mechaniky. Medicínská data jsou získávána ve spolupráci s Fakultní nemocnicí v Plzni (FNPL).

2 Teoretické pozadí

V této kapitole nejprve definuji termín metadat a podávám s tím související informace. Po té čtenáře seznámím s problematikou dat v rozsáhlých výzkumných medicínských a zdravotnických informačních systémech. Především zmiňuji používané datové formáty pro výměnu informací. Jedná se o různé typy klinických zpráv, laboratorní vyšetření, ale i obrazovou dokumentaci.

2.1 Metadata

Slovo metadata pochází z řeckého *meta = mezi, za* a latinského *data = to, co je dáno* a představují strukturovaná data. Význam těchto strukturovaných dat je velmi často definován ve významu *data o datech*. Tedy strukturovaná data popisující data. Tato definice však není přesná. V případě strukturálních metadat, jako je návrh a specifikace datových struktur, se nemůžeme o datech bavit, protože v době návrhu žádná data aplikace neobsahuje. Správná definice pro tento případ by byla *data o kontejneru pro data*. Přitom popisná metadata jsou o individuálních instancích aplikačních dat, tedy obsahu dat (data content). Vhodná definice pro popisná metadata může být *data o obsahu dat* (data about data contents) nebo *obsah o obsahu* (content about content) a můžeme definovat pojem *metaobsah* (metacontent).

Metadata (metaobsah) zvyšují kvalitu dat/obsahu, protože umožňují popis obsahu a kontextu libovolného objektu a tím přidávání dalších, rozšiřujících informací k původnímu objektu. Jako popisovaný původní objekt může být cokoli. V našem případě rozsáhlých informačních systémů v oblasti výzkumu v medicíně to může být např. lékařská zpráva nebo radiologické obrazové vyšetření. Lékařská zpráva může obsahovat metadata o tom, jaké nástroje byly pro její tvorbu použity, kdy byla vytvořena, v jakém jazyce, zda obsahuje přílohy a jaké, ale také u obrazových dat i způsob vizualizace. Metadata přináší výhody při tvorbě katalogů, rejstříků, pro vyhledávání a třídění.

I metadata jsou data a hranice mezi nimi nemusí být vždy jasná. Stejně jako data, můžeme i metadata uložit např. do databáze nebo jiného úložiště. [8] Pouhým pohledem na data není možné rozlišit, zda se jedná o data nebo metadata. Uživatel nemůže vědět, kdy jsou data metadaty nebo pouze data. [2] Tvorba metadat může být provedena automatickým zpracováním informací nebo manuálně uživatelem. Některá základní metadata o času vzniku nebo poslední změny, tvůrci, vlastníkově a velikosti a typu souboru lze zjistit automaticky.

2.1.1 Typy metadat

Rozdělení podle typu metadat se u různých autorů liší. Bretheron & Singley (1994) uvádí dvě skupiny:

1. strukturální/řídící
2. a klíčová metadata.

Strukturální popisují strukturu počítačového systému, jako jsou tabulky, sloupce a indexy. Klíčová metadata pomáhají lidem najít konkrétní položky a jsou tvořena klíčovými slovy v přirozeném jazyce.

Ralph Kimball dělí metadata na tři typy:

1. technická (jinak též interní),
2. obchodní (business, jinak též externí)
3. a procesní metadata.

Na závěr National Information Standards Organization (NISO) dělí metadata na tři typy:

1. popisná,
2. strukturální,
3. administrativní – dále má podskupiny metadat:
 - (a) práva ke zdrojům – intelektuální vlastnictví, ochrana soukromí a důvěrnost zdrojů a informací
 - (b) ochrana zdrojů – metadata zahrnující archivaci a ochranu zdrojů

Popisná metadata slouží k vyhledání a lokalizaci objektu. Strukturální metadata popisují jak jsou komponenty objektu propojeny a organizovány. Administrativní metadata nesou technické informace jako je např. typ souboru. [18]

2.1.2 Struktura a zápis

Struktura metadat je dána použitým schématem. Schéma metadat v sobě zahrnuje standardy a modely metadat. Konkrétní podoba zápisu metadat vychází z pravidel vytvořených pro strukturování polí a elementů metadat (meta obsahu). [3] Schéma pro metadata může být vyjádřeno v množství značkovacích nebo programovacích jazyků, kdy každý vyžaduje odlišný způsob zápisu. Například schéma Dublin Core může být zapsáno v čistém textu, HTML, XML a RDF. [6]

Schéma metadat může být hierarchické, lineární (jednorozměrné) nebo rovinné (dvourozměrné). *Hierarchické* je schéma pokud existují vzájemné vztahy mezi elementy metadat a elementy jsou vnořeny tak, že mezi nimi existuje vazba rodič–potomek. Příkladem hierarchického schéma je IEEE LOM, kde elementy metadat mohou náležet do rodičovského elementu. *Lineární* schéma je takové, kde každý element je zcela nezávislý na ostatních elementech. Stejně tak je každý element klasifikován pouze v jednom rozměru, jako je např. Dublin Core schéma. *Rovinné* schéma má každý element zcela nezávislý od ostatních elementů, ale klasifikace je ve dvou vzájemně ortogonálních rozměrech. U rovinného schéma je nutný i typ mapování, který nám umožní zobrazení na metadata podle zvoleného aspektu. Často se jedná o vrstvení geografických a geologických informací.

Úroveň strukturování metadat se označuje jako **granularita**. Vyšší granularita umožňuje hlubší strukturování informací a současně s tím i lepší možnosti pro následující technickou manipulaci. Naopak, nižší granularita znamená méně detailní strukturování informací. Výhodou nižší granularity je značně levnější tvorba ale i následná údržba metadat. Granularita se totiž projeví i v případě, kdy metadata zastarají nebo je bude nutné transformovat do jiné podoby.

Pro další úroveň strukturování a sjednocení metadat mohou být použity nástroje jako jsou řízené slovníky, systematické slovníky, thesaury, datové slovníky a další.

2.1.3 Standardy

Standardizaci metadat se zabývaly především American National Standards Institute (ANSI) a International Organization for Standardization (ISO). Základním standardem je ISO/IEC 11179-1:2004 [11] a související standardy ISO/IEC 11179 [10]. Standard se zabývá definicí metadat a neposkytuje zmínku o konkrétním strukturování nebo manipulaci s metadaty. Definuje pojem metadata jako *data o kontejneru pro data* a jedná se o jednoduchou datovou položku, která neobsahuje žádné komplexní struktury.

2.1.4 Sémantický web

Dalším subjektem zabývajícím se metadaty a popisem zdrojů je World Wide Web Consortium (W3C), kde je zřejmá orientace na webové stránky už z názvu konzorcia. W3C Metadata Activity definuje metadata jako *strojově zpracovatelné/srozumitelné informace pro web* (machine-understandable information). Výsledkem W3C Metadata Activity byl Resource Description Framework (RDF) pro vyjádření a uplatnění metadat o informacích na webu – model a možnosti zápisu metadat. Snaha W3C Metadata Activity byla později nahrazena W3C Semantic Web Activity, kde vývoj RDF pokračoval.

2.2 Resource Description Framework

Resource Description Framework (RDF) lze volně přeložit jako systém pro popis zdrojů. Jedná se o model pro výměnu dat na webu. Výhodou je možnost slučování dat i v případě odlišných základních schémat a podporuje vývoje schémat v čase bez potřeby změn u cílového klienta. [25]

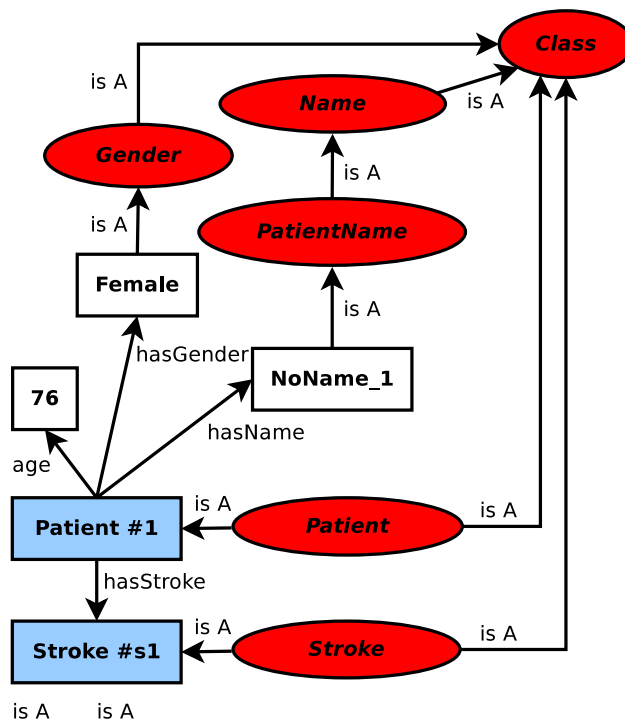
RDF rozšiřuje propojování dokumentů známé z webových stránek. Používá Uniform Resource Identifier (URI) pro pojmenování zdrojů – zdrojového i cílového. Pro pojmenování vazeb mezi zdrojovým a cílovým zdrojem používá také URI. Toto se označuje jako *trojice* (triple). Z ilustrace je zřejmé, že tento jednoduchý princip RDF umožňuje bez problému sdílet metadata s různou granularitou mezi různými aplikacemi.

Spojová struktura představuje pojmenovaný graf. Hrany grafu reprezentují pojmenovanou vazbu mezi dvěma zdroji – uzly. Vizualizace RDF jako graf je nejsnazší možností jak tento model pochopit.

Na obrázku 1 je reálný příklad ilustrující vizualizaci RDF dat grafem. Pro jednoduchost nejsou uvedeny úplné URI. Červeně jsou uvedeny třídy, světle modrá barva je pro instance třídy, bez barvy jsou uvedeny hodnoty atributu. Šipky ukazují pojmenované vztahy. Z obrázku např. vidíme, že existuje pacientka *Patient#1*, jmenuje se *NoName_1* ve věku 76 let a měla blíže nespécifikovanou cévní mozkovou příhodu *Stroke#s1*.

2.3 Medicínská data

Výzkum cévní mozkové příhody, kterým se zabýváme a pro nějž je primárně výzkumný informační systém navrhován, je založen na medicínských datech reálných pacientů a obsahuje klinická terapeutická a obrazová data. Uvedená data jsou heterogenní a neexistuje jeden datový formát, který by byl učen pro jejich



Obrázek 1: Příklad RDF dat zobrazených ve formě grafu

přenos a uchování. Pro klinická a terapeutická data jsou určeny formáty DASTA a HL7. Pro obrazová vyšetření je používán DICOM formát. Jedná se např. o vyšetření CT, MR, SONO, apod. Mimo tato data jsou informace o léčbě cévní mozkové příhody v registru Safe Implementation of Treatments in Stroke (SITS). Další, obecně, libovolné dokumenty mohou být přiloženy k lékařské zprávě. Existence a formát takového dokumentu nemusí být v době návrhu znám. Ve výzkumném informačním systému je však nutné mít možnost i s neznámým dokumentem manipulovat a popsat jej metadaty, např. vhodným rozšířením existujících metod pro zpracování dat.

