Datový sklad ELFis
Projektová dokumentace

Příloha č. 2 závěrečné zprávy projektu

Autoři:	P. Klika, M. Blaha, M. Komenda Instituce: Lékařská fakulta Masarykovy univerzity, Institut biostatistiky a analýz
Zpracováno na základě	Smlouva o spolupráci ze dne 29.8. 2019 (projekt ELFis)
Verzování dokumentu	Verze 3.0 (10.8. 2024)

Příloha dokumentuje verzi 3.0 původně budovaného datového skladu ELFis, která pilotovala a implementovala systém doplňující interní informační systémy nemocnic o sběry dat v dotaznících a statistických šetřeních. Tento systém je stále platný a funkční a umožňuje nad současnou verzí ELFis implementovat dílčí sběry dat či realizovat observační studie. Nová verze ELFis se nicméně dominantně opírá o dobudované centralizované sběry dat a zpětný reporting pro nemocnice. Základní funkčnost tohoto systému nevyžaduje další sběry data a negeneruje tak další administrativní zátěž nemocnic a poskytovatelů obecně.

NZIS | Národní zdravotnický informační systém

nzip.cz

Obsah

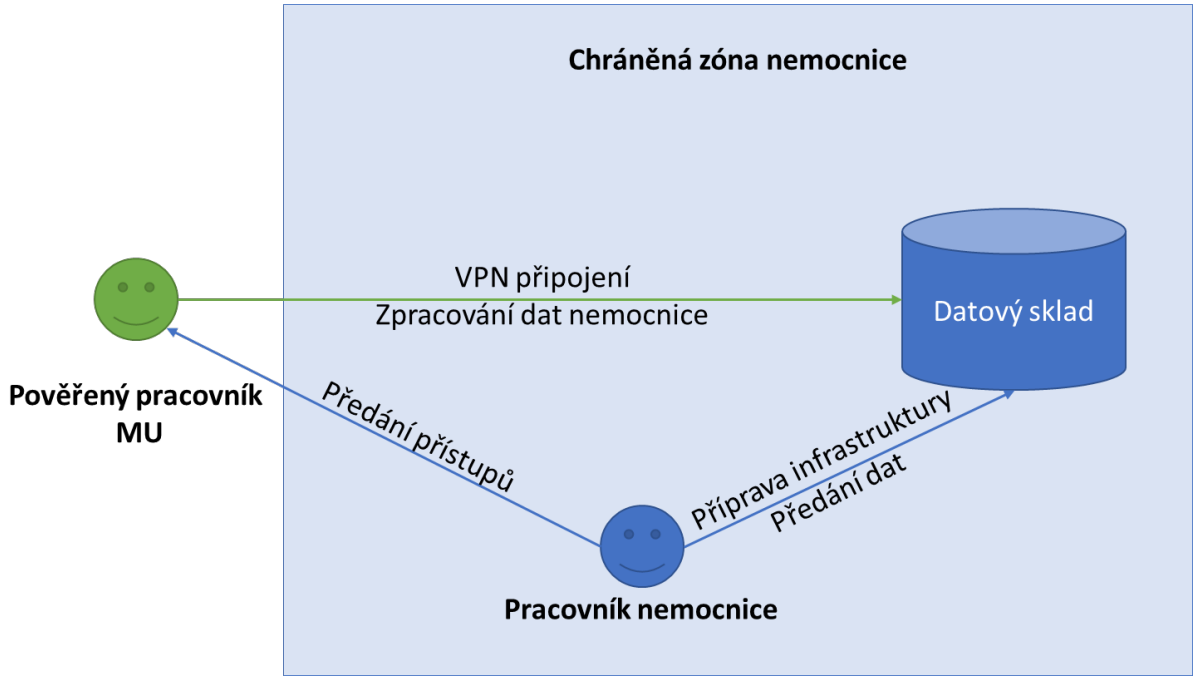
1. OBECNÁ ARCHITEKTURA DATOVÉHO SKLADU ELFIS ..... - 4 -
   1.1. *Datový sklad ELFis* ..... *- 4 -*
   1.2. *Princip zpracování dat v síti nemocnic a nakládání s daty v Datovém skladu ELFis* ..... *- 6 -*
   1.3. *Používané datové zdroje z nemocnice* ..... *- 7 -*
   1.4. *Uživatelé a další subjekty v projektu Datového skladu ELFis a jejich role* ..... *- 11 -*
2. PŘÍSTUPY DO NEMOCNIC ..... - 13 -
   2.1. *Správa přístupů* ..... *- 13 -*
   2.2. *Evidence přístupů pověřených pracovníků LF MU* ..... *- 13 -*
3. AGENT DATOVÉHO SKLADU ELFIS ..... - 14 -
   3.1. *Architektura Agenta Datového skladu ELFis* ..... *- 14 -*
   3.2. *Parametry vstupních dat* ..... *- 20 -*
   3.3. *Funkce Agenta Datového skladu ELFis* ..... *- 24 -*
   3.4. *Správa systému a přístupů* ..... *- 29 -*
4. ZPRACOVÁNÍ DAT PACIENTŮ Z DOTAZNÍKOVÉHO ŠETŘENÍ ELFIS ..... - 30 -
   4.1. *Údaje z dotazníků ELFIS* ..... *- 30 -*
   4.2. *Údaje o poskytnutých zdravotních službách pacientů z ELFis* ..... *- 31 -*
5. ZAJIŠTĚNÍ BEZPEČNOSTI OSOBNÍCH DAT V NEMOCNICÍCH ..... - 34 -
   5.1. *Technická opatření* ..... *- 34 -*
   5.2. *Organizační opatření* ..... *- 34 -*

1. Obecná architektura Datového skladu ELFis

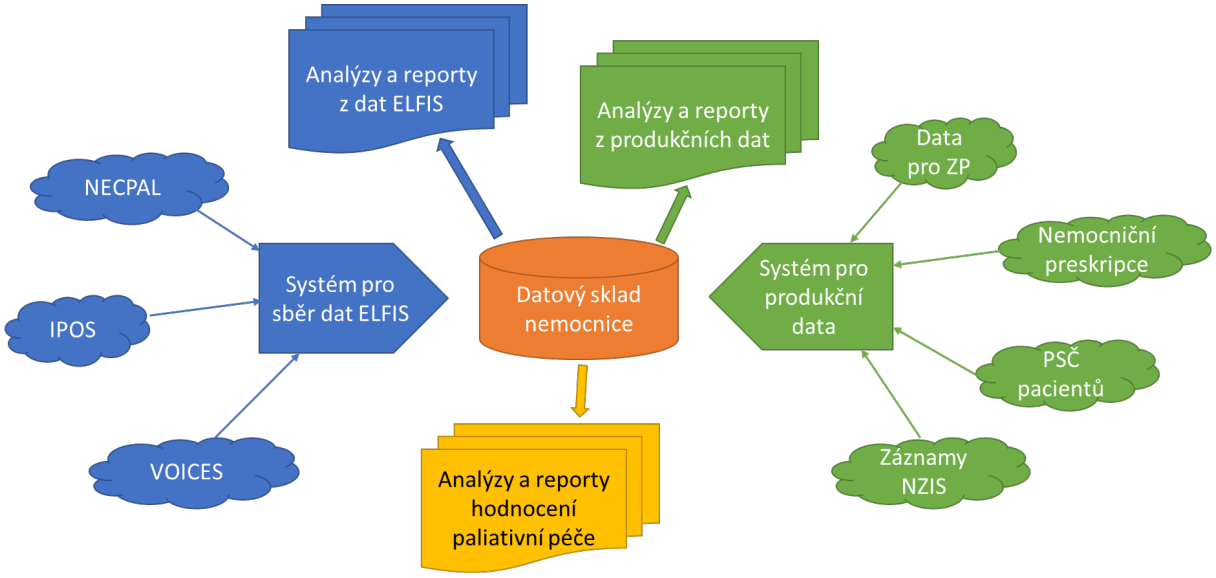
1.1. Datový sklad ELFis

Navržený systém je postaven na architektuře datového skladu, provozovaného uvnitř nemocnice. Komponenty systému jsou umístěny na infrastruktuře spravované nemocnicí a primární data nejsou přenášena jinam. Celý systém je postaven na zdarma dostupných technologiích, které nemocnici negenerují další vícenáklady na pořízení nebo provoz. Centrálním prvkem datového skladu je databáze, postavená na technologii MySQL. Dále se používají skripty a nástroje vytvořené v programovacích jazycích PHP nebo Java. Je podporován běh na operačních systémech Windows i Linux. Přesná specifikace požadavků na infrastrukturu je uvedena v příloze tohoto dokumentu.

Data sbíraná v tomto systému jsou zcela neosobní, tedy neobsahují identifikace žádných osob (pacientů, zaměstnanců nebo externích subjektů) ani žádné osobní nebo jiné citlivé údaje. Data se do systému předávají formou jednorázových exportů textových souborů dle zadaného datového rozhraní, zajištěný pracovníky nemocnice, nikoliv přímým přístupem do informačních systémů nebo jiných databází. Rozsah sbíraných dat je uveden v samostatné příloze. Přístup do nemocnice se řídí pravidly a bezpečnostní politikou nemocnice, dodavatel systému je povinen dodržovat veškeré požadavky na zabezpečení dat a informačních systémů nemocnice. Přístup do infrastruktury nemocnice je podmíněn platnou smlouvou mezi dodavatelem a nemocnicí a vytváří se pro předem definovaného pracovníka dodavatele. Technicky je obvykle chráněn pomocí VPN připojení a má vyhrazený přístup pouze k vybranému serveru s provozovaným systémem. Typická architektura řešení je zachycena následujícím diagramem:



Jednotlivé komponenty a datové zdroje, které jsou integrovány do podoby výstupů a analýz v rámci datového skladu, jsou zobrazeny na následujícím diagramu.



Klíčovým procesem je sběr dat v dotaznících v rámci projektu ELFis. Zde je nutná součinnost s pracovníky příslušných oddělení nemocnice pro zajištění procesu vyhodnocení a pořizování dotazníkového šetření.

Následuje fáze zpracování dodaných materiálů do podoby použitelné pro analýzy, včetně validací, čištění, doplnění a integraci předaných údajů (např. číselníky, referenční data atd). V poslední fázi probíhá příprava analýz a výstupů ze zpracovaných dat. Výstupy jsou ve formě datových tabulek (excel) a dokumentů (analýzy vybraných oblastí – powerpoint), přímý přístup ke zpracovaným datům v datovém skladu nemocnice nebo pomocí nástroje pro jejich prohlížení (business intelligence nástroj).

1.2. Princip zpracování dat v síti nemocnic a nakládání s daty v Datovém skladu ELFis

Architektura systému Datového skladu ELFis je postavena na síti nemocnic Kraje Vysočina. Ve vnitřní síti každé z nich je zprovozněna aplikace, která zajišťuje zpracování a nevratnou de-identifikaci nemocničních dat. Výsledným produktem je datový sklad, který obsahuje integrovaná nemocniční data do podoby vhodné pro další analytické zpracování.

