

Syntetická data jako modalita zpřístupňování dat Národního zdravotnického informačního systému (NZIS)

METODICKÁ DOKUMENTACE

Autoři:	Ústav zdravotnických informací a statistiky ČR RNDr. Martin Komenda, Ph.D., MBA RNDr. Daniel Klimeš, Ph.D. RNDr. Jiří Jarkovský, Ph.D. Ing. Tomáš Májek RNDr. Ondřej Májek, Ph.D. MUDr. Miroslav Zvolský Mgr. Darina Valeková JUDr. Mgr. Vladimíra Těšitelová, LL.M. JUDr. Richard Galuška RNDr. Jan Mužík, Ph.D. Mgr. Jakub Gregor, Ph.D. Mgr. Ondřej Ngo, Ph.D. Mgr. Renata Chloupková Mgr. Jakub Kubát, DiS. Ing. Lucie Čačíková Mgr. Petr Klika prof. RNDr. Ladislav Dušek, Ph.D.
Odborní recenzenti:	České vysoké učení technické v Praze doc. Dr. rer. nat. Aleš Tichopád, Ph.D. Právnická fakulta, Univerzita Karlova doc. JUDr. Petr Šustek, Ph.D. JUDr. Lucie Šířoká
Verzování dokumentu:	Verze 1 (07.12.2024) Verze 2 (13.02.2025) Verze 3 (17.03.2025) Verze 4 (14.10.2025)

1 Úvod

Syntetická data, která jsou poskytována Ústavem zdravotnických informací a statistiky ČR (ÚZIS) jakožto správcem Národního zdravotnického informačního systému (NZIS), představují uměle vygenerovanou databázi odvozenou z NZIS. Tato databáze je sice založena na reálných strukturách a vazbách z primárních dat NZIS, nicméně obsahuje pouze fiktivní záznamy s modifikovaným obsahem, což zaručuje ochranu citlivých osobních informací a znemožňuje identifikaci jednotlivců.

Syntetická data NZIS jsou určena pro zájemce, kteří se zabývají pokročilými analýzami zdravotnických dat NZIS a potřebují data, která nejsou dostupná v [Katalogu zdravotnických dat](#) (otevřená data, datové souhrny). Z dostupných syntetických dat nelze přímo vyvodit žádné reálné výsledky. Tento inovativní systém je založen na přípravě výpočetních SQL skriptů a následné žádosti ÚZIS o jejich spuštění na primárních datech, což umožňuje efektivně zpracovávat citlivá data v souladu s platnou legislativou, aniž by byla porušena pravidla ochrany osobních údajů. Dodržením metodických pravidel získá žadatel požadované výsledky od ÚZIS, aniž by musel přistupovat k primární databázi. ÚZIS tak zefektivnil proces přípravy skriptů, neboť pověřený zaměstnanec pouze ověří, zda skripty splňují požadavky na strukturu, výpočetní náročnost a ochranu osobních údajů, a následně výsledky zašle zabezpečenou formou žadateli. Systém předávání konečných výsledků je založen na několika stupních kontrol zaměřených zejména na ochranu osobních údajů. Výsledná předávaná data jsou agregovaná a nevratně anonymizovaná; výstupy tak neumožní jakoukoli přímou nebo nepřímou identifikaci jedince.

Celý systém si zakládá na otevřeném přístupu a směřuje k dlouhodobému budování spolupráce, sdílení ověřených řešení a budování komunity v oblasti zdravotnických dat. Mezi nosné principy patří:

