
I-02 Projektový zámer

naposledy upravil Peter Saraka

- 2026/10/12 10:17

Obsah

PROJEKTOVÝ ZÁMER

Vzor pre manažérsky výstup I-02

podľa vyhlášky MIRRI č. 401/2023 Z. z. (účinnosť od 1.4.2025)

Povinná osoba	Štatistický úrad Slovenskej republiky				
Názov projektu	AI platforma pre Štatistický úrad Slovenskej republiky				
Zodpovedná osoba za projekt					
Realizátor projektu	Štatistický úrad Slovenskej republiky				
Vlastník projektu	Štatistický úrad Slovenskej republiky				
Schvaľovanie dokumentu					
Položka	Meno a priezvisko	Organizácia	Pracovná pozícia	Dátum	Podpis
					(alebo elektronický súhlas)

1. HISTÓRIA DOKUMENTU

Verzia	Dátum	Zmeny	Meno a priezvisko
--------	-------	-------	-------------------

2. ÚČEL DOKUMENTU, SKRATKY (KONVENCIE) A DEFINÍCIE

V súlade s vyhláškou MIRRI č. 401/2023 Z. z. v znení neskorších predpisov je tento výstup I-02 Projektový zámer určený na rozpracovanie detailných informácií prípravnej a iniciačnej fázy projektu z pohľadu aktuálneho stavu, budúceho stavu a navrhovaného riešenia.

Dokument Projektový zámer obsahuje manažérske zhrnutie, rozsah, ciele a motiváciu na realizáciu projektu, zainteresované strany, návrh merateľných ukazovateľov a obsahuje aj:

1. detailný opis požadovaných projektových výstupov,
2. detailný opis obmedzení, predpokladov, tolerancií a návrh organizačného zabezpečenia projektu,
3. detailný opis rozpočtu projektu a jeho prínosov,
4. harmonogram projektu,
5. vyhodnotenie rizík a závislostí,
6. architektúru riešenia projektu na úrovni biznis vrstvy, aplikačnej vrstvy, dátovej vrstvy, technologickej vrstvy a bezpečnostnej architektúry,
7. vyhodnotenie alternatív riešenia projektu pre každú vrstvu architektúry riešenia,

8. špecifikáciu a klasifikáciu údajov spracovaných v projekte,
9. požiadavky na prevádzku a údržbu výstupov projektu,
10. požiadavky na technologickú infraštruktúru a posúdenie alternatív prevádzky infraštruktúry cloud computingom,
11. požiadavky na zdrojové kódy,
12. opis implementácie projektu a preberania výstupov projektu.

Súbežne sa vyhotovujú dokumenty I-04 Katalóg požiadaviek (súčasť M-05), M-05 Analýza nákladov a prínosov a M-06 Evidencia komponentov v MetalS.

2.1 Použité skratky a pojmy

Skratka/pojem	Popis
ABAC	Attribute-Based Access Control – riadenie prístupu na základe atribútov
AI	Artificial Intelligence – umelá inteligencia
ML	Machine Learning – strojové učenie
AI/ML	Systémy umelej inteligencie a strojového učenia
AIML4OS	European AI and Machine Learning Framework for Official Statistics
API	Application Programming Interface – aplikačné rozhranie
AS	aplikačná služba
BCR	Benefit-Cost Ratio – pomer prínosov a nákladov
CMS	Content Management System
CSAT	Customer Satisfaction Score – ukazovateľ spokojnosti používateľov
CSV	Comma Separated Values
DLP	Data Loss Prevention
DPIA	Data Protection Impact Assessment – posúdenie vplyvu na ochranu osobných údajov
EIRR	Economic Internal Rate of Return – ekonomická vnútorná výnosová miera
ENPV	Economic Net Present Value – ekonomická čistá súčasná hodnota
FIRR	Financial Internal Rate of Return – finančná vnútorná výnosová miera
FNPV	Financial Net Present Value – finančná čistá súčasná hodnota
GDPR	General Data Protection Regulation
HIL	Human-in-the-Loop – ľudská kontrola výstupov umelej inteligencie
IAM	Identity and Access Management
ISVS	informačný systém verejnej správy
JSON	JavaScript Object Notation
KPI	Key Performance Indicator – kľúčový ukazovateľ výkonnosti
KS	koncová služba
LLM	Large Language Model – veľký jazykový model
MFA	Multi-Factor Authentication – viacfaktorová autentifikácia

MLOps	Machine Learning Operations – správa životného cyklu modelov
NKIVS	Národná koncepcia informatizácie verejnej správy
NLG	Natural Language Generation
OAuth2	Open Authorization 2.0
OE	objekt evidencie
OpenAPI	Štandard na popis REST API
PII	Personally Identifiable Information
PIP	postimplementačná podpora
RBAC	Role-Based Access Control – riadenie prístupu na základe rolí
REST	Representational State Transfer
RPO	Recovery Point Objective – maximálna prípustná strata dát
RTO	Recovery Time Objective – cieľový čas obnovy
SDC	Statistical Disclosure Control – kontrola štatistického utajenia
SIEM	Security Information and Event Management – správa bezpečnostných informácií a udalostí
SLA	Service Level Agreement – dohoda o úrovni služieb
SLO	Service Level Objective
SQL	Structured Query Language
TLS	Transport Layer Security
UCP	Use Case Points – metóda odhadu prácnosti na základe prípadov použitia
WAF	Web Application Firewall
XAI	Explainable Artificial Intelligence – vysvetliteľná umelá inteligencia
CMP	Component – architektonický komponent platformy
UC	Use Case – scenár použitia
IN	Integration Point – integračný bod
DB	databáza
ŠÚ SR	Štatistický úrad Slovenskej republiky
UC-01	Interný AI asistent
UC-02	Automatická klasifikácia odpovedí respondentov
UC-03	Generovanie syntetických mikrodasetov
UC-04	Bezpečný AI analytik nad mikroúdajmi
UC-05	Verejný AI analytik
UC-06	Inteligentné vyhľadávanie štatistických údajov
UC-07	Automatické generovanie textových súhrnov
UC-08	Viacjazyčné publikovanie štatistických údajov

Tabuľka 1 Skratky a pojmy

2.2 Konvencie pre typy požiadaviek (príklady)

Hlavné kategórie požiadaviek v zmysle katalógu požiadaviek, rozdelíme na funkčné (funkcionálne), nefunkčné (kvalitatívne, výkonové a pod.). Podskupiny v hlavných kategóriách je možné rozšíriť podľa potrieb projektu, napríklad:

Funkcionálne (používateľské) požiadavky majú nasledovnú konvenciu:

FPID_xx

- FP – funkcionálna alebo používateľská požiadavka
- ID – identifikátor
- xx – číslo požiadavky

Nefunkčné (kvalitatívne, výkonové - Non Functional Requirements - NFR) požiadavky majú nasledovnú konvenciu:

NPID_xx

- NP – nefunkčná požiadavka
- ID – identifikátor
- xx – číslo požiadavky

Technické požiadavky majú nasledovnú konvenciu:

TPID_xx

- TP – technická požiadavka
- ID – identifikátor
- xx – číslo požiadavky

Ostatné typy požiadaviek môžu byť ďalej definované Objednávateľom/PM.

3. DEFINOVANIE PROJEKTU

3.1 Manažérske zhrnutie

Cieľom projektu je vybudovať robustnú, bezpečnú a škálovateľnú platformu umelej inteligencie pre Štatistický úrad Slovenskej republiky, ktorá zásadne modernizuje spracovanie štatistických údajov, zvyšuje kvalitu interných procesov a výrazne zlepšuje dostupnosť a interpretáciu oficiálnych štatistík pre verejnosť. Platforma umožní systematické využitie moderných technológií umelej inteligencie a strojového učenia pri dodržaní najvyšších štandardov ochrany údajov, štatistickej dôveryhodnosti a kybernetickej bezpečnosti, v súlade s odporúčaniami iniciatívy AIML4OS a aktuálnymi trendmi v Európskom štatistickom systéme. Navrhované riešenie reflektuje rastúci objem spracovávaných údajov, tlak na rýchlejšie zverejňovanie výsledkov a zvyšujúce sa očakávania verejnosti, médií a výskumného prostredia v oblasti dostupnosti a zrozumiteľnosti dát. Súčasné procesy spracovania štatistických údajov sú v mnohých prípadoch založené na manuálnych postupoch (napr. kódovanie voľných textových odpovedí, vyhľadávanie metodických informácií, príprava komentárov k tabuľkám), čo predstavuje významnú časovú záťaž, zvyšuje riziko nejednotnosti výsledkov a obmedzuje schopnosť úradu flexibilne reagovať na nové požiadavky používateľov dát. Projekt na tieto výzvy reaguje tým, že vytvára jednotný technologický a procesný rámec, v ktorom AI podporuje kľúčové časti dátového reťazca – od zberu a spracovania údajov až po ich interpretáciu a disemináciu.

Pre interné prostredie úrad získava súbor nástrojov, ktoré posilňujú analytické a metodické kapacity a znižujú závislosť od manuálnych činností. Interný AI asistent poskytuje zamestnancom okamžitú metodickú podporu v prirodzenom jazyku nad internými znalostnými databázami, čím zjednocuje aplikáciu interných predpisov a postupov a znižuje riziko interpretačných rozdielov medzi útvarmi. Automatická klasifikácia textových odpovedí výrazne skraca čas spracovania zisťovaní a prináša konzistentné kódovanie podľa štandardných klasifikácií (napr. SK NACE, ISCO), vďaka čomu je možné spracovať väčší objem otvorených odpovedí a publikovať detailnejšie štruktúrne štatistiky. Bezpečný AI analytik nad mikroúdajmi umožňuje štatistikom a analytikom klásť komplexné dotazy nad citlivými mikroúdajmi v chránenej infraštruktúre, pričom zabudované mechanizmy automaticky aplikujú pravidlá štatistického utajenia na všetky výstupy. Výsledkom je rýchlejšie získavanie analytických poznatkov bez kompromisov v oblasti dôveryhodnosti údajov. Tieto interné prínosy majú priamy dopad na kvalitu a dostupnosť služieb pre externé prostredie. Rýchlejšie spracovanie údajov, jednotnejšie uplatňovanie metodík a nižšia chybovosť vytvárajú predpoklad pre včasnejšie a spoľahlivejšie publikovanie štatistík, konzistentnejšie odpovede na dopyty médií a vyššiu dôveryhodnosť oficiálnych čísel v očiach verejnosti. Generovanie syntetických mikrodátsetov otvára nové možnosti pre akademickú obec, výskumné organizácie

a inovačný ekosystém – poskytuje im realistické, ale neidentifikovateľné dátové súbory, ktoré možno bezpečne využívať na testovanie hypotéz, porovnávanie metodík či vývoj analytických nástrojov bez zložitého procesu získavania prístupu k ostrým mikroúdajom. Tým sa posilňuje otvorená veda, zrýchľuje príprava výskumných projektov a zároveň ostáva plne zachovaný súlad s GDPR a štatistickou dôvernosťou.

Pre verejnosť, médiá a širšie analytické prostredie platforma podstatne znižuje bariéry pri práci s oficiálnymi štatistikami. Verejný AI analytik nad dátovými kockami umožní kľasť otázky v prirodzenom jazyku a okamžite získavať odpovede podložené publikovanými agregátmi, často doplnené o vizualizácie a odkazy na zdrojové tabuľky, bez potreby poznať štruktúru databáz či technické rozhrania. Inteligentné sémantické vyhľadávanie a odporúčanie štatistík výrazne zlepšuje dohľadateľnosť údajov a pomáha používateľom objavovať relevantné datasety, publikácie a analytické materiály, čím podporuje vyššiu návštevnosť portálu a intenzívnejšie využívanie údajov v rozhodovacích procesoch. Automaticky generované textové súhrny (NLG) pridávajú k číselným výstupom stručné a konzistentné komentáre, vďaka čomu sú štatistiky ľahšie interpretovateľné aj pre neodborné publikum a znižuje sa riziko nesprávneho výkladu údajov v mediálnych výstupoch. Viacjazyčná diseminácia zabezpečuje, že kľúčové výstupy budú dostupné aj v angličtine a ďalších jazykoch, čo odstraňuje jazykové bariéry pre zahraničných používateľov, posilňuje medzinárodnú reputáciu ŠÚ SR a zjednodušuje zapojenie slovenských dát do medzinárodných porovnaní.

Z pohľadu strategického smerovania by úrad získal nástroj, ktorý podporuje medzinárodnú spoluprácu a otvorenú vedu. Generovanie syntetických mikrodasetov a zavádzanie AI v procesoch klasifikácie, analýzy a diseminácie je v súlade s prioritami projektov ESSnet AIML4OS a iniciatívami UNECE a Eurostatu, ktoré podporujú zdieľanie know-how, modelov a dobrých praktík medzi národnými štatistickými inštitúciami. Platforma umožňuje ŠÚ SR aktívne sa zapájať do týchto diskusií, ponúkať vlastné skúsenosti (napr. s implementáciou AI pre jazykovo špecifické prostredie slovenčiny) a využívať zahraničné open-source nástroje a metodiky, čím sa posilňuje postavenie Slovenska v komunite oficiálnych štatistík.

Bezpečnosť a ochrana údajov predstavujú jeden z kľúčových pilierov navrhovanej architektúry a sú zároveň v súlade so strategickým zameraním projektu na posilnenie kybernetickej bezpečnosti pri spracovaní a sprístupňovaní štatistických údajov. Riešenie využíva dvojzónový bezpečnostný model, v ktorom sú citlivé mikroúdaje spracovávané výlučne v chránenej Zóne A v infraštruktúre ŠÚ SR, oddelenej od externých systémov a verejne dostupných služieb. Táto zóna je určená na spracovanie dôverných údajov a interných analytických procesov a podlieha najprísnejším bezpečnostným opatreniam. Verejné AI služby sú naopak prevádzkované v Zóne B, ktorá pracuje výhradne s publikovanými agregovanými údajmi, dátovými kockami, metaúdajmi, schválenými syntetickými datasetmi a verejným obsahom. Architektúra tak zabezpečuje, že citlivé mikroúdaje nikdy neopúšťajú zabezpečené prostredie úradu a nie sú priamo dostupné externým službám ani používateľom. Prestupy údajov medzi zónami sú striktné jednosmerné (z A do B) a obmedzené výlučne na publikované agregované údaje, dátové kocky, metaúdaje a schválené syntetické datasety, ktoré už prešli procesom metodologickej kontroly a schválenia na zverejnenie. Každý takýto prenos podlieha schválenému egress workflow, ktorý zahŕňa mechanizmy prevencie úniku dát (DLP), auditné logovanie a kontrolu podľa princípu „four eyes“. Tým sa zabezpečuje, že žiadne citlivé údaje nemôžu byť omylom alebo neoprávnene prenesené do verejnej časti systému. Architektúra zároveň uplatňuje princípy security by design a privacy by default, čo znamená, že bezpečnostné a ochranné mechanizmy sú integrované priamo do návrhu riešenia od jeho začiatku. Súčasťou riešenia je prísna klasifikácia dát podľa úrovne citlivosti (mikroúdaje, interné agregáty, publikované agregáty), centralizované riadenie identít a prístupov prostredníctvom IAM, využívanie modelov riadenia prístupov RBAC a ABAC, šifrovanie údajov počas prenosu aj uloženia, auditné logovanie všetkých kritických operácií a nasadenie ochranných mechanizmov na úrovni infraštruktúry a aplikačných rozhraní, ako sú webové aplikačný firewall a API brána. Dôležitou súčasťou bezpečnostného modelu je aj kontrola výstupov generovaných systémami umelej inteligencie. Všetky AI generované výstupy určené na publikovanie alebo externé využitie prechádzajú procesom Human in the Loop, v rámci ktorého ich kontrolujú oprávnení pracovníci ŠÚ SR. Tento proces zabezpečuje nielen metodickú správnosť a interpretáciu údajov, ale aj súlad s pravidlami ochrany dôverných informácií a legislatívnymi požiadavkami. Takto navrhnutá architektúra spája moderné analytické a AI nástroje s robustným bezpečnostným rámcom a podporuje strategický cieľ projektu – rozvoj dátovo orientovaných služieb pri súčasnom posilnení kybernetickej bezpečnosti, ochrany osobných údajov a dôvery verejnosti v oficiálnu štatistiku.

Realizácia projektu je plánovaná postupne v niekoľkých fázach. Úvodná pilotná fáza overí koncept vybraných verejných služieb v izolovanom prostredí verejného cloudu v rámci EÚ nad necitlivými, syntetickými a verejnými údajmi bez akéhokoľvek sieťového prepojenia na internú infraštruktúru ŠÚ SR. Na základe výsledkov pilotu budú komponenty nasadené do produkčného prostredia v dátovom centre ŠÚ SR v plnom dvojzónovom modeli, kde zabezpečia prevádzku všetkých interných aj verejných AI služieb s garantovanou rezidenciou citlivých údajov na území Slovenskej republiky. Projekt tak predstavuje nielen technologickú modernizáciu analytického prostredia ŠÚ SR, ale aj zásadný krok k tomu, aby oficiálne štatistiky boli pre verejnosť dostupnejšie, zrozumiteľnejšie a

využiteľnejšie, aby výskumné prostredie získalo bezpečnejšie možnosti práce s dátami a aby Slovenská republika posilnila svoju pozíciu v oblasti moderných a dôveryhodných štatistických služieb v európskom priestore.



Obrázok 1 Znáznornenie konceptu projektu

Obrázok znázorňuje koncept AI platformy ŠÚ SR, ktorá prepája interné spracovanie mikroúdajov, analytické nástroje umelej inteligencie, generovanie syntetických mikrodatabáz a verejné analytické služby nad publikovanými údajmi. Architektúra je založená na dvojzónovom bezpečnostnom modeli, ktorý oddeľuje spracovanie citlivých mikroúdajov v internej infraštruktúre od verejných analytických služieb.

3.2 Motivácia a rozsah projektu

Projekt reaguje na rastúcu potrebu modernizácie spracovania, analýzy a sprístupňovania štatistických údajov v prostredí Štatistického úradu SR. Súčasný stav je charakterizovaný vysokou mierou manuálneho spracovania údajov, najmä pri práci s voľnými textovými odpoveďami v štatistických zisťovaniach, kde je potrebné ich klasifikovať podľa štandardizovaných klasifikácií, ako sú ISCO alebo SK NACE. Tento proces je časovo náročný, personálne zaťažujúci a zvyšuje riziko nejednotnosti a chybovosti. Podobne aj pri internej analytickej práci s mikroúdajmi alebo pri tvorbe verejných výstupov ostáva značná časť činností závislá od manuálnych krokov, individuálnej expertízy a opakovanej metodologickej kontroly. Projekt zároveň reaguje na rastúce očakávania externého prostredia. Verejnosť, médiá, analytici, výskumné organizácie aj partneri verejnej správy očakávajú rýchlejší, jednoduchší a zrozumiteľnejší prístup k oficiálnym štatistikám. Hoci sú dnes mnohé údaje formálne dostupné, ich reálne využitie je často obmedzené tým, že používateľ musí rozumieť štruktúre dátových koviek, odbornej terminológii a spôsobu vyhľadávania v existujúcich databázach a publikačných rozhraniach. Významnou bariérou je preto nielen samotná dostupnosť dát, ale aj ich interpretovateľnosť, dohľadateľnosť a používateľská prístupnosť.

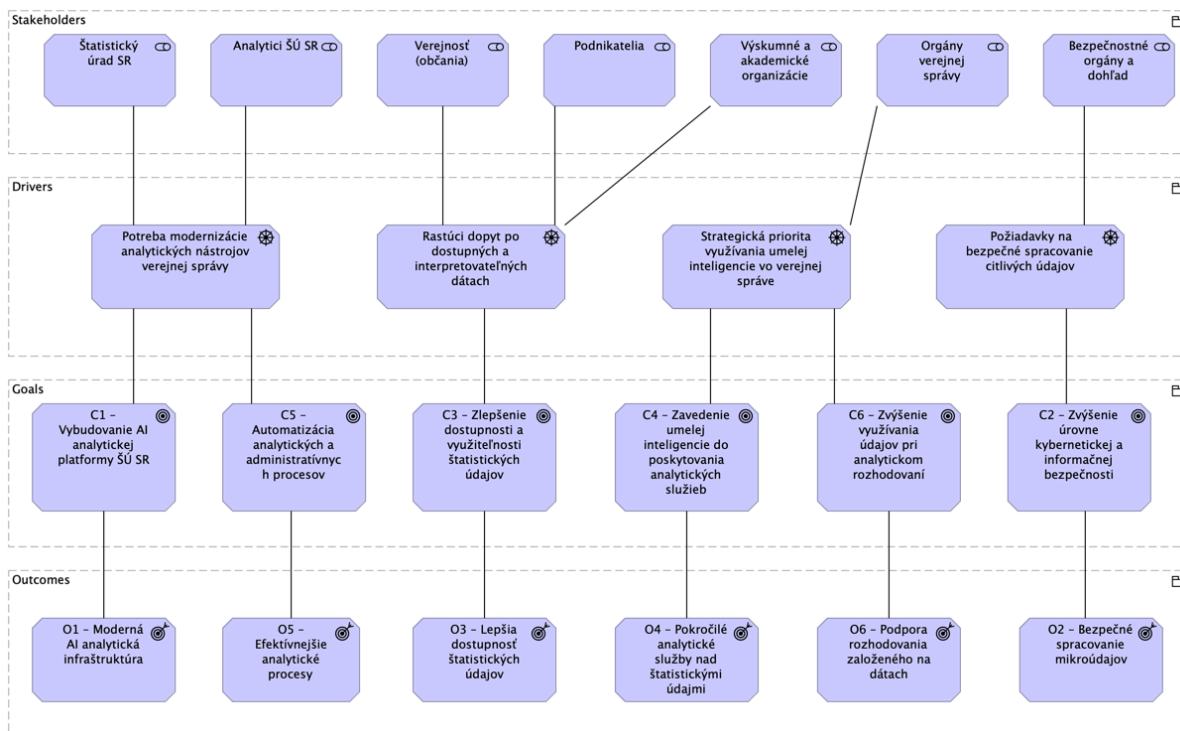
Osobitnú výzvu predstavuje bezpečné využívanie detailných štatistických údajov na analytické a výskumné účely. ŠÚ SR pracuje s citlivými mikroúdajmi, ktorých spracovanie a prípadné sprístupňovanie je viazané na prísne pravidlá štatistickej dôverylosti, ochrany osobných údajov a kybernetickej bezpečnosti. Tradičné prístupy k práci s mikroúdajmi sú preto nevyhnutne obmedzené a administratívne náročné. To znižuje flexibilitu analytickej práce a zároveň obmedzuje možnosti širšieho bezpečného využitia údajov vo výskumnom a inovačnom prostredí. Projekt preto navrhuje vytvorenie bezpečnej AI platformy, ktorá umožní systematické využitie umelej inteligencie v prostredí oficiálnej štatistiky a odstráni kľúčové identifikované limity súčasného stavu. Navrhované riešenie kombinuje automatizáciu vybraných interných procesov, rozšírenie analytických možností úradu, nové spôsoby interakcie

verejnosti s publikovanými údajmi a bezpečné formy sprístupňovania analyticky hodnotných dát. Projekt teda nerieši iba technologickú modernizáciu jednotlivých nástrojov, ale vytvára nový integrovaný rámec pre prácu s údajmi, ktorý prepája spracovanie dát, analytiku, bezpečnosť, publikovanie a používateľské služby do jedného konzistentného architektonického modelu.

Predmetom projektu sú vybrané scenáre použitia, ktoré predstavujú kľúčové oblasti využitia umelej inteligencie v prostredí oficiálnej štatistiky. Ide najmä o interný AI asistent pre metodickú podporu zamestnancov, automatickú klasifikáciu textových odpovedí zo štatistických zisťovaní, generovanie syntetických mikrodatabáz pre výskumné a testovacie účely, bezpečný AI analytický nástroj pre internú prácu s mikroúdajmi, verejného AI analytika nad publikovanými agregovanými údajmi, inteligentné vyhľadávanie a odporúčanie štatistických informácií, automatické generovanie textových interpretácií štatistických výstupov a viacjazyčnú disemináciu štatistických informácií. Tieto scenáre reprezentujú modernizačný rámec digitálnych služieb Štatistického úradu SR a reflektujú aktuálne trendy využívania umelej inteligencie v rámci európskeho štatistického systému. Z pohľadu rozsahu sa projekt týka najmä interných procesov spracovania, analýzy a interpretácie štatistických údajov v Štatistickom úrade SR, ako aj súvisiacich služieb pre externých používateľov. Do projektu vstupujú viaceré skupiny používateľov vrátane štatistikov, metodikov, analytikov a IT prevádzky úradu, ale aj výskumných organizácií, akademickej obce, médií, partnerov verejnej správy a širokej verejnosti. Riešenie bude integrované do existujúcich informačných systémov a dátových zdrojov úradu, najmä zberových systémov, databáz mikroúdajov, interných agregovaných databáz, publikovaných agregovaných databáz a dátových kociek, ako aj do existujúceho webového portálu a redakčného systému. Významnou vlastnosťou navrhovaného riešenia je práve jeho prepojenosť: jednotlivé služby nebudú fungovať izolovane, ale budú zdieľať spoločný dátový, bezpečnostný a integračný rámec, čo umožní konzistentné riadenie údajov, modelov aj používateľských výstupov.

Motiváciou projektu je zvýšiť mieru automatizácie spracovania údajov, skrátiť čas prípravy štatistických výstupov, zlepšiť dostupnosť a interpretáciu oficiálnych štatistických informácií a zároveň posilniť bezpečnosť a kontrolu nad prácou s citlivými údajmi. Projekt má zároveň ambíciu rozšíriť možnosti bezpečného využívania údajov vo výskumnom a analytickom prostredí. Kľúčovým prvkom v tomto smere je generovanie syntetických mikrodatabáz, ktoré predstavujú bezpečný kompromis medzi ochranou dôvernosti a snahou o širšie sprístupnenie analyticky hodnotných dát. Syntetické mikrodáta umožnia realizovať analytické experimenty, testovanie modelov, metodické overovanie alebo prípravu výskumných aktivít bez potreby práce s reálnymi identifikovateľnými mikroúdajmi mimo chráneného prostredia. Tým sa zvyšuje využiteľnosť údajov, podporuje sa otvorená veda a zároveň sa zachováva plná ochrana respondentov. Architektúra riešenia preto využíva zónový bezpečnostný model, v ktorom sú citlivé mikroúdaje spracovávané výhradne v chránenej internej zóne infraštruktúry Štatistického úradu SR. Táto zóna predstavuje bezpečnostne kontrolované prostredie pre všetky procesy pracujúce s mikroúdajmi, internými agregátmi, tréningovými dátami a modelmi, ktoré môžu prísť do kontaktu s citlivým obsahom. Verejné služby umelej inteligencie pracujú iba s publikovanými agregovanými údajmi a verejným obsahom a sú prevádzkované v oddelenej verejnej zóne bez prístupu k citlivým dátam. Bezpečnosť tu preto nie je len ochrannou vrstvou, ale základným princípom návrhu riešenia a zároveň podmienkou dôveryhodného nasadenia AI v prostredí oficiálnej štatistiky. Prenos údajov medzi jednotlivými zónami je prísne kontrolovaný a povolený len pre publikované agregované údaje, dátové kocky, metaúdaje a schválené syntetické databázy, pričom každý takýto prestup podlieha bezpečnostným kontrolám, auditovaniu, mechanizmom prevencie úniku dát a metodickému schvaľovaniu. Tým sa zabezpečuje, že architektúra projektu bude podporovať nielen nové služby, ale aj dlhodobú udržateľnosť a regulačný súlad riešenia. Súčasťou návrhu je aj uplatňovanie princípu Human-in-the-Loop, ktorý zabezpečuje, že všetky AI generované výstupy určené na publikovanie budú pred zverejnením kontrolované oprávneným pracovníkom. Tento princíp je dôležitý z pohľadu kvality, interpretácie aj dôvery verejnosti, keďže zaručuje, že AI bude použitá ako riadený podporný nástroj, nie ako autonómny nekontrolovaný zdroj verejných informácií.

Projekt sa venuje oblasti oficiálnej štatistiky a jej podporných analytických, diseminačných a metodických procesov. Z pohľadu biznis vrstvy ide najmä o procesy súvisiace so spracovaním štatistických zisťovaní, klasifikáciou údajov, analytickým využívaním mikroúdajov, tvorbou agregovaných výstupov, publikovaním štatistických informácií a ich sprístupňovaním verejnosti. Hoci projekt nie je orientovaný na jednu konkrétnu životnú situáciu v klasickom zmysle poskytovania transakčných elektronických služieb občanovi, jeho výstupy sa priamo dotýkajú oblasti poskytovania informácií a analytických služieb občanom, podnikateľom, výskumníkom a iným inštitúciám. Obmedzenia projektu vyplývajú najmä z potreby zachovať rezidenciu citlivých údajov v infraštruktúre ŠÚ SR na území Slovenskej republiky, z požiadaviek na štatistickú dôvernosť, z legislatívnych pravidiel ochrany osobných údajov a z nevyhnutnosti oddeliť interné spracovanie mikroúdajov od verejných služieb. Tieto obmedzenia však zároveň formujú architektúru riešenia a zabezpečujú, že modernizácia bude realizovaná spôsobom, ktorý je bezpečný, auditovateľný a kompatibilný s prostredím verejnej správy.



Obrázok 2 Motivácie pre realizáciu projektu

Motivačný pohľad architektúry projektu AI platformy pre Štatistický úrad SR zachytáva hlavné dôvody, strategické ciele a očakávané výsledky implementácie riešenia. Diagram vychádza z identifikácie kľúčových stakeholderov projektu, medzi ktorých patria samotný Štatistický úrad SR, jeho analytici, orgány verejnej správy, verejnosť, podnikateľský sektor, výskumné a akademické organizácie, ako aj bezpečnostné a dohľadové orgány. Každá z týchto skupín má odlišné potreby a očakávania od budúcej dátovej a analytickej infraštruktúry, čo sa následne premieta do identifikovaných motivačných faktorov (drivers) projektu.

Jedným z hlavných motivačných faktorov je potreba modernizácie analytických nástrojov verejnej správy. Štatistický úrad SR spracováva veľké množstvo štatistických údajov, ktorých objem, komplexnosť a frekvencia aktualizácie sa v posledných rokoch výrazne zvyšujú. Tradičné analytické nástroje a manuálne postupy už nedokážu efektívne reagovať na tieto požiadavky. Modernizácia analytickej infraštruktúry preto predstavuje základný predpoklad na zavedenie pokročilých analytických metód, využívanie umelej inteligencie a automatizáciu spracovania údajov. Tento motivačný faktor sa premieta najmä do cieľa vybudovať AI analytickú platformu a do cieľa automatizácie analytických a administratívnych procesov. Výsledkom je vznik modernej analytickej infraštruktúry a efektívnejšie analytické procesy.

Ďalším významným faktorom je rastúci dopyt po dostupných a interpretovateľných dátach. Štatistické údaje už dnes nevyužívajú len odborní analytici, ale aj široká verejnosť, podnikateľský sektor či akademická obec. Používatelia očakávajú nielen prístup k dátam, ale aj nástroje, ktoré im umožnia tieto údaje jednoducho vyhľadávať, analyzovať a interpretovať. AI platforma má preto umožniť poskytovanie nových typov služieb nad publikovanými údajmi, napríklad prostredníctvom inteligentného vyhľadávania alebo automaticky generovaných interpretácií dát. Tento motivačný faktor sa premieta do cieľa zlepšiť dostupnosť a využiteľnosť štatistických údajov, pričom výsledkom je lepšia dostupnosť štatistických údajov pre rôzne skupiny používateľov.

Významnú úlohu v návrhu architektúry zohráva aj strategická priorita využívania umelej inteligencie vo verejnej správe. Strategické dokumenty informatizácie verejnej správy zdôrazňujú potrebu využívať moderné analytické nástroje a umelú inteligenciu pri poskytovaní verejných služieb aj pri podpore rozhodovacích procesov. V kontexte štatistiky to znamená najmä využitie AI na pokročilú analýzu dát, automatizáciu spracovania textových odpovedí, podporu analytikov pri interpretácii údajov a poskytovanie inteligentných analytických služieb používateľom. Tento

motivačný faktor sa premieta do cieľov zavedenia umelej inteligencie do analytických služieb a zvýšenia využívania údajov pri analytickom rozhodovaní. Očakávaným výsledkom je vznik AI analytických služieb pre používateľov a zároveň podpora rozhodovania založeného na dátach.

Samostatný motivačný faktor predstavujú požiadavky na bezpečné spracovanie citlivých údajov. Štatistický úrad SR pracuje s citlivými mikroúdajmi o jednotlivcoch a podnikoch, ktoré podliehajú prísnyh pravidlám ochrany podľa legislatívy o ochrane osobných údajov a pravidiel štatistickej dôvernosti. Implementácia AI platformy preto musí byť navrhnutá tak, aby zabezpečila vysokú úroveň kybernetickej a informačnej bezpečnosti, vrátane riadenia prístupov, auditovateľnosti operácií, monitorovania bezpečnostných udalostí a oddelenia spracovania citlivých údajov od verejných analytických služieb. Kybernetická bezpečnosť preto nie je len technickým aspektom riešenia, ale predstavuje jeden zo základných architektonických princípov platformy. Tento motivačný faktor sa premieta do cieľa zvýšiť úroveň kybernetickej a informačnej bezpečnosti, pričom očakávaným výsledkom je bezpečné spracovanie mikroúdajov v prostredí platformy.

Vzťahy medzi jednotlivými motiváciami, cieľmi a výsledkami zároveň ilustrujú, že projekt AI analytickej platformy nie je izolovanou technologickou iniciatívou, ale súčasťou širšieho procesu modernizácie analytických kapacít verejnej správy. Modernizácia analytickej infraštruktúry, zavádzanie umelej inteligencie, zvyšovanie dostupnosti dát a posilňovanie kybernetickej bezpečnosti sú navzájom prepojené oblasti, ktoré spoločne prispievajú k efektívnejšiemu využívaniu štatistických údajov a k podpore rozhodovania založeného na dátach. Výsledný motivačný model tak poskytuje logické prepojenie medzi potrebami stakeholderov, strategickými prioritami informatizácie verejnej správy a konkrétnymi cieľmi projektu AI analytickej platformy. Diagram zároveň ukazuje, ako jednotlivé motivačné faktory vedú k definovaným cieľom projektu a aké konkrétne výsledky má implementácia platformy priniesť.

3.3 Zainteresované strany (Stakeholderi)

Id	Aktér / stakeholder	Subjekt (názov / skratka)	Rola
1	Štatistický úrad SR	ŠÚ SR	Vlastník projektu, vlastník dát, hlavný používateľ systému
2	Interní používatelia ŠÚ SR	ŠÚ SR	Primárni používatelia AI nástrojov pre analýzu dát, klasifikáciu odpovedí a tvorbu výstupov (štatistici, metodici, analytici)
3	Výskumníci/akademická obec	Univerzity, výskumné inštitúcie	Používatelia syntetických dát a analytických výstupov
4	Externí používatelia (verejnosť, médiá)	Občania SR, médiá	Používatelia publikovaných agregovaných štatistických údajov
5	Ministerstvo investícií, regionálneho rozvoja a informatizácie SR	MIRRI SR	Dohľad nad projektovým riadením a súladom s architektonickými pravidlami ISVS
6	Analytici orgánov verejnej správy	Ministerstvá, ostatné ústredné orgány štátnej správy, samosprávne kraje	Používatelia verejného AI analytika a inteligentného vyhľadávania pri príprave verejných politík a hodnotení dopadov so štatistickými údajmi ŠÚ SR

Tabuľka 2 Zainteresované strany (Stakeholderi)

3.4 Ciele projektu

ID	Názov cieľa	Názov strategického cieľa	Spôsob realizácie strategického cieľa
----	-------------	---------------------------	---------------------------------------

C1	Vybudovanie AI analytickej platformy ŠÚ SR	Modernizácia analytických nástrojov verejnej správy	Vybudovanie centralizovanej platformy umelej inteligencie pre spracovanie, analýzu a interpretáciu štatistických údajov vrátane podpory analytických procesov a sprístupnenia analytických služieb pre verejnosť.
C2	Zvýšenie úrovne kybernetickej a informačnej bezpečnosti pri spracovaní údajov	Strategická priorita NKIVS: Kybernetická a informačná bezpečnosť – komplexná ochrana digitálneho priestoru Cieľ NKIVS: Cieľ 2 – Bezpečné informačné systémy verejnej správy	Implementácia bezpečnej architektúry AI platformy vrátane oddelenia spracovania mikroúdajov a verejných služieb, riadenia prístupov, auditovateľnosti operácií, monitorovania bezpečnostných udalostí a ochrany citlivých štatistických údajov.
C3	Zlepšenie dostupnosti a využiteľnosti štatistických údajov pre občanov a podnikateľov	Strategická priorita NKIVS: Služby pre občanov a podnikateľov Cieľ NKIVS: Cieľ 3 – Fungujúca a používateľsky prívetivá verejná správa	Sprístupnenie moderných analytických nástrojov pre prácu s publikovanými štatistickými údajmi vrátane sémantického vyhľadávania, interaktívnej analýzy dát a automaticky generovaných interpretácií štatistických výstupov.
C4	Zavedenie umelej inteligencie do poskytovania verejných analytických služieb	Strategická priorita NKIVS: Služby pre občanov a podnikateľov Cieľ NKIVS: Cieľ 3 – Aktívne zavádzanie a využívanie umelej inteligencie vo verejnom záujme (AI asistenti v službách)	Implementácia AI analytických nástrojov a asistenta pre prácu s publikovanými štatistickými údajmi, ktorý umožní používateľom jednoduchšie vyhľadávanie, interpretáciu a analýzu dát.
C5	Automatizácia vybraných analytických a administratívnych procesov	Strategická priorita NKIVS: Digitálny úrad Cieľ NKIVS: Cieľ 3 – Aktívne zavádzanie a využívanie umelej inteligencie vo verejnom záujme (AI automatizácia administratívnych procesov)	Automatizácia spracovania textových odpovedí, klasifikácie údajov a podpory analytikov prostredníctvom AI nástrojov, čím sa znižuje manuálna záťaž pracovníkov a zrýchľuje spracovanie štatistických údajov.
C6	Zvýšenie využívania údajov pri analytickom rozhodovaní	Strategická priorita NKIVS: Využívanie hodnoty v údajoch Cieľ NKIVS: Cieľ 3 – Aktívne zavádzanie a využívanie umelej inteligencie vo verejnom záujme (dátová veda v rozhodovaní)	Využitie analytických modelov a nástrojov dátovej vedy pre podporu analytických procesov Štatistického úradu SR a pre lepšie sprístupnenie štatistických údajov pre výskumné a analytické účely.

Tabuľka 3 Ciele projektu

3.5 Merateľné ukazovatele (KPI)

ID	ID/Názov cieľa	Názov ukazovateľa (KPI)	Popis ukazovateľa	Merná jednotka	AS IS merateľné hodnoty (aktuálne)	TO BE Merateľné hodnoty (cieľové hodnoty)	Spôsob ich merania a Pozn.
KPI-01	C1 – Vybudovanie AI platformy	Počet AI služieb nasadených v produkcii	Počet produkčne nasadených služieb platformy UC-01 až UC-08: interný AI asistent, automatická klasifikácia textových odpovedí, generovanie syntetických mikrodasetov, bezpečný AI analytik nad mikroúdajmi, verejný AI analytik, inteligentné vyhľadávanie, automatické textové súhrny a viacjazyčné publikovanie	počet	0	8	Evidencia release artefaktov a prevádzkových runbookov.
KPI-02	C5 – Automatizácia procesov	Podiel automaticky klasifikovaných odpovedí	Podiel textových odpovedí (ISCO/ NACE...) spracovaných AI bez manuálnej klasifikácie	%	0 %	min. 70 %	Logy klasifikačného modulu + porovnanie s manuálnym zásahom.
KPI-03	C5 – Automatizácia procesov	Úspora času pri klasifikácii	Zníženie priemerného času spracovania klasifikácie oproti manuálnemu procesu	% / s	60,8 s na záznam	min. 50 %	Time tracking v pracovnom postupe + systémové metriky.
KPI-04	C6 – Dátami riadené rozhodovanie	Počet interných analytických úloh podporených AI	Počet interných analytických úloh, kde AI analytik asistoval (dotaz, návrh výpočtu, interpretácia)	počet/rok	0	min. 500	Logy AI analytika + tagging „AI-assisted“.
KPI-05	C3 – Služby pre občanov	Počet používateľov verejných AI služieb	Počet unikátnych používateľov verejného AI	počet/rok	0	min. 5 000	Web analytika + API gateway metriky.

			analytika / vyhľadavania				
KPI-06	C3 – Služby pre občanov	Počet dotazov nad publikovanými dátami	Počet dopytov verejnosti, výskumníkov a analytikov orgánov verejnej správy cez verejného AI analytika a inteligentné vyhľadávanie nad publikovanými agregátmi	počet/rok	0	min. 127 600	API gateway, logy vyhľadávacej vrstvy.
KPI-07	C3 – Používateľská prívetivosť	Spokojnosť používateľov s verejnými AI službami (CSAT)	Priemerné hodnotenie spokojnosti s vyhľadaním/ interpretáciou dát	1–5	N/A	min. 4,0/5	Krátky in-app dotazník (po interakcii).
KPI-08	C4 – AI asistenti v službách	Počet publikovaných interpretácií s HIL kontrolou	Počet textových interpretácií, ktoré prešli ľudskou kontrolou a boli publikované	počet/rok	0	min. 1 000	Workflow schvaľovania + audit trail.
KPI-09	C6 – Hodnota v údajoch	Počet datasetov dostupných pre AI analytiku	Počet publikovaných agregovaných datasetov/ metadát sprístupnených cez platformu	počet	baseline	+20	Dátový katalóg + evidované zdroje.
KPI-10	C6 – Hodnota v údajoch	Počet syntetických datasetov pre výskum	Počet syntetických mikrodátových datasetov poskytnutých pre výskumné účely	počet/rok	0	min. 10	Evidencia výstupov syntetizácie + schvaľovací proces.
KPI-11	C2 – Bezpečné ISVS	Podiel prístupov k citlivým dátam s MFA	Percento administrátorských a analytických prístupov do Zóny A chránených MFA	%	baseline	100 %	IAM logy.
KPI-12	C2 – Bezpečné ISVS	Auditovateľnosť operácií (log coverage)	Podiel kľúčových akcií (prístup k dátam, export, zmena modelu, publikovanie) pokrytých audit logom	%	baseline	100 %	SIEM/central logging kontrola.
KPI-13	C2 – Kyberbezpečnosť	Priemerný čas reakcie na incident	Čas od detekcie po	hod.	baseline	≤ 1 hod.	Incident management + SIEM.

		v rámci platformy	triáž/eskaláciu incidentu				
KPI-14	C2 – Kontinuita	Periodické testy obnovy a záloh	Počet vykonaných a zdokumentovaných testov obnovy ročne	počet/rok	baseline	min. 2	DR test reporty.
KPI-15	C3 – Lokalizácia	Podiel verejných výstupov dostupných v EN	Podiel verejných interpretácií a UI prvkov dostupných v EN	%	0 %	min. 90 %	Prekladový coverage report + QA checklist.
KPI-16	C5 – Zručnosti	Počet interných používateľov zaškolených na AI platformu	Počet zamestnancov ŠÚ SR preškolených na používanie platformy a bezpečné používanie AI	počet/rok	0	min. 100	Prezenčné listiny/e-learning.

Tabuľka 4 Merateľné ukazovatele (KPI)

3.6 Špecifikácia potrieb koncového používateľa

Projekt zahŕňa vývoj a rozvoj digitálnych služieb, ktoré budú využívané rôznymi skupinami koncových používateľov, vrátane zamestnancov verejnej správy, výskumných organizácií, médií a širokej verejnosti. Projekt je zároveň navrhnutý tak, aby pri napĺňaní potrieb používateľov striktne dodržiaval požiadavky na štatistickú dôvernosť, ochranu osobných údajov a kybernetickú bezpečnosť v súlade s rezidenčným a zónovým modelom platformy (Zóna A pre citlivé mikroúdaje, Zóna B pre publikované agregáty). Návrh používateľských služieb je realizovaný v súlade s požiadavkami vyhlášky č. 547/2021 Z. z. o elektronizácii agendy verejnej správy a metodikou tvorby používateľsky kvalitných elektronických služieb.

Pri návrhu funkcionalít platformy sa vychádzalo z identifikácie potrieb jednotlivých skupín používateľov a z analýzy ich typických scenárov používania. Návrh rozhrania a funkcionalít platformy reflektuje princípy používateľsky orientovaného dizajnu (user-centered design), pričom zohľadňuje používateľské cesty, očakávané interakcie so systémom a požiadavky na prístupnosť a použiteľnosť služieb. Zároveň sa pri návrhu interakcií systematicky uplatňujú bezpečnostné a ochranné princípy – najmä princíp najmenších oprávnení, oddelenie interných a verejných prostredí, povinná auditovateľnosť operácií a HIL (Human-in-the-loop) pri výstupoch určených na zverejnenie. Používateľské požiadavky boli identifikované na základe analýzy existujúcich procesov spracovania štatistických údajov, konzultácií s odbornými pracovníkmi Štatistického úradu SR a analýzy potrieb externých používateľov štatistických informácií. Výsledkom tejto analýzy je identifikácia viacerých skupín koncových používateľov, ktorých potreby a očakávania sú zohľadnené pri návrhu riešenia a následne premietnuté do požiadaviek.

Skupiny koncových používateľov

Interní analytici a metodici Štatistického úradu SR sú odborní pracovníci zodpovední za spracovanie štatistických údajov, analýzu dát a prípravu štatistických výstupov. Táto skupina používateľov pracuje priamo s mikroúdajmi a analytickými nástrojmi a vyžaduje pokročilé funkcionality pre prácu s dátami, automatizáciu klasifikácie údajov a metodickú podporu pri interpretácii výsledkov. Platforma ich potreby napĺňa najmä prostredníctvom use casov UC-01 (interný AI asistent), UC-02 (automatická klasifikácia textových odpovedí), UC-03 (generovanie syntetických mikrodatabáz) a UC-04 (bezpečný AI analytik nad mikrodátami), ktoré sú implementované v chránenej Zóne A. Interný AI asistent pracuje nad internou dokumentáciou, metodikami a znalostnou bázou úradu v internej zóne a ostatné tri use casey pracujú nad citlivými mikroúdajmi. Všetky dotazy a operácie nad mikrodátami sú plne auditované, prechádzajú definovanými mechanizmami štatistického utajenia a sú v súlade s bezpečnostnými požiadavkami katalógu požiadaviek na klasifikáciu a rezidenciu údajov a riadenie prestupov medzi zónami.

IT pracovníci a správcovia systémov zabezpečujú prevádzku a správu platformy, vrátane monitorovania jej dostupnosti, správy prístupov a bezpečnostných mechanizmov. Ich požiadavky sa týkajú najmä prevádzkovej stability systému, auditovateľnosti operácií, možností správy používateľských oprávnení a schopnosti efektívne reagovať na incidenty a zmeny konfigurácie. Platforma preto poskytuje nástroje pre správu identít a oprávnení (IAM), monitorovanie a observabilitu (logy, metriky, tracing), správu DLP a egress kontrol, ako aj jasne definované prevádzkové a bezpečnostné politiky v oboch zónach. Tieto potreby sú pokryté najmä nefunkčnými požiadavkami na dostupnosť, obnovu, observabilitu a bezpečnosť, pričom všetky kľúčové operácie majú auditnú stopu a sú riadené procesmi SRE, incident/problem management a change management.

Výskumné organizácie a akademická obec využívajú štatistické údaje na výskumné a analytické účely. Ich potreby zahŕňajú najmä dostupnosť analyticky hodnotných datasetov a možnosť experimentovania s dátami bez rizika porušenia štatistickej dôvernosti alebo ochrany osobných údajov. Platforma preto poskytuje mechanizmy generovania syntetických mikroúdajov vhodných na výskumné účely (UC-03), pričom generovanie prebieha výlučne v Zóne A nad pôvodnými mikroúdajmi a výsledné syntetické datasety sú klasifikované a sprístupňované v Zóne B len vtedy, ak spĺňajú definované metodické a bezpečnostné kritériá (utility, riziko reidentifikácie, aplikované pravidlá štatistického utajenia). Proces generovania a publikovania syntetických dát je viazaný na bezpečnostné požiadavky katalógu požiadaviek na klasifikáciu a rezidenciu údajov a riadenie prestupov medzi zónami, vrátane formálneho schvaľovania a auditovateľnosti prestupov medzi zónami.

Médiá a analytické organizácie patria medzi významných používateľov štatistických údajov. Ich potreby spočívajú najmä v rýchlom prístupe k aktuálnym štatistickým údajom a v jednoduchej interpretácii dátových výstupov bez potreby detailnej znalosti štruktúry dátových koviek alebo štatistických metodík. Platforma pre túto skupinu poskytuje AI analytika nad publikovanými dátovými kockami (UC-05), inteligentné sémantické vyhľadávanie na webe (UC-06) a automatické generovanie textových súhrnov (UC-07), ktoré sú prevádzkované výhradne v Zóne B nad publikovanými agregátmi a verejným obsahom. Žiadny z týchto nástrojov nemá prístup k mikroúdajom ani interným agregátom. V Zóne B platí zákaz spracovania PII a všetky výstupy sú jednoznačne späté so zdrojovými publikovanými datasetmi. Textovo generované výstupy (komentáre a súhrny) zároveň podliehajú povinnému HIL procesu – t. j. ľudskej kontrole a schváleniu pred zverejnením – v súlade s definíciou HIL v architektúre riešenia.

Široká verejnosť využíva štatistické údaje najmä na získavanie informácií o ekonomických a spoločenských trendoch, regionálnych rozdieloch alebo vývoji vybraných ukazovateľov. Používatelia z tejto skupiny spravidla nemajú pokročilé analytické znalosti, preto je dôležité zabezpečiť intuitívne rozhranie, ktoré umožňuje jednoduché vyhľadávanie štatistických informácií, ich zrozumiteľnú interpretáciu a prezentáciu v kontexte. Potreby tejto skupiny sú napĺňané najmä prostredníctvom verejného AI analytika nad dátovými kockami (UC-05), inteligentného vyhľadávania a odporúčaní (UC-06) a textových súhrnov (UC-07), pričom všetky tieto funkcie pracujú výhradne s publikovanými agregátmi v Zóne B. Všetky verejné výstupy generované AI, vrátane vysvetľujúcich textov a viacjazyčných verzií, prechádzajú HIL kontrolou pred zverejnením, čím sa zabezpečuje vecná správnosť, terminologická konzistentnosť a minimalizácia rizika nesprávnej interpretácie dát.