Základním datovým zdrojem jsou administrativní data nemocnic, která tyto nemocnice vykazují zdravotním pojišťovnám, tzv. k-dávky, doplněné o případné další datové zdroje (nemocniční preskripce, PSČ bydliště pacientů). Nemocniční data jsou procesována na vlastním serveru každé partnerské nemocnice zvlášť, všechny nemocnice mají tedy pod kontrolou svá vlastní data. Spojování dat za účelem vzájemného srovnávání center projekt neumožňuje. Na zmíněném serveru také probíhá spojení administrativních dat onkologických pacientů s diagnostickými záznamy, které daná nemocnice hlásí do Národního onkologického registru (NOR). Tyto záznamy Národního onkologického registru (NOR) pacientů léčených v dané nemocnici jsou spojeny se záznamy o léčbě do jedné databáze uvnitř zdravotnického zařízení. Záznamy NOR jsou k nemocničním datům přiřazovány na základě šifry jejich rodných čísel, které vznikají v obou případech stejným způsobem. Všechny operace s daty se týkají výhradně záznamů pacientů léčených v dané nemocnici a probíhají výhradně na interních serverech dle bezpečnostních protokolů dané nemocnice.

Software Datového skladu ELFis, vyvinutý na Masarykově Univerzitě (MU), pod dohledem pověřeného IT experta nemocnice tato data v interní databázi nemocnice transformuje a provádí jejich anonymizaci (nevratnou de-identifikaci): čísla pojištěnců jsou nahrazena šifrou, vzniklou jednosměrnou hešovací funkcí (SHA) s tajným heslem (salt). Všechny ostatní osobní údaje v databázi pro analýzy jsou nevratně smazány. Výsledná de-identifikovaná data jsou přesunuta do oddělené části databáze, která je přístupná pověřenému pracovníkovi LF MU a ve které se již žádná osobní data nevyskytují. Veškeré analýzy jsou prováděny pouze nad anonymizovanými a agregovanými daty.

Veškerá práce s primárními daty, obsahujícími osobní údaje, probíhá v rámci servisu a údržby systému Datového skladu ELFis na serveru nemocnice. Systém je nastaven tak, aby přístup k osobním údajům měl pod kontrolou pouze a jedině pověřený pracovník nemocnice. Ve všech fázích procesu práce s daty je aplikována celá řada opatření (smluvních, organizačních i technických) pro zajištění bezpečnosti, zvláště u osobních dat, ale i všech ostatních citlivých nemocničních dat: šifrování přístupů, oddělené účty a přístupová práva, hesla pro šifrování čísel pojištěnců, bezpečné mazání atd. Přístupy do nemocnic jsou vždy řízeny bezpečnostní politikou každé jednotlivé nemocnice a jsou dodržovány její požadavky a standardy.

Ochrana primárních dat je zajištěna robustními mechanismy, mj. smluvně (včetně podmínky naprosté mlčenlivosti všech pracovníků dodavatele), jak je obvyklé v případech, kdy dodavatel spravuje a provozuje v nemocnici systém pracující s čísly pojištěnců, jako například nemocniční informační systém či jiné provozní systémy v nemocnicích. Nastavený model práce zde plně odpovídá tomuto plošně aplikovanému modelu. Osobní data nikdy neopouští server nemocnice a bezprostředně po jejich transformaci jsou pro analyticky využívanou databázi bezpečně a nevratně smazána.

Jelikož principem projektu je poskytovat nemocnicím zejména referenční srovnání formou předpřipravených reportů, jsou de-identifikovaná data nemocnice přenášena do referenčního datového skladu na serveru spravovaném LF MU. Na tomto serveru se nikdy nevyskytovala a nevyskytují žádná osobní data pacientů a záznamy slouží k poskytování agregovaných podkladů pro analytická zpracování referenčních hodnot pro všechny zapojené nemocnice. Na tento server jsou uplatňována interní pravidla LF MU pro zabezpečení citlivých dat, která jsou v souladu s certifikací ISO

27000. Také veškeré výstupy z tohoto datového skladu jsou řízeny a evidovány technickými prostředky. Ztotožnění identity jedince není z agregovaných referenčních dat možné.

Platí tedy, že žádná data, obsahující osobní údaje, neopouštějí za žádných okolností server nemocnice a LF MU přistupuje k tomu systému na základě uzavřené smlouvy analogicky k provozovatelům podobných informačních systémů v nemocnicích, za dodržení bezpečnostních požadavků nemocnice. Kompletní dokumentace systému Datového skladu ELFis je k dispozici jako samostatný dokument, kde je podrobně popsán princip, metody a opatření pro práci s daty a jejich ochranu.

1.3. Používané datové zdroje z nemocnice

Pro základní hodnocení nemocničních dat v oblasti onkologie jsou uvnitř nemocnic zpracovávány dva hlavní datové zdroje: data předávaná pojišťovnám (administrativní data nemocnice, „k-dávky“) a záznamy hlášené do Národního onkologického registru (NOR) o pacientech léčených v dané nemocnici. Dále jsou používány doplňující interní datové zdroje, jako jsou údaje o nemocničních preskripcích, data z nemocničního informačního systému s PSČ bydliště pacientů, různé číselníky apod. Jejich popis je popsán v následujících částech.

Administrativní data nemocnic

Nemocniční informační systémy (NIS) obsahují řadu cenných informací, jejich přímé a jednotné využití pro analýzy však bývá problematické. Různé nemocnice bohužel provozují rozdílné NIS, které obvykle neobsahují data ve strukturované podobě. Navíc data z NIS nejsou vždy snadno dostupná za rozumných nákladů pro jejich provozovatele. Proto projekt Datového skladu ELFis využívá jako zdroj administrativních dat interní výkazy plátcům zdravotní péče, tzv. k-dávky. Tyto výkazy jsou povinné, dostupné v nemocnici za několik let zpětně a zcela nezávislé na konkrétním NIS.

Technicky vzato jsou k-dávky obyčejné textové soubory (viz následující obrázek) s definovanou strukturou, která je dána metodikou a datovým rozhraním Všeobecné zdravotní pojišťovny (VZP) [https://www.vzp.cz/poskytovatele/vyuctovani-zdravotni-pece/metodika-vyuctovani-aktualni-stav]. Tato struktura je ovšem proměnná v čase, s čímž je nutné počítat při jejich zpracování. V k-dávkách lze nalézt zejména údaje o provedených výkonech a o podaných přípravcích v rámci hospitalizační i ambulantní péče.

```
DP98[redacted]2300200502101636315      57911          1406.481
A 43418600 1111123101235152501502     11338R100            0          0.00    119
V10022005520231          119
A 44119900 2111123101286171501701     11339H660            0          0.00    189
V19022005710221          152
V19022005715331           37
A 43592300 3111123101286171501701     11340H650            0          0.00    189
V13022005710221          152
V13022005715331           37
A 43056000 4111123101160131501301     11341J303            0          0.00    714
V07022005310211          389
V07022005272409          126
V07022005272409          126
V07022005272402           28
V28022005095131           45
Z 43056100 5231011601315013O1     11341                140.41
L070220051 84296      0.660    140.41
A 42528800 6111123101525122301202     11342D759            0          0.00    229
V02022005220221          229
A 42685700 7111123101286171501701     11342J00             0          0.00    152
V02022005710221          152
A 44574500 8111123101206151501501     11343S619            0          0.00    162
V25022005510221          162
A 44584300 9111123101255166502606     11343S424            0          0.00     73
V26022005660231           73
A 42900200 1011123101286171501701     11344H680            0          0.00    181
V03022005710221          152
V03022005730171           29
GH730
A 43823000 1111123101160131501301     11345J450            0          0.00     45
V15022005095131           45
A 42112600 1211123101160131501301     11346J459            0          0.00     45
V10022005095131           45
A 44381400 1311123101286171501701     11346J352            0          0.00    266
V23022005710221          152
V23022005713171          114
A 41344100 1411123101328144501405     11347B07             0          0.00    138
V07022005450221          138
A 42811900 1511123101235152501502     11348R100            0          0.00    203
V03022005520221          203
A 43546700 1611123101135129501209     11349R51             0          0.00    847
V11022005291231          625
V11022005291251          222
A 41966500 1711123101235152501502     11350S823            0          0.00    119
V01022005520231          119
A 44131300 1811123101286171501701     11350H650            0          0.00    189
V20022005710221          152
```

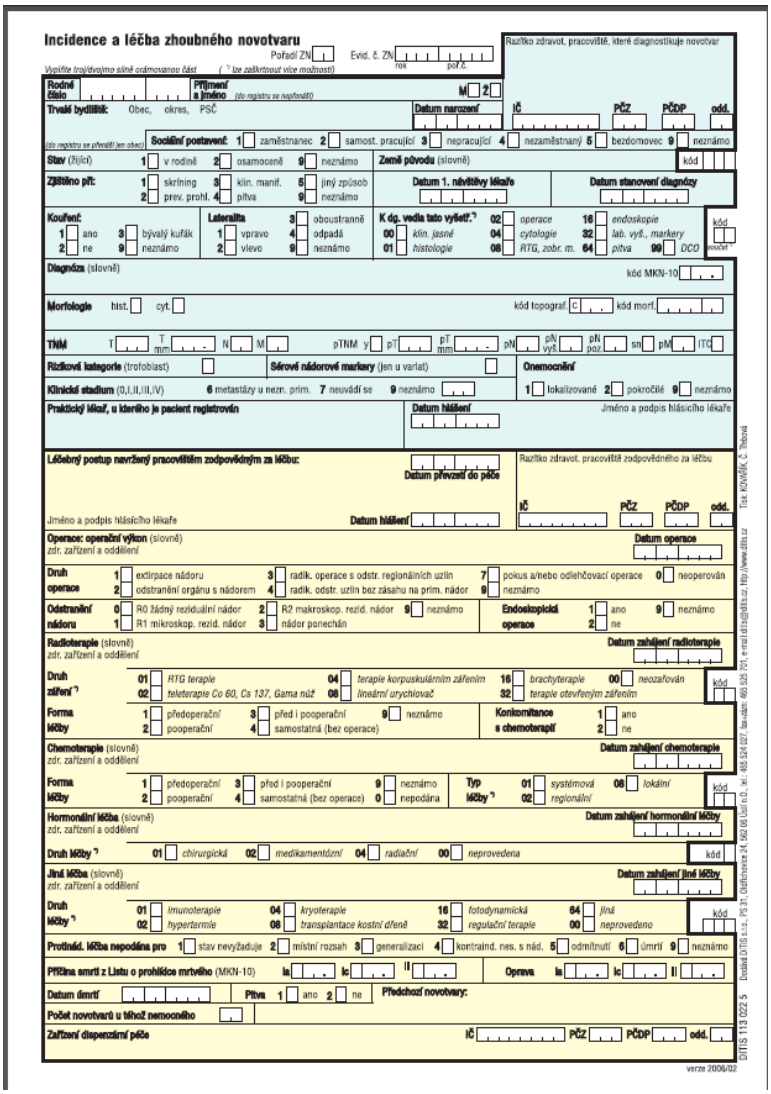
Struktura interně analyzovaných administrativních dat nemocnice je hierarchická. Na nejvyšší úrovni je tzv. hlavička dávky, která popisuje nemocnici a období, za která jsou data předávána. Pod ní jsou evidovány jednotlivé doklady – výkazy o formě poskytnuté péče pacientovi. Základními doklady jsou 01 – Vyúčtování výkonů v ambulantní péči, 02 – Vyúčtování výkonů v ústavní péči, 03 Zvlášť účtované léčivé přípravky a ZP, 06 – Poukaz na vyšetření a ošetření a 10 – Recept. Na nejnižší úrovni jsou pak jednotlivé řádky dokladů – konkrétní detailní údaje o poskytnuté péči, zejména provedené výkony a aplikovaná/vydaná léčiva a materiál.

