Datová a informační základna hodnocení objemu a kvality péče v nemocnicích Jihomoravského kraje

Metodická dokumentace projektu: architektura systému I-COP – integrace datových zdrojů – implementace datových skladů

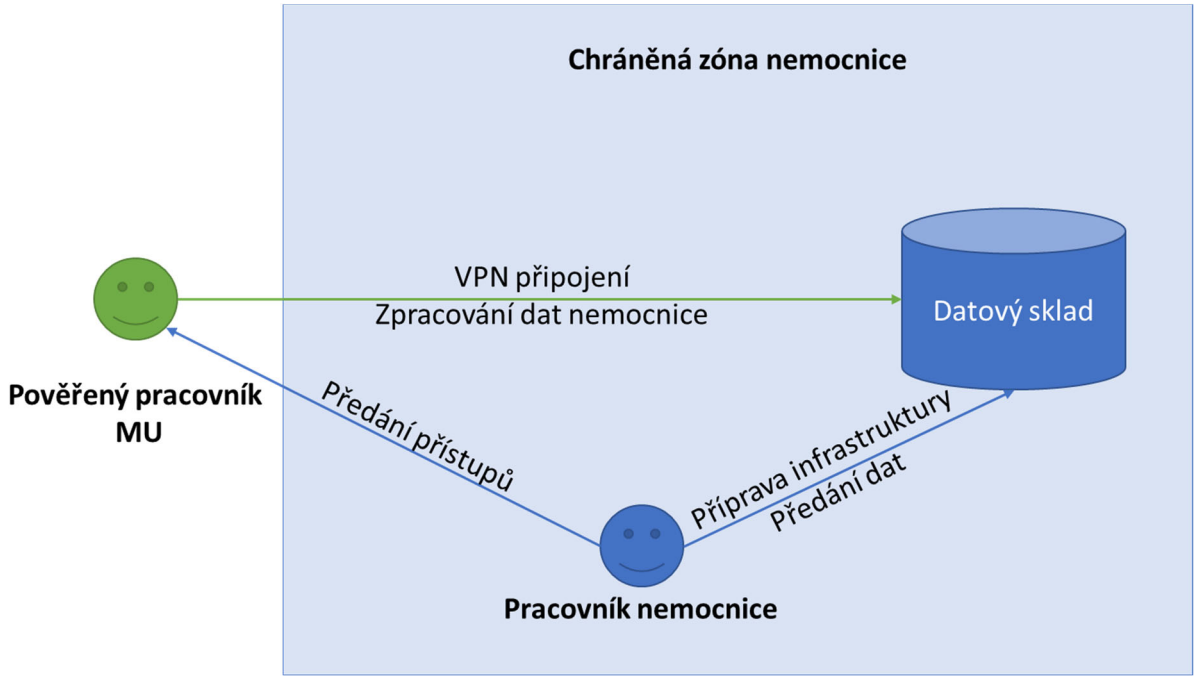
Obsah

1. Obecná architektura systému I-COP

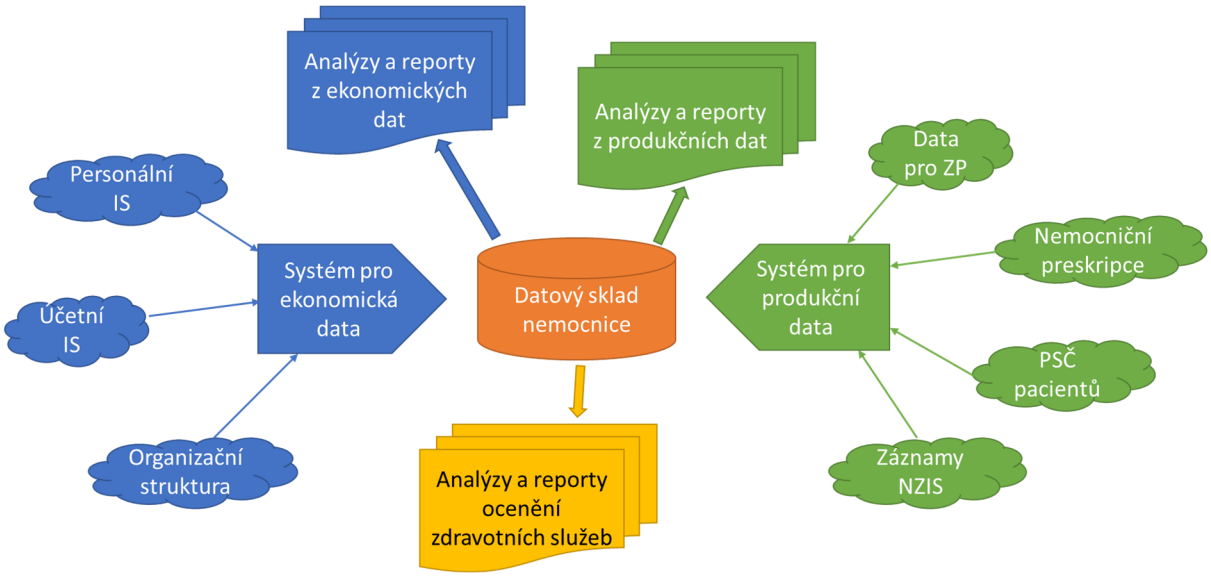
1.1. Datový sklad I-COP

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 příprava požadovaných ekonomických a provozních dat. Zde je nutná součinnost s pracovníky příslušných oddělení nemocnice (ekonomické a personální) pro zajištění výstupů z těchto informačních systémů. Dále je nutná tvorba číselníku organizační struktury s typologií jednotlivých pracovišť a účetní osnovy s klasifikací účtů dle přiložených číselníků. S touto fází se předpokládá významná spolupráce s pracovníky dodavatele, kteří zajistí metodické podklady, školení a případně přímou podporu při přípravě výstupů.

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 se předpokládá příprava analýz a výstupů ze zpracovaných dat. Výstupy jsou ve formě dokumentů (analýzy vybraných oblastí), přímý přístup ke zpracovaným datům v datovém skladu nemocnice nebo pomocí nástroje pro jejich prohlížení (business intelligence nástroj).

1.2.Princip fungování sítě I-COP a nakládání s daty v projektu

Architektura systému I-COP je postavena na síti zapojených nemocnic Jihomoravského kraje. 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 I-COP, 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 I-COP 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 I-COP 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.