Používateľské potreby a ciele

Potreby jednotlivých skupín používateľov boli transformované do používateľských scenárov, ktoré opisujú očakávané využitie elektronických služieb. Tieto scenáre zodpovedajú use casom UC-01 až UC-08 a ich požiadavky sú v katalógu požiadaviek zaradené do príslušných modulov, čo umožňuje sledovanie plnenia potrieb používateľov počas návrhu, implementácie aj testovania.

Príklad používateľských scenárov:

Interný analytik - Ako analytik Štatistického úradu SR potrebujem automaticky klasifikovať textové odpovede zo štatistických zisťovaní, aby som znížil čas potrebný na manuálne spracovanie údajov a zvýšil konzistentnosť výsledkov. Tento scenár je pokrytý use casom UC-02, pričom spracovanie prebieha výlučne nad mikroúdajmi v Zóne A a všetky operácie sú vedené v súlade s bezpečnostnými požiadavkami (rezidencia dát, klasifikácia dát, zákaz presunu mikroúdajov mimo Zónu A).

Výskumník - Ako výskumník potrebujem pracovať s datasetmi, ktoré majú analytickú hodnotu, ale neobsahujú osobné údaje, aby som mohol realizovať výskum bez porušenia pravidiel ochrany údajov. Tento scenár je pokrytý

use casom UC-03. Platforma generuje syntetické mikroúdaje v Zóne A a po metodickom a bezpečnostnom posúdení ich publikuje ako bezpečné syntetické datasey v Zóne B. Celý proces je viazaný na bezpečnostné požiadavky a podlieha formálnemu schvaľovaniu vrátane posúdenia rizika reidentifikácie a uplatnenia pravidiel štatistického utajenia.

Novinár - Ako novinár potrebujem rýchlo nájsť relevantné štatistické údaje a ich interpretáciu, aby som mohol pripraviť článok alebo analýzu bez potreby hlbokej znalosti štatistických databáz. Tento scenár je pokrytý use casmi UC-05 (verejný AI analytik nad dátovými kockami), UC-06 (inteligentné vyhľadávanie) a UC-07 (textové súhrny), ktoré sú implementované v Zóne B nad publikovanými agregátmi. Platforma zabezpečuje transparentnosť zdrojov (odkazy na zdrojové datasetové kocky, verzie dát) a uplatňuje povinný HIL proces – generované textové komentáre a vysvetlenia sú pred publikovaním vždy skontrolované a schválené odborným pracovníkom ŠÚ SR, čím sa znižuje riziko nesprávnej alebo zavádzajúcej interpretácie.

Občan - Ako občan potrebujem jednoducho vyhľadať štatistické informácie o vybranej téme a pochopiť ich význam, aby som získal prehľad o spoločenských alebo ekonomických trendoch. Tento scenár je pokrytý najmä use casmi UC-05 a UC-06, pričom používateľské rozhranie je navrhnuté tak, aby umožňovalo zadávanie otázok v prirodzenom jazyku a získanie odpovedí v zrozumiteľnej forme, doplnenej o jednoduché vizualizácie a krátke textové vysvetlenia. Z pohľadu bezpečnosti a ochrany údajov je dôležité, že občan pracuje výhradne s publikovanými agregátmi v Zóne B, ktoré neobsahujú osobné údaje ani kombinácie vedúce k zvýšenému riziku reidentifikácie, a všetky AI-generované texty prechádzajú HIL kontrolou.

Meranie napĺňania potrieb používateľov

Úspešnosť elektronických služieb bude hodnotená prostredníctvom merateľných indikátorov používania a spokojnosti používateľov. Medzi základné sledované indikátory budú patriť napríklad:

- čas potrebný na nájdenie relevantnej informácie,
- miera dokončenia používateľských scenárov (napr. úspešné dokončenie prieskumu, úspešné nájdenie a zobrazenie dátového výstupu),
- počet využitií jednotlivých služieb platformy (napr. počet dotazov do AI analytika, počet vyhľadávaní, počet generovaných súhrnov),
- spokojnosť používateľov so službou meraná prostredníctvom používateľskej spätnej väzby (dotazníky, ratingy, kvalitatívna spätná väzba).

Tieto indikátory budú technicky realizované kombináciou webovej a aplikačnej analytiky a observability – napr. prostredníctvom logov, metrík a tracovania v rámci jednotného monitoringu platformy, doplneného o nástroje webovej analytiky (napr. meranie času do prvej relevantnej odpovede, počtu krokov potrebných na dokončenie scenára). Pri meraní správania používateľov budú spracúvané len agregované a prípadne pseudonymizované údaje o používaní služieb, bez identifikácie jednotlivých osôb, v súlade s GDPR a bezpečnostnou architektúrou platformy. Identifikátory používateľských relácií budú navrhnuté tak, aby podporovali zlepšovanie používateľskej skúsenosti, nie sledovanie jednotlivcov. Tieto indikátory budú využívané na priebežné vyhodnocovanie kvality poskytovaných služieb a na identifikáciu oblastí, v ktorých je potrebné zlepšiť používateľskú skúsenosť, dostupnosť alebo výkon. Výsledky merania budú vstupom do procesov kontinuálneho zlepšovania (continuous improvement) – napr. úprav používateľského rozhrania, zmien v konfigurácii modelov AI, aktualizácie obsahových šablón alebo doplnenia vysvetľujúcich textov. Výsledné požiadavky identifikované na základe analýzy potrieb používateľov budú zapracované do katalógu požiadaviek projektu v dokumente I-04 Katalóg požiadaviek (kde budú evidované ako funkčné, integračné a bezpečnostné požiadavky s väzbou na konkrétne use casey a stakeholderov). Tým sa zabezpečí, že investície do AI platformy budú priamo viazané na hodnotu pre koncových používateľov a zároveň zostanú v súlade s požiadavkami na bezpečnosť, ochranu súkromia a rezidenciu dát na území SR.

3.7 Detailný opis obmedzení a predpokladov

Rozsah projektu je vymedzený na návrh a implementáciu platformy umelej inteligencie, ktorá podporí spracovanie, analýzu a disemináciu štatistických údajov v prostredí Štatistického úradu Slovenskej republiky. Projekt sa zameriava na vybrané scenáre využitia umelej inteligencie identifikované v analytickej fáze projektu a nepokrýva komplexnú modernizáciu všetkých existujúcich informačných systémov Štatistického úradu SR. Cieľom projektu je vytvoriť technologickú platformu, ktorá umožní postupné rozširovanie funkcionalít umelej inteligencie

bez potreby zásadných zásahov do existujúcich systémov. Jedným z hlavných obmedzení projektu je požiadavka na zachovanie rezidencie citlivých štatistických údajov na území Slovenskej republiky. Citlivé mikroúdaje a údaje podliehajúce štatistickej dôvernosti musia byť spracovávané výlučne v infraštruktúre dátového centra Štatistického úradu SR. Z tohto dôvodu architektúra riešenia využíva zónový model, v ktorom sú komponenty pracujúce s citlivými údajmi umiestnené v chránenej internej zóne infraštruktúry, zatiaľ čo verejné služby umelej inteligencie pracujú výhradne s publikovanými agregovanými údajmi, dátovými kockami, metaúdajmi a schválenými syntetickými datasetmi.

Ďalším významným obmedzením je legislatívny rámec ochrany osobných údajov a štatistickej dôvernosti. Platforma musí byť navrhnutá a prevádzkovaná v súlade s požiadavkami GDPR, zákona o štátnej štatistike a interných bezpečnostných politík Štatistického úradu SR. Tieto požiadavky limitujú spôsob spracovania údajov, prístup k mikroúdajom a možnosti ich sprístupňovania externým subjektom. Rozsah projektu je zároveň ohraničený tým, že riešenie sa primárne zameriava na analytické a diseminačné služby nad existujúcimi dátovými zdrojmi. Projekt nepredpokladá zásadnú zmenu architektúry existujúcich zberových systémov ani ich kompletnú náhradu. Integrácia platformy bude realizovaná prostredníctvom definovaných integračných rozhraní a dátových tokov. Pre úspešnú realizáciu projektu je potrebné splnenie viacerých technických, organizačných a legislatívnych predpokladov. Na technickej úrovni je nevyhnutné zabezpečiť dostupnosť infraštruktúrnych kapacít v dátovom centre Štatistického úradu SR vrátane výpočtových zdrojov, úložných kapacít a bezpečnostných mechanizmov potrebných na prevádzku AI platformy. Rovnako je potrebné zabezpečiť pripravenosť integračných bodov na napojenie platformy na existujúce databázy mikroúdajov, publikované databázy a ďalšie interné systémy.

Z organizačného hľadiska je predpokladom úspešnej realizácie projektu dostupnosť odborných kapacít na strane Štatistického úradu SR, ktoré budú schopné participovať na analýze požiadaviek, validácii riešenia a následnej prevádzke platformy. Dôležitým predpokladom je aj zabezpečenie spolupráce medzi analytickými, metodickými a IT útvarmi úradu. Z pohľadu bezpečnosti a ochrany údajov je nevyhnutné zaviesť mechanizmy klasifikácie údajov, riadenia prístupov a auditu spracovania údajov. Súčasťou implementácie platformy musí byť aj nastavenie procesov riadenia modelov umelej inteligencie, monitorovania ich výkonu a zabezpečenia transparentnosti ich fungovania. Ďalším predpokladom je zabezpečenie súladu riešenia s európskymi iniciatívami a metodickými rámcami využívania umelej inteligencie v oblasti oficiálnej štatistiky, najmä v rámci Európskeho štatistického systému. Tieto rámce poskytujú metodické usmernenia pre bezpečné a transparentné využívanie umelej inteligencie pri práci so štatistickými údajmi. Splnenie uvedených predpokladov umožní implementovať AI platformu ako bezpečný a škálovateľný nástroj podporujúci modernizáciu procesov Štatistického úradu SR a zvyšujúci dostupnosť štatistických informácií pre rôzne skupiny používateľov.

Kľúčové predpoklady realizácie AI platformy:

- Citlivé štatistické dáta sa musia uchovávať a spracúvať výhradne na infraštruktúre ŠÚ SR na území SR. Necitlivé/verejné časti sa prevádzkujú ako verejné služby vo verejnej zóne dátového centra ŠÚ SR bez presakovania citlivých dát. Pravidlá klasifikácie dát a povolených tokov medzi internou a verejnou zónou musia byť dodržané pri všetkých aktivitách projektu.
- Riešenie počíta s napojením na existujúce dátové kocky a agregované databázy ŠÚ SR. Verejné komponenty pracujú len s publikovanými agregovanými údajmi, dátovými kockami, metaúdajmi a schválenými syntetickými datasetmi.
- Z hľadiska bezpečnosti a súladu musia riešenia a procesy rešpektovať dôvernosť štatistických údajov a GDPR. Žiadne citlivé dáta nesmú byť prenesené mimo infraštruktúry ŠÚ SR (napr. do verejného cloudu).

3.8 Vyhodnotenie rizík a závislostí

ID	Názov rizika / závislosti	Kategória rizika	Potenciálny dopad	Opatrenia na zmiernenie rizika (mitigácia)
R1	Nebude možné naplniť všetky kvalitatívne požiadavky projektu	B2	Nebudú plne dosiahnuté očakávané benefity projektu.	Zorganizovať viacero kôl iniciačných stretnutí k definovaniu požiadaviek z katalógu požiadaviek s dodávateľom.

R2	Nedostatočná kompatibilita komponentov platformy	B2	Nekompatibilita komponentov internej a verejnej zóny môže spôsobiť dodatočné úpravy, predĺženie implementácie a zvýšenie nákladov.	Dôkladná kontrola plnenia požiadaviek pri dodávaní produktov vo fáze Analýza a dizajn.
R3	Nedostupnosť komponentov novej infraštruktúry	C2	Nedostupnosť komponentov novej infraštruktúry, prípadne ich veľmi dlhé dodacie lehoty môžu mať negatívny vplyv na termín ukončenia projektu.	Zistiť dostupnosť komponentov a dobu dodania.
R4	Projekt nebude realizovaný a nasadený podľa plánu.	C1	V prípade omeškania dodávok alebo nasadenia výstupov projektu nebude možné v plánovanom termíne sprístupniť interné a verejné AI služby platformy a prínosy projektu sa začnú prejavovať neskôr, ako predpokladá analýza nákladov a prínosov.	Čo najskoršia realizácia implementácie. Dodržiavanie jednotlivých fáz projektu, hlavne plánu nasadenia.
R5	Komplikácie s verejným obstarávaním	B2	V prípade neskorého vypísania VO, prípadne komplikácií v procese VO by bol ohrozený začiatok implementačnej fázy.	Počas prípravnej a iniciačnej fázy zapojiť útvar verejného obstarávania do prípravy súťažných podkladov. Cieľom je informovať príslušných pracovníkov s predstihom o predmete obstarávania. Pozitívne sa to odrazí na príprave podkladov VO, z ktorých časť môžu zodpovedné osoby pripravovať paralelne a nemusia sa zoznamovať s obsahom projektu až po oficiálnom ukončení iniciačnej fázy.
R6	Neúplné požiadavky	B2	Neúplné požiadavky môžu spôsobiť predĺženie trvania projektu, navýšenie nákladov, z dôvodu potreby dodatočného obstarávania komponentov.	Zabezpečiť kvalitné podklady v procese obstarania a tvorbe manažérskych produktov. Dodat' čo najpodrobnejší popis zákazky.
R7	Vysoké náklady na prevádzku	C3	Náklady na prevádzku budú vyššie ako plánované. Prekročenie plánovaných nákladov na prevádzku. Potreba dodatočných finančných zdrojov.	Priebežne kontrolovať prevádzkové náklady voči plánu a zmluve o podpore prevádzky a venovať zvýšenú pozornosť dodávateľským zmluvám.

R8	Ohrozenie prevádzky a dostupnosti systému	C2	Ohrozenie prevádzky a dostupnosti systému z dôvodu možných kybernetických útokov.	Vykonať bezpečnostné testovanie pred nasadením a zabezpečiť priebežný bezpečnostný monitoring prevádzky.
R9	Nedostatočné vyhodnotenie kvality	C2	Neodhalenie slabých miest v jednotlivých fázach implementácie projektu.	Zabezpečiť proces a postupy kvality tak, aby boli pre projekt dostatočné.
R10	Nesúčinnosť a nespoľahlivosť dodávateľa	B2	Predčasné ukončenie projektu, alebo predĺženie doby realizácie projektu.	Definovanie sankcií v rámci dodávateľskej zmluvy.
R11	Nedostatok ľudských zdrojov	B2	Predĺženie doby realizácie projektu. Výstupy projektu budú dodané v nedostatočnej kvalite.	Včas vyčleniť odborné kapacity ŠÚ SR pre projekt a zabezpečiť ich zastupiteľnosť.
R12	Únik alebo presakovanie citlivých dát zo Zóny A	C2	Porušenie štatistickej dôvernosti a GDPR, reputačné a právne následky, strata dôvery verejnosti.	Implementácia DLP mechanizmov, audit logov, kontrola egress komunikácie, pravidelné bezpečnostné testovanie a monitoring exportov dát.
R13	Nedostatočná klasifikácia dát a chybné označenie dátových tokov	C2	Zvýšené riziko reidentifikácie alebo neúmyselného zverejnenia citlivých údajov.	Zavedenie jednotného katalógu dát, validácia klasifikácie datasetov, pravidelné audity dátových tokov.
R14	Slabé riadenie prístupov (RBAC/ABAC) a privilegované účty	C2	Neautorizovaný prístup k citlivým údajom, zmena konfigurácie systému alebo exfiltrácia dát.	Zavedenie princípu least-privilege, MFA, pravidelný audit oprávnení, monitoring privilegovaných operácií.
R15	Neúplné uplatnenie princípov štatistického utajenia	C2	Riziko odvodenia identifikácie jednotlivcov alebo publikácie citlivých kombinácií údajov.	Implementácia kontrol prahových hodnôt, metodická kontrola agregácií, pravidelné testovanie anonymizačných mechanizmov.
R16	Nedostatočná kvalita dát a metaúdajov	B2	Nesprávne analytické výsledky, chybné odpovede AI systému a znížená dôveryhodnosť publikovaných údajov.	Zavedenie procesov riadenia kvality dát, validácia metaúdajov, pravidelné kontroly konzistencie datasetov.
R17	Zlyhanie alebo degradácia AI modelov (model drift)	C2	Pokles presnosti modelov, nesprávne klasifikácie alebo zavádzajúce analytické výstupy.	Monitoring výkonu modelov, pravidelné preškolenie modelov, validácia výsledkov pred publikovaním.
R18	Nedostatočná kapacita infraštruktúry a výkonové limity	C2	Zvýšené latencie služieb, výpadky systému alebo nedodržanie SLA.	Kapacitné plánovanie infraštruktúry, monitoring výkonu, implementácia škálovania infraštruktúry.

R19	Zraniteľnosti alebo nesprávne konfigurácie integračných rozhraní	C2	Zneužitie API rozhraní, útoky na dostupnosť služby alebo únik metaúdajov.	Implementácia bezpečnostných testov, TLS komunikácie, rate limiting a pravidelný bezpečnostný audit.
R20	Procesné a organizačné riziká (change management, SRE)	B2	Predĺžené výpadky systému alebo neefektívne riešenie incidentov.	Zavedenie procesov change managementu, definovanie runbookov, implementácia monitoringu a incident managementu.
R21	Riziká súladu a auditovateľnosti	B2	Negatívne auditné zistenia, riziko sankcií alebo obmedzenia prevádzky systému.	Zavedenie auditných logov, dokumentovanie procesov, realizácia DPIA a pravidelných bezpečnostných auditov.
R22	Riziká používania verejných AI výstupov (NLG / preklady)	B2	Publikovanie nesprávnych alebo zavádzajúcich informácií, reputačné škody.	Uplatnenie princípu Human-in-the-Loop, kontrola výstupov AI pred publikovaním.
R23	Nízka adopcia nástrojov používateľmi	B2	Nízke využitie systému a nedosiahnutie plánovaných prínosov projektu.	Školenia používateľov, metodická podpora a postupné zavádzanie nástrojov do pracovných procesov.
R24	Závislosť od externých komponentov a licencovania	B3	Prevádzkové riziko alebo dodatočné náklady v prípade zmeny licenčných podmienok.	Preferovanie otvorených štandardov, sledovanie životného cyklu komponentov a plánovanie aktualizácií.
R25	Komunikačné a reputačné riziká pri publikovaní údajov	B2	Zníženie dôvery verejnosti v publikované štatistiky.	Zavedenie metodických kontrol publikovaných výstupov a jednotnej komunikačnej stratégie.
R26	Vendor lock-in	B3	Závislosť na jednom dodávateľovi alebo technológii a obmedzenie možnosti migrácie riešenia.	Používanie otvorených štandardov, dokumentovaných API a zmluvných garancií prenositeľnosti riešenia.
R27	Regulačný drift	B2	Nesúlad riešenia s novými legislatívnymi požiadavkami.	Pravidelné legislatívne revízie, implementácia procesu riadenia zmien.
R28	Závislosť od dostupnosti infraštruktúrnych kapacít v dátovom centre ŠÚ SR	C2	Realizácia a prevádzka platformy závisí od dostupnosti výpočtových zdrojov, úložných kapacít a bezpečnostných mechanizmov v dátovom centre ŠÚ SR. Ich nedostupnosť by oddialila nasadenie platformy.	Overiť a rezervovať potrebné kapacity dátového centra počas analýzy a dizajnu.
R29	Závislosť od sprístupnenia údajov z existujúcich databáz ŠÚ SR	C2	Bez sprístupnenia údajov z databáz mikroúdajov a publikovaných databáz ŠÚ SR nebude možné naplniť znalostnú vrstvu	Počas analýzy a dizajnu identifikovať dátové zdroje a dohodnúť spôsob a termíny ich sprístupnenia.

R30	Závislosť od dostupnosti odborných kapacít ŠÚ SR	C2	a analytické služby platformy.	Nedostatočná súčinnosť metodikov a IT pracovníkov ŠÚ SR pri špecifikácii, testovaní a akceptácii predži realizáciu projektu.	Určiť kľúčových používateľov a ich zastupiteľnosť v projektovom tíme a plánovať ich zapojenie v harmonograme projektu.
R31	Závislosť od centrálnej správy identít ŠÚ SR	C2		Riadenie prístupov k službám platformy závisí od dostupnosti a pripravenosti centrálnej správy identít ŠÚ SR (AD/LDAP).	Overiť pripravenosť centrálnej správy identít ŠÚ SR počas analýzy a dizajnu.

Tabuľka 5 Zoznam rizík a závislostí projektu

3.9 Detailný opis rozpočtu projektu a jeho prínosov

3.9.1 Sumarizácia nákladov a prínosov

Sumarizácia nákladov:

Náklady na realizáciu projektu v rokoch 2027 a 2028 predstavujú podľa analýzy nákladov a prínosov 8 637 433 € s DPH. Ročné prevádzkové náklady predstavujú 337 016 € s DPH a vznikajú od roku 2029.

Typ aktivity	Oblasť výdavku	Suma	OPEX / CAPEX	Int / ext	Suma	Zdroj financovania - %						
						Suma %	Vlastné OPEX	Vlastné CAPEX	ESIF OPEX	ESIF CAPEX	POO OPEX	POO CAPEX
Hlavné etapy	Vývoj aplikácií	3,201,899 €	OPEX	Externé	- €							
			CAPEX	Externé	2,786,234 €	100%			100%			
			OPEX	Interné	415,665 €	100%						
	Nákup HW a SW	4,849,726 €	OPEX	Externé	- €							
			CAPEX	Externé	4,849,726 €	100%				100%		
			OPEX	Interné	- €							
Prevádzka	Aplikácie	1,114,494 €	OPEX	Externé	557,247 €	100%	100%					
			CAPEX	Externé	557,247 €	100%		100%				
			OPEX	Interné	- €							
	HW a SW	1,581,632 €	OPEX	Externé	969,945 €	100%	100%					
			CAPEX	Externé	611,687 €	100%		100%				
			OPEX	Interné	- €							
Podporné oblasti	Projektový manažment	163,152 €	OPEX	Externé	0 €							
			OPEX	Interné	163,152 €	100%			100%			
			OPEX	Externé	- €							
	Publicita	422,656 €	OPEX	Externé	- €							
			OPEX	Interné	6,526 €	100%			100%			
			OPEX	Externé	- €							
Ostatné výdavky			Ostatné zmiešané výdavky		416,130 €	100%			100%			
Výstupné náklady		0 €		Externé	- €							
				Interné	- €							
SPOLU		11,333,559 €		-	11,333,559 €							

Sumarizácia nákladov a prínosov riešenia za obdobie 10 rokov:

TO BC - AS IS (€, SUM)													
	Spolu	Asistent pre zamestnancov	Klasifikácia sedových odpovedí	Analýza nad mikroslužbami	Governance syntetických údajov	Vývojný analýzy	Inteligentné vyhľadávacie	Automatické testové účiny	Vieciročné publikovanie	Znaková a vyhľadávacia vlna	Správa modelov a kalibrácia	Kontrolovaný prenos medzi záznammi	Bezpečnosť, prístup a audit
Náklady (DPH)	10,456,815 €	625,681 €	514,718 €	541,798 €	383,441 €	1,033,361 €	373,809 €	305,053 €	352,401 €	3,604,610 €	783,131 €	590,234 €	711,816 €
Výrobcový materiál	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €
IT - CAPEX	8,074,274 €	460,150 €	378,285 €	401,048 €	281,805 €	751,222 €	271,393 €	224,154 €	268,985 €	2,904,085 €	584,025 €	451,399 €	558,872 €
Aplikácie	3,343,481 €	324,839 €	378,285 €	356,768 €	281,805 €	303,322 €	271,393 €	224,154 €	239,405 €	271,393 €	289,157 €	208,359 €	164,502 €
SW	3,670,123 €	- €	- €	- €	- €	7,380 €	- €	- €	- €	2,530,212 €	284,888 €	96,340 €	250,202 €
HW	1,759,250 €	144,320 €	- €	44,280 €	- €	400,520 €	62,320 €	- €	29,520 €	31,460 €	336,200 €	127,100 €	141,450 €
IT - OPEX	1,786,783 €	123,888 €	110,078 €	112,671 €	82,002 €	185,597 €	78,972 €	65,238 €	75,583 €	520,870 €	134,529 €	107,880 €	117,978 €
Aplikácie	972,912 €	94,524 €	110,078 €	103,815 €	82,002 €	88,363 €	78,972 €	65,238 €	69,681 €	78,972 €	87,051 €	66,450 €	47,868 €
SW	613,687 €	- €	- €	- €	- €	1,230 €	- €	- €	- €	423,202 €	47,478 €	15,990 €	41,820 €
HW	200,184 €	29,364 €	- €	9,856 €	- €	96,194 €	12,464 €	- €	5,504 €	18,698 €	47,480 €	24,430 €	28,290 €
Riadenie projektu	580,808 €	33,134 €	26,357 €	28,080 €	19,635 €	56,543 €	23,444 €	15,621 €	18,833 €	179,664 €	62,570 €	30,975 €	36,967 €
Výstupné náklady	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €
Prínosy	19,653,196 €	10,005,580 €	1,141,763 €	3,474,432 €	1,017,900 €	2,348,690 €	1,664,832 €	- €	- €	- €	- €	- €	- €
Finančné prínosy	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €
Administratívne poplatky	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €
Ostatné darčové a neďarčové príjmy	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €
Ekonomické prínosy	19,653,196 €	10,005,580 €	1,141,763 €	3,474,432 €	1,017,900 €	2,348,690 €	1,664,832 €	- €	- €	- €	- €	- €	- €
Občania (O)	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €
Úradníci (E)	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €	- €
Úradníci (FE)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Kvalitatívne prínosy	19,653,196 €	10,005,580 €	1,141,763 €	3,474,432 €	1,017,900 €	2,348,690 €	1,664,832 €	- €	- €	- €	- €	- €	- €

Pomer prínosov a nákladov (BCR): 1,42

Rozpočet projektu: 8 637 433 € s DPH

Celkové náklady na zabezpečenie riešenia boli vypočítané prostredníctvom UCP a priemerom troch ponúk.

907 - Paušálna sadzba na nepriame výdavky podľa článku 54 písm. a) NSU: 585 808,35 € s DPH

Prevádzkové náklady/rok: 337 016 € s DPH

Prínosy

Navrhovaný projekt AI platformy pre Štatistický úrad SR prináša súbor vzájomne prepojených prínosov, ktoré sa prejavujú vo vnútornom fungovaní organizácie, v kvalite a rýchlosti spracovania štatistických údajov, v bezpečnosti práce s citlivými dátami a zároveň aj vo vzťahu k externému prostrediu – teda voči verejnosti, médiám, výskumným organizáciám, akademickej obci, partnerom a iným inštitúciám verejnej správy. Prínosy projektu preto nemožno redukovať iba na technologickú modernizáciu infraštruktúry. Ich podstata spočíva v tom, že vytvorenie centralizovanej AI analytickej platformy umožní ŠÚ SR systematicky meniť spôsob spracovania, interpretácie a diseminácie štatistických údajov a zároveň vytvárať nové typy služieb, ktoré v súčasnom stave nie je možné efektívne poskytovať.

Kvantifikovateľné prínosy projektu

Kvantifikovateľné prínosy projektu sa budú prejavovať predovšetkým v oblasti časových úspor, zvýšenia produktivity, poklesu manuálnej záťaže, zlepšenia kvality a presnosti výstupov a vo vyššej miere využívania poskytovaných služieb. Tieto prínosy je možné vyjadriť prostredníctvom KPI a prevádzkových metrík, ktoré budú monitorované počas pilotnej aj produkčnej prevádzky. V internej časti riešenia sa očakáva najmä skrátenie času spracovania vybraných analytických a metodických úloh. Interný AI asistent umožní rýchlejšie vyhľadávanie metodických odpovedí a pracovných postupov, čím sa zníži čas potrebný na orientáciu v internej dokumentácii. V praxi to znamená, že pracovníci budú vedieť získať odpoveď na štandardizované otázky v priebehu sekúnd až minút namiesto dlhšieho manuálneho vyhľadávania v dokumentoch. Merateľným ukazovateľom v tejto oblasti môže byť priemerný čas na získanie relevantnej metodической odpovede, podiel úspešne vyriešených dopytov bez eskalácie na kolegu alebo počet využití asistenta v porovnaní s paralelnými manuálnymi kanálmi. Pri automatickej klasifikácii textových odpovedí sa kvantifikovateľný prínos prejaví predovšetkým v poklese manuálneho objemu práce a v zrýchlení spracovania zberov. Namiesto plne manuálneho kódovania voľných textových odpovedí bude možné určitý podiel odpovedí klasifikovať automaticky alebo poloaautomaticky s následnou revíziou. Tu možno sledovať napríklad presnosť top-1 a top-5 klasifikácie, podiel odpovedí vybavených bez dodatočného manuálneho zásahu, čas spracovania dávky oproti východiskovému stavu alebo pokles chybovosti pri opakovanom kódovaní. Pri bezpečnom AI analytikovi nad mikroúdajmi možno očakávať zrýchlenie ad-hoc analytických úloh a zníženie bariéry prístupu k interným analýzám. Kým v súčasnosti si mnohé analýzy vyžadujú explicitnú prácu s databázami, ručné písanie dotazov a osobitné uplatnenie pravidiel utajenia, nový nástroj umožní analytikom formulovať otázky prirodzenejšie a získať agregovaný výstup v kratšom čase. Kvantifikácia sa tu môže opierať o priemerný čas vybavenia definovanej sady dotazov, počet úspešne vybavených analytických požiadaviek, počet výstupov, pri ktorých bolo automaticky uplatnené štatistické utajenie, alebo počet prípadov, v ktorých systém zablokoval rizikový výstup. Pri syntetických mikroúdatoch je kvantifikovateľný prínos spojený s rozšírením bezpečného prístupu k analyticky využiteľným dátam. Ukazovateľmi tu môžu byť počet vytvorených syntetických datasetov, počet oprávnených používateľov alebo inštitúcií, ktoré ich využijú, počet analytických alebo testovacích scenárov realizovaných nad syntetickými dátami a rozdiel medzi časom potrebným na začatie práce so syntetickým datasetom a časom potrebným na tradičné sprístupnenie reálnych mikroúdajov.

Vo verejnej časti riešenia sa kvantifikovateľné prínosy prejavujú najmä v zvýšení využívania služieb a v znížení bariér prístupu k štatistickým údajom. Pri verejnom AI analytikovi bude možné sledovať počet dotazov, podiel úspešne vybavených požiadaviek, latenciu odpovedí, mieru eskalácií na manuálnu podporu alebo počet používateľov, ktorí úspešne získali požadovaný údaj bez potreby ďalšej navigácie portálom. Pri inteligentnom vyhľadávaní a odporúčaní budú relevantné ukazovatele ako úspešnosť nájdenia relevantného obsahu, pokles počtu nulových výsledkov, zvýšenie CTR odporúčaných položiek a dĺžka používateľskej relácie na portáli. Pri automatických textových súhrnoch možno kvantifikovať najmä úsporu redakčného času a rozšírenie počtu datasetov, ku ktorým je možné poskytnúť sprievodný komentár. Relevantným ukazovateľom bude podiel AI generovaných textov, ktoré prejdú ľudskou revíziou s minimálnymi úpravami, priemerný čas prípravy textu pred a

po zavedení riešenia a počet publikovaných textových súhrnov za definované obdobie. Pri viacjazyčnej diseminácii bude možné merať najmä čas potrebný na sprístupnenie cudzojazyčnej verzie po publikovaní slovenskej verzie, počet publikovaných viacjazyčných výstupov a mieru potrebných zásahov pri ľudskej revízii prekladov.

V analýze nákladov a prínosov sú kvantifikovateľné prínosy vyčíslené ako úspora pracovného času. Úspora času zamestnancov ŠÚ SR a analytikov orgánov verejnej správy je ocenená osobnými nákladmi zamestnanca verejnej správy vo výške 22,62 € za hodinu, ktoré vychádzajú z priemernej mesačnej mzdy vo verejnej správe a odvodov zamestnávateľa. Úspora času verejnosti a výskumníkov je ocenená priemernou hodinovou mzdou v národnom hospodárstve vo výške 9,31 € za hodinu, ktorá vychádza z priemernej mesačnej mzdy za rok 2025 podľa ŠÚ SR. Pri službách, pri ktorých sa po zavedení riešenia zvýši počet používateľov, sa uplatňuje pravidlo polovice, teda prírastok používateľov sa započítava polovičnou váhou. Prínosy sa prejavujú po skončení realizácie projektu, od roku 2029 do konca referenčného obdobia. Sú zoskupené do šiestich oblastí podľa toho, komu a pri akej činnosti úspora vzniká, a spolu predstavujú 2 456 650 € ročne.

Úspora času zamestnancov pri vyhľadávaní informácií a príprave textov

Zamestnanci ŠÚ SR dnes pri spracovaní zisťovaní, analytickej práci a príprave výstupov opakovane dohľadávajú metodické pokyny, interné postupy a predchádzajúce rozhodnutia v dokumentoch a smerniciach. Odpovede na štandardné otázky často závisia od skúseností konkrétnych kolegov. Komentáre a interpretácie k publikovaným tabuľkám pripravujú pracovníci ručne a ich rozsah je obmedzený kapacitami redakcie. Cudzojazyčné verzie výstupov vznikajú ručným prekladom. Interný AI asistent odpovedá zamestnancom v prirodzenom jazyku nad znalostnou bázou úradu a pri odpovedi odkazuje na zdrojový dokument. Pracovník tak namiesto prehľadávania dokumentov dostane odpoveď v priebehu sekúnd až minút. Platforma zároveň pripravuje návrhy interpretácií k publikovaným údajom a strojový preklad výstupov do angličtiny, ktoré pracovník pred publikovaním kontroluje a upravuje. Výpočet vychádza zo 500 zamestnancov, ktorí budú asistenta pravidelne používať, a z úspory 25 minút za pracovný deň počas 230 pracovných dní v roku. Hodnota úspory je pod výsledkom pilotu Microsoft 365 Copilot britskej vlády s 20 000 štátnymi zamestnancami, ktorý uvádza 26 minút denne, a nad výsledkom pilotu britského ministerstva práce s 19 minútami denne. K tomu sa pripočítava úspora 4 hodín na každú z 1 000 interpretácií ročne, čo zodpovedá cieľovej hodnote ukazovateľa KPI-08, a úspora 2,25 hodiny na každý z 1 500 výstupov publikovaných v cudzom jazyku. Ročná úspora predstavuje 55 292 hodín, teda 1 250 698 €.

Úspora času pri spracovaní zisťovaní a ochrane dôverných údajov

Voľné textové odpovede respondentov, najmä opis povolania a ekonomickej činnosti, dnes pracovníci kódujú ručne do klasifikácií ISCO a SK NACE. Ide o rutinnú činnosť s veľkým objemom, pri ktorej sa výsledok môže líšiť podľa pracovníka. Pred zverejnením alebo poskytnutím výstupov z internej zóny pracovníci pri každom výstupe ručne kontrolujú dodržanie pravidiel štatistického utajenia. Klasifikačný model automaticky priradí kód väčšine odpovedí a pracovník posudzuje len prípady, pri ktorých model nedosiahne požadovanú istotu, pričom mu systém ponúkne návrh kódov. Pri prestupe výstupov zo Zóny A do Zóny B sa pravidlá štatistického utajenia, napríklad minimálne veľkosti buniek a potláčanie nízkych početností, uplatňujú automaticky a pracovník výsledok kontroly potvrdzuje.

Objem kódovania vychádza zo zisťovaní v domácnostiach. Výberové zisťovanie pracovných síl pracuje so vzorkou 10 250 bytov štvrťročne, zisťovanie rodinných účtov s 8 450 domácnosťami a EU-SILC s 11 986 osobami vo veku 16 a viac rokov. Pri dvoch kódoch na zamestnanú osobu, povolanie a ekonomickej činnosti, predstavuje objem približne 131 865 kódovaných záznamov ročne. Ručné kódovanie jedného záznamu trvá 60,8 sekundy a kódovanie s návrhom systému 38 sekúnd podľa štúdie OPERAS (Santé publique France, 2025). Pri 70 % automaticky klasifikovaných odpovedí podľa ukazovateľa KPI-02 sa ušetrí 1 809 hodín ročne. K tomu sa pripočítava úspora 1,5 hodiny na každú z 3 000 kontrol výstupov pri prestupe medzi zónami. Ročná úspora predstavuje 6 309 hodín, teda 142 720 €.

Úspora času pri analytickej práci s mikroúdajmi

Analýzy nad mikroúdajmi dnes vyžadujú ručné písanie dotazov a štatistického kódu, prácu s viacerými databázami a následné ručné uplatnenie pravidiel štatistického utajenia pri každom výstupe. Príprava analytického výstupu pre interné potreby, ministerstvá alebo médiá tak trvá niekoľko pracovných dní. Sprístupnenie mikroúdajov pre výskum vyžaduje ručnú prípravu anonymizovaného súboru a údaje zo systémov úradu sa na analytické

spracovanie pripravujú manuálnymi exportmi. Bezpečný AI analytik umožní analytikovi formulovať otázku v prirodzenom jazyku, vygeneruje dotaz a štatistický kód a vráti agregovaný výstup, ktorý už prešiel automatizovaným štatistickým utajením. Čas na prípravu analytického výstupu sa tak skraca z dní na hodiny. Syntetické datasety pre výskum generuje model a pracovník kontroluje ich kvalitu a riziko odhalenia. Údaje zo systémov úradu sú pre analytické spracovanie dostupné cez integrácie platformy. Výpočet vychádza z 500 analytických úloh ročne podporených AI podľa ukazovateľa KPI-04, pri ktorých sa čas prípravy skraca zo 4 pracovných dní na 4 hodiny. Ďalej z 10 syntetických datasetov ročne podľa ukazovateľa KPI-10, pri ktorých sa čas prípravy skraca zo 160 na 40 hodín, a z 2 000 exportov údajov ročne, pri ktorých sa čas prípravy skraca z 3 hodín na pol hodiny. Ročná úspora predstavuje 19 200 hodín, teda 434 304 €.

Úspora času pracovníkov informačných služieb pri vybavovaní dopytov verejnosti

Verejnosť, médiá a podnikatelia dnes posielajú požiadavky na štatistické údaje e-mailom alebo telefonicky. Pracovníci informačných služieb a hovorca úradu ich vybavujú ručne, pričom veľkú časť tvoria bežné dopyty na základné údaje, ktoré sú už publikované, no používateľ ich nevie nájsť. Verejný AI analytik odpovedá na otázky v prirodzenom jazyku nad publikovanými údajmi nepretržite a s odkazom na zdrojovú tabuľku. Bežné dopyty tak používateľ vybaví sám a na pracovníka úradu sa obracia len so zložitejšími požiadavkami. Výpočet vychádza z 11 250 dopytov ročne, z ktorých polovicu vybaví verejný AI analytik bez eskalácie na pracovníka, a z jednej hodiny potrebnej na vyhľadanie údajov, prípravu a odoslanie odpovede. Ročná úspora predstavuje 5 625 hodín, teda 127 238 €. Počet dopytov a podiel vybavený bez pracovníka sa spresnia podľa evidencie ŠÚ SR a výsledkov pilotnej fázy. Napĺňanie prínosu sa bude sledovať počtom dopytov evidovaných informačnými službami pred zavedením riešenia a po ňom a mierou eskalácií z verejného AI analytika na pracovníka.

Úspora času verejnosti a výskumníkov pri vyhľadávaní a získavaní údajov

Používateľ, ktorý hľadá štatistický údaj, dnes musí poznať štruktúru dátových kociek, odbornú terminológiu a spôsob filtrovania a vyhľadanie na portáli je založené na kľúčových slovách. Výskumníci získavajú dôverné údaje na vedecké účely na základe zmluvy a pracujú s nimi v bezpečnej miestnosti úradu, pričom získanie prístupu trvá jeden až dva mesiace. Verejný AI analytik a sémantické vyhľadanie umožnia nájsť údaj a jeho výklad na základe otázky v prirodzenom jazyku. Syntetické datasety sú dostupné bez zmluvného prístupu do bezpečnej miestnosti, takže ich môžu využiť výskumníci, univerzity, firmy aj mimovládne organizácie na prípravu analýz a overenie analytických postupov. Výpočet vychádza z rastu počtu dotazov verejnému AI analytikovi z 30 000 na 50 000 ročne s úsporou 24 minút na dotaz. Pri raste počtu dotazov sa uplatňuje pravidlo polovice, a preto sa započítava 40 000 dotazov ročne, čo predstavuje 16 000 hodín. Ďalej vychádza z 50 000 vyhľadávaní ročne, teda z 5 000 používateľov verejných AI služieb podľa ukazovateľa KPI-05 s desiatimi vyhľadávaniami, s úsporou 10 minút na vyhľadávanie. Vyhľadávania sa započítavajú celé, čo predstavuje 8 333 hodín. Napokon vychádza z rastu počtu výskumníkov, ktorí pracujú so syntetickými údajmi, zo 100 na 300 s úsporou 36 hodín na získanie údajov. Aj pri raste počtu výskumníkov sa uplatňuje pravidlo polovice, a preto sa započítava 200 výskumníkov, čo predstavuje 7 200 hodín. Ročná úspora predstavuje 31 533 hodín, teda 293 586 €, a je ocenená priemernou hodinovou mzdou v národnom hospodárstve. Cieľových 50 000 dotazov verejnému AI analytikovi a 50 000 vyhľadávaní spolu s 27 600 dopytmi analytikov orgánov verejnej správy z prínosu 6 tvorí cieľovú hodnotu ukazovateľa KPI-06.

Úspora času analytikov orgánov verejnej správy pri získavaní a interpretácii štatistických údajov

Analytici ministerstiev, ostatných ústredných orgánov štátnej správy a samosprávnych krajov pracujú so štatistikami ŠÚ SR pri príprave verejných politík a hodnotení dopadov. Údaje dnes hľadajú v dátových kockách a publikáciách alebo o ne žiadajú úrad, čo pri špecifických otázkach vyžaduje prácu analytikov ŠÚ SR. Cez verejného AI analytika a inteligentné vyhľadanie získajú analytici orgánov verejnej správy požadovaný údaj, jeho výklad a odkaz na zdrojovú tabuľku priamo, bez potreby poznať štruktúru databáz úradu. Výpočet vychádza z 300 analytikov orgánov verejnej správy, z ktorých každý položí dva dopyty týždenne počas 46 pracovných týždňov, a z úspory 20 minút na dopyt. Ide o 27 600 dopytov ročne a ročnú úsporu 9 200 hodín, teda 208 104 €. Úspora je ocenená osobnými nákladmi zamestnanca verejnej správy.

Prínosy pre verejnú správu, spoločnosť a externé prostredie

Osobitne dôležité je, že projekt neprináša prínosy len samotnému Štatistickému úradu SR, ale aj širšiemu verejnému prostrediu. Pre občanov a podnikateľov bude hlavným prínosom jednoduchší prístup k údajom, ktoré

môžu využiť pri orientácii v ekonomických a sociálnych témach. Pre médiá sa zlepší dostupnosť dôveryhodných a rýchlo dohľadateľných štatistických informácií, čo znižuje riziko chýb v interpretácii a podporuje fakticky podloženú verejnú debatu. Pre výskumné a akademické inštitúcie projekt vytvára predpoklad pre bezpečnejší a flexibilnejší prístup k údajom a analytickým nástrojom. Pre orgány verejnej správy sa zvyšuje dostupnosť dátovo podložených výstupov, ktoré možno využiť pri rozhodovaní, príprave verejných politík a hodnotení dopadov. V tomto zmysle projekt nevytvára len jednorazový prínos vo forme nasadenia technológie, ale zavádza dlhodobý udržateľný rámec, v ktorom bude možné priebežne merať a preukazovať zlepšenie efektívnosti, bezpečnosti, kvality a dostupnosti štatistických služieb. Práve táto kombinácia kvalitatívnych a kvantifikovateľných prínosov predstavuje kľúčový argument pre realizáciu projektu a jeho priamu väzbu na strategické ciele modernizácie verejnej správy, kybernetickej bezpečnosti, digitálneho úradu a využívania hodnoty v údajoch.

Kvalitatívne prínosy projektu

Jedným z najvýznamnejších kvalitatívnych prínosov projektu je modernizácia analytického prostredia Štatistického úradu SR. Projekt vytvára technologický a procesný základ, vďaka ktorému bude možné zaviesť umelú inteligenciu do oblastí, kde boli doteraz pracovníci odkázaní najmä na manuálne spracovanie, vlastnú expertízu alebo nejednotné pracovné postupy. Tento posun má zásadný význam najmä v prostredí oficiálnej štatistiky, kde rastie objem údajov, zvyšuje sa tlak na rýchlosť spracovania a zároveň sa očakáva vysoká kvalita, reprodukovateľnosť a metodická konzistentnosť výstupov.

Projekt zároveň prináša významné posilnenie kybernetickej a informačnej bezpečnosti pri práci s citlivými údajmi. Bezpečnosť v tomto projekte nie je vedľajšou technickou vlastnosťou, ale jedným zo základných architektonických princípov. Zónový model riešenia, oddelenie Zóny A a Zóny B, prísne riadenie prístupov, auditovateľnosť operácií, mechanizmy DLP a schvaľovanie prestupov medzi zónami vytvárajú rámec, v ktorom možno využívať AI aj pri práci s mikroúdajmi bez toho, aby sa oslabili požiadavky na štatistickú dôvernú, GDPR alebo vnútornú bezpečnostnú pravidlá. Význam tohto prínosu je dvojitý: na jednej strane chráni samotný úrad a citlivé údaje, na druhej strane posilňuje dôveru používateľov a respondentov, že aj pri zavádzaní nových technológií ostáva ochrana údajov zachovaná.

Osobitný kvalitatívny prínos predstavuje zvýšenie konzistentnosti a metodickkej jednotnosti interných procesov. Interný AI asistent, automatická klasifikácia textových odpovedí a bezpečný AI analytik nad mikroúdajmi znižujú mieru závislosti od individuálnej expertízy konkrétnych pracovníkov a podporujú jednotné uplatňovanie metodických pravidiel naprieč organizáciou. To má pozitívny dopad nielen na vnútornú efektívnosť, ale aj na kvalitu publikovaných výstupov, pretože externé publikum napokon vníma výsledky práce úradu ako jednotný celok. Ak sú metodické interpretácie a analytické postupy konzistentnejšie interne, zvyšuje sa aj konzistentnosť voči verejnosti, médiám a partnerom.

Ďalším dôležitým kvalitatívnym prínosom je lepšia dostupnosť a zrozumiteľnosť štatistických údajov pre externé prostredie. V súčasnom stave sú mnohé publikované údaje síce formálne dostupné, no ich využitie predpokladá znalosť štruktúry dátových kociek, terminológie a spôsobu vyhľadávania. Navrhované riešenie tento problém rieši prostredníctvom verejného AI analytika, inteligentného vyhľadávania, automaticky generovaných textových súhrnov a viacjazyčnej diseminácie. Výsledkom bude prístup k oficiálnym štatistikám, ktorý je bližší potrebám používateľa: nevyžaduje takú vysokú mieru technickej alebo analytickej gramotnosti a umožňuje pracovať s údajmi prostredníctvom prirodzeného jazyka, zrozumiteľných vysvetlení a odporúčaného obsahu.

Z hľadiska spoločenského dopadu projekt prispieva aj k posilneniu dátovej gramotnosti a verejnej informovanosti. Ak sa štatistické údaje sprístupnia vo forme, ktorá je prístupnejšia pre nešpecialistov, zvyšuje sa pravdepodobnosť ich využitia v médiách, vo verejnej debate, v rozhodovaní organizácií aj v osobnej orientácii občanov. Projekt teda neprináša iba „lepšie IT riešenie“, ale aj nový spôsob, ako sa oficiálne štatistiky môžu dostať do spoločenskej diskusie v podobe, ktorá je použiteľná, pochopiteľná a dôveryhodná.

Dôležitou kvalitatívnou oblasťou je aj podpora otvorenej vedy a výskumného prostredia. Generovanie syntetických mikrodátových súhrnov vytvára možnosť, ako sprístupniť analyticky hodnotné dátové zdroje bez porušenia ochrany súkromia alebo štatistickej dôvernosti. Ide o významný posun oproti tradičnému modelu, v ktorom sú detailné dáta dostupné len v silne obmedzenom režime. Syntetické dáta, ak budú metodicky správne navrhnuté a validované, môžu urýchliť prípravu analýz, testovanie hypotéz, experimentovanie s metodikami a vývoj nových analytických nástrojov. To má pozitívny dopad na univerzity, výskumné inštitúcie aj inovačné prostredie. Riešenia

tohto typu sa overujú aj v zahraničí. Napríklad Švajčiarsko testuje Statbot, ktorý prekladá otázky do SQL nad otvorenými dátami, MMF vyvíja StatGPT na dotazovanie sa na makroštatistiky a viaceré krajiny v rámci UNECE experimentujú s chatbotmi nad štatistickými databázami. ŠÚ SR prinesie vlastnú skúsenosť (napr. pre jazykovo špecifické výzvy slovenčiny) a zároveň sa môže poučiť z medzinárodných pilotov. Projekt tak vytvára predpoklad na zapojenie ŠÚ SR do medzinárodnej výmeny skúseností s AI asistentmi v štatistike. Verejný AI analytik síce neponúka surové dáta, ale robí oficiálne dáta prístupnejšími. To je dôležitý aspekt otvorenej vedy – dáta môžu slúžiť širšej komunite. Napríklad učiteľia môžu so študentmi využívať verejného AI analytika a učiť sa pracovať s dátami formou dialógu. Tým sa štatistiky dostanú viac do vzdelávania a povedomia. Keďže verejný AI analytik poskytuje odkazy na zdrojové tabuľky, môže motivovať pokročilejších používateľov prejsť na samostatnú prácu s dátami, čo opäť podporuje rozvoj dátovej kultúry. Verejný AI analytik predstavuje nový komunikačný kanál medzi úradom a verejnosťou. Umožní komukoľvek kľásť otázky a hneď dostať odpoveď podloženú oficiálnymi štatistikami. To zvyšuje transparentnosť a užitočnosť štatistiky pre občanov v každodennom živote. Očakáva sa, že to zvýši návštevnosť štatistických portálov a frekvenciu využívania údajov v médiách (pretože ak novinár vie, že odpoveď dostane za pár sekúnd, bude častejšie overovať fakty práve u verejného AI analytika). Celkovo tak stúpne informovanosť verejnosti a prítomnosť faktov v spoločenskej diskusii.

