Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Bakalářská práce

Bezpečnostní analýza v oblasti elektronické výměny informací o pacientech

Plzeň, 2015 Jaroslav Malát

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 25. června 2015

Jaroslav Malát

Poděkování

Úvodem bych chtěl poděkovat své vedoucí bakalářské práce doc. Dr. Ing. Janě Klečkové za důležité připomínky a rady k formální i obsahové stránce práce.

Abstract

Theme: Security analysis of the electronic exchange of patient's information

The presented bachelor's thesis is focused on the problem of security in the electronic exchange of patient's information. Over the years were developed methods and algorithms for securing private data in the electronic exchange of information between two nodes.

The theoretical part of bachelor thesis is focused on legal regulations and laws for handling private data about patients and presents the most widely used standard for image files - DICOM. There is also described the history of deploying DICOM standard in Czech Republic.

The practical part is focused on the development of program for anonymizing non-standard entries of private data directly into the results of the examinations (CT, X-ray, etc.). The program is developed in programming language Java and I operates on system Linux Ubuntu 14.04LTS.

Abstrakt

Téma: Bezpečnostní analýza v oblasti elektronické výměny informací o pacientech

Předkládaná bakalářská práce se zaměřuje na problém bezpečnosti v oblasti elektronické výměny informací o pacientech. V průběhu let byly vytvořeny metody a algoritmy pro zabezpečení citlivých údajů v oblasti elektronické výměny informací mezi dvěma uzly.

Teoretická část bakalářské práce se zaměřuje na právní předpisy a zákony pro práci s citlivými údaji o pacientech a seznámením s nejpoužívanějším standardem pro obrazové soubory – DICOM. Dále je zde popsána historie nasazení DICOM standardu v České republice.

Praktická část je zaměřena na vývoj programu pro anonymizaci nestandardních zápisů citlivých dat přímo do výsledků vyšetření (CT, rentgen, atd.). Program je vyvíjen v programovacím jazyce Java a na operačním systému LINUX UBUNTU 14.04LTS.

Obsah

1. Úvod

V průběhu let byly vytvořeny metody a algoritmy pro zabezpečení citlivých údajů v oblasti elektronické výměny informací mezi dvěma uzly. Na základě analýzy byla zjištěna část, které nebyla řešena a to nestandardní zápis citlivých údajů přímo do výsledku vyšetření (CT, rentgen, atd.).

Původní téma bylo již částečně zpracováno, rozhodl jsem se tedy v této práci zaměřit spíše na implementaci řešení pro nestandardní zápis citlivých dat přímo do výsledků vyšetření.

Cílem mojí bakalářské práce je tedy analýza legislativních bezpečnostních požadavků pro zpracování medicínských dat a vytvoření anonymizačního programu pro zpracování citlivých dat obsažených v obrazové části DICOM souborů. Bude taky popsáno nasazení DICOM standardu v České republice. Můj anonymizační program vyhodnotí, zda obrazový soubor obsahuje jakákoliv citlivá data, ať už o lékaři nebo o pacientovi, odpovídající předem daným předpisům a následně tato data anonymizuje.

Na úvod provedu výtah z legislativy České republiky platné k začátku roku 2015, která se dotýká práce s citlivými daty.

Realizace je rozdělena do stejného počtu bodů. Jako jednotlivé body v zadání bakalářské práce jsem stanovil:

- Analýza legislativních bezpečnostních požadavků pro zpracování medicínských dat, nařízení a omezení při práci s daty pacientů
- Seznámení se standardem DICOM a jeho nasazení v ČR
- Analýza anonymizace obrazových dat
- Implementace

2. Úvod do bezpečnostní problematiky

Na úvod bych rád definoval pojem zdravotnické dokumentace.
Zdravotnická dokumentace je souhrn informací o pacientovi (klientovi) zdravotnického zařízení, vedený v jakékoliv podobě.

Tato dokumentace má především sloužit jako pracovní nástroj při léčbě, ale případně i jako doklad či dokonce důkaz v případě forenzního projednávaní postupu lékaře při léčení.

Nesprávně vedená dokumentace může pomoci v utvrzení o chybném postupu nebo přinejmenším znemožnit dokázání postupu správného.

Většina zdravotnických zařízení ať prvního styku:

- praktický lékař pro dospělé
- zubní lékař resp. stomatologická ambulance
- praktický lékař pro děti a dorost - pediatrie, gynekologická ambulance, lékařská služba první pomoci

či ambulantních:

- oční lékař resp. oftalmologická ambulance
- ortopedie
- psychiatrie
- neurologie
- dermatologie
- rehabilitace
- urologie
- klinická psychologie a logopedie
- ORL – otorhinolaryngologie
- alergologie
- dermatovenerologie
- zdravotní rehabilitace

popř. hospitalizačních:

- nemocnice
- porodnice
- nemocnice následné péče
- fakultní nemocnice
- léčebna dlouhodobě nemocných
- odborný léčebný ústav
- psychiatrická léčebna

nevyjímaje lékárny, laboratoře a lázeňská zdravotní zařízení, dnes využívají informační systémy s údaji o pacientech včetně jména a adresy, rodného čísla, platebních informací a zejména pak citlivé údaje o průběhu léčby.
Tato data jsou nesporně mnohonásobně více ohrožena oproti papírové formě.

Slabými místy při procesu nakládání se zdravotnickými informacemi nejsou technologie, ale lidský prvek, který bývá často opomíjen.

Z konkrétních případů lze uvést například hackerský útok na informační síť určitého zdravotnického zařízení. Pokud půjdeme do krajnosti, mohla by být ohrožena i celková zdravotnická péče. Můžeme tvrdit, že absolutní bezpečnost informačních systémů je vždy pouze teoretickým pojmem.

Jako argument uvádím překvapující zjištění z kontrolní činnosti ÚOOÚ (Úřadu pro ochranu osobních údajů), více na [1].

- Elektronická podoba zdravotnické dokumentace nebyla totožná s tištěnou.
- Informační systém nemocnice neumožňoval aktivní verzi sledování přístupu do něj, což je ze zákona povinné a musí to být zaznamenáváno.
- Do informačního systému nemocnice vstoupila neoprávněná osoba, přičemž ji nebylo možno ověřit kvůli vypnuté funkci monitorování přístupu.
- Velmi neuváženým počinem bylo zveřejnění záznamu z operačního zákroku na webu nemocnice. Byla zřejmá identifikace osoby a k tomu byl ještě záznam doplněn jménem, částí příjmení a dokonce rodným číslem. I přes doklad o písemném souhlasu musel být záznam okamžitě odstraněn.

- Došlo k úniku citlivých informací o pacientovi. Lékař neznající bezpečnostní prvky hesel komunikoval s pacienty přes e-mail s velmi jednoduchým a nedostatečným heslem, které bylo prolomeno.
- Těžko uvěřitelný je také případ nestátního zdravotnického zařízení, které mělo registrační skříně v čekárně pro pacienty. V ordinaci se střídaly dvě lékařky a nedocházelo k důslednému zamykání skříní. Výsledkem bylo, že kdokoliv v čekárně mohl mít snadný přístup ke zdravotnické dokumentaci ostatních pacientů.

Dalším rizikovým faktorem je bezesporu to, že k citlivým datům přistupují nejen zdravotníci, ale i různé dodavatelské firmy, správci informační sítě apod.

Dále je k datům vyžadován četnější přístup ve srovnání s klasickou papírovou dokumentací.

Jak jsem již řekl, bohužel nelze docílit dokonalé ochrany informačního systému, ale snahou by mělo být dosáhnout optimální úrovně zabezpečení. Zdravotnická zařízení by měla mít vypracován plán kybernetické bezpečnosti a plán krizové připravenosti, který popisuje možné rizikové situace jak vně perimetru (mimo nemocnici – živelná pohroma, teroristický útok, velká dopravní nehoda), tak uvnitř (požár, teroristický útok). Kybernetický útok je nebezpečí hrozící stále většímu množství lidí. Zvenku je to především možnost průniku internetem, WI-FI sítěmi nebo GSM sítí z chytrých zařízení. Zevnitř jde o lidský prvek - nevzdělaného nebo nedisciplinovaného uživatele. Ve většině případů zdravotnických zařízení nejde pouze o snahu ochránit osobní údaje pacientů, ale také o udržení jejich chodu. V tomto ohledu zůstávají zdravotnická zařízení mnohem citlivějším místem k útoku než například státní instituce či banka. Velmi důležité je školit uživatele IT systémů a eliminovat tak případné chyby lidského faktoru.

Dalšími důvody většího zabezpečení dat ve zdravotnictví lépe než v jiných oborech lidské činnosti je například chybná diagnóza na základě pozměněných údajů, následné pochybení v léčbě a bezprostřední ohrožení zdraví i života samého.

Požadována je dostupnost jak životně důležitých dat a údajů v případě ohrožení zdraví či přímo života pacienta, tak dostupnost údajů pacienta různými odděleními, dále dostupnost pro služby, změny personálu, zástupy a rozlišení různé citlivosti dat pacienta.

Aby ochrana byla účinná, měli bychom znát potenciální slabá místa a možné útočníky (vnitřní i vnější), míru ohrožení, případné postupy, nutné náklady na eliminaci rizik a typ hrozby či útoku, jaká IS vrstva je ohrožena (infrastruktura, OS, DB, aplikace), výstupy mimo informační systém zdravotnického zařízení a způsoby zabezpečení výměny dat mezi zdravotnickými zařízeními, pojišťovnami a dalšími subjekty.

3. Právní předpisy

Z hlediska práce se zdravotnickou dokumentací jsou v platné české legislativě roku 2015 tyto důležité platné zákony a vyhlášky:

- Zákon č. 96/2001 Sb., o lidských právech a biomedicíně
- Zákon č. 372/2011 Sb., o zdravotních službách
- Vyhláška č. 98/2012 Sb., o zdravotnické dokumentaci
- Zákon č. 227/2000 Sb., o elektronickém podpisu
- Zákon č. 101/2000 Sb., o ochraně osobních údajů
- Vyhláška č. 62/2015 Sb., o provedení některých ustanovení zákona o zdravotnických prostředcích
- Zákon č. 181/2014 Sb., o kybernetické bezpečnosti
- Vyhláška č. 316/2014 Sb., o kybernetické bezpečnosti
- Vyhláška č. 317/2014 Sb., o významných informačních systémech a jejich určujících kritériích

Více informací zákonech na [2] nebo o kybernetickém zákoně na [3].

