

8. 9. 2026



WORKSHOP: ZDRAVOTNICKÁ DATA V PRAXI

SYSTÉM SYNTETICKÝCH DAT

Lucie Čačíková, Daniel Klimeš

www.uzis.cz

OBSAH

- 1. SYNTETIKA: Systém syntetických dat (SSD)
- 2. SYNTETIKA: Základní struktura
- 3. SYNTETIKA: Jak na to?
- 4. Laboratorní data

SYNTETIKA

System syntetických dat
(SSD)

- Přístupové údaje do SSD
- user: zkusebni_ucet
password: 1travnata_plochA
- Pouze po dobu semináře
- <https://www.nzip.cz/synteticka-data#uvod>
 - [Technická dokumentace](#)
 - 3.2 Technické zajištění přístupu

Syntetika: Systém syntetických dat (SSD)

- Unikátní přístup ke zdravotnickým datům
- Vytvořen a spravován výhradně ÚZIS ČR
(bez externích dodavatelů)
- Samotný přístup není nijak limitován
(o přístup může požádat kdokoliv)
- Není zpoplatněn

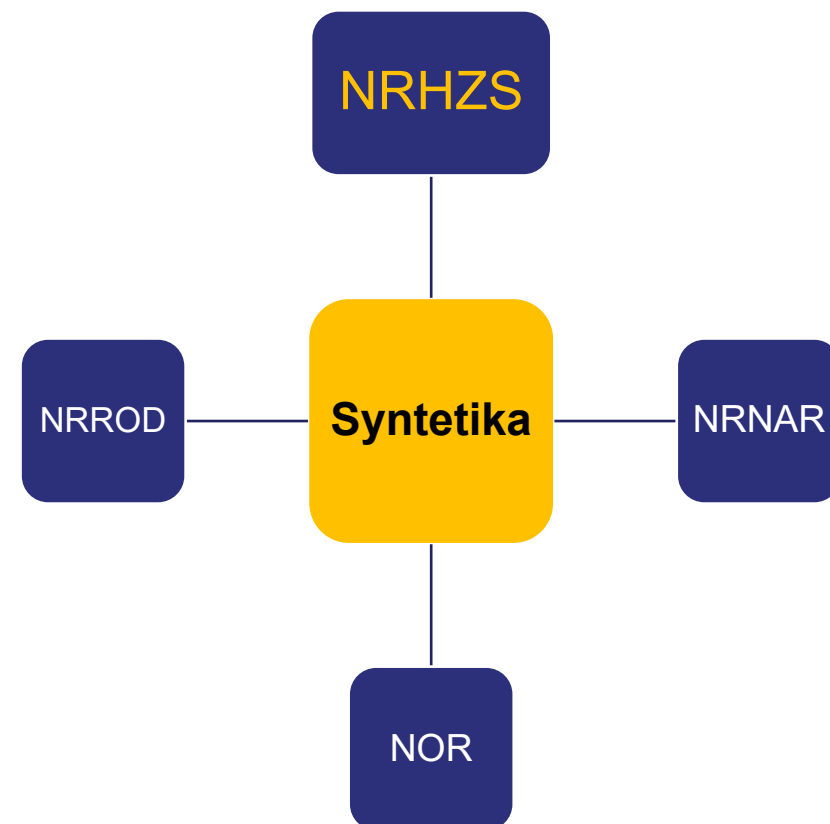


Základ syntetiky: NRHZS

- Obsahuje nejúplnější data (všechny ZZ / ZS DG na jednom místě)
- Obsahuje kompletní množinu pacientů (všichni pojištěnci ZP)
- Obsahuje detailní data (na úrovni jednotlivých položek hrazených ZS)
- Má své limitace:
 - Data až od roku 2010
 - Zpoždění dat
 - Složitá struktura

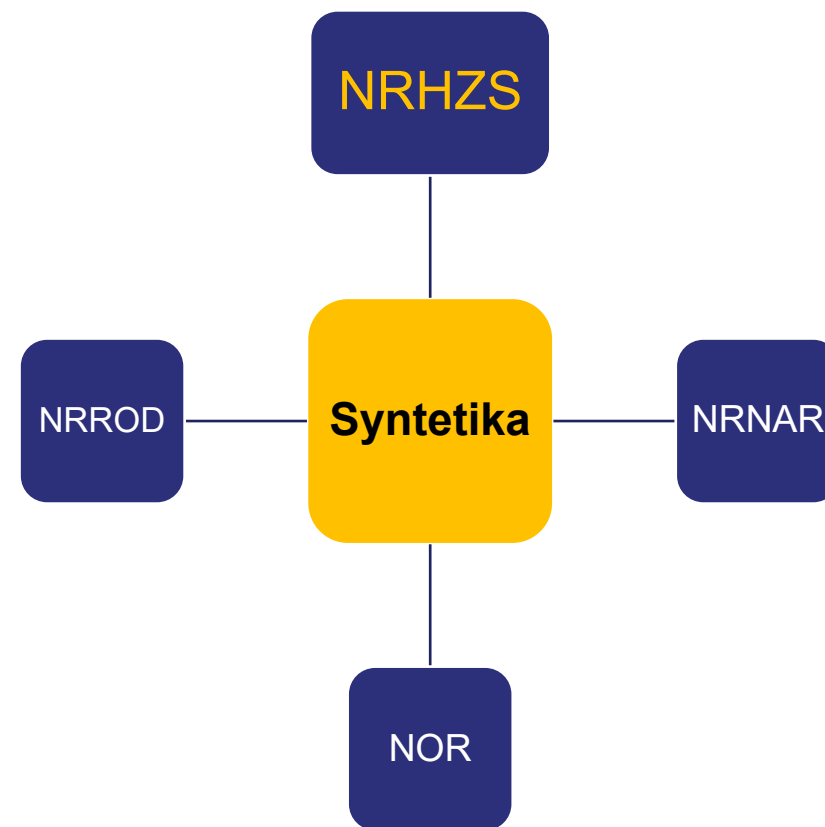
Rozšiřují klinické registry

- NOR
- NRNAR
- NRROD



7 základních kroků aneb co je potřeba znát pro pochopení syntetiky:

- 1) Základní rozdíly NRHZS a ostatními klinickými registry
- 2) Způsob tvorby syntetických dat
- 3) Struktura syntetické databáze
- 4) Způsob propojování (pacientů a zdravotnických zařízení)
-
- 5) Znalost SQL a práce s databází
- 6) Znalost zdravotnických dat
-
- 7) Znat omezení jednotlivých registrů
a respektovat doporučení při sestavování scriptů



1) Základní rozdíly mezi NRHZS a klinickými registry

ZDROJE dat:

Frekvence aktualizace

Forma předávání

Rozsah dat

Způsob anonymizace

Granularita dat

Velikost dat (největší tabulka)

A) Klinické registry

NRNAR, NRHOSP, NOR, ...