Projekt zároveň prináša posilnenie medzinárodnej kompatibility a reputácie ŠÚ SR. Tým, že sa AI platforma opiera o princípy AIML4OS, otvorené štandardy, MLOps a transparentné governance mechanizmy, vytvára sa predpoklad, že ŠÚ SR bude vnímaný ako inštitúcia, ktorá drží krok s aktuálnymi trendmi v európskom štatistickom systéme. To zvyšuje potenciál zapájania sa do medzinárodných iniciatív, zdieľania know-how a využívania spoločných metodických prístupov. Najmä pri témach ako automatická klasifikácia, syntetické dáta, NLG alebo AI asistenti v oficiálnej štatistike je medzinárodná výmena skúseností veľmi dôležitá a projekt vytvára predpoklad, že ŠÚ SR v tejto oblasti nebude iba pasívnym príjemcom, ale aj aktívnym partnerom.

3.9.2 Zdroj financovania

Projekt bude financovaný z Programu Slovensko.

3.10 Harmonogram projektu

Nasledujúca tabuľka uvádza fázy a aktivity projektu s odhadovaným začiatkom a koncom.

ID	Fáza/aktivita	Začiatok (odhad termínu)	Koniec (odhad termínu)	Poznámka
1.	Prípravná fáza a Iniciačná fáza	11/2026	12/2026	Schválenie projektovej dokumentácie a príprava súťažných podkladov
2.	Realizačná fáza	01/2027	12/2028	
2a	Analýza a Dizajn	01/2027	06/2027	
2b	Nákup infraštruktúrnych služieb, programových a technických prostriedkov	02/2027	09/2027	Pilotné prostredie na 12 mesiacov, technologická infraštruktúra a licencie pre internú a verejnú zónu
2c	Implementácia a testovanie	04/2027	09/2028	
2d	Nasadenie a PIP	07/2028	12/2028	PIP - 3 mesiace po nasadení
3.	Dokončovacia fáza	10/2028	12/2028	
4.	Podpora prevádzky (SLA)	01/2029	12/2036	Zmluva o podpore prevádzky platformy a AI modelov

Tabuľka 6 Harmonogram projektu

Plánovaná dĺžka realizačnej fázy projektu bude 24 mesiacov. Projekt je členený do logicky nadväzujúcich etáp, ktoré zabezpečujú postupnú, riadenú a bezpečnú implementáciu AI platformy v prostredí Štatistického úradu SR. Celkový harmonogram zahŕňa prípravnú a iniciačnú fázu, analyticko-návrhovú fázu, implementačnú fázu rozdelenú na čiastkové kroky podľa jednotlivých funkčných oblastí, fázu testovania a overenia, nasadenie do produkčného prostredia a následnú prevádzkovú podporu.

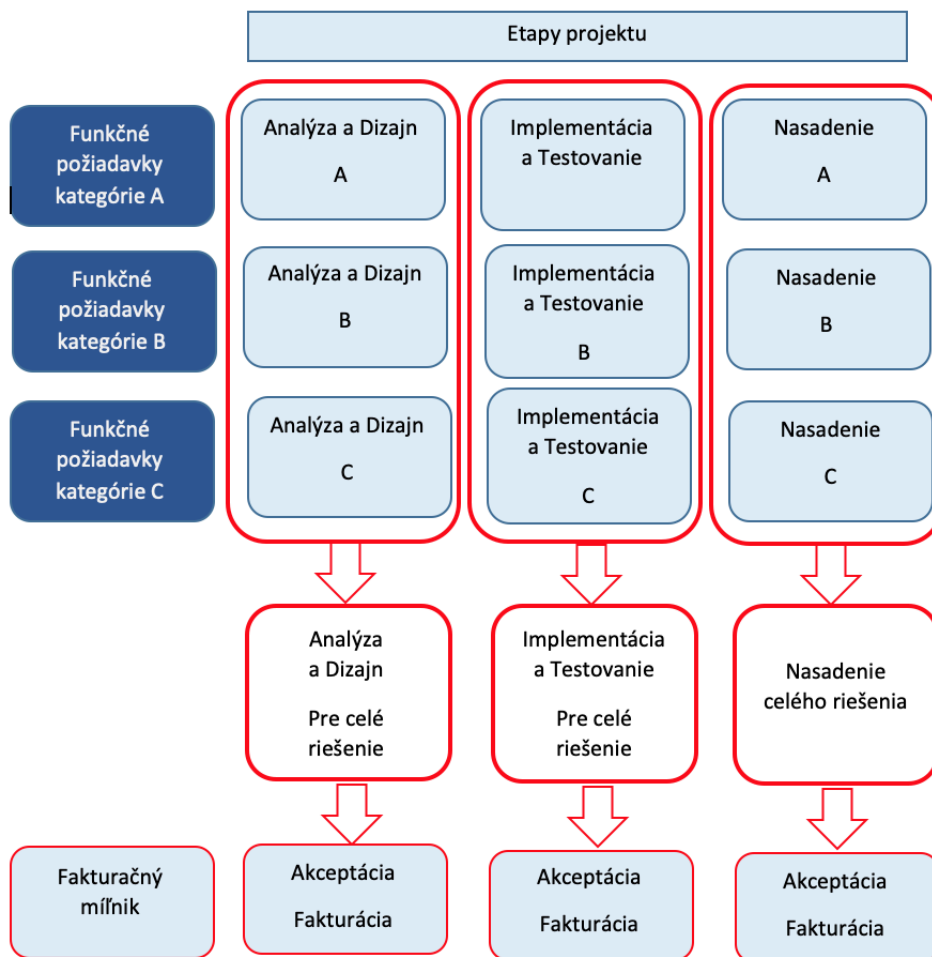
Realizácia prebieha v dvoch inkrementoch podľa analýzy nákladov a prínosov. Inkrement 1 (01/2027 – 12/2027) zahŕňa moduly Asistent pre zamestnancov, Klasifikácia textových odpovedí, Analytik nad mikroúdajmi, Generovanie syntetických údajov, Verejný analytik, Inteligentné vyhľadávanie a Automatické textové súhrny. Inkrement 2 (01/2028 – 12/2028) zahŕňa moduly Viacjazyčné publikovanie, Znalostná a vyhľadávacia vrstva, Správa modelov a kalibrácia, Kontrolovaný prestup medzi zónami, Bezpečnosť, prístupy a audit, Integrácie na systémy úradu a Prevádzkové vlastnosti a monitoring. Výstupy Inkrementu 1 sú dodané do 12/2027. Do produkčnej prevádzky sa platforma nasadzuje ako celok po dodaní Inkrementu 2 vo fáze nasadenia 07/2028 – 12/2028.

Projekt bude realizovaný postupným, etapovým spôsobom, pričom jednotlivé fázy budú na seba nadväzovať a ich ukončenie bude viazané na dosiahnutie definovaných výstupov a schvaľovacích míľnikov. Tento prístup je vhodný vzhľadom na charakter riešenia, ktoré kombinuje bezpečnostne citlivé spracovanie mikroúdajov, integráciu do existujúcich informačných systémov ŠÚ SR a zavádzanie nových AI služieb pre interné aj externé použitie. Sekvenčné riadenie etáp umožní dôslednú kontrolu kvality výstupov, overenie súladu s bezpečnostnými a metodickými požiadavkami a minimalizáciu rizík pri zavádzaní riešenia do produkčnej prevádzky.

Projekt bude riadený v súlade s princípmi metodicky riadeného projektového manažmentu, s dôrazom na transparentnosť, auditovateľnosť, kontrolu rozsahu a schvaľovanie kľúčových výstupov. Pre jednotlivé etapy budú definované míľniky, akceptačné kritériá a rozhodovacie body. Súčasťou riadenia bude aj priebežné vyhodnocovanie rizík, kontrola plnenia bezpečnostných požiadaviek, dohľad nad kvalitou dát a modelov a koordinácia medzi biznis, metodickou, bezpečnostnou a technologickou zložkou projektu.

Osobitnou súčasťou harmonogramu bude pilotná fáza, ktorá bude zameraná na overenie konceptu vybraných verejných služieb v izolovanom prostredí verejného cloudu v rámci EÚ nad necitlivými, syntetickými a verejnými údajmi. Táto fáza bude realizovaná v úplnej izolácii od internej infraštruktúry ŠÚ SR. Cieľom pilotu bude overiť funkčný koncept, používateľský prínos, kvalitu odpovedí a prevádzkové parametre vybraných služieb ešte pred ich nasadením do cieľového on-premise prostredia. Následne bude realizované produkčné nasadenie platformy v dátovom centre Štatistického úradu SR, kde bude zabezpečená plnohodnotná prevádzka chránenej internej zóny aj verejnej zóny služieb.

Po ukončení implementácie prejde riešenie do režimu zmluvnej servisnej podpory a prevádzkovej stabilizácie. Podpora bude zahŕňať správu aplikačných a integračných komponentov, monitoring prevádzky, riešenie incidentov, bezpečnostné aktualizácie, správu modelov a dátových pipeline, ako aj priebežné vyhodnocovanie kvality poskytovaných služieb. Rozsah a trvanie servisnej podpory budú definované v súlade s výsledkami verejného obstarávania a prevádzkovými požiadavkami ŠÚ SR. Zvolený model realizácie zabezpečuje riadenú implementáciu AI platformy, minimalizuje dopad na existujúcu prevádzku informačných systémov a vytvára predpoklady pre dlhodobu udržateľnosť, bezpečnú a rozšíriteľnú analytické prostredie v oblasti oficiálnej štatistiky.



Objednávateľ špecifikuje funkčné požiadavky a kategórie A, B, C (pričom A = must have, B = nice to have, C= zvyšné)

Obrázok 3 Ilustrácia rozdelenia realizácie projektu metódou Waterfall

3.11 Návrh organizačného zabezpečenia projektu (projektový tím)

Riadiaci výbor projektu (RV) predstavuje najvyšší riadiaci orgán zodpovedný za strategické usmerňovanie a dohľad nad realizáciou projektu v súlade so schválenou projektovou dokumentáciou. Riadiaci výbor sa riadi dokumentom „Štatút Riadiaceho výboru“, ktorý upravuje jeho právomoci, zloženie, rozhodovacie procesy a kompetencie.

Úlohy Riadiaceho výboru

- Dohľad nad hospodárením, efektívnym a účelným využívaním finančných prostriedkov projektu.
- Schvaľovanie kľúčových dokumentov a výstupov projektu (napr. projektový plán, zmenové požiadavky, odovzdávacie protokoly).
- Prispôsobenie metodiky riadenia podľa špecifik realizovaného projektu.
- Riešenie eskalovaných problémov a podpora pri rozhodovaní v kritických bodoch.

Zloženie Riadiaceho výboru

Riadiaci výbor má minimálne 5 členov. Väčšinu členov s hlasovacím právom nominuje objednávateľ a títo zástupcovia reprezentujú jeho záujmy. Zloženie výboru je nasledovné:

- Predseda RV
- Biznis vlastník
- Zástupca prevádzky
- Projektový manažér realizátora projektu
- Zástupca dodávateľa

V prípade neprítomnosti predsedu je možné písomne delegovať jeho právomoci na iného člena výboru prostredníctvom splnomocnenia.

Zasadnutia a hlasovanie

Riadiaci výbor zasadá pravidelne, spravidla raz mesačne, minimálne však raz za tri po sebe nasledujúce mesiace. Zasadnutie zvoláva predseda. Hlavné výstupy zo zasadnutí sú zdokumentované v zápisnici, ktorej prílohou je prezenčná listina a prípadné splnomocnenia. Program zasadnutia a podklady pripravuje a distribuuje administratívny projektový manažér, a to najneskôr 3 pracovné dni pred zasadnutím. O rozhodnutiach sa hlasuje. Pre schválenie je potrebná nadpolovičná väčšina prítomných členov s hlasovacím právom. V prípade rovnosti hlasov rozhoduje hlas predsedu, ktorý má v takom prípade dvojnásobnú váhu.

Ukončenie činnosti Riadiaceho výboru

Činnosť Riadiaceho výboru sa ukončí dňom zavedenia projektu do produkčnej prevádzky, t. j. po jeho plnohodnotnej implementácii. Zoznam aktuálnych členov Riadiaceho výboru je evidovaný v dokumente Komunikačná matica, ktorý je uložený na zdieľanom projektovom úložisku.

Projektový tím a riadenie projektu

Projekt bude po organizačnej stránke riadený kombináciou administratívneho a odborného riadenia, ktoré zabezpečia poverení pracovníci Objednávateľ a počas celej doby realizácie. Administratívne riadenie bude vykonávané najmä prostredníctvom projektového manažéra a finančného manažéra, ktorí budú zodpovední za priebežnú koordináciu aktivít, sledovanie naplňovania harmonogramu a kontrolu rozpočtu. Zabezpečia aj spracovanie dokumentácie v súlade s požiadavkami Príručky pre prijímateľa a ďalších usmernení poskytovateľa. Odborné riadenie zabezpečí IT projektový manažér, ktorý bude vykonávať dohľad nad vývojom a implementáciou dodaného diela. Bude koordinovať vývoj s požiadavkami objednávateľa, zabezpečovať odbornú komunikáciu s dodávateľom a dohliadať na dodržanie technických a bezpečnostných štandardov.

Projektové roly

- IT projektový manažér – dohľad nad odbornou časťou riadenia a technickou realizáciou.
- Kľúčový používateľ – zabezpečenie spätnej väzby z pohľadu koncových užívateľov.
- Finančný manažér – kontrola rozpočtu, oprávnenosti výdavkov a finančné riadenie.

Počas implementácie projektu a po jeho skončení (počas obdobia udržiavateľnosti) bude k dispozícii interný personál na obsluhu, prevádzku a rozvoj riešenia. Zároveň bude zabezpečené financovanie tohto personálu.

ID	Rola v projekte	Meno a Priezvisko	Pracovné zaradenie	Org. útvar
1.	IT projektový manažér	Doplňte meno a priezvisko	Doplňte pozíciu (pracovné zaradenie v línii)	Doplňte názov org. útvaru
2.	Kľúčový používateľ	Doplňte meno a priezvisko	Doplňte pozíciu (pracovné zaradenie v línii)	Doplňte názov org. útvaru
3.	Finančný manažér	Doplňte meno a priezvisko	Doplňte pozíciu (pracovné zaradenie v línii)	Doplňte názov org. útvaru
4.	Manažér kybernetickej a informačnej bezpečnosti	Doplňte meno a priezvisko	Doplňte pozíciu (pracovné zaradenie v línii)	Doplňte názov org. útvaru

Tabuľka 7 Projektový tím

Zodpovednosti, povinnosti a kompetencie IT projektového manažéra uvádza nasledujúca tabuľka.

Projektová rola:

Detailný popis rozsahu zodpovedností, povinností a kompetencií:

IT projektový manažér

Strategické riadenie a plánovanie:

- Zodpovedá za komplexné riadenie projektu v mene objednávateľa – plánovanie, organizovanie, monitorovanie, vyhodnocovanie a ukončenie projektu.
- Definuje projektové ciele, hlavné míľniky, časový harmonogram, rozpočet a personálne zdroje projektu.
- Navrhne metodiku a rámec riadenia projektu, vrátane výberu nástrojov a spôsobov riadenia (napr. PRINCE2, PMBOK, SCRUM).
- Zabezpečuje súlad projektu so strategickými cieľmi organizácie a štátnymi/digitálnymi politikami.

Riadenie realizácie projektu:

- Koordinuje prácu interných a externých členov projektového tímu, dodávateľov a partnerov.
- Riadi realizáciu jednotlivých fáz životného cyklu projektu (iniciácia, návrh, vývoj, testovanie, nasadenie).
- Vykonáva kontrolu plnenia míľnikov a výstupov (deliverables) projektu v zmysle časového plánu a rozpočtu.
- Prijíma rozhodnutia v mene objednávateľa v rozsahu kompetencií stanovených projektovým plánom alebo riadiacim výborom.

Zabezpečenie zhody s legislatívou a metodikami:

- Zodpovedá za splnenie požiadaviek podľa zákona č. 95/2019 Z. z. o IT vo verejnej správe a súvisiacich predpisov (zákon č. 69/2018 Z. z., vyhláška č. 78/2020 Z. z. atď.).
- Zabezpečuje vypracovanie projektovej dokumentácie v súlade s požiadavkami MetaIS, vrátane výstupov M-01 až M-12.

Riadenie komunikácie a informovanosti:

- Predkladá pravidelné správy o stave a pokroku projektu riadiacemu výboru a vedeniu organizácie.
- Zabezpečuje komunikáciu so zúčastnenými stranami – internými oddeleniami, verejnosťou, médiami (prostredníctvom komunikačného oddelenia).
- Vypracúva a aktualizuje komunikačný plán projektu, zodpovedá za medializáciu a interné informovanie.

Riadenie rizík a zmien:

- Identifikuje a analyzuje riziká projektu, navrhuje opatrenia na ich minimalizáciu.
- Zodpovedá za aktualizáciu registra rizík a plánov reakcie na riziká.
- Riadi proces zmien – prijíma zmenové požiadavky, vyhodnocuje ich dopad, navrhuje schválenie riadiacemu výboru.

Riadenie kvality a výstupov projektu:

- Zabezpečuje, že výstupy projektu spĺňajú kvalitatívne požiadavky podľa špecifikácie.
- Dohliada na testovanie, overovanie funkčnosti, bezpečnostné testy (napr. penetračné testy).
- Overuje výsledky a garantuje pripravenosť projektu na prechod do produkčnej prevádzky.

Riadenie ľudských a finančných zdrojov:

- Navrhuje zloženie projektového tímu, prideluje úlohy a kontroluje ich plnenie.
- Motivuje, hodnotí a podporuje členov tímu. V prípade potreby vykonáva rolu tímového manažéra.
- Zodpovedá za finančné riadenie projektu – sledovanie rozpočtu, verifikácia dodávateľských faktúr, zabezpečenie rozpočtových opatrení.

Archív a dokumentácia:

- Vede a archivuje všetku projektovú dokumentáciu v súlade s archivačnými a internými predpismi.
- Zodpovedá za publikovanie výstupov a zápisov do MetaIS podľa aktuálne platnej metodiky a štruktúry výstupov.
- Garantuje verziovanie a formálnu správnosť dokumentácie.

Tabuľka 8 Opis projektovej roly IT projektový manažér

Zodpovednosti, povinnosti a kompetencie kľúčového používateľa uvádza nasledujúca tabuľka.

Projektová rola:

Stručný popis:

Kľúčový používateľ

Zodpovedá za zastupovanie záujmov budúcich koncových používateľov projektových produktov alebo výstupov. Aktívne sa podieľa na špecifikácii požiadaviek, testovaní a akceptácii riešenia. Poskytuje vstupy pre zabezpečenie kvality, použiteľnosti a funkčnosti systému z pohľadu používateľov. Podporuje projektový tím pri nastavovaní prevádzkových a podporných procesov.

Detailný popis rozsahu zodpovedností, povinností a kompetencií

Zastupovanie koncových používateľov:

- Zastupuje potreby a očakávania cieľovej skupiny používateľov, poskytuje spätnú väzbu na navrhované riešenia, reprezentuje používateľskú perspektívu počas celého životného cyklu projektu.

Spolupráca na špecifikácii:

- Spolupodieľa sa na tvorbe funkčných, nefunkčných a technických požiadaviek.
- Zabezpečuje ich vecno-procesnú a legislatívnu správnosť a jasnosť.
- Definuje očakávané prínosy systému (kvalitatívne aj kvantitatívne) z pohľadu používateľa.

Definovanie vstupov do projektového návrhu:

- Prispieva k návrhu akceptačných kritérií a používateľských scenárov.
- Identifikuje riziká, závislosti a rozhrania medzi systémom a používateľmi, medzi jednotlivými procesmi, ako aj so súvisiacimi externými systémami.

Účast' na testovaní:

- Zodpovedá za vykonanie používateľského testovania (UX testovania) a overenie súladu používateľského rozhrania s očakávaniami a definovanými požiadavkami.
- Spolupodieľa sa na návrhu a realizácii akceptačných testov (UAT), ktoré potvrdzujú pripravenosť systému na nasadenie.

Akceptácia a spustenie:

- Vydáva návrh na akceptáciu výstupov projektu a následné spustenie systému do produkčnej prevádzky.
- Podieľa sa na príprave prechodových aktivít vrátane školenia a informovania koncových používateľov.

Zmenový manažment:

- Iniciuje požiadavky na zmeny funkcionality produktov na základe spätnej väzby od používateľov alebo zmenených legislatívnych či prevádzkových podmienok.

Spolupráca s projektovým tímom:

- Zúčastňuje sa projektových stretnutí, aktívne komunikuje s projektovým manažérom a ďalšími členmi tímu.
- Poskytuje vstupy do manažérskej a špecializovanej dokumentácie.
- Plní úlohy podľa pokynov projektového manažéra a záverov zo stretnutí.

Podpora pre prevádzkové nastavenia:

- Poskytuje súčinnosť pri nastavovaní interných riadiacich aktov súvisiacich s prevádzkou, servisom a podporou nového systému.

Tabuľka 9 Opis projektovej roly Kľúčový používateľ

Zodpovednosti, povinnosti a kompetencie finančného manažéra uvádza nasledujúca tabuľka.

Projektová rola:

Detailný popis rozsahu zodpovedností, povinností a kompetencií

Finančný manažér

Finančné plánovanie a monitoring:

- Koordinuje a monitoruje napĺňanie finančných ukazovateľov projektu.
- Sleduje priebežné čerpanie finančných prostriedkov v súlade so schváleným rozpočtom a časovým harmonogramom.

Rozpočtová kontrola:

- Zodpovedá za správne, efektívne a oprávnené čerpanie rozpočtu podľa Zmluvy o NFP a príručky pre prijímateľa.
- Overuje oprávnenosť výdavkov pred ich zúčtovaním.

Metodická a odborná podpora:

- Poskytuje usmernenia členom projektového tímu ohľadom oprávnených výdavkov, postupov financovania a zmenových konaní.
- Sleduje a aplikuje aktuálne zmeny legislatívy a riadiacich dokumentov týkajúcich sa financovania z fondov EÚ.

Žiadosti o platbu:

- Zodpovedá za prípravu, formálnu kontrolu a predloženie žiadostí o platbu vrátane všetkých príloh podľa príručky pre prijímateľa.
- Zabezpečuje evidenciu a administráciu žiadostí v ITMS2021+.

Dokumentácia a evidencia:

- Vedie kompletnú finančnú dokumentáciu projektu.
- Pripravuje pravidelné finančné prehľady, reporty a štatistiky o čerpaní.

- Zabezpečuje archiváciu dokumentov v súlade s pravidlami fondov EÚ (Program Slovensko) a internými smernicami.

Komunikácia a kontrola:

- Koordinuje prípravu na externé finančné kontroly a audit, poskytuje potrebnú súčinnosť.
- Vystupuje za prijímateľa v komunikácii s poskytovateľom NFP a ďalšími kontrolnými subjektmi v oblasti financií.

Administratíva a interná spolupráca:

- Spolupracuje s organizačnými útvarmi (napr. ekonomické, právne, verejné obstarávanie) pri zabezpečovaní činností súvisiacich s implementáciou.
- Zúčastňuje sa projektových porád, predkladá návrhy a zistenia týkajúce sa rozpočtu, financií a výdavkov.

Zabezpečovanie ďalších úloh:

- Zodpovedá za prípravu ďalších dokumentov a podkladov pre audit, evaluáciu alebo zmenu projektu z pohľadu financií.
- Riadi finančnú časť zmenového riadenia v projekte (napr. rozpočtové opatrenia, presuny prostriedkov).
- Dohliada na splnenie pravidiel publicity týkajúcich sa čerpania finančných prostriedkov.

Tabuľka 10 Opis projektovej roly Finančný manažér

4. LEGISLATÍVA

Navrhovaný projekt je koncipovaný tak, aby bol realizovateľný v rámci existujúceho legislatívneho rámca Slovenskej republiky a Európskej únie a aby jeho implementácia nevyžadovala zásadné legislatívne zmeny ako podmienku realizácie projektu. Architektúra platformy umelej inteligencie a spôsob spracovania údajov sú navrhnuté v súlade s požiadavkami platnej legislatívy, najmä v oblasti ochrany osobných údajov, štatistickej dôverylosti, kybernetickej bezpečnosti a riadenia informačných systémov verejnej správy. Projekt sa opiera najmä o nasledovné právne predpisy:

- zákon č. 540/2001 Z. z. o štátnej štatistike, ktorý upravuje zber, spracovanie a ochranu štatistických údajov a princípy štatistickej dôverylosti,
- nariadenie Európskeho parlamentu a Rady (EÚ) 2016/679 (GDPR), ktoré upravuje ochranu osobných údajov pri spracovaní údajov,
- zákon č. 69/2018 Z. z. o kybernetickej bezpečnosti, ktorý stanovuje požiadavky na riadenie kybernetických rizík a bezpečnostné opatrenia v informačných systémoch,
- nariadenie (EÚ) 2024/1689 (akt o umelej inteligencii) a aktualizovanú smernicu NIS2, ktoré definujú požiadavky na bezpečné, transparentné a auditovateľné využívanie umelej inteligencie a prevádzku kritickej digitálnej infraštruktúry.

Navrhovaná platforma je navrhnutá tak, aby rešpektovala princípy štatistickej dôverylosti a ochrany osobných údajov, pričom citlivé mikroúdaje zostávajú spracovávané výhradne v infraštruktúre Štatistického úradu SR na území Slovenskej republiky. Architektúra riešenia využíva zónový bezpečnostný model, v ktorom sú komponenty pracujúce s citlivými údajmi oddelené od verejných služieb umelej inteligencie, ktoré pracujú výhradne s publikovanými agregovanými údajmi. S cieľom zabezpečiť dlhodobú legislatívnu kompatibilitu platformy je návrh riešenia založený na princípoch security by design a privacy by default. Platforma zároveň obsahuje mechanizmy, ktoré umožňujú flexibilne reagovať na budúce legislatívne zmeny bez potreby zásadných architektonických úprav. Medzi tieto mechanizmy patria najmä riadenie prístupových práv prostredníctvom modelov RBAC a ABAC, kontrola tokov údajov prostredníctvom mechanizmov DLP a egress politiky, auditovateľnosť prístupov a operácií nad dátami, ako aj implementácia princípu Human-in-the-Loop, ktorý zabezpečuje, že výstupy generované systémami umelej inteligencie určené na publikovanie podliehajú ľudskej kontrole. Platforma zároveň využíva API-first architektúru a verzovanie rozhraní, čo umožňuje implementovať legislatívne požiadavky prednostne prostredníctvom úprav politik,

konfigurácií a integračných pravidiel, bez nutnosti zásadných zmien v dátovej alebo aplikačnej architektúre. Tento prístup umožňuje pružne reagovať na legislatívne zmeny a zároveň zachovať stabilitu prevádzky informačných systémov.

Súčasťou prevádzky platformy bude aj zavedenie pravidelných compliance revízií, ktoré budú vyhodnocovať dopady legislatívnych zmien na procesy a konfiguráciu platformy. Revízny cyklus bude zahŕňať najmä aktualizáciu dokumentácie ochrany údajov (DPIA), revíziu bezpečnostných politík, audit logov prístupov a prestupov údajov medzi jednotlivými bezpečnostnými zónami a aktualizáciu interných prevádzkových postupov. Výstupy týchto revízií budú slúžiť ako podklad pre riadené zmeny konfigurácie platformy a aktualizáciu bezpečnostných pravidiel. Aplikácia legislatívnych zmien bude realizovaná prostredníctvom riadeného procesu zmien, ktorý zahŕňa bezpečnostné posúdenie, testovanie dopadov na výkon a bezpečnosť systému a rozhodovanie v rámci definovaného Go/No-Go rámca. Tento proces zabezpečí, že legislatívne úpravy budú implementované postupne a kontrolovane bez narušenia dostupnosti a integrity poskytovaných služieb.

V prípade, že by v budúcnosti došlo k legislatívnym zmenám, ktoré by zásadne obmedzili využívanie umelej inteligencie pri spracovaní štatistických údajov alebo by sprísnil podmienky prístupu k dátam, mohlo by to ovplyvniť rozsah funkcionalít platformy, najmä v oblasti automatizovanej analýzy dát alebo sprístupňovania syntetických datasetov. V takom prípade by bolo potrebné upraviť konfiguráciu alebo funkcionalitu platformy tak, aby zostala plne v súlade s legislatívnymi požiadavkami. Návrh riešenia však počíta s vysokou mierou legislatívnej adaptovateľnosti a s implementáciou mechanizmov riadenia súladu, ktoré umožnia reagovať na legislatívne zmeny bez významného dopadu na základné ciele projektu a jeho časový harmonogram.

5. ARCHITEKTÚRA RIEŠENIA PROJEKTU

Architektúra riešenia projektu je navrhnutá ako robustná, bezpečná a škálovateľná platforma umelej inteligencie podporujúca spracovanie, analýzu a disemináciu štatistických údajov v prostredí Štatistického úradu Slovenskej republiky. Architektonický návrh vychádza z požiadaviek na ochranu štatistickej dôvernosti, ochranu osobných údajov a kybernetickú bezpečnosť a zároveň rešpektuje požiadavku rezidencie citlivých údajov na území Slovenskej republiky. Architektúra nadväzuje na definované scenáre použitia pričom zabezpečuje, aby každá funkcionalita bola prevádzkovaná v primeranej zóne a s primeranými bezpečnostnými kontrolami. Platforma je navrhnutá ako modálne riešenie prevádzkované v infraštruktúre dátového centra Štatistického úradu SR, pričom využíva princípy bezpečnej architektúry, riadenia prístupov a oddelenia spracovania citlivých údajov od verejných analytických služieb. Základným architektonickým princípom riešenia je využitie dvojzónového bezpečnostného modelu, ktorý zabezpečuje oddelenie spracovania citlivých údajov od verejných služieb platformy a umožňuje jasne definovať pravidlá dátových tokov, auditovateľnosť a prevádzkové politiky. Architektúra je navrhnutá v súlade s princípmi security-by-design a privacy-by-default, t. j. bezpečnostné a ochranné mechanizmy sú integrované priamo do návrhu systému, nie dopĺňané dodatočne. Základným princípom je prísne oddelenie spracovania citlivých mikroúdajov od verejných analytických služieb. Tento princíp je implementovaný prostredníctvom rozdelenia architektúry do bezpečnostných zón, ktoré umožňujú kontrolovaný prenos údajov a minimalizujú riziko neoprávneného prístupu k citlivým dátam. Architektúra zároveň rešpektuje princíp API-first, podľa ktorého sú všetky aplikačné služby platformy publikované prostredníctvom stabilných a verziovaných rozhraní. Tento prístup umožňuje jednoduchú integráciu s existujúcimi informačnými systémami ŠÚ SR a zároveň podporuje interoperabilitu s externými systémami. Dôležitým princípom je aj modulárnosť riešenia. Jednotlivé komponenty platformy sú navrhnuté ako samostatné služby, ktoré je možné nezávisle rozvíjať, nasadzovať a škálovať podľa aktuálnych potrieb. Architektúra taktiež podporuje princípy otvorenosti a predchádzania vendor lock-in. Riešenie využíva štandardizované rozhrania, otvorené formáty dát a otvorené štandardy, čím sa zabezpečuje dlhodobá udržiateľnosť a možnosť ďalšieho rozvoja systému.

Navrhované riešenie rozdeľuje platformu na dve oddelené bezpečnostné domény, ktoré umožňujú bezpečné spracovanie údajov s rôznou úrovňou citlivosti a zároveň zabezpečujú súlad s legislatívnymi požiadavkami na ochranu osobných údajov a štatistickej dôvernosti.

- Zóna A – interná chránená zóna predstavuje bezpečnostne izolované prostredie, v ktorom sú spracovávané citlivé mikroúdaje a interné agregované údaje Štatistického úradu SR. V tejto zóne sú prevádzkované komponenty zabezpečujúce tréning a inferenciu modelov umelej inteligencie nad citlivými dátami, interné analytické nástroje a úložiská mikroúdajov. Zóna A je navrhnutá ako prostredie bez priamej externej sieťovej konektivity pre citlivý obsah – nie je sieťovo prepojená s verejným internetom ani prostriedmi verejného

cloudu, prístupy sú striktné riadené prostredníctvom IAM, RBAC/ABAC a všetky operácie sú plne auditované. Tým sa minimalizuje riziko úniku dát a zabezpečuje ochrana štatistického tajomstva a súlad s požiadavkami.

- Zóna B – verejná zóna obsahuje komponenty platformy určené pre poskytovanie verejných služieb nad publikovanými agregovanými údajmi. Ide najmä o služby umožňujúce sémantické vyhľadávanie v štatistických údajoch, generovanie textových interpretácií štatistických výstupov alebo analytické nástroje dostupné pre verejnosť, médiá a výskumné organizácie. Tieto služby nemajú priamy prístup k mikroúdajom v Zóne A a pracujú výhradne s publikovanými agregátmi a verejnými dátovými zdrojmi. V Zóne B platí zákaz spracovania osobných údajov a PII a všetky verejné výstupy generované AI (textové súhrny, preklady, vysvetlenia) podliehajú povinnému procesu HIL (Human-in-the-loop), t. j. ľudskej kontrole a schváleniu pred zverejnením.

Logická vrstvová architektúra

Architektúra platformy je zároveň navrhnutá ako viacvrstvový systém, ktorý umožňuje jasné oddelenie jednotlivých funkčných oblastí riešenia. Tento prístup zvyšuje bezpečnosť, modularitu a udržateľnosť riešenia a zároveň uľahčuje jeho ďalší rozvoj, testovanie a auditovateľnosť. Dátová vrstva zabezpečuje správu a ukladanie dátových zdrojov využívaných platformou. V Zóne A sú prevádzkované databázy mikroúdajov a interných agregovaných údajov, zatiaľ čo v Zóne B sú spravované úložiská publikovaných agregovaných údajov a verejných datasetov. Súčasťou tejto vrstvy sú aj dátové katalógy, metadátové registre a mechanizmy riadenia kvality a klasifikácie údajov, vrátane vedenia rodokmeňa dát (lineage) a evidencie verzií datasetov. Pred každým presunom agregovaných údajov do Zóny B sa uplatňujú pravidlá štatistického utajenia (minimálne veľkosti buniek, potlačanie nízkych početností, zaokrúhľovanie, kontrola kombinovateľnosti dimenzií) a klasifikácia dát podľa bezpečnostných politík, aby sa minimalizovalo riziko reidentifikácie. AI/ML vrstva poskytuje prostredie pre tréning, validáciu a prevádzku modelov umelej inteligencie. Táto vrstva zahŕňa komponenty pre riadenie životného cyklu modelov, experimentálne prostredia, monitoring výkonu modelov a nástroje pre MLOps (orchestráciu tréningových a nasadzovacích pipeline, registráciu modelov a datasetov, deployment modelov do produkcie). V Zóne A sú prevádzkované modely pracujúce s citlivými údajmi, zatiaľ čo v Zóne B sú nasadzované modely pracujúce výhradne s publikovanými agregátmi a verejným obsahom. AI/ML vrstva zároveň zabezpečuje vysvetliteľnosť vybraných modelov (XAI) a kontinuálne monitorovanie výkonu a driftu modelov, aby bolo možné včas reagovať na zmeny dát, korigovať rizikové správanie modelov a podporiť zodpovedné využívanie AI v oficiálnej štatistike.

Aplikačná vrstva obsahuje aplikačné služby platformy, ktoré sprostredkujú funkcionality jednotlivých AI nástrojov a analytických služieb. V internej zóne sú prevádzkované služby podporujúce interné analytické procesy a metodickú podporu pracovníkov Štatistického úradu SR (napr. interný AI asistent, interný AI analytik nad mikrodátami, služba pre generovanie syntetických mikrodasetov). Vo verejnej zóne sú dostupné služby sprístupňujúce štatistické údaje prostredníctvom aplikačných rozhraní a API brán (verejný AI analytik nad dátovými kockami, sémantické vyhľadávanie, generovanie textových súhrnov, viacjazyčná diseminácia), pričom všetky tieto služby pracujú výlučne s publikovanými agregátmi. Prezentácia vrstva zabezpečuje používateľské rozhrania platformy. Zahŕňa webové rozhranie pre verejnosť, ktoré umožňuje vyhľadávanie a interpretáciu štatistických údajov v prirodzenom jazyku a formou vizualizácií, ako aj pracovné rozhrania určené pre interných analytikov a metodikov Štatistického úradu SR v internej zóne. Prezentácia výsledkov je navrhnutá tak, aby zohľadňovala požiadavky na prístupnosť a použiteľnosť elektronických služieb, a zároveň jasne komunikovala zdroje údajov, verziu datasetov a prípadné uplatnené pravidlá utajenia (napr. upozornenia na zaokrúhlenie a potlačené hodnoty).

Pravidlá dátových tokov a prestupov

Dátové toky medzi jednotlivými bezpečnostnými zónami sú prísne regulované technickými a organizačnými kontrolami. Z architektonického hľadiska je povolený iba jednosmerný prestup dát zo Zóny A do Zóny B, pričom prenášané môžu byť výhradne publikované agregované údaje, dátové kocky, metaúdaje a schválené syntetické datasety, ktoré prešli kontrolou rizika reidentifikácie, automatizovanými testami štatistického utajenia a metodickou revíziou. Prestupy sú definované v špecifických egress workflow, ktoré zahŕňajú označenie datasetu ako publikovateľného, spustenie automatizovaných kontrol, odbornú revíziu a schválenie. Každý takýto prestup podlieha bezpečnostným kontrolám vrátane mechanizmov prevencie úniku dát (DLP), schvaľovaciemu procesu podľa princípu „four-eyes“ a úplnej auditovateľnosti všetkých operácií. Kritické dátové toky a prestupy sú predmetom posúdenia vplyvu na ochranu osobných údajov (DPIA) a sú monitorované v rámci centralizovaného SIEM riešenia, čo umožňuje včasnú detekciu anomálií a incidentov. Technické kanály sú zabezpečené pomocou šifrovania (TLS/mTLS), sieťových politík typu deny-by-default a kontrolovaných egress brán. Priame dotazovanie interných databáz zo Zóny B alebo z externého prostredia nie je povolené. Tento model zabezpečuje, že citlivé mikroúdaje

nikdy neopustia chránené prostredie Štatistického úradu SR a že všetky publikované výstupy sú v súlade s bezpečnostnými politikami a pravidlami štatistického utajenia.

Architektonické princípy a technologické riešenia

Pri návrhu platformy sú uplatnené moderné architektonické princípy podporujúce bezpečnosť, škálovateľnosť a dlhodobú udržateľnosť riešenia. Platforma je navrhnutá v súlade s princípmi security-by-design a privacy-by-default, pričom ochrana údajov a bezpečnostné mechanizmy (klasifikácia dát, riadenie prístupov, šifrovanie, DLP, auditovanie, HIL) sú implementované priamo na úrovni architektúry systému a prevádzkových procesov. Riešenie odzrkadľuje odporúčania iniciatívy AIML4OS (Eurostat/ESSnet) pre využitie AI v oficiálnej štatistike, vrátane dôrazu na zodpovednú AI, vysvetliteľnosť, reprodukciu a governance. Riešenie využíva API-first prístup, v rámci ktorého sú všetky funkcionality platformy sprístupnené prostredníctvom stabilných a verziovaných rozhraní. Tento prístup umožňuje jednoduchú integráciu platformy s existujúcimi systémami a podporuje jej ďalší rozvoj, ako aj auditovateľnosť prístupov a sledovanie verzií služieb. Verejné API nikdy neposkytujú prístup k mikroúdajom ani interným agregátom a prístupujú výhradne k publikovaným agregátom v Zóne B. Interné API v Zóne A sú chránené prísnyimi politikami riadenia prístupu a sú plne auditovateľné. Technologická infraštruktúra platformy je založená na kontajnerizovaných komponentoch prevádzkovaných v prostredí orchestrácie kontajnerov, čo umožňuje škálovanie jednotlivých služieb podľa aktuálneho zaťaženia a zabezpečuje vysokú dostupnosť riešenia. Pri implementácii sú preferované otvorené technologické štandardy a štandardizované rozhrania, ktoré znižujú závislosť na konkrétnych dodávateľoch, podporujú prenositeľnosť riešenia (cloud-agnostic prístup) a dlhodobú udržateľnosť.

Fázovanie implementácie

Implementácia platformy bude realizovaná postupne v dvoch základných etapách, ktoré sú ďalej detailne rozpracované do jednotlivých fáz projektu. Pilotná fáza slúži na overenie konceptu vybraných verejných služieb v izolovanom prostredí verejného cloudu v rámci EÚ nad necitlivými, syntetickými a verejnými údajmi. Pilotná fáza prebieha v plnej sieťovej izolácii od Zóny A – neexistuje žiadne sieťové prepojenie na infraštruktúru spracúvajúcu mikroúdaje – a slúži na technické overenie konceptu a funkcionality vybraných „B-komponentov“ (verejných služieb) platformy. Cieľom tejto fázy je overenie technického konceptu, funkčných vlastností a škálovateľnosti vybraných komponentov bez rizika ohrozenia citlivých dát. Po úspešnom overení konceptu budú jednotlivé komponenty platformy nasadené do produkčného prostredia v dátovom centre Štatistického úradu SR, kde budú prevádzkované v rámci definovaného dvojzónového modelu. Produkčné spracovanie citlivých mikroúdajov bude prebiehať výlučne v Zóne A na on-premise infraštruktúre, zatiaľ čo verejné služby nad publikovanými agregátmi budú prevádzkované v Zóne B s možnosťou postupného rozširovania kapacít podľa potrieb a za podmienky zachovania bezpečnostných a legislatívnych požiadaviek. Takto navrhnutý architektonický model zabezpečuje, že citlivé štatistické údaje zostávajú trvalo v chránenom prostredí Štatistického úradu SR, zatiaľ čo verejnosti sú sprístupnené moderné nástroje umožňujúce jednoduchšie a intuitívnejšie využívanie štatistických informácií. Zároveň umožňuje riadené fázovanie implementácie, priebežné vyhodnocovanie rizík a prínosov a prispôbovanie riešenia vyvíjajúcim sa požiadavkám a odporúčaniam na úrovni národnej aj európskej štatistickej komunity.

5.1 Stanovenie alternatív architektúry riešenia

Alternatívy architektúry riešenia boli identifikované a posudzované na jednotlivých vrstvách architektúry riešenia – biznisovej, aplikačnej a technologickej. Cieľom tohto procesu bolo určiť variant riešenia, ktorý najlepšie napĺňa ciele projektu, požiadavky jednotlivých stakeholderov a zároveň rešpektuje legislatívne, bezpečnostné a prevádzkové obmedzenia. V biznis vrstve bol výber alternatívy realizovaný prostredníctvom multikriteriálnej analýzy (MCA), ktorá umožňuje systematické porovnanie jednotlivých variantov riešenia na základe definovaných kritérií vychádzajúcich z motivácie projektu, potrieb používateľov a strategických cieľov organizácie. Alternatívy v aplikačnej a technologickej vrstve sú posúdené slovné. Posudzovanie alternatív prebiehalo postupne na jednotlivých architektonických vrstvách, pričom každá vrstva nadväzuje na výsledky predchádzajúcej.

5.1.1 Stanovenie alternatív v biznisovej vrstve architektúry

Na úrovni biznis architektúry boli identifikované alternatívy riešenia spôsobu spracovania a diseminácie štatistických údajov. Alternatívy predstavujú rôzne prístupy k modernizácii procesov spracovania štatistických dát a k využívaniu umelej inteligencie v prostredí Štatistického úradu SR.

Alternatíva B0 – nulový variant

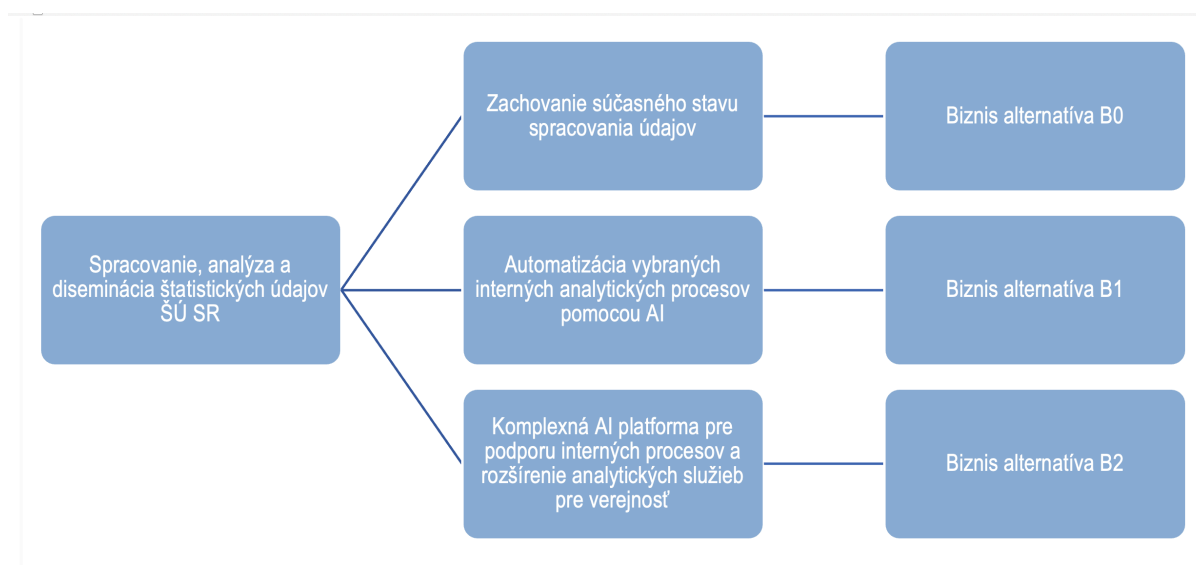
Nulový variant predstavuje zachovanie súčasného stavu bez implementácie nových nástrojov umelej inteligencie a bez zásadných zmien v procesoch spracovania a diseminácie štatistických údajov. Existujúce procesy zberu, spracovania a publikovania údajov zostávajú realizované prostredníctvom aktuálne používaných nástrojov a pracovných postupov. Spracovanie textových odpovedí respondentov, klasifikácia údajov a príprava analytických výstupov by naďalej prebiehali prevažne manuálne alebo s využitím existujúcich čiastkových nástrojov bez systematickej podpory umelej inteligencie. Analytické procesy by zostali závislé od manuálnej práce analytikov, čo by naďalej predstavovalo významnú časovú a kapacitnú záťaž. V oblasti sprístupňovania štatistických údajov by nedošlo k rozšíreniu existujúcich služieb. Verejnosť, výskumné organizácie a ďalší používatelia by mali aj naďalej prístup k údajom najmä prostredníctvom tradičných publikačných kanálov a existujúcich databázových nástrojov bez možnosti využívať pokročilé analytické služby založené na umelej inteligencii. Nevýhodou tohto variantu je najmä pretrvávajúce vysokej manuálnej záťaže pracovníkov, dlhší čas spracovania údajov a obmedzené možnosti analytického využitia štatistických údajov zo strany externých používateľov. Zároveň by nedošlo k využitiu potenciálu moderných analytických technológií na zvýšenie efektivity spracovania dát a zlepšenie dostupnosti štatistických informácií.

Alternatíva B1 – čiastočná automatizácia interných procesov

Táto alternatíva predpokladá využitie technológií umelej inteligencie predovšetkým na podporu interných analytických procesov Štatistického úradu SR. Hlavným cieľom je zefektívniť vybrané činnosti spojené so spracovaním štatistických údajov a znížiť manuálnu záťaž analytikov pri rutinných úlohách. Riešenie by sa zameriavalo najmä na automatizáciu vybraných procesov spracovania dát, napríklad na automatickú klasifikáciu textových odpovedí respondentov podľa štandardných klasifikačných schém (napr. ISCO alebo NACE), podporu analytikov pri vyhľadávaní informácií v interných databázach alebo asistenciu pri interpretácii vybraných analytických výstupov. Zavedenie týchto nástrojov by prispelo k zrýchleniu spracovania dát, zníženiu chybovosti manuálnych procesov a zvýšeniu efektivity práce analytikov. V rámci tejto alternatívy by však nedochádzalo k zásadnej transformácii spôsobu sprístupňovania štatistických údajov verejnosti. Externé analytické služby založené na umelej inteligencii by neboli implementované a publikovanie údajov by naďalej prebiehalo prostredníctvom existujúcich mechanizmov a publikačných nástrojov. Alternatíva tak prináša čiastočné zlepšenie efektivity interných procesov, avšak nevyužíva plný potenciál umelej inteligencie na rozšírenie analytických služieb pre externých používateľov ani na modernizáciu spôsobu práce s verejne dostupnými štatistickými údajmi.

Alternatíva B2 – komplexná AI platforma pre interné aj verejné služby

Táto alternatíva predstavuje komplexnú modernizáciu procesov spracovania, analýzy a diseminácie štatistických údajov prostredníctvom zavedenia centrálnej platformy umelej inteligencie. Riešenie je zamerané na podporu interných analytických procesov Štatistického úradu SR a zároveň na rozšírenie spôsobov sprístupňovania štatistických údajov pre verejnosť, výskumnú komunitu a ďalších používateľov. V rámci tejto alternatívy dochádza k postupnej transformácii vybraných analytických a spracovateľských procesov z manuálnych alebo poloautomatizovaných postupov na procesy podporované algoritmi umelej inteligencie. Platforma umožňuje automatizované spracovanie textových odpovedí respondentov, podporu analytikov pri interpretácii dát a využívanie pokročilých analytických nástrojov nad štatistickými údajmi. Analytici získavajú podporu vo forme AI nástrojov, ktoré umožňujú rýchlejšie vyhľadávanie informácií, interpretáciu údajov a generovanie analytických výstupov. Dôležitou súčasťou tejto alternatívy je aj rozšírenie analytických služieb pre externých používateľov. Platforma umožní verejnosti a výskumným organizáciám jednoduchšie pracovať s publikovanými štatistickými údajmi prostredníctvom moderných analytických nástrojov, sémantického vyhľadávania a automatizovaných interpretácií dát. Používatelia tak budú môcť efektívnejšie analyzovať dostupné štatistické informácie bez potreby hlbokých technických znalostí. Alternatíva zároveň zohľadňuje vysoké požiadavky na ochranu štatistického tajomstva a osobných údajov. Platforma preto implementuje princípy bezpečného sprístupňovania údajov, najmä prostredníctvom práce s publikovanými agregovanými údajmi a využívania syntetických datasetov pre výskumné účely. Tým sa zabezpečuje, že citlivé mikrodáta zostávajú chránené v internom prostredí úradu, zatiaľ čo verejnosti sú sprístupňované analytické možnosti nad bezpečnými dátovými reprezentáciami. Táto alternatíva zároveň vytvára technologický a procesný základ pre ďalší rozvoj analytických služieb založených na umelej inteligencii, podporuje efektívnejšie využívanie štatistických údajov v spoločnosti a prispieva k zvyšovaniu dostupnosti a využiteľnosti oficiálnej štatistiky.