3.1. Zákon č. 96/2001 Sb., o lidských právech v biomedicíně

Každý má právo na ochranu soukromí ve vztahu k informacím o svém zdraví.

Každý je oprávněn znát veškeré své informace shromažďované o jeho zdravotním stavu a je nutno respektovat i přání každého nebýt takto informován.

3.2 Zákon č. 372/2011 Sb., o zdravotních službách

Část šestá zákona se věnuje přímo zdravotnické dokumentaci a národnímu zdravotnickému informačnímu systému.

Hlava první určuje zapracování osobních údajů, hlava druhá zdravotnickou dokumentaci, její vedení a nakládání s ní, možnosti do jejího nahlížení či pořizování kopií.

Hlava třetí je pak věnována přímo Národnímu zdravotnickému informačnímu systému (NZIS).

NZIS je jednotný celostátní informační systém veřejné správy určený k:

- zpracování údajů o zdravotním stavu obyvatelstva, o činnosti poskytovatelů, o zdravotnických pracovnicích
- vedení Národních zdravotních registrů a zpracování údajů v nich uvedených
- vedení Národních registrů poskytovatelů a zdravotnických pracovníků a zpracování údajů v nich uvedených
- zpracování údajů pro statistické účely a k zpracování různých šetření.

NZIS je de facto systémem, který sdružuje a spravuje informace obsažené v:

- národních zdravotních registrech (Národní onkologický registr, Národní registr hospitalizovaných, Národní registr kloubních náhrad a jiné)
- Národním registru poskytovatelů

- Národním registru zdravotnických pracovníků
- Národních registrech podle zákona o transplantacích
- informačních systémech infekčních nemocí vedených podle zákona o ochraně veřejného zdraví

Současná právní úprava po novele rozšiřuje určení NZIS a kromě Národních zdravotních registrů předpokládá i existenci Národního registru poskytovatelů a Národního registru zdravotnických pracovníků.

Dle původní úpravy bylo možné také v registrech zpracovávat bez souhlasu subjektu jejich osobní a další údaje, tyto však byly přesně vymezené. V novém zákoně je oproti tomu rozsah zpracovávaných údajů rozšířen, dle něj se předávají i údaje související se zdravotním stavem pacienta ve vztahu k onemocnění a jeho léčbě, a to zejména:

- údaje socio-demografické a diagnostické
- údaje o osobní, rodinné a pracovní anamnéze pacienta související s onemocněním včetně posouzení jeho aktuálního zdravotního stavu
- údaje o poskytovaných zdravotních službách pacientovi
- údaje o výkonu povolání nebo zaměstnání, popřípadě o výkonu služebního poměru, potřebné pro posouzení zdravotního stavu pacienta.

Jedná se tedy o demonstrativní výčet, který je, oproti úpravě v zákoně o péči o zdraví lidu, rozšiřující a nelze úplně jasně určit, které všechny informace bude možné pod toto ustanovení podřadit. Zákon dále uvádí rozsáhlý seznam subjektů povinných poskytovat informace.

Osoby oprávněné pracovat s informacemi z NZIS jsou určeny přímo v zákoně o zdravotních službách. Jedná se o oprávněné pracovníky správce a zpracovatele registru, zdravotnické pracovníky, oprávněné pracovníky Koordinačního střediska transplantací a oprávněné pracovníky institucí, které jsou zákonem zmocněny k využívání dat z NSIZ.

Strany zastávající úpravu NZIS tvrdí, že tato pouze kodifikuje stav, který už existoval předtím. Odpůrci tvrdí, že rozsah sdělovaných dat překračuje potřebnou

míru, zákonná úprava je příliš obecná, a přenáší tak rozhodovaní o tom, jaká citlivá data budou sdělována na subjekt jiný, než je zákonodárce. Kritici dále tvrdí, že v zákoně chybí účelné a přesně stanovené postupy zpracovávání citlivých údajů občanů a také řešení otázky zabezpečení údajů. Až následná praxe nejspíš ukáže, jak bude staronový systém fungovat.

3.3. Vyhláška č. 98/2012 Sb., o zdravotnické dokumentaci

V § 1 se určuje, co má obsahovat zdravotnická dokumentace (identifikační údaje poskytovatele a pacienta, pacientovo pohlaví, data zápisu, razítka, pracovní závěry a konečnou diagnózu, rozsah poskytnutých služeb, aktuální vývoj zdravotního stavu pacienta, návrh léčebných postupů, podání léčivých přípravků, lékařské posudky, záznam o pracovním neschopnosti…)

§ 2 definuje součásti zdravotnické dokumentace, § 3 co musí být uvedeno na každém listu ZD a kdo je zodpovědný za zápis.

V dalších paragrafech pak nalezneme povinnosti k uchování ZD a nutnost elektronického podpisu v elektronické ZD.

§6 Ukládá, že technické prostředky pro vedení zdravotnické dokumentace v elektronické podobě zaručí:

- zabezpečení výpočetní techniky softwarovými a hardwarovými prostředky před přístupem neoprávněných osob ke zdravotnické dokumentaci
- vedení evidence všech přístupů ke zdravotnické dokumentaci včetně jejich oprav, změn a mazání.

Příloha č. 1 k vyhlášce č. 98/2012 Sb. pak určuje minimální obsah samostatných částí zdravotnické dokumentace, přílohy 2 a 3 určují zásady pro dobu a samotné uchování a následné zničení ZD.

3.4. Zákon č. 227/2000 Sb., o elektronickém podpisu

V zák. č. 227/2000 Sb., o elektronickém podpisu s novelizací 227/2009 Sb., 101/2010 Sb., lze nalézt důležité informace týkající se zabezpečeného přenosu dat se zaručeným podpisem.

V § 17 jsou definovány prostředky pro bezpečné vytváření a ověřování elektronických podpisů:

- Data pro vytváření podpisu se mohou vyskytnout pouze jednou a jejich utajení je náležitě zajištěno
- Data pro vytváření podpisu nelze při náležitém zajištění odvodit ze znalosti způsobu jejich vytváření a podpis je chráněn proti padělání s využitím existující dostupné technologie
- Data pro vytváření podpisu mohou být podepisující osobou spolehlivě chráněna proti zneužití třetí osobou
- Prostředky pro bezpečné vytváření podpisu nesmí měnit data, která se podepisují, ani zabraňovat tomu, aby tato data byla předložena podepisující osobě před vlastním procesem podepisování
- Prostředky pro bezpečné vytváření elektronických podpisů musí být před svým použitím bezpečným způsobem vydány a data pro vytváření elektronických podpisů musí být důvěryhodným způsobem v těchto prostředcích vytvořena nebo do nich přidána
- Prostředek pro bezpečné ověřování podpisu musí za pomoci odpovídajících technických a programových prostředků a postupů minimálně zajistit, aby:
    - Data používaná pro ověření podpisu odpovídala datům zobrazených osobě provádějící ověření
    - Podpis byl spolehlivě ověřen a výsledek tohoto ověření byl řádně zobrazen
    - Ověřující osoba mohla spolehlivě zjistit obsah podepsaných dat

- Pravost a platnost certifikátu při ověřování podpisu byly spolehlivě zjištěny

Výsledek ověření a totožnost podepisující osoby byly řádně zobrazeny

- Bylo jasně uvedeno použití pseudonymu
- Bylo možné zjistit veškeré změny ovlivňující bezpečnost

3.5. Zákon č. 101/2000 Sb., o ochraně osobních údajů

Důvodem vzniku zákona o ochraně osobních údajů bylo Listinou lidských práv a svobod zaručené právo na ochranu občana před neoprávněným zasahováním do jeho soukromého a osobního života neoprávněným shromažďováním, zveřejňováním nebo jiným zneužíváním osobních údajů.

Zákon se vztahuje na osobní údaje, které zpracovávají státní orgány, samospráva, fyzické a právnické osoby automatizovaně nebo jinými prostředky. Nevztahuje se na zpracování údajů fyzickou osobou nebo osobní potřebu a ve vymezených případech též na zpravodajské služby a policii.

V §13 jsou tyto důležité údaje týkající se ochrany osobních údajů:

- Správce a zpracovatel jsou povinni přijmout taková opatření, aby nemohlo dojít k neoprávněnému nebo nahodilému přístupu k osobním údajům, k jejich změně, zničení či ztrátě, neoprávněným přenosům, k jejich jinému neoprávněnému zpracování, jakož i k jinému zneužití osobních údajů. Tato povinnost platí i po ukončení zpracování osobních údajů
- Správce nebo zpracovatel je povinen zpracovat a dokumentovat přijatá a provedená technická a organizační opatření k zajištění ochrany osobních údajů v souladu se zákonem a jinými právními předpisy
- Správce nebo zpracovatel posuzuje rizika týkající se:
    - Plnění pokynů pro zpracování osobních údajů osobami, které mají bezprostřední přístup k osobním údajům

- Zabránění neoprávněným osobám přistupovat k osobním údajům a k prostředkům pro jejich zpracování
- Zabránění neoprávněnému čtení, vytváření, kopírování, přenosu, úpravě či vymazání záznamů obsahujících osobní údaje opatření, která umožní určit a ověřit, komu byly osobní údaje předány

- V oblasti automatizovaného zpracování osobních údajů je správce nebo zpracovatel v rámci opatření podle odstavce 1 povinen také:
  - Zajistit, aby systémy pro automatizovaná zpracování osobních údajů používaly pouze oprávněné osoby
  - Zajistit, aby fyzické osoby oprávněné k používání systémů pro automatizovaná zpracování osobních údajů měly přístup pouze k osobním údajům odpovídajícím oprávnění těchto osob, a to na základě zvláštních uživatelských oprávnění zřízených výlučně pro tyto osoby
  - Pořizovat elektronické záznamy, které umožní určit a ověřit, kdy, kým a z jakého důvodu byly osobní údaje zaznamenány nebo jinak zpracovány
  - Zabránit neoprávněnému přístupu k datovým nosičům

3.6. Vyhláška č. 62/2015 Sb. o provedení některých ustanovení zákona o zdravotnických prostředcích

Vyhláška č. 62/2015 Sb., o provedení některých ustanovení zákona o zdravotnických prostředcích.