Zpracovány jsou vždy doklady, které byly vykázány danou nemocnicí a případně její ústavní lékárnou. U ní platí, že jsou zde vykázány všechny recepty v této lékárně vydané. Mohou zde být proto recepty pacientů, které byly předepsány v jiném zdravotnickém zařízení (tyto jsou ze zpracování dále vyřazeny). Naopak, pokud si pacient nemocnice předepsaný recept vyzvedne v jiné lékárně, tuto informaci se z těchto dat nedozvíme. Pro tento účel je vhodnější datový zdroj nemocniční preskripce (viz část 1.2.3).

V zásadě lze konstatovat, že k-dávky popisují kompletně proces péče o konkrétního pacienta v daném zdravotnickém zařízení, byť spíše s ohledem na provozní stránku péče a se zanedbáním některých konkrétních detailů.

Záznamy národního onkologického registru hlášené nemocnicí

Národní onkologický registr je strukturovaná databáze, která tvoří jednu ze základních částí Národního zdravotního informačního systému. Do tohoto registru musí být ze zákona povinně zaznamenán každý nově diagnostikovaný novotvar v ČR již od roku 1976 (viz obrázek hlášenky NOR). Tato epidemiologická databáze obsahuje základní klinické parametry, jako diagnózu a stadium, které rozhodují o prognóze pacienta, jakožto i základní údaje o schématu jeho léčby.


Incidence a léčba zhoubného novotvaru

Záznamy NOR pacientů léčených danou nemocnicí jsou interně napojeny k administrativním datům nemocnice a obohacují interní elektronickou zdravotnickou dokumentaci zejména o klinické stadium v době diagnózy. Každý záznam v NOR je v datovém skladu napojen na velké množství číselníků (pro pohlaví, diagnózy, léčebné modality apod.). Mezi základní údaje patří detailní údaje o diagnóze, rozsahu onemocnění (TNM a stadium), datum diagnózy, data zahájení léčebných modalit a jejich povaha.

Nemocniční preskripce

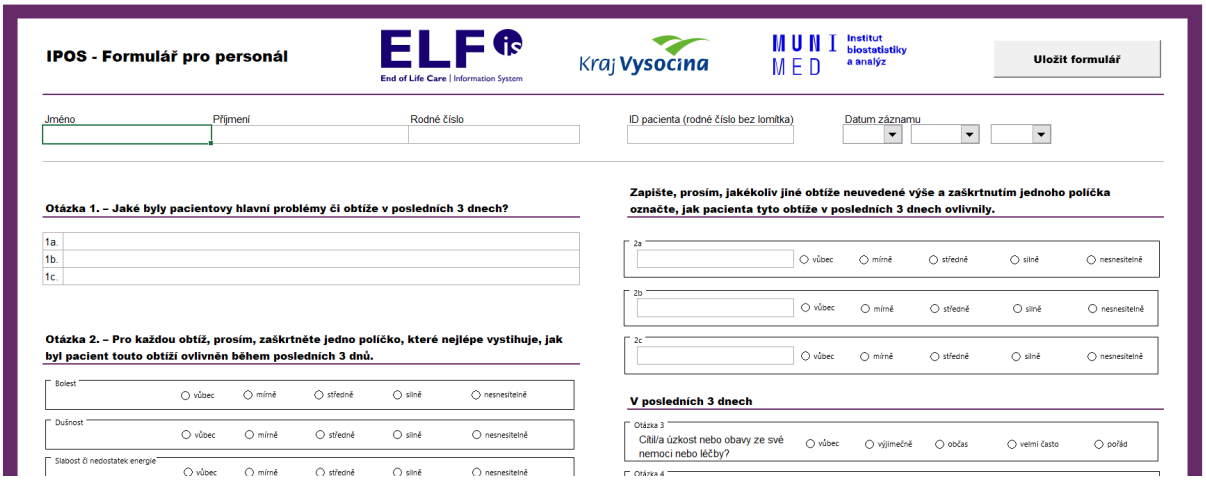
Nemocniční preskripce jsou záznamem o předepsání léčiva nebo zdravotnického materiálu pacientovi lékařem nemocnice. K jeho evidenci se obvykle používá samostatný modul NIS – evidence nemocničních preskripcí. Obsahuje údaje o všech receptech, které lékaři této nemocnice pacientům předepsali, bez ohledu na to zda a ve které lékárně si léčivo nebo materiál vyzvedli. V tomto případě tedy nedochází ke ztrátám dat o předepsaných léčivech, jak tomu hrozí v případě dokladů Recepty z datového rozhraní VZP (viz část 3.3.3).

Data dotazníků ELFIS

Data dotazníků ELFIS se sbírají do připravených šablon nástroje MS Excel. Každý pacient má pro každý dotazník vlastní soubor, který obsahuje veškeré zaznamenané údaje v podobě, která je dále programově zpracovatelná. Dále jsou ukázky 3 základních formulářů, které se v rámci projektu ELFIS zpracovávají – NECPAL, IPOS a VOICES.

Formulář NECPAL

Formulář IPOS:



Formulář VOICES:

VOICES

1. Jak dlouho byl(a) pacient(ka) nemocný(á) předtím, než zemřel(a)?

2. Strávil(a) pacient(ka) alespoň nějaký čas doma v posledních třech měsících života?

Domácí péče

3. V době, kdy byl(a) pacient(ka) během posledních tří měsíců života doma, byla mu/jí tam poskytnuta/poskytována nějaká z níže uvedených služeb?

- [x] sestra domácí péče
- [ ] pečovatel(ka), odlehčovací
- [ ] sociální pracovník(ce)
- [ ] duchovní
- [ ] rozvoz jídla
- [ ] domácí hospic (tedy služba, která má vlastního lékaře součástí týmu)
- [ ] ergoterapeut(ka)
- [ ] zapůjčení pomůcek (polohovací postel, antidekubitní matrace, koncentrátor kyslíku)
- [ ] nebyl poskytován žádný typ péče
- [ ] nevím
- [ ] jiné

Více o dotaznících a sběru dat v projektu ELFIS je možno zjistit na webu http://elfis.iba.muni.cz.

1.4. Uživatelé a další subjekty v projektu Datového skladu ELFis a jejich role

Systém Datového skladu ELFis a jeho výstupy používají následující skupiny uživatelů:

Tým Datového skladu ELFis

Vývojářský tým celého řešení datového skladu v roli věcného a technického správce systému. Zodpovídají za návrh, vývoj a údržbu celého systému, předávání dat oprávněným subjektům apod.

Analytický tým ELFis

Pracovníci LF MU, kteří mají přístup k předaným datům zapojených center ELFis a zodpovídají za provádění analytických výstupů z předaných dat. Data jsou jim předávána týmem Datového skladu ELFis buď jako standardizovaný export do statistického nástroje, ad-hoc definované exporty pro účely konkrétních analýz nebo je jim v některých případech zařízen přímý přístup do databáze k vybraným datovým tabulkám. Za data předaná analytickému týmu zodpovídá hlavní věcný správce systému Datového skladu ELFis. Předávaná data jsou vždy nevratně anonymizovaná, bez jakýchkoliv osobních údajů o pacientech.

Analytický tým ELFis plní požadavky oprávněných subjektů pro přístup k výstupům ze systému Datového skladu ELFis. Těmi jsou výhradně hlavní management a odborní garanti jednotlivých zapojených center a jimi pověření pracovníci nemocnice.

Pověřený IT pracovník nemocnice

Úkolem pověřeného IT pracovníka nemocnice je zajišťovat aktualizaci dat Datového skladu ELFis centra na základě dohody s týmem Datového skladu ELFis, obvykle jednou ročně. Získává data z informačních systémů nebo jiných oddělení nemocnice, zajišťuje jejich iniciální zpracování a de-identifikaci. Představuje hlavní kontaktní osobu pro tým Datového skladu ELFis na další specialisty v oblasti IT. Některé činnosti může delegovat na další spolupracovníky.

Vedoucí management a odborní garanti center ELFis

Nejvyšší vedení zapojených nemocnic a pověření zástupci pro projekt ELFis. Jsou oprávněni žádat o předání výstupů ze systému Datového skladu ELFis, ať již v podobě statistických přehledů, nebo analytických výstupů z nich. Schvalují využití anonymizovaných dat pro publikace.

Další role

Vedením LF MU je určen hlavní manažer projektu, který zajišťuje koordinaci zapojených center v rámci projektu a poskytuje jim metodickou podporu. Nepodílí se na předávání dat ani jejich analytickém hodnocení a nemá fyzický přístup k žádným komponentám systému Datového skladu ELFis.

Na straně zapojených nemocnic jsou definovány týmy provozovatelů systému (IT support), kteří zajišťují přístupy pro členy týmu Datového skladu ELFis, zajišťují výpočetní prostředí pro provoz Agent datového skladu ELFis, aktualizace a zálohování, bezpečnostní politiku nemocnice atd.

2. Přístupy do nemocnic

Možnost přístupů do nemocnic přes zabezpečené komunikační kanály je pro fungování projektu klíčová.

2.1. Správa přístupů

Vzdálené přístupy jsou pověřeným pracovníkům dodavatele zřizovány pracovníky nemocnice na základě smlouvy o spolupráci mezi oběma institucemi. O zřízení přístupu pro konkrétní osobu žádá hlavní manažer projektu na straně LF MU, který předává zodpovědnému pracovníkovi nemocnice požadované kontaktní údaje osoby, pro niž se přístup zřizuje (jméno, email, telefon).

Vlastní proces zřízení přístupu a jeho technická implementace je čistě v kompetenci pracovníků nemocnice, kteří se řídí interními pravidly pro poskytování přístupů a jejich zabezpečení. Standardně je zabezpečený vzdálený přístup do nemocnice zajištěn pomocí specifikované VPN sítě. K jejímu zřízení bývá obvykle požadováno vyplnění protokolu o zřízení VPN, v některých případech i smlouva mezi institucemi. Na samotný server je pak přístup přes klienta Remote desktop (RDP na Windows Server) nebo SSH klienta (Linux Server).

Mohou být požadovány i další doplňující prvky ochrany, např. periodické obnovování žádostí o VPN přístup, pravidelná změna hesla na server aj.