Obrázok 4 Znáznornenie alternatív riešenia v biznis vrstve architektúry

	Kritérium	Zdôvodnenie kritéria	ŠÚ SR	Orgány verejnej správy	Verejnosť / výskumníci
BIZNIS VRSTVA	Kritérium A – Zachovanie štatistickej dôverylosti (KO)	Spracovanie údajov musí zachovať ochranu štatistického tajomstva a zabrániť identifikácii respondentov	X	X	X
	Kritérium B – Súlad s GDPR a legislatívou (KO)	Spracovanie údajov musí byť v súlade s GDPR a národnou legislatívou ochrany údajov	X	X	
	Kritérium C – Zníženie manuálnej práce analytikov (KO)	Riešenie musí automatizovať časovo náročné analytické procesy pri spracovaní štatistických zisťovaní	X		
	Kritérium D – Možnosť využitia AI pri spracovaní dát (KO)	Riešenie musí umožniť využitie AI modelov pri spracovaní štatistických údajov	X		
	Kritérium E – Zlepšenie dostupnosti štatistických údajov	Projekt má zlepšiť dostupnosť štatistických údajov prostredníctvom moderných digitálnych nástrojov	X	X	X
	Kritérium F – Podpora výskumu a analytiky	Projekt má podporiť využitie štatistických údajov pre výskumné a analytické účely		X	X

Tabuľka 11 Kritériá MCA v biznis vrstve

Zoznam kritérií	B0	Spôsob Dosiahnutia	B1	Spôsob Dosiahnutia	B2	Spôsob Dosiahnutia
Kritérium A (KO)	Áno	Spracovanie údajov zostáva v existujúcich systémoch ŠÚ SR	Áno	Spracovanie údajov v interných systémoch	Áno	Ochrana mikroúdajov cez oddelené dátové zóny
Kritérium B (KO)	Áno	Spracovanie údajov prebieha v súlade s existujúcimi legislatívnymi pravidlami	Áno	Spracovanie údajov v kontrolovanom prostredí	Áno	Implementácia GDPR a bezpečnostných kontrol
Kritérium C (KO)	Nie	Procesy zostávajú manuálne	Áno	Automatizácia klasifikácie textových odpovedí	Áno	AI klasifikácia a analytické nástroje
Kritérium D (KO)	Nie	AI nástroje sa nepoužívajú	Áno	Využitie AI modelov pre klasifikáciu	Áno	Využitie AI modelov pre analýzu a interpretáciu
Kritérium E	Nie	Nové analytické služby pre verejnosť nevznikajú	Nie	Riešenie je zamerané iba na interné procesy	Áno	Verejný AI analytik a sémantické vyhľadávanie
Kritérium F	Nie	Nové nástroje pre výskum nevznikajú	Nie	Platforma neposkytuje nástroje pre výskum	Áno	Syntetické datasetové nástroje

Tabuľka 12 Vyhodnotenie MCA v biznis vrstve

Nulový variant B0 predstavuje zachovanie súčasného stavu bez implementácie navrhovaného riešenia. Tento variant nespĺňa kritériá zamerané na zníženie manuálnej práce analytikov ani využitie umelej inteligencie pri spracovaní dát, a preto neumožňuje riešiť identifikované problémy. Alternatíva B1 síce spĺňa kritické kritériá súvisiace s automatizáciou interných procesov, avšak neprináša zlepšenie dostupnosti údajov pre verejnosť ani podporu výskumných aktivít. Alternatíva B2 spĺňa všetky kritické kritériá a zároveň prináša dodatočné prínosy v oblasti dostupnosti štatistických údajov, analytických služieb a podpory výskumu. Na základe výsledkov multikritériálnej analýzy bola preto ako preferovaná alternatíva vybraná alternatíva B2 – komplexná AI platforma.

Na základe výsledkov multikritériálnej analýzy bola ako preferovaná biznis alternatíva vybraná alternatíva B2 – komplexná AI platforma pre interné aj verejné služby, ktorá najlepšie naplňuje ciele projektu, prináša najvyšší prínos pre jednotlivých stakeholderov a zároveň vytvára predpoklady pre ďalší rozvoj analytických služieb založených na umelej inteligencii.

5.1.2 Stanovenie alternatív v aplikačnej vrstve architektúry

Na základe výsledku multikritériálnej analýzy v biznisovej vrstve architektúry boli definované alternatívy riešenia v aplikačnej vrstve. Alternatívy reflektujú zvolený biznis scenár modernizácie procesov spracovania a publikovania štatistických údajov prostredníctvom nástrojov umelej inteligencie a dátovej analytiky. Pri návrhu alternatív v aplikačnej vrstve bolo cieľom identifikovať taký rozsah aplikačných modulov, ktorý umožní dosiahnuť definované ciele projektu, pričom jednotlivé alternatívy sa odlišujú najmä rozsahom implementovaných funkcionalít a mierou podpory jednotlivých scenárov použitia. Alternatívy v aplikačnej vrstve boli definované nasledovne.

Alternatíva A0 – nulový variant

Nulový variant predstavuje zachovanie súčasnej aplikačnej architektúry bez implementácie AI platformy. Existujúce informačné systémy ŠÚ SR naďalej zabezpečujú zber, spracovanie a publikovanie štatistických údajov a nevzniknú aplikačné moduly pre interného AI asistenta, automatickú klasifikáciu textových odpovedí, analytika nad mikroúdajmi, generovanie syntetických údajov ani verejné AI služby nad publikovanými údajmi. Tento variant slúži ako referenčný scenár pre porovnanie prínosov navrhovaného riešenia v rámci analýzy nákladov a prínosov.

Alternatíva A1 – základná aplikačná platforma

Alternatíva A1 predstavuje implementáciu základnej aplikačnej platformy zameranej predovšetkým na podporu interných analytických procesov Štatistického úradu Slovenskej republiky. Cieľom tejto alternatívy je čiastočná automatizácia vybraných analytických úloh a zvýšenie efektivity spracovania dát prostredníctvom využitia nástrojov umelej inteligencie, pričom rozsah implementovaných funkcionalít je obmedzený na interné potreby organizácie. V rámci tejto alternatívy sa predpokladá implementácia základných aplikačných modulov, ktoré umožnia automatizovať najnáročnejšie manuálne procesy súvisiace so spracovaním štatistických údajov. Kľúčovým komponentom riešenia je modul automatickej klasifikácie textových odpovedí, ktorý umožní efektívnejšie spracovanie otvorených odpovedí respondentov a ich automatické priradovanie ku klasifikačným schémam používaným v štatistike, napríklad ISCO alebo NACE. Tento modul umožňuje výrazne znížiť časovú náročnosť manuálneho kódovania odpovedí a zároveň zvyšuje konzistentnosť výsledkov klasifikácie. Súčasťou riešenia je aj interný analytický nástroj určený pre analytikov Štatistického úradu SR, ktorý umožňuje efektívnejšiu prácu s dátami, podporu analytických úloh a využitie modelov umelej inteligencie pri spracovaní dát. Tento nástroj poskytuje analytikom centralizované prostredie pre prácu s dátovými zdrojmi a pre aplikáciu analytických metód pri spracovaní štatistických údajov. Ďalším významným komponentom tejto alternatívy je modul správy modelov a experimentov (MLOps), ktorý umožňuje riadenie životného cyklu modelov umelej inteligencie. Tento modul zabezpečuje evidenciu experimentov, správu verzií modelov a kontrolované nasadzovanie modelov do analytických procesov, čím podporuje systematické využívanie umelej inteligencie v analytickej činnosti organizácie.

Súčasťou tejto alternatívy je aj základná dátová integrácia medzi existujúcimi informačnými systémami Štatistického úradu SR. Táto integrácia zabezpečuje prístup analytických nástrojov k relevantným dátovým zdrojom a umožňuje efektívnejšie využitie existujúcich dátových aktív organizácie. Napriek uvedeným prínosom je rozsah tejto alternatívy obmedzený výhradne na interné analytické procesy. Platforma neposkytuje nové analytické služby pre externých používateľov, ako sú výskumné inštitúcie, médiá alebo verejnosť, a neobsahuje ani pokročilé nástroje pre interaktívnu prácu s publikovanými štatistickými údajmi. Táto alternatíva preto predstavuje minimalistický scenár implementácie aplikačnej architektúry, ktorý síce zlepšuje efektivitu interných procesov, avšak nevyužíva plný potenciál analytickej platformy pre širšie využitie štatistických údajov.

Alternatíva A2 – rozšírená analytická platforma

Alternatíva A2 predstavuje rozšírenie základnej aplikačnej platformy o nové analytické služby určené nielen pre interných používateľov, ale aj pre externých používateľov, ako sú výskumné inštitúcie, analytické organizácie, médiá alebo širšia verejnosť. Cieľom tejto alternatívy je zvýšiť dostupnosť a využiteľnosť štatistických údajov prostredníctvom moderných analytických nástrojov a zároveň podporiť efektívnejšie spracovanie dát v interných procesoch Štatistického úradu SR. V rámci tejto alternatívy sú implementované všetky základné moduly potrebné na automatizáciu interných analytických procesov. Kľúčovým komponentom je modul automatickej klasifikácie textových odpovedí, ktorý umožňuje efektívne spracovanie otvorených odpovedí respondentov a ich automatické priradovanie k relevantným klasifikačným schémam, napríklad ISCO alebo NACE. Tento modul výrazne znižuje manuálnu záťaž analytikov a prispieva k zvýšeniu konzistentnosti klasifikácie údajov.

Dôležitou súčasťou platformy je aj interný AI analytik nad mikrodátami, ktorý poskytuje analytikom nástroje na efektívnejšiu prácu s internými dátovými zdrojmi. Tento nástroj umožňuje využívať modely umelej inteligencie pri analýze dát, identifikácii vzorcov alebo príprave analytických výstupov, čím podporuje modernizáciu analytických procesov organizácie. Súčasťou riešenia je aj modul správy modelov a experimentov (MLOps), ktorý zabezpečuje riadenie životného cyklu modelov umelej inteligencie. Tento modul umožňuje evidenciu experimentov, správu verzií modelov, monitorovanie ich výkonu a kontrolované nasadzovanie modelov do produkčného prostredia. Na rozdiel od minimalistickej alternatívy však táto alternatíva rozširuje platformu aj o služby určené pre externých používateľov. Súčasťou riešenia je verejný analytický nástroj pracujúci s publikovanými agregovanými údajmi, ktorý umožňuje interaktívnu analýzu štatistických datasetov bez prístupu k citlivým mikroúdajom. Tento nástroj poskytuje používateľom možnosť jednoduchšieho vyhľadávania, kombinovania a interpretácie štatistických údajov.

Platforma zároveň obsahuje modul sémantického vyhľadávania štatistických údajov, ktorý umožňuje používateľom vyhľadávať relevantné datasetové výstupy alebo analytické informácie prostredníctvom prirodzeného jazyka. Tento prístup výrazne zjednodušuje prácu s rozsiahlymi štatistickými databázami a zvyšuje dostupnosť údajov pre širšie spektrum používateľov. Ďalším komponentom je modul automatického generovania textových interpretácií štatistických výstupov, ktorý umožňuje automatizovanú tvorbu základných analytických komentárov alebo opisov výsledkov štatistických analýz. Tento modul môže podporiť analytikov pri príprave publikácií alebo sprievodných textov k štatistickým výstupom. Alternatíva A2 tak výrazne zlepšuje dostupnosť štatistických údajov pre verejnosť a podporuje ich analytické využívanie v akademickom, analytickom aj mediálnom prostredí, pričom zároveň zachováva bezpečnostné princípy ochrany citlivých údajov tým, že externé služby pracujú výhradne s publikovanými agregovanými údajmi.

Alternatíva A3 – komplexná AI analytická platforma (preferovaný variant)

Alternatíva A3 predstavuje plnú implementáciu navrhovanej aplikačnej architektúry, ktorá podporuje modernizáciu interných analytických procesov Štatistického úradu SR a zároveň rozširuje dostupnosť analytických služieb pre externých používateľov. Táto alternatíva nadväzuje na funkcionality definované v predchádzajúcich variantoch a rozširuje ich o ďalšie moduly, ktoré umožňujú komplexné využitie umelej inteligencie pri spracovaní, interpretácii a publikovaní štatistických údajov. Základom tejto alternatívy je implementácia všetkých nutných modulov, ktoré zabezpečujú automatizáciu kľúčových analytických procesov. Modul automatickej klasifikácie textových odpovedí umožňuje efektívne spracovanie otvorených odpovedí respondentov a ich priradovanie k štandardizovaným klasifikačným schémam, napríklad ISCO alebo NACE. Tento modul výrazne znižuje manuálnu záťaž analytikov a zvyšuje konzistentnosť klasifikácie údajov. Interný AI analytik nad mikrodátami poskytuje analytikom prostredie na pokročilú analýzu dát s využitím modelov umelej inteligencie, pričom umožňuje identifikáciu vzorcov v dátach, podporu analytických úloh a prípravu analytických výstupov. Súčasťou riešenia je aj modul správy modelov a experimentov (MLOps), ktorý zabezpečuje riadenie životného cyklu modelov umelej inteligencie vrátane evidencie experimentov, správy verzií modelov a kontrolovaného nasadzovania modelov do produkčného prostredia. Súčasťou alternatívy A3 je aj interný AI asistent pre zamestnancov, ktorý poskytuje metodickú podporu nad internou dokumentáciou, metodikami a znalostnou bázou úradu.

Na rozdiel od predchádzajúcich alternatív však táto alternatíva zahŕňa širší rozsah preferovaných modulov, ktoré rozširujú funkcionality platformy a vytvárajú nové možnosti využívania štatistických údajov. Platforma obsahuje verejný AI analytik nad publikovanými agregovanými údajmi, ktorý umožňuje externým používateľom interaktívne analyzovať publikované štatistické údaje bez prístupu k citlivým mikrodátom. Súčasťou riešenia je aj modul sémantického vyhľadávania štatistických údajov, ktorý umožňuje používateľom vyhľadávať relevantné datasetové výstupy alebo analytické informácie prostredníctvom prirodzeného jazyka. Ďalším významným prvkom tejto alternatívy je modul automatického generovania textových interpretácií štatistických výstupov, ktorý umožňuje automatizovanú tvorbu základných analytických komentárov alebo opisov výsledkov štatistických analýz. Tento modul môže podporiť analytikov pri príprave publikácií alebo sprievodných textov k štatistickým výstupom a zároveň zlepšiť dostupnosť interpretácií štatistických údajov pre širšie spektrum používateľov. Súčasťou platformy je aj modul generovania syntetických mikrodát, ktorý umožňuje vytvárať datasetové súbory zachovávajúce štatistické vlastnosti pôvodných údajov bez rizika identifikácie konkrétnych respondentov. Takto generované syntetické datasetové súbory môžu byť využívané najmä na výskumné alebo analytické účely bez potreby sprístupnenia citlivých mikrodát. Platforma zároveň umožňuje viacjazyčnú interpretáciu štatistických výstupov, čím podporuje širšiu dostupnosť štatistických informácií pre rôzne skupiny používateľov vrátane medzinárodnej odbornej komunity.

Alternatíva A3 tak predstavuje najkomplexnejšie riešenie aplikačnej architektúry, ktoré zabezpečuje modernizáciu interných analytických procesov, rozširuje dostupnosť štatistických údajov pre verejnosť a výskumnú komunitu a zároveň vytvára technologický základ pre ďalší rozvoj analytických služieb založených na umelej inteligencii. Z uvedených dôvodov bola táto alternatíva identifikovaná ako preferovaný variant riešenia.

Vyhodnotenie alternatív v aplikačnej vrstve

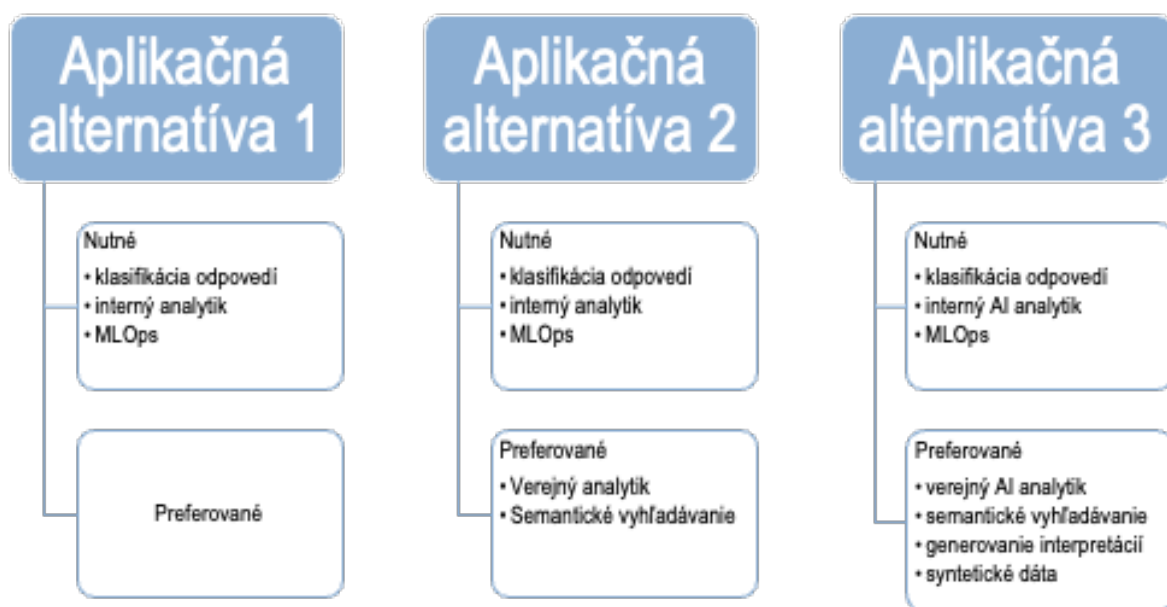
Alternatívy definované v aplikačnej vrstve boli posudzované z hľadiska rozsahu implementovaných funkcionalít, prínosov pre jednotlivé skupiny stakeholderov, podpory interných analytických procesov a súladu s cieľmi projektu definovanými v biznisovej vrstve architektúry. Hodnotená bola najmä schopnosť jednotlivých alternatív prispieť k automatizácii spracovania štatistických údajov, zlepšeniu dostupnosti analytických nástrojov pre interných používateľov a k rozšíreniu analytických služieb poskytovaných verejnosti.

Alternatíva A0 predstavuje nulový variant, ktorý zachováva súčasný stav bez implementácie novej aplikačnej platformy. Tento variant neumožňuje riešiť identifikované problémy súvisiace s manuálnym spracovaním údajov, obmedzenými analytickými nástrojmi ani s rastúcimi požiadavkami používateľov na dostupnosť a interpretáciu štatistických údajov. Z tohto dôvodu slúži nulový variant iba ako referenčný scenár pre porovnanie prínosov navrhovaného riešenia v rámci analýzy nákladov a prínosov.

Alternatíva A1 predstavuje minimalistický variant riešenia, ktorý obsahuje iba nevyhnutné aplikačné moduly potrebné na podporu interných analytických procesov. Táto alternatíva umožňuje čiastočnú automatizáciu vybraných analytických úloh a znižuje manuálnu záťaž analytikov, avšak neumožňuje rozšírenie analytických služieb pre externých používateľov. Z uvedeného dôvodu bude táto alternatíva použitá ako minimalistický variant v rámci analýzy nákladov a prínosov.

Alternatíva A2 rozširuje základnú platformu o analytické služby pre externých používateľov a umožňuje efektívnejšiu prácu s publikovanými štatistickými údajmi. Táto alternatíva významne zlepšuje dostupnosť štatistických informácií pre verejnosť a podporuje ich analytické využívanie v akademickom a analytickom prostredí, avšak rozsah implementovaných funkcionalít je stále obmedzený v porovnaní s komplexným riešením.

Na základe celkového posúdenia bola ako preferovaná alternatíva vybraná alternatíva A3 – komplexná AI analytická platforma, ktorá zabezpečuje najvyššiu mieru automatizácie analytických procesov, podporuje systematické využívanie umelej inteligencie pri spracovaní štatistických údajov a zároveň umožňuje výrazné rozšírenie analytických služieb pre verejnosť a výskumnú komunitu. Táto alternatíva najlepšie naplňa strategické ciele projektu a vytvára technologický základ pre ďalší rozvoj analytických a dátových služieb v prostredí Štatistického úradu SR.



Obrázok 5 Znáznornenie alternatív riešenia v aplikačnej vrstve

5.1.3 Stanovenie alternatív v technologickej vrstve architektúry

Alternatívy v technologickej vrstve architektúry nadväzujú na výsledok analýzy alternatív v aplikačnej vrstve a reflektujú identifikované požiadavky na spracovanie dát, bezpečnosť, výkon a integráciu s existujúcimi

informačnými systémami Štatistického úradu SR. Pri návrhu technologických alternatív bolo potrebné zohľadniť najmä charakter spracovávaných údajov, požiadavky na ochranu štatistického tajomstva, požiadavky na ochranu osobných údajov podľa GDPR a legislatívne požiadavky na zabezpečenie informačných systémov verejnej správy. Technologická architektúra bola navrhovaná v súlade s metodickým usmernením pre klasifikáciu informačných systémov verejnej správy, pričom jednotlivé aplikačné moduly boli posudzované z hľadiska citlivosti spracovávaných údajov, požiadaviek na dostupnosť a bezpečnostných požiadaviek. Na základe uvedených požiadaviek boli identifikované tri základné technologické alternatívy.

Alternatíva T0 – nulový variant

Nulový variant predstavuje zachovanie existujúceho technologického prostredia bez implementácie novej infraštruktúry alebo technologickej platformy. Analytické procesy zostávajú realizované v existujúcich systémoch a prostredníctvom manuálnych postupov bez implementácie nových technologických komponentov potrebných pre prevádzku modelov umelej inteligencie. Tento variant neumožňuje implementáciu navrhovaných aplikačných modulov ani využitie umelej inteligencie pri spracovaní štatistických údajov a slúži ako referenčný scenár pre porovnanie prínosov navrhovaného riešenia v rámci analýzy nákladov a prínosov.

Alternatíva T1 – nasadenie riešenia vo vládnom cloude

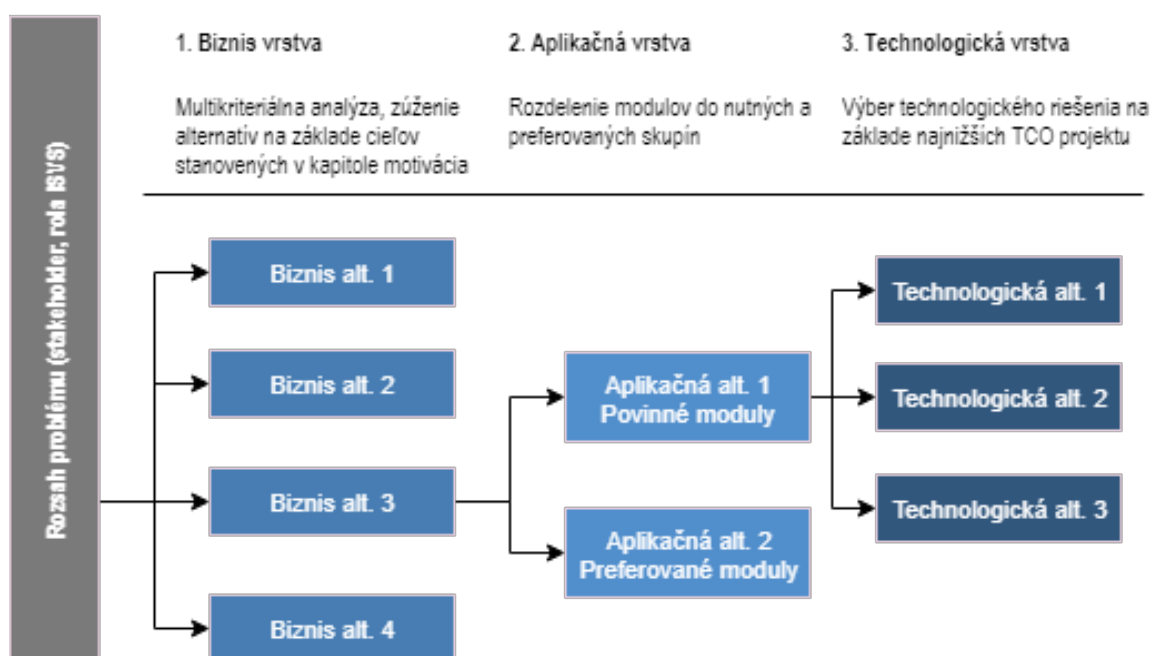
Táto alternatíva predpokladá implementáciu technologickej infraštruktúry platformy prostredníctvom služieb vládneho cloudu dostupných v katalógu cloudových služieb. Výhodou tohto prístupu je možnosť flexibilného škálovania výpočtových kapacít a využitie existujúcich cloudových služieb pre prevádzku aplikačných komponentov. Zároveň by bolo možné využiť centralizovanú správu infraštruktúry a štandardizované cloudové služby. Napriek týmto výhodám však tento variant nie je vhodný vzhľadom na charakter spracovávaných údajov. Navrhovaná platforma pracuje s citlivými štatistickými mikroúdajmi, ktoré podliehajú režimu ochrany štatistického tajomstva a prísnyh pravidlám ochrany osobných údajov. Spracovanie týchto údajov vyžaduje vysokú mieru kontroly nad infraštruktúrou, sieťovou architektúrou a bezpečnostnými mechanizmami. V prípade plného nasadenia riešenia vo vládnom cloude by nebolo možné zabezpečiť požadovanú úroveň izolácie medzi spracovaním citlivých mikroúdajov a verejne dostupnými analytickými službami. Zároveň by bolo potrebné implementovať dodatočné bezpečnostné opatrenia na zabezpečenie ochrany citlivých údajov, čo by výrazne zvyšovalo komplexnosť riešenia a jeho prevádzkové náklady.

Alternatíva plného nasadenia riešenia vo vládnom cloude bola detailnejšie posudzovaná aj z pohľadu bezpečnostných, prevádzkových a technologických požiadaviek navrhovanej platformy. Analýza identifikovala viacero zásadných obmedzení, ktoré znemožňujú využitie vládneho cloudu ako primárneho prostredia pre implementáciu riešenia. Prvým významným faktorom je charakter spracovávaných údajov. Navrhovaná platforma pracuje so štatistickými mikroúdajmi, ktoré podliehajú režimu ochrany štatistického tajomstva a prísnyh pravidlám ochrany osobných údajov. Spracovanie takýchto údajov vyžaduje plnú kontrolu nad infraštruktúrou, sieťovou architektúrou a bezpečnostnými mechanizmami, vrátane možnosti implementovať izolované spracovateľské prostredie bez priamej externej konektivity. Takýto model je štandardne implementovaný v prostredí interných dátových centier štatistických úradov a je založený na princípe striktnnej izolácie prostredí spracovania mikroúdajov od prostredí poskytujúcich verejné služby. Druhým faktorom je požiadavka na architektúru založenú na prísnom oddelení dátových domén. Navrhované riešenie je koncipované ako dvojzónový model, v ktorom je spracovanie citlivých mikroúdajov realizované v internej chránenej zóne a analytické služby pre verejnosť sú prevádzkované v oddelenej zóne pracujúcej výhradne s publikovanými agregovanými údajmi. Medzi týmito zónami je povolený iba kontrolovaný jednosmerný tok údajov po metodologickej kontrole a posúdení rizika reidentifikácie. Takýto architektonický model vyžaduje vysokú mieru kontroly nad sieťovou architektúrou a bezpečnostnými mechanizmami, ktorú nie je možné plnohodnotne zabezpečiť v prípade centralizovaného cloudového prostredia. Tretím faktorom sú technologické požiadavky na prevádzku modelov umelej inteligencie. Platforma predpokladá využívanie výpočtovo náročných procesov, najmä tréning modelov umelej inteligencie a experimentálnu prácu s modelmi v rámci MLOps prostredia. Takéto procesy vyžadujú špecializovanú infraštruktúru vrátane GPU akcelerácie, vysokokapacitných dátových úložísk a flexibilného riadenia výpočtových zdrojov. V prostredí vládneho cloudu nie je v súčasnosti dostupná infraštruktúra optimalizovaná pre tento typ výpočtových úloh v rozsahu potrebnom pre prevádzku analytickej platformy.

Z uvedených dôvodov bola alternatíva plného nasadenia riešenia vo vládnom cloude vyhodnotená ako nevhodná. Preferovaným riešením je on-premise dvojzónová architektúra v dátovom centre Štatistického úradu SR, v ktorom je prevádzkovaná interná aj verejná zóna platformy.

Alternatíva T2 – on-premise dvojzónová architektúra v dátovom centre ŠÚ SR (preferovaný variant)

Táto alternatíva predpokladá prevádzku oboch zón platformy na on-premise infraštruktúre v dátovom centre Štatistického úradu SR. Technologická architektúra je navrhnutá ako dvojzónový model, ktorý zabezpečuje oddelenie spracovania citlivých údajov od verejne dostupných analytických služieb. Interná zóna je určená na spracovanie citlivých mikroúdajov, tréning modelov umelej inteligencie a prevádzku interných analytických nástrojov a nie je priamo prístupná z externých sietí. Verejná zóna je určená pre služby pracujúce s publikovanými agregovanými údajmi, dátovými kockami, metaúdajmi a schválenými syntetickými datasetmi. V tejto zóne sú prevádzkované analytické nástroje dostupné pre verejnosť, napríklad nástroje pre interaktívnu analýzu štatistických údajov alebo sémantické vyhľadávanie. Prenos údajov medzi týmito zónami je striktné kontrolovaný a je umožnený iba smerom z internej zóny do verejnej zóny, pričom prenášané údaje podliehajú metodologickej kontrole a posúdeniu rizika reidentifikácie. Verejný cloud v rámci EÚ sa využíva len pre izolované pilotné prostredie, ktoré slúži na overenie konceptu vybraných verejných služieb v izolovanom prostredí verejného cloudu v rámci EÚ nad necitlivými, syntetickými a verejnými údajmi. Takto navrhnutá architektúra umožňuje bezpečné spracovanie citlivých údajov v kontrolovanom prostredí Štatistického úradu SR a zároveň poskytuje dostatočnú flexibilitu pre prevádzku verejných analytických služieb.



Obrázok 6 Znáznornenie alternatív riešenia v technologickej vrstve

Výsledok posúdenia technologických alternatív

Na základe posúdenia technologických alternatív bola ako preferovaná vybraná alternatíva T2 – on-premise dvojzónová architektúra v dátovom centre ŠÚ SR, v ktorom sú prevádzkované obe zóny platformy. Verejný cloud v rámci EÚ sa využije len pre izolované pilotné prostredie. Táto alternatíva najlepšie spĺňa požiadavky na ochranu citlivých údajov, bezpečnosť spracovania štatistických dát a zároveň umožňuje efektívnu prevádzku analytických služieb pre interných aj externých používateľov. Alternatíva plného nasadenia vo vládnom cloudu bola z ďalšieho posudzovania vylúčená z dôvodu nevhodnosti pre spracovanie citlivých štatistických mikroúdajov a nedostatočnej možnosti zabezpečiť požadovanú úroveň izolácie medzi internými dátovými procesmi a verejnými analytickými službami.

5.2 Náhľad architektúry a popis budúceho cieľového produktu

Budúcim cieľovým produktom projektu je robustná, bezpečná a škálovateľná AI platforma pre Štatistický úrad SR (ŠÚ SR), ktorá integruje moderné technológie umelej inteligencie a strojového učenia do procesov oficiálnej štatistiky. Produkt je navrhnutý podľa európskeho rámca AIML4OS a rešpektuje striktné bezpečnostné obmedzenia týkajúce sa rezidencie dát na území SR. Navrhované riešenie je implementované postupným spôsobom prostredníctvom pilotnej fázy a následného plného nasadenia v infraštruktúre ŠÚ SR. Pred samotnou implementáciou cieľového riešenia sa realizuje pilotná fáza, ktorej cieľom je overiť vybrané funkcionality platformy bez potreby okamžitých investícií do infraštruktúry. Pilotná fáza slúži na overenie konceptu vybraných verejných služieb v izolovanom prostredí verejného cloudu v rámci EÚ nad necitlivými, syntetickými a verejnými údajmi. Pilotné prostredie je úplne izolované od interných systémov ŠÚ SR. Neexistuje žiadna konektivita do interných databáz ani do chránenej zóny mikroúdajov. Cieľom pilotnej fázy je overenie kvality odpovedí modelu, definovanie promptových šablón, nastavenie základných metrik výkonu a príprava prevádzkových postupov (runbookov) pre následné nasadenie riešenia v infraštruktúre úradu. Po úspešnom overení pilotnej fázy nasleduje implementácia cieľovej platformy v infraštruktúre dátového centra ŠÚ SR, kde je riešenie prevádzkované v plne kontrolovanom on-premise prostredí. Realizácia prebieha v dvoch inkrementoch a do produkčnej prevádzky sa platforma nasadzuje ako celok po dodaní Inkrementu 2. Na rozdiel od pilotnej fázy produkčné prostredie pracuje aj s citlivými mikroúdajmi, ktoré sa spracúvajú výhradne v chránenej internej zóne v infraštruktúre ŠÚ SR.

Biznis architektúra

Biznis architektúra platformy je definovaná prostredníctvom súboru kľúčových scenárov použitia (Use Cases), ktoré pokrývajú hlavné oblasti využitia umelej inteligencie v prostredí ŠÚ SR. Navrhované riešenie podporuje osem hlavných scenárov použitia, ktoré prispievajú k zvyšovaniu efektivity interných analytických procesov a zároveň rozširujú možnosti interakcie s verejnosťou. Medzi najvýznamnejšie scenáre patrí automatizované kódovanie textových odpovedí respondentov (UC-02), ktoré výrazne znižuje manuálnu záťaž analytikov pri spracovaní údajov. Interný AI asistent (UC-01) poskytuje metodickú podporu zamestnancom pri práci so štatistickými metodikami a klasifikáciami. Bezpečný AI analytik nad mikroúdajmi (UC-04) umožňuje kontrolované analytické dopyty nad citlivými údajmi pri zachovaní pravidiel štatistického utajenia. Na strane verejných služieb platforma poskytuje moderné nástroje na prístup k štatistickým údajom, napríklad verejného AI analytika (UC-05), ktorý umožňuje interaktívne analytické dopyty nad publikovanými agregovanými údajmi. Ďalšie scenáre zahŕňajú inteligentné sémantické vyhľadávanie v publikovaných datasetoch (UC-06), automatické generovanie textových súhrnov štatistických výstupov (UC-07) a viacjazyčnú disemináciu údajov (UC-08).

Aplikačná architektúra

Z aplikačného hľadiska je platforma organizovaná do štyroch základných vrstiev:

- dátová vrstva,
- AI/ML vrstva,
- aplikačná vrstva,
- prezentačná vrstva.

Dátová vrstva zabezpečuje prístup k primárnym štatistickým údajom, publikovaným agregátom a metaúdajom. AI/ML vrstva poskytuje prostredie pre tréning, správu a prevádzku modelov umelej inteligencie vrátane nástrojov pre MLOps. Aplikačná vrstva implementuje jednotlivé analytické služby a biznis logiku platformy. Prezentačná vrstva zabezpečuje interakciu používateľov so systémom prostredníctvom webových portálov, analytických rozhraní a aplikačných API. Architektúra platformy je zároveň rozdelená do dvoch logických bezpečnostných zón. Zóna A je určená pre spracovanie citlivých údajov, najmä mikroúdajov a interných analytických procesov. Zóna B poskytuje verejné analytické služby a prístup k publikovaným agregovaným údajom. Komunikácia medzi zónami je prísne riadená prostredníctvom kontrolovaných dátových prestupov a bezpečnostných mechanizmov. Všetky aplikačné služby platformy sú publikované prostredníctvom stabilných a verzovaných rozhraní podľa princípu API-first, čo zabezpečuje vysokú mieru interoperability a umožňuje jednoduchú integráciu s ďalšími systémami.

Technologická architektúra

Technologická architektúra platformy je založená na modernom cloud-native prístupe využívajúcom kontajnerizačné technológie. Jednotlivé komponenty systému sú prevádzkované v kontajneroch v prostredí

orchestrácie kontajnerov, ktoré zabezpečuje škálovanie, vysokú dostupnosť a odolnosť systému. Pre správu životného cyklu modelov umelej inteligencie sa využívajú nástroje z oblasti MLOps vrátane registra modelov. Vyhľadávacie a analytické funkcie platformy sú podporené fulltextovým a vektorovým vyhľadávacím nástrojom, ktorý umožňuje efektívne indexovanie a sémantické vyhľadávanie nad publikovanými údajmi. Bezpečnostná architektúra platformy využíva správu tajomstiev a kryptografických kľúčov a nástroj na vyhodnocovanie prístupových politík pre vynucovanie bezpečnostných politík na úrovni služieb a API. Celá infraštruktúra platformy je umiestnená v dátovom centre ŠÚ SR, čím je zabezpečená plná kontrola nad spracovaním údajov a dodržanie požiadaviek na rezidenciu dát. Zóna A je navyše sieťovo izolovaná od externej konektivity, aby sa minimalizovalo riziko úniku citlivých údajov.

Realizácia projektu predstavuje významný krok v modernizácii informačných systémov ŠÚ SR. Zavedením systematického riadenia dátových zdrojov a modelov umelej inteligencie (MLOps) vzniká nová analytická vrstva nad existujúcimi systémami úradu. Navrhovaná architektúra zároveň zavádza jasné oddelenie medzi transakčnými systémami spracovania údajov a analytickými službami poskytovanými používateľom. Toto oddelenie je realizované prostredníctvom riadených dátových prestupov, bezpečnostných brán a mechanizmov ochrany údajov, čím sa zvyšuje ochrana štatistického tajomstva pri súčasnom rozširovaní prístupu k štatistickým údajom pre verejnosť.

Architektonický návrh AI platformy pre ŠÚ SR je postavený na striktnom rozdelení do dvoch bezpečnostných domén, ktoré sa líšia typom spracúvaných dát, mierou zabezpečenia a účelom ich využitia. Hlavné rozdiely medzi Zónou A a Zónou B sú nasledovné:

1. Charakter a klasifikácia spracúvaných dát

- Zóna A (Interná chránená zóna): Je určená výhradne na spracovanie citlivých mikroúdajov (individuálne záznamy o osobách alebo podnikoch) a interných agregátov. Mikroúdaje v tejto zóne podliehajú štatistickej dôvernosti a nariadeniu GDPR.
- Zóna B (Verejná zóna): Pracuje výhradne s publikovanými agregovanými údajmi, dátovými kockami, metaúdajmi, schválenými syntetickými datasetmi a verejným obsahom. Tieto dáta sú schválené na zverejnenie a nesmú obsahovať žiadne osobne identifikovateľné údaje (PII) ani kombinácie vedúce k reidentifikácii.

2. Účel a cieľové skupiny

- Zóna A: Hostí všetky interné AI/ML pracovné zaťaženia, ako je tréning modelov nad citlivými dátami, automatická klasifikácia odpovedí respondentov (UC-02) a bezpečná analýza mikroúdajov pre interných metodikov (UC-04).
- Zóna B: Je optimalizovaná na poskytovanie služieb verejnosti, médiám a partnerom. Zabezpečuje funkcie ako verejný AI analytik (UC-05), inteligentné vyhľadávanie (UC-06) a viacjazyčnú disemináciu (UC-08).

3. Sieťová konektivita a prístup

- Zóna A: Nemá žiadnu externú sieťovú konektivitu pre citlivý obsah. Prístup do nej je prísne riadený, auditovaný a chránený viacfaktorovou autentifikáciou (MFA) a princípom najmenších oprávnení.
- Zóna B: Je vystavená do internetu cez verejnú API bránu (CMP-B8), ktorá je chránená pomocou WAF (Web Application Firewall), filtrácie obsahu a obmedzovania počtu požiadaviek (rate limiting).

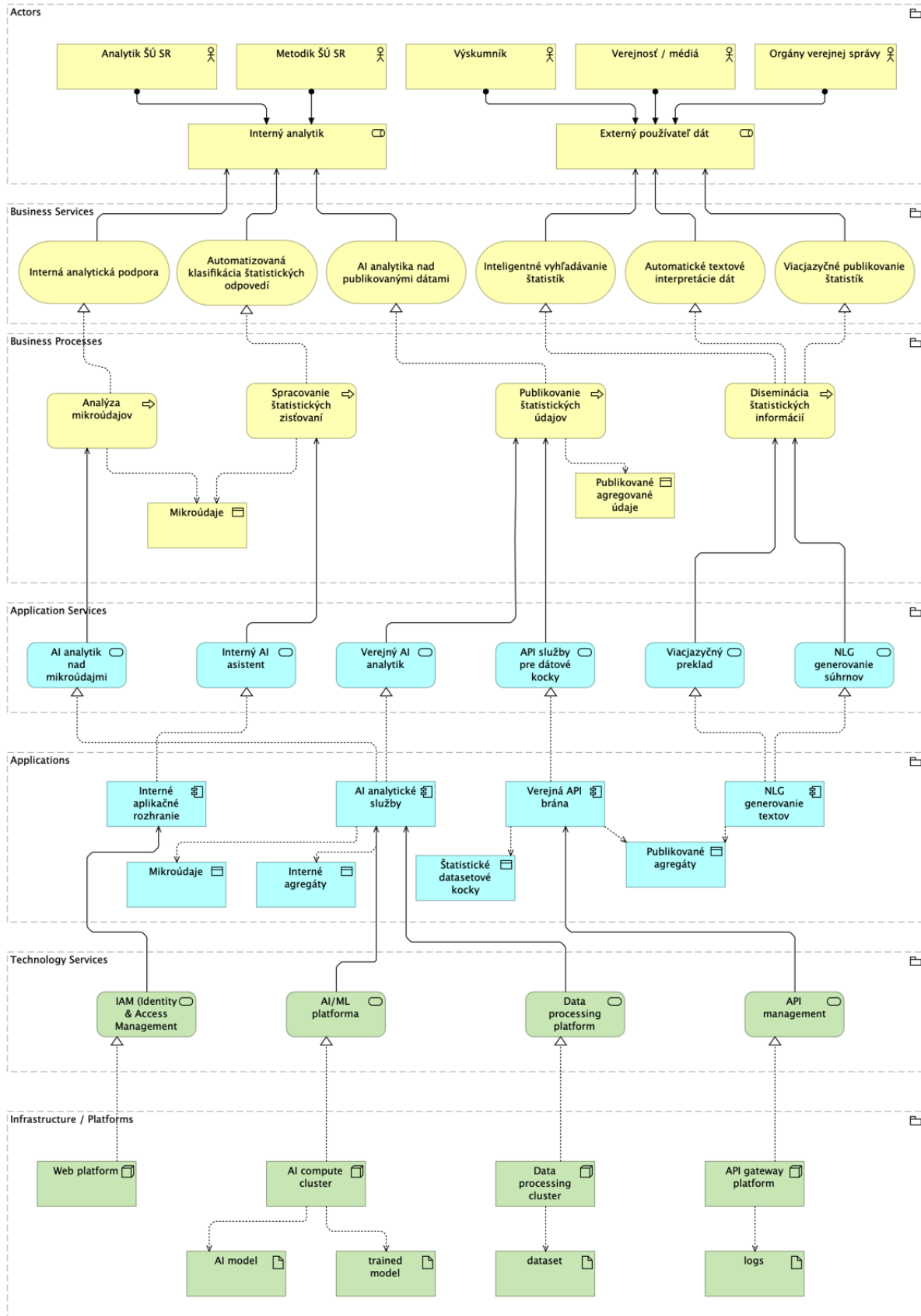
4. Kľúčové komponenty

- Komponenty v Zóne A: Patria sem Mikrodátová databáza (CMP-A1), Katalóg dát a metaúdajov (CMP-A3), Register modelov (CMP-A9) a Služba bezpečného AI analytika (CMP-A7).
- Komponenty v Zóne B: Zahŕňajú Úložisko publikovaných agregátov (CMP-B1), Službu sémantického vyhľadávania (CMP-B5), Službu NLG súhrnov (CMP-B6) a Web portál (CMP-B10).

5. Pravidlá prenosu dát (Prestupy)

- Smer toku: Prestupy medzi zónami sú povolené len v smere zo Zóny A do Zóny B. Spätné volania alebo priamy prístup zo Zóny B do Zóny A (k mikrodátam) sú zakázané.
- Kontrolné mechanizmy: Každý presun dát podlieha procesu schvaľovania „štyrmi očami“ (metodik a vlastník dát), automatizovaným testom rizika reidentifikácie a kontrole cez DLP (prevencia úniku dát) a egress brány.
- Podmienka prestupu: Pred exportom do Zóny B sa musia uplatniť pravidlá štatistického utajenia, ako sú minimálne veľkosti buniek, potláčanie nízkych počtov a zaokrúhľovanie.

Zóna A je teda zameraná na maximálnu bezpečnosť a ochranu súkromia, zatiaľ čo Zóna B je navrhnutá pre vysokú dostupnosť a škálovateľnosť verejných informačných služieb.



Obrázok 7 Náhľad celkovej architektúry

Celková architektúra AI platformy Štatistického úradu SR je navrhnutá ako viacvrstvový model, ktorý prepája používateľské skupiny, biznis služby, procesy, aplikačné služby, aplikačné komponenty, technologické služby a infraštruktúrne zdroje. Takéto usporiadanie nie je samoúčelné. Jeho cieľom je ukázať, že navrhované riešenie vytvára konzistentný reťazec od potrieb stakeholderov až po konkrétne technické a prevádzkové prostriedky, ktorými sa tieto potreby naplňujú. Diagram tak súčasne vysvetľuje logiku budúceho cieľového stavu, preukazuje väzbu medzi prínosmi a architektúrou riešenia a podporuje argumentáciu, prečo sú jednotlivé komponenty a služby v projekte nevyhnutné.

Na úrovni aktérov sú v architektúre identifikované dve základné skupiny používateľov: interní používatelia reprezentovaní analytikmi a metodikmi ŠÚ SR a externí používatelia reprezentovaní výskumníkmi, verejnosťou, médiami a orgánmi verejnej správy. Toto rozlíšenie je dôležité, pretože práve rozdielne potreby týchto skupín formujú celý návrh platformy. Interné prostredie potrebuje najmä nástroje na efektívnejšie spracovanie dát, metodickú podporu a bezpečnú analytiku nad citlivými mikroúdajmi. Externé prostredie naopak potrebuje rýchly, zrozumiteľný a bezpečný prístup k publikovaným údajom, ich interpretácii a lepšej dohľadateľnosti. Zaradenie týchto aktérov do diagramu preto nie je len formálnou identifikáciou používateľov, ale vyjadruje, že architektúra je od začiatku navrhovaná ako odpoveď na rôzne režimy používania, rôzne úrovne oprávnení a rôzne typy hodnôt, ktoré má platforma priniesť.

Na úrovni biznis služieb sa premietajú hlavné oblasti hodnoty, ktoré boli identifikované v analýze prínosov navrhovaných AI riešení. Interná analytická podpora je v architektúre zdôvodnená tým, že interný AI asistent má zvyšovať rýchlosť a konzistentnosť práce zamestnancov ŠÚ SR, zjednocovať metodické postupy a nepriamo prispievať k vyššej kvalite údajov a rýchlejšim výstupom pre verejnosť. Automatizovaná klasifikácia štatistických odpovedí je opodstatnená tým, že znižuje manuálnu záťaž pri kódovaní otvorených textových odpovedí, zrýchľuje spracovanie zisťovaní a umožňuje publikovať detailnejšie a kvalitnejšie štrukturálne štatistiky. AI analytika nad publikovanými dátami, inteligentné vyhľadávanie štatistík, automatické textové interpretácie dát a viacjazyčné publikovanie tvoria samostatné biznis služby preto, lebo každý z týchto prvkov adresuje inú bariéru pri využívaní oficiálnych štatistík v externom prostredí. Verejný AI analytik odstraňuje bariéru zložitosti dátových kociek a potreby odbornej znalosti databázovej štruktúry. Inteligentné vyhľadávanie rieši problém dohľadateľnosti údajov a orientácie v rozsiahlej publikačnej vrstve.

Toto riešenie využíva technológiu Natural Language Generation (NLG) na tvorbu krátkych textových zhrnutí k vybraným štatistickým výstupom. Znamená to, že k tabuľkám alebo grafom, ktoré ŠÚ SR publikuje (napr. mesačný vývoj inflácie, porovnanie regiónov v nezamestnanosti), sa automaticky vygeneruje slovný komentár popisujúci hlavné zistenia, napríklad trend oproti minulému obdobiu, percentuálne zmeny alebo výrazné odlišnosti medzi jednotlivými skupinami. Takýto text je pripravený modelom vytrénovaným na historických komentároch a dátach a následne skontrolovaný človekom pred zverejnením, aby sa zaistila faktická presnosť a vhodná formulácia. NLG komentáre reagujú na potrebu interpretácie čísel a zrozumiteľnosti štatistických výstupov pre neobdobné publikum, keďže umožňujú previesť číselné údaje do stružnej textovej interpretácie vysvetľujúcej hlavné trendy a význam publikovaných štatistík. Automaticky generované súhrny tak prispievajú k tomu, aby boli oficiálne štatistiky prístupnejšie širšiemu publiku vrátane médií, študentov, analytikov alebo bežnej verejnosti, ktorí nemusia mať odborné štatistické vzdelanie. Automatizácia textových komentárov zároveň zabezpečuje konzistentnú komunikáciu kľúčových faktov, pretože ku každej publikovanej tabuľke alebo grafu môže byť priradené základné vysvetlenie, ktoré zvýrazňuje najdôležitejšie zistenia a kontext vývoja. Vďaka tomu sa znižuje riziko nesprávnej interpretácie údajov a zvyšuje sa schopnosť používateľov rýchlo pochopiť význam prezentovaných dát. NLG nástroje zároveň umožňujú rýchlejšie publikovanie a širší záber tém, keďže automatizované generovanie textov umožňuje pripraviť komentáre aj k takým štatistickým výstupom, pri ktorých by manuálne spracovanie nebolo časovo alebo kapacitne možné. Okrem toho môžu generované súhrny slúžiť aj ako základ pre ďalšie formy komunikácie štatistických údajov, napríklad pre infografiky, mediálne výstupy alebo prezentácie na sociálnych sieťach, čím sa zvyšuje dosah štatistických informácií v rôznych komunikačných kanáloch. V kombinácii s viacjazyčným publikovaním zároveň takéto riešenie pomáha odstraňovať jazykovú bariéru a zvyšuje využiteľnosť slovenských štatistík v medzinárodnom prostredí. Zaradenie týchto biznis služieb do architektúry riešenia je preto priamo podložené očakávanými externými benefítmi identifikovanými v analytickej časti dokumentu, najmä v oblasti zlepšenia dostupnosti štatistických údajov, podpory výskumu a zvyšovania informovanosti verejnosti.