Náležitosti dokumentace používaných zdravotnických prostředků

(K provedení § 59 odst. 4 zákona)

Dokumentace používaných zdravotnických prostředků, u kterých musí být provedena instruktáž, nebo u kterých musí být na základě pokynu výrobce provedena odborná údržba, či které jsou jiným právním předpisem označeny jako pracovní měřidla, obsahuje následující údaje:

- obchodní název zdravotnického prostředku
- doplněk názvu označující variantu zdravotnického prostředku, pokud existuje
- identifikaci zdravotnického prostředku uvedením čísla výrobní dávky, před kterou je uveden symbol „LOT“ nebo sériové číslo, pokud jsou výrobcem určeny
- katalogové číslo varianty zdravotnického prostředku přidělené výrobcem, pokud toto číslo existuje
- označení rizikové třídy nebo skutečnosti, že se jedná o aktivní implantabilní zdravotnický prostředek nebo diagnostický zdravotnický prostředek in vitro
- jméno nebo název výrobce a distributora
- umístění zdravotnického prostředku ve zdravotnickém zařízení poskytovatele zdravotních služeb, jedná-li se o pevně instalovaný zdravotnický prostředek
- datum uvedení do provozu a
- informace o provedených instruktážích, provedené odborné údržbě, provedených opravách a provedených revizích.

3.7. Zákon č. 181/2014 Sb., o kybernetické bezpečnosti

První část (§1-§33) se týká kybernetické bezpečnosti.

V základních ustanoveních je formulován předmět úpravy. Předmětem úpravy tohoto zákona jsou práva a povinnosti osob a působnost a pravomoci orgánů

veřejné moci v oblasti kybernetické bezpečnosti, nevztahuje se na informační nebo komunikační systémy , jenž nakládají s utajovanými informacemi.

§2 vysvětluje a definuje pojmy kybernetický prostor, kritická informační infrastruktura, významný informační systém, významná síť, správce informačního a komunikačního systému, bezpečnostní opatření a bezpečnost informací, §3 pak určuje osoby a orgány, kterým se ukládají povinnosti v oblasti kybernetické bezpečnosti.

Hlava II popisuje systém zajištění kybernetické bezpečnosti a ten zahrnuje bezpečnostní opatření, kybernetickou bezpečnostní událost a kybernetický bezpečnostní incident a jeho hlášení.

Dále pak evidence, opatření, varování, reaktivní a ochranné opatření a možné kontaktní údaje. Vymezuje úkoly národního i vládního CERT a jeho provozovatele. CERT (Computer Emergency Response Team) vznikl v roce 1988 na základě aféry s jedním z prvních počítačových červů, kterým byl tzv. Morrisův červ, jenž využil k svému šíření celosvětové sítě internetu. Od té doby CERT monitoruje všechny internetové průlomy, informuje o zranitelných místech v různých systémech a na základě toho zveřejňuje maximální množství bezpečnostních rad.

Hlava III se pak přímo zaobírá stavem kybernetického nebezpečí. Definuje jej jako stav, kdy je ve velkém rozsahu ohrožena bezpečnost informací v informačních systémech nebo bezpečnost a integrita služeb nebo sítí elektronických komunikací. Nalezneme zde i podmínky pro vyhlášení nouzového stavu.

Hlava IV určuje Úřad jako hlavního vykonavatele státní správy a hlav V řeší kontrolu, nápravná opatření a správní delikty. Zmocňovací, přechodná a společná ustanovení jsou pak tématem závěrečných ustanovení.

Druhá část (§34) obsahuje změnu zákona o ochraně utajovaných informací a o bezpečnostní způsobilosti.

Třetí část (§35) zahrnuje změnu zákona o elektronických komunikacích.

Čtvrtá část (§36) týká se změny zákona o svobodném přístupu k informacím.

Pátá část (§37) upravuje znění zákona o provozování rozhlasového a televizního vysílání.

Šestá část (§38) určuje účinnost zákona od 1. ledna 2015.

Všechny prováděcí předpisy k zákonu č. 181/2014 Sb., o kybernetické bezpečnosti, které platí stejně jako zákon od 1. 1. 2015, byly dne 19. 12. 2014 uveřejněny ve Sbírce zákonů v částce 127 pod tímto označením:

317/2014 Vyhláška o významných informačních systémech a jejich určujících kritériích

316/2014 Vyhláška o bezpečnostních opatřeních, kybernetických bezpečnostních incidentech, reaktivních opatřeních a o stanovení náležitostí podání v oblasti kybernetické bezpečnosti (vyhláška o kybernetické bezpečnosti)

315/2014 Nařízení vlády, kterým se mění nařízení vlády č. 432/2010 Sb., o kritériích pro určení prvku kritické infrastruktury

S prvními dvěma vás následně stručně seznámím.

3.8. Vyhláška č. 316/2014 Sb., o kybernetické bezpečnosti

Vyhláška o bezpečnostních opatřeních, kybernetických bezpečnostních incidentech, reaktivních opatřeních a o stanovení náležitostí podání v oblasti kybernetické bezpečnosti (vyhláška o kybernetické bezpečnosti).

Vyhláška obsahuje čtyři části.

V první z nich (v úvodních ustanoveních v §1) se objasňuje, co je předmětem této vyhlášky a to, že stanovuje obsah a strukturu bezpečnostní dokumentace pro informační systém kritické informační infrastruktury, její komunikační systém nebo významný informační systém. Dále pak obsah bezpečnostních opatření, rozsah jejich zavedení, typy a kategorie kybernetických bezpečnostních incidentů,

náležitosti a způsob hlášení kybernetického bezpečnostního incidentu, náležitosti oznámení o provedení reaktivního opatření a jeho výsledku a vzor oznámení kontaktních údajů a jeho formu.

§2 se věnuje vymezení pojmů: systém řízení bezpečnosti informací, aktivum, primární, podpůrné a technické aktivum, riziko a jeho hodnocení a řízení, hrozba, zranitelnost, přijatelné riziko, garant, uživatel a administrátor.

Druhá část řeší bezpečnostní opatření.

Má tři hlavy, v první z nich nalezneme organizační opatření, v jednotlivých paragrafech pak přímo systém řízení bezpečnosti informací, řízení rizik, bezpečnostní politiku, organizační bezpečnost, stanovení bezpečnostních požadavků pro dodavatele, řízení aktiv, bezpečnost lidských zdrojů, řízení provozu a komunikací, řízení přístupu a bezpečné chování uživatelů, akvizici, vývoj a údržbu, zvládání kybernetických bezpečnostních událostí a incidentů, řízení kontinuity činností, kontrolu a audit kritické informační infrastruktury a významných informačních systémů.

Předmětem hlavy druhé jsou technická opatření, jako je fyzická bezpečnost a její prostředky, nástroj pro ochranu integrity komunikačních sítí, nástroj pro ověřování identity uživatelů, nástroj pro řízení přístupových oprávnění, nástroj pro ochranu před škodlivým kódem, nástroj pro zaznamenávání činností kritické informační infrastruktury a významných informačních systémů, jejich uživatelů a administrátorů, nástroj pro detekci kybernetických bezpečnostních událostí, nástroj pro sběr a vyhodnocení kybernetických bezpečnostních událostí, aplikační bezpečnost, kryptografické prostředky, nástroj pro zajišťování úrovně dostupnosti a bezpečnost průmyslových a řídicích systémů.

V hlavě třetí se řeší bezpečnostní dokumentace a prokázání certifikace.

V třetí části najdeme problematiku kybernetického bezpečnostního incidentu, jeho typy, kategorie, formy a náležitosti hlášení.

3.9. Vyhláška č. 317/2014 Sb., o významných informačních systémech a jejich určujících kritériích

Vyhláška definuje předmět úpravy - stanoví se významné informační systémy a jejich určující kritéria podle § 6 písm. d) zákona. Taktéž se zde uvádí dopadová a oblastní určující kritéria a závěrečný paragraf určuje účinnost vyhlášky od 1. ledna 2015.

§ 1 Předmět úpravy

Touto vyhláškou se stanoví významné informační systémy a jejich určující kritéria podle § 6 písm. d) zákona.

§ 2 Významné informační systémy

Významný informační systém naplňující určující kritéria uvedená v § 3 je uveden v příloze č. 1 k této vyhlášce.

§ 3 Určující kritéria

- Určující kritéria významného informačního systému se člení na
    - dopadová určující kritéria
    - oblastní určující kritéria
- Významným informačním systémem není informační systém, jehož správcem je obec
    - a při výkonu působnosti obce hlavní město Praha.
- Naplnění určujících kritérií významného informačního systému, který není uveden v příloze č. 1 k této vyhlášce, posuzuje správce informačního systému.

§ 4 Dopadová určující kritéria

Dopadovým určujícím kritériem je skutečnost, že

- úplná nebo částečná nefunkčnost informačního systému způsobená narušením bezpečnosti informací by mohla mít negativní vliv na

- fungování orgánu veřejné moci,
- poskytování služeb nebo informací orgánem veřejné moci veřejnosti,
- hospodaření orgánu veřejné moci nebo hospodaření orgánu veřejné moci, který je správcem významného informačního systému, anebo hospodaření orgánu nebo osoby, která je správcem informačního nebo komunikačního systému kritické informační infrastruktury, nebo
- provoz jiného významného informačního systému využívajícího služeb hodnoceného informačního systému, který je nefunkční, přičemž omezení činnosti takového systému by mohlo mít za následek omezení výkonu působnosti orgánu veřejné moci po dobu delší než 3 pracovní dny, nebo výrazné ohrožení výkonu působnosti orgánu veřejné moci, které lze odvrátit za vynaložení nepřiměřených nákladů na provoz nebo obnovu informačního systému

- úplná nebo částečná nefunkčnost informačního systému způsobená narušením bezpečnosti informací by mohla způsobit
  - ohrožení nebo narušení prvku kritické infrastruktury2),
  - oběti na životech s mezní hodnotou více než 10 mrtvých nebo 100 zraněných osob vyžadujících lékařské ošetření, s případnou hospitalizací s dobou delší než 24 hodin,
  - finanční nebo materiální ztráty s mezní hodnotou více než 5 % stanoveného rozpočtu orgánu veřejné moci,
  - zásah do osobního života nebo do práv fyzických nebo právnických osob postihující nejméně 50000 osob, nebo
  - výrazné ohrožení nebo narušení veřejného zájmu,

přičemž následky podle bodů 1 až 4 nedosáhnou hodnot pro určení prvku kritické infrastruktury podle průřezových kritérií stanovených krizovým zákonem.