B) NRHZS

ZDROJE dat:

▪ Zdravotnická zařízení

▪ Zdravotní pojišťovny

Frekvence aktualizace

▪ **Průběžně** (i denně)

▪ **Kvartálně** (s 2-měsíčním zpožděním)

Forma předávání

Rozsah dat

Způsob anonymizace

Granularita dat

Velikost dat (největší tabulka)

A) Klinické registry

NRNAR, NRHOSP, NOR, ...

B) NRHZS

ZDROJE dat:

▪ Zdravotnická zařízení

▪ Zdravotní pojišťovny

Frekvence aktualizace

▪ **Průběžně** (i denně)

▪ **Kvartálně** (s 2-měsíčním zpožděním)

Forma předávání

▪ **Přímo do webového formuláře / dávkou**

▪ **Webový Portál** (archiv s csv soubory)

Rozsah dat

Způsob anonymizace

Granularita dat

Velikost dat (největší tabulka)

A) Klinické registry

NRNAR, NRHOSP, NOR, ...

B) NRHZS

ZDROJE dat:

▪ Zdravotnická zařízení

▪ Zdravotní pojišťovny

Frekvence aktualizace

▪ Průběžně (i denně)

▪ Kvartálně (s 2-měsíčním zpožděním)

Forma předávání

▪ Přímě do webového formuláře / dávkou

▪ Webový Portál (archiv s csv soubory)

Rozsah dat

▪ 1 ZZ: jednotky až tisíce záznamů

▪ Všechny smluvní ZZ, miliony záznamů

Způsob anonymizace

Granularita dat

Velikost dat (největší tabulka)

A) Klinické registry

NRNAR, NRHOSP, NOR, ...

B) NRHZS

ZDROJE dat:

▪ Zdravotnická zařízení

▪ Zdravotní pojišťovny

Frekvence aktualizace

▪ Průběžně (i denně)

▪ Kvartálně (s 2-měsíčním zpožděním)

Forma předávání

▪ Přímě do webového formuláře / dávkou

▪ Webový Portál (archiv s csv soubory)

Rozsah dat

▪ 1 ZZ: jednotky až tisíce záznamů

▪ Všechny smluvní ZZ, miliony záznamů

Způsob anonymizace

▪ Přímé identifikátory (RČ) => nzis_hash

▪ Nepřímé identifikátory (AIFO) => pseudonym

Granularita dat

Velikost dat (největší tabulka)

A) Klinické registry

NRNAR, NRHOSP, NOR, ...

B) NRHZS

ZDROJE dat:

▪ Zdravotnická zařízení

▪ Zdravotní pojišťovny

Frekvence aktualizace

▪ Průběžně (i denně)

▪ Kvartálně (s 2-měsíčním zpožděním)

Forma předávání

▪ Přímě do webového formuláře / dávkou

▪ Webový Portál (archiv s csv soubory)

Rozsah dat

▪ 1 ZZ: jednotky až tisíce záznamů

▪ Všechny smluvní ZZ, miliony záznamů

Způsob anonymizace

▪ Přímé identifikátory (RČ) => nzis_hash

▪ Nepřímé identifikátory (AIFO) => pseudonym

Granularita dat

▪ Pacient/Pobyt (hospitalizace, operace, ...)

▪ 1 záznam=1 zdrav. služba (výkon, LP, ..)

Velikost dat (největší tabulka)

A) Klinické registry

NRNAR, NRHOSP, NOR, ...

B) NRHZS

ZDROJE dat:

▪ Zdravotnická zařízení

▪ Zdravotní pojišťovny

Frekvence aktualizace

▪ Průběžně (i denně)

▪ Kvartálně (s 2-měsíčním zpožděním)

Forma předávání

▪ Přimo do webového formuláře / dávkou

▪ Webový Portál (archiv s csv soubory)

Rozsah dat

▪ 1 ZZ: jednotky až tisíce záznamů

▪ Všechny smluvní ZZ, miliony záznamů

Způsob anonymizace

▪ Přímé identifikátory (RČ) => nzis_hash

▪ Nepřímé identifikátory (AIFO) => pseudonym

Granularita dat

▪ Pacient/Pobyt (hospitalizace, operace, ...)

▪ 1 záznam=1 zdrav. služba (výkon, LP, ..)

Velikost dat (největší tabulka)

▪ 1.8 GB / 3,6 mil řádků 148 sloupců

▪ 3.2T nebo 18 mld řádků / 46 sloupců

Garant registru: nadefinuje základní strukturu

- Výběr sloupců (=> přehlednost)
- Úprava názvů (=> srozumitelnost)
- U klinických registrů jsou základem data transformovaná do 1 tabulky
- V NRHZS jsou data jednotlivých položek rozdělena do samostatných view, doplněných o základní číselníkové údaje
- V NRHZS jsou veškeré údaje z dokladu doplněny k položkám, aby se zamezilo joinování dokladu a položek
- V NRHZS je PSEUDONYM transformován na nzis_hash, aby se uživateli zjednodušilo propojování registrů

Oddělení datamanagementu

- Vytvoří uměle vygenerovaná syntetická data k danému registru
- Vytvoří syntetickou strukturu / view nad syntetickými daty dle definice garanta
- Doplní potřebné číselníky
- Shodné view vytvoří i v ostré DB, napojené do ostrých dat

Syntetická data NRHZS

- Jsou vygenerovaná za období 2015–2019 a odpovídající cca **0,1 % objemu** dat produkčního prostředí (~100 tis. pacientů, 10 mil. dokladů, 20 mil. položek)
- Zachovávají vazby mezi zdravotními službami
- Obsahují pouze nejčtenější zdravotní služby (podle reálných dat NRHZS):
 - 100 nejčtenějších odborností na dokladech
 - 100 nejčtenějších 3znakových MKN kódů diagnóz
 - 100 nejčtenějších zdravotních výkonů, HVLP, PZT
 - 10 nejčtenějších IPLP, stomatologických prostředků, dopravních výkonů
 - 5 ošetrovacích kategorií pacientů , 3 nejčtenější druhy nákladů na lázeňské pobyty

Syntetická data klinických registrů

- Základem data transformovaná do 1 tabulky
- Obsahují všechny kombinace vyskytujících se hodnot, doplněny jsou základní logické vazby mezi vybranými údaji
- Identifikátory pacientů, zařízení a období jsou upraveny tak, aby data se dala alespoň částečně propojovat

Výhody

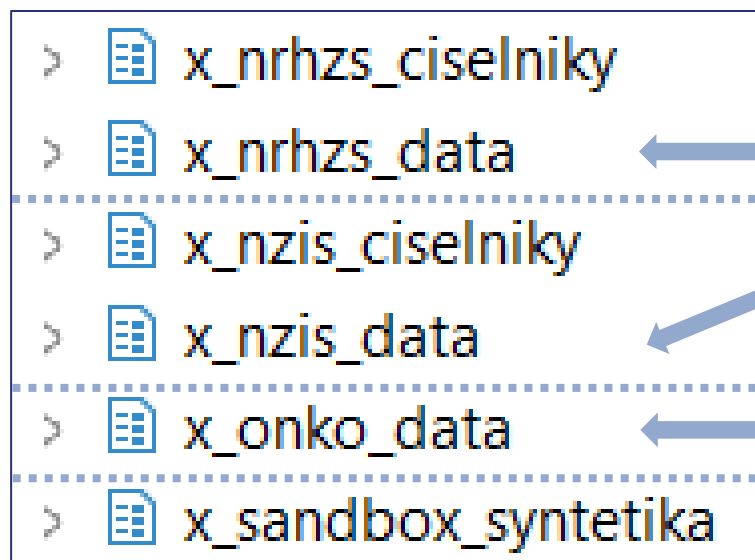
- Ideální pro jednoduché scripty
- Umožňuje přístup k agregovaným „ostrým“ datům dle vlastní definice
- Upravená struktura (názvy tabulek a sloupců) usnadňuje orientaci v datech a zamezuje nadbytečným „joinum“
- Pravidelná aktualizace dat
- Jednotný přístup k datům z různých registrů