Na úrovni biznis procesov diagram ukazuje, že navrhované AI služby nie sú izolované funkcionality, ale sú vložené do reálnych procesov oficiálnej štatistiky. Proces analýzy mikroúdajov je v architektúre nevyhnutný preto,

že interný AI analytik má hodnotu len vtedy, ak podporuje reálnu internú analytickú prácu s citlivými údajmi a vedie k rýchlejšiemu, presnejšiemu a bezpečnejšiemu analytickému výstupom. Proces spracovania štatistických zisťovaní je zase základom pre automatickú klasifikáciu odpovedí, keďže práve v tejto časti životného cyklu údajov vzniká vysoká manuálna záťaž, ktorú má AI riešenie redukovať.

Proces publikovania štatistických údajov je dôležitý preto, že verejný AI analytik, API prístup, NLG súhrny a viacjazyčné publikovanie sú všetko služby, ktoré nadväzujú na moment, keď sú údaje už metodicky pripravené a schválené na zverejnenie. V tejto fáze životného cyklu údajov sa publikované agregované údaje a dátové kocky stávajú základným zdrojom pre služby orientované na externých používateľov. Verejný AI analytik nad dátovými kockami umožňuje, aby používatelia z radov verejnosti, médií, výskumníkov alebo podnikateľského prostredia mohli s publikovanými údajmi pracovať prostredníctvom prirodzeného jazyka bez potreby znalosti štruktúry databáz alebo analytických nástrojov. Používateľ tak môže položiť otázku vlastnými slovami a systém na základe publikovaných agregovaných údajov identifikuje relevantné štatistické indikátory, poskytne číselnú odpoveď a v prípade potreby doplní aj grafické znázornenie, kontext alebo odkaz na zdrojové tabuľky. Takýto prístup výrazne znižuje bariéru vstupu pri práci so štatistickými údajmi a umožňuje využívať oficiálne dáta širšiemu okruhu používateľov vrátane tých, ktorí nemajú špecializované štatistické alebo analytické vzdelanie. Verejný AI analytik zároveň umožňuje poskytovať informácie v reálnom čase a v režime nepretržitej dostupnosti, čo je významné najmä pre mediálne prostredie alebo analytické pracoviská, ktoré potrebujú rýchlo získať relevantné štatistické údaje. Automatizované spracovanie dopytov zároveň odľahčuje pracovníkov štatistického úradu od rutinných otázok, ktoré dnes prichádzajú prostredníctvom e-mailu alebo telefonických požiadaviek, a umožňuje im sústrediť sa na komplexnejšie analytické úlohy. Okrem samotných čísel môže systém poskytovať aj kontext a interpretáciu údajov, napríklad medziročné porovnanie alebo porovnanie s celonárodným priemerom, čím podporuje lepšie pochopenie štatistických trendov a prispieva k rozvoju dátovej gramotnosti spoločnosti. V širšom kontexte tak verejný AI analytik predstavuje nový komunikačný kanál medzi štatistickým úradom a verejnosťou, ktorý zvyšuje transparentnosť, dostupnosť a praktickú využiteľnosť oficiálnych štatistických údajov a zároveň podporuje ich širšie využívanie v médiách, výskume a verejnej diskusii.

Proces diseminácie štatistických informácií je v architektúre samostatne zdôraznený preto, lebo projekt nie je len o výrobe údajov, ale aj o ich lepšej komunikácii, interpretácii a prístupnosti pre rôzne skupiny používateľov. Zaradenie publikovaných agregovaných údajov ako biznis objektu medzi procesmi a službami zároveň ukazuje, že agregované údaje predstavujú kľúčový prechodový bod medzi interným spracovaním a externým využitím a že práve nad nimi sa odohráva väčšina verejných AI služieb. V tejto vrstve architektúry zohráva významnú úlohu aj viacjazyčná diseminácia štatistických výstupov, ktorá umožňuje automatizované prekladanie vybraných publikácií, súhrnov, názvov indikátorov alebo metodických vysvetlení do cudzích jazykov, najmä do angličtiny a potenciálne aj do ďalších jazykov používaných v medzinárodnom prostredí. Takýto prístup umožňuje sprístupniť slovenské štatistiky širšiemu okruhu používateľov vrátane zahraničných výskumníkov, analytikov, investorov, novinárov alebo medzinárodných organizácií, ktoré potrebujú rýchlo pochopiť význam publikovaných údajov bez nutnosti detailnej znalosti slovenského jazyka. Viacjazyčné publikovanie zároveň znižuje jazykovú bariéru pri využívaní údajov zo Slovenska v medzinárodných porovniatiach a výskumných projektoch a prispieva k lepšiemu zapojeniu slovenských štatistik do globálneho vedeckého a analytického prostredia. Automatizované preklady podporené nástrojmi umelej inteligencie umožňujú pripraviť cudzojazyčné verzie výstupov takmer okamžite po publikovaní slovenského originálu, pričom výsledné texty prechádzajú následnou ľudskou revíziou, aby bola zabezpečená terminologická presnosť a stylistická kvalita. Tento prístup zvyšuje efektivitu publikovania a umožňuje rozšíriť rozsah dokumentov, ktoré môžu byť dostupné v cudzom jazyku. Okrem medzinárodného publika má viacjazyčná diseminácia význam aj pre jazykovo zmiešané prostredie alebo cezhraničné regióny, kde môžu byť štatistické informácie využívané médiami, výskumnými inštitúciami alebo verejnými organizáciami v susedných krajinách. Tým sa zároveň posilňuje transparentnosť a reputácia štatistického úradu ako modernej a otvorenej inštitúcie, ktorá sa snaží sprístupňovať svoje údaje čo najširšiemu okruhu používateľov. Prepojenie viacjazyčnej diseminácie s ostatnými službami diseminačnej vrstvy architektúry tak priamo podporuje cieľ projektu zvyšovať dostupnosť štatistických informácií, podporovať výskum a otvorenú vedu a zlepšovať informovanosť verejnosti prostredníctvom kvalitne komunikovaných a medzinárodne využiteľných dát.

Na úrovni aplikačných služieb je architektúra rozdelená tak, aby bolo zrejmé, akými konkrétnymi softvérovými schopnosťami sa budú biznis služby realizovať. AI analytik nad mikroúdajmi je tu ako aplikačná služba preto, že interná biznis potreba bezpečnej analýzy mikroúdajov si vyžaduje osobitnú analytickú službu s kontrolovaným prístupom, podporou dotazovania a mechanizmami utajenia. Interný AI asistent je oddelený ako samostatná aplikačná služba preto, že jeho charakter je iný – nejde o analytiku nad dátami, ale o metodickú a znalostnú podporu nad internou dokumentáciou. Verejný AI analytik je samostatnou službou preto, že pracuje nad publikovanými agregátmi, nie nad mikroúdajmi, a jeho požiadavky na bezpečnosť, výkon a používateľské

rozhranie sú odlišné od interného analytika. API služby pre dátové kocky sú v architektúre nevyhnutné preto, že bez štandardizovaného programového prístupu by verejný analytik, portál ani externí konzumenti nedokázali konzistentne pracovať s publikovanými agregátmi. Viacjazyčný preklad a NLG generovanie súhrnov sú samostatné aplikačné služby preto, lebo predstavujú špecifické vrstvy transformácie publikovaných údajov do používateľsky zrozumiteľného a jazykovo dostupného obsahu. Toto členenie vyjadruje modularitu riešenia, podporuje škálovateľnosť a zodpovedá tomu, že jednotlivé prínosy projektu vyžadujú rôzne typy AI služieb.

Na úrovni aplikačných komponentov architektúra prechádza od logických služieb ku konkrétnym implementačným modulom. Interné aplikačné rozhranie je v diagrame opodstatnené tým, že interní pracovníci potrebujú kontrolované používateľské rozhranie pre metodickú podporu, klasifikáciu, analytiku a prístup k interným dátam. AI analytické služby sú ako aplikačný komponent kľúčové preto, že predstavujú miesto, kde sa sústreďuje analytická logika nad mikroúdajmi a internými agregátmi. Verejná API brána je v architektúre zásadná preto, že zabezpečuje bezpečné, verzované a štandardizované sprístupnenie verejných služieb a dát externým konzumentom. NLG generovanie textov je samostatný komponent, pretože generovanie komentárov a interpretácií si vyžaduje vlastnú logiku, pracovné toky a väzbu na publikačný proces. Publikované agregáty, mikroúdaje, interné agregáty a štatistické datasetové kocky sú v aplikačnej vrstve zdôvodnené ako základné dátové objekty a úložiská, bez ktorých by jednotlivé služby nemali nad čím fungovať. Mikroúdaje sú nevyhnutné pre internú analytiku, klasifikáciu a syntetizáciu. Interné agregáty predstavujú medzivrstvu medzi citlivým detailom a publikovateľným výstupom. Publikované agregáty a kocky sú nevyhnutné preto, že verejné AI služby musia byť striktné postavené na už zverejnených, schválených a neidentifikujúcich údajoch.

Na úrovni technologických služieb sa architektúra opiera o štyri hlavné podporné piliere. Identity and Access Management je v diagrame nevyhnutné preto, že bez riadenia identít, rolí a oprávnení by nebolo možné oddeliť interných používateľov od externých, ani bezpečne kontrolovať prístup k mikroúdajom, interným agregátom a publikačným službám. AI/ML platforma je zdôvodnená tým, že jednotlivé AI služby potrebujú spoločné prostredie pre tréning, inferenciu, správu modelov a ich prevádzkové riadenie. Data processing platform je potrebná preto, že medzi zdrojovými údajmi, agregovanými výstupmi a verejnými službami musí existovať spracovateľská vrstva schopná zabezpečiť prípravu dát, transformácie, integrácie a napojenie na datasetové kocky. API management je v architektúre nevyhnutný z dôvodu riadenia životného cyklu rozhraní, bezpečnosti API volaní, limitov, verzovania a monitorovania spotreby služieb. Tieto technologické služby teda priamo podporujú kvalitu, bezpečnosť, prevádzkovateľnosť a škálovateľnosť celého riešenia.

Na úrovni infraštruktúry a platforiem diagram ukazuje minimálne technické predpoklady, bez ktorých by sa navrhovaná AI platforma nedala prevádzkovať. Web platform je potrebná pre portálové a rozhraniové časti riešenia. AI compute cluster je v architektúre nevyhnutný preto, že tréning a inferencia AI modelov si vyžadujú samostatné výpočtové kapacity a kontrolované prevádzkové prostredie. Data processing cluster je opodstatnený tým, že spracovanie údajov, ich príprava, transformácia a integrácia si vyžadujú robustné dátové spracovanie oddelené od samotných AI modelov. API gateway platform je potrebná na bezpečné publikovanie a riadenie API rozhraní. Artefakty ako AI model, trained model, dataset a logs sú v diagrame prítomné preto, lebo ukazujú, že platforma nepracuje len s aplikáciami, ale aj s reálnymi digitálnymi artefaktmi, ktoré treba spravovať, verzovať, auditovať a opakovane používať. Tým sa posilňuje argument, že projekt musí riešiť nielen „funkcionalitu“, ale aj správu modelového a dátového životného cyklu.

Celý diagram má pritom ešte jednu dôležitú vlastnosť: ukazuje, že hodnota architektúry nevzniká iba v jednotlivých prvkoch, ale najmä v ich vzájomnom prepojení. Interné služby zvyšujú kvalitu a rýchlosť spracovania údajov, čo sa následne pozitívne premieta do externých výstupov. Bezpečná analytika nad mikroúdajmi a automatizovaná klasifikácia odpovedí zlepšujú internú efektivitu a kvalitu dát, čo podporuje dôveryhodnejšie publikované agregáty. Verejný AI analytik, inteligentné vyhľadávanie, NLG komentáre a viacjazyčné publikovanie potom nad týmito výstupmi vytvárajú novú vrstvu používateľskej hodnoty pre občanov, médiá, výskumníkov a partnerov verejnej správy. Generovanie syntetických mikrodatasetov navyše rozširuje hodnotu architektúry smerom k otvorenej vede a inovačnému prostrediu, pričom zároveň rešpektuje požiadavky na dôvernosť a rezidenciu dát.

5.3 Biznis vrstva

Táto kapitola opisuje navrhované riešenie z pohľadu biznis vrstvy – teda z hľadiska činností, služieb a procesov, ktoré projekt mení a podporuje. Zameriava sa na to, ako AI platforma ovplyvní každodennú prácu zamestnancov,

spôsob vzniku a spracovania štatistických údajov a formy, akými budú výsledné informácie poskytované rôznym skupinám používateľov.

5.3.1 Návrh riešenia v biznis vrstve architektúry

V súčasnom stave je výkon kľúčových činností oficiálnej štatistiky vo veľkej miere závislý od manuálnej práce odborných zamestnancov a od fragmentovaných nástrojov, ktoré nie sú navrhnuté na systematické využívanie umelej inteligencie pri striktnom dodržiavaní štatistickej dôvernosti. Pri práci s mikroúdajmi je bezpečnostný režim prirodzene prísny, ale analytická práca, metodická podpora a diseminácia sú oddelené tak, že prevod poznatkov z mikrodát do verejných výstupov vyžaduje viacnásobné manuálne kroky, opakované kontroly a individuálne posudzovanie rizika úniku. Kľúčové interné procesy, ktoré sa v AS-IS stave opakujú a predstavujú významnú časovú záťaž, zahŕňajú najmä:

- manuálne vyhľadávanie metodických postupov v dokumentoch a smerniciach,
- interpretáciu interných pravidiel pri konkrétnych prípadoch,
- ručné kódovanie voľných textových odpovedí do klasifikácií,
- ad-hoc analýzy nad mikroúdajmi realizované tradičnými nástrojmi a následné manuálne uplatňovanie pravidiel štatistického utajenia pri príprave každého výstupu.

V oblasti diseminácie sú agregované údaje poskytované najmä cez existujúce portálové a publikačné mechanizmy, pričom používateľ musí poznať štruktúru dátových kociek, terminológiu a spôsoby filtrovania, aby získal odpoveď. Vyhľadávanie je prevažne založené na kľúčových slovách, viacjazyčnosť je riešená manuálnym prekladom a textové interpretácie výstupov sú závislé od kapacít redakcie. Z pohľadu e-Government komponentov sa projekt dotýka najmä digitálnych služieb súvisiacich s disemináciou štatistických výstupov a interných biznis služieb pre zamestnancov ŠÚ SR. V AS-IS stave sú tieto služby poskytované bez jednotnej AI vrstvy, bez systematicky spravovaného katalógu rozhraní a bez formálneho modelu prestupov medzi citlivými mikroúdajmi a publikovanými agregátmi. Pravidlá pre prácu s dátami sú síce nastavené, ale nie sú „zabudované“ priamo do služieb. To obmedzuje možnosť zavádzať nové typy interakcie (chatboty nad štatistikou, sémantické vyhľadávanie, syntetické dáta) a zároveň zvyšuje interné náklady na rutinné spracovanie a kontrolu. Optimalizačné príležitosti v AS-IS stave spočívajú v odstránení opakovaných manuálnych krokov, skrátení času potrebného na vyhľadanie a aplikáciu interných metodík, automatizácii kódovania voľných textov, zavedení bezpečného self-service analytického prístupu nad mikroúdajmi s automatizovaným štatistickým utajením, zrýchlení diseminácie (vrátane viacjazyčnosti a generovania sprievodných textov) a v sprístupnení publikovaných agregátov verejnosti spôsobom, ktorý nevyžaduje znalosť dátových štruktúr ani špecializovaný softvér.

V cieľovom stave projekt zavádza jednotnú AI platformu ako novú nosnú vrstvu biznis služieb ŠÚ SR, ktorá kombinuje dve ambície:

- systematické využitie AI v oficiálnej štatistike
- dôslednú ochranu citlivých mikroúdajov

Biznis model riešenia je postavený na jasnej separácii spracovania citlivých mikroúdajov v chránenej Zóne A a prevádzky verejných služieb v Zóne B, ktorá pracuje výhradne s publikovanými agregovanými údajmi, dátovými kockami, metaúdajmi, schválenými syntetickými datasetmi a verejným obsahom. Všetky činnosti nad mikroúdajmi – vrátane tréningu a inferencie modelov, generovania syntetických dát či bezpečnej AI analytiky – prebiehajú v Zóne A pod prísny dohľadom, kým Zóna B poskytuje iba read-only služby bez akéhokoľvek prístupu k pôvodným záznamom respondentov. Prestupy medzi zónami sú jednosmerné (A→B), obmedzené na publikované agregované údaje, dátové kocky, metaúdaje a schválené syntetické datasety a realizované cez egress workflow s automatizovanými testami rizika, DLP kontrolami, princípom „four-eyes“ a úplnou auditnou stopou. Vnútorne služby pre zamestnancov sú navrhnuté tak, aby znížovali interné náklady a riziká pri spracovaní citlivých dát, kým verejné služby rozširujú spôsob, akým môže verejnosť, médiá a výskumníci pristupovať k oficiálnym štatistikám – moderným, konverzačným a medzinárodne porovnateľným spôsobom.

Portfólio TO-BE biznis služieb vychádza z MoSCoW prioritizácie:

Must služby – jadro bezpečného a efektívneho spracovania dát:

- Interný AI asistent pre metodickú podporu (UC-01) – poskytuje konzistentné odpovede z interných metodík, znižuje závislosť od jednotlivcov a skracuje čas pri rozhodovaní.

- Automatická klasifikácia textových odpovedí (UC-02) – automatizuje kódovanie veľkého objemu otvorených odpovedí do klasifikácií ako SK NACE a ISCO s konzistentnou kvalitou.
- Bezpečný AI analytik nad mikroúdajmi (UC-04) – umožňuje interným analytikom pýtať sa „akékoľvek“ otázky nad mikrodátami, ale vracia len agregované výstupy, ktoré prešli automatizovaným štatistickým utajením. Spája tak LLM asistenta so zabudovanou kontrolou štatistického utajenia (SDC).

Should služby – rozšírenie smerom k otvorenej vede a verejnosti:

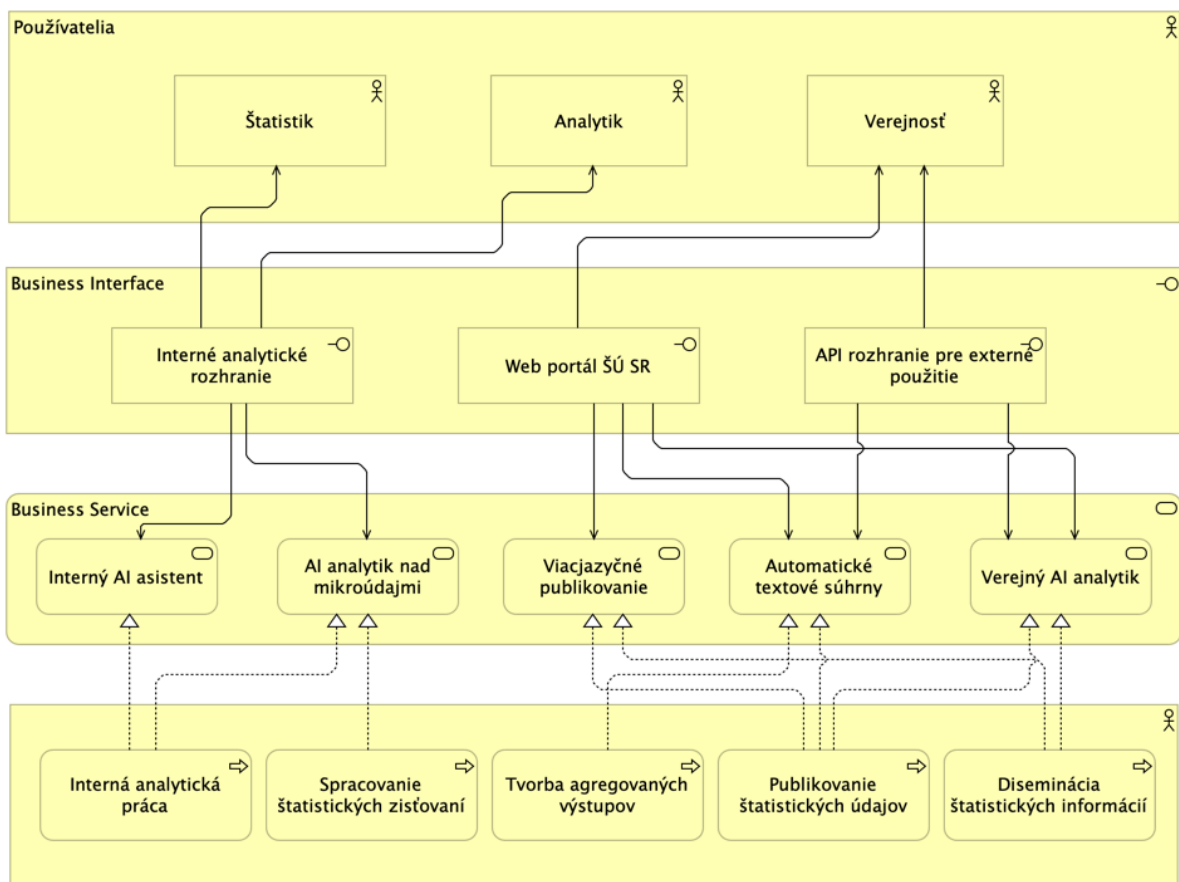
- Generovanie syntetických mikrodasetov (UC-03) – buduje novú triedu produktov, ktoré štatisticky napodobňujú skutočné mikrodáta, ale sú bezpečne zdieľateľné s výskumníkmi, akademikmi a partnermi. Podporuje tak otvorenú vedu bez porušenia dôvernosti.
- Verejný AI analytik nad dátovými kockami (UC-05) – prináša „štatistický chatbot“, ktorý odpovedá na otázky v prirodzenom jazyku nad publikovanými agregátmi, podobne ako piloty StatGPT/StatBot v iných krajinách, ale špecificky prispôbený pre jazyk a štruktúru štatistiky SR.
- Inteligentné vyhľadávanie a odporúčania (UC-06) – zavádza sémantické vyhľadávanie v dátach a publikáciách v súlade s medzinárodnou praxou.
- Viacjazyčná diseminácia (UC-08) – využíva AI preklady s ľudskou revíziou, aby boli kľúčové výstupy rýchlo dostupné v angličtine (a ďalších jazykoch), čím umožňuje, aby sa štatistiky SR stali prirodzenou súčasťou medzinárodných analýz.

Could služba – doplnenie diseminácie podľa kapacít:

- Automatické textové súhrny (UC-07) – generujú konzistentné, metodicky správne komentáre k štatistikám, ktoré zvyšujú zrozumiteľnosť údajov a zároveň tvoria nový korpus textov využiteľný aj pre ďalšie AI aplikácie.

Z procesného pohľadu TO-BE stav mení spôsob, akým zamestnanci pracujú s metodikami, kódujú otvorené odpovede, analyzujú mikroúdaje a pripravujú výstupy. Interný AI asistent nahrádza zdĺhavé prehľadávanie dokumentov interaktívnou konzultáciou. Automatická klasifikácia odsúva manuálne kódovanie do výnimočných prípadov. Bezpečný AI analytik skracuje čas na prípravu analytických výstupov z dní na hodiny pri zachovaní rovnakých alebo vyšších štandardov dôvernosti. Verejnosť v TO-BE stave získava nové digitálne kanály: otázky v prirodzenom jazyku nad oficiálnymi štatistikami, sémantické vyhľadávanie naprieč portálom, automatické komentáre vysvetľujúce čísla a rýchle viacjazyčné vydania kľúčových výstupov. Tým ŠÚ SR rozširuje využitie AI pri sprístupňovaní údajov.

Hlavný rozdiel medzi AS-IS a TO-BE stavom spočíva v prechode od izolovaných, manuálnych a prevažne nástrojovo nesúrodých činností k jednotnej, bezpečne riadenej AI vrstve, ktorá je pevnou súčasťou biznis architektúry úradu. V internom prostredí sa znižuje závislosť od individuálnych expertov a „tichého know-how“, skracuje sa čas potrebný na aplikáciu metodiky, automatizujú sa rutinné činnosti vo veľkom objeme (najmä kódovanie textových odpovedí) a do analytických nástrojov sa priamo zabudúva štatistické utajenie, namiesto toho, aby bolo aplikované až manuálne na konci procesu. Vo verejnej časti sa mení charakter diseminácie z pasívneho zverejňovania tabuliek na interaktívny, vysvetľujúci a viacjazyčný servis: používateľ už nepotrebuje poznať štruktúru dát, môže sa pýtať v prirodzenom jazyku a systém mu v bezpečnom režime poskytne číselnú odpoveď, vizualizáciu aj textový kontext.



Obrázok 8 Znáznornenie biznis architektúry TO BE

Obrázok znázorňuje logickú architektúru biznis vrstvy AI platformy Štatistického úradu Slovenskej republiky. Diagram ilustruje väzby medzi používateľmi systému, prístupovými rozhraniami, poskytovanými službami umelej inteligencie a hlavnými procesmi spracovania štatistických údajov. Cieľom modelu je ukázať, akým spôsobom nové digitálne služby podporované umelou inteligenciou zlepšujú interné analytické procesy a zároveň zvyšujú dostupnosť štatistických informácií pre externých používateľov.

Najvyššiu úroveň modelu predstavujú používatelia systému. Medzi hlavné skupiny používateľov patria štatistici, analytici a verejnosť. Štatistici a analytici predstavujú interných používateľov platformy, ktorí využívajú analytické nástroje pri spracovaní štatistických zisťovaní a tvorbe analytických výstupov. Verejnosť zahŕňa externých používateľov, ako sú občania, médiá, výskumné organizácie alebo podnikateľské subjekty, ktorí prístupujú k publikovaným štatistickým údajom a analytickým službám. Používatelia prístupujú k jednotlivým službám prostredníctvom troch hlavných rozhraní. Interné analytické rozhranie slúži zamestnancom Štatistického úradu SR na analytickú prácu s údajmi, využívanie interného AI asistenta a realizáciu pokročilej analytiky nad štatistickými datasetmi. Web portál ŠÚ SR predstavuje hlavný komunikačný kanál pre verejnosť, prostredníctvom ktorého sú sprístupňované štatistické informácie, analytické nástroje a komentáre k štatistickým výstupom. API rozhranie pre externé použitie umožňuje programový prístup k publikovaným štatistickým údajom a podporuje integráciu dát do externých analytických alebo výskumných nástrojov.

Na úrovni biznis služieb sú definované hlavné služby umelej inteligencie, ktoré platforma poskytuje. Interný AI asistent je určený pre interných pracovníkov úradu a poskytuje metodickú podporu pri spracovaní údajov, interpretácii štatistických výsledkov alebo pri práci s analytickými nástrojmi. AI analytik nad mikroúdajmi podporuje analytickú prácu s mikroúdajmi v kontrolovanom internom prostredí a umožňuje identifikovať vzory a vzťahy v štatistických dátach. Pre externých používateľov sú určené služby verejného AI analytika, automatického generovania textových súhrnov a viacjazyčného publikovania štatistických výstupov. Verejný AI analytik umožňuje

používateľom pracovať s publikovanými agregovanými údajmi prostredníctvom prirodzeného jazyka a získavať analytické odpovede na otázky týkajúce sa štatistických údajov. Automatické generovanie textových súhrnov zabezpečuje vytváranie komentárov a interpretácií štatistických údajov, ktoré pomáhajú používateľom lepšie pochopiť prezentované údaje. Viacjazyčné publikovanie umožňuje sprístupniť štatistické výstupy v rôznych jazykových verziách, čím sa zvyšuje dostupnosť údajov pre širšie publikum.

Tieto služby podporujú hlavné procesy spracovania štatistických údajov. Medzi základné procesy patria interná analytická práca, spracovanie štatistických zisťovaní, tvorba agregovaných výstupov, publikovanie štatistických údajov a diseminácia štatistických informácií. Interné analytické služby podporujú najmä procesy spracovania štatistických zisťovaní a analytickej práce s mikroúdajmi. Služby určené pre verejnosť podporujú proces publikovania štatistických údajov a diseminácie štatistických informácií. Model zároveň znázorňuje, že jednotlivé služby umeľej inteligencie nie sú izolované komponenty, ale sú priamo prepojené s procesmi spracovania údajov a s rozhraniami, prostredníctvom ktorých k nim pristupujú používatelia. Tým sa vytvára jednotné analytické prostredie, ktoré umožňuje efektívnejšie využívanie štatistických údajov v interných procesoch úradu a zároveň zlepšuje dostupnosť a interpretáciu štatistických informácií pre verejnosť.

5.3.2 Prehľad koncových služieb – budúci stav (TO BE):

Kód KS (z MetalS)	Názov KS	Používateľ KS (G2C/ G2B/G2G/G2A)	Životná situácia (+ kód z MetalS)	Úroveň elektronizácie KS
ks_382577	Sprístupnenie AI asistenta pre interných zamestnancov ŠÚ SR	G2A		4
ks_382578	Automatizovaná klasifikácia a kódovanie textových odpovedí respondentov	G2A		4
ks_382579	Poskytovanie bezpečného analytického prístupu k mikroúdajom s automatickým štatistickým utajením	G2A		4
ks_382580	Poskytovanie syntetických datasetov pre výskum, testovanie a analytické účely	G2B / G2A		4
ks_382581	Sprístupnenie verejného AI analytika nad publikovanými agregovanými štatistickými údajmi	G2C / G2B / G2G		4
ks_382582	Inteligentné vyhľadávanie a odporúčanie obsahu štatistického portálu	G2C / G2B / G2G		4
ks_382583	Automatické generovanie textových komentárov k publikovaným štatistickým údajom	G2C / G2B		4
ks_382584	Sprístupnenie viacjazyčných verzií publikovaných štatistických výstupov	G2C / G2B		4

Tabuľka 13 Prehľad koncových služieb

5.3.3 Organizačné zmeny a Procesy dotknuté navrhovaným riešením

Navrhované riešenie AI platformy Štatistického úradu Slovenskej republiky prináša zmeny v spôsobe spracovania štatistických údajov, najmä v oblasti analytických procesov, práce s mikroúdajmi a poskytovania štatistických informácií verejnosti. Základný princíp zmeny spočíva v postupnom prechode od manuálnych analytických postupov k automatizovaným analytickým procesom podporovaným nástrojmi umelej inteligencie. Tento prístup umožní zvýšiť efektivitu spracovania údajov, zlepšiť kvalitu analytických výstupov a zároveň zabezpečiť bezpečné sprístupňovanie štatistických informácií rôznym skupinám používateľov. Navrhované koncové služby uvedené v kapitole 5.3.2 sú priamo naviazané na jednotlivé fázy spracovania štatistických údajov. Procesný model znázornený na obrázku predstavuje logický tok spracovania dát od zberu údajov až po publikovanie a interpretáciu štatistických výstupov. Zároveň ilustruje oddelenie interného spracovania citlivých mikroúdajov v chránenej zóne a poskytovania verejných analytických služieb nad publikovanými agregovanými údajmi.

Prvá skupina koncových služieb je zameraná na podporu interných analytických procesov v rámci Štatistického úradu SR. Ide najmä o služby sprístupnenia AI asistenta pre interných zamestnancov, automatizovanej klasifikácie textových odpovedí respondentov a poskytovania bezpečného analytického prístupu k mikroúdajom s automatickým štatistickým utajením. Tieto služby podporujú procesy spracovania a klasifikácie údajov, analytickej práce s mikroúdajmi a tvorby interných agregovaných výstupov. V praxi to znamená, že analytické a metodické útvary budú môcť využívať nástroje umelej inteligencie na automatizáciu činností, ktoré sú dnes vykonávané manuálne, napríklad pri klasifikácii otvorených odpovedí zo štatistických zisťovaní alebo pri analytickej interpretácii výsledkov. Ďalšia koncová služba sa týka poskytovania syntetických datasetov pre výskum, testovanie a analytické účely. Táto služba podporuje procesy bezpečného sprístupňovania dát pre výskumné a analytické účely bez rizika porušenia štatistickej dôvernosti. V rámci procesu spracovania údajov bude možné generovať syntetické mikrodáta, ktoré zachovávajú štatistické vlastnosti pôvodných dát, ale neobsahujú identifikovateľné informácie. Tretia skupina koncových služieb je orientovaná na externých používateľov a zlepšenie prístupu k oficiálnym štatistikám. Patria sem služby sprístupnenia verejného AI analytika nad publikovanými agregovanými údajmi, inteligentného vyhľadávania štatistických informácií, automatického generovania textových komentárov k štatistickým údajom a viacjazyčného publikovania štatistických výstupov. Tieto služby nadväzujú na proces publikovania agregovaných výstupov a diseminácie štatistických informácií. Ich cieľom je umožniť používateľom jednoduchšie pracovať s oficiálnymi štatistikami a interpretovať publikované údaje prostredníctvom moderných analytických nástrojov.

Prehľad procesov dotknutých projektom

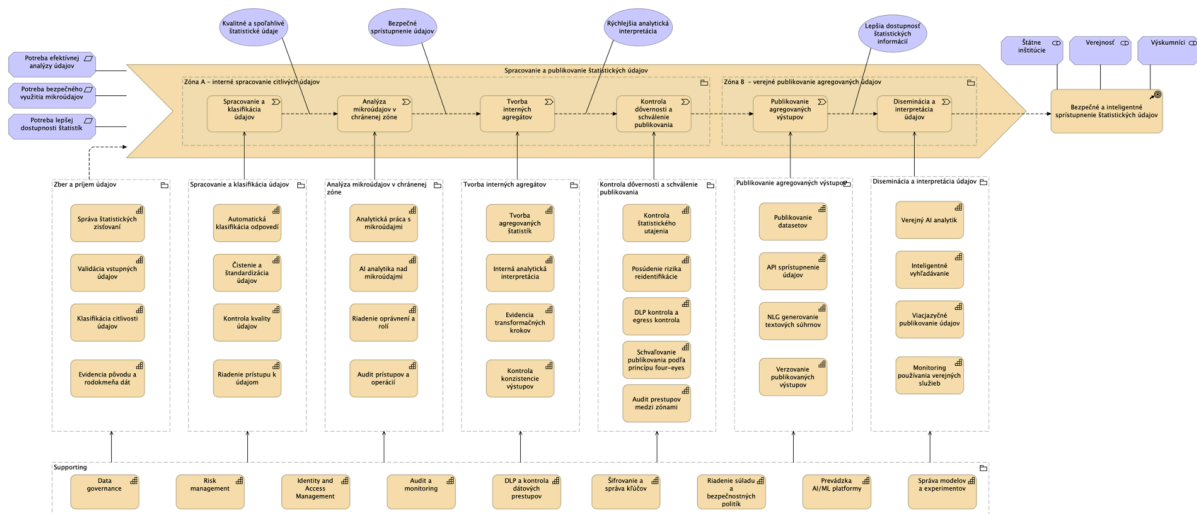
Nasledujúca tabuľka porovnáva súčasný a budúci stav procesov dotknutých projektom a uvádza súvisiace koncové služby.

Proces	Súčasný stav	Budúci stav (TO-BE)	Súvisiace koncové služby
Spracovanie a klasifikácia údajov	Manuálne kódovanie a kontrola kvality údajov	Automatická klasifikácia odpovedí pomocou AI	Automatizovaná klasifikácia textových odpovedí
Analytická práca s mikroúdajmi	Tradičné analytické nástroje	AI analytika nad mikroúdajmi v chránenej zóne	Bezpečný analytický prístup k mikroúdajom
Interná metodická podpora	Manuálne vyhľadávanie metodických informácií	AI asistent pre interných zamestnancov	Sprístupnenie AI asistenta
Bezpečné zdieľanie dát	Obmedzené možnosti poskytovania dát	Generovanie syntetických mikrodasetov	Poskytovanie syntetických datasetov
Publikovanie štatistických údajov	Publikovanie tabuliek a datasetov	Automatizované publikovanie a generovanie komentárov	NLG komentáre k štatistikám
Vyhľadávanie štatistických údajov	Tradičné vyhľadávanie na portáli	Inteligentné vyhľadávanie a odporúčanie obsahu	Inteligentné vyhľadávanie
Interpretácia údajov	Manuálna interpretácia analytikmi	Verejný AI analytik nad publikovanými dátami	Verejný AI analytik
Medzinárodná dostupnosť údajov	Publikovanie prevažne v jednom jazyku	Automatické viacjazyčné publikovanie	Viacjazyčné publikovanie

Tabuľka 14 Prehľad procesov dotknutých projektom

Navrhované riešenie zároveň zachováva prísne pravidlá ochrany štatistickej dôvernosti. Proces spracovania údajov je preto rozdelený do dvoch logických zón. Interná zóna slúži na spracovanie citlivých mikroúdajov a realizáciu analytických procesov v chránenom prostredí. Verejná zóna je určená na publikovanie agregovaných

výstupov a poskytovanie analytických služieb pre verejnosť. Prenos údajov medzi týmito zónami je striktné kontrolovaný a je povolený iba pre publikované agregované údaje, dátové kocky, metaúdaje a schválené syntetické datasety.



Obrázok 9 Znáznornenie procesov

Procesný model znázornený na obrázku ilustruje celý životný cyklus spracovania štatistických údajov od ich zberu až po publikovanie a interpretáciu výstupov. Model zároveň zdôrazňuje rozdelenie procesov do dvoch logických zón. Interná zóna je určená na spracovanie citlivých mikroúdajov a realizáciu analytických procesov v chránenom prostredí. Verejná zóna je určená na publikovanie agregovaných štatistických výstupov a poskytovanie analytických služieb pre externých používateľov. Proces začína zberom a prijímaním údajov zo štatistických zisťovaní. V tejto fáze sú údaje prijímané, validované a klasifikované podľa úrovne citlivosti. Súčasťou procesu je aj evidencie pôvodu dát a zabezpečenie základných princípov dátovej správy. Po prijatí údajov nasleduje fáza spracovania a klasifikácie dát, ktorá zahŕňa čistenie údajov, ich štandardizáciu a kontrolu kvality. V rámci navrhovaného riešenia bude táto fáza doplnená o automatickú klasifikáciu textových odpovedí prostredníctvom nástrojov umelej inteligencie, čím sa výrazne zníži rozsah manuálneho spracovania.

Nasleduje analytická fáza spracovania mikroúdajov v chránenej internej zóne. V tejto fáze analytici vykonávajú analýzu dátových súborov a pripravujú podklady pre tvorbu agregovaných štatistických výstupov. Navrhovaná AI platforma umožní podporiť tieto činnosti prostredníctvom nástrojov umelej inteligencie, ktoré umožnia pokročilú analytiku nad mikroúdajmi, identifikáciu vzorov v dátach a podporu analytického rozhodovania. V tejto fáze sú zároveň striktné riadené prístupové práva používateľov a auditované všetky operácie vykonávané nad citlivými údajmi.

Po analytickej fáze nasleduje tvorba interných agregovaných štatistických výstupov. Tento proces zahŕňa tvorbu agregovaných štatistík, internú analytickú interpretáciu výsledkov a evidenciu transformačných krokov, ktoré boli vykonané počas spracovania údajov. Súčasťou tejto fázy je aj kontrola konzistencie výsledných štatistických výstupov.

Pred publikovaním štatistických údajov prebieha kontrola štatistickej dôveryhodnosti a schvaľovanie výstupov. Táto fáza zahŕňa kontrolu rizika reidentifikácie, aplikáciu pravidiel štatistického utajenia a realizáciu bezpečnostných kontrol vrátane mechanizmov prevencie úniku dát. Výstupy sú publikované až po schválení oprávnenými pracovníkmi podľa princípu human-in-the-loop.

Po schválení výstupov nasleduje publikovanie agregovaných údajov vo verejnej zóne platformy. Publikovanie zahŕňa poskytovanie dátových súborov, sprístupnenie údajov prostredníctvom aplikačných rozhraní a generovanie textových súhrnov štatistických výstupov. V tejto fáze sa využívajú aj nástroje umelej inteligencie na automatické generovanie komentárov k štatistickým údajom a na podporu interpretácie dát.

Poslednou fázou procesu je diseminácia a interpretácia štatistických údajov pre externých používateľov. Platforma umožní poskytovať nové digitálne služby, ako je verejný AI analytik nad publikovanými agregovanými údajmi, inteligentné vyhľadávanie štatistických informácií alebo viacjazyčné publikovanie štatistických výstupov. Tieto služby umožnia širšej verejnosti jednoduchšie pracovať s oficiálnymi štatistikami a získavať analytické informácie bez potreby detailnej znalosti dátových štruktúr.

Navrhované riešenie si vyžiada aj určité organizačné zmeny v rámci Štatistického úradu SR. Zavedenie AI platformy bude vyžadovať posilnenie kompetencií v oblasti dátovej správy, správy analytických modelov a riadenia bezpečnosti dát. Predpokladá sa zapojenie existujúcich analytických a metodických útvarov do prevádzky nových analytických nástrojov a vytvorenie kompetencií pre správu AI modelov a experimentov. Zároveň bude potrebné zabezpečiť úzku spoluprácu medzi analytickými útvarmi, IT útvarom a útvarmi zodpovednými za ochranu údajov a štatistickú dôvernosť, ktoré sa zapoja do kontroly publikovania výstupov a monitorovania prenosov údajov medzi zónami. Z pohľadu procesov projekt zasahuje najmä do procesov spracovania štatistických zisťovaní, analytickej práce s mikroúdajmi, tvorby štatistických výstupov a diseminácie štatistických informácií. Navrhované zmeny procesov sú v súlade s cieľmi digitalizácie verejnej správy a prispievajú k modernizácii analytických procesov oficiálnej štatistiky.

5.3.4 Jazyková podpora a lokalizácia

Navrhovaná AI platforma Štatistického úradu SR je koncipovaná ako riešenie, ktoré zvyšuje dostupnosť a zrozumiteľnosť oficiálnej štatistiky pre rôzne skupiny používateľov. Z tohto dôvodu sa v TO-BE stave uvažuje s jazykovou podporou a lokalizáciou tak používateľského rozhrania, ako aj textových výstupov generovaných platformou. Jazyková podpora sa bude uplatňovať najmä v časti verejných služieb (Zóna B), ktoré pracujú výhradne s publikovanými agregovanými údajmi, metaúdajmi a verejným obsahom, a tým sú vhodné na poskytovanie služieb aj mimo územia SR. V internej časti platformy (Zóna A) sa jazyková podpora realizuje primárne pre potreby interných používateľov úradu. Východiskovým jazykom riešenia je slovenský jazyk, ktorý bude plne podporovaný vo všetkých používateľských rozhraniach, systémových hláseniach, help textoch a v štandardných výstupoch. Ako druhý povinný jazyk sa uvažuje anglický jazyk, najmä z dôvodu medzinárodnej porovnateľnosti a dostupnosti štatistických informácií pre používateľov z členských štátov EÚ, zahraničných výskumníkov, medzinárodnej organizácie a médiá. Rozsah jazykovej podpory je navrhnutý tak, aby bol rozšíriteľný aj o ďalšie jazyky bez zásahu do architektúry – prostredníctvom konfigurovateľných jazykových balíkov a centralizovanej správy prekladových reťazcov.

Lokalizácia používateľského rozhrania zahŕňa preklad navigácie, ovládacích prvkov, formulárov, systémových oznamov, validácií a používateľských inštrukcií. Súčasťou lokalizačnej podpory je aj správne zobrazovanie dátumov, času, číselných formátov, mien a jednotiek podľa zvoleného jazyka a lokality. Pre používateľov je dostupný prepínač jazyka na úrovni používateľského rozhrania, pričom systém si vie preferenciu jazyka pamätať v rámci používateľského profilu alebo session. V prípade verejných služieb sa jazyková preferencia zohľadní aj na úrovni API (napr. hlavička Accept-Language), aby bolo možné rovnaké výstupy konzistentne poskytovať aj cez integračné rozhrania. Osobitnú pozornosť si vyžadujú textové výstupy generované platformou, najmä automatické interpretácie štatistických výsledkov, súhrny, odpovede AI analytika a vysvetľujúce texty pre používateľa. Pri týchto výstupoch bude jazyková podpora riešená tak, aby sa zachovala obsahová presnosť a zároveň sa minimalizovalo riziko skreslenia interpretácie štatistických údajov. Základným princípom je, že viacjazyčné výstupy sú generované alebo prekladané výhradne nad publikovanými agregovanými údajmi a verejným obsahom. Súčasne sa uplatňuje princíp HIL (Human-in-the-loop) pre verejne publikované výstupy – teda povinná ľudská kontrola pred zverejnením v prípadoch, keď sa výstup používa ako oficiálny textový sprievodný materiál alebo publicita. Platforma bude evidovať jazykový režim výstupu, verziu prekladového modelu alebo prekladových pravidiel a auditnú stopu, aby bolo možné späťne preukázať spôsob vzniku konkrétneho viacjazyčného výstupu.

Služby platformy budú poskytované aj používateľom z členských štátov EÚ, a to najmä prostredníctvom verejných analytických služieb nad publikovanými agregovanými údajmi, sémantického vyhľadávania a viacjazyčných interpretácií. Z toho vyplývajú hlavné požiadavky na riešenie: zabezpečenie minimálne dvojazyčného režimu (SK/EN), konzistentná lokalizácia používateľského rozhrania a verejných výstupov, kompatibilita s európskymi štandardmi pre publikovanie a popis dát (najmä v oblasti metaúdajov a otvorených dát) a zabezpečenie dostupnosti služieb pre používateľov mimo SR pri zachovaní bezpečnostných princípov architektúry. Pri poskytovaní služieb pre zahraničných používateľov sa zároveň rešpektuje, že platforma v externom režime pracuje výlučne s publikovanými agregátmi, pričom citlivé mikroúdaje ostávajú spracovávané len v chránenej internej zóne v infraštruktúre ŠÚ SR.

Pre účely Katalógu požiadaviek sa požiadavky na jazykovú podporu a lokalizáciu evidujú ako samostatná skupina funkčných a nefunkčných požiadaviek. Patria sem minimálne požiadavky na podporu jazykov SK a EN pre verejné používateľské rozhrania a výstupy, mechanizmus prepínania jazyka a perzistenciu preferencií, lokalizáciu formátov dátumu/času/čísel, podporu viacjazyčných výstupov generovaných AI nad publikovanými agregátmi, API podporu pre jazykové preferencie, a požiadavky na auditovateľnosť a kontrolu kvality viacjazyčných výstupov vrátane uplatnenia HIL pre vybrané typy publikovaných textov.

5.4 Aplikačná vrstva

5.4.1 Návrh riešenia v aplikačnej vrstve architektúry

Súčasná aplikačná vrstva informačných systémov ŠÚ SR je tvorená súborom špecializovaných informačných systémov a analytických nástrojov, ktoré podporujú jednotlivé fázy spracovania oficiálnych štatistických údajov. Ide najmä o systémy zabezpečujúce zber údajov zo štatistických zisťovaní, ich validáciu, spracovanie, analytickú interpretáciu a následnú disemináciu výsledkov prostredníctvom publikačných a dátových služieb. Tieto informačné systémy sú navrhnuté primárne na podporu tradičných analytických a publikačných procesov oficiálnej štatistiky. Aplikačné služby poskytované jednotlivými systémami sú orientované najmä na spracovanie dátových vstupov zo štatistických zisťovaní, správu dátových zdrojov, tvorbu agregovaných štatistických výstupov a publikovanie štatistických informácií pre odbornú verejnosť, médiá a výskumné organizácie. Súčasná aplikačná architektúra tak podporuje základné koncové služby úradu, ktoré zahŕňajú najmä zber štatistických údajov, spracovanie štatistických zisťovaní, tvorbu štatistických výstupov a publikovanie oficiálnych štatistických informácií.

V rámci existujúcej architektúry sú jednotlivé informačné systémy navzájom prepojené prostredníctvom integračných mechanizmov a rozhraní umožňujúcich výmenu údajov medzi systémami zberu údajov, analytickými nástrojmi a publikačnými platformami. Integrácie sú realizované najmä prostredníctvom databázových prístupov, interných aplikačných rozhraní a integračných služieb, ktoré umožňujú prenášať agregované údaje medzi jednotlivými systémami. Aplikačná vrstva zároveň poskytuje rozhrania na komunikáciu s externým prostredím, napríklad s verejnými portálmi, analytickými nástrojmi alebo systémami partnerských organizácií verejnej správy. Tieto rozhrania umožňujú poskytovanie štatistických údajov a výstupov vo forme publikovaných agregovaných dát, štatistických tabuliek alebo analytických prehľadov.