2.3.1 DASTA

DASTA nebo také DS je zkratka pro Datový standard Ministerstva zdravotnictví České republiky a slouží k předávání dat mezi zdravotnickými informačními systémy. První verze (DS 1.1) byla do praxe zavedena v roce 1997 a jednalo se o čistě textový soubor. Následující verze DS 02.01 (rok 2002) a DS 03.01 (rok 2003) vycházejí z Extensible Markup Language (XML) a používají Document Type Definition (DTD) soubory. Od verze DS 03.01 jsou zavedeny číselníky Národního zdravotnického informačního systému (NZIS). Teprve verze DS 04.01 uvedená v roce 2007 využívá XML Schéma a jmenné prostory. [27]

DASTA nyní (DS verze 04.06.05) umožňuje výměnu dat o pacientech v rozsahu:

- identifikace pacienta,
- základní informace (příjmení, jméno, adresa, výška, hmotnost, atd.),
- urgentní informace (alergie, diagnóza, krevní skupina),

- anamnéza,
- aktuální i trvalé diagnózy,
- léky, očkování,
- klinické události:
 - laboratorní vyšetření,
 - radiologické vyšetření (RTG, CT, MR, PET,MR, PET, ...),
 - operační zpráva,
 - konzilium,
 - dekurz,
 - propouštěcí zpráva,
 - ambulantní zpráva,
 - a další,
- podklady pro vyúčtování a manažerský informační systém,
- platební vztahy, pojišťovny, pracovní neschopnost,
- speciální datové bloky,
- a další.

Klinické události byly zavedeny od DS 04.01 včetně jejich fáze. Fázemi jsou např. objednávka realizace, potvrzení objednávky, přijetí nebo nepřijetí objednávky, zpracovávání zakázky průběžné a finální, převzetí zakázky, předání dat třetí straně (např. IZIP) nebo opravy a storno.

DASTA v datových blocích využívá interní a externí číselníky, kterých je přes 300. Obsahují především číselníky pro NZIS a blok číselníků pro laboratorní komplement s Národním číselníkem laboratorních položek (NČLP). [27]

Zásadním nedostatkem je pouze národní podpora v České republice. V sousední Slovenské republice je DASTA využíván, protože jsou použity informační systémy z České republiky, které tento standard implementují. Oficiální podpora DASTA na Slovensku chybí. Ze strany DASTA však byly zařazeny datové bloky speciálně pro použití ve Slovenské republice. V ostatních zemích EU stejně jako mimo EU není DASTA využíván vůbec. [27]

2.3.2 HL7

Health Level Seven (HL7) je nezisková organizace zabývající se vývojem mezinárodních zdravotnických standardů. HL7 organizace vyvíjí stejnojmenný konceptuální standard, standard pro dokumenty, aplikační standardy a standardy pro výměnu zpráv. Uvedené standardy, návody a příručky mají za úkol zjednodušit a sjednotit komunikaci mezi rozličnými zdravotnickými systémy. [9]

Pro nás je významný Clinical Document Architecture (CDA), který je používán pro výměnu dokumentů jako jsou lékařské zprávy, poznámky lékaře a další materiály. CDA je standard využívající Extensible Markup Language (XML) určený ke specifikaci kódování, struktury a sémantiky klinických dokumentů. CDA dokument obsahuje povinné textové a nepovinné strukturované části. Povinné textové části dokument obsahuje kvůli čitelnosti člověkem. Strukturované části

závisí na kódových systémech jako jsou *Systematized Nomenclature of Medicine–Clinical Terms* (SNOMED CT) a *Logical Observation Identifiers Names and Codes* (LOINC). Strukturované části jsou určeny pro snadnou strojovou interpretaci dokumentu. [5]

2.3.3 DICOM

Digital Imaging and Communications in Medicine (DICOM) je standard pro distribuci, výměnu a zobrazování medicínských obrazových vyšetření nezávisle na jejich původu. DICOM umožňuje výměnu mezi počítači i technickým vybavením jako je počítačová tomografie, magnetická rezonance a další. DICOM může uchovávat množství různých obrazových i neobrazových (např. signály) dat. Jedná se velmi populární formát pro reprezentaci a přenos radiologických snímků a vyšetření. [17]

DICOM je binární soubor složený z několika atributů zahrnující položky jako je ID snímku, ID pacienta, jméno pacienta, modalita, datum a čas (studie, série i konkrétního snímku), protokol, popis série, pozici snímku, způsob zobrazení, tloušťku řezu, rozměry snímku a především vlastní snímek – obrazová data. Obvykle každý soubor představuje jeden prostý snímek. Ve skutečnosti formát DICOM podporuje více snímků, takže podporuje i filmové smyčky (cine-loop). Obrazová data mohou být ve třech nebo čtyřech rozměrech a mohou být komprimována řadou standardů – JPEG, bezztrátový JPEG, JPEG2000 a Run-Length Encoding (RLE). Dokonce může být použita LZW komprese na všechna obsažená data.

2.3.4 SITS

Safe Implementation of Treatments in Stroke (SITS) je akademická, nezisková a mezinárodní iniciativa pro spolupráci zdravotnických odborníků pro zrychlení klinických pokusů v akutní léčbě a prevenci mrtvice. Registr SITS pochází z Karolinska Institutu ve Švédsku, ale do sítě registru SITS přispívá široká skupina nemocnic z celého světa včetně nemocnic z České republiky. Registr SITS inicioval interaktivní internetový registr trombózy sloužící klinickým centřům ke sledování jejich vlastní léčebných výsledků a možnost porovnání s ostatními centry v zemi i mimo ní. [23]

Pro naše potřeby považujeme registr SITS za důležitý, protože obsahuje klinické a léčebné záznamy ve strukturované podobě. Obsahují terapeutické záznamy, reakci léčiv, skórování stavu pacienta podle National Institutes of Health (NIH) stupnice, popis snímků počítačové tomografie (CT) a magnetické rezonance (MR), popis příčin úmrtí a mnoho dalších.

2.3.5 Ostatní

Pro rozšíření výzkumného zaměření navrhovaného informačního systému v budoucnu je nutné zahrnout i podporu předem neznámých datových formátů. Je zřejmé, že nebude možné neznámé soubory zpracovávat a analyzovat bez doplnění potřebných nástrojů do informačního systému. Mělo by však být reálné i takovéto soubory uložit a evidovat jejich existenci v základní možné podobě – datum vzniku, autor a zdroj, související dokumenty, atd.

Jedním z formátů, o který by měla být v budoucnu podpora rozšířena, je European Data Format (EDF). Především jeho novější podobu EDF+ z roku 2003.

Oproti EDF umožňuje EDF+ textové anotace s časovou značkou, podněty (stimuly), průměrné hodnoty signálů, parametry elektrokardiogramu, dočasná zástava dýchání (apnoea) a další. Dále EDF+ podporuje uchování elektromyografie, evokovaných potenciálů, elektroneurografie, elektrokardiografie a řadu dalších vyšetření. [14]

2.4 Interoperabilita

Interoperabilita je schopnost systémů vzájemně si poskytovat služby a efektivně spolupracovat. *Syntax* se zabývá výrazy a jejich skladbou. *Sémantika* studuje významy výrazů. Úrovně interoperability lze odstupňovat a v roce 2009 přinesla iniciativa Dublin Core Metadata Initiative [19] v následující podobě:

1. Sdílené termíny – jedná se o sdílené slovníky definované v přirozeném jazyce.
2. Formální sémantika – sdílené slovníky založené na formální sémantice. Jsou definovány vztahy mezi termíny a pravidla pro použití výrazu. Umožňuje automatické odvozování (logical inferences). Bez ohledu na nativní kódování, specifikace může být označena jako sémanticky interoperabilní, když je zajištěno úplné mapování na RDF trojice (např. z XML souboru použitím Gleaning Resource Descriptions from Dialects of Languages transformací).
3. Sdílené formální slovníky, které si mohou strany vyměňovat a poskytují základní validaci a výměnu záznamů s metadaty.
4. Sdílené formální slovníky včetně definovaných strukturálních omezení nad slovníky.

3 Aktuální stav a související práce

3.1 Stav DASTA vs. HL7

DASTA je využíván ve zdravotnických informačních systémech v České republice řadu let a všechny tyto informační systémy jej podporují, ale o jednotnosti implementace nemůže být řeč. Podkladem pro toto tvrzení je průzkum z února 2011 publikovaný na Českém národním fóru pro eHealth. [26]

- osloveno 50 firem
- odpovědělo 11 tvůrců NIS (Nemocniční IS), 8 tvůrců LIS (Lékařský IS) a 3 tvůrci AIS (Ambulantní IS)

Ze zúčastněných 8 tvůrců Lékařských informačních systémů jsou detaily následující.

- Implementovány verze DS 3.10 až 3.15.
- Zásadní úpravy provedené v DS 3.14 provedly 2 firmy.
- Webové služby pro aktualizaci slovníků používají pouze 2 firmy.
- Verzi DS 4 implementovala pouze 1 firma a ostatní čekají, až bude DS 4 požadováno uživateli.

Z uvedeného výčtu a verzí DS 3 je zřejmé, že v LIS jsou používány historické datové standardy z roku 2006 až po aktuální z roku 2011. Problémem je chybějící legislativní podpora DASTA a číselníků, kdy není kladen žádný tlak na tvůrce ani uživatele IS. [26]

Dalším negativem DASTA je neexistující mezinárodní podpora. Standard je popsán pouze v českém jazyce. Dokonce datové elementy a atributy jsou v češtině (bez diakritiky). Řešením by bylo provést revizi DASTA standardu a vše pojmenovat v anglickém jazyce. Vznikla by zcela nová verze standardu. S jejím nasazením by ke zlepšení bez legislativní opory a vynucení si reflektování aktuální verze standardu také nedošlo. Další nevýhodou by byla nutnost konverze stávajících dokumentů na novou verzi standardu.

Pro HL7 existuje v České republice občanské sdružení jehož cílem je prosazování standardů HL7 v České republice prostřednictvím diskusního prostoru, osvěty a výuky a vytvářením národního přízpůsobení a překladu. HL7 Česká republika je mezinárodní přidruženou organizací Health Level Seven (HL7 International Affiliate). Podpora HL7 je v jediném nemocničním informačním systému v České republice. [24]

3.2 Formát SITS-XML

Komplikací při práci lékaře se SITS registrem je v nemožnosti jednoduchým způsobem přebírat data zpět, data vhodně vizualizovat nebo nad nimi provádět např. testování hypotéz.

Veškerá do registru vyplněná data jsou přístupná pouze ve formě webové stránky. To je důvodem, proč data transformujeme z webové stránky do vlastního formátu SITS-XML s XSD schématem. Soubor ve formátu SITS-XSD používáme jako vstupní vedle formátů DASTA, HL7 a DICOM.

Pro uvedenou transformaci jsem využil základních nástrojů dostupných na operačních systémech GNU/Linux. Kombinace nástrojů *wget*, *links*, *sed*, *csplit*, *cut* a *paste* posloužila pro tento převod. Názvy elementů jsou transformovány z webové stránky. Realizované řešení je přes rok v ostrém provozu a osvědčilo se jako bezproblémové i v případě, kdy SITS registr změní strukturu dat.