Omezení

- Není vhodné používat u složitějších analýz, kdy je nezbytné analyzovat jednotlivé „mezivýstupy“
- Nutnost optimalizace scriptu pomocí „Explain“(náročnost) – na syntetice je cca 1000x menší než v ostré db
- NRHZS: Při analýze položky (konkrétní dg, výkon, léčivo), která není v syntetice zahrnuta, se nutně ji dočasně nahradit existujícím kódem
- Z důvodu složitosti zdravotnických dat a specifik jejich vykazování se **doporučuje** využívat syntetickou DB „odborné veřejnosti“

SYNTETIKA

Základní struktura

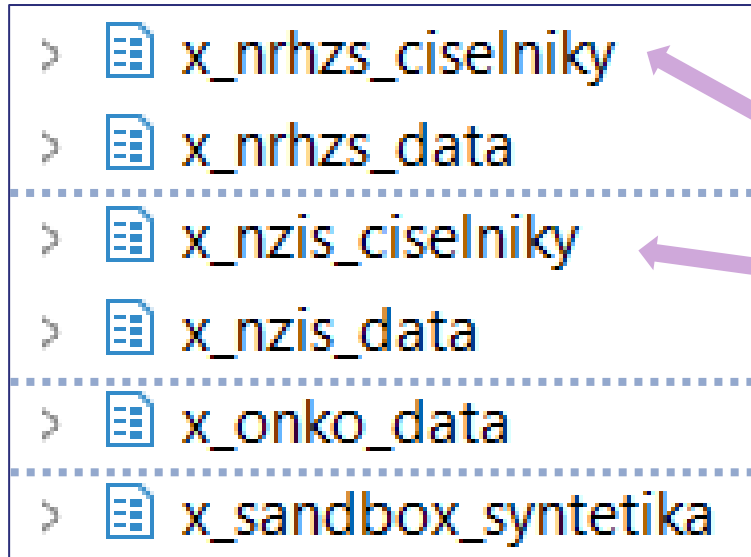


- Data NRHZZ a .základních klinických NZIS registrů jsou v samostatných schématech: ***_data**
- Související číselníky jsou ve schématech: ***_ciselniky**
- Samostatně jsou navíc vyčleněna onkologická data: **x_onko_data** (v návaznosti na <https://onconet.cz/>)

- Ve schématech:

***_data** a ***_ciselniky**

jsou přístupná pouze view. Nelze zde zasahovat do dat ani do struktury.

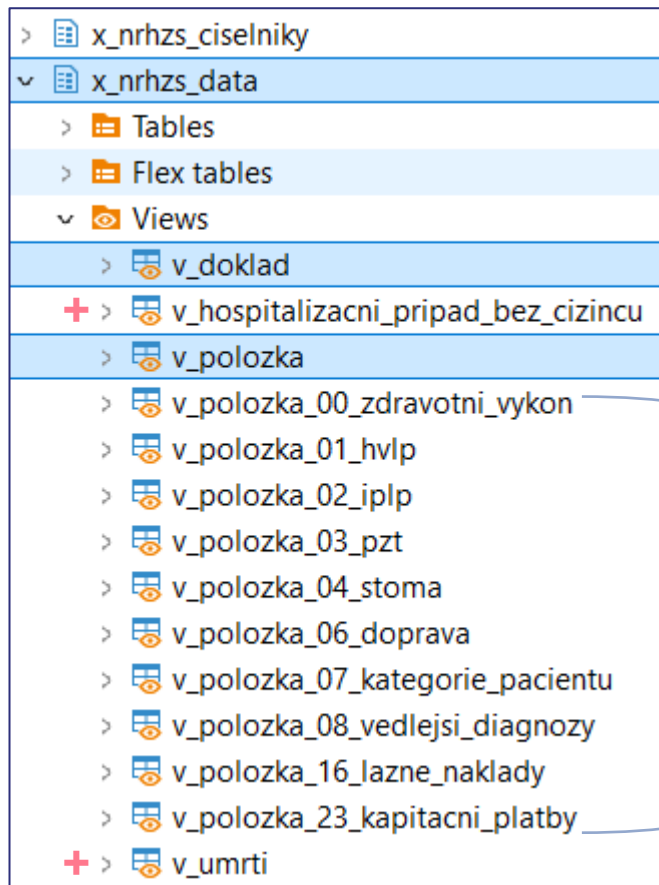


- Data NRHZS a .základních klinických NZIS registrů jsou v samostatných schématech: ***_data**
- Související číselníky jsou ve schématech: ***_ciselniky**
- Samostatně jsou navíc vyčleněna onkologická data: **x_onko_data** (v návaznosti na <https://onconet.cz/>)

- Pracovní schema:

x_sandbox_syntetika

(Pouze v tomto schema lze vytvářet a upravovat tabulky a view)



▪ **v_doklad:** základem je věta AB dle DR NRHZS

▪ **v_polozka:** základem je věta AC dle DR NRHZS

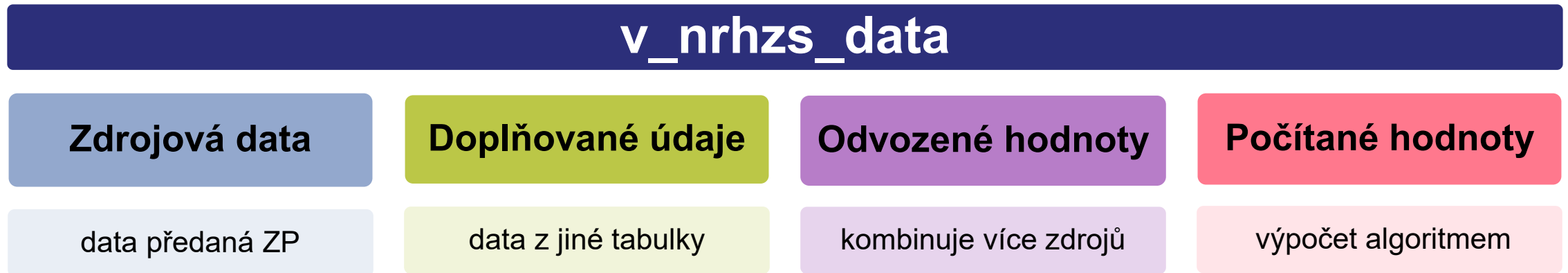
▪ **v_polozka_*:**

- pouze vybraný druh zdravotní služby
- doplněný o číselníkové údaje

+ 2 specifická view:

- **v_umrti:** základem je registr LPZ (List o prohlídce zemřelého)
- **v_hospitalizacni_pripad_bez_cizincu:** případy spočítané dle CZ_DRG na datech NRHZS

- Pro snadnější orientaci v datech NRHZZ používáme čtyřbarevné schéma:



Syntetika: v_doklad

druh_pojisteni_dokladu
druh_dokladu
prijeti_k_hospitalizaci
ukonцени_hospitalizace
diagnoza_dokladu
icp
lcp_zadatele
odbornost
odbornost_zadatele
datum_vystaveni_dokladu
datum_prijeti
datum_propusteni
id_dokladu
pojistovna
rok_davky

Zdroj dat

AB.DRUH_POJ
AB.DRUDOK
AB.DOPOR
AB.ZPUKON
AB.ZDG
AB.ICP
AB.ICP_ZAD
AB.ODB
AB.ODB_ZAD
AB.DAT_VYST
AB.PRI
AB.PRO
Počítaná hodnota
Predavaci_Protokol.CPO
AA.ROK

Syntetika: v_doklad

pseudonym_nzis
pohlavi
rok_narozeni
pacient_okres
ico
icz
icz_okres
icp_okres
skupina_diagnoz_seznam
pocet_diagnoz
den_prvni_polozky
den_posledni_polozky
odb_prevazujici
odb_prevazujici_zadatele
id_pripadu_bez_cizincu

Zdroj dat

Odvozená hodnota
AE.POHLAVI
AE.NAROZENI
Odvozená hodnota
Odvozená hodnota
AB.ICP
Odvozená hodnota
Odvozená hodnota
Odvozená hodnota
Počítaná hodnota
Odvozená hodnota
Odvozená hodnota
Počítaná hodnota
Počítaná hodnota
Počítaná hodnota

Syntetika: v_polozka

druh_polozky
kod_polozky
mnozstvi
den
body
kc
zvlastni_uhrada
diagnoza_polozky
odbornost_polozky
druh_pojisteni_polozky
Id_polozky
Id_dokladu
pojistovna
rok_davky
pseudonym_nzis
pohlavi
rok_narozeni
pacient_okres
druh_dokladu
prijeti_k_hospitalizaci

Zdroj dat

AC.DRUH_POL
AC.KOD
AC.MNOZSTVI
AC.DEN
AC.BODY
AC.KC
AC.ZVLU
AC.DG_RAD
AC.ODB_RAD
AC.DRUH_POJ_RAD
Počítaná hodnota
v_doklad.id_dokladu
Predavaci_Protokol.CPO
AA.ROK
v_doklad.pseudonym_nzis
AE..POHLAVI
AE.NAROZENI
v_doklad.pacient_okres
AB.DRUDOK
AB.DOPOR

Syntetika: v_polozka

ukonceni_hospitalizace
datum_vystaveni_dokladu
datum_prijeti
datum_propusteni
ico
icz
icz_okres
icp
icp_okres
lcp_zadatele
odbornost_dokladu
odbornost
odbornost_prevazujici
odbornost_zadatele
odbornost_prevazujici_zadatele
diagnoza_dokladu
diagnoza
druh_pojisteni_dokladu
druh_pojisteni
id_pripadu_bez_cizincu

Zdroj dat

AB.ZPUKON
AB.DAT_VYST
AB.PRI
AB.PRO
v_doklad.ico
AA.ICZ
v_doklad.icz_okres
AB.ICP
v_doklad.icp_okres
AB.ICP_ZAD
AB.ODB
Odvozená hodnota
v_doklad.odb_prevazujici
AC.ODB_ZAD
v_doklad.odb_prevazujici_zadatele
AB.ZDG
Odvozená hodnota
AB.DRUH_POJ
Odvozená hodnota
v_doklad.id_pripadu_bez_cizincu



NRHZS Odvozené a počítané hodnoty

Syntetika: v_doklad

pseudonym_nzis

pacient_okres

ico

icz_okres

icp_okres

skupina_diagnoz_seznam

den_prvni_polozky

den_posledni_polozky

Syntetika: v_doklad

id_dokladu

pocet_diagnoz

odb_prevazujici

odb_prevazujici_zadatele

id_pripadu_bez_cizincu

Odvozené hodnoty

Hash z RČ nebo CP pacienta

Z údajů ve větě AE

Z údajů ve větě CA

Z údajů ve větě CA

Z údajů ve větě CE

AB.ZDG a AC.VDG (druh_pol=8)

min(AC.DEN) dokladu

max(AC.DEN) dokladu

Počítaná hodnota

Interní unikátní identifikátor dokladu

Počet ZDG+VDG

Nejčastější odbornost pro icp a rok

Nejčastější odbornost pro icp a rok

Identifikátor HP dle CZ-DRG

Syntetika: v_polozka

odbornost

diagnoza

druh_pojisteni

Syntetika: v_polozka

id_polozky

Odvozené hodnoty

coalesce

(AC.ODB_RAD, AB.ODB)

coalesce

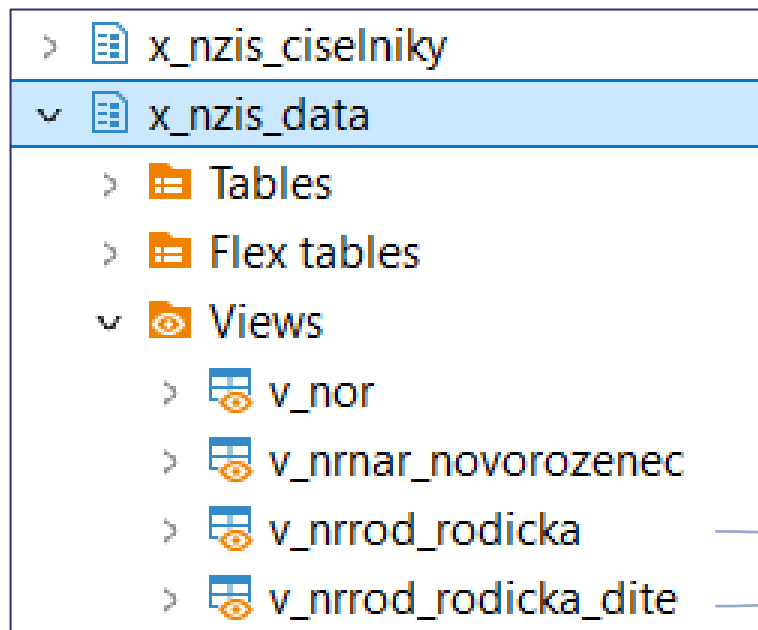
(AC.DG_RAD, AB.ZDG)