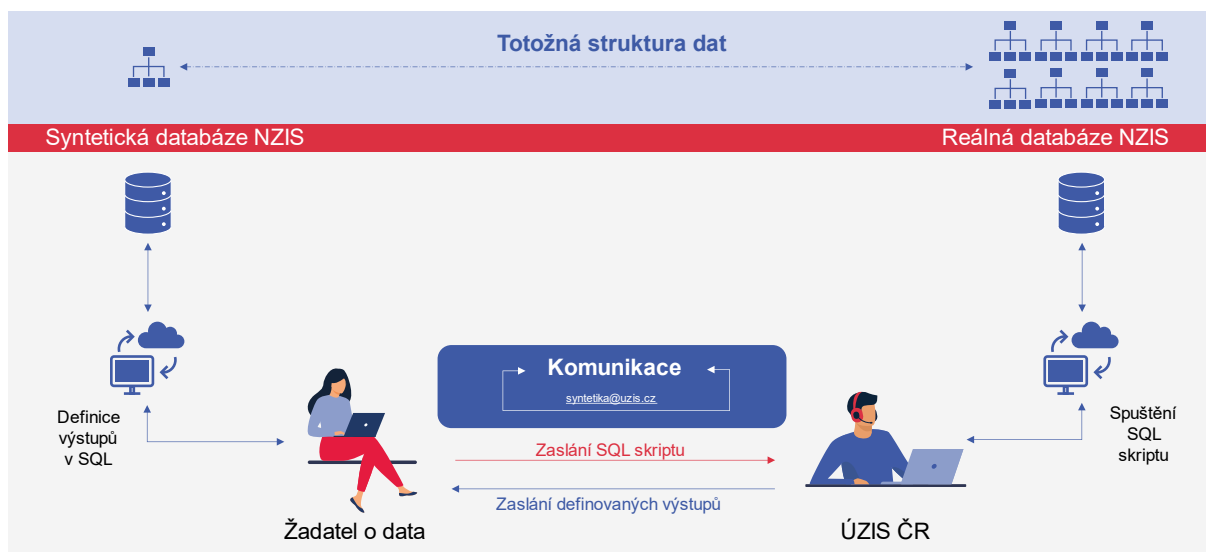
- **Transparentní přehled** požadavků na zastřešující platformě v podobě [Národního zdravotnického informačního portálu](#) (NZIP) v souladu s pravidly pro ochranu osobních údajů žadatele.
- **Spravedlivý přístup k žadateli**, kdy každý žadatel může odeslat v daný okamžik právě jednu žádost; po jejím vyřízení (zamítnuto nebo publikováno), může podat další žádost. Žádosti jsou zpracovávány v pořadí, v jakém byly doručeny.
- **Pravidelná komunikace s odbornou komunitou** zahrnuje organizaci online webinářů a prezenčních konferencí s cílem diskutovat zkušenosti a náměty od uživatelů zdravotnických dat.

2 Registrace a žádost o data

Aktivní zapojení a zpřístupnění syntetických dat NZIS žadatelům je rozděleno do dvou kroků, které pokrývají formální evidenci požadavků a podporují efektivní práci s nastaveným systémem.

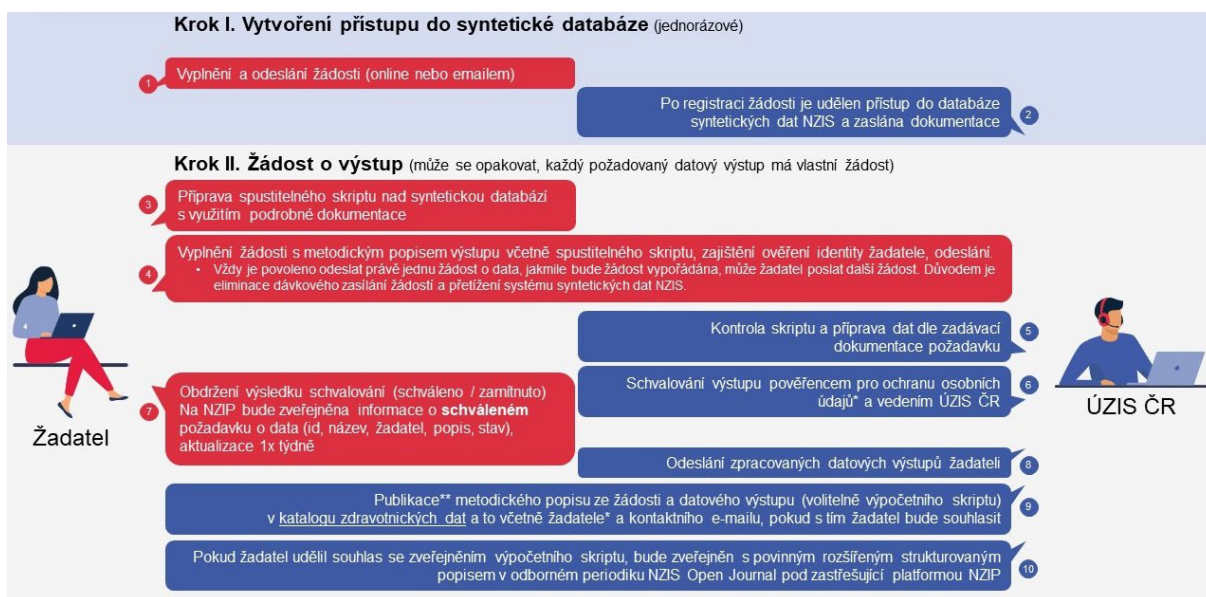
1. Vytvoření přístupu (jednorázová žádost a registrace uživatele, neveřejná) a předání základních informací o způsobu zveřejňování výstupů získaných ze systému syntetických dat NZIS.
2. Žádost o výstup (každý požadovaný datový výstup má vlastní žádost, v případě předání výstupu publikováno dle nastavených pravidel – kapitola 3).