Listing 1: Ukázka SITS-XML formátu souboru

```
<?xml version="1.0" encoding="UTF-8"?>
<sits>
  <Baseline__TreatmentFile>CZUHP2006053001</Baseline__TreatmentFile>
  <Baseline__Aspirin75MgOverDay>No</Baseline__Aspirin75MgOverDay>
  <Baseline__DipyridamoleSlowRelease>No</Baseline__DipyridamoleSlowRelease>
  <Baseline__Clopidogrel>No</Baseline__Clopidogrel>
  <Baseline__OtherAntiplatelet>No</Baseline__OtherAntiplatelet>
  <Baseline__SystolicBloodPressure>125</Baseline__SystolicBloodPressure>
  <Baseline__DiastolicBloodPressure>73</Baseline__DiastolicBloodPressure>
  <Baseline__Glucose>5.4</Baseline__Glucose>
  <Baseline__TotalCholesterol>5.4</Baseline__TotalCholesterol>
  <Baseline__WeightMeasured>58</Baseline__WeightMeasured>
  <Baseline__CTDone>Yes</Baseline__CTDone>
  <Baseline__CTDateandtime>2006-01-14 T10:20:00:00</Baseline__CTDateandtime>
  <Baseline__CTCurrentInfarct>No</Baseline__CTCurrentInfarct>
  <Baseline__CTDenseArterySign>Yes</Baseline__CTDenseArterySign>
  <Baseline__CTPerfusionDeficit>Yes</Baseline__CTPerfusionDeficit>
  <Treatment__DoseActilyse>50.0</Treatment__DoseActilyse>
  <After2H__SystolicBloodPressure>137</After2H__SystolicBloodPressure>
  <After2H__DiastolicBloodPressure>71</After2H__DiastolicBloodPressure>
  <After24H__GlobalOutcome>Better</After24H__GlobalOutcome>
  <After24H__SystolicBloodPressure>126</After24H__SystolicBloodPressure>
  <After24H__DiastolicBloodPressure>71</After24H__DiastolicBloodPressure>
  <After24H__CTDone>Yes</After24H__CTDone>
  <After24H__CTDateandtime>2006-01-15 T10:20:00:00</After24H__CTDateandtime>
  <After24H__CTCurrentInfarct>Yes</After24H__CTCurrentInfarct>
  <After24H__CTDenseArterySign>No</After24H__CTDenseArterySign>
  <After24H__CTLocalHaemorrhage>No</After24H__CTLocalHaemorrhage>
  <After24H__CTRemoteHaemorrhage>No</After24H__CTRemoteHaemorrhage>
  <After7D__Aspirin75MgOverDay>No</After7D__Aspirin75MgOverDay>
  <After7D__DipyridamoleSlowRelease>No</After7D__DipyridamoleSlowRelease>
  <After7D__Clopidogrel>No</After7D__Clopidogrel>
  <After7D__OtherAntiplatelet>No</After7D__OtherAntiplatelet>
  <After7D__TypeOfStroke>Ischaemic stroke</After7D__TypeOfStroke>
  <After7D__ICDCode>I63.3</After7D__ICDCode>
  <After7D__GlobalOutcome>Much better</After7D__GlobalOutcome>
  <After3M__RankinScore>2 Slight disability</After3M__RankinScore>
  <Free__CTDone>Yes</Free__CTDone>
  <Free__CTDateandtime>2006-01-16 T10:20:00:00</Free__CTDateandtime>
  <Free__CTCurrentInfarct>Yes</Free__CTCurrentInfarct>
  <Free__CTDenseArterySign>No</Free__CTDenseArterySign>
  <Free__CTAOcclusion>No</Free__CTAOcclusion>
  <Free__CTPerfusionDeficit>No</Free__CTPerfusionDeficit>
  <Free__CTLocalHaemorrhage>No</Free__CTLocalHaemorrhage>
  <Free__CTRemoteHaemorrhage>No</Free__CTRemoteHaemorrhage>
</sits>
```

V posledním roce vznikl v rámci diplomové práce [7] nový formát navržený nad HL7 CDA a plně respektující údaje v SITS. Navržený formát má být diskutován s autory SITS registru. V případě jeho nasazení by zjednodušil výměnu dat a umožnil vyřešit duplicitní činnost spojenou se zakládáním záznamů v registru SITS. Lékař totiž veškeré údaje zapisuje do lékařské zprávy, kde jsou údaje ve formě prostého textu. Při vyplňování záznamu v SITS registru lékař tyto informace v lékařské zprávě dohledává a pracně přepisuje.

3.3 Vyšetření a léčba cévní mozkové příhody

Průběh léčby pacienta s cévní mozkovou příhodou (CMP) je popsán v této sekci. Při léčbě CMP jsou klíčové body v okamžiku příjmu (baseline), po 2 hodinách, 24 hodinách, 7 dnech a po 3 měsících je výstupní kontrola.

Bezprostředně po příjmu pacienta s podezřením na CMP je provedeno CT vyšetření a sestra provede hodnocení stavu pacienta dle metodiky National Institutes of Health Stroke Scale (NIHSS). Jde o testování, zda pacient může mluvit, porozumí otázkám, orientuje v okolí, má motorické schopnosti apod. Výsledkem je číselná hodnota udávající úroveň postižení. Čím více se hodnota blíží k 0, tím menší postižení u pacienta existuje. Vyhodnocení NIHSS je prováděno ve všech klíčových bodech léčby. Stejně tak se eviduje krevní tlak, teplota a další údaje. O pacientovi se zaznamenávají i rizikové faktory a provedená léčba.

Na základě CT vyšetření je rozhodnuto o typu CMP a provedena odpovídající léčba. Kontrolní CT vyšetření se provádí přibližně po 24 hodinách a ze snímků se vyhodnocuje další postup. Také se opakuje vyhodnocení podle NIHSS a další údaje. Po výstupním vyšetření, které může obsahovat také i CT vyšetření se léčba pacienta ukončuje.

3.4 Struktura dat pro výzkum CMP

Úplný popis CMP pacienta je popsán množinou souborů obsahující heterogenní medicínská data, jejichž formáty DASTA a DICOM již byly popsány v sekci 2.3 a SITS-XML viz 3.2.

Každé z CT vyšetření představuje samostatnou klinickou událost a popisuje nález na CT snímcích. Údaje o průběhu CMP lékař eviduje v lékařské zprávě. Tyto údaje také duplicitně vyplňuje do registru SITS. Na rozdíl od lékařské zprávy, kde je zápis v podobě čistě textové, je SITS ve strukturované podobě. Soubory CT vyšetření jsou součástí lékařské zprávy v podobě příloh. Jedná se o tisíce souborů a jsou přiloženy uvedením názvu souboru, nikoliv přímo v XML dokumentu DASTA.

Každý soubor DASTA představuje jednu klinickou událost, např. vstupní CT vyšetření. V případě kontrolního CT vyšetření je v nemocničním systému vytvořena další klinická událost a samostatný soubor DASTA.

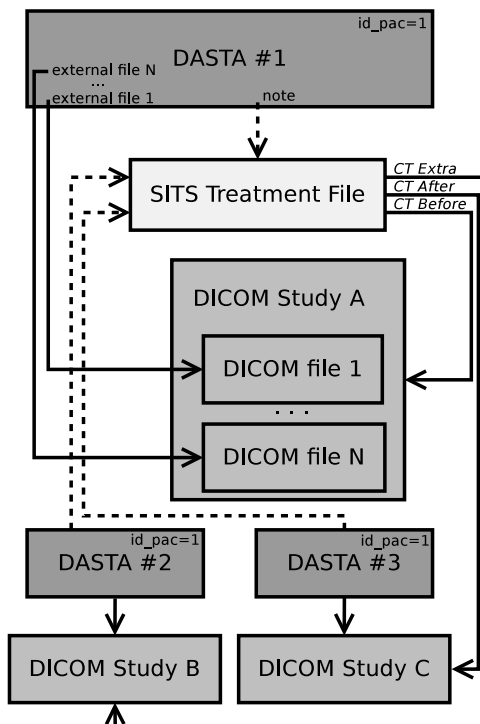
Každý soubor DASTA obsahuje číslo klinické události, které je i součástí všech DICOM souborů se snímky CT vyšetření. Vazba mezi těmito soubory je proto bez komplikací. Klinické události jsou však samostatné a na další, především přímo související, události přímou vazbu nemají. Na první pohled je proto společná pouze identifikace pacienta.

Do registru SITS lékař zapisuje postupně celou historii léčby pacienta s CMP. Pro jeden případ CMP proto existuje jeden záznam v tomto registru. Mezi velkým množstvím údajů v SITS obsahuje i datum a čas provedeného CT vyšetření pro

konkrétní časový úsek. Použitím data a času jsme schopni zpětně dohledat a propojit související klinické události – CT vyšetření vztahující se k jednomu případu. Řešení komplikuje existence nezávislých identifikátorů v SITS, kdy SITS vytváří svůj vlastní identifikátor. SITS umožňuje lékařům i evidenci MR vyšetření, které však není prováděno.

Pro mapování mezi identifikací pacienta v nemocnici a v SITS používáme převodní mapovací tabulku. Mapování vzniká dynamicky z označení na kartě pacienta nebo manuálně ručním doplněním této tabulky.

Ilustrační schéma vnitřního propojení výzkumných dat mezi sebou je na obrázku 2. Pro 3 možná CT vyšetření – klinické události – existují tři samostatné DASTA soubory, kde je pacient identifikován atributem *id_pac*. Každý DASTA soubor obsahuje výčet příloh, tedy všech názvů souborů se snímky v DICOM formátu. Mimo výčet příloh obsahuje také číslo klinické události (současně DICOM studie), které je ve všech DICOM souborech obvykle zapsáno. Posledním prvkem struktury je tzv. záznam Treatment File získaný ze SITS registru. Pomocí něho lze rozlišit které DICOM studie spolu souvisí, tj. tvoří vstupní (CT Before – před léčbou), kontrolní (CT After – po léčbě) a volitelně výstupní (CT Extra) vyšetření.



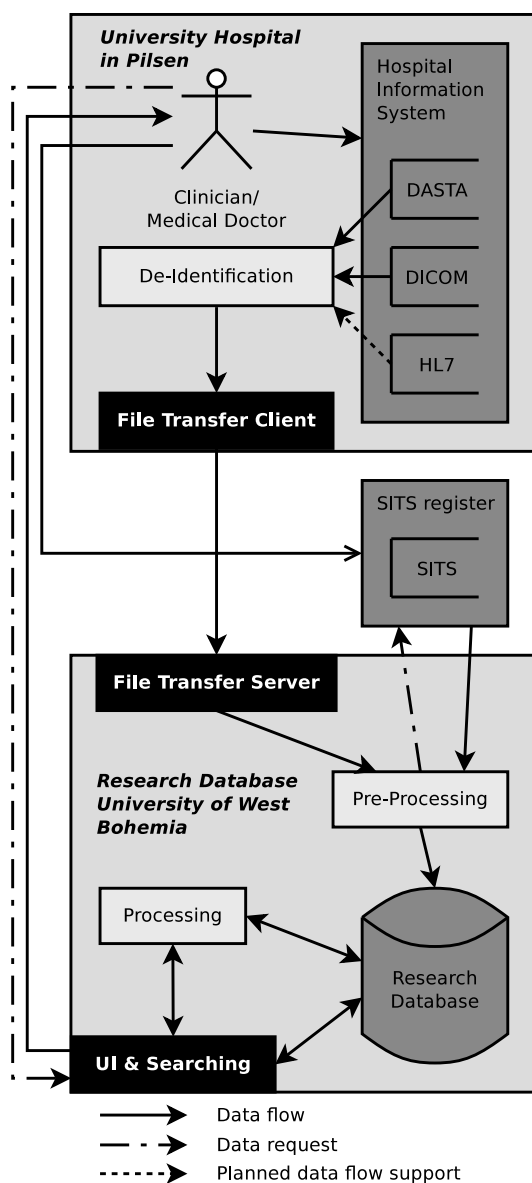
Obrázek 2: Struktura a vnitřní vazby ve výzkumných datech pro jeden případ cévní mozkové příhody

3.5 Získání výzkumných dat

Schéma získávání medicínských dat, tedy přenosu z Fakultní nemocnice v Plzni a Západočeskou univerzitou, je na obrázku 3. Lékař zapisuje informace o lékařské péči do nemocničního informačního systému (NIS, anglicky Hospital Information System) a SITS registru.