coalesce

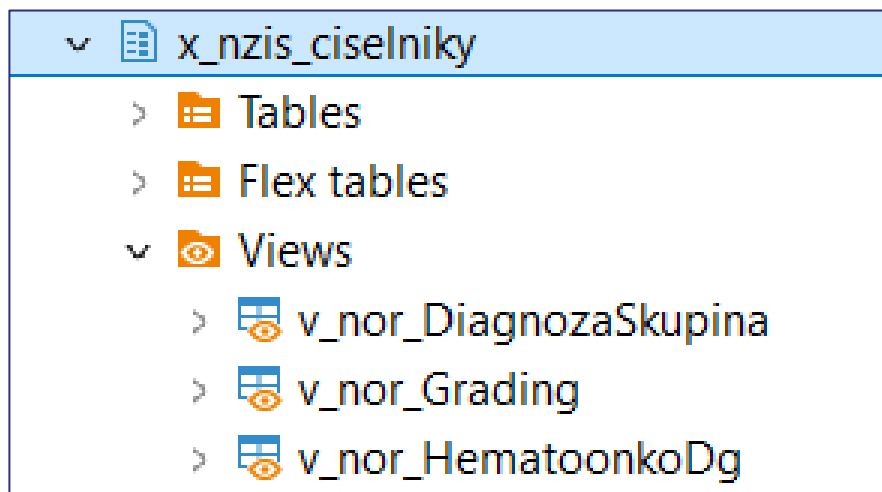
(AC.DRUH_POJ_RAD, AB.DRUH_POJ)

Počítaná hodnota

Interní unikátní identifikátor položky



- **v_nor**: základem je Národní onkologický registr
- **v_nrnar_novorozec**: základem je Národní registr novorozenců
- **v_nrrod_***: základem je Národní registr rodiček



Používají prefix registru ke snadnější orientaci v číselnících

- **v_nor_***: pouze pro NOR
- **v_nrnar_***: pouze pro NRNAR
- **v_nrrod_***: pouze pro NRROD
- **v_nrrz_***: společné pro NRNAR a NRROD
- **v_uzis_***: společné pro všechny klinické registry ve schématu x_nzis_data

- >  v_nor_MorfologieZdroj
- >  v_nor_Stadium
- >  v_nor_VekovaKategorie
- >  v_nrnar_DuvodUkonceniZpravyONovorozenci
- >  v_nrnar_PolohaPlodu
- >  v_nrnar_PorodKde

- >  v_nrrod_VzdelaniRodicka
- >  v_nrrod_ZhodnoceniCTG
- >  v_nrrz_Vitalita
- >  v_uzis_KrajOkres
- >  v_uzis_MKN10_5

Pacient (nzis_pseudonym):

- Pseudonymy používané v NRHZS jsou převedeny na nzis_pseudonym.
- Pokud pacient nebyl v NRHZS ztotožněn, nelze je jej přímo propojit s klinickými registry (cca 3 %)

Zdravotnické zařízení:

- NRHZS: ICP a ICZ (+ doplněné ICO přes počítaný číselník ICZ, 1 ICZ = 1 ICO)
- Klinické registry : ICO + PCZ

NRNAR vs. NRROD:

- Propojení přes datum narození dítěte + hash matky a Zdravotnické zařízení

NRHZS vs. NOR:

- Propojení pouze přes pacienta, příp. období

NRHZS vs. NRNAR/NRROD :

- Propojení přes pacienta (hash matky, hash dítěte + datum), zpřesnění přes datumové položky

SYNTETIKA

Jak na to ?

- **1. krok: Podat žádost ke zřízení přístupu do SSD**
 - V rámci 1 organizace může 1 přístup používat více uživatelů
 - Žádost o data musí být podávána jménem registrovaného žadatele
- **2. krok: Připojit se do syntetické databáze**
 - Návod je součástí technické dokumentace
- **3. krok: Připravit a otestovat sql script**
 - Nutno dodržovat požadované limity a respektovat omezení
 - Výsledné scripty optimalizovat
- **4. krok: Podat žádost o výstup**
 - Žádost je přijata po formální kontrole zařazena do seznamu
 - Pro každou analýzu je nutno vyplnit formulář žádosti

Komunikace pouze formou e-mailu:
syntetika@uzis.cz

Formuláře obou žádostí jsou k dispozici na webu: <https://www.nzip.cz/synteticka-data#metodika>

1. Jednorázová registrace

Nejprve se zaregistrujte do Systému syntetických dat NZIS, což vám umožní vytvářet vlastní analytické skripty.

K registraci:

- vyplňte [online formulář](#), nebo
- odešlete vyplněný a podepsaný [dokument MS Word](#) v elektronické podobě na e-mailovou adresu syntetika@uzis.cz;
- vyčkejte na kontaktování ze strany technických administrátorů Systému syntetických dat NZIS.

Seznam subjektů registrovaných do Systému syntetických dat NZIS není veřejně dostupný.

Žádost o registraci

- > [Registrace do systému \(online\)](#)
- > [Registrace do systému \(docx, 260 kB\)](#)

- Žádost je přijata po formální kontrole (**správné vyplnění** všech údajů urychlí proces získání přístupu do SSD)
- Nový uživatel obdrží **přístupové údaje** včetně odkazů na dokumentaci na registrovaný e-mail (heslo se předává formou SMS)

Žádost o přístup

webový formulář / dokument (docx)

Kontrola žádosti

formální kontrola (příp. vrácení k doplnění)

Přijetí žádosti

evidence, zaslání informace žadateli

Zřízení přístupu

vytvoření přístupových práv

Předání přístup. údajů žadateli

login (e-mail)
heslo (SMS)

3.2 Technické zajištění přístupu

Datový sklad syntetických dat NZIS je přístupný přes internet na adrese:

Adresa	Port	Význam	Databáze
syntetika.uzis.cz	5433	Syntetický sklad pro data NZIS	HP Vertica

Pro přístup k databázi je doporučován DB klient DBeaver (<https://dbeaver.io/download/>). Vyberte ikonu New Database Connection (Ctrl+Shift+N) vlevo. Jako typ připojení vyberte Vertica.

Syntetická data jako modalita zpřístupňování dat NZIS: technická dokumentace

Strana 19 z 44

nzis.open

Jak se připojit?

Návod je součástí technické dokumentace, kterou naleznete na úvodní stránce webu.

Dokumentace

- > [Metodická dokumentace \(pdf, 760 kB\)](#)
- > [Technická dokumentace \(pdf, 2,8 MB\)](#)
- > [Technická dokumentace – onkologie \(pdf, 490 kB\)](#)

<https://www.nzip.cz/synteticka-data#uvod>

■ Příprava scriptů

- Sestavení funkční sady scriptů
- Kontrola náročnosti pomocí „explain“ (Cost ideálně <100 k, max nižší stovky)
- Kontrola, zda se ve výsledku neduplikují data
- Kontrola, zda je výsledek dostatečně agregovaný
- Odstranění nepotřebných částí scriptů (např. samostatné kontrolní selecty, zbytečné prázdné řádky), popisy a poznámky v „zakomentované“ podobě lze ponechat
- Odstranění všech pomocných tabulek (drop table, po spuštění kompletního scriptů v DB zůstane pouze požadovaná výstupní tabulka)
- Číselníkové popisné (textové) hodnoty připojovat až k výslednému agregovanému výstupu

■ Doporučení

- Je vhodné používat jednoduché pomocné tabulky, např. s výčtem pacientů nebo diagnóz, které splňují základní podmínky výběrové množiny (ta se následně využije k filtrování dat)
- Používání zástupných kódů položek, pokud se tyto v syntetických datech nevyskytují
- Sledované veličiny parametrizovat
- Dodržovat konvenci názvů vytvářených tabulek (jednotný prefix označující žadatele)

UZIS NZIS SYNTETIKA V PRAXI: 3. krok: Příprava sql scriptu

Před spuštěním SQL dotazu se doporučuje provést kontrolu příkazem EXPLAIN, který vyhodnotí:

- **Korektnost** – zda dotaz proběhne bez chyby (rychlejší ověření než při jeho samotném spuštění).