2.2. Evidence přístupů pověřených pracovníků LF MU

Veškeré přístupové údaje, které byly pracovníkům dodavatele předány ze strany nemocnic, jsou ošetřeny v souladu s pravidly maximální ochrany citlivých údajů, odpovídající ISO 27000. Nikdy se nevyskytují zapsané v otevřené podobě přístupné jiným než oprávněným uživatelům. Jsou ukládány v bezpečném úložišti hesel, zabezpečeným hlavním heslem. Přístupové údaje jsou k dispozici pouze osobám, kterým byly pracovníky nemocnic předány.

V okamžiku, kdy jakákoliv osoba dodavatele v roli správce komponenty Agent Datového skladu ELFis s přístupy do nemocnic z projektu odejde nebo změní roli, jsou spolupracující centra o tomto faktu informována, všechny účty jsou jí zablokovány a změněna hesla k přístupům, které měla dotyčná osoba k dispozici.

3. Agent Datového skladu ELFis

Agent Datového skladu ELFis (Agent) je komponenta, která je provozována na serveru uvnitř spolupracující nemocnice a je zodpovědná za zpracování primárních dat nemocnice, které mohou obsahovat osobní údaje. Agent je používán k nevratné de-identifikaci záznamů.

3.1. Architektura Agenta Datového skladu ELFis

Agent se skládá z několika základních částí. Jádrem celého systému je databáze, která provádí většinu procesu zpracování primárních nemocničních dat a jejich de-identifikaci. Doplněna je sadou obslužných knihoven v PHP, Linux/Windows shell skriptů, archivačním programem pro zabezpečené ukládání citlivých dat atd.

Databáze

Použita je databáze MySQL licencovaná jako open-source a free, která je standardem při používání v nekomerčních projektech.

Schémata

Pro funkcionalitu Agenta jsou potřeba v databázi 3 databázová schémata:

icop_dw1_koc_import

Slouží pro zpracování primární dat nemocnice, která obsahují osobní údaje pacientů. Má do ní přístup pouze uživatel pověřeného pracovníka nemocnice (obvykle “root”) za účelem zpracování nových dat. Všechna data obsahující osobní údaje se po skončení importu nových dat (jednou ročně) mažou.

icop_dw1_koc_import_anonym

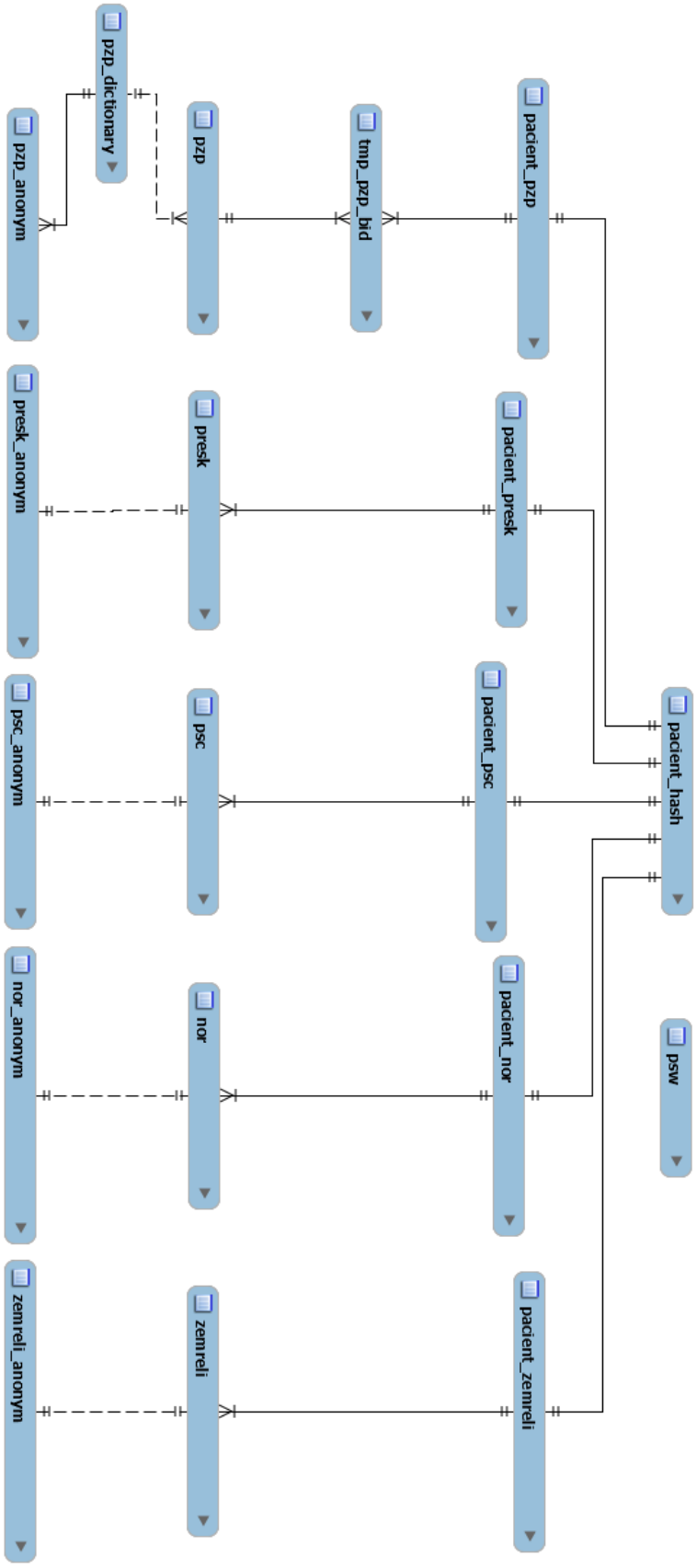
Slouží pro uložení výsledku iniciálního zpracování nemocničních dat, která již neobsahují žádné osobní údaje pacientů. Jsou přístupná týmu Datového skladu ELFis.

icop_access

Schéma sloužící k uchování parametrů procesu importu nových dat, archivaci logů z průběhu jejich zpracování apod. Neobsahuje ani žádná primární data nemocnice, ani žádné osobní nebo jiné citlivé údaje. Je přístupná pověřenému pracovníkovi nemocnice i týmu Datového skladu ELFis.

ERD schématu icop_dw1_koc_import

Obrázek 1 ER diagram zachycující základní komponenty Agenta s daty obsahujícími osobní údaje



pacient_hash

Slouží jako převodník mezi číslem pojištěnce a jeho šifrou. Ke každému číslu pojištěnce je přiřazena odpovídající šifra. Po každém použití je vždy archivován v zašifrovaném archivu na disku nemocnice (pod kontrolou pracovníka nemocnice) a z databáze je nevratně smazán. Analytický tým ELFis nemá žádnou možnost identifikovat osobu pacienta.

Název	Datový typ	Integritní omezení	Popis
rodne_cislo	varchar(10)	NN, PK	Číslo pojištěnce
rodne_cislo_hash	varchar(64)	NN, U	Šifra pacienta vzniklá jednosměrnou šifrovací funkcí

psw

Ukládá otisk hesla (soli), které se používá pro šifrování čísel pojištěnců při procesu de-identifikace. Slouží jako kontrola hesla, aby nedošlo k šifrování různými hesly, které by znemožnily jednoznačné přiřazení pacientů mezi různými importy dat. Plně pod kontrolou pracovníků nemocnice.

Název	Datový typ	Integritní omezení	Popis
id	int	NN, PK, AI	Surrogate key
psw	varchar(32)	NN, U	Otisk hesla (soli) používaný při šifrování čísel pojištěnců

pacient_pzp, pacient_presk, pacient_psc, pacient_nor, pacient_zemreli

Tyto tabulky mají stejnou strukturu. Uchovávají v primárních datech nemocnice základní údaje o pacientech, kteří byli dohledáni v primárních datech nemocnice. Tyto údaje jsou odvozeny z čísla pojištěnce. Z něj jsou odvozeny údaje o datu narození, pohlaví a státní příslušnosti. V opačném případě jsou tyto údaje neznámé.

Název	Datový typ	Integritní omezení	Popis
id	Int	NN, PK, AI	Surrogate key
rodne_cislo	varchar(10)	NN, U	Číslo pojištěnce dohledané v primárních datech
rodne_cislo_hash	varchar(64)		Šifra čísla pojištěnce po procesu de-identifikace
pohlavi	char(1)		0 = muž, 1 = žena, 2 = neznámo
datum_narozeni	varchar(10)		Datum narození ve formátu YYYYMMDD
je_cizinec	char(1)		0 = ne, 1 = ano, 2 = neznámo

pzp_dictionary

Slovník pro identifikaci známých typů vět v datech pojišťoven. Obsahuje informaci pro rozpoznání typu věty podle počátečního písmene a délky řádku. Dále obsahuje údaj o případném obsahu osobních údajů v daném typu věty – jaké údaje se zde nacházejí a na jaké pozici v řádku. Tyto údaje pak slouží v procesu de-identifikace pod kontrolou pracovníků nemocnice.

Název	Datový typ	Integritní omezení	Popis
id_sys	int	NN, PK, AI	Surrogate key
firstchar	varchar(1)	NN	První znak na řádku věty
sentencelength	smallint	NN	Délka věty v počtu znaků
begin	smallint		Počáteční pozice případného osobního údaje v tomto typu věty
length	smallint		Koncová pozice případného osobního údaje v tomto typu věty
flag	varchar(1)		Typ případného osobního údaje (N=nic, R=číslo pojištěnce, D=jiný osobní údaj)

notice	varchar(255)	Poznámka

pzp, pzp_anonym

Obsahuje záznam pro každý načtený řádek primárních dat vykázaných zdravotní pojišťovně. Kromě vlastního znění řádku obsahuje již odvozené údaje pro identifikaci typu věty, nadřazených záznamů (dávka, doklad) a identifikace konkrétního spuštění procesu importu primárních dat.

Název	Datový typ	Integritní omezení	Popis
id_sys	int	NN, PK, AI	Surrogate key
radka	varchar(400)	NN	Celý obsah řádku věty PZP v původním znění; v případě tabulky pzp_anonym jsou všechna osobní data v atributu radka vymazána
kod_vety	char(1)		Úvodní znak řádku určující typ věty
delka_vety	tinyint		Délka řádku věty určující typ věty
kod_dokladu	char(2)		Výsledná klasifikace typu věty dle typu a délky
davka	varchar(50)		Identifikátor dávky, ve které byl řádek vykázán
doklad	varchar(30)		Identifikátor dokladu, ve kterém byl řádek vykázán
doklad_subs	varchar(30)		Část identifikátoru dokladu
nadrazeny_doklad	varchar(30)		Pro podřízené doklady (03 – zvlášť účtovaná léčiva a materiál) odkazuje na nadřízený doklad (01, 02, 06)
id_pacient	int	FK -> pacient_pzp (id_pacient)	Identifikátor pacienta, ke kterému byl doklad vykázán
zz_importu	char(10)	NN	Identifikátor zařízení, ze kterého pocházejí primární data
datum_importu	datetime	NN	Čas zahájení importu primárních dat
flag	tinyint		Příznak dohledání typu věty ve slovníku známých typů vět (pzp_dictionary)
rok_uzavreni	smallint		Rok uzavření dokladu
mesic_uzavreni	tinyint		Měsíc uzavření dokladu

tmp_pzp_bid

Slouží k vytvoření asociace mezi řádkem záznamu v tabulce pzp a identifikátorem pacienta. Obsahuje záznam pro každý řádek tabulky PZP, kde se číslo pojištěnce vyskytuje a je používána ke snadnému dohledání a nahrazení těchto identifikátorů při procesu de-identifikace.