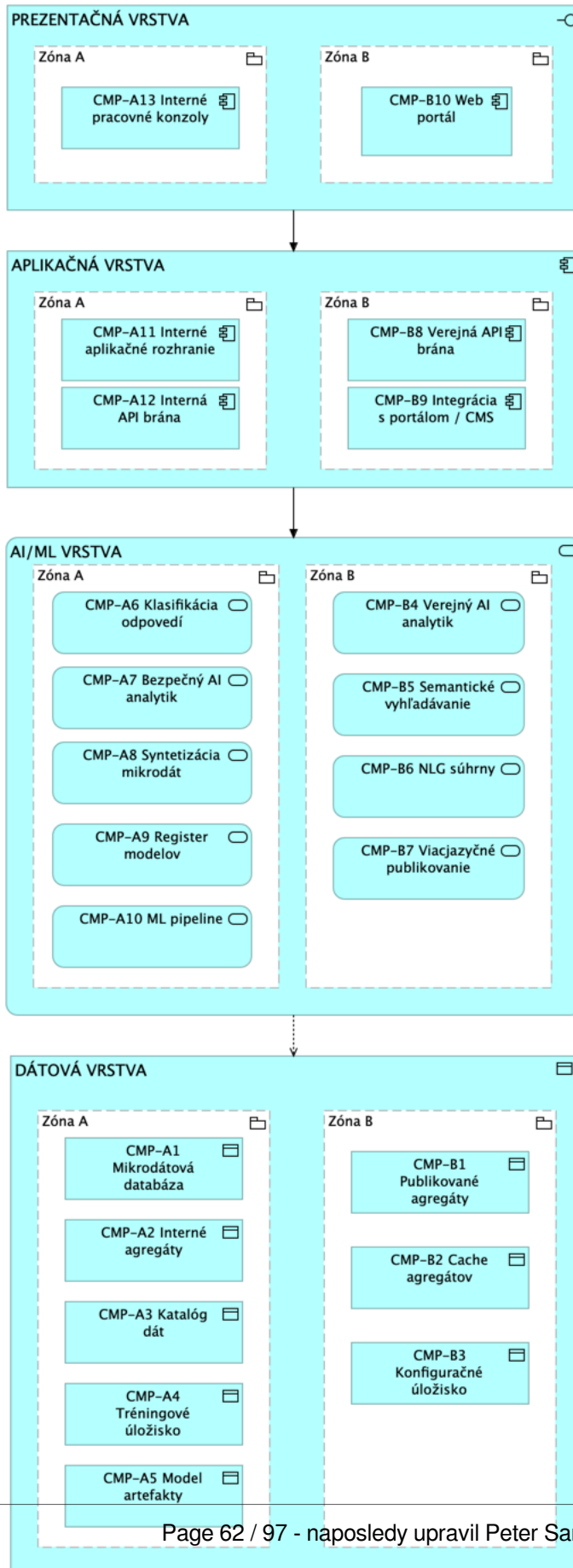
Súčasná architektúra však neobsahuje komplexnú aplikačnú platformu určenú na systematické využívanie umelej inteligencie pri spracovaní štatistických údajov. Chýbajú najmä aplikačné služby umožňujúce automatizovanú klasifikáciu textových odpovedí zo štatistických zisťovaní, inteligentné vyhľadávanie štatistických informácií, generovanie textových interpretácií dát alebo bezpečnú analytickú prácu nad citlivými mikroúdajmi s využitím AI nástrojov. Súčasná aplikačná architektúra tak poskytuje stabilný základ pre spracovanie a publikovanie štatistických údajov, avšak neumožňuje plnohodnotné využívanie moderných analytických nástrojov umelej inteligencie ani bezpečné experimentovanie s novými analytickými metódami. Z tohto dôvodu vzniká potreba rozšírenia aplikačnej vrstvy o nové komponenty a aplikačné služby, ktoré budú podporovať moderné analytické a diseminačné procesy.

Navrhovaná aplikačná architektúra rozširuje existujúce informačné systémy o novú platformu umelej inteligencie, ktorá bude poskytovať súbor špecializovaných aplikačných služieb určených na podporu moderných analytických procesov a zlepšenie dostupnosti štatistických údajov pre verejnosť. Aplikačná vrstva bude pozostávať z viacerých logických komponentov, ktoré budú poskytovať aplikačné služby podporujúce jednotlivé biznis procesy a koncové služby Štatistického úradu SR. Tieto komponenty budú implementované ako modulárne aplikačné funkcie platformy umelej inteligencie, ktoré budú integrované s existujúcimi informačnými systémami úradu. Navrhovaná platforma bude podporovať najmä interné analytické procesy spojené so spracovaním štatistických údajov. Jednou z hlavných aplikačných služieb bude interný AI asistent určený pre zamestnancov úradu, ktorý bude poskytovať metodickú podporu pri spracovaní štatistických údajov a práci s analytickými nástrojmi. Ďalšou významnou aplikačnou službou bude systém automatickej klasifikácie textových odpovedí zo štatistických zisťovaní, ktorý umožní výrazne znížiť rozsah manuálnej práce pri spracovaní otvorených odpovedí respondentov. Platforma bude zároveň obsahovať aplikačné služby určené na bezpečnú analytickú prácu s mikroúdajmi v kontrolovanom prostredí. Tieto služby umožnia analytikom využívať nástroje umelej inteligencie pri práci s citlivými dátami, pričom budú rešpektované všetky pravidlá štatistickej dôverylosti a ochrany osobných údajov.

Významnou súčasťou navrhovanej aplikačnej architektúry bude aj generovanie syntetických mikrodát. Táto aplikačná služba umožní vytvárať dátové súbory, ktoré zachovávajú analytické vlastnosti pôvodných mikrodát, ale neobsahujú identifikovateľné informácie. Syntetické mikrodáta bude možné využívať na experimentálne analytické práce, výskumné projekty alebo testovanie analytických modelov bez rizika porušenia štatistickej dôvernosti. Okrem interných analytických služieb bude aplikačná architektúra poskytovať aj služby orientované na verejnosť. Patria sem najmä služby inteligentného vyhľadávania štatistických informácií, verejný AI analytik nad publikovanými agregovanými údajmi a nástroje na automatické generovanie textových súhrnov štatistických výstupov. Tieto služby umožnia používateľom jednoduchšie pracovať s oficiálnymi štatistikami a získavať analytické informácie bez potreby detailnej znalosti dátovej štruktúry. Aplikačné služby platformy budú prepojené s existujúcimi informačnými systémami prostredníctvom integračných rozhraní a dátových služieb. Integrácie budú umožňovať bezpečný prístup k dátovým zdrojom, publikovaným agregovaným údajom a analytickým výstupom. Súčasne bude zabezpečené, aby citlivé mikrodáta zostali spracovávané výlučne v chránenom prostredí a neboli sprístupnené externým systémom alebo verejným službám.

Logická architektúra navrhovanej AI platformy je znázornená na nasledujúcom obrázku a predstavuje vrstvený model riešenia, ktorý oddeľuje prezentačnú, aplikačnú, analytickú (AI/ML) a dátovú vrstvu systému. Tento architektonický prístup umožňuje jasné oddelenie zodpovedností jednotlivých komponentov platformy, zvyšuje modularitu riešenia a podporuje bezpečné spracovanie citlivých štatistických údajov. Architektúra je zároveň rozdelená do dvoch logických bezpečnostných zón – internej analytickej zóny (Zóna A) a verejnej publikačnej zóny (Zóna B). Toto rozdelenie zabezpečuje oddelenie spracovania citlivých mikrodát od verejne dostupných analytických služieb.

Navrhovaná logická architektúra riešenia je znázornená na nasledujúcom diagrame. Architektúra je organizovaná do štyroch hlavných vrstiev – dátovej vrstvy, AI/ML vrstvy, aplikačnej vrstvy a prezentačnej vrstvy. Každá z týchto vrstiev plní špecifickú úlohu v spracovaní štatistických údajov a poskytovaní analytických služieb používateľom. Architektúra zároveň rešpektuje bezpečnostné rozdelenie infraštruktúry Štatistického úradu SR do dvoch bezpečnostných zón. Zóna A je určená na spracovanie citlivých mikrodát a interných analytických procesov, zatiaľ čo Zóna B poskytuje služby verejnosti a pracuje výhradne s publikovanými agregovanými údajmi.



Obrázok 10 Logická architektúra AI platformy ŠÚ SR

Prezentačná vrstva zabezpečuje interakciu používateľov s platformou a poskytuje používateľské rozhrania pre interné aj externé skupiny používateľov. V internej zóne (Zóna A) je prezentačná vrstva reprezentovaná internými pracovnými konzolami (CMP-A13), ktoré poskytujú analytikom a metodikom ŠÚ SR prístup k analytickým nástrojom a AI službám nad internými dátami. Vo verejnej zóne (Zóna B) je prezentačná vrstva realizovaná prostredníctvom webového portálu ŠÚ SR (CMP-B10), ktorý poskytuje prístup k publikovaným štatistickým údajom, analytickým nástrojom a interaktívnym službám pre verejnosť.

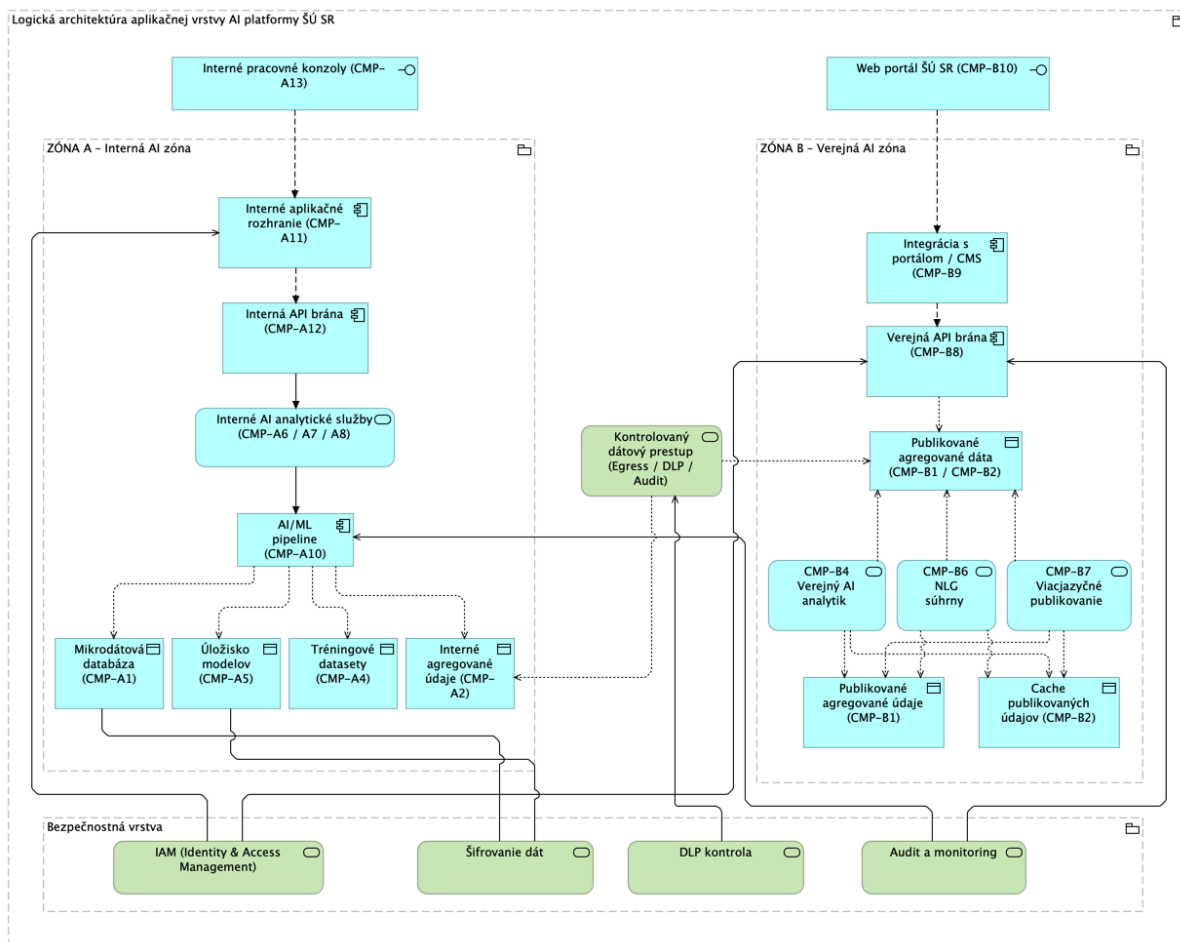
Aplikačná vrstva implementuje biznis logiku platformy a zabezpečuje integráciu medzi prezentačnou vrstvou a analytickými službami. V internej zóne je aplikačná vrstva tvorená interným aplikačným rozhraním (CMP-A11), ktoré poskytuje služby interným analytickým nástrojom, a internou API bránou (CMP-A12), ktorá zabezpečuje riadený prístup k analytickým službám a dátovým zdrojom. Vo verejnej zóne je aplikačná vrstva realizovaná prostredníctvom verejnej API brány (CMP-B8), ktorá poskytuje externé rozhrania pre prístup k publikovaným údajom a analytickým službám. Integrácia s existujúcim webovým portálom a redakčným systémom je zabezpečená komponentom integrácie s portálom alebo CMS (CMP-B9).

AI/ML vrstva predstavuje analytické jadro platformy a zabezpečuje implementáciu služieb umelej inteligencie a strojového učenia. V internej zóne (Zóna A) sú implementované služby určené na spracovanie citlivých mikroúdajov a podporu interných analytických procesov. Patria sem najmä služby klasifikácie textových odpovedí respondentov (CMP-A6), bezpečný AI analytik nad mikroúdajmi (CMP-A7), generovanie syntetických mikroúdajov (CMP-A8), register modelov umelej inteligencie (CMP-A9) a pipeline pre tréning a nasadzovanie modelov (CMP-A10). Vo verejnej zóne (Zóna B) sú implementované analytické služby určené pre prácu s publikovanými agregovanými údajmi. Medzi tieto služby patrí verejný AI analytik (CMP-B4), sémantické vyhľadávanie v štatistických údajoch (CMP-B5), automatické generovanie textových súhrnov (CMP-B6) a viacjazyčné publikovanie štatistických výstupov (CMP-B7).

Dátová vrstva zabezpečuje správu a uchovávanie dátových zdrojov využívaných platformou. V internej zóne sú spracúvané citlivé mikroúdaje a interné analytické výstupy. Medzi hlavné komponenty tejto vrstvy patrí mikrodátová databáza (CMP-A1), úložisko interných agregovaných údajov (CMP-A2), katalóg dát a metaúdajov (CMP-A3), tréningové úložisko datasetov (CMP-A4) a úložisko modelových artefaktov (CMP-A5). Vo verejnej zóne sú uložené publikované agregované štatistické údaje, ktoré sú určené na disemináciu a analytické využitie verejnosťou. Táto časť obsahuje úložisko publikovaných agregátov (CMP-B1), cache agregovaných údajov pre zrýchlenie analytických dotazov (CMP-B2) a konfiguračné úložisko platformy (CMP-B3).

Zásadným princípom architektúry je oddelenie internej analytickej zóny a verejnej publikačnej zóny. Zóna A je určená na spracovanie citlivých mikroúdajov a je prevádzkovaná v striktno kontrolovanom prostredí bez priamej externej konektivity. Zóna B poskytuje verejné analytické služby nad publikovanými agregovanými údajmi a je optimalizovaná pre vysokú dostupnosť a škálovateľnosť. Prenos údajov medzi zónami je povolený výhradne v smere zo Zóny A do Zóny B a podlieha kontrolným mechanizmom ochrany údajov vrátane pravidiel štatistického utajenia, procesov schvaľovania publikovania a bezpečnostných kontrol prevencie úniku dát.

Navrhovaná aplikačná architektúra tak vytvára modernú analytickú platformu, ktorá podporuje digitálnu transformáciu procesov spracovania štatistických údajov, zvyšuje efektivitu interných analytických činností a zároveň umožňuje poskytovať nové digitálne služby pre verejnosť. Rozdiel medzi súčasným a budúcim stavom aplikačnej architektúry spočíva najmä v rozšírení existujúcej architektúry o novú vrstvu aplikačných služieb umelej inteligencie. Kým súčasná architektúra podporuje najmä tradičné procesy zberu, spracovania a publikovania štatistických údajov, navrhovaná architektúra umožňuje automatizáciu vybraných analytických procesov a zavedenie nových služieb založených na umelej inteligencii. Nové aplikačné komponenty budú dopĺňať existujúce informačné systémy a rozširovať ich funkcionality bez zásadného zásahu do ich základnej architektúry. Platforma umelej inteligencie bude fungovať ako analytická nadstavba nad existujúcimi dátovými zdrojmi a publikačnými systémami. Z pohľadu biznis vrstvy bude navrhovaná aplikačná architektúra podporovať najmä procesy spracovania štatistických zisťovaní, tvorby analytických výstupov a diseminácie štatistických informácií. Nové aplikačné služby tak priamo prispievajú k efektívnejšiemu poskytovaniu koncových služieb úradu a k lepšej dostupnosti štatistických údajov pre verejnosť.



Obrázok 11 TO BE aplikačnej architektúry a závislostí (dátových tokov) medzi systémami

Obrázok znázorňuje logickú architektúru aplikačnej vrstvy navrhovanej AI platformy Štatistického úradu Slovenskej republiky. Diagram prezentuje hlavné aplikačné komponenty platformy, ich vzájomné prepojenia, väzby na dátové zdroje a bezpečnostné mechanizmy, ktoré zabezpečujú ochranu citlivých údajov a kontrolovaný prístup k analytickým službám. Architektúra je navrhnutá podľa princípu zónovej segmentácie, ktorá oddelí spracovanie citlivých mikroúdajov od verejných analytických služieb. Platforma je rozdelená na dve logické časti – internú AI zónu (Zóna A) a verejnú AI zónu (Zóna B). Medzi týmito zónami existuje striktno kontrolovaný dátový prechod, ktorý zabezpečuje, že citlivé mikroúdaje neopúšťajú chránené prostredie.

Zóna A – interná analytická zóna

Zóna A predstavuje chránené interné prostredie určené na spracovanie citlivých mikroúdajov a prevádzku interných analytických nástrojov umelej inteligencie. Táto časť platformy je dostupná výhradne interným používateľom Štatistického úradu SR prostredníctvom pracovných analytických konzol. Používatelia prístupujú k analytickým nástrojom prostredníctvom interného aplikačného rozhrania a internej API brány, ktoré zabezpečujú riadený prístup k jednotlivým analytickým službám. Tieto komponenty zároveň umožňujú riadenie prístupových práv, audit prístupov a integráciu s ďalšími internými systémami. Jadrom analytickej vrstvy v internej zóne sú interné AI analytické služby. Tieto služby umožňujú využívanie umelej inteligencie pri spracovaní štatistických údajov a podporujú viaceré analytické scenáre, napríklad klasifikáciu textových odpovedí zo štatistických zisťovaní, analytickú prácu nad mikroúdajmi alebo generovanie syntetických dátových súborov. Analytické služby sú implementované prostredníctvom AI/ML pipeline, ktorá zabezpečuje životný cyklus modelov umelej inteligencie. Táto pipeline zahŕňa tréning modelov, ich nasadzovanie, správu verzií modelov a realizáciu inferencie nad dátami. Pipeline pracuje s viacerými dátovými zdrojmi. Medzi najdôležitejšie patrí mikrodátová databáza obsahujúca

citlivé štatistické údaje, tréningové datasety určené na tréningovanie modelov umelej inteligencie a úložisko modelov, v ktorom sú uložené jednotlivé verzie analytických modelov. Okrem mikroúdajov platforma využíva aj interné agregované údaje, ktoré slúžia na analytické spracovanie a prípravu dát určených na publikovanie.

Kontrolovaný dátový prestup medzi zónami

Prepojenie medzi internou a verejnou zónou platformy je realizované prostredníctvom kontrolovaného dátového prestupu. Tento mechanizmus zabezpečuje, že do verejnej zóny môžu byť prenesené výhradne publikované agregované údaje, dátové kocky, metaúdaje a schválené syntetické datasety. Prestup údajov je riadený bezpečnostnými mechanizmami, ktoré zahŕňajú kontrolu prevencie úniku dát (DLP), audit prenosov a monitorovanie prístupov. Tento mechanizmus zabezpečuje dodržanie pravidiel štatistickej dôvernosti a ochrany osobných údajov.

Zóna B – verejná analytická zóna

Zóna B predstavuje verejnú časť AI platformy, ktorá poskytuje analytické služby nad publikovanými štatistickými údajmi. Táto zóna je určená pre externých používateľov, ako sú médiá, výskumné organizácie, analytici alebo široká verejnosť. Vstupným bodom do verejnej zóny je webový portál Štatistického úradu SR, ktorý poskytuje používateľské rozhranie pre prístup k analytickým službám. Portál je prepojený s integračnou vrstvou a verejnou API bránou, ktorá zabezpečuje riadený prístup k jednotlivým aplikačným službám platformy. Jadrom verejnej zóny sú analytické a diseminačné služby umelej inteligencie. Medzi ne patrí verejný AI analytik, ktorý umožňuje interaktívnu analýzu publikovaných agregovaných údajov, nástroje na automatické generovanie textových súhrnov štatistických výstupov a služby na viacjazyčné publikovanie štatistických informácií. Tieto služby pracujú výhradne s publikovanými agregovanými údajmi uloženými v dátových úložiskách verejnej zóny. Pre zvýšenie výkonu a rýchlosti odozvy platforma využíva aj cache publikovaných dát, ktorá umožňuje rýchle sprístupnenie často používaných analytických výstupov.

Bezpečnostná vrstva

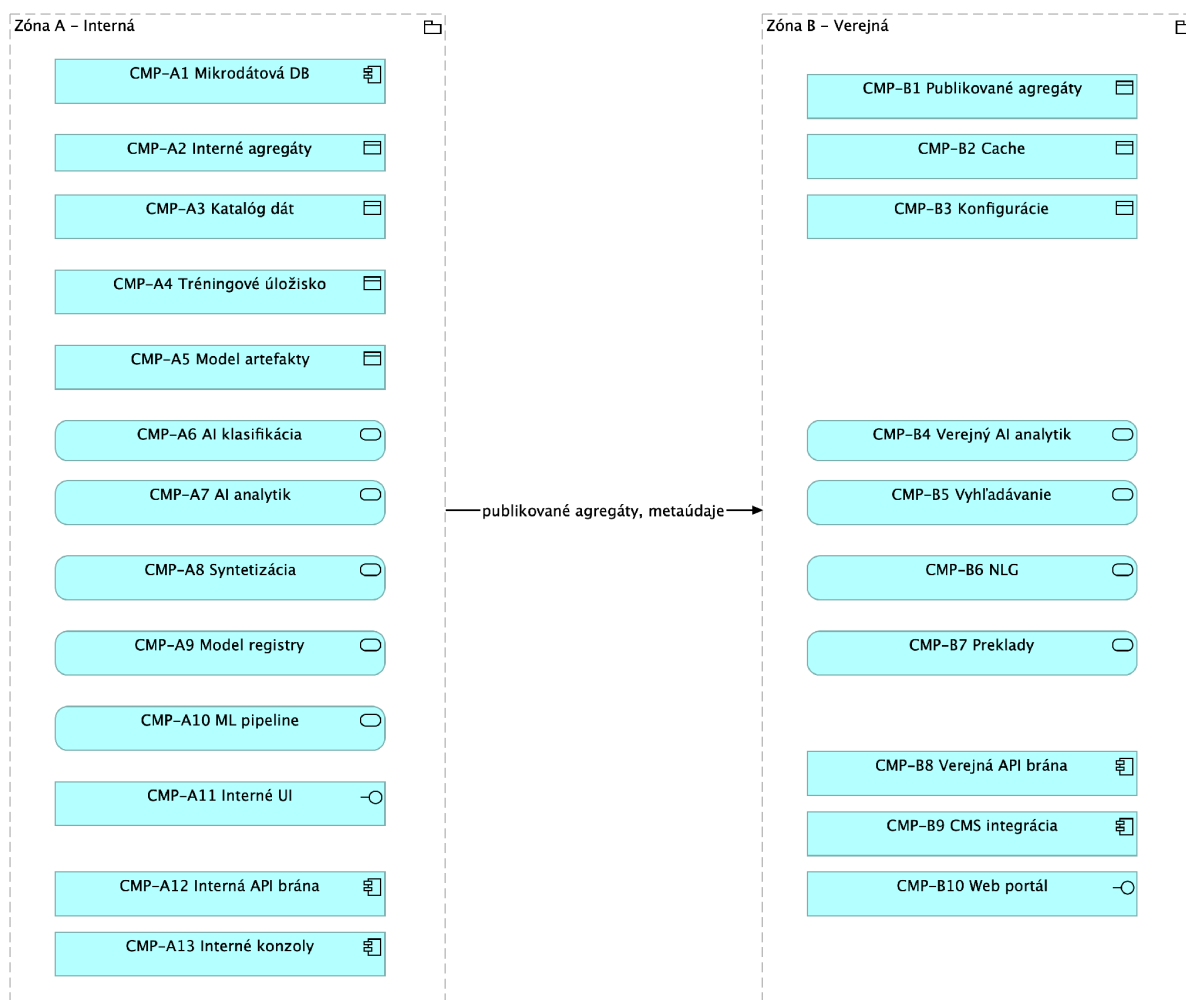
Celá aplikačná architektúra platformy je podporená bezpečnostnou vrstvou, ktorá zabezpečuje ochranu dát, kontrolu prístupov a monitorovanie prevádzky systému. Medzi hlavné bezpečnostné mechanizmy patrí systém riadenia identít a prístupov (IAM), ktorý kontroluje autentifikáciu a autorizáciu používateľov platformy. Dôležitou súčasťou architektúry je aj šifrovanie dát, ktoré zabezpečuje ochranu údajov počas prenosu aj počas ich uloženia. Platforma zároveň implementuje mechanizmy prevencie úniku dát (DLP), ktoré kontrolujú všetky prenosy údajov medzi jednotlivými komponentmi systému. Prevádzka platformy je monitorovaná prostredníctvom auditných a monitorovacích nástrojov, ktoré umožňujú sledovanie prístupov, detekciu bezpečnostných incidentov a zabezpečenie auditovateľnosti celého systému.

Význam navrhovanej aplikačnej architektúry

Navrhovaná aplikačná architektúra vytvára technologický základ pre bezpečné využívanie umelej inteligencie v prostredí oficiálnej štatistiky. Oddelenie interných analytických procesov od verejných analytických služieb umožňuje kombinovať moderné analytické nástroje s prísnyimi požiadavkami na ochranu citlivých údajov. Platforma zároveň umožňuje rozšíriť existujúce služby Štatistického úradu SR o nové digitálne analytické nástroje, ktoré zjednodušia prístup k štatistickým informáciám, zlepšia interpretáciu údajov a podporia širšie využívanie oficiálnych štatistík vo výskumnom a analytickom prostredí.

Komponentový model

Komponentový model opisuje menovité stavebné bloky cieľovej AI platformy Štatistického úradu SR (ŠÚ SR), ich zodpovednosti, poskytované rozhrania, umiestnenie v zónach a prevádzkové väzby. Model je navrhnutý tak, aby jednoznačne oddelil spracovanie citlivých mikroúdajov v Zóne A od verejných služieb v Zóne B a zároveň vytvoril stabilný rámec pre integrácie, bezpečnostné mechanizmy a budúce rozširovanie. Každý komponent je identifikovaný kódom CMP, pričom kódovanie reflektuje umiestnenie (A/B) a funkčnú úlohu v rámci dátovej, AI/ML, aplikačnej a prezentačnej vrstvy.



Obrázok 12 Znáozornenie komponentov

V Zóne A (interná chránená zóna) sa sústreďujú komponenty, ktoré spracúvajú mikroúdaje, interné agregáty a tréningové dáta, vrátane AI/ML pipeline, modelového registra a interných pracovných rozhraní. Základným dátovým úložiskom je CMP-A1 Mikrodátová databáza, ktorá spravuje primárne mikroúdaje a transakčné tabuľky používané v kurátorských procesoch. Podporuje SQL prístup (read/write pre kuráciu) a read-only prístup pre analytické dotazy realizované cez bezpečné aplikačné služby. Na úrovni medzivýstupov a kontroly kvality platforma využíva CMP-A2 Úložisko interných agregátov, ktoré obsahuje zoskupené údaje určené na interné analýzy a metodické overenia. Podporované sú interné rozhrania (SQL/Parquet) a väzby na zdroje mikroúdajov a metadát. Kľúčovým prvkom pre riadenie dátovej správy je CMP-A3 Katalóg dát a metaúdajov, ktorý eviduje klasifikáciu datasetov, rodokmeň spracovania (lineage), verziovanie, vlastníctvo a väzby na publikačné a integračné procesy. Katalóg je navrhnutý ako referenčný zdroj pravdy pre identifikáciu údajov v rámci platformy a podporuje rozhrania typu REST/JSON. Pre potreby tréningu a validácie modelov platforma obsahuje CMP-A4 Tréningové úložisko a CMP-A5 Úložisko modelových artefaktov. Tréningové úložisko uchováva tréningové a validačné datasety vo formátoch Parquet/CSV s registráciou v katalógu. Úložisko modelových artefaktov spravuje verzie modelov, ich balíčky a metadáta v integrácii s CI/CD a MLOps nástrojmi. Riadenie experimentov a modelového životného cyklu zabezpečuje CMP-A9 Register modelov a experimentov, pričom vykonávanie a orchestrácia tréningov, validácií a nasadení prebieha v rámci CMP-A10 AI/ML pipeline.

Na aplikačnej úrovni Zóny A sú definované služby, ktoré realizujú konkrétne use-case funkcie. CMP-A6 Služba klasifikácie otvorených odpovedí (UC-02) zabezpečuje automatizované kódovanie voľných textov a podporuje synchronné REST/JSON volania aj dávkové spracovanie CSV/Parquet. CMP-A7 Služba bezpečného AI analytika (UC-04) poskytuje interný analytický prístup k mikroúdajom a interným agregátom s vynútením

pravidiel štatistického utajenia. Služba prijíma dotazy cez aplikačné rozhranie a vykonáva read-only dotazy nad CMP-A1/CMP-A2. Pre generovanie syntetických datasetov je určený komponent CMP-A8 Služba syntetizácie mikrodát (UC-03), ktorý pracuje dávkovo a vytvára výstupy vo formátoch Parquet/CSV, pričom pôvodné mikroúdaje neopúšťajú Zónu A. Používateľské rozhrania a prevádzkové nástroje v Zóne A zabezpečujú CMP-A11 Interné aplikačné rozhranie (UI pre analytikov a metodikov) a CMP-A13 Interné pracovné konzoly (konzoly pre prevádzku, monitoring a kuráciu). Všetky interné volania sú sprostredkované cez CMP-A12 Internú API bránu, ktorá vynucuje autentizáciu/autorizáciu, trasovanie, kvóty, auditovateľnosť a jednotné bezpečnostné politiky v internom segmente. V Zóne B (verejná zóna) sa nachádzajú komponenty, ktoré poskytujú verejné služby nad publikovanými agregátmi a verejným obsahom, bez prístupu k mikroúdajom. Zdrojom publikovaných výstupov je CMP-B1 Úložisko publikovaných agregátov, ktoré predstavuje „zdroj pravdy“ pre verejné dotazy. Na dosiahnutie nízkej latencie je doplnené o CMP-B2 Cache publikovaných agregátov s definovanými TTL politikami. Konfiguračné a prezentačné artefakty verejných služieb spravuje CMP-B3 Konfiguračné úložisko verejných služieb. Interného AI asistenta (UC-01) realizuje komponent CMP-A14 Interný AI asistent, ktorý odpovedá zamestnancom v prirodzenom jazyku výhradne z obsahu sprístupneného znalostnou vrstvou a pri každej odpovedi uvádza zdrojový dokument, jeho verziu a odkaz na príslušnú časť dokumentu. Obsah pre asistenta poskytuje CMP-A15 Znalostná a vyhľadávacia vrstva internej zóny, ktorá rozdeľuje interné dokumenty na pasáže, získava ich metaúdaje a vytvára fulltextový a vektorový index, ktorý nie je dostupný z verejnej zóny. Prestupy údajov do verejnej zóny zabezpečuje CMP-A16 Brána kontrolovaného prestupu medzi zónami, ktorá vynucuje jednosmerný prestup z internej do verejnej zóny, vykonáva automatizované kontroly štatistického utajenia a prevencie úniku údajov a prestup umožní až po schválení podľa pravidiel štyroch očí so zaznamenaním do auditného záznamu.

Verejné aplikačné služby realizujú najmä UC-05 až UC-08. CMP-B4 Verejný AI analytik (UC-05) poskytuje interaktívne odpovede a vizualizácie nad publikovanými agregátmi, pričom číta výhradne z CMP-B1/B2. CMP-B5 Sémantické vyhľadávanie a odporúčania (UC-06) zabezpečuje vyhľadávacie služby nad verejným obsahom a metaúdajmi, vrátane indexačných pipeline. CMP-B6 NLG súhrny (UC-07) umožňujú generovanie textových súhrnov k publikovaným dátam v dávkovom aj online režime so zapojením publikačného workflow. CMP-B7 Viacjazyčné publikovanie (UC-08) podporuje automatizované preklady a terminologickú konzistenciu s human-in-the-loop (HIL) schvaľovaním. Všetky verejné volania prechádzajú cez CMP-B8 Verejnú API bránu, ktorá zabezpečuje autentizáciu/autorizáciu podľa typu služby, rate limiting, kvóty a integráciu s WAF. Integrácia s portálom a publikačné workflow sú realizované cez CMP-B9 Integráciu s portálom/CMS, pričom samotné prezentačné rozhranie pre verejnosť tvorí CMP-B10 Web portál. Vyhľadávacie služby vo verejnej zóne využívajú CMP-B11 Verejný index znalostnej a vyhľadávacej vrstvy, ktorý obsahuje fulltextový a vektorový index publikovaného obsahu a metaúdajov, je oddelený od indexu internej zóny a aktualizuje sa prírastkovo.

Dátové toky a prestupy

Platforma definuje záväzné pravidlá pohybu údajov v rámci platformy a medzi zónami s cieľom zabezpečiť ochranu štatistickej dôvernosti, rezidenciu dát a konzistentné uplatňovanie metodických pravidiel publikovania. Návrh vychádza zo zónového modelu, v ktorom sa všetky citlivé spracovania realizujú výhradne v Zóne A a všetky verejné služby v Zóne B pracujú iba s publikovanými agregovanými údajmi, dátovými kockami, metaúdajmi a schválenými syntetickými datasetmi. Pravidlá uvedené v tejto kapitole sa uplatňujú na všetky integračné body, pipeline, publikačné procesy a aplikačné služby. Platforma rozlišuje tri základné kategórie údajov podľa miery citlivosti a spôsobu použitia. Mikroúdaje predstavujú individuálne záznamy o osobách alebo podnikoch a sú spracúvané výhradne v Zóne A. Interné agregáty sú zoskupené údaje určené na internú analýzu, metodické overenie a kontrolu kvality. Zostávajú v Zóne A a nie sú určené na publikovanie. Publikované agregáty sú zoskupené údaje schválené na zverejnenie. Do Zóny B môžu byť sprístupnené len publikované agregované údaje, dátové kocky, metaúdaje a schválené syntetické datasety. Z hľadiska prevádzkových tokov platforma pracuje s internými tokmi v Zóne A a verejnými tokmi v Zóne B. Medzi typické interné toky v Zóne A patrí analytické čítanie mikroúdajov a interných agregátov pre potreby bezpečného analytika (CMP-A1 → CMP-A7), spracovanie mikroúdajov pre automatickú klasifikáciu (CMP-A1 → CMP-A6) a dávkové spracovanie pre syntetizáciu (CMP-A1 → CMP-A8). Verejné toky v Zóne B sú striktné read-only a zahŕňajú najmä sprístupnenie publikovaných agregátov cez portál a verejné služby (CMP-B10 → CMP-B4 → CMP-B1/2) a vyhľadávacie služby (CMP-B10 → CMP-B5). Prestupy medzi zónami sú definované ako jednosmerné A → B a slúžia výhradne na sprístupnenie publikovaných agregovaných údajov, dátových kociek, metaúdajov a schválených syntetických datasetov. Prestup sa realizuje len z interných zdrojov agregácií a metaúdát (CMP-A2/CMP-A3) a zo služby syntetizácie mikrodát (CMP-A8) do verejných úložísk (CMP-B1/CMP-B3). Pri verejných tokoch IN-03 je prístup realizovaný cez verejnú API bránu CMP-B8, pričom odpovede nesú referenciu na zdroj (sourceRef) a verziu datasetu (datasetVersion). Všetky požiadavky sú logované s korelačným identifikátorom a podliehajú kvótam a rate limiting mechanizmom. Interné

agregáty sa môžu stať publikovanými iba po absolvovaní riadeného procesu: automatické testy rizika reidentifikácie, metodická revízia, schválenie podľa zásady „four-eyes“ a následný export cez egress/DLP mechanizmus.

Povolené prestupy zo Zóny A do Zóny B

Povolené prestupy zo zóny A do zóny B prebiehajú výhradne v rámci definovaného egress workflow. Proces začína označením datasetu ako „publikovateľný“ v katalógu CMP-A3, pokračuje automatizovanými kontrolami rizika reidentifikácie a validačnými testami, nasleduje metodická revízia a schválenie podľa zásady štyroch očí. Až po schválení sa realizuje export cez DLP/egress bránu a uloženie do CMP-B1 (publikované agregáty) a CMP-B3 (sprievodné metaúdaje/konfigurácie), pričom celý postup je zakončený potvrdením s auditnou stopou. Prenosové formáty pre publikované balíky sú CSV/JSON/Parquet a každý balík obsahuje povinné metadáta: čas publikácie, identifikátor zdroja, verziu, klasifikáciu a kontrolné súčty. Pred exportom sa vykonávajú validačné kroky vrátane kontroly absencie PII, uplatnenia minimálnych veľkostí buniek, kontroly kombinovateľnosti dimenzií, integrity a podpisu dátového balíka. Priamy SQL prístup z externého perimetra do databáz v Zóne B sa nepovoľuje. Externé prístupy sú sprostredkované aplikačnými službami za verejnou bránou CMP-B8. Interné read-only DB roly sú používané výhradne na komunikáciu medzi službami v rámci Zóny B.

Zakázané toky a výnimky

Zakázaný je akýkoľvek prenos mikroúdajov alebo interných agregátov do Zóny B, rovnako aj priame dotazovanie verejných komponentov na interné zdroje v Zóne A. Zakázané je aj spätné prenášanie obsahu z verejných služieb do Zóny A bez formálneho schválenia a mechanizmov, ktoré by garantovali súlad so zónovým modelom. Návrh nepredpokladá priame učenie modelov v Zóne A na základe nespracovaných verejných interakcií. Ak by sa výnimočne uvažovalo o odchýlke, vyžadovalo by to formálne posúdenie rizík, schválenie a časové obmedzenie. Pri pokuse o export mikroúdajov, obchádzanie egress brány alebo obchádzanie pravidiel štatistického utajenia je operácia zablokovaná, vzniká záznam v SIEM a spúšťa sa eskalačný postup podľa runbooku.

Schvaľovanie a audit

Každý prestup medzi zónami podlieha schvaľovaciemu procesu zahŕňajúcemu kontrolu klasifikácie dát, verifikáciu pravidiel publikovania a konfiguráciu DLP/egress politik. Prestupy sa vykonávajú výlučne cez schválené brány s vynúteným šifrovaním a autentizáciou. Auditné záznamy obsahujú identitu, čas, rozsah a účel prestupu a sú uchovávané počas definovaných retenčných lehôt pre potreby auditov a forenznej analýzy. Periodické kontroly vyhodnocujú súlad prestupov s politikami a identifikujú odchýlky.

Agregačné a anonymizačné pravidlá pred prestupom

Pred publikovaním sa uplatňujú pravidlá štatistického utajenia vrátane minimálnych veľkostí buniek, potláčania nízkych početností, zaokrúhľovania a kontroly rizikových kombinácií dimenzií. Prahové hodnoty a pravidlá sú spravované ako verzovaný konfiguračný artefakt metodiky v CMP-A3. Validácia rizika zahŕňa výpočet ukazovateľov reidentifikácie, kontrolu linkability s inými publikovanými zdrojmi a metodickú revíziu s výsledkom „OK / OK s výhradou / NEPUBLIKOVAŤ“. Uplatnenie týchto pravidiel prebieha jednak v pipeline CMP-A7 (pre ad-hoc analytické výstupy), jednak v publikačnom procese exportu A → B. Všetky zásahy sú logované s uvedením pravidla a rozsahu úprav.

Technické kanály a protokoly

Všetky API používajú HTTPS s TLS 1.2+ a interné kanály v Zóne A vyžadujú mTLS. Certifikáty sú spravované a rotované podľa internej politiky. Read-only prístupy v Zóne B sú realizované cez vyhradené DB roly s SELECT oprávneniami a pripravené pohľady alebo endpointy pre publikované agregáty. Ochrany zahŕňajú limity veľkosti payloadov, timeouty, retry politiku s backoff a rate limiting ochranu na verejnej bráne CMP-B8. Pre pilotné prevádzkové prostredie vo verejnom cloude v rámci EÚ sa uplatňuje povinný režim: TLS 1.2+, WAF a rate limiting na perimetri, žiadne inbound spojenia do Zóny A a dátové toky výhradne v rozsahu necitlivých, syntetických a verejných údajov.

Prevádzkové politiky

Cieľové SLO pre kľúčové toky sú definované najmä pre verejné dotazy v Zóne B, kde sa predpokladá latencia do 2 sekúnd pre bežné prípady a do 5 sekúnd pre zložitejšie dotazy. Prestupy A → B prebiehajú v rámci publikačných okien podľa veľkosti datasetu a interných kapacít. Monitoring sleduje latenciu a chybovosť API, priepustnosť brán, počet odmietnutých prestupov, chybovosť validačných kontrol rizika a vyťaženie cache. Alerting je naviazaný na prahové hodnoty s definovanými reakčnými lehotami a eskaláciami.

Referenčné toky

- Referenčný tok IN-01 „Verejný AI analytik → publikované agregáty“ je realizovaný tak, že používateľ zadá otázku na Web portáli (CMP-B10), ktorá je odoslaná do Verejného AI analytika (CMP-B4) cez Verejnú API bránu (CMP-B8). Služba následne mapuje dotaz na publikované agregáty v CMP-B1 s využitím cache CMP-B2, vytvorí odpoveď a vizualizáciu a odošle výsledok späť portálu. Všetky volania sú read-only, logované s correlationId a limitované kvótami.
- Referenčný tok IN-09 „Bezpečný AI analytik → mikrodátové databázy“ prebieha tak, že analytik zadá dotaz cez Interné UI (CMP-A11), Služba bezpečného AI analytika (CMP-A7) ho transformuje na schémový dotaz, vykoná ho nad Mikrodátovou DB (CMP-A1) a Internými agregátmi (CMP-A2), aplikuje pravidlá utajenia a vráti agregovaný výsledok v rámci Zóny A. Tok je plne logovaný a auditovaný a nevyužíva externú konektivitu.
- Referenčný tok IN-05 „Portál ↔ Verejný AI analytik“ zabezpečuje obojsmernú výmenu otázok a odpovedí medzi CMP-B10 a CMP-B4 cez integračnú vrstvu CMP-B9 a Verejnú API bránu CMP-B8. Odpovede (vrátane grafov) sú doručené rovnakou cestou späť na zobrazenie. Kvóty a ochrany na CMP-B8 bránia zneužitiu.
- Referenčný tok IN-03 je realizovaný ako programový prístup klienta k publikovaným agregátom cez endpoint /api/v1/aggregates/query, pričom volanie prechádza cez CMP-B8, kde sa aplikuje autentizácia (ak je vyžadovaná), kvóty a WAF. Služba načíta dáta z CMP-B1/CMP-B2 a vráti odpoveď v štruktúre data[], columns[], sourceRef a datasetVersion. Zapisujú sa correlationId a metriky latencie a chýb.

Pre všetky referenčné toky sú definované kontrolné body a minimálne logované polia: correlationId, identifikátor používateľa alebo relácie (podľa typu služby), časové pečiatky, status kódy, identifikátor zdrojového datasetu a jeho verzia. Detailné sekvenčné diagramy a dátové kontrakty budú súčasťou príloh a kapitoly Integrácie a dátové rozhrania.

Integrácie a dátové rozhrania

Integračný návrh vychádza z princípu API-first, pričom všetky integračné rozhrania sú navrhované ako štandardizované a dokumentované aplikačné služby. Architektúra zároveň rešpektuje princíp dôslednej separácie Zóny A a Zóny B v infraštruktúre dátového centra Štatistického úradu SR, ako aj jednotné pravidlá bezpečnosti, verziovania, monitorovania a auditovateľnosti integračných rozhraní.

Verejné integrácie realizované v Zóne B sú striktne navrhnuté ako read-only služby a pracujú výhradne s publikovanými agregovanými údajmi a metaúdajmi. Naopak integrácie v Zóne A sú určené pre interné spracovanie mikroúdajov a analytických datasetov. Tieto integračné toky sú prevádzkované výlučne v internom segmente infraštruktúry, riadené na základe rolí a prístupových oprávnení a podliehajú plnej auditovateľnosti. Platforma využíva jednotnú API stratégiu založenú na štandardizovanom názvosloví rozhraní a konzistentnom verziovaní. Všetky aplikačné rozhrania sú publikované pod jednotnou konvenciou /api/v{major}/... a sú dokumentované prostredníctvom špecifikácie OpenAPI. Zmeny rozhraní, ktoré nie sú spätne kompatibilné, vedú k zvýšeniu hlavnej verzie API, pričom predchádzajúca verzia zostáva dostupná počas definovaného obdobia deprecácie, aby sa zabezpečila kontinuita služieb pre existujúcich konzumentov. Každé volanie API obsahuje korelačný identifikátor prenášaný v hlavičke X-Correlation-Id, ktorý slúži na sledovanie požiadaviek v rámci distribuovaných systémov. Ak tento identifikátor nie je odoslaný klientom, je automaticky generovaný integračnou bránou a následne vrátený v odpovedi.

Mechanizmy autentizácie a autorizácie sú diferencované podľa bezpečnostnej zóny. V Zóne A sa využívajú zabezpečené komunikačné kanály s obojstrannou autentifikáciou (mTLS) a tokeny viazané na konkrétnu rolu alebo dataset. V Zóne B sú verejné API chránené kombináciou šifrovanej komunikácie TLS, aplikačných tokenov, ochrany na úrovni webovej aplikačnej brány (WAF) a mechanizmov limitovania požiadaviek a kvót. Všetky integračné kanály vyžadujú používanie protokolu TLS 1.2 alebo vyššieho, pričom správa certifikátov je realizovaná v súlade s internou bezpečnostnou politikou organizácie. Formáty prenášaných údajov sú jednotné naprieč celou platformou.

Synchronné aplikačné služby využívajú dátový formát JSON s kódovaním UTF-8, zatiaľ čo dávkové prenosy používajú formáty CSV alebo Parquet, doplnené o metaúdaje vo formáte JSON. Chybové odpovede API sú štandardizované a obsahujú minimálne polia code, message, correlationId, timestamp a prípadne aj doplňujúce pole details.

Pre integračné toky s väčším objemom údajov sú definované limity časov spracovania, maximálne veľkosti prenášaných dát a mechanizmy opakovaných pokusov (retry) s exponenciálnym backoff algoritmom. Idempotencia operácií je zabezpečená najmä pri dávkových operáciách prostredníctvom identifikátora idempotency key. V prípade verejných read-only API v Zóne B sa idempotentnosť operácií predpokladá implicitne.

Riadenie životného cyklu API (API governance) pokrýva celý proces návrhu, publikovania a údržby integračných rozhraní. Každé API má definovaného vlastníka, je evidované v katalógu rozhraní a obsahuje opis účelu, dátové schémy, príklady použitia, cieľové parametre služby (SLO) a plán prípadnej deprecácie. Zmeny kontraktov rozhraní prechádzajú formálnym schvaľovacím procesom, testami spätnej kompatibility a publikáciou zmien prostredníctvom changelogu. Nasadzovanie nových verzií prebieha v kontrolovaných oknách a môže využívať režimy canary deployment alebo shadow testing, pričom sú pripravené aj postupy pre rýchly rollback v prípade incidentov.

Bezpečnosť a observabilita integračných služieb sú integrálnou súčasťou architektúry. Všetky volania API prechádzajú cez integračné brány CMP-A12 v Zóne A alebo CMP-B8 v Zóne B, kde sa uplatňujú mechanizmy autentizácie, autorizácie, limitovania požiadaviek, detekcie anomálií a distribuovaného trasovania požiadaviek. Logovacie mechanizmy zaznamenávajú minimálne korelačný identifikátor, identitu volajúceho subjektu, identifikátor endpointu, časové pečiatky, kódy odpovedí a v prípade dávkových operácií aj počet spracovaných záznamov. Monitorovanie prevádzky sleduje najmä latenciu služieb, chybovosť, priepustnosť, veľkosti prenášaných dát a vyťaženie cache vrstiev. Na základe definovaných cieľových parametrov služby (SLO) sú nastavené automatické alerty a eskalačné postupy prevádzky. Pre verejné API sú nastavené štandardné limity používania na úrovni 100 požiadaviek za minútu a maximálne 10 požiadaviek za sekundu na klienta. Individuálne kvóty je možné upraviť prostredníctvom autorizácie pomocou API key alebo OAuth2. Verejné endpointy sú navrhnuté výhradne ako read-only služby a sú chránené proti zneužitiu prostredníctvom mechanizmov ochrany proti botom, TLS šifrovania a WAF filtrov. Prevádzka v pilotnom prostredí vo verejnom cloude v rámci EÚ alebo v produkčnom prostredí dátového centra ŠÚ SR nevyžaduje zmenu API kontraktov. Rozhrania, verzovanie, OpenAPI dokumentácia aj cieľové parametre služby zostávajú identické, takže konzumenti API nemusia rozlišovať medzi jednotlivými prostrediami.

Katalóg integračných bodov

Katalóg integračných bodov sumarizuje hlavné dátové toky medzi jednotlivými komponentmi platformy a externými alebo internými systémami, ktoré poskytujú alebo konzumujú údaje v rámci navrhovaného riešenia. Každý integračný bod je definovaný prostredníctvom svojho účelu, smeru dátového toku, rozsahu prenášaných údajov, očakávanej frekvencie alebo latencie komunikácie, umiestnenia v rámci bezpečnostných zón architektúry a relevantných bezpečnostných opatrení. Integračné rozhrania sú implementované prostredníctvom štandardizovaných API alebo dátových pipeline, ktoré umožňujú riadený prístup k údajom a ich spracovanie jednotlivými aplikačnými komponentmi platformy. Endpointy sú obsluhované komponentmi umiestnenými v Zóne B, ktoré sú v produkčnej prevádzke nasadené v dátovom centre ŠÚ SR a v pilotnej fáze v izolovanom prostredí verejného cloudu v rámci EÚ. Pre všetky verejné alebo externé volania sú definované štandardné limity používania, pričom základná kvóta je nastavená na 100 požiadaviek za minútu s maximálnym burst limitom 10 požiadaviek za sekundu. Výnimky z týchto limitov môžu byť udelené prostredníctvom autorizácie pomocou API kľúčov alebo OAuth2 mechanizmov na základe schválenia prevádzkovateľom systému.