Cílová strana v podobě Západočeské univerzity je v NIS dostupná jako samostatná instituce na úrovni spolupracujícího zařízení, kam lze odesílat data. Odeslání dat (klinické události) do výzkumného centra je proto z pohledu lékaře stejné, jako by data odesílal např. do jiné nemocnice. Je možná i varianta, kdy se v NIS nastaví trigger (spouštěč), který vygeneruje odeslání dat automaticky po splnění definovaných podmínek triggeru.

Ve skutečnosti je cílová stanice, která přijímá data, umístěna přímo ve Fakultní nemocnici v Plzni. Je to z důvodu zajištění ochrany osobních údajů pacientů a detaily popisuje sekce 3.6. Jedná se o anonymizační stroj. Po provedení úkonů zmíněných v uvedené části, jsou data připravena k přenosu. Přenos je prováděn dávkově v pravidelných intervalech. Každý přenos souboru je zaznamenán včetně



Obrázek 3: Schéma předávání dat mezi Fakultní nemocnicí v Plzni a Západočeskou univerzitou při poskytování výzkumných medicínských dat.

informace o úspěšnosti. Soubor na anonymizačním stroji je přesunut do archívu v případě, že se přenos podaří. V opačném případě zůstává soubor v adresáři určeném pro přenos.

Pro přenos dat mezi nemocnicí a univerzitou bylo testováno několik nástrojů. Z počátku byla vyvinuta i vlastní klient-server aplikace pro případ použití operačního systému Microsoft Windows. Ve finální podobě bylo nasazeno řešení na operačním systému GNU/Linux a byly využity nástroje *rsync* a *scp* z důvodu dosahované přenosové rychlosti.

Využití registru SITS, tedy stažení a zpracování dat podle sekce 3.2 je provedeno automaticky v pravidelných časových intervalech. Pro rychlejší získání těchto dat lze aktualizaci údajů ze SITS přímo navázat na skripty zajišťující přenos dat mimo nemocnici.

3.6 Ochrana osobních údajů

Ze strany nemocnice je nutné zajistit dostatečnou ochranu osobních údajů – ze všech dat odstranit identitu pacientů – **anonymizovat** data. Při opětovném použití medicínských dat reálných pacientů ve výzkumu není potřeba, aby taková data obsahovala identitu pacientů. K výzkumným datům se mohou dostat osoby bez lékařského vzdělání jako jsou výzkumní pracovníci a studenti, mající za úkol implementovat testování hypotéz nebo provést jiné požadované zpracování takových dat.

Osobní údaje definuje zákon [12] jako *jakékoli informace týkající se určeného nebo určitelného subjektu údajů* (§4 písm. a) Zákona o ochraně osobních údajů). Osobním údajem proto jsou jakékoli údaje, které se ke konkrétní fyzické osobě vážou.

Každý zdravotnický pracovník má časově neomezenou povinnost zachovávat mlčenlivost o všech skutečnostech, o nichž se dozvěděl v přímé souvislosti s výkonem povolání. Mlčenlivost se vztahuje na skutečnosti medicínského i jiného charakteru. Zdravotnický personál nesmí sdělovat informace o diagnóze, léčebných postupech či rodině třetí straně.

Výzkumná data proto získáme odstraněním identity pacienta a to tak, aby nebylo možné pacienta zpětně identifikovat. Identifikátory budou nahrazeny novými bez vztahu k původní identifikaci a dalšími informacemi se zabývá následující sekce 3.6.1. Tuto problematiku řešila organizace *The Health Insurance Portability and Accountability Act* (HIPAA) v roce 1996 a vytvořila *Privacy and Security Rules*. Informace o pravidlech jsou také ve zdroji [13].

3.6.1 Osobní údaje pacienta

V DASTA základní XML element *dsip : ip* nese data vztažená k jedinému pacientovi. Ve verzi DS 04.06.05 jsou to následující data rozdělená na dvě části. První obsahuje osobní data zahrnující především identifikaci pacienta a další jeho vztahy k okolí. Druhá obsahuje osobní data, která jsou nutná pro výzkum a je nutno je zachovat pro další zpracování (příp. upravit). [15]

- Identifikace pacienta a osobní údaje – všechna tato data je nutné odstranit:
 - identifikace pacienta v IS odesilatele,
 - rodné číslo,

- jméno,
 - příjmení,
 - rodné příjmení,
 - titul před jménem,
 - titul za jménem,
 - jiné identifikační údaje,
 - adresy vázané k pacientovi,
 - platební vztah aktuální,
 - údaje o zdravotní pojišťovně (zdravotních pojišťovnách),
 - údaje pro NZIS,
 - léky vydané lékárnou,
 - a pracovní neschopnosti.
- Data potřebná pro výzkum:
 - urgentní informace o pacientovi neformalizované,
 - datum a čas narození,
 - datum a čas úmrtí,
 - pohlaví,
 - výška a hmotnost standardní,
 - anamnéza souhrnná neformalizovaná,
 - očkování,
 - diagnózy trvalé a přechodné,
 - podávané léky,
 - a klinické události.

I data určená pro výzkum lze dále upravit. Příkladem je datum a čas narození – je-li vyplněn i čas narození, můžeme jej zanedbávat stejně jako den (nebo i měsíc) narození. Preferovanou možností je datum narození zcela odstranit a pro klinickou událost jen vypočítat věk pacienta. Záleží, jak vysokou míru přesnosti pro výzkum potřebujeme.

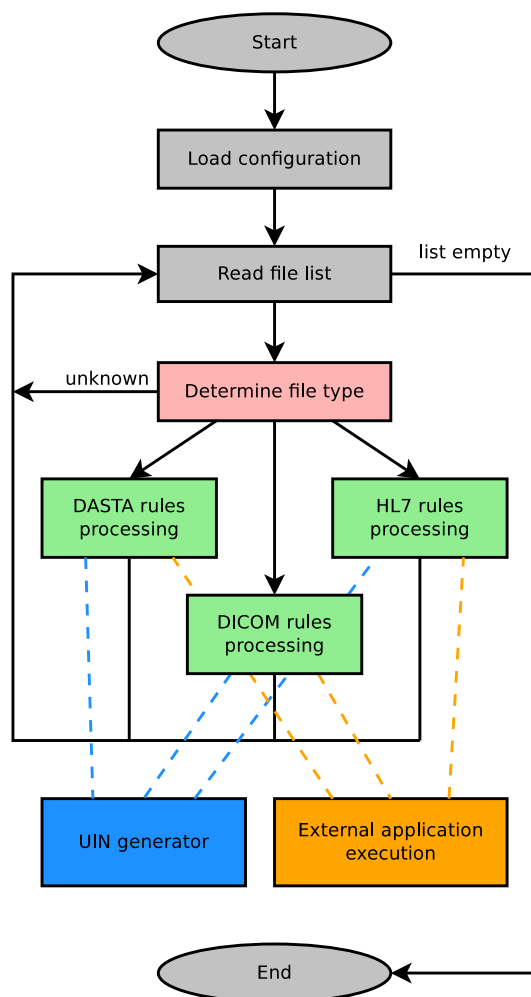
Adresa pacienta je významným identifikačním osobním údajem, ale část adresy může být vhodná i pro výzkum. Např. poštovní směrovací číslo (PSČ) může být účelem výzkumu při zkoumání v jaké oblasti je nejvyšší počet zkoumaných onemocnění u pacientů. Přitom PSČ je informace, která může poskytnout zásadní informaci pro identifikaci pacienta především v případě obcí s menším počtem obyvatel. Pro rozhodování, zda je množina dat anonymní zavádím v sekci 3.6.4 definici k-anonymity.

3.6.2 Proces anonymizace

Anonymizace počítá se strukturovanými, nestrukturovanými i obrazovými daty. Strukturovaná data jsou v XML souborech (DASTA a HL7) v definovaných elementech a attributech. Nestrukturovaná data spočívají v lékařské zprávě ve formě čistého textu, případně stejného textu v podobě formátovaného textu v RTF dokumentu vloženém přímo v DASTA souboru v Base64 kódování. Posledním případem

obrazových dat je uložena identifikace pacienta přímo v DICOM snímcích (viz 3.6.3) a v budoucnu plánované doplnění anonymizace o možnost deformace obličejových rysů v DICOM sériích.

Umístění anonymizace při zpracovávání dat je uvedeno v sekci 3.5 a situaci ilustruje schéma na obrázku 3. Ze schéma je zřejmé, že anonymizace je prováděna ještě v nemocnici. Proto data s identitou pacientů nebo jejich osobními údaji nemocnici v žádném případě neopustí. Teprve po úspěšném provedení anonymizace mohou být data přenesena do výzkumného centra pro další zpracování. Metody anonymizace, jako je např. odstranění identifikace z obrazové informace (viz obrázek 5), které nemusí mít vždy 100% spolehlivost, umístí anonymizovaný soubor do vyhrazeného adresáře. Soubory z vyhrazeného adresáře nejsou automaticky přenášeny mimo anonymizační stroj – je nutný zásah osoby, která úroveň anonymizace ověří a schválí. V případě nedosažení kvality anonymizace, má osoba možnost provést anonymizaci opakovaně nebo manuálně.



Obrázek 4: Schéma činnosti aplikace AnonMed pro anonymizaci dat a metadat

Proces anonymizace je realizován vlastní aplikací AnonMed. Schéma funkce aplikace je ilustrováno na obrázku 4. Po spuštění je načtena konfigurace a pravidla pro anonymizaci pro jednotlivé datové formáty a jejich verze. Vstupní soubory jsou zpracovávány postupně, ale v případě více procesorů/jader může být nastaveno zpra-

cování paralelní. Provede se rozpoznání typu souboru podle jeho obsahu a nikoliv pouze jeho koncovky názvu souboru. Podle zjištěného typu se provede příslušná série operací anonymizace souboru. Původní soubor je uložen do výstupního adresáře. Pro zajištění vyšší spolehlivosti anonymizace jsou neznámé typy souborů přeskočeny a zůstávají pouze ve vstupním adresáři. Pro anonymizace je možnost využít i libovolné externí aplikace. To umožňuje zavést i dodatečně podporu neznámých formátů nebo nových metod.