V prvním kroku se uživatel seznámí s databází syntetických dat NZIS a vyzkouší si použití skriptovacího jazyka. Zjednodušené schéma komunikace (obrázek 1) znázorňuje, jak probíhá interakce mezi žadatelem a správcem dat.



Obrázek 1. Syntetická data NZIS: Schéma komunikace

V druhém kroku žadatel zasílá finální verzi funkčního SQL skriptu společně s metodickým listem požadovaného výstupu na ÚZIS. Následně probíhá kontrola, spuštění, validace a schválení připravených výstupů, které následně žadatel obdrží.



Obrázek 2. Detailní komunikace mezi žadatelem a ÚZIS

Orientační lhůta pro vypořádání žádostí o data je typicky do 30 pracovních dní od podání finální verze žádosti včetně dodání validního skriptu, případně dle výpočetní složitosti/náročnosti zpracování a dle dohody se žadatelem.

Pokud skript samotný nebo požadovaný výstup nesplňuje základní požadavky popsané v technické dokumentaci (kapitola 3.3 a 3.4) je o tom žadatel informován nejpozději do 5 pracovních dní od doručení žádosti. O technických problémech a obsahových nedostacích zjištěných v průběhu zpracování skriptu bude žadatel informován neprodleně. Nezbytné/dodatečné úpravy skriptů provádí vždy žadatel na základě doporučení ÚZIS.

3 Režimy zveřejnění

Syntetická data NZIS představují velmi silný nástroj pro budování otevřené spolupráce a sdílení výstupů napříč komunitou zabývající se analýzou zdravotnických dat. Hlavním cílem je umožnit transparentní a svobodné bádání nad daty NZIS při zajištění maximální možné ochrany osobních údajů. Syntetická data mají potenciál vytvořit zcela novou komunitu datových výzkumníků, kteří budou vzájemně sdílet, respektovat a citovat své výsledky v podobě algoritmů, výpočetních skriptů a také vlastních výsledků bádání. Především za tímto účelem byla nastavena dále uvedená pravidla zveřejňování a publikace výsledků.

Pro pokrytí požadavků žadatelů o tato data byly stanoveny tři úrovně zveřejnění (informace o žádosti, datové výstupy, SQL skript), které jsou níže popsány:

- **Zveřejnění „úroveň 1“** = Přehledová informace o schválené žádosti s rozlišením, zda je žadatelem fyzická osoba nebo právnická osoba
- **Zveřejnění „úroveň 2“** = Datové výstupy jsou publikovány vždy s odstupem 6 měsíců od zaslání dat žadateli v souladu s ochranou osobních údajů a dle metodického popisu v [Katalogu zdravotnických dat](#)
- **Zveřejnění „úroveň 3“** = SQL skript (volitelně dle preference žadatele)
 - Při explicitním souhlasu žadatele je možné publikovat SQL skript v odborném periodiku [NZIS Open Journal](#) po 6 měsících od zaslání dat žadateli včetně rozšířeného abstraktu a s vazbou na související datový výstup v [Katalogu zdravotnických dat](#).
 - Odborné periodikum [NZIS Open Journal](#) je další přidanou hodnotou v implementaci a metodickém rozvoji syntetických dat NZIS. Publikací svých metodických přístupů, skriptů a koncepcí získávají tvůrci plně autorizovaný výstup svého duševního vlastnictví ve formátu, který je standardním výstupem vědecké práce (plně citovatelná publikace v odborném časopise). Cílem časopisu bude zejména rozvíjet vzájemné sdílení výstupů a algoritmických postupů při zpracování dat NZIS.

4 Aktualizace dat

Výstupy ze syntetických dat budou publikovány v [Katalogu zdravotnických dat](#) jako datové souhrny, případně jako otevřená data dle jejich struktury a informační hodnoty. V metodickém popisu žadatel vybere ze dvou níže uvedených možností z pohledu navazující aktualizace dat:

- **Ad hoc připravený analytický výstup** v podobě datového souhrnu nebo otevřené datové sady, který nevyžaduje pravidelnou aktualizaci.
- **Pravidelně aktualizovaný datový souhrn nebo otevřená datová sada.** Pokud je výstup navázán na zveřejnění SQL skriptu „úroveň 3“ (viz výše), tedy skript bude volně k využití jako publikované autorské dílo, bude v souladu s otevřenou licencí zajištěna podobně jako u jiných výstupů stejného formátu automatická aktualizace ze strany ÚZIS. V opačném případě bude nutné podat novou žádost o výstup ze systému syntetických dat, která se bude řídit stejnými pravidly jako v případě žádosti o zcela nový výstup.

5 Stornování žádostí

Žadatelé mohou výstupy stornovat kdykoliv v průběhu podávání žádosti, dokud nemá tato žádost statut „Uzavřeno“. Výstupy, které již žadatel obdržel, nelze stornovat a budou zveřejněny v [katalogu zdravotnických dat](#) na NZIP v příslušném režimu a lhůtách. Pokud se žadatel rozhodne upravit již publikované výstupy (například z důvodu chyb), může podat novou žádost v běžném režimu jako korekci výstupu. Tyto výstupy budou zveřejněny jako korekce předchozího (korigovaného) výstupu, a to v termínu, který přísluší původní žádosti. Po uplynutí 6 měsíců od uzavření původní žádosti již korekce uplatnit nelze.

6 Dokumentace

V rámci portálu [nzip.cz](#) vznikl průvodce zdravotnickými daty, který obsahuje všechny důležité informace o Národním zdravotnickém informačním systému, dostupných datech a žádostech o jednotlivé typy dat, a to včetně detailního popisu systému [syntetických dat NZIS](#). Nedílnou součástí je technická dokumentace, která podrobně rozebírá strukturu databáze, způsoby práce a limitace v používání.

7 Přílohy

- Syntetická data jako modalita zpřístupňování dat NZIS: technická dokumentace
 - Popis struktura a obsah databáze společně s možnostmi použití v praxi
- NZIS Open Journal: recenzované publikační médium zaměřené na zpřístupňování, sdílení a zpracování zdravotnických dat
- Žádosti o registraci a výstupy ze syntetických dat NZIS, metodický popis výstupu ze syntetických dat NZIS

8 Citace

Syntetická data NZIS. *Národní zdravotnický informační portál* [online]. Praha: Ministerstvo zdravotnictví ČR a Ústav zdravotnických informací a statistiky ČR, 2025. Dostupné z: <https://www.nzip.cz/clanek/2230-synteticka-data-nzis>. ISSN 2695-0340.

Agenda syntetických dat NZIS je dedikována níže uvedené projektové podpoře:

- Systematická edukace a zvyšování zdravotní gramotnosti populace prostřednictvím NZIP (CZ.03.02.02/00/22_005/0001137)
- Zvyšování informovanosti prostřednictvím Národního zdravotnického informačního portálu (CZ.03.4.74/0.0/0.0/15_025/0016090)