§ 5 obsahuje oblastní určující kritéria

- U orgánu veřejné moci
    - vedení správního řízení,
    - databáze obsahující osobní údaje,
    - hospodaření orgánu veřejné moci,
    - výkon spisové služby,
    - státní dozor,
    - kontrolní a inspekční činnost,
    - příprava na krizové situace a jejich řešení,
    - tvorba právních předpisů,
    - elektronická pošta,
    - vedení internetových stránek,
    - mezirezortní spolupráce,
    - mezinárodní spolupráce,
    - zadávání veřejných zakázek,
    - státní statistická služba.
- U orgánu veřejné moci - kraje v rámci přenesené působnosti
    - databáze obsahující osobní údaje,
    - vedení správního řízení,
    - hospodaření orgánu veřejné moci,
    - elektronická pošta,
    - vedení internetových stránek,
    - příprava na krizové situace a jejich řešení,
    - mezinárodní spolupráce,
    - státní dozor,
    - kontrolní a inspekční činnost,
    - zadávání veřejných zakázek.

4. DICOM

DICOM (Digital Imaging and Communications in Medicine) je mezinárodní standard pro komunikaci a správu obrazových medicínských dat a data s nimi spojené (ISO 12052). Definuje formáty pro obrazová medicínská data, aby mohly být posílány v kvalitě nezbytné pro lékařské účely.

DICOM můžeme najít v každém radiologickém, kardiologickém a i v zařízení pro radioterapii, mezi ně patří například – X-ray, CT, MRI, ultrazvuk atd. Zvyšuje se však využití i v dalších oblastech lékařství, například v očním a zubním lékařství.

S desítkami tisíc zobrazovacími zařízeními v provozu a s miliardy lékařských snímků se stává DICOM jedním z nejrozšířenějších zdravotních standardů po celém světě.

Další informace viz [4].

4.1. Historie vzniku DICOM

Když bylo poprvé představeno CT společně s dalšími digitálními diagnostickými zobrazovacími metodami a se zvyšovaným využíváním počítačů pro klinické aplikace, ARC (American College of Radiology) a NEMA (National Electrical Manufacturers Association), rozpoznali potřebu vytvoření standardu pro přenos snímků a s nimi souvisejícími informacemi, mezi zařízeními od různých výrobců.

ACR a NEMA tedy vytvořili společný výbor v roce 1983, aby vytvořili standard, který by:

- Podporoval komunikaci digitálních obrazových dat bez ohledu na výrobce zařízení
- Usnadnil rozvoj a rozšíření archivace obrazu a komunikačních systémů (PACS), které mohou také komunikovat s jinými systémy nemocničních informací

- Umožnil vytvoření diagnostických informačních databází, které mohou být čteny velkým rozsahem geograficky distribuovaných zařízeních

Od prvního zveřejnění v roce 1993, DICOM způsobil revoluci v praxi radiologie, kdy možné vyměnit X-ray filmy za plně digitální workflow.

Ať už u oddělení urgentního příjmu, u srdečního zátěžového testování nebo u detekce rakoviny prsu, DICOM je standard, který usnadňuje práci při komunikaci s medicínskými daty a usnadňuje tak práci pro lékaře a tedy i pro pacienty.

4.2. Základní části standardu DICOM

DICOM standard se původně skládal z 20 základních částí, avšak část PS3.9 a část PS3.13 byly postupem času odstraněny.

Více informací na [5].

4.2.1. PS 3.2 Shoda

V této části standardu jsou definovány principy, které musí zařízení nebo informační systém splňovat, aby dosáhl shody se standardem.

- Požadavky na shodu – část PS3.2 specifikuje obecné požadavky, které musí být v procesu implementace splněny. Konkrétní požadavky pro jednotlivé funkce, data i příkazy jsou pak uvedeny ve specifických částích standardu
- Prohlášení o shodě – část PS3.2 definuje strukturu dokumentu Prohlášení o shodě. Specifikuje informace, které musí být v dokumentu obsaženy, včetně vazeb na konkrétní požadavky uvedené v ostatních částech standardu

4.2.2. PS 3.3 Definice informačních objektů

V této části standardu jsou specifikovány třídy informačních objektů (Information Object Classes), které umožňují realizovat abstraktní definici entit reálného světa aplikovatelnou při komunikaci a přenosu medicínských obrazů a informace s nimi spojené (křivky, strukturalizované nálezy, dávky radiační terapie, atd.). Každá definice třídy informačních objektů je tvořena popisem jejího určení a atributů, pomocí kterých je definice realizována.

Standard rozlišuje dva typy tříd informačních objektů:

- Normalizované třídy informačních objektů – obsahují pouze atributy, které jsou vlastní reprezentované entitě reálného světa
- Kompozitní třídy informačních objektů – mohou i obsahovat atributy, související s entitou reálného světa, které nejsou vlastní (cizorodé)

Kompozitní třídy informačních objektů udávají strukturalizovaný rámec pro realizaci komunikačních požadavků pro zajištění úzké vazby mezi obrazovou informací a informacemi s nimi souvislými.

4.2.3. PS 3.4 Specifikace servisních tříd

Tato část definuje řadu servisních tříd. Servisní třída spojuje jeden nebo víc informačních objektů s jedním nebo více příkazy, které nad těmito informačními objekty mají být vykonány.

Mezi servisní třídy například patří:

- Management tisku
- Uložení informací
- Dotaz/opověď
- Základní management worklistu
- Management pacienta

- Management výsledků

4.2.4. PS 3.5 Datové struktury a kódování

V této části se specifikuje vytváření a kódování datových souborů (Data set) DICOM aplikací, které vycházejí z užití informačních objektů a servisních tříd. V části se také specifikuje, jaké jsou použité kompresní techniky.

4.2.5. PS 3.6 Datový slovník

V této části se specifikuje centrální registr DICOM datových elementů a jejich definic. Datové elementy představují základní entitu reprezentované informace včetně jejich unikátní identifikace v rámci standardu DICOM.

Každý datový element je specifikován:

- Jednoznačným tagem, tvořeným z čísla skupiny a z čísla elementu
- Jménem
- Hodnota multiplicity (číslo, udávající kolik hodnot může být zakódováno do datového elementu)
- Typem hodnoty (integer číslo, řetězec znaků, atd.)

Každý unikátní identifikátor je specifikován:

- Složen ze dvou částí, které jsou odděleny desetinou tečkou
    - <org root> - unikátní číselná hodnota pro organizaci
    - <suffix> - unikátní číselná hodnota v rámci organizace

Příklad:

UID = <org root>.<suffix>

4.2.6. PS 3.7 Výměna zpráv

Tato část specifikuje služby a protokoly používané aplikacemi medicínských zobrazovacích metod při výměně zpráv v rámci DICOM komunikace. Tyto zprávy jsou složeny z posloupnosti příkazů a z navazujícího datového streamu.

PS 3.7 dále udává:

- Operace a informace o stavu (nebo případné změně stavu) entity (DIMSE služby – DICOM Message Service Element), které jsou k dispozici jednotlivým třídám služeb definovaných v části PS 3.4
- Pravidla pro ovládání příkazů realizujících komunikaci na principu požadavek/odezva
- Pravidla pro navázání a ukončení spojení zajišťovaného komunikačními službami
- Kódovací pravidla nezbytná pro tvorbu posloupností příkazů a zpráv

4.2.7. PS 3.8 Podpora síťové komunikace pro výměnu zpráv

V této části se specifikují komunikační služby a protokoly nejvyšší komunikační vrstvy nezbytné pro komunikaci mezi DICOM aplikacemi, které zajišťují, aby komunikace byla prováděna efektivně a koordinovaně v daném síťovém prostředí. Uvedená specifikace služeb vrchní komunikační vrstvy (Upper Layer Service) je podmnožinou služeb zajišťovaných sedmivrstvovým komunikačním modelem ISO/OSI. Její definice specifikuje použití protokolu DICOM horní vrstvy ve spojení s TCP/IP transportním protokolem.

4.2.8. PS 3.10 Paměťová média a formát souboru pro výměnu médií

Tato část specifikuje obecný model ukládání medicínských obrazových dat na výměnných médiích. Hlavním účelem této části je poskytnout rámec umožňující vzájemnou výměnu různých typů medicínských obrazových dat i s nimi souvisejícími informacemi na různé typy paměťových médií.

4.2.9. PS 3.11 Aplikační profily paměťových médií

V této části se specifikuje aplikační podmnožina DICOM standardu, pro kterou implementace může dosáhnout shody. Takovéto prohlášení shody je aplikováno na funkčnost procesu výměny medicínských obrazových dat a s nimi souvisejícími informacemi na paměťových médiích pro specifické klinické využití.

4.2.10. PS 3.12 Formáty medií a fyzická média pro výměnu medií

Tato část podporuje a usnadňuje výměnu informací mezi medicínskými aplikacemi a specifikuje:

- Charakteristiku specifických fyzických medií a jejich formátů
- Strukturu pro popis vzájemných vztahů mezi obecným modelem paměťových medií a specifickými fyzickými médii a jejich formátem

4.2.11. PS 3.14 Zobrazovací funkce standardní stupnice šedi

V této části se specifikují standardizované zobrazovací funkce, které jsou nezbytné pro konzistentní zobrazování obrazových dat založených na stupnici šedi. Zobrazovací funkce poskytují metody kalibrace konkrétních zobrazovacích systémů, umožňující zajistit konzistentní prezentaci obrazových dat na různých mediích (displeje, tiskárny, atd.). Zobrazovací funkce jsou založeny na lidském vizuálním vnímání (Bartenův model).