Aplikace AnonMed modifikuje nebo odstraňuje data a metadata ve zdrojových souborech. Umožňuje-li to datový formát souboru, lze i nová metadata přidat. Použitá aplikace AnonMed musí podporovat všechny datové formáty a jejich verze, které jsou v nemocnicích používány. Na základě informací uvedených v sekci 3.1 se může zdát taková podpora verzí DASTA značně komplikovanou.

Při reálném nasazení vycházíme z použitého nemocničního informačního systému pro dané zdravotnické zařízení a použijeme přímo anonymizační pravidla určená pro verze podporované v NIS. Odlišnosti jsou především mezi verzí DS 03 a DS 04. Struktura elementů a atributů s identifikací a osobními údaji pacienta se v rámci skupiny podverzí DS 03 nebo DS 04 neliší.

3.6.3 Identifikace pacienta v obrazových datech

Některá zařízení umožňují zápis identifikace pacienta přímo do obrazových dat, která jsou uložena v DICOM souborech. V metadatach DICOM souboru je v tagu *Burned-in Annotation* (0028, 0301) zápis do obrazové části indikován hodnotou *Yes*. Objevily se však DICOM soubory z několika zařízení, která i přes vypálenou identifikaci pacienta tento tag vyplněný neměly.

Odstranění identifikace pacienta z obrazové části DICOM snímku je možné podle předdefinovaných pravidel pro typ a verzi zařízení nebo použít optické rozpoznávání textu. V našem případě je jméno pacienta, rodné číslo, datum vyšetření i další informace před anonymizací v DICOM souboru v podobě metadat dostupné a metoda rozpoznávání textu může hledat předem známé textové řetězce. Nalezený textový řetězec je v DICOM snímku překryt černou výplní. Obrázek 5 ukazuje DICOM soubor po odstranění identifikace a osobních údajů z obrazových dat.

CT vyšetření hlavy v DICOM souborech informace v obrazové části neobsahují. Jedná se o zkušební nasazení tohoto mechanismu pro ověření funkce pro budoucí použití. Otestováno bylo několik DICOM souborů z dalších zdrojových zařízení, např. vyšetření ultrazvukem.

3.6.4 K-anonymita

Snažíme se zajistit co nejvyšší anonymitu informačního systému a identifikovat parametry a atributy, které jednoznačně ukazují na konkrétní osobu – pacienta. Jedná se o způsob, jak zjistit, zda je soubor dat dostatečně anonymní.

Definice 1 (K-anonymita). *Mějme K celé číslo větší než 1. Podmínka K-anonymity je splněna právě tehdy, když vyhledávání v souboru dat vrátí výsledky, jejichž počet je větší nebo roven číslu K .*

Číslo K musí být větší než 1, protože pokud by existoval jediný výsledek vyhledávání, již by se jednalo o konkrétního pacienta.

Důsledek definice lze ilustrovat na následujícím příkladu. Hledáme v souboru 4 žen a 6 mužů podle pohlaví. Pro případ vyhledávací podmínky na osoby ženského pohlaví je vrácen výsledek o velikosti 4 záznamů. Soubor dat poskytuje 4-anonymitu, protože je $K = 4$. Zpřesníme-li vyhledávání doplněním podmínky na věk pacienta a obdržíme ve výsledcích jediný záznam, pak jsme v souboru pacientů dokázali identifikovat konkrétní osobu.

Z uvedeného textu je zřejmé, že musí být počítáno při návrhu výzkumného informačního systému s možností některé informace považovat za citlivé i v jinak anonymních datech. Příkladem takové informace je v sekci 3.6.1 uvedené PSČ, které může velmi snadno identifikovat fyzickou osobu.

3.7 Navrhovaný výzkumný informační systém

Hlavním cílem je vytvořit informační systém pro univerzální podporu výzkumu v oblasti medicíny (VIS). Univerzalita je míněna ve vztahu k podpoře výzkumu ve více oblastech medicíny. Současná řešení lékařských nebo nemocničních IS jsou zaměřena primárně na zajištění zdravotnické péče a s tím souvisejících agend. Chybí v nich podpora pro výzkum a tomu odpovídající zpracování dat. Veškerá data z vyšetření jsou proto pouze archivována. Požadavky a návrh IS zaměřeného na výzkum s medicínskými daty je uveden v této sekci.

Výzkumný informační systém (VIS) má sloužit pro organizaci výzkumných medicínských dat, testování hypotéz, zpracování dat a provádění dalších analýz a experimentů včetně trvalého uchování výsledků. Zdrojem dat budou existující zdravotní zařízení, která se do výzkumu zapojí. Pro využití výsledků zdravotnickým zařízením nebo ošetřujícím lékařem je nutná existence zpětné vazby.

3.7.1 Motivace a princip

Výzkumný informační systém má být společnou platformou pro odborníky a pracovníky výzkumu z oblasti medicíny, radiologie, informačních technologií a dalších. Lékaři mají mít možnost prezentovat svoje požadavky, hypotézy, případně metody a nutné podmínky. Požadavky prezentují v rámci společné platformy a vznikne tým lidí, kteří na řešení daného problému spolupracují. Programátoři, výzkumníci z oblasti informačních technologií ale i třeba studenti navrhnou způsoby, jak takové ověření hypotézy realizovat. Výzkumný informační systém jim poskytne zázemí pro přístup k datům. Zabezpečí i testování a uchovávání dočasných i konečných výsledků. Takový výzkum může být přímo konzultován s lékaři a ti mohou sledovat postup prací. Možnost vidět průběh, umožní lékařům včas rozhodnout o změnách. Vše při možnosti zpracovávat rozsáhlé množství dat reálných pacientů.

Obecné schéma spolupráce je uvedeno na obrázku 6. Nemocnice (hospital) poskytuje data pro výzkum. Výzkumník vytváří a implementuje algoritmy, které spuštěním tvoří experimenty. Lékař navrhuje požadavky na experimenty a má možnost prohlížet data a upravovat je. Do budoucna je počítáno i se studenty medicíny, kteří by měli možnost využít výzkumný informační systém pro studium, neboť by obsahoval řadu ohodnocených dat včetně diagnóz, provedené léčby, atd.

Princip systému je v definování požadavku, čeho má být dosaženo a jeho realizace v jednotném prostředí, které řadu dílčích problémů zjednoduší nebo řeší centrálně.

Uvedený princip ilustruje následující požadavek lékařů: *Nalezení objemu mrtvé tkáně z CT vyšetření mozku pacienta po prodělané CMP.* [21] Předpokládáme,

že VIS potřebná data obsahuje a máme i osoby, které takový algoritmus budou vyvíjet a testovat. Vývojáři s lékaři konzultují, jaká konkrétní data od daného pacienta potřebují. Prvním krokem je vytvoření dotazu pro VIS, kterým získají seznam CT vyšetření a DICOM snímků. V následujícím kroku VIS realizuje přenos požadovaných dat na výpočetní stanici, kde bude výpočet probíhat. Další práce je v režii vývojáře, aby s daty prováděl operace takové, aby zjistil hledaný objem. Může využít služby VIS pro trvalé uchování dočasných dat (např. pro dodatečné porovnání) a výsledků a to i v několika verzích. VIS poskytne nástroje pro manipulaci se všemi uchovávanými daty. Lékař má možnost prohlížet výsledky, komentovat a ohodnotit jejich kvalitu nebo i úroveň a místo, kde vznikají chyby. Každý uživatel má možnost přidávat vlastní značky k libovolným datům a i ty lze použít pro výběr dat.

3.7.2 Stručný popis a vymezení hranice

Výzkumný informační systém je navržen jako samostatný IS, přebírající data z některého z existujících lékařských (LIS) nebo nemocničních (NIS) informačních systémů. LIS a NIS musí zajistit export dat, která budou sloužit pro výzkum, a nemusejí zajišťovat jejich anonymizaci. V exportovaných datech musí být zachovány vnitřní vazby, aby bylo možné automaticky strojově rozpoznat, která data tvoří celek a patří ke konkrétní klinické události pacienta. VIS přebírá zodpovědnost za zajištění ochrany osobních údajů provedením anonymizace dat. I další zpracování dat, počínaje jejich přenosem do výzkumného centra, je plně v režii VIS.

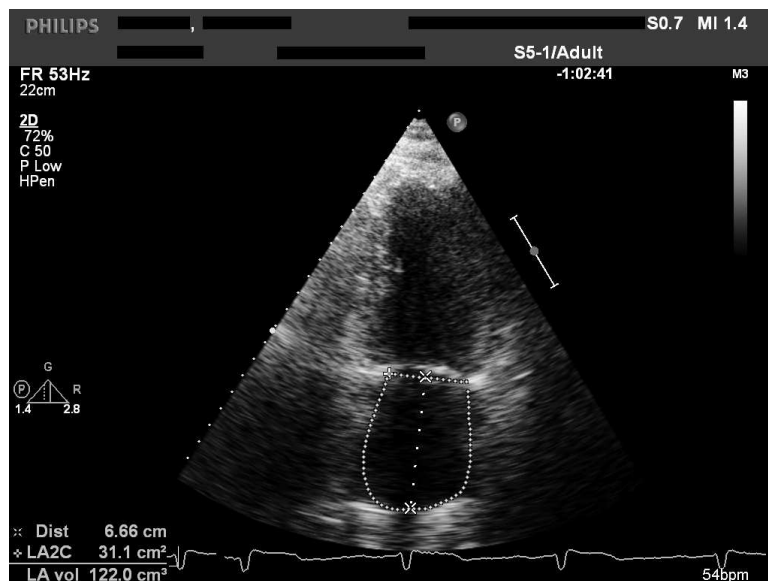
Výzkum probíhá ve VIS, do kterého mohou mít spolupracující zdravotníci přidělen přístup. Mohou se na výzkumu podílet a ovlivňovat směr výzkumu, hodnotit a komentovat výzkumná data i výsledky. Nástroje a software umožňující provádění požadovaného zpracování dat jsou volitelnou a rozšiřitelnou součástí VIS. Musí se řídit pravidly a doporučeními pro jejich tvorbu a komunikaci ve VIS.

Zpětná vazba pro lékaře nebo celé zdravotnické zařízení je rozdělena na dvě části – nepřímá a přímá. K nepřímé zpětné vazbě má přístup spolupracující lékař automaticky. Může výsledky a ohodnocení z VIS použít pro svoji praxi a léčbu pacientů. Naopak přímá zpětná vazba spočívá v dodatečném propojení VIS s LIS nebo NIS. Některá ohodnocení pacientů, klinických událostí, nebo vazby mezi souvisejícími či podobnými klinickými událostmi by se mohly přímo promítnout v LIS nebo NIS. Nutno podotknout, že přímá zpětná vazba by byla za stále platné podmínky anonymity dat ve VIS – tj. namapování výzkumných dat zpět na reálné pacienty bude možné pouze na straně zdravotnického zařízení.