V případě špatně napsaného dotazu, se zobrazí chybové hlášení,

Např.:

 SQL Error [2624] [42703]: [Vertica][VJDBC](2624) ERROR: Column "ab.id_doklad" does not exist

- **Náročnost** (využití paměti) a **velikost** (počet řádků) – dotazy s hodnotu Cost: > 300K nebo Rows v „B“ se vhodné optimalizovat. Tabulky v řádech „B“ (miliardy řádků) nelze v ostré db vytvářet!

```
EXPLAIN
SELECT
  id_polozky,
  druh_polozky
FROM x_nrhzs_data.v_polozka;
```

SSD:

Access Path: » Provedení trvá cca 221 ms
+-JOIN HASH [LeftOuter] [Cost: 254K, Rows: 14M] (PATH ID: 2)

NRHZZ:

Access Path: » Provedení trvá cca 21 s
+-JOIN HASH [Cost: 275M, Rows: 17B] (PATH ID: 2)

```
EXPLAIN
SELECT
  ac.id_polozky,
  ac.druh_polozky,
  ab.id_dokladu,
  ab.odbornost
FROM x_nrhzs_data.v_polozka ac
LEFT JOIN x_nrhzs_data.v_doklad ab
  ON ac.id_dokladu = ab.id_dokladu;
```

SSD:

Access Path: » Provedení trvá cca 3 s
+-JOIN HASH [LeftOuter] [Cost: 482K, Rows: 14M] (PATH ID: 1)

NRHZZ:

Access Path: » Provedení trvá cca 23 min
+-JOIN HASH [LeftOuter] [Cost: 594M, Rows: 17B] (PATH ID: 1)

- **Co je nutné si uvědomit při práci v SSD / při tvorbě scriptu**

- Databáze NRHSZ obsahuje mld. záznamu => **BIG DATA**. Tomu je nutno přizpůsobit práci s db.
 - NELZE vytvářet kopie základních položkových tabulek
- ZP předávají produkční data ve větách AA, AB, AC, AE
- V SSD jsou k dispozici pouze věty AB (doklady) a AC (položky). Aby se zamezilo spojování těchto dvou „BIG“ tabulek, jsou všechny důležité údaje z AA a AE doplněny do AB (k dokladu) a od tam pak do AC (k jednotlivým položkám). Zároveň jsou v ID dokladů upravena tak, aby nešlo tyto 2 tabulky na sebe přímo joinovat
 - NELZE spojovat doklady (věta AB) a položky (věta AC)
- SSD obsahuje 0,1 % dat oproti reálné db
- explain v SSD ukazuje náročnost xx Cost / XX row, v ostré DB to bude o 3 řády výše
 - Cost: 250K => cca 250M
 - Row: 7M => cca 7B

▪ Příprava „Žádosti o výstup“

- Pro každý samostatnou analýzu je nutno vyplnit formulář žádosti (k dispozici na webu)
- Formulář musí mj. obsahovat seznam všech výstupních tabulek včetně jejich přesného popisu
- Zaslání vyplněného formuláře žádosti a scriptu na adresu „syntetika@uzis.cz“
- Kompletní script se předává v jednom souboru

▪ Doporučení k žádosti

- Omezit množství jednotlivých výstupních souborů (každý prochází 3-úrovňovou kontrolou)
- Omezit požadavky až na konečné výsledky
- Žádost může být odmítnuta z důvodu výskytu citlivých a příliš detailních údajů (pseudonym, konkrétní věk, ...)
- Žádost může být odmítnuta z důvodu nadměrné velikosti výstupů, ideálně v MB, max statisíce záznamů (nedostatečně agregovaná data)

Počty pacientů na pracovištích u vybraných diagnóz

```
SELECT YEAR (p.den) AS rok,
       p.icp,
       c.icz,
       z.icz_nazev,
       p.odbornost_prevazujici,
       c.okres AS okres_icp,
       p.diagnoza,
       COUNT(DISTINCT p.pseudonym_nzis) AS pocet_uni_pacientu
FROM (SELECT den,
             icp,
             diagnoza,
             pseudonym_nzis,
             odbornost_prevazujici
      FROM x_nrhzs_data.v_polozka
      WHERE druh_dokladu IN ('01', '02', '06')
      AND diagnoza LIKE 'G%'
     ) p
LEFT JOIN x_nrhzs_ciselniky.v_cis_icp c ON p.icp = c.icp
LEFT JOIN x_nrhzs_ciselniky.v_cis_icz z ON c.icz = z.icz
GROUP BY YEAR(p.den), c.icz, p.icp, z.icz_nazev,
         p.odbornost_prevazujici, c.okres, p.diagnoza
ORDER BY rok, icp
;
```

- ❑ Sql vytvořené uživatelem na syntetické databázi
- ❑ Na první pohled jednoduché sql
- ❑ Explain:
 - syntetika [Cost: 6M, Rows: 10M]
 - ostrá db [Cost: **8B**, Rows: **15B**]

=> nelze spustit
- ❑ Problém:
 - joinování doplňujících informací k základnímu neagregovanému výběru

Počty pacientů na pracovištích u vybraných diagnóz – optimalizováno

```
SELECT p.rok,
       p.icp,
       p.odbornost_prevazujici,
       p.diagnoza,
       p.pocet_uni_pacientu,
       c.icz,
       c.okres AS okres_icp,
       z.icz_nazev
FROM ( SELECT year(den) as rok, icp, diagnoza, odbornost_prevazujici,
              COUNT(DISTINCT pseudonym_nzis) AS pocet_uni_pacientu
      FROM x_nrhzs_data.v_polozka
      WHERE druh_dokladu IN ('01', '02', '06')
            AND diagnoza LIKE 'G%'
      GROUP BY year(den), icp, diagnoza, odbornost_prevazujici
    ) p
LEFT JOIN x_nrhzs_ciselniky.v_cis_icp c ON p.icp = c.icp
LEFT JOIN x_nrhzs_ciselniky.v_cis_icz z ON c.icz = z.icz
ORDER BY rok, icp
;
```

- ❑ Agregace dat v základním subselectu „p“
- ❑ Číselníkové/popisné údaje napojeny až nakonec
- ❑ Explain:
 - syntetika [Cost: 315K, Rows: 7K]
 - ostrá db [Cost:832M, Rows: 316K]
 - původně [Cost: **8B**, Rows: **15B**]*
- => Stále náročné, ale lze spustit**
- ❑ Další optimalizace – omezování základní množiny