Název	Datový typ	Integritní omezení	Popis
id_sys	Int	NN, PK, FK -> pzp (id_sys)	Odkaz na řádek v tabulce PZP
bid	varchar(64)	NN, PK, FK -> pacient_pzp (rodne_cislo)	Odkaz na řádek v tabulce Pacient_PZP; po provedení anonymizace je číslo pojištěnce nahrazeno šifrou

presk, presk_anonym

Obsahuje jeden záznam pro každé předepsané léčivo nebo materiál, evidované v NIS. Vztahuje se ke konkrétnímu pacientovi pro daný typ léčiva v daném dni na daném oddělení. V tabulce **PRESK_ANONYM** je sloupec **RC** nahrazen sloupcem **RC_HASH**, který obsahuje šifru čísla pojištěnce.

Název	Datový typ	Integritní omezení	Popis

inscomp	varchar(3)		Číslo pojišťovny pacienta
datrece	varchar(10)		Datum předepsání léčiva
rc	varchar(10)		Číslo pojištěnce
drug	varchar(7)		Kód léčiva
quantity	varchar(10)		Množství předepsaného léčiva
czicz	varchar(8)		IČP předepisujícího oddělení
sk	varchar(1)		Skupina léčiv/materiálu (1=HVLP, 2=IVLP, 3=PZT, 4=STOMAG)
kc	varchar(15)		Vykazovaná cena předepsaného léčiva
atc	varchar(7)		ATC kód předepsaného léčiva
nazevatc	varchar(100)		Název ATC skupiny předepsaného léčiva
nazev	varchar(100)		Název léčiva
dg	varchar(5)		Diagnóza, pro kterou bylo léčivo předepsáno
jine	varchar(100)		Další údaje, vztahující se k preskripci
zz_importu	varchar(10)	NN	Identifikátor zařízení, ze kterého pocházejí primární data
datum_importu	datetime	NN	Čas zahájení importu primárních dat
id_pacient	Int	NN, FK -> pacient_presk (id_pacient)	Identifikátor pacienta

psc, psc_anonym

Obsahuje údaje o bydlišti pacienta z interního NIS nemocnice, případně další údaje. V tabulce **PSC_ANONYM** je sloupec **RC** nahrazen sloupcem **RC_HASH**, který obsahuje šifru čísla pojištěnce. Sloupce **ULICECISLO, ULICE, CP, CO, JMENO, KRESTNI** a **PRIJMENI** jsou smazány.

Název	Datový typ	Integritní omezení	Popis
rc	varchar(10)	NN	Číslo pojištěnce
psc	varchar(6)		PSČ kód bydliště
poj	varchar(3)		Číslo pojišťovny pacienta
ulicecislo	varchar(100)		Ulice a číslo domu
ulice	varchar(100)		Ulice
cp	varchar(10)		Číslo popisné
co	varchar(10)		Číslo orientační
obec	varchar(8)		Obec bydliště pacienta (statistický kód obce)
jmeno	varchar(100)		Jméno a příjmení pacienta
krestni	varchar(50)		Křestní jméno pacienta
prijmeni	varchar(50)		Příjmení pacienta
inicialy	varchar(2)		Iniciály (první písmeno křestního jména a příjmení)
jine	varchar(200)		Jakékoliv jiné údaje o pacientovi
zz_importu	varchar(10)	NN	Identifikátor zařízení, ze kterého pocházejí primární data
datum_importu	datetime	NN	Čas zahájení importu primárních dat
id_pacient	int	NN, FK -> pacient_psc (id_pacient)	Identifikátor pacienta

nor, nor_anonym

Obsahuje záznam pro každý načtený řádek primárních dat Národního onkologického registru (NOR). Obsahuje všechny záznamy pacientů dané nemocnice nahlášené do NOR. V tabulce **NOR_ANONYM** je sloupec **RODCIS** nahrazen sloupcem **RODCIS_HASH**, který obsahuje šifru čísla pojištěnce.

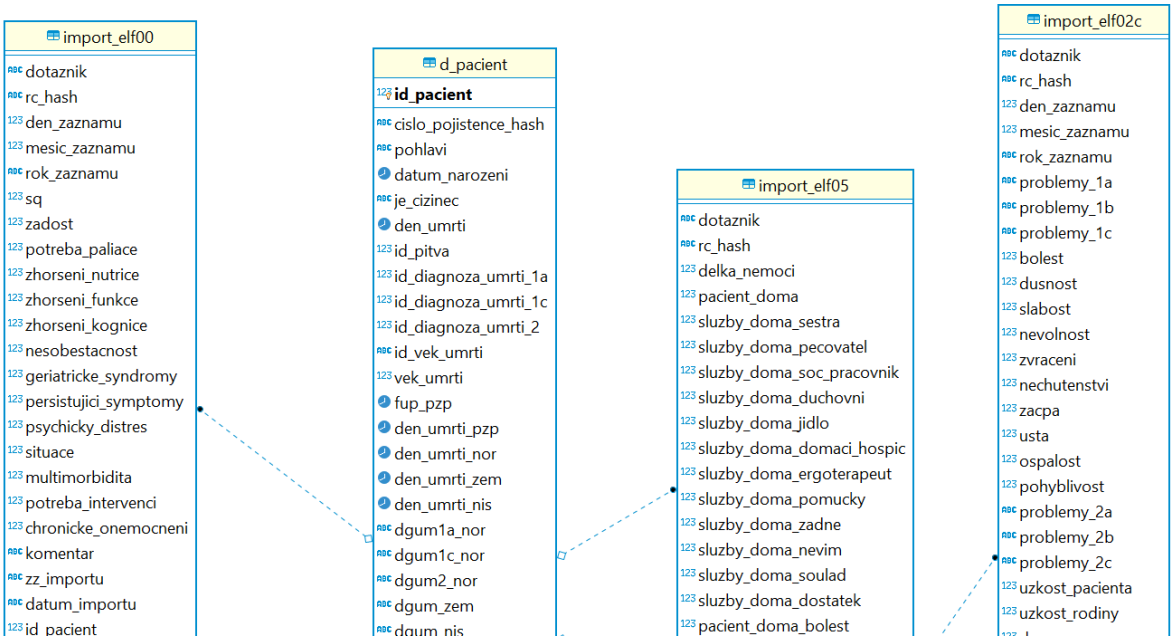
Název	Datový typ	Integritní omezení	Popis
id_sys	int	NN, PK, AI	Surrogate key
id_pacient	int	NN, FK -> pacient_nor (id_pacient)	Identifikátor pacienta
evcislo	varchar(14)		Evidenční číslo novotvaru
pocno	varchar(2)		Pořadové číslo novotvaru u stejného pacienta
stol	varchar(2)		První 2 cifry z roku narození pacienta
datnar	varchar(8)		Datum narození pacienta
rodcis_r	varchar(11)		Číslo pojištěnce
pohlav	varchar(1)		Pohlaví pacienta (1=muž, 2=žena, 9=neznámo)
psc	varchar(5)		PSČ bydliště pacienta
obce	varchar(6)		Statistický kód obce trvalého bydliště pacienta
obecpuv	varchar(6)		Statistický kód obce bydliště v době hlášení
...	...		*Řada dalších parametrů*
zz_importu	char(10)	NN	Identifikátor zařízení, ze kterého pocházejí primární data
datum_importu	datetime	NN	Čas zahájení importu primárních dat

zemreli, zemreli_anonym

Obsahuje záznamy ze seznamu pacientů nemocnice doplněné o případná data úmrtí a úmrtní diagnózy. V tabulce **ZEMRELI_ANONYM** je sloupec **RC** nahrazen sloupcem **RC_HASH**, který obsahuje šifru čísla pojištěnce.

Název	Datový typ	Integritní omezení	Popis
rc	varchar(10)	NN, PK, AI	Číslo pojištěnce
den_umrti	varchar(10)		Datum úmrtí pacienta
dg_umrti	varchar(5)		Hlavní příčina úmrtí pacienta
zz_importu	char(10)	NN	Identifikátor zařízení, ze kterého pocházejí primární data
datum_importu	datetime	NN	Čas zahájení importu primárních dat
id_pacient	int	NN, FK -> pacient_nor (id_pacient)	Identifikátor pacienta

Diagram dat z dotazníků ELFIS



Další komponenty Agenta Datového skladu ELFis

Pro správnou funkcionalitu Agenta jsou zapotřebí ještě následující komponenty:

- Skripty pro spouštění importu nových dat (shell a PHP skripty pro snadné spouštění importu)
- Převodník čísel pojištěnců (shell a 7zip pro bezpečnou archivaci převodníků čísel pojištěnců na šifry)
- Sync (PHP skripty a HTTPS rozhraní pro přesun de-identifikovaných dat do anonymizované databáze

3.2. Parametry vstupních dat

Agent a jeho komponenty využívají parametry pro zpracování rozdílných vstupních dat. Proto se spouští v konzoli (ve verzi pro operační systém Windows také s administrátorským oprávněním). Parametry rozdělujeme na interní, která jsou uložena v konfiguračním souboru a externí, která jsou požadována při spouštění v konzoli.

Interní

Parametry jsou uložena v konfiguračních souborech agent.bat a agent.ini . Konfigurační soubor typu bat sdružuje primárně systémová nastavení jako cesty k adresářové struktuře agenta, databáze, komprimačního nástroje a jiné. Konfigurační soubor typu .ini se zaměřuje na popis datového rozhraní primárních dat pro zpracování. Jedná se nejen o nastavení cest k datovým zdrojům, ale hlavně pro popis struktury vstupních dat. Pro soubory typu CSV například oddělovač, formát data, seznam předávaných atributů a další.

Tyto konfigurační soubory jsou uložena v adresáři etc a všechny komponenty agenta se odvolávají na data obsažena v nich.

Externí

Jedná se o parametry nutné ke spuštění komponent Agent, například o typ primárního zdroje dat (k-dávky, preskripce, nor…).