IN-01 Integrácia s dátovými kockami a agregovanými databázami

Cieľom integračného bodu je zabezpečiť konzistentný a nízkolatenčný prístup k publikovaným agregovaným údajom a metadátam pre verejného AI analytika a súvisiace aplikačné služby v Zóne B. Zdrojom údajov je úložisko publikovaných agregovaných datasetov reprezentované komponentom CMP-B1 a cache vrstvou CMP-B2. Konzumentmi údajov sú analytické a generatívne služby platformy, najmä komponenty CMP-B4 a CMP-B6, ako aj dátový portál CMP-B10, pričom komunikácia je sprostredkovaná integračnou bránou CMP-B8. Prenos údajov je realizovaný prostredníctvom REST rozhraní vo formáte JSON alebo prostredníctvom read-only SQL pohľadov. Synchronné volania využívajú štandardné časové limity a odpovede obsahujú nielen samotné agregované hodnoty, ale aj referenciu na zdrojové dátové štruktúry. Kontrakty integračného rozhrania definujú schému entít, medzi ktoré

patria ukazovatele, dimenzie, členovia dimenzií a časové obdobia, ako aj podporované filtre, stránkovanie a formát agregovaných výsledkov. Cieľová úroveň služby predpokladá latenciu odpovede do dvoch sekúnd pri bežných dotazoch a vysokú úroveň dostupnosti. Prístup k údajom je striktné read-only a je chránený prostredníctvom WAF, mechanizmov limitovania požiadaviek a autentifikačných mechanizmov. Monitoring sleduje najmä latenciu odpovedí, mieru využitia cache vrstvy a chybovosť jednotlivých volaní.

IN-02 Napojenie na zberové systémy a registre

Integračný bod zabezpečuje príjem a kurátorské spracovanie mikroúdajov získaných zo zberových systémov a registrov, ktoré sú následne využívané v analytických procesoch platformy. Zdrojom údajov sú zberové a zdrojové systémy prevádzkované v internom segmente infraštruktúry. Konzumentmi údajov sú príjmové a kurátorské pipeline implementované v komponentoch dátovej a AI/ML vrstvy platformy, najmä CMP-A1, CMP-A4 a CMP-A6. Prenos údajov prebieha v rámci Zóny A buď dávkovým spôsobom alebo v režime nízkej latencie. Komunikácia je zabezpečená prostredníctvom šifrovaných kanálov využívajúcich mTLS a podpisovanie dátových dávok. Podporované formáty údajov zahŕňajú najmä CSV a Parquet súbory doplnené o sprievodné manifesty vo formáte JSON obsahujúce kontrolné súčty a metaúdaje. Integračné kontrakty definujú validačné pravidlá, spracovanie chýb a mechanizmy spätného hlásenia o stave spracovania dávok. Úroveň služby je definovaná najmä časovým oknom kurácie údajov. Všetky dátové toky prebiehajú výlučne v Zóne A a sú plne auditované a riadené prostredníctvom rolí a prístupových práv.

IN-03 Verejný API pre agregované výstupy

Tento integračný bod poskytuje verejné read-only API určené na prístup k publikovaným agregovaným údajom a prípadne k vybraným analytickým výstupom verejného AI analytika. API umožňuje externým konzumentom, ako sú verejnosť, partnerské organizácie alebo médiá, programový prístup k agregovaným štatistickým údajom a súvisiacim metadátam. Všetky poskytované údaje sú anonymizované a agregované tak, aby neobsahovali osobné údaje ani rizikové kombinácie dimenzií. API je dostupné v Zóne B a je chránené prostredníctvom bezpečnostných mechanizmov vrátane limitovania požiadaviek, autentifikácie a ochrany na úrovni aplikačnej brány. Štandardná kvóta používania je stanovená na 100 požiadaviek za minútu s maximálnym burst limitom 10 požiadaviek za sekundu. Výnimky z limitov je možné udeliť prostredníctvom autorizácie pomocou API kľúčov alebo OAuth2 mechanizmov.

IN-04 Generovanie syntetických dát

Integračný bod zabezpečuje prepojenie medzi zdrojovými mikrodátovými datasetmi a nástrojom syntetizácie údajov implementovaným v komponente CMP-A8. Zdrojom údajov sú komponenty CMP-A1 a CMP-A4, ktoré poskytujú vstupné datasety pre generovanie syntetických dát. Proces generovania prebieha dávkovým spôsobom v Zóne A až do momentu schválenia výsledného datasetu. Po schválení sú syntetické datasety spolu s príslušnými metaúdajmi uložené v úložiskách CMP-B1 alebo CMP-B3. Prenos údajov využíva najmä formáty Parquet alebo CSV, pričom súčasťou výstupu sú aj hodnotiace reporty utility a rizika vo formáte HTML alebo PDF. Integračný kontrakt definuje požadované metaúdaje, identifikátory verzií datasetov, hodnotiace metriky a stavový model spracovania pozostávajúci zo stavov „draft“, „posúdenie“, „schválené“ alebo „neschválené“. Bezpečnostné pravidlá zabezpečujú, že originálne mikroúdaje neopúšťajú Zónu A a že prípadný presun syntetických datasetov do Zóny B prechádza kontrolnými mechanizmami egress a DLP.

IN-05 Verejný AI analytik – dátový portál

Integračný bod zabezpečuje obojsmernú komunikáciu medzi dátovým portálom CMP-B10 a analytickou službou verejného AI analytika CMP-B4. Komunikácia prebieha prostredníctvom integračnej brány CMP-B8 a umožňuje prenos používateľských otázok aj generovaných odpovedí. Transport údajov je realizovaný prostredníctvom REST rozhraní vo formáte JSON s podporou streamingových odpovedí pri komplexnejších dotazoch. Kontrakty integračného rozhrania definujú štruktúru otázok, ktoré môžu obsahovať voľný text aj voliteľné parametre, ako aj štruktúru odpovedí, ktoré môžu obsahovať textové vysvetlenie, tabuľky, grafy alebo odkazy na relevantné datasety. Cieľová latencia odpovede pri bežných dotazoch je do dvoch sekúnd. Bezpečnostné opatrenia zahŕňajú ochranu pomocou WAF, mechanizmy rate-limitingu, ochranu proti automatizovaným botom a obsahové filtre. Monitoring sleduje najmä počet dotazov, latenciu odpovedí, typy chýb a mieru eskalácií.

IN-06 Inteligentné vyhľadávanie na webe

Integračný bod zabezpečuje prepojenie medzi portálom alebo CMS systémom CMP-B9 a vyhľadávacou službou CMP-B5. Indexácia obsahu prebieha plánovaným spôsobom alebo na základe udalostí a využíva výhradne verejne dostupný obsah a metaúdaje publikované v Zóne B. Používateľské dotazy sú odosielané z webového rozhrania do vyhľadávacej služby a výsledky sú vracané vo forme zoradeného zoznamu dokumentov doplnených o metaúdaje a náhľady obsahu. Integračné kontrakty definujú štruktúru indexu, formát dotazov, rankingové signály a podporu viacjazyčného vyhľadávania. Cieľová latencia odpovedí sa pohybuje v rozsahu jednej až dvoch sekúnd. Bezpečnostné opatrenia zahŕňajú ochranu pred indexovaním citlivého obsahu, využitie WAF a limitovanie požiadaviek. Monitoring sleduje najmä úspešnosť nájdenia relevantného obsahu, počet nulových výsledkov a mieru preklikov na odporúčané výsledky.

IN-07 NLG súhrny k publikovaným dátam

Tento integračný bod zabezpečuje prepojenie medzi službou generovania prirodzeného jazyka CMP-B6 a zdrojmi publikovaných agregovaných údajov CMP-B1 a CMP-B2, ako aj publikačným workflow v CMS CMP-B9. Generovanie textových súhrnov môže prebiehať dávkovo pri aktualizácii datasetov alebo online na požiadanie používateľa. Integračné kontrakty definujú vstupné parametre generovania, medzi ktoré patrí dataset, šablóna a rozsah generovaného textu. Výstup môže byť poskytovaný vo formáte HTML, Markdown alebo JSON. Súčasťou procesu je aj mechanizmus human-in-the-loop (HIL), ktorý umožňuje manuálne schválenie alebo úpravu generovaných textov. Cieľová latencia pri interaktívnom generovaní je do dvoch sekúnd. Monitoring sleduje najmä podiel textov schválených bez zásadných úprav a počet manuálnych zásahov v procese publikovania.

IN-08 Viacjazyčné publikovanie

Integračný bod zabezpečuje integráciu medzi CMS systémom CMP-B9 a prekladovou službou CMP-B7, ktorá umožňuje automatizované generovanie jazykových mutácií publikovaného obsahu. Komunikácia je obojsmerná – CMS odovzdáva obsah spolu s metaúdajmi a prekladová služba vracia preložené verzie textov a terminologické návrhy. Integračné kontrakty definujú formát vstupného obsahu (HTML, Markdown alebo JSON), podporované jazykové páry, využitie terminologických glosárov a stavový model publikovania obsahujúci stavy „na revíziu“ a „publikované“. Služba pracuje výhradne s verejne publikovaným obsahom. Monitoring sleduje najmä čas spracovania prekladov, počet manuálnych zásahov v procese revízie a konzistentnosť terminológie.

IN-09 Interný AI analytik nad mikrodátami

Integračný bod poskytuje službu bezpečného interného AI analytika CMP-A7 riadený prístup k mikrodátovým databázam a interným agregovaným datasetom. Zdrojom údajov sú komponenty CMP-A1 a CMP-A2, pričom komunikácia prebieha prostredníctvom internej integračnej brány CMP-A12. Prenos údajov využíva read-only SQL pohľady pre dátové dotazy a REST rozhrania pre prístup k metadátam a logovacím službám. Integračné kontrakty definujú mapovanie analytických dotazov na dátové schémy, povolené operácie a rozhranie pre poskytovanie agregovaných výstupov. Úroveň služby závisí od komplexnosti analytických dotazov a pohybuje sa v rozsahu sekúnd až minút. Bezpečnostné pravidlá zabezpečujú aplikovanie princípov štatistického utajenia a kontrolu nad rozsahom poskytovaných výsledkov. Monitoring zahŕňa sledovanie latencie dotazov, objemu výsledkov po aplikovaní pravidiel utajenia a evidenciu auditných záznamov.

Pre lepšiu orientáciu v architektúre riešenia sú vzťahy medzi scenármi použitia, aplikačnými službami a integračnými bodmi sumarizované v nasledujúcich tabuľkách. Tabuľky umožňujú identifikovať, ktoré aplikačné služby realizujú jednotlivé scenáre použitia a prostredníctvom ktorých integračných bodov prístupujú k dátovým zdrojom alebo externým systémom. V ďalšej tabuľke je uvedená väzba medzi aplikačnými službami a architektonickými komponentmi platformy, čím sa zabezpečuje konzistentné mapovanie medzi logickou architektúrou a implementačnými komponentmi systému.

Nasledujúca tabuľka sumarizuje väzby medzi jednotlivými scenármi použitia (use cases), aplikačnými službami platformy a integračnými bodmi, prostredníctvom ktorých sú tieto služby implementované. Cieľom tejto tabuľky je prehľadne identifikovať, aké integračné rozhrania sú využívané jednotlivými funkčnými scenármi platformy a aké dátové toky zabezpečujú realizáciu konkrétnych analytických alebo publikačných funkcionalít. Pre každý scenár použitia je uvedená zodpovedajúca aplikačná služba a integračné body, ktoré zabezpečujú prístup k dátovým

zdrojom alebo komunikáciu medzi komponentmi platformy. Táto väzba umožňuje jednoznačne identifikovať, ktoré časti architektúry sú využívané pri realizácii jednotlivých funkcionalít a zároveň poskytuje základ pre plánovanie prevádzky, monitorovania a riadenia integračných rozhraní. Prehľad väzieb zároveň podporuje architektonickú konzistenciu návrhu, pretože umožňuje sledovať prepojenie medzi biznis požiadavkami projektu, aplikačnými službami platformy a konkrétnymi integračnými rozhraniami implementovanými v architektúre riešenia.

Use case	Aplikačná služba	Integračný bod
Interný AI asistent	Metodická podpora zamestnancov nad znalostnou bázou úradu	bez integračného bodu
Automatická klasifikácia	Klasifikácia textových odpovedí respondentov	IN-02
Verejný AI analytik	AI analytika nad agregovanými štatistikami	IN-01, IN-05
Interný AI analytik	Analýza mikroúdajov pomocou AI	IN-02, IN-09
Syntetické dáta	Generovanie syntetických datasetov	IN-04
Inteligentné vyhľadávanie	Vyhľadávanie štatistických informácií	IN-06
NLG komentáre	Generovanie textových komentárov k dátam	IN-07
Viacjazyčné publikovanie	Automatizovaný preklad štatistických výstupov	IN-08
Verejný API	Poskytovanie agregovaných štatistických dát	IN-03

Tabuľka 15 Väzby scenárov použitia, aplikačných služieb a integračných bodov

Druhá tabuľka predstavuje mapovanie medzi aplikačnými službami platformy a konkrétnymi aplikačnými komponentmi, ktoré tieto služby implementujú v rámci architektúry systému. Účelom tejto tabuľky je zabezpečiť jednoznačné prepojenie medzi logickou architektúrou služieb a implementačnými komponentmi identifikovanými v aplikačnej vrstve architektúry. Každá aplikačná služba reprezentuje funkčnú schopnosť systému poskytovanú používateľom alebo iným systémom. Tieto služby sú implementované prostredníctvom konkrétnych aplikačných komponentov platformy označených identifikátormi CMP. Mapovanie medzi službami a komponentmi umožňuje lepšie pochopiť rozdelenie zodpovedností jednotlivých častí systému a zároveň slúži ako základ pre návrh implementácie, prevádzky a monitorovania platformy. Tento prehľad zároveň podporuje konzistentné prepojenie medzi jednotlivými vrstvami architektúry, keďže aplikačné služby reprezentujú logickú funkcionalitu systému, zatiaľ čo komponenty predstavujú konkrétne implementačné moduly, ktoré túto funkcionalitu realizujú.

Aplikačná služba	Komponent
Metodická podpora zamestnancov nad znalostnou bázou úradu	CMP-A14, CMP-A15
Klasifikácia textových odpovedí respondentov	CMP-A6
AI analytika nad agregovanými dátami	CMP-B4
Analýza mikroúdajov	CMP-A7
Generovanie syntetických dát	CMP-A8
Inteligentné vyhľadávanie	CMP-B5
Generovanie textových komentárov	CMP-B6
Prekladový modul	CMP-B7
Verejný API agregátov	CMP-B8

Tabuľka 16 Väzby aplikačných služieb a komponentov platformy

5.4.2 Rozsah informačných systémov – budúci stav (TO BE)

ISVS AI platforma pre analýzu a disemináciu štatistických údajov ŠÚ SR tvorí 11 modulov. V internej zóne sú umiestnené moduly Asistent pre zamestnancov, ktorý odpovedá zamestnancom nad znalostnou bázou úradu s uvedením zdroja, Klasifikácia textových odpovedí, ktorý zaraďuje textové odpovede do klasifikácií s mierou istoty a pracovným rozhraním kódovača, Analytik nad mikroúdajmi, ktorý generuje a vykonáva dotazy nad mikroúdajmi

s uplatnením pravidiel štatistického utajenia, Generovanie syntetických údajov, ktorý vytvára syntetické datasety s vyhodnotením zhody štatistických vlastností a rizika reidentifikácie, a Správa modelov a kalibrácia, ktorý eviduje modely a ich verzie, riadi ich nasadenie, vyhodnocovanie a kalibráciu. Vo verejnej zóne sú umiestnené moduly Verejný analytik, ktorý odpovedá na otázky nad publikovanými agregátmi s uvedením zdroja a zahŕňa aj verejné webové rozhranie AI služieb, Inteligentné vyhľadávanie, ktorý zabezpečuje sémantické vyhľadávanie a odporúčanie publikovaného obsahu vložené do webového sídla úradu, Automatické textové súhrny, ktorý pripravuje komentáre k tabuľkám a grafom so schválením pred zverejnením, a Viacjazyčné publikovanie, ktorý prekladá publikované výstupy a rozhrania s využitím terminologického slovníka a ľudskej revízie. V oboch zónach pôsobia moduly Znalostná a vyhľadávacia vrstva, ktorý spracúva dokumenty a vytvára fulltextové a vektorové indexy oddelené podľa zón, Kontrolovaný prestup medzi zónami, ktorý zabezpečuje jednosmerný prestup údajov z internej do verejnej zóny s automatizovanými kontrolami a schválením podľa pravidla štyroch očí.

KÓD ISVS (z metais)	NÁZOV ISVS	MODUL ISVS	STAV ISVS	TYP ISVS	KÓD NADRADENÉHO ISVS
isvs_15772	AI platforma pre analýzu a disemináciu štatistických údajov ŠÚ SR	#	Plánujem budovať	Agendový	
isvs_15773	Asistent pre zamestnancov	#	Plánujem budovať	Agendový	isvs_15772
isvs_15774	Klasifikácia textových odpovedí	#	Plánujem budovať	Agendový	isvs_15772
isvs_15775	Analytik nad mikroúdajmi	#	Plánujem budovať	Agendový	isvs_15772
isvs_15776	Generovanie syntetických údajov	#	Plánujem budovať	Agendový	isvs_15772
isvs_15777	Verejný analytik	#	Plánujem budovať	Agendový	isvs_15772
isvs_15778	Inteligentné vyhľadávanie	#	Plánujem budovať	Agendový	isvs_15772
isvs_15779	Automatické textové súhrny	#	Plánujem budovať	Agendový	isvs_15772
isvs_15780	Viacjazyčné publikovanie	#	Plánujem budovať	Agendový	isvs_15772
isvs_15781	Znalostná a vyhľadávacia vrstva	#	Plánujem budovať	Agendový	isvs_15772
isvs_15782	Správa modelov a kalibrácia	#	Plánujem budovať	Agendový	isvs_15772
isvs_15783	Kontrolovaný prestup medzi zónami	#	Plánujem budovať	Agendový	isvs_15772

Tabuľka 17 Rozsah informačných systémov - budúci stav (TO BE)

5.4.3 Využívanie nadrezortných a spoločných ISVS – AS IS

Nie je plánované.

5.4.4 Prehľad plánovaných integrácií na nadrezortné ISVS – spoločné moduly podľa zákona č. 305/2013 Z. z. o e-Governmente – budúci stav (TO BE)

Nie je plánované.

5.4.5 Prehľad plánovaných integrácií na iné ISVS – budúci stav (TO BE)

Nie je plánované. Projekt nebuduje integrácie na ISVS iných správcov. Napojenie platformy na interné zdroje údajov a publikačné kanály ŠÚ SR je súčasťou modulu Integrácie na systémy úradu a je opísané v kapitole Návrh riešenia v aplikačnej vrstve architektúry v časti Integrácie a dátové rozhrania.

5.4.6 Aplikačné služby pre Koncové služby – budúci stav (TO BE)

Kód AS (z MetaIS)	Názov AS	Realizuje ISVS (kód ISVS, ktorý realizuje AS)	Aplikačná služba slúži KS (kód KS z MetaIS)
as_69537	Poskytovanie verejného AI analytika nad publikovanými agregovanými štatistickými údajmi	AI platforma pre analýzu a disemináciu štatistických údajov ŠÚ SR	ks_382581
as_69538	Inteligentné vyhľadávanie a odporúčanie štatistického obsahu	AI platforma pre analýzu a disemináciu štatistických údajov ŠÚ SR	ks_382582
as_69539	Generovanie automatických textových súhrnov k publikovaným štatistickým údajom	AI platforma pre analýzu a disemináciu štatistických údajov ŠÚ SR	ks_382583
as_69540	Poskytovanie viacjazyčných verzií publikovaných štatistických informácií	AI platforma pre analýzu a disemináciu štatistických údajov ŠÚ SR	ks_382584
as_69541	Poskytovanie interného AI asistenta pre metodickú podporu zamestnancov	AI platforma pre analýzu a disemináciu štatistických údajov ŠÚ SR	ks_382577
as_69542	Automatizované kódovanie textových odpovedí respondentov	AI platforma pre analýzu a disemináciu štatistických údajov ŠÚ SR	ks_382578
as_69543	Poskytovanie bezpečného analytického rozhrania nad mikroúdajmi	AI platforma pre analýzu a disemináciu štatistických údajov ŠÚ SR	ks_382579
as_69544	Generovanie syntetických datasetov na analytické a testovacie účely	AI platforma pre analýzu a disemináciu štatistických údajov ŠÚ SR	ks_382580
as_69545	Poskytovanie verejného API pre publikované agregované štatistické údaje	AI platforma pre analýzu a disemináciu štatistických údajov ŠÚ SR	ks_382581
as_69546	Integrácia s portálom a CMS pre publikovanie analytických výstupov	AI platforma pre analýzu a disemináciu štatistických údajov ŠÚ SR	ks_382581

Tabuľka 18 Aplikačné služby pre Koncové služby – budúci stav (TO BE)

5.4.7 Aplikačné služby na integráciu – budúci stav (TO BE)

Nie je plánované.

5.5 Dátová architektúra

Dátová architektúra projektu je navrhnutá tak, aby bezpečne integrovala existujúce dátové zdroje Štatistického úradu SR (ŠÚ SR) do AI platformy a zároveň bola plne kompatibilná s metodikou Ministerstva vnútra SR pre identifikáciu, vizualizáciu a referencovanie údajov pri dátovom modelovaní vo verejnej správe. Jadrom návrhu je jednotný rámec objektov evidencie, referenčných väzieb a životného cyklu údajov, doplnený o systematický manažment dát (data governance), ktorý umožňuje konzistentné modelovanie, auditovateľnosť, opakovateľnosť a bezpečné sprístupňovanie údajov v súlade so strategickými prioritami „Manažment údajov“ a „Otvorené údaje“.

Dátová architektúra zároveň rešpektuje zónový bezpečnostný model platformy – citlivé mikroúdaje zostávajú výhradne v chránenej Zóne A, zatiaľ čo Zóna B pracuje len s publikovanými agregátmi a verejným obsahom.

V stave AS IS ŠÚ SR pracuje s uceleným reťazcom údajov od zberu a validácie cez spracovanie až po disemináciu, pričom z pohľadu logického modelu dominujú tri kľúčové kategórie údajov: mikroúdaje, referenčné klasifikácie (číselníky) a publikované agregáty. Mikroúdaje referencujú klasifikácie ako záväzné referenčné údaje a tvoria základ pre tvorbu interných agregátov a publikovaných výstupov. Tento vzťahový rámec je vhodný na formálne zachytenie v logickom modeli (diagram tried/ER) v súlade s metodikou MV SR, pričom každý objekt evidencie má jednoznačný identifikátor, definované atribúty, väzby na ďalšie objekty a väzbu na procesy a systémy, ktoré s ním pracujú.

Zber a vznik údajov začína v zberových systémoch a napojených registroch v chránenej zóne. Vstupné dáta majú charakter formulárových odpovedí a transakčných záznamov, ktoré obsahujú kombináciu štruktúrovaných položiek a voľných textových odpovedí určených na kódovanie. V ďalšej fáze prechádzajú validáciou, čistením a kuráciou, aby vznikli mikroúdaje ako základný prvok štatistického spracovania. Mikroúdaje sú vedené v mikroúdajovej databáze a sú naviazané na ukazovatele, referenčné obdobia, územné a tematické dimenzie a referenčné kódy klasifikácií (napr. SK NACE, ISCO a ďalšie relevantné číselníky). Klasifikácie a číselníky pritom plnia úlohu referenčných údajov – majú definovanú hierarchiu, verziu, platnosť a vlastníctvo, sú spravované ako riadené objekty evidencie a využívajú sa nielen pri spracovaní a agregácii, ale aj pri automatickej klasifikácii textových odpovedí v rámci AI scenárov.

Nad mikroúdajmi sa vytvárajú interné agregáty využívané na kontrolu kvality, metodické overovanie a interné analýzy. Publikované agregáty a dátové kocky sú následne odvodené buď priamo z mikroúdajov alebo cez interné agregáty, vždy však až po povinnom uplatnení pravidiel štatistického utajenia a posúdení rizika reidentifikácie. Tento krok je koncepčne zakotvený ako bezpečnostno-metodický „gate“ medzi Zónou A a Zónou B: do Zóny B prechádzajú iba publikované agregované údaje, dátové kocky, metaúdaje a schválené syntetické datasety, pričom prenos je jednosmerný, kontrolovaný a auditovaný. Metadáta publikovaných agregátov obsahujú okrem definície ukazovateľa a zdrojov aj informácie o aplikovaných pravidlách utajenia, verziovaní datasetu a konfiguračných parametroch, ktoré sú potrebné pre transparentnosť a dôveryhodnosť diseminácie.

Riadenie životného cyklu údajov je v návrhu pokryté end-to-end, od tvorby až po archiváciu. Pre každý typ datasetu sa definujú fázy spracovania (zber, kúracia, analýza, publikovanie, archivácia, vyradenie) vrátane retenčných lehôt a pravidiel prístupu. Tieto informácie sa evidujú v dátovom katalógu a uplatňujú sa aj v prevádzkových mechanizmoch zálohovania, obnovy a kontinuity. Osobitná pozornosť sa venuje rodokmeňu údajov (lineage): platforma eviduje pôvod dát, transformačné kroky, použité metodiky, verzie a zodpovednosti, aby bolo možné spätne overiť, ako konkrétny publikovaný výstup vznikol, aké pravidlá sa naň uplatnili a kto ho schválil.

Dátová architektúra je zároveň navrhnutá ako „governance-driven“. Katalóg dát a metadát je centrálny komponent, ktorý zabezpečuje jednotnú evidenciu objektov evidencie, číselníkov, datasetov a rozhraní, vrátane prepojenia na procesy a aplikačné služby. Katalóg podporuje klasifikáciu údajov (mikroúdaje, interné agregáty, publikované agregáty), evidenciu právnych a metodických obmedzení, vlastníkov údajov (Data Owners) a správcov metadát, ako aj verziovanie a referencovateľnosť pre technické špecifikácie a API kontrakty. Výsledkom je dátová vrstva, ktorá je konzistentná naprieč systémami a zároveň pripravená na automatizovanú validáciu štruktúr a rozhraní.

Bezpečnostné a compliance mechanizmy sú vložené priamo do dátových tokov. Pri prestupoch medzi zónami sa uplatňuje kontrolovaný egress workflow s DLP kontrolami, šifrovaním, podpisovaním dátových balíkov a schvaľovaním podľa princípu „four-eyes“. Pravidlá prístupu sa opierajú o RBAC/ABAC, pričom citlivé identifikátory respondentov a mikroúdaje ostávajú striktné v Zóne A. Auditné záznamy o prístupoch k údajom, exportoch, zmenách konfigurácií a publikovaní výstupov sa centralizujú do log manažmentu/SIEM a slúžia pre periodické kontroly súladu, forenznú analýzu aj preukázanie napĺňania požiadaviek GDPR, bezpečnostných štandardov a interných politík.

Organizačné zabezpečenie dátovej architektúry je v návrhu riešené explicitne, aby sa správa údajov nestala iba technickou disciplínou, ale riadeným procesom. Projekt predpokladá zavedenie roly dátového kurátora (resp. dátového architekta), ktorý je zodpovedný za kvalitu a konzistentnosť metadát, katalogizáciu, referencovanie objektov evidencie a zosúladenie návrhu s metodikou MV SR. Súčasne sa odporúča vytvorenie rezortnej

dátovej kancelárie ako koordinačného rámca medzi štatistickou metodikou, dátovou správou, IT bezpečnosťou a prevádzkou. Tento rámec zabezpečuje segregáciu povinností (Separation of Duties), stanovuje politiky klasifikácie údajov, riadi procesy publikovania otvorených údajov a dohliada na to, aby žiadna rola nebola zároveň konfiguratorom aj schvaľovateľom citlivých operácií nad údajmi.

Takto koncipovaná dátová architektúra spája logický model objektov evidencie, riadenie životného cyklu údajov, metadátovú a bezpečnostnú governance a organizačné roly do jednotného, auditovateľného a opakovateľného rámca. Výsledkom je dátová vrstva, ktorá umožňuje bezpečné využitie mikroúdajov pre interné AI scenáre, spoľahlivú a transparentnú tvorbu publikovaných agregátov a modernú disemináciu štatistických výstupov pre verejnosť bez narušenia štatistickej dôveryhodnosti, rezidencie dát a regulačných požiadaviek.

5.5.1 Objekty evidencie

V rámci dátovej architektúry AI platformy Štatistického úradu SR je potrebné jednoznačne definovať objekty evidencie (OE), ktoré predstavujú základné dátové entity spracúvané, transformované alebo publikované v rámci riešenia. Identifikácia OE je kľúčová pre zabezpečenie interoperability, referencovania údajov a štandardizovanej výmeny dát medzi existujúcimi informačnými systémami ŠÚ SR, novobudovanou AI platformou a nadväzujúcimi publikovanými rozhraniami. Objekty evidencie zároveň poskytujú stabilnú „dátovú kostru“ projektu, na ktorú sa následne viažu integračné rozhrania, katalóg požiadaviek, definícia dátových tokov, pravidlá prístupu a bezpečnostné opatrenia podľa princípov security by design a privacy by default. Definované OE vychádzajú z reálneho životného cyklu údajov v prostredí oficiálnej štátnej štatistiky – od prijatia údajov zo zberu cez spracovanie mikroúdajov a aplikáciu klasifikácií až po tvorbu agregovaných výstupov určených na publikovanie. Zámerom nie je vytvoriť úplný detailný dátový model všetkých existujúcich systémov, ale vytvoriť doménový model, ktorý zrozumiteľne zachytí rozsah údajov dotknutých projektom a zároveň umožní abstrahovať od nepodstatných technických detailov. Doménový model je definovaný ako platformovo nezávislý návrh dátových prvkov súvisiacich s projektom a slúži ako podklad pre návrh aplikačných služieb, integračných bodov a API kontraktov.

Pri definovaní OE sa zohľadňujú štandardy výmeny údajov medzi informačnými systémami verejnej správy podľa § 30 vyhlášky č. 78/2020 Z. z. o štandardoch pre ITVS a zároveň usmernenia k interoperabilite. V prípadoch, kde je to možné, sú objekty evidencie naviazané na existujúce dátové prvky Centrálného modelu údajov verejnej správy a sú uvádzané referencovateľnými URI identifikátormi v skrátenej forme (napr. pper:PhysicalPerson). Takýto prístup znižuje riziko nekompatibilných alebo duplicitných dátových definícií, zjednodušuje budúce integračné väzby v rámci ekosystému verejnej správy a podporuje konzistentné katalogizovanie údajov. Výber konkrétnych objektov evidencie je zdôvodnený tým, že predstavujú minimálnu, no dostatočnú množinu dátových entít, ktoré sú nevyhnutné pre realizáciu hlavných scenárov použitia AI platformy (automatizované kódovanie a klasifikácia, bezpečná interná analýza, tvorba agregátov a publikovanie, sprístupnenie údajov cez dátové kocky a API). Zároveň reflektujú bezpečnostnú architektúru riešenia: objekty obsahujúce mikroúdaje a identifikátory respondentov sú spracúvané výhradne v chránenej Zóne A, kým publikované agregované údaje a metaúdaje sú sprístupňované v Zóne B. Rozdelenie OE podľa dátovej citlivosti umožňuje uplatniť jednoznačné pravidlá prístupu (RBAC/ABAC), DLP a auditovateľnosť operácií, čím sa podporí súlad s GDPR, štatistickou dôveryhodnosťou a bezpečnostnými požiadavkami informačných systémov verejnej správy. Na základe vyššie uvedených princípov je v nasledujúcej tabuľke uvedený zoznam objektov evidencie relevantných pre projekt spolu s ich popisom a referencovateľným URI identifikátorom, ak je pre daný dátový prvok dostupný v Centrálnom modeli údajov verejnej správy.

ID OE	Objekt evidencie - názov	Objekt evidencie - popis	Referencovateľný identifikátor URI dátového prvku
OE_01	Respondent	Subjekt poskytujúci údaje v rámci štatistických zisťovaní. Môže ísť o domácnosť, fyzickú osobu, podnik alebo inú inštitúciu. Objekt evidencie obsahuje identifikátor respondenta, typ subjektu, sektor alebo odvetvie a územnú príslušnosť. Identifikátory respondentov sa používajú výlučne v chránenej časti systému (Zóna A).	ppper:PhysicalPerson / plegal:LegalPerson

OE_02	Zberový formulár / odpoveď	Záznam odpovede respondenta v rámci konkrétneho štatistického zisťovania. Obsahuje identifikátor formulára, referenciu na respondenta, časový údaj, textové odpovede, číselné hodnoty a indikátory kvality záznamu. Tento objekt predstavuje vstupný údaj pre proces spracovania dát.	Nemá
OE_03	Mikroúdaj	Základná jednotka štatistického spracovania predstavujúca jednotlivý záznam údajov po validácii a kurácii. Mikroúdaj obsahuje identifikátor záznamu, ukazovateľ, hodnotu, referenčné obdobie, klasifikačné kódy a príznaky kvality. Mikroúdaje sú spracúvané výhradne v chránenej infraštruktúre ŠÚ SR.	Nemá
OE_04	Klasifikácia / číselník	Referenčný údaj používaný pri kódovaní údajov a analytických operáciách (napr. SK NACE, ISCO). Obsahuje kód klasifikácie, názov položky, hierarchickú štruktúru, verziu klasifikácie a obdobie platnosti. Používa sa ako referenčný rámec pri klasifikácii odpovedí a tvorbe agregovaných údajov.	Nemá
OE_05	Interný agregát	Agregovaný údaj odvodený z mikroúdajov, ktorý sa používa na interné analytické účely a kontrolu kvality štatistických výstupov. Interné agregáty nie sú určené na publikovanie a sú prístupné iba autorizovaným pracovníkom ŠÚ SR.	Nemá
OE_06	Publikovaný agregát / dátová kocka	Agregovaný dataset určený na publikovanie pre verejnosť. Obsahuje ukazovateľ, dimenzie (čas, územie, odvetvie a ďalšie), agregovanú hodnotu a metaúdaje o zdroji a metodike výpočtu. Pred publikovaním sa na tieto údaje uplatňujú pravidlá štatistického utajenia.	Nemá
OE_07	Dataset	Logická kolekcia údajov publikovaná prostredníctvom dátového portálu alebo API. Dataset obsahuje identifikátor datasetu, názov, opis, verziu, zdroj a metaúdaje potrebné pre jeho interpretáciu.	dcat:Dataset
OE_08	Metaúdaje datasetu	Popisné informácie o datasete alebo údajoch (napr. definícia	

ukazovateľ a, metodika,
zdroj údajov, referenčné
obdobie, verzia datasetu).
Slúžia na zabezpečenie
interpretovateľnosti a
interoperabilného zdieľania
údajov.

Tabuľka 19 Objekty evidencie

5.5.2 Referenčné údaje

Nie je plánované.

5.5.3 Poskytovanie údajov z ISVS do IS CPDI – budúci stav (TO BE)

Nie je plánované.

5.5.4 Konzumovanie údajov z IS CPDI – budúci stav (TO BE)

Nie je plánované.

5.5.5 Identifikácia údajov a subjektov pre konzumovanie alebo poskytovanie údajov do/z CPDI (CSRÚ)

Nie je plánované.

5.5.6 Kvalita a čistenie údajov

Kvalita údajov predstavuje kritický faktor pre správne fungovanie navrhovanej AI platformy Štatistického úradu SR. Keďže riešenie pracuje s mikroúdajmi, klasifikačnými údajmi a agregovanými datasetmi, nesprávna alebo neúplná hodnota údaje môže mať priamy dopad na kvalitu analytických výstupov, štatistických agregátov a následne aj na interpretáciu údajov verejnosťou. Nekvalitné údaje môžu viesť napríklad k nesprávnemu zaradeniu odpovedí respondentov do klasifikácií, k chybnému výpočtu agregovaných ukazovateľov alebo k nesprávnemu publikovaniu štatistických údajov. V navrhovanom riešení budú preto implementované mechanizmy priebežnej kontroly, validácie a čistenia údajov počas celého ich životného cyklu – od prijatia údajov zo zberových systémov až po publikovanie agregovaných datasetov. Validácia údajov bude realizovaná na viacerých úrovniach. Pri prijímaní údajov sa budú kontrolovať základné formálne vlastnosti údajov, ako je formát, rozsah hodnôt, povinné atribúty alebo logická konzistentnosť. V prípade údajov, ktoré využívajú referenčné klasifikácie alebo číselníky (napr. klasifikácie odvetví alebo povolání), bude implementované obmedzenie hodnôt prostredníctvom kontrolovaných číselníkov, čím sa minimalizuje riziko nekonzistentných alebo neplatných hodnôt.

Riešenie zároveň umožní overovanie hodnôt údajov prostredníctvom validačných pravidiel, kontrol kvality a automatizovaných analytických kontrol. Pri spracovaní údajov budú využívané aj metódy identifikácie anomálií a kontroly úplnosti datasetov. V prípade identifikácie nekonzistentných údajov bude možné tieto údaje označiť na ďalšie manuálne overenie alebo ich spracovať prostredníctvom pravidiel čistenia údajov. Projekt zároveň predpokladá migráciu časti údajov z existujúcich informačných systémov ŠÚ SR, najmä v oblasti existujúcich agregovaných datasetov a metaúdajov. Pri migrácii údajov budú aplikované kontrolné mechanizmy zamerané na identifikáciu duplícit, neúplných záznamov a nesprávnych formátov údajov. Proces migrácie bude zahŕňať aj transformáciu dátových štruktúr a validáciu konzistentnosti údajov voči referenčným klasifikáciám. Na zabezpečenie systematického riadenia dátovej kvality bude zavedený proces pravidelného merania kvality údajov pre jednotlivé objekty evidencie. V nasledujúcej tabuľke je uvedené vyhodnotenie významnosti kvality údajov pre jednotlivé objekty evidencie spolu s prioritou ich monitorovania a čistenia.

ID OE	Názov Objektu evidencie (uvádzať OE z tabuľky v kap. 5.5.1)	Významnosť kvality 1 (malá) až 5 (veľmi významná)	Citlivosť kvality 1 (malá) až 5 (veľmi významná)	Priorita – poradie dôležitosti (začnite číslovať od najdôležitejšieho)
-------	--	--	---	---

OE_01	Respondent	5	5	1
OE_02	Zberový formulár / odpoveď	5	5	2
OE_03	Mikroúdaj	5	5	3
OE_04	Klasifikácia / číselník	5	4	4
OE_05	Interný agregát	4	4	5
OE_06	Publikovaný agregát / dátová kocka	4	3	6
OE_07	Dataset	3	3	7
OE_08	Metaúdaje datasetu	3	2	8

Tabuľka 20 Zhodnotenie dátovej kvality objektov evidencie

5.5.7 Otvorené údaje

Projekt sprístupňuje publikované agregované štatistické údaje a metaúdaje prostredníctvom verejného API (integračný bod IN-03) a schválené syntetické datasety pre výskumné účely.

5.5.8 Analytické údaje

Nebudú poskytované.

5.5.9 Moje údaje

Nebudú poskytované.

5.5.10 Prehľad jednotlivých kategórií údajov

ID	Register / Objekt evidencie	Referenčné údaje	Moje údaje	Otvorené údaje	Analytické údaje
	(uvádzať OE z tabuľky v kap. 5.5.1)				
OE_01	Respondent	#	#	#	#
OE_02	Zberový formulár / odpoveď	#	#	#	#
OE_03	Mikroúdaj	#	#	#	#
OE_04	Klasifikácia / číselník	#	#	#	#
OE_05	Interný agregát	#	#	#	#
OE_06	Publikovaný agregát / dátová kocka	#	#	#	#
OE_07	Dataset	#	#	#	#
OE_08	Metaúdaje datasetu	#	#	#	#

Tabuľka 22 Prehľad jednotlivých kategórií údajov

5.6 Technologická architektúra

5.6.1 Návrh riešenia technologickej architektúry

Technologická vrstva Štatistického úradu Slovenskej republiky je v súčasnosti postavená predovšetkým na tradičnej infraštruktúre prevádzkovej v dátovom centre úradu. Táto infraštruktúra je historicky navrhnutá na

podporu štandardných transakčných informačných systémov, databázových riešení a analytických nástrojov určených pre klasické spracovanie štatistických údajov. V takomto prostredí je zabezpečená stabilná prevádzka existujúcich štatistických agend, spracovanie zisťovaní a publikovanie štatistických výstupov.

Rozvoj moderných analytických metód založených na umelej inteligencii však prináša nové technologické požiadavky, ktoré súčasná infraštruktúra pokrýva len čiastočne. Jedným z kľúčových obmedzení je absencia špecializovaného výpočtového výkonu určeného pre spracovanie a tréning modelov umelej inteligencie. Moderné AI a ML algoritmy, najmä modely založené na neurónových sieťach alebo veľké jazykové modely, vyžadujú infraštruktúru s akcelеровaným výpočtovým výkonom, najmä prostredníctvom grafických procesorov alebo podobných výpočtových akceleratorov.

Ďalšou významnou výzvou je nízka miera automatizácie niektorých analytických a metodických procesov. Viaceré úlohy, ako napríklad klasifikácia otvorených textových odpovedí zo štatistických zisťovaní alebo príprava textových interpretácií štatistických výstupov, sú dnes do veľkej miery založené na manuálnom spracovaní. Takýto prístup predstavuje časovú záťaž pre odborných pracovníkov a zároveň obmedzuje schopnosť rýchlo reagovať na nové analytické požiadavky. Súčasná technologická vrstva zároveň neposkytuje komplexnú platformu, ktorá by umožňovala bezpečnú experimentáciu s modernými analytickými nástrojmi, napríklad generovanie syntetických mikrodatabáz alebo interaktívnu analytickú prácu nad citlivými mikrodatabázami v kontrolovanom prostredí. Chýba najmä technologický rámec, ktorý by umožňoval kombinovať moderné AI nástroje s prísnymi požiadavkami na ochranu štatistickej dôveryhodnosti a bezpečné spracovanie citlivých údajov. Tieto technologické obmedzenia predstavujú významný faktor, ktorý limituje možnosti širšieho využívania umelej inteligencie v prostredí oficiálnej štatistiky a zároveň vytvára potrebu systematickej modernizácie technologickej infraštruktúry.

Navrhované riešenie technologickej architektúry vychádza z cieľa vybudovať modernú, bezpečnú a škálovateľnú platformu, ktorá umožní využívanie umelej inteligencie pri spracovaní štatistických údajov a zároveň bude rešpektovať všetky požiadavky na ochranu citlivých dát. Architektúra platformy je navrhnutá na princípoch technologickej modularity, škálovateľnosti a nezávislosti od konkrétneho technologického prostredia. Riešenie je koncipované tak, aby bolo možné jednotlivé komponenty platformy prevádzkovať v rôznych infraštruktúrnych prostrediach a aby bolo možné platformu postupne rozširovať o nové analytické alebo disemináčnе služby. Kľúčovým architektonickým princípom je využitie zónového modelu technologickej infraštruktúry. Platforma bude rozdelená na dve logické bezpečnostné zóny. Prvá zóna predstavuje interné chránené prostredie určené na spracovanie citlivých mikrodatabáz a na prevádzku analytických nástrojov umelej inteligencie, ktoré s týmito údajmi pracujú. Druhá zóna je určená pre komponenty poskytujúce služby nad publikovanými agregovanými údajmi a verejným obsahom, napríklad pre verejné analytické nástroje, inteligentné vyhľadávanie alebo automatické generovanie textových interpretácií štatistických údajov.

Dôležitým prvkom architektúry je striktné kontrolovaný prenos údajov medzi týmito zónami. Prenos je povolený výhradne v smere z internej zóny do verejnej zóny a iba pre publikované agregované údaje, dátové kocky, metaúdaje a schválené syntetické databázy. Každý takýto prenos bude podliehať bezpečnostným kontrolám, auditovaniu a mechanizmom prevencie úniku dát. Technologická architektúra zároveň predpokladá využitie moderných prístupov k prevádzke softvérových systémov, najmä kontajnerových technológií a automatizovaných nástrojov na správu aplikácií a analytických komponentov. Takýto prístup umožní efektívnejšie škálovanie jednotlivých častí platformy, jednoduchšie nasadzovanie nových analytických modelov a lepšiu správu celého technologického prostredia.

Požiadavky na prevádzkové prostredia

Navrhovaná technologická platforma bude prevádzkovaná vo viacerých oddelených prostrediach, ktoré zabezpečia bezpečný vývoj, testovanie a produkčnú prevádzku jednotlivých komponentov. Produkčné prostredie bude prevádzkované v dátovom centre Štatistického úradu SR a bude predstavovať jediný technologický priestor, v ktorom bude povolené spracovanie reálnych citlivých mikrodatabáz. V tomto prostredí budú prevádzkované interné analytické nástroje umelej inteligencie, systémy správy modelov a ďalšie komponenty platformy pracujúce s citlivými údajmi. Vývojové a testovacie prostredia budú od produkčného prostredia striktné oddelené a budú využívať výhradne anonymizované, syntetické alebo publikované agregované údaje. Takýto prístup umožní bezpečné testovanie nových analytických nástrojov bez rizika spracovania citlivých údajov mimo produkčného prostredia. V úvodnej fáze projektu je zároveň možné využitie pilotného prostredia vo verejnom cloude v rámci Európskej únie, ktoré bude slúžiť na overenie konceptu vybraných verejných služieb v izolovanom prostredí

verejného cloudu v rámci EÚ nad necitlivými, syntetickými a verejnými údajmi. Toto prostredie bude prevádzkované v úplnej sieťovej izolácii a nebude obsahovať citlivé mikroúdaje ani prepojenie na internú infraštruktúru úradu.

Model technologickej infraštruktúry

Technologická infraštruktúra platformy bude navrhnutá tak, aby priamo podporovala jednotlivé aplikačné komponenty umelej inteligencie a analytických nástrojov. V chránenej internej zóne budú prevádzkované výpočtové uzly určené pre analytické úlohy, spracovanie mikroúdajov a tréning modelov umelej inteligencie. Tieto uzly budú poskytovať výpočtový výkon potrebný pre moderné AI modely a zároveň budú integrované s internými dátovými zdrojmi. Vo verejnej zóne budú prevádzkované komponenty zabezpečujúce interakciu s verejnosťou, napríklad analytické služby nad publikovanými agregovanými údajmi, vyhľadávacie nástroje alebo rozhrania pre integráciu s externými systémami. Tieto komponenty budú optimalizované najmä na rýchlu odozvu a škálovateľnosť pri poskytovaní verejných služieb. Komunikácia medzi jednotlivými komponentmi platformy bude zabezpečená prostredníctvom bezpečných integračných mechanizmov a bude riadená podľa princípu minimálnych oprávnení. Všetky prístupy k dátam, prenosy údajov a interakcie medzi komponentmi budú auditované a monitorované.

Zdôvodnenie umiestnenia platformy

Produkčné spracovanie citlivých mikroúdajov bude realizované výlučne v lokálnej infraštruktúre dátového centra Štatistického úradu SR. Toto rozhodnutie vychádza z požiadaviek na ochranu štatistickej dôvernosti, z legislatívnych povinností v oblasti ochrany osobných údajov a zo zásady rezidencie citlivých údajov v prostredí pod plnou kontrolou úradu. Priame prepojenie citlivých dátových prostredí s externými cloudovými infraštruktúrami nie je z bezpečnostných dôvodov povolené. Verejný cloud v rámci EÚ sa preto využíva iba pre izolované pilotné prostredie, ktoré slúži na overenie konceptu vybraných verejných služieb v izolovanom prostredí verejného cloudu v rámci EÚ nad necitlivými, syntetickými a verejnými údajmi. Produkčná prevádzka internej aj verejnej zóny je v dátovom centre ŠÚ SR. Takto navrhnutá technologická architektúra umožní bezpečné využívanie umelej inteligencie pri spracovaní štatistických údajov, podporí modernizáciu analytických procesov a zároveň zabezpečí plný súlad s požiadavkami na ochranu citlivých údajov a štatistickej dôvernosti.

5.6.2 Požiadavky na výkonnostné parametre, kapacitné požiadavky – budúci stav (TO BE)

Parameter	Jednotky	Predpokladaná hodnota	Poznámka
Počet interných používateľov	Počet	Všetci zamestnanci ŠÚ SR	Zahrňa štatistov, metodikov, analytikov, IT podporu a novoprijatých zamestnancov.
Počet súčasne pracujúcich interných používateľov v špičke	Počet	Limitované priepustnosťou interných API	Dimenzovanie na úrovni 3 – 4 worker uzlov v Zóne A pre interné služby.
Počet externých používateľov (internet)	Počet	Nereštriktívny prístup	Verejnosť, médiá a akademická obec prístupujúca k publikovaným dátam.
Počet externých používateľov v špičkovom zaťažení	Počet	Špičky návštevnosti portálu	Škálovanie je horizontálne (5 – 7 uzlov) podľa error budgetu a SLO.
Počet transakcií (požiadaviek) za obdobie	Počet/klient	100 požiadaviek / min	Štandardná kvóta na klienta, burst limit 10 požiadaviek/s.
Objem údajov na transakciu	Objem	Limitované politikou API	Definované maximálne veľkosti payloadov pre stabilitu systému.
Objem existujúcich kmeňových dát	Objem	100 – 200 GB	Sizing pre indexy a artefakty v úvodných fázach, nárast na TB pri plnej prevádzke.
Latencia verejných služieb (P90)	Čas	≤ 2 s	Cieľové SLO pre bežné verejné dotazy v Zóne B.
Latencia verejných služieb (P99)	Čas	≤ 5 s	Cieľové SLO pre zložité dotazy (viac filtrov, dlhšie časové rady).

Dostupnosť kľúčových služieb	Percento	≥ 98,5 %	Mesačná úroveň dostupnosti pre kľúčové B-komponenty a API bránu.
Retencia auditných logov	Mesiac	24 – 36 mesiacov	Doba uchovávanía auditných udalostí v SIEM.
Cieľový čas obnovy (RTO)	Hodina	4 hodiny	Pre verejné služby v Zóne B.
Maximálna strata dát (RPO)	Hodina	6 hodín	Závisí od publikačných cyklov a typu dát.