- **Funguje pro jednodušší výstupy**

- Základní agregace (počty pacientů, výkonů, léčiv, diagnóz) v čase a prostoru
- Seznamy pacientů s omezeným počtem atributů

- **Značné problémy s odladováním složitějších výstupů**

- Obsahové – uživatel nezná všechna specifika dat a výstupy pak neodpovídají tomu, co chtěl získat
- Optimalizace – část zaslaných scriptů nelze na ostré DB spustit (práce s „big data DB“ vyžadují pokročilou znalost SQL a procesů optimalizace)

- **Administrativní zátěž**

- Nutná kontrola každého výstupního souboru s *ohledem na ochranu osobních údajů*

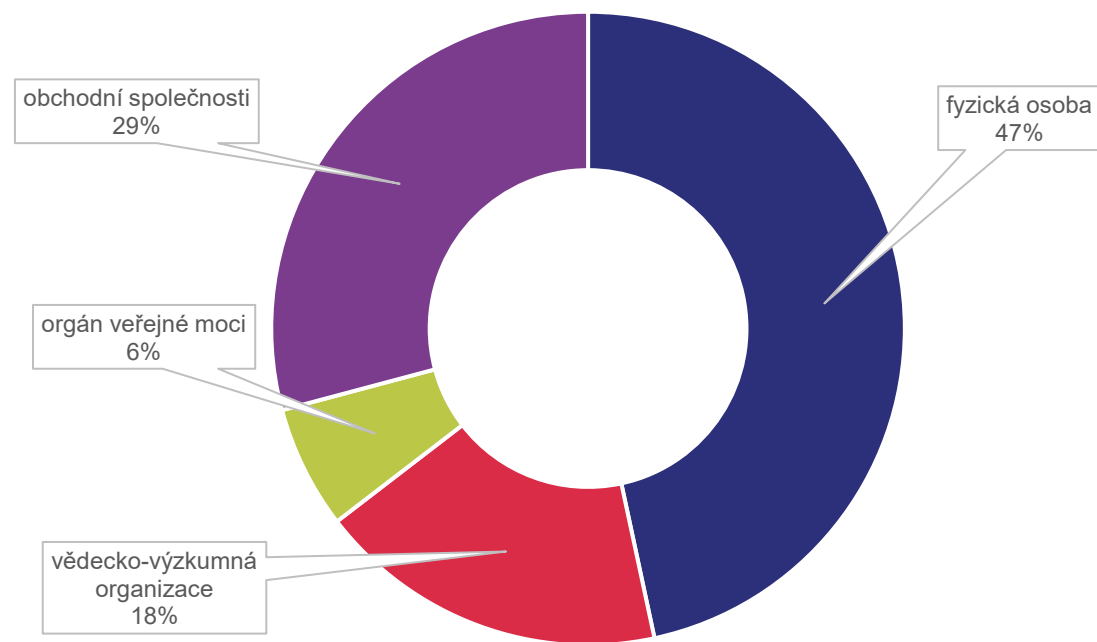
- **Rizika**

- Provozní zátěž databáze
- Přetěžování kapacity i výkonu databáze
- Nepřesná interpretace získaných dat

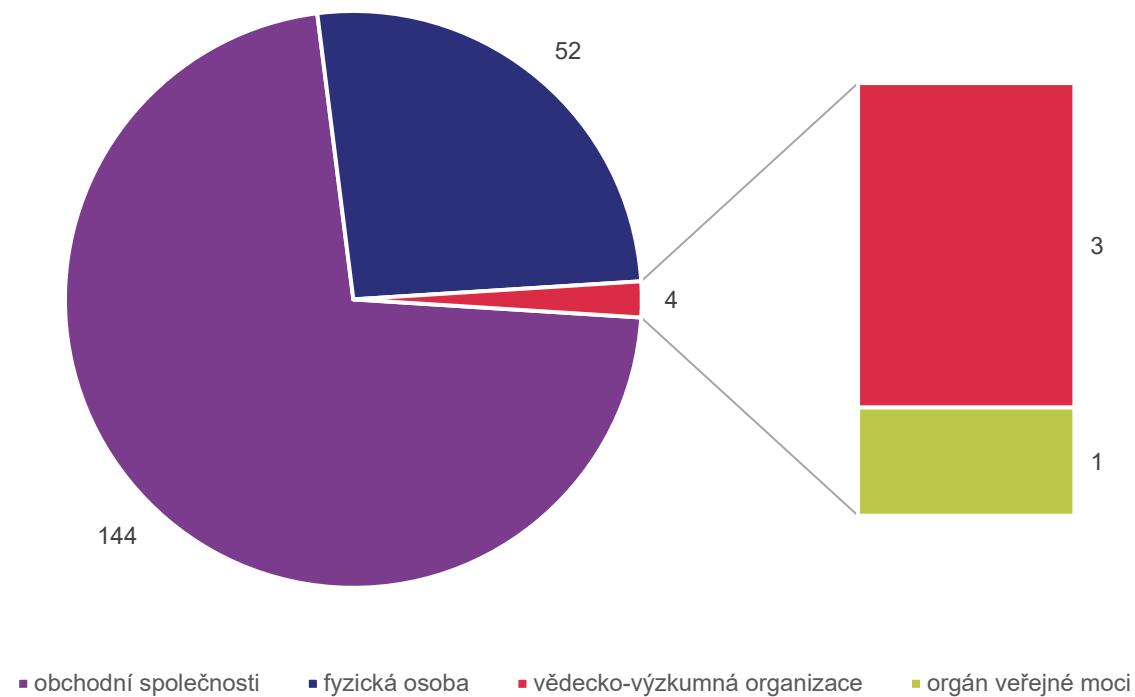
- **Základní problémy předávaných scriptů**

- Rozdělování scriptů do více souborů
- Rozsáhlé scripty – 1 sql soubor, který obsahuje:
 - Téměř 3000 řádků kódu
 - 150x create table
 - 20x insert
 - Každý script se musí prověřit => neúnosná zátěž
- Nesmyslnost výstupů vyplývajících z nedostatečné znalosti dat
- Duplikování výstupů nevhodnými joiny
- Vytváření kopií „BIG“ tabulek

Typy žadatelů o přístup do syntetiky



Zpracované výstupy



LABORATORNÍ DATA

LABORATORNÍ POLOŽKY

- Diabetes
 - glykovaný hemoglobin (HbA1c)
 - glukóza v plazmě/séru (na lačno)
 - C-peptid (sérum i moč)
 - anti-GAD
- Kardiovaskulární onemocnění
 - cholesterol celkový
 - cholesterol HDL
 - cholesterol LDL (vyšetření i výpočet)
 - triacylglyceroly
 - lipoprotein Lp(a)
 - NT-proBNP
- Štítná žláza
 - TSH (Tyreotropin)
 - T4 volný
- Vitamin D
 - 25-hydroxyvitamin D (kalcidiol) celkový
- Onemocnění ledvin
 - U protein celkový
 - urea
 - albumin v moči
 - kreatinin v moči
 - kreatinin
 - ACR (Album/kreatinin ratio) (výpočet)
 - Odhad glomerulární filtrace (eGFR) (výpočet)
- Onkologie
 - okultní krvácení do stolice (TOKS)
 - prostatický specifický antigen (PSA) celkový
 - PSA volný
 - p2PSA
 - vysokorizikové HPV - DNA test