4.2.12. PS 3.15 Bezpečnostní a systémové profily managementu

V této části se specifikuje bezpečnost systémů DICOM standardu a pravidla řízení přístupu k datům, která musí být dodržena pro dosažení shody aplikace se standardem. Tu obstarávají obecně uznávané protokoly, jako jsou například DHCP, LDAP, TSL a další.

4.2.13. PS 3.16 Mapování obsahových zdrojů

Tato část DICOM standardu specifikuje, jaké návrhy formátů strukturovaných dokumentů DICOM informačních objektů lze používat. Dále taky uvádí množinu kódovaných termínů, které jsou využívány informačními objekty a také překlady kódovaných termínů specifických pro jednotlivé země.

4.2.14. PS 3.17 Vysvětlivky

Část PS 3.14 standardu DICOM obsahuje rozsáhlé dodatečné vysvětlivky k předešlým částím. Ostatní části se na ní taktéž odkazují.

4.2.15 PS 3.18 Webový přístup k DICOM objektům (WADO)

V této části je specifikováno, jaké prostředky umožnují realizaci požadavku na povolené DICOM objekty ve formátu http URL/URI (Uniform Resource Locator/ Uniform Resource Identifier). Požadavek musí obsahovat směrník, který odkazuje na příslušný známý a definovaný DICOM objekt ve formě konkrétního UID.

4.3. Historie DICOM v České republice

Zavedení DICOM v České republice se dělí do tří základních etap, které postupně popíši.

Detailnější informace k nalezení na [6].

4.3.1. Zavedení DICOM v ČR – 1. etapa

1. etapa (2007-2010) měla za cíl ověřit možnosti změny při předávání obrazové dokumentace ze stávající „neelektronické“(neefektivní převážení filmů a CD s obrazovou pacientskou dokumentací mezi zdravotnickými zařízeními většinou sanitkami či kurýrem, ať už za účelem odborné konzultace, „druhého čtení“ či převážení pacienta z jednoho zdravotnického zařízení do druhého) na „elektronickou “. Podmínkou bylo právě využití právě DICOM , jakožto celosvětově uznávaného standardu pro tuto oblast. Dále pak se projekt opíral o zkušenosti MeDiMed (brněnské řešení problematiky) a kladl nemalý důraz na bezpečnost a univerzálnost řešení.

Ministerstvo zdravotnictví vyzvalo Všeobecnou fakultní nemocnici v Praze k řízení tohoto projektu a dohodlo účast tří nemocnic v 1. etapě tohoto projektu: Všeobecné fakultní nemocnice v Praze, Fakultní nemocnice Na Bulovce a Ústřední vojenské nemocnice Praha.

Řešení mělo zajišťovat komunikaci s ostatními zdravotnickými zařízeními na úrovni protokolu DICOM verze 3 bez dalších konverzí a převodu dat, při maximální míře zabezpečení a být univerzální a jednoduché pro koncového uživatele.

Charakteristické pro projekt bylo propojení nemocnic pomocí zabezpečených VPN, odeslání vybraných obrazových dat jen z vůle odesílatele (pověřená osoba v konkrétním zdravotnickém zařízení), neumožnění přístupu z jedné nemocniční sítě do druhé, správa přístupových práv pro odesílání a příjem obrazových dat je zcela v pověření konkrétního zdravotnického zařízení, komunikačním protokolem je DICOM a koordinátorem projektu Všeobecná fakultní nemocnice v Praze, kde byl umístěn i centrální komunikační uzel DICOM a jeho správa.

Popis projektu:

Navržené řešení se skládalo z centrálního komunikačního uzlu ve Všeobecné fakultní nemocnici a komunikačního uzlu v jednotlivých zdravotních zařízení. Vlastní komunikace mezi komunikačními uzly a centrálním komunikačním uzlem probíhala v protokolu DICOM přenášená pomocí VPN tunely. Tím byla i zaručena maximální míra bezpečnosti.

Na zabezpečeném serveru byl nahrán software centrálního uzlu, který obstarával následující události:

- propojení VPN tunelem s příslušnými komunikačními uzly
- centrální správa adres odesílatelů a příjemců
- přesměrovávání DICOM packetů dle nastavených konfigurací (od odesílatele k příjemci)
- technický monitoring
- kontrolní funkce.

Samotný komunikační kanál byl vybaven softwarem, který byl nahrán na vyhrazeném zabezpečeném serveru a obstarával následující událost:

- propojení VPN tunelem s centrálním komunikačním uzlem
- přidělování příslušných adres dle konfigurace
- odesílání a přijímání DICOM packetů

Volitelně bylo možné vybavit komunikační kanál „miniPACS“ archivem pro ukládání obdržených obrazových studií, příp. automatickým DICOM přeposíláním došlých obrazových studií k příslušné DICOM entitě.

Vlastní komunikace pak probíhala následovně: Uživatel/odesílatel (oprávněný radiolog, klinik) na svém DICOM prohlížeči zvolil danou studii a „přeposlání“ do zvolené nemocnice. Tato studie byla nasměrována do místního komunikačního uzlu, který ji ve VPN tunelu poslal do centrálního komunikačního uzlu, ten pak provedl základní kontrolu a přesměroval studii k zadanému adresátovi. Uzel adresáta přijmul studii a obvykle ji uložil do „miniPACSu“, kde byla přes protokol DICOM Dotaz/Odpověď dostupná oprávněným uživatelům nebo adresátům.

Díky projektu byla vytvořena infrastruktura pro vzdálené konzultace mezi odborníky na úrovni zapojených nemocnic v Praze, dále pro výměnu obrazové dokumentace při přechodu pacienta mezi zainteresovanými nemocnicemi, zrychlení a zjednodušení přístupu k obrazovým datům pro pacienty, kteří jsou odesláni na vyšetření mimo nemocnice a to vše vedlo k lepší spolupráci radiologických a klinických odborníků.

1. etapa projektu byla úspěšně realizována, byly splněny veškeré cíle.

4.3.2. Zavedení DICOM v ČR – 2. etapa

Následovala tudíž druhá etapa a to v létech 2010 až 2012, jejíž cílem bylo připojení dalších zdravotnických zařízení k projektu a tím maximalizovat efektivnost komunikace mezi nimi.

Ministerstvo zdravotnictví vyzvalo Všeobecnou fakultní nemocnici v Praze ke koordinaci tohoto projektu. Ta vedla evidenci zdravotnických zařízení, která byla připojena do projektu DICOM komunikace.

Navržené řešení v první fázi vycházelo z požadavku nízko-nákladového projektu realizovaného v první etapě. To znamenalo připojení „komunikačních uzlů“ v jednotlivých zdravotnických zařízení k již vybudovanému Centrálního uzlu –

DICOM routeru Všeobecné fakultní nemocnice. Tím byla vytvořena možnost vzájemné komunikace mezi zdravotními zařízeními. To dalo čas pro vznik dalších DICOM routerů v zdravotnických zařízení, které by sloužily pro komunikaci s dalšími zdravotními zařízeními v rámci regionu a zmenšily by tak případnou komunikační zátěž při větším využití této komunikace. Ostatní parametry jsou shodné s 1. etapou (bezpečnost, způsob komunikace a administrace). Projekt byl navrhován tak, aby komunikační uzel zdravotnického zařízení nebyl závislý na typu HW ani SW.

Komunikační uzel musel pouze splňovat:

- Podpora pro VPN na bázi SSL/TLS protokolu s využitím asymetrických klíčů a certifikátů dle X509
- Komunikace mezi komunikačním uzlem a centrálním směrovačem prostřednictvím standardu DICOM verze 3., specifikace min. 2003
- Asociace včetně rozšířeného vyjednávání
- DIMSE služby C-Echo a C-Store
- Podpora min. ImageStorage SOP Class a StructuredReport SOP Class
- Podpora min. Default Transfer Syntax a JPEG Transfer Syntax
- Podpora zabezpečení digitálním podpisem dle specifikace DICOM 3.15 – Security Profiles

Evidenci připojených zdravotních zařízení do projektu vedla Všeobecná fakultní nemocnice. K připojení do projektu byla použita přístupová mezi zařízeními a Fakultními nemocnicemi. Tato smlouva umožnila zdravotnickému zařízení komunikovat elektronickou formou s ostatními zařízeními dle tohoto projektu, nejsou to však smluvní vztahy mezi jednotlivými zařízeními, které si obrazovou dokumentaci vyměňují (např. o poskytování konsiliárním vyšetření apod.). MZČR připomnělo, že výměna obrazové dokumentace musí probíhat v intensi zákona č. 20/1966 Sb., o péči o zdraví lidu, ve znění pozdějších předpisů. Pro využití DICOM komunikace nemuselo mít zdravotní zařízení plně funkční PACS, ale pro odesílání obrazové informace muselo provozovat modality komunikující prostřednictvím protokolu Dicom3.

Zdravotnická zařízení, která se chtěla připojit k projektu, se musela formálně přihlásit přístupovou smlouvou a mít připojení k internetu bez důležitosti rychlosti připojení.

Pro odesílání dat bylo třeba:

- ukládat digitální obrazová data ve formátu DICOM (PACS nebo lokální úložiště)
- stanovit přístupová práva (kdo smí odesílat data)
- pořídit a instalovat komunikační uzel
- nakonfigurovat DICOM entity ve ZZ
- dohodnout se na zprovoznění s VFN

Pro příjemce dat bylo třeba:

- mít alespoň jeden prohlížeč (třeba freeware) či jednu DICOM entitu pro ukládání došlých dat
- stanovit přístupová práva (kdo smí přijímat data)
- pořídit a instalovat komunikační uzel
- nakonfigurovat DICOM entity ve zdravotním zařízení a dohodnout se na zprovoznění s Všeobecnou fakultní nemocnicí

2. etapa projektu měla za cíl vytvořit standard pro komunikaci zdravotnických zařízení v oblasti obrazové komunikace. Její trvání bylo časově neohraničené, jelikož předpokladem bylo postupné zapojení všech zdravotnických pracovišť, které mají potřebu elektronické komunikace v oblasti obrazové informace. Tento projekt byl otevřen všem zdravotním zařízením bez rozdílu právní formy a vlastnictví.