1.3.1. 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 I-COP 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 1011112310128617150l701     11344H680            0              0.00    181
V03022005710221          152
V03022005730171           29
GH730
A 43823000 1111112310116013150l301     11345J450            0              0.00     45
V15022005095131           45
A 42112600 1211112310116013150l301     11346J459            0              0.00     45
V10022005095131           45
A 44381400 1311112310128617150l701     11346J352            0              0.00    266
V23022005710221          152
V23022005713171          114
A 41344100 1411112310132814450l405     11347B07             0              0.00    138
V07022005450221          138
A 42811900 1511112310123515250l502     11348R100            0              0.00    203
V03022005520221          203
A 43546700 1611112310113512950l209     11349R51             0              0.00    847
V11022005291231          625
V11022005291251          222
A 41966500 1711112310123515250l502     11350S823            0              0.00    119
V01022005520231          119
A 44131300 1811112310128617150l701     11350H650            0              0.00    189
V20022005710221          152
```

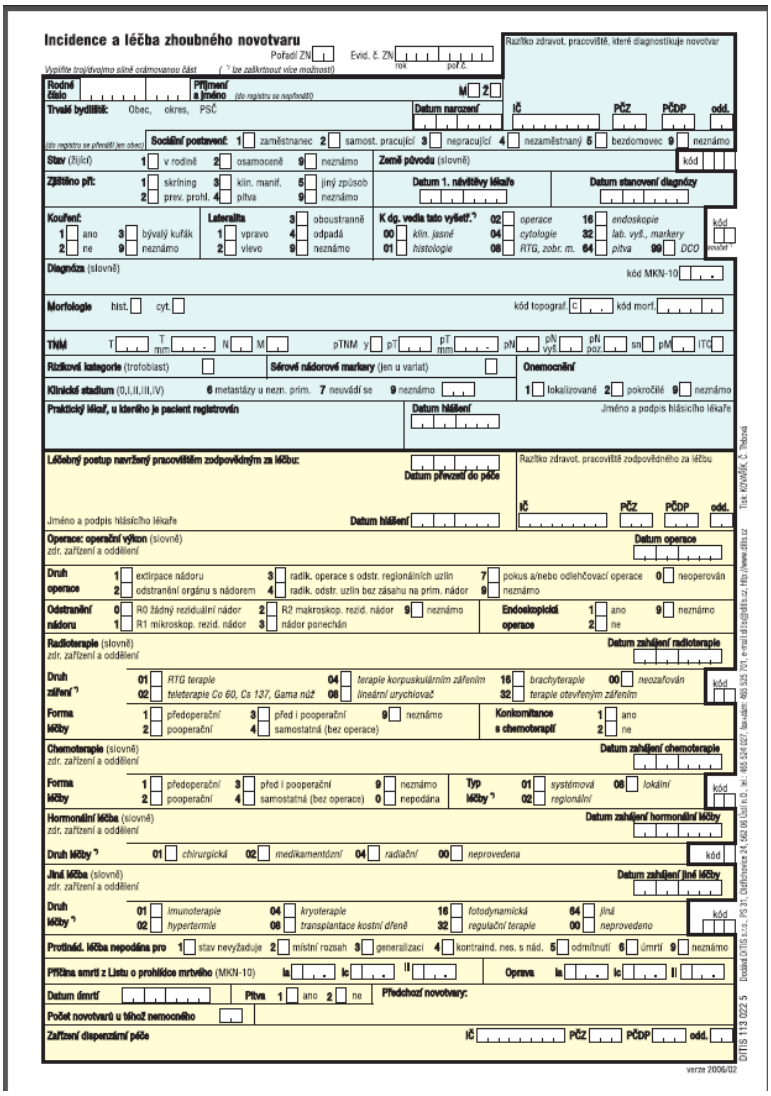
Struktura interně analyzovaných administrativních dat nemocnice je hierarchická. Na nejvyšší úrovni je tzv. hlavička dávky, která popisuje nemocnici a období, za která jsou data předávána. Pod ní jsou evidovány jednotlivé doklady – výkazy o formě poskytnuté péče pacientovi. Základními doklady jsou 01 – 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.

1.3.2. 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.

Napojení dat NOR je metodickým příkladem využití centrálních dat a populačních statistik pro srovnávací analýzy uvnitř nemocnic.

1.3.3. 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).

1.4.Uživatelé a další subjekty v projektu I-COP a jejich role

Systém I-COP a jeho výstupy používají následující skupiny uživatelů:

1.4.1. I-COP tým

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.

1.4.2. Analytický tým I-COP

Pracovníci LF MU, kteří mají přístup k předaným datům zapojených I-COP center a zodpovídají za provádění analytických výstupů z předaných dat. Data jsou jim předávána I-COP týmem 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 I-COP. Předávaná data jsou vždy nevratně anonymizovaná, bez jakýchkoliv osobních údajů o pacientech.

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

1.4.3. Pověřený IT pracovník nemocnice

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

1.4.4. Vedoucí management a odborní garanti I-COP center