Doklady datového rozhraní VZP („k-dávky“)

Doklady odpovídají datovému rozhraní VZP pro individuální doklady, platné ke dni pořizování těchto dat – viz HTTPS://WWW.VZP.CZ/POSKYTOVATELE/VYUCTOVANI-ZDRAVOTNI-PECE/METODIKA-VYUCTOVANI-AKTUALNI-STAV. Obsah předávaných dat zahrnuje především následující soubory:

Hlavička dávky

Název	Popis
davka	ID dávky
Zp	pojišťovna
chardav	charakter dávky
icz	IČZ na dávce
rok	rok dávky
mesic	měsíc dávky

Ambulantní doklad

Název	Popis
doklad	ID dokladu
platce	kód plátce
icp_vykazujici	IČP
varsymbol	variabilní symbol
odbornost_dokladu	odbornost
diagnoza_hlavni	hlavní diagnóza
cp	ID pojištěnce
davka	ID dávky

Hospitalizační doklad

Název	Popis
doklad	ID dokladu
platce	kód plátce
icp_vykazujici	IČP
varsymbol	variabilní symbol
odbornost_dokladu	odbornost
diagnoza_hlavni	hlavní diagnóza
dophosp	doporučení k hospitalizaci
zpusukon	způsob ukončení hospitalizace
den_zahajeni	zahájení hospitalizace
den_ukonceni	ukončení hospitalizace
cp	ID pojištěnce
davka	ID dávky

Poukaz na vyšetření/ošetření

Název	Popis
doklad	ID dokladu
platce	kód plátce
icp_vykazujici	IČP
varsymbol	variabilní symbol
icp_zadajici	IČP žádající
odbornost_dokladu	odbornost
diagnoza_hlavni	hlavní diagnóza
cp	ID pojištěnce
davka	ID dávky

Zvlášť účtované léčivé prostředky, zdravotnický materiál

Název	Popis
doklad	ID dokladu
icp_vykazujici	IČP
varsymbol	variabilní symbol
odbornost_dokladu	odbornost
diagnoza_hlavni	hlavní diagnóza
cp	ID pojištěnce
davka	ID dávky

Provedené výkony

Název	Popis
doklad	ID dokladu
den_provedeni	datum provedení výkonu
vykon	kód výkonu
odbornost_radkova	odbornost
diagnoza_radkova	diagnóza
mnozstvi_vykony	množství provedení
body_vykony	body za výkony

Zvlášť účtované položky

Název	Popis
doklad	ID dokladu
nadrazeny_doklad	ID nadřazeného dokladu
den_podani	datum podání ZUP
skupina	skupina ZUP
zvlu	kód ZVLU
pripravek	kód položky
mnozstvi_pripravky	množství položky
hodnota_pripravky	Kč za položku
dg	diagnóza na položce

Vedlejší diagnózy dokladu

Název	Popis
doklad	ID dokladu
diagnoza_vedlejsi	vedlejší diagnóza

Ošetřovací kategorie pacienta

Název	Popis
doklad	ID dokladu
kategpac	ošetřovací kategorie pacienta
pocet_dnu	počet dnů v kategorii

Vydaný recept

Název	Popis
doklad	ID dokladu
den_podani	datum vydání léčiva

icp	IČP předepisujícího lékaře
skupina1	skupina prvního LP
pripravek1	kód prvního LP
mnozstvi_pripravky1	množství prvního LP
hodnota_pripravky1	Kč za první LP
dg1	diagnóza prvního LP
rpzu1	zvláštní úhrada za první LP
skupina2	skupina druhého LP
pripravek2	kód druhého LP
mnozstvi_pripravky2	množství druhého LP
hodnota_pripravky2	Kč za první LP
dg2	diagnóza druhého LP
rpzu2	zvláštní úhrada za druhý LP
cp	ID pojištěnce
davka	ID dávky

Záznamy nemocniční preskripce

V souborech, které popisují předepsaná léčiva nebo materiál pro pacienty nemocnice, mohou obsahovat následující parametry:

Název	Popis
Datum	datum předepsání léčiva
Poj	pojišťovna pacienta
Icp	IČP předepisujícího lékaře
sk1, sk2	skupina prvního / druhého předepsaného léčivého přípravku
kod1, kod2	kód prvního / druhého předepsaného léčivého přípravku
mnoz1, mnoz2	množství prvního / druhého předepsaného léčivého přípravku
cena1, cena2	úhrada za první / druhý předepsaný léčivý přípravek
atc1, atc2	ATC skupina prvního / druhého předepsaného léčivého přípravku
nazev1, nazev2	název prvního / druhého předepsaného léčivého přípravku
latka1, latka2	účinná látka prvního / druhého předepsaného léčivého přípravku
dg1, dg2	diagnóza prvního / druhého předepsaného léčivého přípravku
j1_*, j2_*	místo * může být libovolný řetězec alfanumerických znaků nebo _) – jakýkoliv další parametr, vztahující se k prvnímu / druhému předepsanému léčivému přípravku, nesmí obsahovat osobní údaje pacientů
cokoliv jiného	jakýkoliv další parametr, vztahující se k preskripci jako celku, nepřenáší se do DB, může obsahovat osobní údaje

Záznamy z Národního onkologického registru (NOR)

Data odpovídají datovému rozhraní NOR dle metodiky, platné ke dni exportu z registru. Rozhraní je součástí předaných dat a obsahuje pro všechny záznamy exportu stejnou datovou strukturu.

3.3.Funkce Agenta Datového skladu ELFis

Import dat ze zdravotnického zařízení

Stručný souhrn procesu

Provádí	Pověřený pracovník nemocnice s oprávněním zpracování dat obsahující osobní údaje
Frekvence	Při zpracování nových dat, obvykle jednou ročně
Popis	Funkce zpracovává a načítá primární data nemocnice z textových souborů do databáze a spouští proces jejich dalšího zpracování
Vstup	Vstupem jsou připravená nemocniční data v požadovaném formátu, jejichž popis je správně nastaven v konfiguračních parametrech nástroje.
Výstup	Výsledkem procesu jsou zpracovaná nemocniční data v DB, která obsahují pouze de-identifikovaná data bez osobních údajů. Všechna ostatní dočasná data, která mohou obsahovat osobní údaje (mimo vlastních vstupních souborů), jsou bezpečně a nevratně smazána.

Podrobný popis procesu

Uvedený postup je popsán na příkladu zpracování administrativních dat nemocnice určených pro pojišťovny (k-dávky, PZP). V ostatních případech je postup velmi podobný, často však výrazně jednodušší.

I. Agent se zeptá na uživatelské heslo do DB (MySQL)
II. Agent se zeptá na heslo k převodníku rodných čísel, které uloží do paměti.
III. Agent se zeptá, který z primárných datových zdrojů se bude zpracovávat:
    a. K-dávky a preskripce vykázané nemocnicí (PZP, PRESK)
    b. Data z NIS nemocnice s lokalitou bydliště (PSČ)
    c. Hlášení do Národního onkologického registru pacientů nemocnice (NOR)
    d. Dotazníky projektu ELFis (ELFIS) – viz kapitola 4.
IV. 7zip rozbalí z archivu pacient-hash.zip soubor pacient-hash.sql pomocí hesla z bodu II.
V. Načte se pacient-hash.sql do tabulky **ICOP_DW1_KOC_IMPORT.PACIENT_HASH**
VI. SDELETE bezpečně smaže pacient-hash.sql
VII. Pomocí PHP skriptu se vybere jeden z primárních datových zdrojů, které uživatel označil v bodu III
VIII. PHP skript projde primární zdrojová data a přetransformuje je do CSV souboru
IX. CSV soubor se nahraje do DB tabulky **ICOP_DW1_KOC_IMPORT.PZP_BULK**
X. Pustí se DB procedura **PUMPA_PZP** (resp. PRESK, PSC, NOR, ZEM v případě jiného typu primárního datového zdroje). Víc sekce **3.3.2. Zpracování primárních dat obsahující osobní údaje v DB**
XI. CSV soubor se smaže
XII. Pokud je další nezpracovaný primární zdroj, vrátí se Agent na bod VII, jinak pokračuje na bod XIII
XIII. Exportuje tabulku **ICOP_DW1_KOC_IMPORT.PACIENT_HASH** do souboru pacient_hash.sql
XIV. Pokud neexistuje soubor pacient-hash.zip (jedná se o prvotní import), přejde Agent na bod XVI
XV. Pokud soubor pacient-hash.zip existuje, zálohuje se do adresáře Archive (…)
XVI. 7zip zabalí soubor pacient-hash.sql do pacient-hash.zip
XVII. SDELETE bezpečně smaže soubor pacient-hash.sql

XVIII. V DB je smazán (truncate) obsah tabulky **ICOP_DW1_KOC_IMPORT.PACIENT_HASH**
XIX. Zastaví se DB
XX. SDELETE bezpečně smaže žurnál DB
XXI. Spustí se DB
XXII. Agent smaže všechny dočasné soubory a vypíše do konzoly závěrečné informace

Zpracování primárních dat obsahující osobní údaje v DB

Stručný souhrn procesu

Provádí	Pověřený pracovník nemocnice s oprávněním zpracování dat obsahující osobní údaje
Frekvence	Při zpracování nových dat, obvykle jednou ročně; spouští se automaticky v rámci procesu „Import dat ze zdravotnického zařízení“
Popis	Funkce zpracovává primární data nemocnice v rámci databáze, validuje jejich obsah, odvozuje další parametry a spouští proces jejich de-identifikace
Vstup	Vstupem jsou nemocniční data nahraná v základním tvaru do databáze
Výstup	Výsledkem procesu jsou zpracovaná nemocniční data v DB, která obsahují pouze de-identifikovaná data bez osobních údajů. Mezivýsledky v DB, které obsahují osobní údaje, jsou smazány. Jsou provedeny validační kontroly vstupních dat.

Podrobný popis procesu

Uvedený postup je popsán na příkladu zpracování administrativních dat nemocnice určených pro pojišťovny (k-dávky, PZP). V ostatních případech je postup velmi podobný, často však výrazně jednodušší.