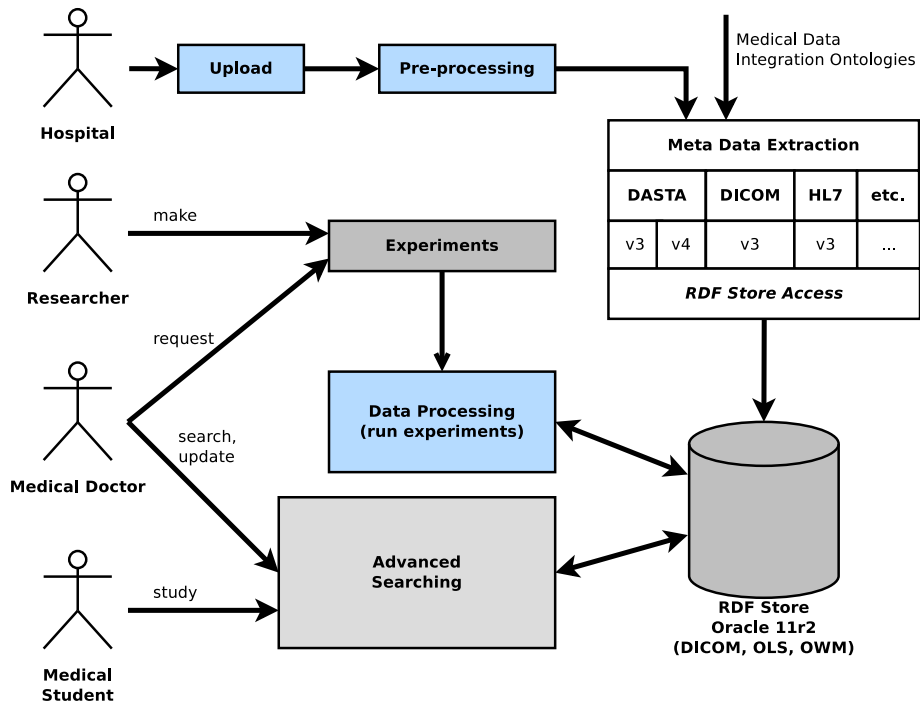
3.7.3 Specifikace požadavků pro podporu výzkumu

1. Příjem dat z NIS, LIS, ... (DICOM Server, NFS, Samba).
2. Ochrana osobních údajů pacientů, zdravotníků i nemocnice (data strukturovaná a nestrukturovaná, obrazová data).
3. Zabezpečený přenos výzkumných dat do výzkumného centra (*scp*, *rsync*, TransferMed).
4. Před-zpracování výzkumných dat.
5. Získání základních metadat z výzkumných dat při importu do výzkumného centra (pacient, klinická událost, přílohy, DICOM studie, série a snímky, SITS data a vazby, další související dokumenty, ...)

6. Model a schéma pro metadata.
7. Řízení přístupu k datům podle role, výzkumné skupiny a oprávnění.
8. Vzájemná spolupráce mezi výzkumnými nástroji.
9. Algoritmus pro výzkum – nástroje a software.
10. Výzkumné experimenty – nastavení pro algoritmus.
11. Podpora triggerů (čas; interval; nová data; manuálně) pro spouštění experimentů.
12. Spouštění experimentů (distribuované výpočty).
13. Uchování výsledků experimentů.
14. Verzování výsledků experimentů.
15. Zpřístupnění heterogenních dat včetně zdrojových výzkumných dat (souborů).
16. Vizualizace uchovávaných dat a výsledků.
17. Administrativní rozhraní.
18. Uživatelé, role, výzkumné skupiny a oprávnění.
19. Podpora vzdělávání studentů medicíny (profil, testy, vyhodnocení).



Obrázek 5: Ukázka DICOM snímku po odstranění identifikace z obrazových dat



Obrázek 6: Spolupráce mezi Fakultní nemocnicí v Plzni a Výzkumným biomedicínským centrem na Západočeské univerzitě v Plzni

4 Návrh zpracování metadat

4.1 Získání metadat

Část získání metadat ilustruje obrázek 6 a jedná se o blok *Meta Data Extraction*. Probíhá bezprostředně po skončení předzpracování přijatých dat z nemocnice.

Pro získání metadat z konkrétního typu souboru je nutné provést následující kroky v pořadí:

1. zjištění typu souboru (DASTA, DICOM, HL7, SITS-XML, SITS-HL7, ...)
2. zjištění verze (pro daný typ souboru)
3. získání metadat ze známého typu a verze souboru

Provedení analýzy pro zjištění typu souboru včetně jeho verze je zásadní pro následující uchování metadat. Rozhodování o typu souboru nemůže být založeno na koncovce souboru a musí být analyzován obsah souboru. Stejně je tomu u verze, která bude zjištěna na základě analýzy obsahu souboru. Je to z důvodu, kdy např. typ souboru DASTA i HL7 jsou oba XML soubory. Implementovaná metoda získávající data ze souborů DASTA a HL7 přitom může být obecná pro všechny XML soubory.

Pro následné vytvoření informace ze získaných metadat, ale mnohdy i dat samotných, je nutno znát původ i kontext. Příkladem může být atribut *ur* v elementu *dasta*, který definuje typ přenášených dat. Na základě typu musí být některé soubory zpracovány přednostně. Jedná-li se o data pacienta, pak rutinní zpracování je indikováno hodnotou *ur* = "R". V případě *ur* = "S" u patientských dat se již jedná o tzv. statimové zpracování, které indikuje nutnost přednostního zpracování dat příjemcem – urgentnost zprávy. Existují i typy přenášených dat pro výkazy a zprávy pro ÚZIS ČR (U), soubor vykazovaných výkonů (V), laboratorní bloky dat (B), číselníky (C), hlášení a zprávy z oblasti "hygiena a epidemiologie" (H), technická (testovací) data (T), data pro Českou správu sociálního zabezpečení (N)

Ve výše navrhovaném výzkumném informačním systému (VIS) je získání metadat a tvorba informací proces, nikoliv jednorázová akce. Prvotní získání metadat je provedeno po předzpracování vstupních anonymních medicínských dat tak, aby bylo získáno co největší možné množství informací – je proveden import vstupních dat do VIS.

4.2 Uchování metadat

K permanentnímu uchování modelu metadat nesoucích informace bude využito Resource Description Frameworku (RDF), které je obecnou metodou pro konceptuální popis nebo modelování informace. RDF je implementováno primárně pro zdroje na internetových stránkách (web) a tento princip provázání informací je velmi výhodný i pro organizaci informací ve VIS. Lze přímo provázat zjištěné nové informace s lékařskou zprávou, pacientem, konkrétní diagnózou nebo např. podmnožinou rizikových faktorů.

Metadata VIS mají obsahovat množinu informací o následujících oblastech zájmu – medicínská data a jejich obsah, podpora výzkumu, podpora vzdělávání a systémové informace a zajištění zabezpečení systému.

1. Medicínská data

- popis zdrojových dat (souborů, přílohy lékařských zpráv, ...),
- lékařská zpráva (z DASTA a HL7),
- pacient,
- diagnóza,
- léčba,
- medikace,
- NIHSS vyšetření (SITS),
- DICOM snímky (studie, série, snímek, zařízení, pacient, ...).

2. Podpora výzkumu

- algoritmus,
- experiment,
- výsledky experimentu,
- verze,
- pravidla spuštění a plánování,
- a plánování experimentu.

3. Podpora vzdělávání

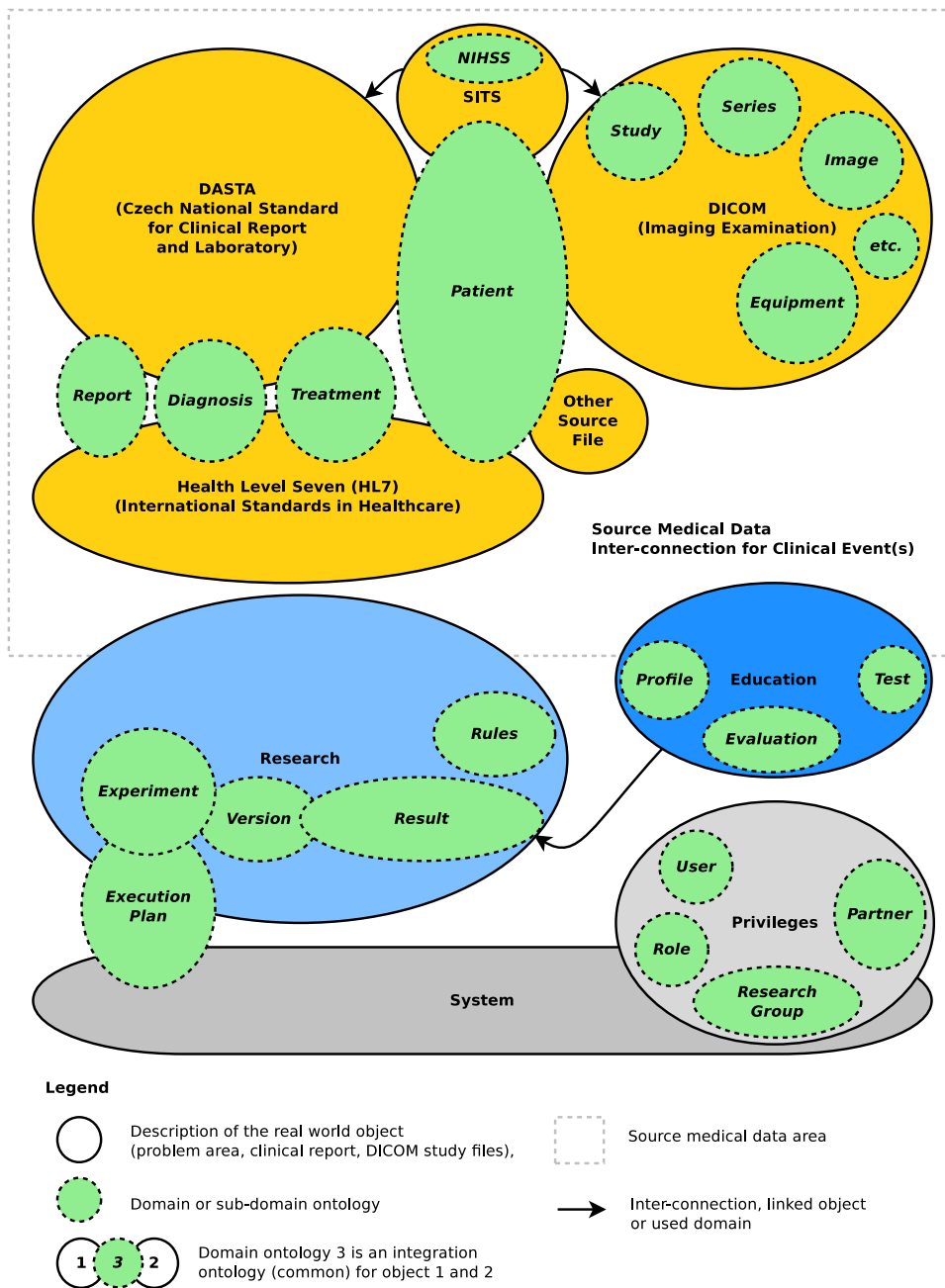
- studijní profil,
- popis testů studenta,
- a hodnocení výsledků.

4. Systémové informace a zabezpečení přístupu

- systémové informace,
- přístupová práva,
- uživatel,
- role uživatele (výzkumník, lékař, programátor, administrátor),
- a výzkumná skupina.