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

1.4.5. 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 I-COP.

Na straně zapojených nemocnic jsou definovány týmy provozovatelů systému (IT support), kteří zajišťují přístupy pro členy I-COP týmu, zajišťují výpočetní prostředí pro provoz I-COP Agent, 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 I-COP Agent 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. I-COP Agent

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

3.1. Architektura I-COP Agent

I-COP 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.

3.1.1. Databáze

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

3.1.1.1. Schémata

Pro funkcionalitu I-COP Agenta jsou potřeba v databázi 3 databázová schémata:

3.1.1.1.1. 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.

3.1.1.1.2. icop_dw1_koc_import_anonym

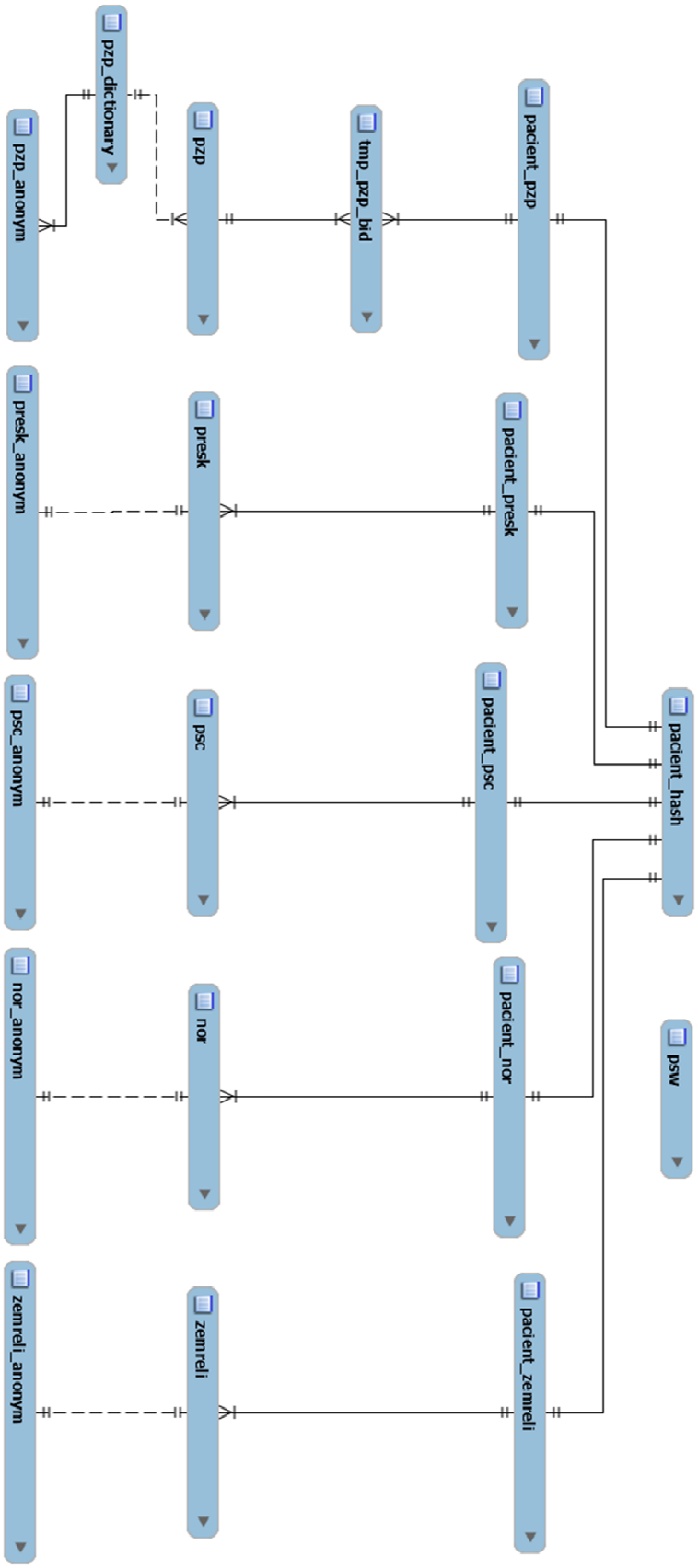
Slouží pro uložení výsledku iniciálního zpracování nemocničních dat, která již neobsahují žádné osobní údaje pacientů. Jsou přístupná I-COP týmu.

3.1.1.1.3. 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 I-COP týmu.

3.1.1.2. ERD schématu icop_dw1_koc_import

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



3.1.1.3. Popis datového modelu

3.1.1.3.1. 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 I-COP 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

3.1.1.3.2. 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

3.1.1.3.3. 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

3.1.1.3.4. 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

3.1.1.3.5. 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

3.1.1.3.6. 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

3.1.1.3.7. 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

3.1.1.3.8. 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

3.1.1.3.9. 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

3.1.1.3.10. 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

3.1.2. Další komponenty I-COP Agent

Pro správnou funkcionalitu I-COP Agent 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)
- I-COP Sync (PHP skripty a HTTPS rozhraní pro přesun de-identifikovaných dat do anonymizované databáze

3.2.Parametry vstupních dat

I-COP 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 I-COP Agent, například o typ primárního zdroje dat (k-dávky, preskripce, nor…).

3.2.1. 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. Jejich struktura a obsah je tam přesně popsán.

3.2.2. 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

3.2.3. 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 I-COP Agenta