I. Smaže se (truncate) obsah tabulek v DB **ICOP_DW1_KOC_IMPORT** primárních datových zdrojů
II. Obsah tabulky **ICOP_DW1_KOC_IMPORT.PZP_BULK** se po měsíčních obdobích překopíruje do tabulky **ICOP_DW1_KOC_IMPORT.PZP**
III. Podle slovníku typů dokladů a vět jsou označeny záznamy se známou strukturou – procedura **CHECK_UNKNOWN_PZP_ROWS()**
IV. Do převodní tabulky **ICOP_DW1_KOC_IMPORT.TMP_PZP_BID** vloží ke každému nalezenému číslu pojištěnce (**BID**) odkaz na řádek v tabulce **ICOP_DW1_KOC_IMPORT.PZP**, ve které se toto číslo nachází
V. Pomocí procedury **CREATE_PACIENT_PZP()** se do tabulky **ICOP_DW1_KOC_IMPORT.PACIENT_PZP** doplní ze všech nalezených čísel pojištěnců tato data:
    a. Číslo pojištěnce
    b. Pohlaví
    c. Je_cizinec (příznak zda je pacient cizinec – podle RČ)
    d. Id_pacienta (automaticky doplněno)
VI. Doplnění tabulky **ICOP_DW1_KOC_IMPORT.PZP** o **ID_PACIENTA** z tabulky **ICOP_DW1_KOC_IMPORT.PACIENT_PZP**
VII. Ověření hash hesla k převodníku rodných čísel (zadaného na začátku procesu 3.2.2.1. bod II
    a. Heslo zadané na začátku procesu se zahashuje pomocí funkce **MD5()** a následně se porovná s obsahem tabulky **ICOP_DW1_KOC_IMPORT.PSW**
    b. Pokud je tabulka **ICOP_DW1_KOC_IMPORT.PSW** prázdná, vloží do ní záznam s hashem hesla (první spuštění agenta)
    c. Pokud je tabulka **ICOP_DW1_KOC_IMPORT.PSW** naplněná:
        i. Pokud se hashe neshodují, ukončí program s chybovou hláškou o nesprávném heslu k převodníku rodných čísel

ii. Pokud se hashe shodují, pokračuje dál

VIII. Vytvoří tabulku **ICOP_DW1_KOC_IMPORT.PZP_ANONYM** jako kopii tabulky PZP ale jenom těch řádků dokladů, které mají příznak známého typu věty (známé typy vět ověřené procedurou **CHECK_UNKNOWN_PZP_ROWS()** – viz bod III)

IX. Provede se de-identifikace všech osobních údajů v primárních datech - procedury **HASH_PATIENTS_PZP()** a **ANONYMIZE_PATIENTS_PZP()**. Procedury jsou popsány v části 4.2.3. De-identifikace osobních údajů

X. Do převodní tabulky **ICOP_DW1_KOC_IMPORT.PACIENT_HASH** se přidají nově nalezené čísla pojištěnců a jejich hashe

XI. Do anonymizovaného schématu (**ICOP_DW1_KOC_IMPORT_ANONYM**) se naplní obsah tabulek **PZP**, **PACIENT_PZP** a **TMP_PZP_BID** z původního schématu (**ICOP_DW1_KOC_IMPORT**), již bez čísel pojištěnců. Primárním identifikátorem pacienta se stane **RODNE_CISLO_HASH**

XII. Provedou se testy zpracování vstupních dat

XIII. Smaže se obsah tabulek obsahující osobní údaje ve schématu **ICOP_DW1_KOC_IMPORT**

De-identifikace osobních údajů

Stručný souhrn procesu

Provádí	Pověřený pracovník nemocnice s oprávněním zpracování dat obsahující osobní údaje
Frekvence	Při zpracování nově příchozích dat, obvykle jednou ročně; spouští se automaticky v rámci procesu „Zpracování primárních dat obsahující osobní údaje v DB“
Popis	Funkce provádí náhradu atributů, obsahujících osobní údaje, za jejich de-identifikované alternativy. Čísla pojištěnců jsou nahrazena bezvýznamovými identifikátory (jednosměrná šifra), další osobní údaje jsou smazány
Vstup	Vstupem jsou nemocniční data obsahující osobní údaje v DB, které jsou v nich dohledány a označeny.
Výstup	Výsledkem procesu jsou data, kde jsou osobní údaje pacientů de-identifikovány – čísla pojištěnců jsou nahrazeny bezvýznamovými identifikátory, ostatní osobní údaje jsou nevratně odstraněny.

Podrobný popis procesu

Uvedené příklady opět demonstrují funkčnost při zpracování administrativních dat nemocnice (k-dávky, PZP). Ostatní datové zdroje jsou zpracovávány analogicky.

De-identifikace čísel pojištěnců v seznamu pacientů probíhá pomocí DB procedury **HASH_PATIENTS_PZP().** Následující tabulka zobrazuje způsob, jakým se data s osobními údaji pacientů mapují na de-identifikované záznamy**:**

Obrázek 2 Mapování provádějící de-identifikaci čísel pojištěnců v seznamu pacientů na dokladech PZP

pacient_pzp	mapping	pacient_pzp_anonym
cislo_poj	Sha1(cislo_poj + salt)	cislo_poj_hash
datum_narozeni	⇨	datum_narozeni
pohlavi	⇨	pohlavi
je_cizinec	⇨	je_cizinec
id_pacient	⇨	id_pacient

Procedura **ANONYMIZE_PATIENTS_PZP()** provádí náhradu čísel pojištěnců ve vlastních řádcích tabulky **ICOP_DW1_KOC_IMPORT.PZP_ANONYM** – v tomto případě je nahradí za řetězec deseti znaků „#“. Mapování původních datových atributů tabulky **ICOP_DW1_KOC_IMPORT.PZP** na atributy v tabulce **ICOP_DW1_KOC_IMPORT.PZP_ANONYM** jsou popsány v následující tabulce.

Obrázek 3 Mapování atributů dokladů PZP na de-identifikované záznamy

pzp	mapping	pzp_anonym
id_sys	→	id_sys
radka	Replace(find_cp_in_pzp_dict(radka), “##########”)	radka
kod_vety	→	kod_vety
delka_vety	→	delka_vety
kod_dokladu	→	kod_dokladu
davka	→	davka
doklad	→	doklad
doklad_subs	→	doklad_subs
nadrazeny_doklad	→	nadrazeny_doklad
id_pacient	→	id_pacient
zz_importu	→	zz_importu
datum_importu	→	datum_importu
flag	→	flag
rok_uzavreni	→	rok_uzavreni
mesic_uzavreni	→	mesic_uzavreni

Vlastní asociace řádků dokladů na pacienty v tabulce **ICOP_DW1_KOC_IMPORT.TMP_PZP_BID** jsou nahrazeny hashem čísla pojištěnce, získaného procedurou **HASH_PATIENTS_PZP()**.

Obrázek 4 Mapování atributů asociační tabulky s čísly pojištěnců a řádky dokladů do de-identifikované podoby

tmp_pzp_bid	mapping	tmp_pzp_bid_anonym
bid	Sha1(cislo_poj + salt)	bid
id_sys	→	id_sys

V dalších tabulkách jsou uvedeny mapování dalších typů primárních datových zdrojů. Je v nich patrný způsob nakládání s čísly pojištěnců, resp. s dalšími osobními údaji, které se zde mohou vyskytovat.

Obrázek 5 Mapování atributů nemocničních preskripcí na de-identifikované záznamy

presk	mapping	presk_anonym
inscomp	→	inscomp
datrece	→	datrece
cislo_poj	Sha1(cislo_poj + salt)	cislo_poj_hash
drug	→	drug
quantity	→	quantity
czicz	→	czicz
sk	→	sk
kc	→	kc
atc	→	atc
nazevatc	→	nazevatc
nazev	→	nazev
dg	→	dg
jine	→	jine
zz_importu	→	zz_importu
datum_importu	→	datum_importu
id_pacient	→	id_pacient

Obrázek 6 Mapování atributů záznamů o pacientech nemocnice na de-identifikované záznamy

psc	mapping	psc_anonym
cislo_poj	Sha1(cislo_poj + salt)	cislo_poj_hash
psc	→	psc
poj	→	poj
ulicecislo	✖	*nepřenáší se*
ulice	✖	*nepřenáší se*
cp	✖	*nepřenáší se*
co	✖	*nepřenáší se*
obec	→	obec
jmeno	✖	*nepřenáší se*
krestni	✖	*nepřenáší se*
prijmeni	✖	*nepřenáší se*
inicialy	→	inicialy
jine	→	jine
zz_importu	→	zz_importu
datum_importu	→	datum_importu
id_pacient	→	id_pacient

3.4.Správa systému a přístupů

Za správu systému, správu přístupů, zajištění bezpečnosti a zálohování zodpovídá vždy pověřený pracovník nemocnice, na jejichž prostředcích je Agent provozován. Pracovníci dodavatele s přístupy do nemocnic jsou povinni dodržovat veškeré zásady, požadované zodpovědnými provozovateli IT v nemocnici.

4. Zpracování dat pacientů z dotazníkového šetření ELFis

Na základě dat, jejichž zpracování v Agentu je popsáno v bodu 3, je provedena analýza dat z dotazníkového šetření pacientů v terminálním stavu ELFis. Analyzovány jsou jak samostatné údaje z dotazníků, tak jejich návaznost na data o poskytnutých zdravotních službách v dané nemocnici.

Sledují se například údaje o hospitalizační péči (akutní / následná lůžka), intenzivní a resuscitační péče a DUPV, operativa, dialýza, farmakoterapie atd. Údaje o poskytnuté péči jsou přiřazeny k záznamům z dotazníku na základě shodného anonymizovaného identifikátoru pacienta.

Typickým výstupem je např. přehled hospitalizačních případů pacientů v posledních měsících před úmrtím s následující strukturou:

pacient	Anonymní identifikátor pacienta
datum_pri	Datum přijetí k hospitalizaci
datum_pro	Datum ukončení hospitalizace
odb_pri	Přijímající odbornost
odb_pro	Propouštěcí odbornost
doporuceni	Doporučení k hospitalizaci
dg_hlavni	Hlavníé diagnóza hospitalizace
od_stan	Počet OD standardní akutní lůžkové péče
od_int	Počet OD intenzivní akutní lůžkové péče
od_nasl	Počet OD následné lůžkové péče
od_int_aro	Počet OD resuscitační lůžkové péče
od_int_jip	Počet OD intenzivní lůžkové péče nižšího typu
upv_hodin	Počet hodin strávených na UPV
dialyza_pocet	Počet výkonů provedení dialýzy

4.1. Údaje z dotazníků ELFIS

Data z dotazníků ELFIS jsou k dispozici v následující podobě:

NECPAL

Obsahuje údaje z dotazníku NECPAL.

dotaznik	Kód dotazníku
rc_hash	Anonymní identifikátor pacienta
den_zaznamu	Den pořízení záznamu
mesic_zaznamu	Měsíc pořízení záznamu
rok_zaznamu	Rok pořízení záznamu
sq	Surprise question
zadost	Žádost nebo potřeba
potreba_paliace	Potřeba paliativní péče
zhorseni_nutrice	Zhoršení nutričního stavu
zhorseni_funkce	Zhoršení funkčního stavu
zhorseni_kognice	Zhoršení kognitivního stavu
nesobestacnost	Výrazná závislost/nesoběstačnost
geriatricke_syndromy	Geristrické syndromy
persistujici_symptomy	Persistující syndromy

psychicky_distres	Psychosociální aspekty - distres
situace	Psychosociální aspekty – sociální
multimorbidita	Multimorbidita
potreba_intervenci	Potřeba a využívání zdrojů
chronicke_onemocneni	Specifické indikátory závažnosti onemocnění
komentar	Další komentář

IPOS

Obsahuje údaje z dotazníku IPOS pro pacienty nebo personál.