Definováno skrze NČLP kódy

POLOŽKY ROZHRAŇÍ PRO ZP

Pacient

Číslo pojištěnce	Číslo pojištěnce zdravotní pojišťovny (povinné vždy)
Pacient	Datum narození , pohlaví, RID
Zdravotní pojišťovna	Kód zdravotní pojišťovny, které bylo nahlášeno vyšetření
Diagnóza	Kód diagnózy, pro kterou bylo vyšetření provedeno

Pracoviště

Identifikace laboratoře	IČP laboratoře
Identifikace žadatele	IČZ + IČP žadatele

Vyšetření

Datum odběru	Datum a čas odběru vzorku
Datum vyšetření	Datum a čas dokončení vyšetření (vykázání na ZP)
NČLP klíč	Kód provedeného laboratorního vyšetření dle číselníku NČLP
Vysledek	Naměřená numerická hodnota výsledku v uvedených jednotkách
Jednotky	Jednotky, ve kterých je výsledek uveden
Vysledek znamenko	Znaménko při překročení detekčních limitů metody
Vysledek text	Kvalitativní výsledek (POZIT, NEGAT)
Referenční meze	Dle standardu DASTA 4 (skala) jde o 8 atributů, kde povinný je minimálně skala4 a skala5 pro vymezení referenčních mezí https://dastacr.cz/dasta/hypertext/DSABS.htm
Skupina metod	Doplněná položka ÚZIS, agregace příbuzných NČLP kódů

- Data předána **v podobě CSV souborů** zdravotním pojišťovnám
- Historická data budou po kvartálech od roku 2020 až první pololetí 2026.
- Data připravena na úložišti pro zdravotní pojišťovny
- Další období (prospektivní data) budou předávána po měsících vždy k 15. dni v druhém následujícím měsíci (např. data za červenec 2026 budou k dispozici od 15. 9. 2026).

Rok 2025:

☐ 28 skupin

☐ 129 metod

☐ 75 mil. vyšetření

▪ Skupina „Cholesterol celkový“: 5 metod vs 1 zdravotní výkon

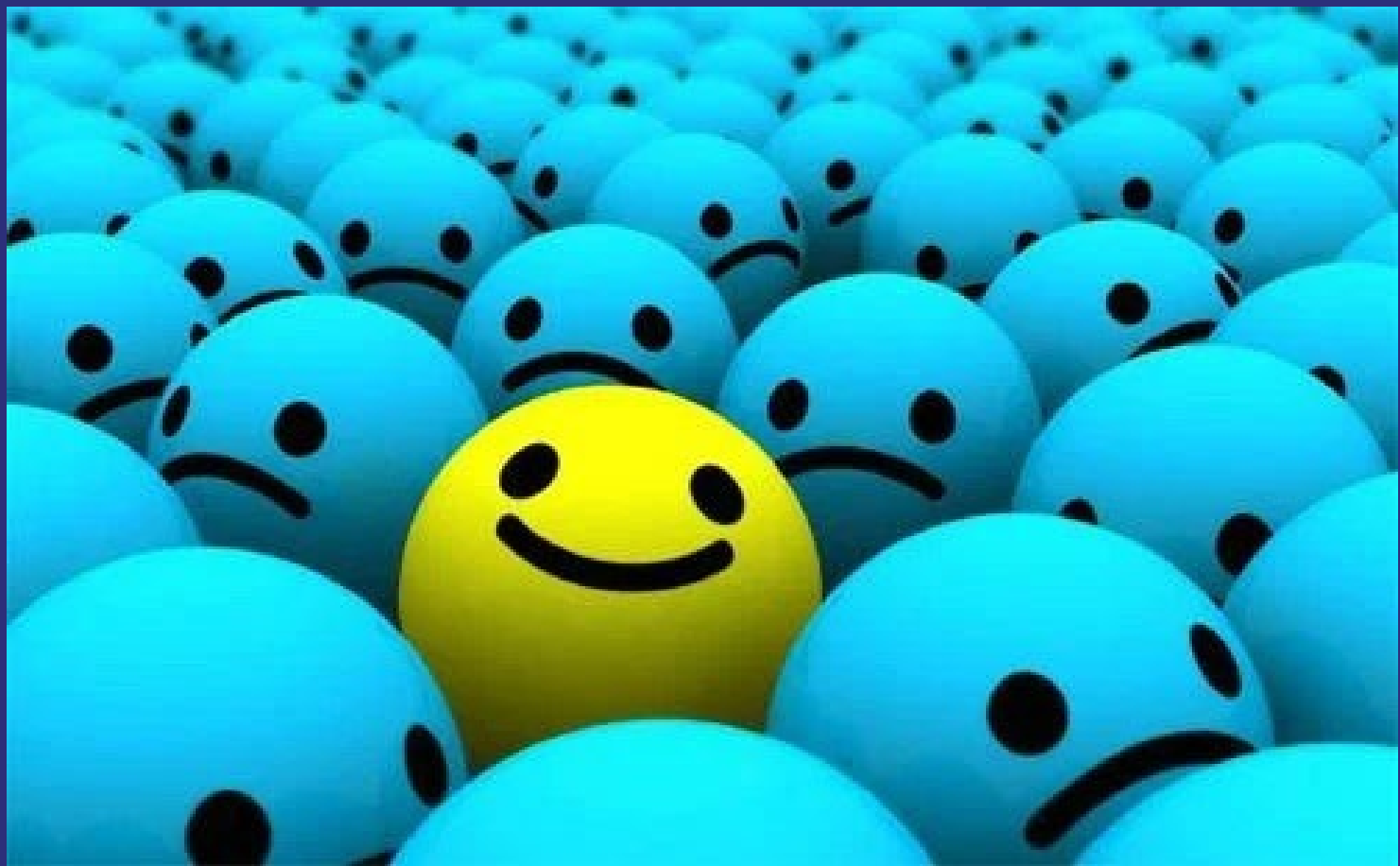
- ✓ NRHZS 5,886,050 výkonů / 182 ICO => => propojeno 96 %
- ✓ LABBDATA 5,625,407 / 157 ICO

▪ Skupina „PSA celkové“: 4 metody vs 3 zdravotní výkony

- ✓ NRHZS 1,240,768 výkonů / 169 ICO => propojeno 83 %
- ✓ LABBDATA 1,030,291 vyšetření / 136 ICO

▪ Skupina „Kreatinin v moči“: 6 metod vs 0 zdravotních výkonů

- ✓ LABDATA 950334 vyšetření / 144 ICO



Děkuji za pozornost.

... to be continue



BLIŽŠÍ INFORMACE

WWW.UZIS.CZ