MZČR doporučilo připojení dalších zdravotních zařízení k tomuto projektu.

2. fáze projektu plně naplnila cíle dané etapy.

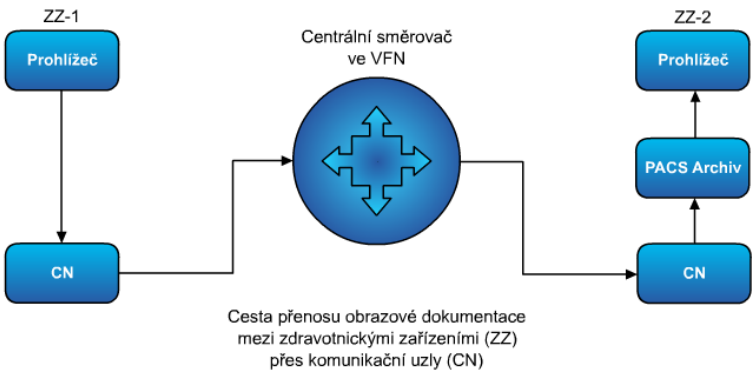
4.3.3. Zavedení DICOM v ČR – 3. etapa

Cílem 3. etapy (začátek v r. 2012) je navázat na 2. etapu, ještě více standardizovat řešení, aby připojení zdravotnických zařízení bylo pro koncového uživatele velmi jednoduché a nenarušovalo jeho vlastní vnitřní infrastrukturu.

3. etapa navazuje na 2. etapu projektu. V rámci 2. etapy byly na základě příspěvků uživatelů, subdodavatelů a zadavatele realizovány mnohé změny a vytvořeny různé standardy vzájemné komunikace, síťové infrastruktury a SW řešení. V současné době je pro připojení k projektu použitelný pouze komunikační uzel AMIS*PACS CommunicationNode nebo schránka určena pouze pro privátní lékaře. O distribuci se starají určité firmy.

Stávající etapa má za cíl ještě více rozšiřovat nabízené služby a zpřístupnit projekt ještě více uživatelům.

Ukázka komunikace mezi zdravotními zařízeními přes komunikační uzly viz obrázek č. 4.1.



Obr. č. 4.1: Komunikace zdravotnických zařízení přes komunikační uzly

5. Analýza anonymizace obrazových dat

Modernizace způsobu nakládání se zdravotnickými obrazovými daty a neustálé vylepšování dostupné technologie pro jejich pořizování, zapříčinily nutnost dalšího zpracování medicínských dat. V této práci je řešen důležitý úkol odstranění zbytkových citlivých dat z obrazových souborů. Po odstranění citlivých dat jsou obrazové části DICOM souborů použitelné i mimo zdravotnická zařízení, např. pro účely výuky nebo statistiky.

V další části mojí bakalářské práce uvedu postupně metody a knihovny, které jsem v anonymizaci docílil.

5.1. Metody hledání textových řetězců

Abych mohl citlivá data vymazat, potřebuji je nejprve v souboru najít. Jelikož však předem nevím přesné řetězce, hledám pouze řetězce splňující určitá kritéria. Jaké metody vyhledávání jsem použil, vysvětlím v další části.

5.1.1. Hammingova vzdálenost

Značí, na kolika pozicích se dva řetězce od sebe navzájem liší, tedy množství záměn, které je potřeba provést, aby byly porovnávané řetězce stejné.

V bakalářské práci je Hammingova vzdálenost použita pro případné ošetření špatného výstupu dat z programu Tesseract, které není 100% při čtení znaků. Program dokáže detekovat až 2 chyby načtení obrazových dat. Například číslice v rodném čísle mohou být přečteny i jako jiné znaky – **1** jako **I**, **7** jako **?** atd.

Hammingova vzdálenost je pro můj program zásadní, jelikož bez její použití by se procentuální úspěšnost anonymizace znaků značně snížila.

5.1.3. Algoritmus hrubé síly

Algoritmus postupuje tak, že pro každou pozici v textu kontroluje, jestli v ní nezačíná hledaný řetězec. Složitost hledání v běžném textu je O(m+n).

Jelikož nevíme, jaké řetězce hledáme, musí existovat určitá pravidla podle kterých se při hledání řetězců řídíme. Algoritmus hrubé síly byl použit při kontrole řetězců, které anonymizovat nepotřebují, avšak splňují podmínky pro anonymizaci (např. všechny znaky jsou velké). Tedy vytvořil jsem sérii řetězců, které při shodě s nalezeným textem (který splnil podmínky pro anonymizaci) nejsou považovány za citlivé informace nutné k anonymizaci.

5.1.5. Regulární výrazy

Regulární výrazy jsou speciální řetězce umožňující hledání celé množiny řetězců v textu odpovídajícím vzorovému výrazu.

Nejčastěji se vyskytuje při:

- vyhledávání v textu, pro určení pozice vzoru v textu, nebo zda vzor odpovídá regulárnímu výrazu
- nalezení a případnou změnu řetězce nebo získání všech shod do proměnné, s kterou pak lze provádět další operace

V mém programu jsem využil regulární výrazy pro hledání speciálních znaků, které oznamují, že bylo splněno pravidlo programu a za těmito speciálními znaky se nachází námi vyžadovaný text pro anonymizaci. Příklad takového použití je třeba výskyt znaku ***** v obrazových datech, který oznamuje, že bude následovat datum narození.

5.2 Použité nástroje

Abych docílil splnění zadání své bakalářské práce, potřeboval jsem využít různé škály open source nástrojů. V následujících kapitolách vysvětlím, jaké open source nástroje jsem použil a také uvedu stručně něco o nich.

Na úpravu obrazových souborů jsem použil **ImageMagick**, distribuovaný pro nejrůznější operační systémy. Po úpravě obrazového souboru jsem použil open source engine **Tesseract-ocr** pro naskenování dat z obrazového souboru a jejich následné uložení do příslušného formátu.

Jakožto vhodný formát pro výstupní data z **Tesseract-ocr** jsem zvolil formát **HOCR**, který je odnoží hypertextových formátů (**HTML**), konkrétněji **XHTML**. Důvodem zvolení **HOCR** formátu bylo, že tento výstupní formát uchovává i pozice nalezeného textu, které jsou nezbytné pro jeho anonymizaci. Textový výstup nebo výstup **PDF** informace o poloze textu neuvádějí.

Pro parsování formátu **HOCR** jsem v mém programu použil Java knihovnu zvanou **JSoup**, díky které jsem byl schopen načtená data uložit do pole stringů, se kterým se mi bude lépe pracovat.

Jako vývojové prostředí pro můj program jsem použil **NetBeans** pro Linux.

5.2.1. Tesseract-Ocr

Tesseract je pravděpodobně jedním z nepřesnějších open source OCR (Optical Character Recognition). V kombinací s Leptonica Image Processing Library je Tesseract schopen přečíst rozsáhlé množství obrazových formátů a dokáže je i převézt do více jak 60 různých jazyků. Tesseract byl považován za jeden ze tří nejlepších engine v UNLV testu přesnosti v roce 1995. Od roku 1995 se vývoj Tesseractu zpomalil až do roku 2006, kdy začal být vývoj Tesseractu sponzorovaný společností Google. Nyní je vydán pod Apache Licencí 2.0.

Tesseract podporuje operační systémy:

- Windows
- Linux
- Mac OS X

Důvodů, proč jsem si vybral Tesseract pro skenování obrazového souboru je a upřednostnil ho tak před ostatními Open Source OCR enginy je hned několik:

- Tesseract je celosvětově považován za jeden z nejlepších OCR enginů
- Výsledky rozpoznávání znaků byly lepší než u ostatních mnou zkoušených enginů (např. GOCR)
- Tesseract lze trénovat pro lepší výsledky rozpoznávání znaků
- Možnost výstupu dat ve formátu XHTML

Výstupním formátem Tesseract může být textový soubor, PDF soubor nebo soubor HOCR. Pro mojí bakalářskou práci jsem použil výstupní formát typu HOCR, který mi umožnil získat pozice mnou hledaných slov (stringů).

Přesnost rozlišování znaků v mém případě se pohybovala okolo 80%.

Více viz [7].

5.2.2. ImageMagick

ImageMagick je balík nástrojů na vytváření, úpravu a zpracování bitmapových obrázků. Dokáže číst a přepisovat rozsáhlou škálu formátů souborů – oficiální stránky udávají, že přes 200 typů souborů – mezi nimiž jsou i formáty použité při zpracování mé bakalářské práce, tedy původní formát PNG a mnou převedený formát JPG.

Důvodem použití tohoto nástroje byla nutná úprava obrazových souborů, kde schopnost enginu Tesseract rozpoznat znaky se pohybovala pod hranicí 50%. To jsem považoval za nepřípustné. Viz více na [8].

5.2.3. JSoup

Při programování praktické části bakalářské práce jsem použil Java knihovnu JSoup, která slouží k parsování dat z hypertextových typů souborů, v mém případě tedy XHTML – hocr. Více informací dostupné na [9].

5.2.4. AWT Graphics

K načtení obrazového souboru do programu a k následné anonymizaci citlivých dat jsem se rozhodl použít základní knihovny AWT Graphics, která je standardně obsažena v Javě.

6. Implementace

Práci na anonymizaci jsem rozdělil do dvou částí:

- Úprava obrazových souborů a získání dat pro anonymizaci obrazových souborů
- Anonymizace obrazových souborů

Program obstarává sám obě dvě části.

Vývoj programu jsem implementoval v operačním systému Linux.

Program spouštím pomocí příkazu:

java –jar program.jar properties.properties

kde první parametr označuje properties soubor, ve kterém jsou uloženy všechny potřebné proměnné, jako je například cesta nebo název původního obrazového souboru pro anonymizaci a také cesta k výstupnímu adresáři, kam bude uložen logovací soubor a anonymizovaný obrazový soubor. Do výstupní složky budou také

uloženy soubory potřebné pro anonymizaci, ty však budou před ukončením programu smazány.