dotaznik	Kód dotazníku
rc_hash	Anonymní identifikátor pacienta
den_zaznamu	Den pořízení záznamu
mesic_zaznamu	Měsíc pořízení záznamu
rok_zaznamu	Rok pořízení záznamu
problemy_1a	Hlavní problémy 1a
problemy_1b	Hlavní problémy 1b
problemy_1c	Hlavní problémy 1c
bolest	Bolest
dusnost	Dušnost
slabost	Slabost či nedostatek energie
nevolnost	Nevolnost (pocit na zvracení)
zvraceni	Zvracení
nechutenstvi	Nechutenství
zacpa	Zácpa
usta	Bolest či sucho v ústech
ospalost	Ospalost
pohyblivost	Snížená pohyblivost
problemy_2a	Jiné obtíže 2a
problemy_2b	Jiné obtíže 2b
problemy_2c	Jiné obtíže 2c
uzkost_pacienta	Cítil jste úzkost z nemoci nebo léčby
uzkost_rodiny	Úzkost někoho z rodiny z nemoci nebo léčby
deprese	Cítil jste se depresivně?
vnitrni_klid	Pociťujete vnitřní klid?
rozhovor_s_rodinou	Byli jste schopni hovořit se svou rodinou?
dostatek_informaci	Dostáváte tolik informací, kolik si přejete?
prakticke_problemy	Byly řešeny praktické problény?
vyplneni_dotazniku	Jak jste vyplnil tento dotazník?
komentar	Další komentář

4.2. Údaje o poskytnutých zdravotních službách pacientů z ELFis

Pro standardizací analýz byly navrženy následující výstupy o poskytnutých zdravotních službách:

PACIENT

Obsahuje základní charakteristiku pacientů ve výběru.

rc_hash	Anonymní identifikátor pacienta
pohlavi	Pohlaví pacienta
datum_narozeni	Datum narození
den_umrti	Datum úmrtí
psc	PSČ bydliště pacienta

DOKLAD01

Obsahuje údaje o péči provedené při ambulantním vyšetření.

rc_hash	Anonymní identifikátor pacienta
kod_odbornost	Odbornost lékaře
kod_diagnoza	Základní diagnóza vyšetření
vdg1	Vedlejší diagnózy
vdg2	Vedlejší diagnózy
vdg3	Vedlejší diagnózy
vdg4	Vedlejší diagnózy
dg_radkova	Diagnóza výkonu
kod_vykon	Kód výkonu
mnozstvi_vykony	Počet provedení výkonu
body_vykony	Body za výkon
pma	Kč za výkon

DOKLAD02

Obsahuje údaje o péči provedené při hospitalizaci.

rc_hash	Anonymní identifikátor pacienta
id_sys_hpripad	Identifikátor hospitalizačního případu
kod_odbornost	Odbornost lůžkového oddělení
kod_diagnoza	Základní diagnóza hospitalizace
vdg1	Vedlejší diagnózy
vdg2	Vedlejší diagnózy
vdg3	Vedlejší diagnózy
vdg4	Vedlejší diagnózy
den_zahajeni	Den zahájení hospitalizace
den_ukonceni	Den ukončení hospitalizace
odbornost_radkova	Odbornost provedení výkonu
kod_vykon	Kód výkonu
mnozstvi_vykony	Počet provedení výkonu
body_vykony	Body za výkon
pma	Kč za výkon

DOKLAD03

Seznam podaných zvlášť účtovaných léčiv, léčivých přípravků a zdravotnického materiálu.

rc_hash	Anonymní identifikátor pacienta
kod_odbornost	Odbornost podávajícího lékaře
kod_diagnoza	Základní diagnóza léčby

skupina	Skupina zvlášť účtovaných položek (HVLP/IPLP/ZP/STOM)
kod_pripravek	Kód přípravku
kod_atc_skupina	ATC skupina léčiv
mnozstvi_pripravky	Množství podaného přípravku/materiálu
hodnota_pripravky	Kč za aplikovaný přípravek/materiál

DOKLAD06

Obsahuje údaje o vyžádané péči.

rc_hash	Anonymní identifikátor pacienta
kod_odbornost	Odbornost lékaře
kod_diagnoza	Základní diagnóza vyšetření
vdg1	Vedlejší diagnózy
vdg2	Vedlejší diagnózy
vdg3	Vedlejší diagnózy
vdg4	Vedlejší diagnózy
odbornost_zadajici	Odbornost žádajícího lékaře / oddělení
Id_sys_hpripad	Identifikátor hospitalizačního případu, pokud byla péče provedena za hospitalizace
kod_vykon	Kód výkonu
mnozstvi_vykony	Počet provedení výkonu
body_vykony	Body za výkon
pma	Kč za výkon

DOKLAD10

Seznam léčiv a léčivých přípravků, vydaných nebo předepsaných na recept.

rc_hash	Anonymní identifikátor pacienta
kod_odbornost	Odbornost předepisujícího lékaře
kod_diagnoza	Základní diagnóza léčby
skupina	Skupina zvlášť účtovaných položek (HVLP/IPLP/ZP/STOM)
kod_pripravek	Kód přípravku
kod_atc_skupina	ATC skupina léčiv
mnozstvi_pripravky	Množství předepsaného přípravku
hodnota_pripravky	Kč za předepsaný přípravek

5. Zajištění bezpečnosti osobních dat v nemocnicích

Pro zajištění bezpečnosti a ochrany osobních a jiných citlivých dat je aplikována celá řada opatření – od smluvních opatření, přes technická až po organizační pravidla. Dohromady vytvářejí velmi robustní systém ochrany nemocničních dat, který zcela vylučuje možnost jejich zneužití nebo úniku (viz kapitola 1).

5.1. Technická opatřeni

Přístupy ke všem datům chráněny heslem

DB MySQL rozdělena na neanonymní a anonymmní část, pro které existují dva rozdílné uživatelské účty. Zpracování vstupních dat s osobními údaji se standardně spouští pod uživatelem root (pracovníci nemocnice), který má přístup ke všem datům. Při vytváření výstupních dat a přenosu centrálního úložiště anonymizovaných dat se používa účet icop_admin, který má přístup jenom do anonymizované části.

Archivována data jsou zašifrovaná s heslem. Přihlášení k serveru a přístup k adresářům je řízen správou uživatelských účtů operačního systému dle politiky dané nemocnice.

Nahrazení čísel pojištěnců bezpečnou šifrou

Šifrování čísel pojištěnců je prováděno následujícím postupem.

```
BEGIN

    if (heslo is not null) then
        update pacient_pzp
        set rodne_cislo_hash = sha1(concat(trim(rodne_cislo), '#', heslo));
    end if;

end
```

Čísla pojištěnců, ve spojení s heslem (sůl), jsou šifrována pomocí jednosměrné šifrovací funkce SHA-1 (160-bit) na bezvýznamový identifikátor. Pro možnost zpětného dohledání čísel pojištěnců pro interní potřeby nemocnice je uchováván převodník mezi číslem pojištěnce a jeho šifrou. Způsob práce s tímto převodníkem je popsán v části 3.4.2.4.

Šifrování přenosu dat mezi Agentem a centrálním úložištěm anonymizovaných dat

Při přenosu de-identifikovaných dat z nemocnice anonymizovanou databázi pro analýzy se používá zabezpečené spojení pomocí protokolu HTTPS s certifikátem, spravovaným dodavatelem. Pro komunikaci jsou povoleny pouze spojení z předem povolených IP adres jednotlivých serverů ve spolupracujících nemocnicích.

5.2. Organizační opatření

Smluvní ochrana osobních dat s nemocnicí

Mezi každou nemocnicí a MU je uzavřena smlouva, jejíž nedílnou součástí je dohoda o ochraně citlivých údajů nemocnice pracovníky zpracovatele a povinnost jejich mlčenlivosti.

Dohoda o mlčenlivosti zaměstnanců

Každý zaměstnanec LF MU má podepsanou dohodu o mlčenlivosti a ochraně osobních a jiných citlivých údajů, se kterými přichází při práci do kontaktu.

Bezpečná evidence přístupů do nemocnic

Evidence přístupů do nemocnic je popsána v části 2.2.

Oddělení přístupů, řízení rolí

Zpracování dat, obsahujících osobní údaje pacientů, probíhá v databázi, která je přístupná pouze pod uživatelským účtem administrátora DB (obvykle root). Zodpovědností pověřeného pracovníka nemocnice je zpřístupnění vstupních dat a jejich zpracování do podoby de-identifikovaných záznamů, které jsou již v oddělené části databáze, přístupné členům týmu Datového skladu ELFis. S uživatelským účtem icop_admin, který mají členové týmu Datového skladu ELFis k dispozici, není možné se k žádným datům s osobními údaji dostat.

Mazání a šifrování osobních dat v době, kdy nejsou třeba

Všechna data obsahující citlivé informace jsou uložena v databázi pouze po dobu nezbytně nutnou pro provedení požadovaných funkcí.

Na začátku procesu zpracování primárních dat (K-dávky, preskripce, seznam pacientů s lokalitou bydliště, atd.) se rozbalí archív převodníku čísel pojištěnců a načte do DB. Rozbalený soubor se okamžitě smaže bezpečným způsobem pomocí programu SDelete. Při procesu importu nových nemocničních dat jsou do převodníku v DB přidáni noví pacienti. Po skončení importu se soubor archívu převodníku zálohuje (přejmenuje podle aktuálního data a přesune do archívu). Následně se z DB exportuje aktualizovaný převodník do souboru, který je následně zabalen jako nový šifrovaný archív převodníku. Vstupní soubor převodníku i tabulka z databáze jsou následně bezpečně smazány.

Minimalizace práce s osobními údaji

S daty, která obsahují osobní údaje pacientů (zejména číslo pojištěnce) se manipuluje co nejméně je to možné. Primární data jsou pouze základně zpracována (načtena jejich struktura a obsah) a uložena do DB. Ihned poté jsou čísla pojištěnců zašifrována na bezvýznamový identifikátor a tímto novým identifikátorem jsou původní čísla pojištěnců nahrazena. Již jen nové identifikátory jsou pak spolu s vlastními daty nahrány do oddělené části DB, kde probíhá další zpracování. Veškerá další práce již probíhá nad de-identifikovanými daty, bez osobních údajů – tyto jsou nevratně smazány.

Vyřazení dat, která mohou obsahovat osobní údaje, z dalšího zpracování

Záznamy, které mohou obsahovat osobní údaje (jiné než číslo pojištěnce, které se bezpečně šifruje), se z dalšího zpracování vyřazují. Jedná se primárně o zpracování seznamu pacientů s lokalitou bydliště, kde je možné očekávat výskyt osobních údajů (jméno a příjmení, přesná adresa apod.). Do de-identifikované části DB se tyto údaje nepřenáší.

Při zpracování administrativních dat nemocnice (PZP) se každý doklad ověřuje přes slovník (seznam známých typů dokladů). Pokud je daný doklad ve slovníku nalezen, je označen pro další zpracování. V další části se pak zpracovávají jen doklady s tímto příznakem (jenom rozpoznané typy dokladů). Ostatní doklady jsou smazány. Tímto se zabezpečí ochrana neznámých dokladů, které by mohly obsahovat osobní údaje.

Výhradní použití de-identifikovaných dat pro analýzy

Agent pro přesun dat mezi nemocnicí a databází pro analýzy využívá DB účet icop_admin. Tento účet je omezen přístupem jenom do anonymizované DB. Z tohoto důvodu nehrozí únik citlivých osobních údajů mimo nemocnici, natož pak jejich použití při analýzách.