Metadata budou organizována s použitím řady ontologií pro jednotlivé oblasti, které bude z velké části nutné nejprve navrhnout a vytvořit. Ilustrační schéma na obrázku 7 ukazuje navrhované ontologie včetně jejich struktury. Cílem tohoto návrhu je sjednotit přístup k informacím získaných ze zdrojových souborů, v nichž existuje řada duplicit a přes počítačové zpracování i neúplných vazeb. Příklad duplicitního uvedení informace je viditelný na obrázku 7, kde jsou informace o pacientovi definovány v každém DASTA, HL7 i DICOM souboru. Liší se pouze množstvím detailů, které soubor obsahuje. Nejvíce detailů obsahuje lékařská zpráva (DASTA a HL7). Identifikaci pacienta (anonymizovanou) ponecháme stranou, neboť ta je nutná pro provázání zdrojových dat mezi sebou. Mezi častými duplicitami se objevuje datum narození a pohlaví.

Zdrojové DICOM soubory obsahují značně množství duplicit, které je dáno formátem souboru. Každý DICOM soubor obsahuje veškerá metadata nutná nejen pro jeho interpretaci. Přitom DICOM souborů je v jedné CT studii i několik tisíc. Všechny soubory pak obsahují shodné informace o použitém zařízení, jeho výrobcí,



Obrázek 7: Navrhované ontologie výzkumného informačního systému včetně ontologií integrujících duplicitní informace

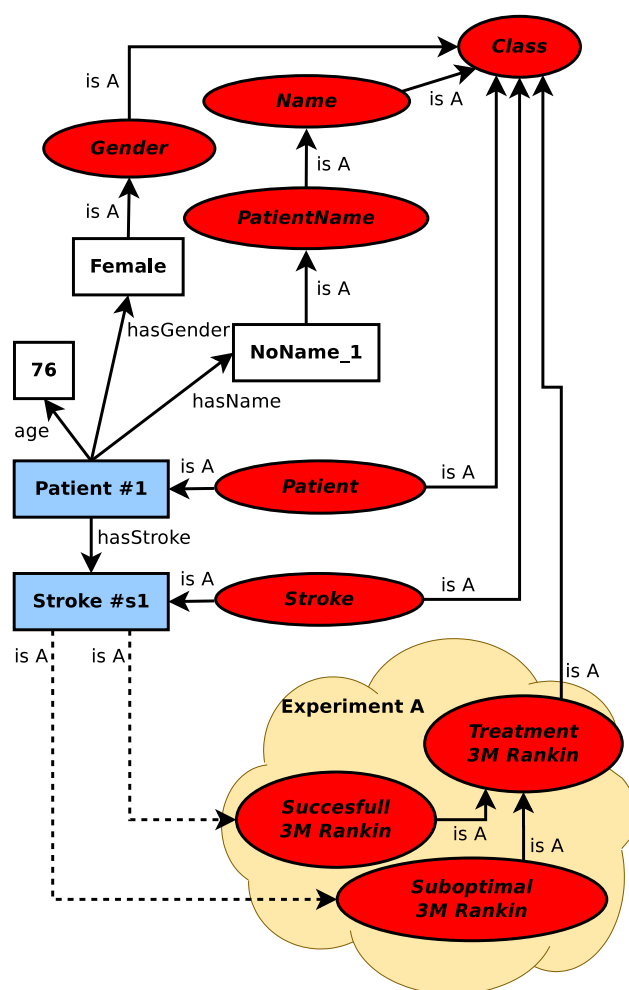
verzi, datum kalibrace, apod. Při uchování metadat je proto vhodné a s použitím RDF i možné odstranit duplicitu uchovávaných informací ve VIS.

Pro fyzické uchování metadat v RDF bude použito vhodné úložiště. Jeho volba bude závislá především na rychlostech odezvy při vkládání a vyhledávání RDF dat a možnostech zabezpečení.

4.3 Interoperabilita a použití metadat

Kvůli zajištění interoperability metadat, tedy především možnosti jejich opakovaného použití, je největší důraz kladen na použití ontologií. Ontologii lze považovat za sdílený nebo systematický slovník, který modeluje popisovanou doménu – definici objektů a/nebo konceptů včetně jejich vazeb a vlastností. Ontologie je tedy nástroj pro organizaci informací a jako forma reprezentace znalostí.

Cílem je zajistit vnitřní i vnější interoperabilitu metadat – informací ve VIS. Vnitřní interoperabilitou VIS je myšleno využití primárně pro různé výzkumné algoritmy a experimenty, které informace a data z VIS zpracovávají. Vnější interoperabilitou je míněna přenositelnost informace i mimo VIS, např. v národním i mezinárodním měřítku.



Obrázek 8: Princip rozšiřování výzkumného informačního systému s novým experimentem označeným *Experiment A*. Pro provedení experimentu vznikne nejvýše jedna z vazeb naznačených čárkovaně.

Schéma z obrázku 7 popisuje ontologie, které budou ve VIS používány a pomocí nichž mohou další algoritmy efektivně spolupracovat. Schéma z obrázku by pro obecné využití ve VIS bylo omezující, a proto je nutné jej brát jako základ pro další

rozšiřování systému. Růst rozsahu informací a znalostí o zkoumané problematice je možný přidáním dalších ontologií. Princip použití a rozšiřování systému ilustruje obrázek 8, kde rozšíření o nové informace přináší použití experimentu *A*. Experiment *A* v tomto případě nejen že využívá existující ontologie, ale přináší rozšíření o nové třídy *Treatment3MRankin* (hodnocení výsledku léčby po 3 měsících), *Succesfull3MRankin* (léčba hodnocena za úspěšnou) a *Suboptimal3MRankin* (léčba hodnocena jako neoptimální). Uvedené třídy nebyly ve VIS do zavedení experimentu *A* známy. Experiment *A* provede ohodnocení jednotlivých případů cévní mozkové příhody po jeho spuštění. Jakýkoliv další experiment už může takové ohodnocení přímo použít například tím, že zkoumá podobnost případů, kdy byla léčba neoptimální (ohodnocení *Suboptimal3MRankin*), a o výsledky opět rozšířit dosavadní informace ve VIS. Uvedený příklad ilustruje navrhovaný princip rozšiřování znalostí systému a primárně vnitřní interoperabilitu.

Krokem k zajištění vnější interoperability je použití sdílených a systematických slovníků. Mezinárodními slovníky jsou např. *International Classification of Diseases (ICD)*, *Logical Observation Identifiers Names and Codes (LOINC)* a *Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT)*. V České republice je ICD převzato pod názvem *Mezinárodní statistická klasifikace nemocí a přidružených zdravotních problémů* a je označována zkratkou MKN. Dále v České republice existuje vlastní *Národní číselník laboratorních položek (NČLP)* zavedený Ministerstvem zdravotnictví České republiky. NČLP a MKN jsou používány v souborech DASTA. Mezinárodní interoperabilita NČLP neexistuje.

Novou informaci získáme řízeným zpracováním zdrojových dat jako jsou přímo CT snímky s využitím již známých informací z VIS experimentem např. ověření hypotéz. Na základě výsledku platnosti hypotézy může až lékař nebo i právě probíhající experiment definovat nové informace, které rozšíří znalosti VIS. Význam VIS je v tom, že existenci informace nebo vazby mezi informacemi předem neznáme a vznikne až zpracováním jako výsledek výzkumu (experimentu).

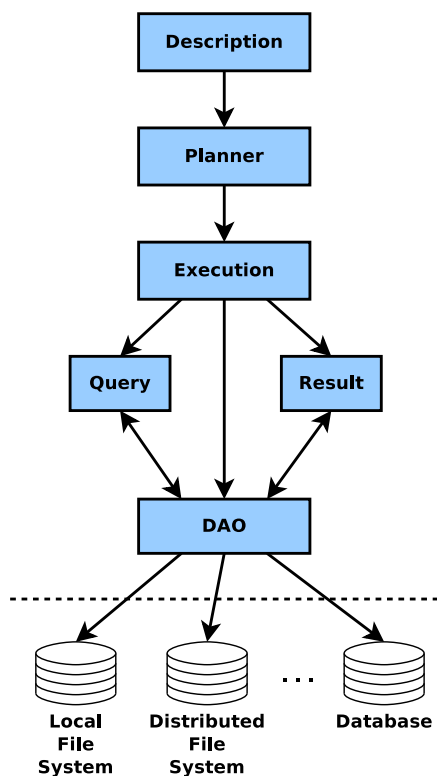
4.4 Schéma provedení experimentu

Obrázek 9 ukazuje navrhovaný princip provedení zpracování dat experimentem. Každý experiment je popsán (Description) sadou informací týkající se názvu, příkazu pro vykonání, parametrů, volby spouštěče apod. Plánovač (Planner) ve VIS má za úkol v očekávaném čase experiment spustit. Plánovač rozhoduje na základě definovaného spouštěče (např. příchod nových dat, definovaný čas nebo manuální vyvolání uživatelem), zatížení a priorit (mohou se objevit data, která mají být přednostně zpracována a vyhodnocena).

Spuštěný experiment na počátku vytvoří novou identifikaci verze spuštění a potřebuje získat data. Proto provede dotazování (Query) ve VIS. Dotaz je proveden přes vrstvu pro přístup k datům (DAO), která zajišťuje oddělení od fyzického uložení dat. Dotaz je proveden, vyhodnocen a výsledek dotazu se vrátí do běžícího experimentu. Následuje výkonná část, kdy dochází k vlastnímu zpracování dat. V této fázi může experiment dotazovat libovolné informace i zdrojová data jako jsou např. přímo CT snímky. Pro uložení výsledků (Result) je opět definováno aplikační rozhraní zjednodušující tuto činnost. Dočasné i trvalé výsledky mohou být ukládány zpět do VIS. Všechny výsledky budou zaobaleny a uloženy včetně informace o identifikaci verze a příslušnosti k experimentu.

Uchovávat ve VIS lze čistě textové informace, soubory a to i binární. Část uchování dat zajišťuje zmíněná vrstva pro přístup k datům (DAO) a fyzicky je možné pracovat s daty v lokálním nebo distribuovaném souborovém systému, či v databázi.

Použitím identifikace verze spuštění lze docílit možnosti mít více výsledků od jednoho experimentu a archivovat je v čase. Tento princip bude výhodný při vývoji, ladění a testování experimentů. Z jakých informací bude identifikace verze spuštění složena definuje každý autor experimentu. Volbou identifikace pomocí parametrů experimentu lze docílit možnosti porovnávat změny výsledku v závislosti na použitých parametrech. Volbou data a času lze např. docílit uchování historie změn výsledku/hodnoty v čase. Lze kombinovat i více metod pro identifikaci verze spuštění (např. datum a čas společně s parametry).



Obrázek 9: Experiment a aplikační rozhraní pro jeho vykonání v navrhovaném výzkumném informačním systému.

5 Ověření navrženého zpracování metadat

V této sekci diskutuji způsoby ověření správnosti navrženého zpracování metadat ve výzkumném informačním systému pro medicínská data.

5.1 Anonymní medicínská data

Pro zachování ochrany osobních údajů je nutné provést anonymizaci a zajistit anonymitu výzkumných dat. Anonymizace je prvním a nutným krokem, kdy dochází ke zpracování dat a metadat.

Na anonymizaci dat je navržen software, který před přenosem dat ze zdravotnického zařízení odstraní identitu pacienta. Úroveň anonymizace závisí na požadavcích zdravotnického zařízení na výzkum. Některé údaje mohou nepřímo napomoci re-identifikaci pacienta neboť se může jednat o unikátní výskyt údajů v rámci množiny výzkumných dat.