6.1. Výběr programovacího jazyka

Pro zpracování požadavků bakalářské práce jsem se rozhodl použít programovací jazyk Java, protože je jedním z nejrozšířenějších programovacích jazyků a také proto, že mám nejvíc zkušenosti s programovacím jazykem Java z průběhu mého studia.

6.2. Úprava obrazových souborů

V mojí bakalářské práci byl ImageMagick použit pro změnu rozlišení původního obrazového souboru. Důvodem téhle změny byla příprava souboru pro práci s programem Tesseract. Zjistil jsem totiž, že na obrazových souborech, které mi byly pro práci přiděleny, není úspěšnost rozeznávání písmen a číslic zdaleka uspokojivá, tudíž před použitím programu Tesseract změním rozlišení obrazového souboru viz tabulka č. 6.1. K dosažení mého cíle jsem tedy použil **–resize** z knihovny ImageMagick.

S příkazem –resize jsem zkoušel zadat několik hodnot a porovnával jsem, z jaké hodnoty dostanu nejlepší výsledek – tedy poměr velikosti souboru a schopnosti Tesseractu rozeznat vyžadované znaky.

Tab. č. 6.1: Porovnání velikosti a úspěšnosti rozeznávání textu pro formát PNG

Rozlišení v pixelech	Velikost	Ukázka textu
Originální (980x980)	500 kB	Fakunm nemacmce men
5000x5000	5 000 kB	Fakuitni nemocnice Plzen
10000x10000	14 000 kB	Fakuitni nemoonioe Plzen

Z tabulky výše lze krásně vidět velký rozdíl mezi originálním obrazovým souborem a mnou upraveným souborem.

Rozdíl mezi rozlišením 5000 a 10000 není už zase tak razantní, avšak lepší hodnoty jsem dostával z rozlišení 5000. Při rozhodování také záleželo na velikosti souboru a jeho zpracování. Rozhodl jsem se tedy každý upravovaný obrazový soubor nejdříve změnit na 5000x5000 pixelů.

Formát PNG však v takto velkém rozlišení nabývá velké velikosti, viz tabulka výše, která zpomalovala chod programu až na několik vteřin. Rozhodl jsem se tedy při změně rozlišení také změnit formát souboru, v mém případě jsem zvolil formát JPG. Zkoumal jsem dále, jestli změna formátu neovlivní i změnu úspěšnosti načítání znaků, viz tabulka č. 6.2.

Tab. č. 6.2: Porovnání velikosti a úspěšnosti rozeznávání textu pro formát PNG

Rozlišení v pixelech	Velikost	Ukázka textu
Originální (980x980)	500 kB	Fakunm nemacmce men
5000x5000	2 000 kB	Fakultni nemocnice Plzen
10000x10000	5 000 kB	Fakuttni nemocnioe Plzen

Tabulka uvedená výše tedy potvrzuje, že moje obavy ohledně snížení rozpoznatelnosti znaků se nepotvrdily, ba naopak, v určitých pasážích textu se rozpoznávání znaků dokonce i zlepšilo.

Pro změnu formátu a rozlišení obrazového souboru tedy použiji následující příkaz

convert obraz.png –resize 5000 pred.jpg

Obrazové soubory už mám tedy připravené pro další postup v mojí práci. Jako další krok jsem nechal program Tesseract naskenovat obrazový soubor a uložit jej do mnou zvoleným výstupním formátem. Defaultně program Tesseract používá jako výstupní formát typ TXT, ten je však pro mojí bakalářskou práci nedostačující, jelikož neobsahuje pozice nalezeného textu.

Zvolil jsem tedy výstup ve formátu hOCR. Jedná se o soubor typu XHTML.

hOCR soubor jsem získal pomocí příkazu:

tesseract pred.jpg data hocr

kde hodnota „data“ značí výstupní soubor formátu hOCR, který obsahuje veškerý nalezený text v upraveném obrazovém souboru a také pozice nalezeného textu.

Příkaz **convert** i příkaz **tesseract** si program volá automaticky sám při spuštění.

Výstupem jsou tedy dva soubory:

- pred.jpg
- data.hocr

Ukládány jsou do předem zvoleného adresáře ze souboru properties a po skončení programu jsou automaticky smazány.

6.3. Nalezení citlivých dat v obrazových souborech

Po získání hOCR soubor a upraveného obrazového souboru jsem připraven na anonymizaci citlivých dat v původním obrazovém souboru.

Nejprve hOCR soubor však musím zpracovat. Abych toho dosáhl, použil jsem následující metody z knihovy JSoup.

org.jsoup.nodes.Document doc = Jsoup.parse(input, "UTF-8");

Díky této metodě jsem si uložil celý obsah HOCR souboru do dokumentu **doc**, ze kterého pak budu načítat mnou požadovaná data.

Načítání dat probíhá v cyklu **for**. Pro uložení získaných řetězců textu jsem vytvořil dvě pole. Do prvního pole budu ukládat nalezené řetězce textu a do druhého pole budu ukládat pozice nalezeného textu.

Parsování pak vypadá takto:

```
for (Element ocrxWord : doc.select(".ocrx_word")) {

    jmena[i] = ocrxWord.text();  //JMENO, PRIJMENI, CISLA

    pozice[i] = ocrxWord.attr("title");  //bbox 250 192 1606
    375; x_wconf 70
```

Cyklus hledá pouze text v souboru hOCR, tedy hodnoty „**.ocrx_word**“.

Do pole **jmena[]** se pomocí příkazu **ocrxWord.text();** uloží všechny stringy z dokumentu, které byly naskenovány pomocí Tesseractu.

ocrxWord.attr(“title“); uloží do pole **pozice[]** všechny hodnoty atributu **title**. Mezi nimiž jsou i námi vyžadované pozice načtených stringů.

Každý string je po získání kontrolován, zda neodpovídá předem daným předpisům pro anonymizaci dat, tedy jestli získaný string není datum narození, rodné číslo, jméno nebo příjmení.

Pro zjišťování citlivých dat jsem vytvořil tři metody:

- **zjistiJmeno(i);**
- **zjistiDatum(i);**
- **zjistiCislo(i)**;

Index **i** odkazuje na hodnotu momentálně testovaného řetězce. Všechny metody jsou **void**, tudíž nemají žádnou návratovou hodnotu.

Metoda **zjistiJmeno(i);**

Tato metoda, jak již název napovídá, je určena pro nalezení jména nebo příjmení v získaném stringovém řetězci.

Při prohlížení obdržených materiálů k vývoji mého programu jsem zjistil, že takřka vždy je jméno nebo příjmení napsáno velkými písmeny. Proto jsem se rozhodl pro jméno a příjmení kontrolovat řetězce, ve kterých jsou pouze velká písmena.

Pro procházení řetězce jsem použil cyklus **for**, který kontroluje každý char v řetězci funkcí:

Character.isUpperCase(jmena[i].charAt(k));

Hodnota **k** udává pozici charu v testovaném řetězci.

Návratová hodnota funkce je typu boolean, tedy true nebo false. Pokud je nalezeno číslo, nebo jiný znak, cyklus **for** se „breakne“ a přechází se na další nalezený řetězec.

V této metodě jsem také musel ošetřit případy, kdy se z Tesseractu načetl řetězec, který obsahoval na konci jména znak tečky, nebo čárky viz tabulka č. 6.3.

Tab. č. 6.3: Příklad ukládání řetězců do pole stringů

Původní text	**První řetězec**	**Druhý řetězec**
PŘIJMENÍ, JMÉNO	PŘIJMENÍ,	JMÉNO

Metoda tedy dokáže akceptovat tečku nebo čárku, vyskytuje-li se na konci řetězce a zároveň, pokud je řetězec delší než tři znaky. Je tomu tak z důvodu, aby se nebral řetězec složen ze dvou čárek jako jméno.

Pokud splňuje nalezený řetězec všechny pravidla, je program rozhodne, že je nutné tento řetězec anonymizovat. Zavolá se tedy příslušná metoda

anonymizujJmeno(i), která se o anymizaci postará. Jak tato metoda funguje, a jak probíhá anonymizace, vysvětlím v další části.

Metoda **zjistiDatum(i);**

Další věcí, kterou jsem si při prohlížení dodaných materiálů bylo, že datum narození vždy začíná znakem hvězdičky „*****“. Metoda proto při nalezení tohoto znaku předpokládá, že dále bude následovat datum narození. Opět musí platit pravidlo, že řetězec musí být delší než tři znaky.

Po splnění kritérií pro anonymizaci se zavolá metoda **anonymizujDatum(i)**. Opět fungování metody a anonymizaci vysvětlím v dalších části.

Metoda **zjistiCislo(i);**

Metoda určená pro zjištění rodného čísla. Funguje na principu metody **zjistiJmeno(i)** s tím rozdílem, že místo kontroly, zda je char velké písmeno (UpperCase), kontroluji shodu znaku s číslem, viz funkce:

Character.isDigit(jmena[i].charAt(k));

Hodnota **k** udává pozici charu v testovaném řetězci.

V metodě jsem také musel ošetřit případy, kdy načtená hodnota z Tesseractu nemusela odpovídat originální hodnotě, viz tabulka č. 6.4.

Tab. č. 6.4: Příklad nepřesně načteného řetězce z Tesseractu

Původní text	**Načtený text**
12/23/4567	1272374567
123456789	123456?89

Metoda tedy dokáže akceptovat i s takto nepovedeně načtené řetězce. Dokáže stejně jako metoda **zjistiJmeno(i)** ošetřit tečku nebo čárku na konci řetězce.

Musí být také zavedeno pravidlo, že řetězec musí být delší než 7 znaků. Zmíněné pravidlo je zapotřebí, nastane-li podobný případ jako ve výše uvedené tabulce, tedy kdy Tesseract špatně přečte znak lomítka „**/**“ a pokládá ho za číslici **7**. Z řetězce o 8 číslicích se pak rázem stane řetězec obsahující 10 číslic.

Po nalezení citlivých dat se zavolá funkce **anonymizujCislo(i)**. Zavolanou funkci vysvětlím v další části.

6.4 Anonymizace nalezených citlivých dat

Pro anonymizaci nalezených citlivých dat jsem vytvořil tři metody:

- **anonymizujJmeno(i);**
- **anonymizujDatum(i);**
- **anonymizujCislo(i);**

Index **i** ukazuje odkaz na hodnotu v poli **jmena[]** i v poli **pozice[]**.