3.3.1. Import dat ze zdravotnického zařízení

3.3.1.1. 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.

3.3.1.2. 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. I-COP agent se zeptá na uživatelské heslo do DB (MySQL)
II. I-COP agent se zeptá na heslo k převodníku rodných čísel, které uloží do paměti.
III. I-COP 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)
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 I-COP 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 I-COP 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. I-COP agent smaže všechny dočasné soubory a vypíše do konzoly závěrečné informace

3.3.2. Zpracování primárních dat obsahující osobní údaje v DB

3.3.2.1. 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.

3.3.2.2. 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**

3.3.3. De-identifikace osobních údajů

3.3.3.1. 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.

3.3.3.2. 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 I-COP Agent provozován. Pracovníci dodavatele s přístupy do nemocnic jsou povinni dodržovat veškeré zásady, požadované zodpovědnými provozovateli IT v nemocnici.

3.5. 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).

3.5.1. Technická opatřeni

3.5.1.1. 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 na I-COP Central 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.

3.5.1.2. Nahrazení čísel pojištěnců bezpečnou šifrou

Šifrování čísel pojištěnců je prováděno následujícím postupem.

```
1  BEGIN
2
3      if (heslo is not null) then
4          update pacient_pzp
5          set rodne_cislo_hash = sha1(concat(trim(rodne_cislo), '#', heslo));
6      end if;
7
8  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.

3.5.1.3. Šifrování přenosu dat mezi I-COP Agent a I-COP Central

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.

3.5.2. Organizační opatření

3.5.2.1. 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.

3.5.2.2. 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.

3.5.2.3. Bezpečná evidence přístupů do nemocnic

Evidence přístupů do nemocnic je popsána v části 2.2.

3.5.2.4. 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 I-COP týmu. S uživatelským účtem icop_admin, který mají členové I-COP týmu k dispozici, není možné se k žádným datům s osobními údaji dostat.

3.5.2.5. 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.

3.5.2.6. 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.

3.5.2.7. Vyřazení dat, která mohou obsahovat osobni 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.

3.5.2.8. Výhradní použití de-identifikovaných dat pro analýzy

I-COP 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.