Ověření anonymity dat lze rozdělit na případy, kdy dojde:

1. k přenosu anonymních dat – veškeré údaje identifikující pacienta byly odstraněny,
2. k přenosu částečně anonymních dat – například:
 - (a) lidský faktor, tj. lékař zapíše pacienta identifikující údaje do textové části lékařské zprávy a tyto se nepodaří úplně odstranit,
 - (b) identifikace pacienta je zapsána také do obrazové části dat,
3. a k přenosu neznámých a potenciálně neanonymních dat –např. přílohy neznámé přílohy k lékařské zprávě, kterou anonymizační proces nezná.

Při automatickém zpracování nelze uvedeným případům zcela zabránit. Pro ověření účinnosti anonymizace určíme hranici 0,01 % jako maximální povolený počet selhání. Už v návrhu problém minimalizujeme vynecháním souborů z automatického zpracování u takových případů, kde není jistota dokonalé anonymizace, např. v obrazové informaci v DICOM snímcích.

Z uvedeného seznamu je první případ jediný požadovaný stav. Poslední případ je také v pořádku, protože by neměl nastat. Neznámý typ souboru je při anonymizaci ponechán bez změny a není předáván na výstup.

5.2 Získání základních metadat při importu

Hodnotit úspěšně získání základních metadat vstupních dat ze zdravotnického zařízení, prováděné po případném před-zpracování, lze pouze v kompletně nahraných datech. Úspěšné je, když jsou všechny související informace propojeny vazbami. Lze se přes vazby dostat na veškeré nahrané soubory pacienta, klinické události, atd.

Je-li chyba v přijatých datech a nelze data zcela propojit, není možno je postoupit k výzkumným účelům.

5.3 Tvorba metadat experimenty

Ověření navrženého rozšiřování metadat ve VIS lze prováděním spouštění experimentů. Když experiment může uložit výsledky a rozšířit uchovávané informace ve VIS i o zcela nové, pak je úkol tvorby metadat experimenty splněn.

Experimenty musí mít možnost dotázat a získat požadované informace z VIS i přímo zdrojových souborů. Stejně tak lze jako výsledek experimentu uložit textovou informaci, i komplikované datové struktury včetně celých souborů.

Může existovat závislost mezi experimenty ve smyslu, že experiment X využívá výsledků/ontologie, které vytváří experiment Y . Na tomto principu je VIS založen a proto musí všechny výsledky v organizovány v podobě ontologií, o které lze VIS rozšiřovat. Při rozšiřování o nové ontologie nesmí dojít k případu, kdy spolu budou dvě nebo více ontologií kolidovat.

5.4 Interoperabilita metadat

Interoperabilní metadata jsou jasně definována v dokumentaci ontologie – význam, popis, hodnota a určení obsahu informace. V ontologii nesmí existovat definice, které jsou ve vzájemném rozporu.

Ve VIS musí být striktně používáno již existujících identifikátorů metadat se stejným významem a obsahem, před zaváděním duplicitních.

Pro dosažení interoperability mezi VIS a vnějším okolím lze doplnit integrační ontologii nebo mapování definující synonyma.

5.5 Uchování metadat a rychlost odezvy

Pro uchování metadat a předpokládaný velký rozsah zpracovávaných medicínských dat je nutné vybrat úložiště poskytující vysokou rychlost při přidávání nových RDF dat, nejkratší odezvy na vykonání dotazu a podporující použití nástroje pro odvozování informací.

Cílem je identifikovat a odstranit kritická místa propustnosti systému a dosažení nejvyššího výkonu pro poskytnutí rychlého zpracování prioritních klinických událostí.

6 Závěr

V předložené práci byl popsán navržený způsob zpracování metadat v informačním systému pro podporu výzkumu v oblasti medicíny. Kladem je možnost rozšiřování rozsahu uchovávaných informací. V sekci 2 je problematika metadat, RDF, používaných a podporovaných medicínských typů dat a interoperabilita. Aktuální stav a související práce je diskutována v sekci 3. Je popsán stav formátu DASTA a HL7, popis vlastního formátu pro data ze SITS registru. Nechybí stručná osnova vyšetření a léčby pacientů s cévní mozkovou příhodou. Zmíněna je ochrana osobních údajů pacientů i nutný proces anonymizace dat. Pro ilustraci je uvedena motivace a princip výzkumného informačního systému včetně stručné specifikace požadavků. Sekce 4 popisuje navrhované metody pro zpracování metadat – získání, trvalé uložení a popisuje, jak je zajištěna interoperabilita a použití metadat. V poslední sekci 5 jsou diskutovány možnosti ověření správnosti navrženého řešení zpracování metadat a jejich interoperability.

Zpracování metadat je navrhováno pro přímé využití při výzkumu cévních mozkových příhod (CMP). Primární zaměření na výzkum CMP vedlo k požadavku na výzkumný informační systém. Rozšíření na výzkum v dalších oblastech medicíny spočívá pouze v získání odpovídajících medicínských dat a implementaci požadovaných výzkumných experimentů, které budou provádět požadované analýzy či např. testování hypotéz. V budoucnu je reálné i další zobecnění výzkumného informačního systému pro podporu provádění výzkumu s jinými než medicínskými daty. Pro uvedená rozšíření a zobecnění systému zůstávají metody a principy navrženého zpracování metadat v platnosti.

Zpracovaná metadata a získané informace o datech ve výzkumném informačním systému mohou sloužit pro navazující znalostní, expertní ale i výukové systémy. Ontologie pomáhají zajistit přiřazení jednoznačného významu každé uchovávané informaci, které vede k zajištění interoperability informací s obecně dalšími systémy.

7 Návrh cílů dizertační práce

1. Návrh a tvorba ontologií pro popis medicínských, výzkumných, systémových i výukových metadat ve výzkumném informačním systému.
2. Implementace získávání metadat v prostředí výzkumného informačního systému a uložení s využitím existujících a navržených ontologií.
3. Volba úložiště pro RDF data.
4. Testování, diskuze výsledků.

Reference

- [1] P. Awareness, E. Referral, P. Transfer, S.U. Care, D. Imaging, G. Principles, B. Tests, P. Prevention, H.B. Pressure, C. Smoking, et al. Guidelines for Management of Ischaemic Stroke and Transient Ischaemic Attack 2008. *Cerebrovasc Dis*, 25:457–507, 2008.
- [2] Bruce E. Bargmeyer and Daniel W. Gillman. Metadata standards and metadata registries. 2002.
- [3] Warwick Cathro. Metadata: An overview. 1997.
- [4] CSU – Czech Statistical Office. Český statistický úřad: Úmrtnostní tabulky (Death-rate Statistics). Online, 2010-03-02. http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky, 2010.
- [5] R.H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F.M. Behlen, P.V. Biron, and A. Shabo Shvo. HL7 clinical document architecture, release 2. *Journal of the American Medical Informatics Association*, 13(1):30, 2006.
- [6] Dublin Core Metadata Initiative. Semantic recommendations. 2011.
- [7] Boris Eninger. Přenos dat z registru SITS ve formátu HL7. Diplomová práce. Západočeská univerzita v Plzni., 2011.
- [8] K. Hüner, B. Otto, and H. Österle. Collaborative management of business metadata. 2011.
- [9] Health Level Seven International. HL7 faqs. Online, 2011-07-20. <http://www.hl7.org/about/FAQs/index.cfm>, 2011.
- [10] International Organization for Standardization. ISO/IEC 11179, 2009.
- [11] International Organization for Standardization. ISO/IEC 11179-1:2004 Information technology – Metadata registries (MDR) – Part 1: Framework, 2009.
- [12] E. Janečková and Česko. *Zákon o ochraně osobních údajů s komentářem*. Právo (ANAG). ANAG, 2010.
- [13] Jules J. Bernman. HHS Workshop on the HIPAA Privacy Rule’s De-Identification Standard. HHS Workshop, March 8–9, 2010, Marriot at Metro Center, Washington, DC, March 8, 2010.
- [14] Bob Kemp and Jesus Olivan. European data format [‘]plus’ (EDF+), an EDF alike standard format for the exchange of physiological data. *Clinical Neurophysiology*, 114(9):1755 – 1761, 2003.
- [15] Ministerstvo zdravotnictví České republiky. Datový blok ip – pacient – varianta pro ds4. Online, 2011-07-20. http://ciselniky.dasta.mzcr.cz/CD_DS4/hypertext/MZAU0.htm, 2011.
- [16] MZCR – Ministry of Health of the Czech Republic. Ministerstvo zdravotnictví České Republiky: Věstník č. 2/2010: Péče o pacienty s cerebrovaskulárním onemocněním v České republice. Online, 2010-03-01. http://legislativa.mzcr.cz/File.ashx?id=233&name=V%C4%9Bstn%C3%ADk_%20%C4%8D_02_2010.pdf, 2010.
- [17] National Electrical Manufacturers Association (NEMA). Digital Imaging and Communications in Medicine (DICOM). Online, 2011-07-20. <http://medical.nema.org>, 2011.

- [18] National Information Standards Organization (NISO). Understanding metadata. 2002.
- [19] Mikael Nilsson, Thomas Baker, and Pete Johnston. Interoperability levels for dublin core metadata. 2009.
- [20] J.T. O'Brien, T. Erkinjuntti, B. Reisberg, G. Roman, T. Sawada, L. Pantoni, J.V. Bowler, C. Ballard, C. DeCarli, P.B. Gorelick, et al. Vascular cognitive impairment. *The Lancet Neurology*, 2(2):89–98, 2003.
- [21] V. Rohan, P. Sevcik, J. Polivka, Z. Ambler, B. Kreuzberg, and J. Ferda. Klinický pohled na výpočetní tomografii u akutní ischemie mozku (A clinical Approach to Computed Tomography in Acute Cerebral Ischemia). *Česká a slovenská neurologie a neurochirurgie*, 2007.
- [22] PM Rothwell, AJ Coull, LE Silver, JF Fairhead, MF Giles, CE Lovelock, JNE Redgrave, LM Bull, SJV Welch, FC Cuthbertson, et al. Population-based study of event-rate, incidence, case fatality, and mortality for all acute vascular events in all arterial territories (oxford vascular study). *The Lancet*, 366(9499):1773–1783, 2005.
- [23] Safe Implementation of Thrombolysis in Stroke (SITS). About safe implementation of thrombolysis in stroke. Online, 2011-07-20. <http://www.acutestroke.org>.
- [24] Libor Seidl. H17 Česká republika. Online, 2011-07-20. <http://www.h17.cz/cz/h17/about.html>, 2011.
- [25] W3C. Resource Description Framework (RDF): Concepts and Abstract Syntax. Online, 2010-03-02. <http://www.w3.org/TR/rdf-concepts/>, 2010.
- [26] Miroslav Zámečník. Aktuální otázky k Datovému standardu MZ ČR. 2011.
- [27] Miroslav Zámečník, Luděk Straka, and Antonín Jabor. Dasta a projekty e-health – současný stav, rozvoj, otázky. Online, 2011-07-20. <http://ciselniky.dasta.stapro.cz/hypertext/201130/hypertext/MZAWF.htm>, 2008.