Metody fungují na podobném principu. Jedná se o metody void, takže nevrací žádnou návratovou hodnotu.

Metoda obsahuje pole **rozlozeniPozic[]**, do kterého se pomocí funkce split rozdělí řetězec na podřetězce. Ukázka kódu:

```
rozlozeniPozic = pozice[i].split(" |\\;|\\,|\\^");
```

Tento proces je nezbytný k získání pozic řetězce, neboť původní text obsažený v poli **pozice[]** nemá správnou formu (viz část Nalezení citlivých dat – parsování).

Příklad řetězce pole **pozice[i]**:

```
bbox 250 192 1606 375; x_wconf 70
```

Pomocí funkce split tedy získáme hodnoty, viz tabulka č. 6.5.

Tab. č. 6.5: Rozdělení hodnot pomocí funkce split

k = 0	**k = 1**	**k = 2**	**k = 3**	**k = 4**	**k = 5**	**k = 6**
bbox	250	192	1606	375	x_wconf	70

Kde hodnota **k** značí index pole **rozlozeniPozic[]**.

Pro anonymizaci jsou důležité pouze souřadnice nalezeného textu, tedy **X1**, **Y1**, **X2** a **Y2**. Získaná hodnota z funkce split je však stále string, tudíž se musí stringové hodnoty převést na integerové.

Po převedení dosadím hodnoty do proměnných **X1**, **Y1**, **X2** a **Y2**.

Získání souřadnic ovšem není poslední krok, protože tyto souřadnice jsou určené pro upravený obrazový soubor, tedy ten s rozlišením 5000x5000. Tudíž dalším krokem je, převést souřadnice pro anonymizaci původního obrazového souboru.

Na to jsem vytvořil metodu **prepocti();**.

Metoda funguje na jednoduchém principu trojčlenky, kdy vím, že upravený soubor bude vždy v rozlišení 5000x5000, takže jen stačí procentuálně přepočíst souřadnice pro původní obrazový soubor.

Při načtení původního obrazového souboru si deklaruji dvě proměnné – **width** a **height**. Do **width** dosadím šířku a do **height** dosadím výšku. Proměnné pak dosadím do trojčlenky a přepočítám souřadnice.

Po získání potřebných souřadnic pak zavolám metodu

paintComponent(Graphics g, int X1, int Y1, int X2, int Y2);

kde **g** je grafické plátno (původní obrazový soubor) a **X1**, **Y1**, **X2**, **Y2** jsou pozice nalezeného textu.

Při běhu program vypisuje na obrazovku všechny úkony, co provádí. Také se vytvoří logovací soubor **log.txt**, do kterého se zapisují nejdůležitější data o anonymizaci.

Na začátku logovacího souboru se uvede obrázek a informace, zda obsahuje citlivá data, či nikoliv (Private data/Clean).

Přiklad zápisu logovacího souboru:

obrazek.png Private data

obrazek.png 25 19 160 37 JMENO

Jednotlivé údaje jsou odděleny tabulátorem.

7. Závěr

Cílem této práce bylo analýza legislativních bezpečnostních požadavků pro zpracování medicínských dat a vytvoření anonymizačního programu pro zpracování citlivých dat obsažených v obrazové části DICOM souborů.

Vytvořené metody splňují požadavek zvýšené bezpečnosti citlivých údajů při přenosu. Kromě očekávaných údajů ve formě formátu DASTA a DICOM jsou identifikovány a odstraněny citlivé údaje z obrazových souborů.

Omezujícím prvkem při tvorbě programu byl OCR program Tesseract, který přes vysokou úspěšnost rozeznávání znaků (přibližně okolo 80%), nebyl perfektní. Tuto vadu jsem se snažil do jisté míry v programu ošetřit.

V dalších verzích programu by bylo možné rozšířit funkčnost programu, např. vytvoření učenlivé databáze pro jména a příjmení, která by usnadnila anonymizaci dat, např. nebude-li jméno nebo příjmení velkými písmeny. Dále by bylo možné vytvořit grafického rozhraní s možností prohlížení upravených obrazových souborů. Také by bylo možné lépe trénovat OCR engine Tesseract, který by pak dával lepší výsledky.

Program byl vyvíjen a testován na platformě Linux Ubuntu 14.04LTS.

Literatura

[1] Konference ICT ve zdravotnictví | Inflow. *Inflow | magazín nejen pro knihovníky* [online]. 2013 [cit. 2015-04-24]. Dostupné z: http://www.inflow.cz/konference-ict-ve-zdravotnictvi.

[2] Zákony pro lidi - Sbírka zákonů ČR v aktuálním konsolidovaném znění [online]. [cit. 2015-04-24]. Dostupné z: http://www.zakonyprolidi.cz/.

[3] Kybernetický zákon [online]. [cit. 2015-04-24]. Dostupné z: http://www.kybernetickyzakon.cz/.

[4] DICOM: About DICOM. DICOM Homepage [online]. [cit. 2015-04-24]. Dostupné z: http://medical.nema.org/Dicom/about-DICOM.html.

[5] DICOM Homepage [online]. [cit. 2015-04-24]. Dostupné z: http://medical.nema.org/standard.html

[6] O projektu. *EPACS - DICOM komunikace mezi zdravotnickými zařízeními* [online]. [cit. 2015-04-24]. Dostupné z: http://www.epacs.cz/faces/pages/o-projektu.xhtml.

[7] Tesseract-ocr An OCR Engine that was developed at HP Labs between 1985 and 1995.. and now at Google. - Google Project Hosting. [online]. [cit. 2015-04-24]. Dostupné z: https://code.google.com/p/tesseract-ocr/.

[8] *ImageMagick: Convert, Edit, Or Compose Bitmap Images* [online]. [cit. 2015-04-24]. Dostupné z: http://www.imagemagick.org/script/index.php

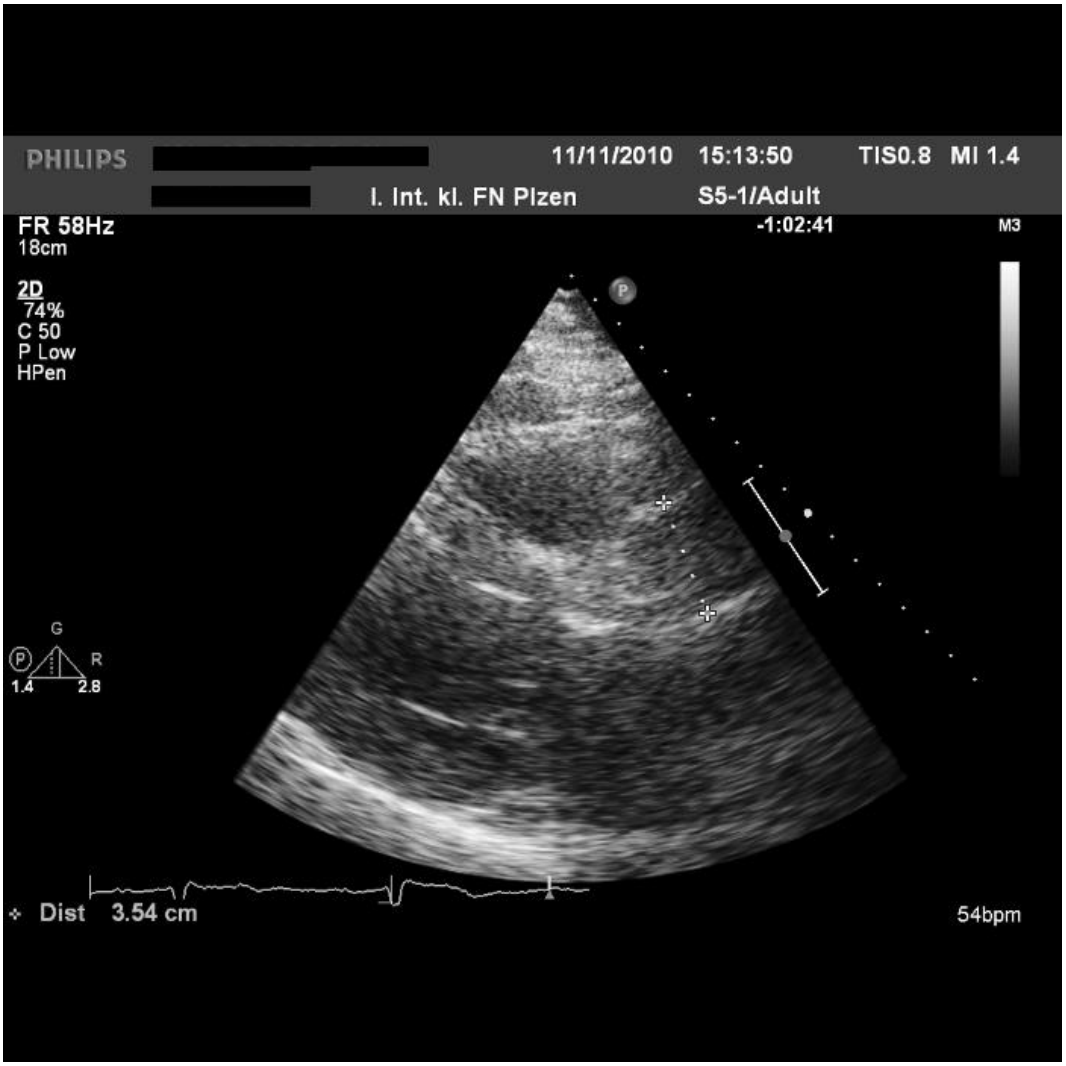
[9] Jsoup Java HTML Parser, with best of DOM, CSS, and jquery [online]. [cit. 2015-04-24]. Dostupné z: http://jsoup.org/

Přílohy

K této práci jsou přiloženy následující přílohy:

CD obsahující tuto dokumentaci ve formátu PDF a XDOC, zdrojové soubory a podpůrné knihovny.

Dále pro upřesnění a lepší představu vloženy následující obrazové soubory:



Obr. č. 7.2: Příklad výstupu anonymizačního programu.