Tabuľka 23 Požiadavky na výkonnostné parametre, kapacitné požiadavky – budúci stav (TO BE)

Doplňujúce informácie k výkonu a sizingu:

- Verejné služby v Zóne B sú navrhnuté na horizontálne škálovanie pomocou orchestrácie kontajnerov, zatiaľ čo v Zóne A sa uplatňuje kombinácia vertikálneho a horizontálneho škálovania podľa náročnosti tréningu a inferencie modelov.
- Systém využíva striktnú sieťovú segmentáciu (L2/L3 segmenty) s oddelenými doménami pre Zónu A a Zónu B. Prestup dát A → B je technicky vynútený cez DLP/egress brány na úrovni aplikačnej vrstvy (L7).
- Na zníženie latencie sa v Zóne B využíva viacúrovňová cache (vyrovnávacia pamäť pre kľúčové výsledky a reverzný proxy server s vyrovnávacou pamäťou pre statický obsah) a read-replica topológie databáz.
- Výkonnostné parametre sú sledované v reálnom čase (monitoring a vizualizácia metrík) s alertami naviazanými na porušenie definovaných error budgetov.

5.6.3 Využívanie služieb z katalógu služieb vládneho cloudu

V rámci projektu sa neplánuje využitie vládneho cloudu. Zdôvodnenie je uvedené v kapitole Stanovenie alternatív v technologickej vrstve architektúry.

5.7 Bezpečnostná architektúra

5.7.1 Návrh riešenia bezpečnosti

Súčasná bezpečnostná architektúra sa opiera o tradičné on-premise spracovanie v dátovom centre Štatistického úradu SR (ŠÚ SR) s dôrazom na fyzickú a logickú izoláciu citlivých systémov. Spracovanie mikroúdajov prebieha v uzavretom internom prostredí, pričom publikovanie štatistických agregátov na webový portál je realizované manuálnymi alebo poloaufomatizovanými procesmi. Chýba však jednotná AI platforma, ktorá by integrovala moderné metódy automatizovaného štatistického utajenia (SDC) a riadené prestupy dát medzi zónami v reálnom čase. Hlavným rozdielom oproti budúcemu stavu je absencia automatizovanej DLP/egress kontroly a centralizovanej správy modelov (MLOps), ktoré sú nevyhnutné pre bezpečnú prevádzku umelej inteligencie.

Navrhované riešenie zavádza robustnú architektúru postavenú na princípoch security-by-design a privacy-by-default. Základom je dvojzónový model v rámci infraštruktúry ŠÚ SR:

- Zóna A (Interná chránená zóna): Hostí citlivé mikroúdaje a všetky interné AI/ML pracovné zaťaženia. Nemá externú konektivitu pre citlivý obsah a prístupy sú striktné auditované.
- Zóna B (Verejná zóna): Obsahuje komponenty pracujúce výhradne s publikovanými agregovanými údajmi, dátovými kockami, metaúdajmi, schválenými syntetickými datasetmi a verejným obsahom. Je izolovaná od Zóny A a chránená pomocou WAF a rate-limitingu.

Postupy na dosiahnutie bezpečnosti a ochrana aktív

Na zabezpečenie aktív projektu a dosiahnutie požadovanej úrovne bezpečnosti sa uplatňujú nasledovné postupy:

- Riadenie prístupu: Implementácia kombinácie RBAC (role-based) a ABAC (attribute-based) s povinným viacfaktorovým overením (MFA) pre prístup k citlivým údajom v Zóne A.
- Ochrana dát: Povinné šifrovanie v pokoji aj pri prenose (TLS 1.2+, mTLS pre interné kanály) s využitím centralizovanej správy kľúčov (KMS) a hardvérových modulov (HSM).

- Kontrolované prestupy (Egress): Zo Zóny A do Zóny B smú prechádzať len publikované agregované údaje, dátové kocky, metaúdaje a schválené syntetické datasety a všetky podliehajú automatizovaným predpublikačným testom rizika reidentifikácie, DLP kontrole a schvaľovaniu podľa princípu „štyroch očí“.
- Štatistické utajenie: Vynútenie minimálnych veľkostí buniek, potláčanie nízkych počtov a zaokrúhľovanie v rámci analytických modulov.
- Zodpovedné AI: Povinná ľudská kontrola (Human-in-the-loop - HIL) pred publikovaním akýchkoľvek výstupov generovaných AI do verejnej zóny.

Zabezpečenie na vrstvách architektúry

1. Dátová vrstva: Striktná segmentácia úložísk. Mikroúdaje v Zóne A sú šifrované na úrovni databázy (TDE) a izolované od verejných agregátov v Zóne B.
2. AI/ML vrstva: Modely sú registrované v registri modelov v Zóne A, kde sa monitoruje ich verzia a reprodukovateľnosť. Na výstupy analytika v Zóne A sa uplatňujú kontroly prahových hodnôt štatistického utajenia.
3. Aplikačná vrstva: Použitie API brán (CMP-A12 a CMP-B8), ktoré vynucujú autentizáciu, validáciu schém a ochranu proti DoS útokom.
4. Prezentačná vrstva: Interné konzoly sú prístupné len cez zabezpečený bastion jump host s auditom príkazov.

Aspekty bezpečnosti (CIA triáda)

- Dôvernosť: Zabezpečená rezidenciou citlivých dát výlučne v Zóne A na území SR, šifrovaním a riadením prístupov podľa najmenších potrebných oprávnení.
- Integrita: Dosiahnutá pomocou kryptografických kontrolných súčtov a podpisov balíkov pri prestupoch medzi zónami, ako aj sledovateľnosťou rodokmeňa (lineage) dát a modelov.
- Dostupnosť: Zaručená redundantným nasadením v klastroch orchestrácie kontajnerov, aktívno-aktívnu konfiguráciu API brán a replikáciou databáz s automatickým failoverom.

Návrh riešenia reflektuje legislatívne požiadavky GDPR, nariadenia (EÚ) 2024/1689 (akt o umelej inteligencii) a smernice NIS2 a je v súlade s medzinárodným rámcom AIML4OS.

5.7.2 Určenie obsahu bezpečnostných opatrení

Obsah bezpečnostných opatrení podľa vyhlášky ÚPVII č. 179/2020 Z. z	Aplikované opatrenia	Aplikovaná legislatíva
Minimálne bezpečnostné opatrenia Kategórie III	Architektúra riešenia je navrhnutá ako bezpečná AI analytická platforma prevádzkovaná v infraštruktúre dátového centra Štatistického úradu SR. Riešenie využíva dvojzónový bezpečnostný model, ktorý striktno oddeľuje spracovanie citlivých mikroúdajov od verejne prístupných analytických služieb. Interná zóna je určená na spracovanie mikroúdajov a tréning analytických modelov, pričom prístup k týmto údajom je obmedzený na autorizovaných používateľov a kontrolovaný prostredníctvom viacfaktorovej autentifikácie a riadenia prístupových práv. Verejná zóna poskytuje analytické služby nad publikovanými agregovanými údajmi a neumožňuje priamy prístup k citlivým dátam. Platforma zabezpečuje centralizované riadenie prístupov, auditovateľnosť všetkých operácií, monitorovanie bezpečnostných udalostí, evidenciu incidentov a integráciu s existujúcimi bezpečnostnými nástrojmi organizácie. Súčasťou riešenia je aj ochrana dátových tokov medzi jednotlivými komponentmi, šifrovanie komunikácie, správa kryptografických kľúčov a podpora bezpečnostného	§ 3 ods. 3 písm. a) až d) vyhlášky 179/2020 Z. z.; Vyhláška MIRRI č. 401/2023 Z. z. o riadení projektov

	monitoringu a forenznej analýzy incidentov.	
Bezpečnostný projekt	Pre navrhované riešenie bude vypracovaný bezpečnostný projekt informačného systému verejnej správy v súlade s požiadavkami zákona o IT vo verejnej správe. Bezpečnostný projekt bude obsahovať kategorizáciu systému, analýzu rizík, definíciu bezpečnostnej architektúry, opis bezpečnostných opatrení a postupy riadenia bezpečnostných incidentov. Súčasťou dokumentácie bude aj definovanie bezpečnostných rolí, pravidiel správy prístupov, ochrany údajov a postupov bezpečnej prevádzky systému.	§ 23 ods. 1 a 2 zákona č. 95/2019 Z. z.
Bezpečnostné opatrenia podľa osobitného predpisu	Riešenie rešpektuje požiadavky právnych predpisov týkajúcich sa ochrany osobných údajov a spracovania dát vo verejnej správe. Pri spracovaní údajov sa uplatňujú princípy privacy-by-design a privacy-by-default, pričom citlivé mikroúdaje zostávajú spracovávané výlučne v chránenej internej zóne. Verejné analytické služby pracujú výhradne s publikovanými agregovanými údajmi, dátovými kockami, metaúdajmi a schválenými syntetickými datasetmi. Súčasťou riešenia sú mechanizmy riadenia prístupov, auditovateľnosť operácií, evidovanie prístupov k údajom a ochrana dát počas prenosu aj uloženia.	Zákon č. 18/2018 Z. z.; Nariadenie Európskeho parlamentu a Rady (EÚ) 2016/679 (GDPR); Zákon č. 305/2013 Z. z. o e-Governmente

Tabuľka 24 Určenie zdrojov a obsahu minimálnych bezpečnostných opatrení

5.7.3 Legislatívne, právne, štatutárne, regulačné a zmluvné požiadavky

Navrhovaná bezpečnostná architektúra platformy umelej inteligencie je navrhnutá tak, aby bola plne v súlade s legislatívnym rámcom Slovenskej republiky a Európskej únie upravujúcim prevádzku informačných systémov verejnej správy, ochranu osobných údajov, kybernetickú bezpečnosť a ochranu kritickej infraštruktúry. Bezpečnostné opatrenia platformy vychádzajú z princípov security by design a privacy by default a sú implementované na úrovni architektúry systému, riadenia prístupov, ochrany dátových tokov, monitorovania prevádzky a auditovateľnosti spracovania údajov.

Pri návrhu bezpečnostnej architektúry boli zohľadnené najmä nasledovné právne predpisy a regulačné rámce:

- Zákon č. 95/2019 Z. z. o informačných technológiách vo verejnej správe
- Zákon č. 69/2018 Z. z. o kybernetickej bezpečnosti
- Vyhláška 362/2018 Z. z. ktorou sa ustanovuje obsah bezpečnostných opatrení, obsah a štruktúra bezpečnostnej dokumentácie a rozsah všeobecných bezpečnostných opatrení
- Zákon č. 367/2024 Z. z. o kritickej infraštruktúre
- vyhláška Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 78/2020 Z. z. o štandardoch pre informačné technológie verejnej správy
- vyhláška Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 179/2020 Z. z., ktorou sa ustanovuje spôsob kategorizácie a obsah bezpečnostných opatrení informačných technológií verejnej správy
- vyhláška Úradu na ochranu osobných údajov Slovenskej republiky č. 158/2018 Z. z. o postupe pri posudzovaní vplyvu na ochranu osobných údajov
- Nariadenie Európskeho parlamentu a Rady (EÚ) 2016/679 z 27. apríla 2016 o ochrane fyzických osôb pri spracúvaní osobných údajov a o voľnom pohybe takýchto údajov, ktorým sa zrušuje smernica 95/46/ES (všeobecné nariadenie o ochrane údajov)
- Zákon č. 18/2018 Z. z. o ochrane osobných údajov a o zmene a doplnení niektorých zákonov.

Uvedené legislatívne požiadavky sú v návrhu riešenia implementované prostredníctvom kombinácie architektonických, organizačných a technických opatrení. Platforma využíva viacvrstvový bezpečnostný model zahŕňajúci riadenie prístupových práv prostredníctvom modelov RBAC a ABAC, centralizované logovanie bezpečnostných udalostí, auditovateľnosť operácií nad dátami, ochranu komunikačných rozhraní prostredníctvom API brány a implementáciu mechanizmov ochrany pred únikom dát (DLP a egress kontrol).

Z hľadiska ochrany údajov je architektúra platformy navrhnutá s dôrazom na klasifikáciu údajov a oddelenie bezpečnostných domén. Citlivé mikroúdaje a údaje podliehajúce štatistickej dôvernosti sú spracovávané výlučne v chránenej infraštruktúre dátového centra Štatistického úradu SR na území Slovenskej republiky. Verejné služby platformy pracujú iba s publikovanými agregovanými údajmi, dátovými kockami, metaúdajmi a schválenými syntetickými datasetmi a sú prevádzkované v oddelenej bezpečnostnej zóne bez priameho prístupu k citlivým dátam. Súčasťou implementácie platformy bude aj realizácia posúdenia vplyvu na ochranu osobných údajov (DPIA) v súlade s požiadavkami vyhlášky Úradu na ochranu osobných údajov SR č. 158/2018 Z. z. Platforma zároveň implementuje princíp Human-in-the-Loop, ktorý zabezpečuje, že výstupy generované systémami umelej inteligencie určené na publikovanie alebo využitie v interných procesoch podliehajú ľudskej kontrole a validácii.

Okrem legislatívnych požiadaviek sa návrh bezpečnostnej architektúry opiera aj o metodické odporúčania a bezpečnostné štandardy publikované národným bezpečnostným tímom CSIRT.SK, najmä v oblasti hardeningu systémov, bezpečnostného monitorovania a riadenia bezpečnostných incidentov. Pri implementácii a prevádzke platformy budú zároveň uplatňované interné bezpečnostné smernice a metodické materiály Štatistického úradu SR, ktoré upravujú klasifikáciu údajov, riadenie prístupových práv, správu identít používateľov, prevádzkové bezpečnostné postupy, monitorovanie bezpečnostných udalostí a riadenie bezpečnostných incidentov.

Architektúra platformy je zároveň navrhnutá tak, aby umožňovala pružnú reakciu na legislatívne a regulačné zmeny. Požiadavky vyplývajúce z legislatívnych zmien budú implementované prednostne prostredníctvom úprav bezpečnostných politík, konfigurácií a procesov riadenia prístupov, bez nutnosti zásadných zásahov do architektúry alebo dátovej vrstvy systému. Implementácia takýchto zmien bude prebiehať prostredníctvom riadeného procesu zmien zahŕňajúceho bezpečnostné posúdenie, testovanie dopadov a schvaľovací proces v rámci definovaného Go/No-Go rámca, čím sa zabezpečí kontinuita prevádzky a súlad s aktuálnymi regulačnými požiadavkami.

5.7.8 Riešenie autentifikácie a prístupov používateľov

Návrh riešenia autentifikácie a riadenia prístupov pre AI platformu Štatistického úradu SR (ŠÚ SR) vychádza z princípov centralizovaného riadenia identít, bezpečného jednotného prihlásenia (Single Sign-On) a dôsledného riadenia prístupových práv v súlade s klasifikáciou údajov a bezpečnostnou architektúrou riešenia. Cieľom navrhovaného riešenia je zabezpečiť vysokú úroveň ochrany citlivých štatistických údajov, najmä mikroúdajov spracovávaných v chránenej infraštruktúre ŠÚ SR, pri zachovaní komfortného a jednotného prístupu používateľov k jednotlivým službám platformy.

Autentifikácia používateľov bude realizovaná prostredníctvom centrálného systému správy identít (IAM), ktorý bude poskytovať jednotné prihlásenie pre všetky komponenty AI platformy vrátane analytických nástrojov, aplikačných rozhraní a administrátorských konzol. Riešenie bude postavené na technológiách podporujúcich moderné autentifikačné štandardy, najmä protokoly SAML 2.0 a OpenID Connect (OIDC). Na implementáciu IAM vrstvy sa navrhuje riešenie správy identít s podporou SAML 2.0/OIDC, ktoré poskytuje podporu pre federáciu identít, správu používateľských rolí a integráciu s existujúcou infraštruktúrou organizácie. Centrálny IAM systém bude integrovaný s referenčným zdrojom identít ŠÚ SR, ktorým je adresárová služba Active Directory (AD) / LDAP. Táto integrácia umožní využívať existujúce používateľské účty, skupiny a organizačné štruktúry bez potreby vytvárania duplicitných identít. Identita používateľa bude pri autentifikácii overovaná voči týmto adresárovým službám a následne bude IAM systém vydávať autentifikačné tokeny využívané jednotlivými komponentmi platformy. Pre prístup k citlivým údajom a privilegovaným operáciám bude povinne implementovaná viacfaktorová autentifikácia (MFA). Tá sa bude vyžadovať najmä pri prístupe do interných analytických nástrojov pracujúcich s mikroúdajmi v Zóne A, pri administrácii platformy alebo pri operáciách, ktoré umožňujú export alebo spracovanie citlivých údajov. Ako druhý autentifikačný faktor budú podporované štandardné mechanizmy, napríklad TOTP aplikácie, hardvérové bezpečnostné kľúče alebo WebAuthn.

Riešenie je navrhnuté tak, aby umožňovalo aj budúcu federáciu identít s externými systémami verejnej správy alebo partnerských organizácií. V prípade potreby bude možné integrovať autentifikačné mechanizmy napríklad s

identitnou infraštruktúrou ÚPVS, alebo s identitnými poskytovateľmi partnerských výskumných organizácií. Riadenie prístupov k jednotlivým službám platformy bude realizované kombináciou modelov RBAC (Role-Based Access Control) a ABAC (Attribute-Based Access Control). RBAC model bude definovať základné používateľské role a oprávnenia, zatiaľ čo ABAC mechanizmus umožní jemnozrné riadenie prístupov na základe atribútov používateľa, klasifikácie údajov, kontextu prístupu alebo bezpečnostnej zóny systému.

Interní používatelia ŠÚ SR budú rozdelení do viacerých kategórií:

- Štatistickí a analytici budú mať prístup k analytickým nástrojom v chránenej Zóne A, napríklad k bezpečnému AI analytikovi umožňujúcemu dopyty nad mikroúdajmi. Tento prístup bude vždy kontrolovaný mechanizmami štatistického utajenia a pravidlami ochrany dôverných údajov.
- Metodici budú zodpovední za správu pravidiel štatistického utajenia, kontrolu kvality údajov a metodické schvaľovanie prestupov údajov medzi zónami systému podľa princípu „four-eyes“.
- Administrátori a prevádzkoví pracovníci (SRE) budú zabezpečovať prevádzku platformy, správu infraštruktúry a klastrov orchestrácie kontajnerov. Ich prístup do citlivých systémov bude realizovaný prostredníctvom riešenia riadeného privilegovaného prístupu (bastion host alebo jump server) s úplným auditom vykonaných operácií.
- Bezpečnostní pracovníci budú mať prístup k bezpečnostným logom a nástrojom SIEM, kde budú monitorovať bezpečnostné udalosti a riadiť bezpečnostné politiky platformy.

Externí používatelia budú pristupovať predovšetkým k službám poskytovaným v Zóne B, ktoré pracujú výlučne s publikovanými agregovanými údajmi. Verejnosť a médiá budú mať možnosť využívať verejný AI analytický portál a vyhľadávacie služby najmä anonymne prostredníctvom webového rozhrania. Na úrovni API brány budú implementované mechanizmy ochrany proti zneužitiu služieb, napríklad rate-limiting, kvóty alebo monitoring objemu požiadaviek. V prípade partnerov a výskumníkov môže byť pre prístup k pokročilým službám (napr. programový prístup k datasetom alebo syntetickým mikrodátam) vyžadovaná registrácia a autentifikácia prostredníctvom OAuth2 alebo API kľúčov.

Dôležitou súčasťou návrhu je aj zabezpečenie správneho riadenia prístupových práv a segregácie povinností (Segregation of Duties – SoD). Žiadna rola nebude oprávnená vykonávať zároveň konfiguráciu bezpečnostných pravidiel a schvaľovanie operácií, ktoré tieto pravidlá využívajú. Tento princíp minimalizuje riziko neoprávnených zásahov do systému alebo zneužitia prístupových práv. Na úrovni aplikačných rozhraní a dátových služieb bude implementovaný mechanizmus centrálného vynucovania bezpečnostných politík prostredníctvom nástroja na vyhodnocovanie prístupových politík. Tento komponent umožní dynamicky vyhodnocovať pravidlá prístupu k API a dátovým zdrojom na základe identity používateľa, typu operácie a klasifikácie údajov. Každá operácia autentifikácie a autorizácie bude zaznamenaná v auditných logoch systému. Logy budú obsahovať identitu používateľa, typ operácie, čas operácie a korelačný identifikátor požiadavky. Tieto záznamy budú centralizovane odosielané do systému SIEM, kde budú slúžiť na detekciu bezpečnostných incidentov, forenznú analýzu a audit súladu s bezpečnostnými politikami organizácie. Takto navrhnutý model autentifikácie a riadenia prístupov zabezpečuje bezpečné využívanie AI platformy, chráni citlivé štatistické údaje a zároveň poskytuje flexibilný rámec pre integráciu nových služieb a používateľských skupín v budúcnosti.

6. PREVÁDZKA A ÚDRŽBA VÝSTUPOV PROJEKTU

6.1 Návrh riešenia prevádzky a údržby

V súčasnom stave je prevádzka a údržba informačných systémov Štatistického úradu SR zabezpečovaná prostredníctvom existujúcej infraštruktúry dátového centra a štandardných prevádzkových procesov IT oddelenia. Prevádzkové činnosti zahŕňajú najmä správu serverovej infraštruktúry, databázových systémov, sieťových komponentov a aplikačných služieb, pričom monitoring dostupnosti a základných prevádzkových parametrov je realizovaný prostredníctvom existujúcich nástrojov infraštruktúrneho dohľadu. Incidents a požiadavky používateľov sú riešené prostredníctvom interných procesov technickej podpory a servisných zásahov administrátorov.

Úroveň poskytovania služieb (SLA) je v súčasnosti definovaná najmä na úrovni dostupnosti základných informačných systémov a reakčných časov pri riešení incidentov. Prevádzka existujúcich systémov je orientovaná primárne na zabezpečenie kontinuity zberu a spracovania štatistických údajov, pričom diseminačné služby

a analytické nástroje sú spravované ako samostatné komponenty bez jednotnej architektúry služieb a bez centralizovaného riadenia životného cyklu AI alebo analytických modelov. Monitoring výkonnosti a využívania služieb je realizovaný prevažne na infraštruktúrnej úrovni a menej na úrovni aplikačných služieb alebo používateľských scenárov.

Z pohľadu budúceho rozvoja predstavuje výzvu najmä rastúca komplexnosť systémov, potreba integrácie moderných analytických nástrojov a požiadavka na vyššiu dostupnosť služieb pre verejnosť. Súčasný model prevádzky je orientovaný najmä na tradičné informačné systémy, zatiaľ čo budúca platforma založená na dátovej analytike a umelej inteligencii bude vyžadovať pokročilejšie mechanizmy monitoringu, riadenia modelov, správy API rozhraní a zabezpečenia kontinuity služieb.

6.2 Zabezpečenie podpory používateľov a prevádzky

Cieľom je zabezpečiť stabilnú, bezpečnú a auditovateľnú prevádzku AI platformy Štatistického úradu SR, vrátane podpory interných analytických služieb, spracovania mikroúdajov a poskytovania verejných AI služieb nad agregovanými dátami. Prevádzkový model rešpektuje architektonický princíp oddelenia citlivých a verejných častí riešenia a je navrhnutý v súlade s požiadavkami na ochranu štatistickej dôverylosti, rezidenciu dát na území Slovenskej republiky a kontrolu prístupu k mikroúdajom.

Prevádzkový model platformy

Platforma bude prevádzkovaná v dvojzónovej architektúre pozostávajúcej z internej bezpečnej zóny (Zóna A) a verejnej zóny (Zóna B).

- Zóna A predstavuje chránené interné prostredie určené na spracovanie citlivých štatistických mikroúdajov, tréning a inferenciu AI modelov nad citlivými dátami a analytické činnosti interných používateľov ŠÚ SR. Prístup do tejto zóny je striktné riadený prostredníctvom riadenia identít a prístupov (IAM), auditovaných prístupových politík a bezpečnostných kontrol. Zóna A neumožňuje externú konektivitu pre citlivé údaje a všetky operácie sú plne auditovateľné.
- Zóna B predstavuje verejnú prezentačnú vrstvu platformy a poskytuje služby pre externých používateľov, napríklad AI analytika nad publikovanými štatistickými agregátmi, vyhľadávanie informácií, generovanie textových súhrnov alebo vizualizácie dát. Táto zóna je optimalizovaná na dostupnosť, škálovanie a bezpečný prístup pre verejnosť, pričom neobsahuje ani nespracúva citlivé mikroúdaje.

Oddelenie týchto dvoch zón zabezpečuje, že nedochádza k presakovaniu citlivých údajov do verejného prostredia a zároveň umožňuje poskytovať moderné dátové služby verejnosti.

Dvojfázové nasadenie platformy

Nasadenie platformy je navrhnuté ako dvojfázový proces, ktorý umožňuje postupné overenie technického riešenia a minimalizáciu prevádzkových rizík. Prvá fáza slúži na overenie konceptu vybraných verejných služieb v izolovanom prostredí verejného cloudu v rámci EÚ nad necitlivými, syntetickými a verejnými údajmi. Cieľom tejto fázy je overiť správanie AI modelov a spôsob práce s dátami bez citlivých mikroúdajov. Po úspešnom pilotnom overení bude riešenie v druhej fáze nasadené do produkčného on-premise prostredia v dátovom centre ŠÚ SR, kde bude zabezpečené spracovanie citlivých štatistických údajov v súlade s legislatívnymi a bezpečnostnými požiadavkami. Tento prístup umožňuje minimalizovať implementačné riziká, zrýchliť vývoj riešenia a zároveň zachovať plnú kontrolu nad citlivými dátami v produkčnej prevádzke.

Riadenie prevádzky a zodpovednosti

Prevádzka platformy bude riadená podľa princípov moderných DevOps a MLOps postupov, ktoré zabezpečujú kontrolované nasadzovanie zmien, automatizáciu testovania a auditovateľnosť konfigurácie.

Zodpovednosti za jednotlivé oblasti budú rozdelené nasledovne:

- Prevádzkový tím (SRE/IT prevádzka ŠÚ SR) bude zodpovedný za stabilitu infraštruktúry, monitoring systémov, riadenie kapacity, dostupnosť služieb a dodržiavanie SLA parametrov.
- Bezpečnostný tím bude zabezpečovať riadenie prístupov, audit bezpečnostných udalostí, ochranu dát a dohľad nad dodržiavaním bezpečnostných politík a legislatívnych požiadaviek.
- Dátová správa a metodické tímy budú zodpovedné za kvalitu dát, správu dátových katalógov, klasifikáciu údajov a uplatňovanie pravidiel štatistického utajenia.

Vlastníci jednotlivých služieb budú riadiť funkčný rozvoj služieb, backlog požiadaviek a koordináciu zmien. Každá služba bude mať definované SLO/SLA, prevádzkový runbook, kontakty na on-call, schválené okná zmien, špecifikáciu závislostí, požiadavky na zálohy a obnovu a zoznam známych rizík a mitigácií.

Riadenie zmien a nasadzovanie

Riadenie zmien v platforme bude realizované prostredníctvom princípov Infrastructure as Code a GitOps, čo umožní plnú sledovateľnosť konfigurácie a kontrolované nasadzovanie nových verzií systému. Každá zmena konfigurácie, aplikácie alebo AI modelu bude prechádzať schvaľovacím procesom, automatizovaným testovaním a bezpečnostnou kontrolou pred nasadením do produkčného prostredia. Nasadzovanie zmien bude realizované postupne pomocou stratégií ako canary deployment alebo shadow deployment, ktoré umožňujú minimalizovať dopad zmien na používateľov a zabezpečiť rýchly rollback v prípade problémov.

Riadenie kapacity a výkonu

Prevádzka platformy bude kontinuálne monitorovať využitie výpočtových zdrojov, vrátane CPU, GPU, pamäte, úložiska a sieťovej priepustnosti. V prípade verejných služieb v Zóne B bude využívané horizontálne škálovanie, cache mechanizmy a optimalizácia dotazov nad dátovými kockami. V internej zóne bude výkon riadený prostredníctvom plánovania analytických úloh, riadenia dávkových procesov a prioritizácie interaktívnych požiadaviek analytikov.

Podpora používateľov a incident management

Podpora používateľov a prevádzky bude organizovaná v troch úrovniach podpory.

- Podpora L1 predstavuje prvú líniu podpory a slúži ako jednotné kontaktné miesto pre používateľov platformy. Jej úlohou je prijímať požiadavky, vykonávať základnú diagnostiku a riešiť jednoduché problémy, napríklad problémy s prihlásením alebo dostupnosťou služieb.
- Podpora L2 je zabezpečovaná špecializovanými prevádzkovými tímami, ktoré riešia technické problémy aplikácií, integrácií alebo dátových procesov. L2 vykonáva podrobnejšiu analýzu incidentov a spolupracuje s L1 na ich riešení.
- Podpora L3 predstavuje najvyššiu úroveň podpory a je zabezpečovaná dodávateľom riešenia alebo vývojovým tímom platformy. Táto úroveň rieši komplexné systémové problémy, chyby v aplikáciách alebo zmeny v architektúre riešenia.

Podpora infraštruktúrnych služieb

Prevádzka platformy zahŕňa aj podporu infraštruktúrnych komponentov, ako sú výpočtové servery, úložiská, sieťová infraštruktúra a prípadné cloudové služby využité v pilotnej fáze. Za prevádzku infraštruktúry bude zodpovedný prevádzkovateľ dátového centra ŠÚ SR alebo zmluvný poskytovateľ infraštruktúrnych služieb. Komunikácia medzi prevádzkovateľom infraštruktúry a prevádzkovým tímom platformy bude definovaná v servisných zmluvách a prevádzkových postupoch.

Podpora	Poskytovateľ (subjekt zodpovedný za poskytnutie podpory)	Požadovaný Čas dostupnosti	STAV zabezpečenia	Pozn. (napr. známe obmedzenia služby, špeciálne zodpovednosti, a pod.)
L1 – service desk	IT podpora ŠÚ SR	8x5	Existujúca služba	Jednotný kontaktný bod

L2 – aplikačná prevádzka	IT prevádzka / dátový tím	8×5		Riešenie incidentov
L3 – vývoj a architektúra	Dodávateľ riešenia	8×5 / SLA	Súčasť obstarania	Riešenie komplexných problémov
Infraštruktúra	Prevádzkovateľ DC ŠÚ SR	24×7	Existujúce služby	HW, sieť, virtualizácia

Tabuľka 25 Prehľad riešenia zabezpečenia podpory používateľov a prevádzky

6.3 Riešenie incidentov v prevádzke - parametre úrovni služby

Monitoring pokrýva infraštruktúru, platformu a aplikácie v Zóne A aj Zóne B. Metriky výkonu (latencia, priepustnosť, využitie zdrojov), dostupnosti a kvality modelov sú zbierané a vizualizované na rolovo orientovaných dashboardoch. Logy a trasovanie požiadaviek umožňujú koreláciu udalostí naprieč vrstvami a zónami. Každý request nesie korelačný identifikátor prenášaný celým spracovaním. Kľúčové bezpečnostné udalosti (autentizácia/autorizácia, prestupy A → B, pokusy o porušenie politík) sú odosielané do centrálnych auditných úložísk a SIEM. Nastavené alerty vychádzajú z definovaných SLO a z detekcie anomálií. Eskalácie smerujú na on-call, bezpečnosť alebo vlastníkov služieb podľa typu incidentu.

SLO/SLA sú definované pre každú kľúčovú službu. Pre verejného AI analytika a vyhľadávanie v Zóne B je cieľová latencia bežných dotazov do dvoch sekúnd, dostupnosť je nastavená na vysokú spoľahlivosť s vymedzeným error budgetom a jasne danými údržbovými oknami. Pre bezpečného AI analytika v Zóne A sa latencia prispôbuje zložitosti dopytov, s dôrazom na správnosť a uplatnenie štatistického utajenia. Dostupnosť zohľadňuje plánované dávkové okná a náročné analytické scenáre. SLO sú prepojené s kapacitnými plánmi, testami výkonu a prevádzkovými rozhodnutiami (napríklad navýšenie replík, rozšírenie cache, optimalizácia dotazov).

Incident management popisuje postup od detekcie cez triáž, mitigáciu, komunikáciu a obnovu až po uzavretie incidentu. Závažnosť incidentov sa triedi podľa dopadu na používateľov a bezpečnosť. Každý incident má vlastníka, časovú os, komunikáciu k zainteresovaným stranám a post-incidentnú analýzu s nápravnými opatreniami a termínmi. Problem management zaisťuje, že opakujúce sa príčiny sú odstránené trvalými zmenami v konfiguráciách, kóde, infraštruktúre alebo procesoch. Výstupom sú aktualizované runbooky, zmeny v testovacích scenároch a upravené SLO. Na zraniteľnosti sa uplatňuje pravidelný cyklus skenovania, prioritizácie a nápravy. Kritické zistenia majú definované maximálne časy nápravy a sú sledované až do uzavretia.

Riadenie verzií a release management nadväzuje na monitoring a incidentné procesy. Nasadenia prebiehajú cez schválené pipeline s automatizovanými testami, kontrolami kompatibility rozhraní a bezpečnostnými kontrolami. Pre Zónu B sa preferujú postupné stratégie (canary/shadow) s detailným monitoringom dopadu. V prípade regresie sa uplatní rýchly návrat na predchádzajúcu verziu. Pre Zónu A platia prísnejšie podmienky schvaľovania, plánovania okien a izolácie vplyvu na mikroúdaje a interné analytické toky.

Parametre služby riešenia incidentov v prevádzke sú špecifikované na základe určenia priority incidentu pomocou kombinácie jeho naliehavosti a dopadu podľa najlepších skúseností z praxe (best practice) z oblasti manažmentu IT služieb (Information Technology Infrastructure Library - ITIL V3) nasledovným spôsobom:

Incident - za incident je považovaná každá nahlásená alebo inak zistená relevantná skutočnosť týkajúca sa aktíva (informačného systému) alebo jeho častí, ktorého nedostupnosť alebo nefunkčnosť má vplyv na poskytovanie služieb.

Klasifikácia naliehavosti incidentu	Závažnosť incidentu	Popis naliehavosti incidentu
A	Kritická	Kritické chyby, ktoré spôsobia úplné zlyhanie systému ako celku a nie je možné používať ani jednu jeho časť, nie je možné poskytnúť požadovaný výstup z IS.
B	Vysoká	Chyby a nedostatky, ktoré zapríčinia čiastočné zlyhanie systému a neumožnia používať časť systému.
C	Stredná	Chyby a nedostatky, ktoré spôsobia čiastočné obmedzenia používania systému.

D Nízka Kozmetické a drobné chyby.

Tabuľka 26 Klasifikácia naliehavosti incidentu

Klasifikácia závažnosti incidentu	Dopad	Popis dopadu
1	katastrofický	katastrofický dopad, priamy finančný dopad alebo strata dát
2	značný	značný dopad alebo strata dát
3	malý	malý dopad alebo strata dát

Tabuľka 27 Klasifikácia závažnosti incidentu

Určenie priority incidentu je kombináciou dopadu a naliehavosti podľa nasledovnej matice:

Matica priority incidentov		Dopad		
		Katastrofický - 1	Značný - 2	Malý - 3
Naliehavosť	Kritická - A	1	2	3
	Vysoká - B	2	3	3
	Stredná - C	2	3	4
	Nízka - D	3	4	4

Tabuľka 28 Určenie priority incidentu

Parametre služby Riešenia incidentov v prevádzke:

Označenie priority incidentu	Reakčná doba ⁽¹⁾ od nahlásenia incidentu po začiatok riešenia incidentu	Doba konečného vyriešenia incidentu od nahlásenia incidentu (DKVI) ⁽²⁾	Spoľahlivosť ⁽³⁾ (počet incidentov za mesiac)
1	1 hod.	4 hodiny	1
2	1 hod.	12 hodín	2
3	1 hod.	24 hodín	10
4	1 hod.	Vyriešené a nasadené v rámci plánovaných releasov (vydaní novej verzie programového vybavenia a konfigurácie)	

Tabuľka 29 Parametre služby Riešenia incidentov v prevádzke

6.4 Požadovaná dostupnosť informačného systému:

Nasledujúca tabuľka uvádza požadované parametre dostupnosti informačného systému.

Popis	Parameter	Upresnenie
Prevádzkové hodiny	12 hodín	od 6:00 hod. - do 18:00 hod. počas pracovných dní
Servisné okno	10 hodín	od 19:00 hod. - do 5:00 hod. počas pracovných dní
	24 hodín	od 00:00 hod. - 23:59 hod. počas dní pracovného pokoja a štátnych sviatkov Servis a údržba sa bude realizovať mimo pracovného času.
Dostupnosť produkčného prostredia IS	98,5 %	98,5 % z 24/7/365, t. j. max. ročný výpadok je 131 hod. Maximálny mesačný výpadok je 10,9 hodiny.

Vždy sa za takúto dobu považuje čas od 0:00 hod. do 23:59 hod. počas všetkých kalendárnych dní.

Nedostupnosť IS sa počíta od nahlásenia incidentu Zákazníkom v čase dostupnosti podpory Poskytovateľa (t. j. nahlásenie incidentu na L3 v čase od 6:00 hod. - do 18:00 hod. počas pracovných dní). Do dostupnosti IS nie sú započítavané servisné okná a plánované odstávky IS.

V prípade nedodržania dostupnosti IS bude každý ďalší začatý pracovný deň nedostupnosti braný ako deň omeškania bez odstránenia vady alebo incidentu.

RTO vyjadruje množstvo času potrebné pre obnovenie dát a celej prevádzky nedostupného systému

RPO vyjadruje, do akého času (bodu) v minulosti možno obnoviť dáta, t. j. rozsah dát, o ktoré môže organizácia prísť

RTO (Recovery Time Objective) 4 hodiny

RPO (Recovery Point Objective) 6 hodín

Tabuľka 30 Požadovaná dostupnosť informačného systému

6.5 Požiadavky na ľudské zdroje potrebné pre zabezpečenie prevádzky

Prevádzka navrhovanej AI platformy bude zabezpečená kombináciou interných kapacít Štatistického úradu SR a špecializovaných odborných kapacít dodávateľ a riešenia. Personálne zabezpečenie prevádzky je navrhnuté tak, aby pokrývalo technickú prevádzku platformy, správu infraštruktúry, správu dát, správu AI modelov a podporu používateľov.

V rámci prevádzky systému bude potrebné zabezpečiť najmä nasledujúce kompetenčné oblasti.

Prvou oblasťou je prevádzka IT infraštruktúry a platformy, ktorá zahŕňa správu serverovej infraštruktúry, virtualizačnej vrstvy, úložných systémov a sieťovej infraštruktúry v dátovom centre Štatistického úradu SR. Tieto činnosti budú zabezpečované IT prevádzkovým tímom organizácie, ktorý bude zodpovedný za dostupnosť systému, monitoring prevádzky, riadenie kapacity a riešenie technických incidentov.

Druhou oblasťou je správa aplikačnej platformy a integračných komponentov, ktorá zahŕňa správu jednotlivých aplikačných služieb platformy, API rozhraní a integračných mechanizmov medzi jednotlivými systémami. Za túto oblasť budú zodpovední aplikační administrátori a špecialisti na prevádzku informačných systémov.

Treťou oblasťou je správa dát a dátových procesov, ktorá zahŕňa správu dátových tokov, kontrolu kvality dát, správu dátového katalógu a dohľad nad dodržiavaním pravidiel štatistického utajenia. Túto oblasť budú zabezpečovať dátoví kurátori, dátoví analytici a metodickí pracovníci Štatistického úradu SR.

Špecifickou oblasťou prevádzky je správa AI a analytických modelov, ktorá zahŕňa správu modelov umelej inteligencie, ich monitorovanie, aktualizáciu a kontrolu kvality výsledkov. Táto činnosť zahŕňa najmä správu modelového repozitára, validáciu modelov, sledovanie výkonu modelov a riadenie ich opätovného tréningu. Tieto činnosti budú zabezpečované špecialistami na dátovú vedu a AI analytiku.

Samostatnou oblasťou je podpora používateľov a riešenie incidentov, ktorá bude zabezpečovaná prostredníctvom troj úrovňového modelu podpory (L1, L2 a L3). Prvá úroveň podpory bude zabezpečovaná interným servisným pracoviskom IT podpory Štatistického úradu SR, ktoré bude slúžiť ako jednotné kontaktné miesto pre používateľov. Druhá úroveň podpory bude zabezpečovaná internými prevádzkovými tímami a aplikačnými

administrátormi. Tretia úroveň podpory bude zabezpečovaná dodávateľom riešenia alebo vývojovým tímom platformy.

Z pohľadu kompetencií budú pre prevádzku systému potrebné najmä tieto odborné roly:

- administrátor IT infraštruktúry a virtualizačnej platformy
- aplikačný administrátor platformy
- dátový kurátor alebo dátový správca
- dátový analytik alebo AI špecialista
- pracovník IT podpory (service desk)

Okrem samotného personálneho zabezpečenia prevádzky je potrebné zabezpečiť aj adekvátne školenia a rozvoj odborných kompetencií pracovníkov, ktorí budú systém využívať alebo prevádzkovať. Pre používateľov systému (analytici, metodici, štatistici) budú realizované školenia zamerané na používanie analytických nástrojov platformy, prácu s AI asistovanými analytickými funkciami a interpretáciu výsledkov generovaných systémom. Pre pracovníkov IT prevádzky budú realizované odborné školenia zamerané na správu platformy, monitoring systému, správu infraštruktúry a bezpečnostné postupy. V prípade využitia špecializovaných technologických komponentov platformy sa predpokladá aj absolvovanie odborných školení alebo certifikácií poskytovaných výrobcami technológie. Pre pracovníkov správy dát a AI modelov budú realizované školenia zamerané na správu dátových procesov, správu dátového katalógu, správu AI modelov a monitorovanie kvality modelov. Školenia budú realizované najmä v období implementácie systému a pri jeho nasadení do produkčnej prevádzky, pričom v prípade potreby budú zabezpečené aj priebežné školenia pri rozširovaní funkčnosti platformy alebo pri nasadzovaní nových analytických nástrojov.

6.6 Požiadavky na zdrojové kódy

Navrhované riešenie AI platformy pre Štatistický úrad SR predpokladá dodávku kompletných zdrojových kódov všetkých aplikačných častí riešenia vrátane dokumentácie a súvisiacich artefaktov v súlade s § 15 ods. d) zákona č. 95/2019 Z. z. o informačných technológiách vo verejnej správe. Dodávateľ je povinný odovzdať objednávateľovi zdrojové kódy všetkých komponentov vytvorených v rámci projektu vrátane skriptov, konfiguračných súborov, pipeline definícií, databázových schém, integračných rozhraní a dokumentácie potrebnej na kompiláciu, nasadenie a prevádzku riešenia. Zdrojové kódy budú odovzdávané v plnej čitateľnej forme vrátane komentárov, dokumentácie architektúry a návodov na zostavenie a nasadenie systému. Súčasťou dodávky budú aj skripty pre automatizované nasadzovanie aplikácií, konfiguračné súbory prostredí, definície kontajnerov a definície CI/CD pipeline. Dokumentácia bude obsahovať opis architektúry systému, opis jednotlivých modulov, rozhraní API, dátových štruktúr a postupov pre správu a prevádzku riešenia. Zdrojové kódy budú spravované v centralizovanom verziovacom systéme, pričom objednávateľ bude mať plný prístup k repozitárom zdrojových kódov počas celej doby realizácie projektu aj po jeho ukončení. Repozitáre budú obsahovať históriu zmien, evidenciu autorov zmien, dokumentáciu commitov a označenie jednotlivých verzií systému. Štruktúra repozitárov bude reflektovať architektúru riešenia a bude rozdelená podľa jednotlivých aplikačných komponentov platformy.

Preberanie zdrojových kódov bude viazané na jednotlivé míľniky projektu a fakturačné etapy. Pri každom míľniku dodávateľ odovzdá aktuálnu verziu zdrojových kódov vrátane dokumentácie, build skriptov a konfiguračných súborov. Súčasťou preberania bude aj kontrola kvality kódu, overenie možnosti zostavenia systému zo zdrojových kódov a kontrola úplnosti dokumentácie. Finálne odovzdanie zdrojových kódov bude realizované pri akceptácii produkčného riešenia. Zdrojové kódy budú archivované v repozitári objednávateľa alebo v repozitári spravovanom v rámci infraštruktúry verejnej správy, pričom súčasťou archivácie bude aj verzovaná dokumentácia systému, konfiguračné súbory a artefakty nasadzovania. Archivácia bude zabezpečená tak, aby bolo možné kedykoľvek reprodukovať konkrétnu verziu systému vrátane všetkých jeho komponentov. Na zabezpečenie dispozičných práv k riešeniu bude objednávateľovi udelená nevýhradná, časovo neobmedzená a územne neobmedzená licencia na používanie, úpravu a ďalší rozvoj zdrojových kódov. V prípade komponentov vyvinutých špecificky pre projekt bude preferované licencovanie pod otvorenou licenciou kompatibilnou s licenciou European Union Public Licence (EUPL), pokiaľ to charakter komponentu umožňuje.

Navrhovaná architektúra riešenia zároveň minimalizuje riziko vendor lock-inu. Platforma je navrhnutá na princípoch otvorených štandardov a využíva štandardizované rozhrania API. Všetky integračné rozhrania sú dokumentované prostredníctvom OpenAPI špecifikácií a jednotlivé komponenty platformy sú navrhnuté modulárnym spôsobom, čo umožňuje ich nezávislú výmenu alebo rozšírenie bez nutnosti zásahu do celej architektúry systému. Dodávateľ je povinný zabezpečiť, aby riešenie neobsahovalo komponenty alebo licenčné podmienky, ktoré by

znemožňovali ďalší rozvoj systému alebo jeho prevádzku iným dodávateľom. Všetky ustanovenia týkajúce sa dodávky zdrojových kódov, licenčných práv a prevencie vendor lock-inu budú explicitne uvedené v Zmluve o dielo a následne reflektované aj v zmluve o podpore a prevádzke systému. Pri vývoji a odovzdávaní zdrojových kódov budú dodržané požiadavky metodického usmernenia MIRRI SR č. 024077/2023 o kvalite zdrojových kódov a balíkov softvéru, ktoré stanovuje pravidlá pre kvalitu, dokumentáciu a správu softvérových artefaktov vo verejnej správe.

Zálohy, DR a kontinuita

Zálohovanie a obnova sú navrhnuté tak, aby chránili integritu a dostupnosť kritických dát a služieb v Zóne A aj Zóne B. V Zóne A sa zálohujú mikrodátové databázy, tréningové a validačné datasky, modelové artefakty a kritické konfigurácie. V Zóne B sa zálohujú publikované agregáty, vyhľadávacie indexy, konfigurácie, šablóny a obsah potrebný pre verejné služby. Zálohy sú šifrované, retenčné politiky sú definované podľa kritickosti a legálnych požiadaviek a prístup k zálohám je riadený a auditovaný.

Plán obnovy (BC/DR) určuje prioritizáciu služieb a poradie obnovy. Pre Zónu B sa kladie dôraz na rýchlu obnovu verejných služieb s minimálnym výpadkom. Redundancia a vysoká dostupnosť sú dosiahnuté viacnásobnými replikami a automatickým failoverom. Pre Zónu A je prioritou integrita a dôverynosť. Obnova prebieha kontrolované s verifikáciou konzistencie dát, konfigurácií a politík prístupu. BC/DR dokumentácia zahŕňa spúšťacie podmienky, komunikačné scenáre, kompetencie a alternatívne postupy, ak niektoré kapacity nie sú dostupné. Testy obnovy sú pravidelné a povinné. Table-top cvičenia overujú organizačnú pripravenosť a rozhodovacie procesy. Technické DR testy overujú reálne obnovenie databáz, indexov, konfigurácií a kľúčových služieb v definovaných RPO/RTO. Výsledky testov vedú k aktualizácii runbookov, zlepšeniu skriptov a automatizácie a k úpravám hardvérových profilov a kapacít, ak je to potrebné. Integrita a dostupnosť auditných záznamov je súčasťou obnovy. Po incidente alebo DR teste sa kontroluje úplnosť auditov, súlad prestupov a opätovné vynútenie DLP a egress politík.

Kontinuita prevádzky je zabezpečovaná aj preventívnymi opatreniami: pravidelné kontroly konfigurácií, validácie egress a DLP politík, overovania platnosti certifikátov a kľúčov, záťažové a odolnostné testy, ako aj riadené okná zmien. Tímy prevádzky a bezpečnosti koordinujú plánovanie zmien tak, aby nedochádzalo ku konfliktom s publikačnými cyklami ani s internými analytickými uzávierkami. Všetky rozhodnutia a zásahy sú zaznamenané a dohľadateľné. Platforma je tým pripravená na interné aj externé audity, ktoré preverujú bezpečnosť, súlad a pripravenosť na incidenty a obnovu.

7. OPIS IMPLEMENTÁCIE PROJEKTU A PREBERANIA VÝSTUPOV PROJEKTU

Implementácia a preberanie výstupov projektu bude realizované v súlade s vyhláškou Ministerstva investícií, regionálneho rozvoja a informatizácie Slovenskej republiky č. 401/2023 Z. z. o riadení projektov a zmenových požiadaviek v prevádzke informačných technológií verejnej správy v zmysle ustanovení podľa § 5 a nasledovných ustanovení. Požadované výstupy jednotlivých etáp uvádza nasledujúca tabuľka.

Etapa	Požadované výstupy
R1	R-01 Projektový iniciálny dokument (PID) pre všetky funkčné oblasti
Analýza a dizajn	<ul style="list-style-type: none">• Zoznam požiadaviek• Akceptačné kritériá• Rámcová špecifikácia riešenia (Popis produktu, Dekompozícia produktu, Vývojový diagram produktu)• Biznis architektúra• Aplikačná architektúra• Technologická architektúra – časť systémová architektúra• Bezpečnostná architektúra• Stratégia testovania• Plán testovania• Testovacie scenáre a prípady
	R1-1 Detailný návrh riešenia (DNR)

	<ul style="list-style-type: none">(1) Zámer riešenia, analýza požiadaviek, používateľský prieskum a motivačná architektúra(2) Popis postupu analýzy a návrhu riešenia(3) Biznis architektúra<ul style="list-style-type: none">a. Existujúca a cieľová biznis architektúrab. Procesy podporované navrhovaným riešenímc. Vytvorenie informačnej architektúry a mapovanie používateľskej cestyd. Vytvorenie grafického návrhu a prototypu používateľského rozhrania (UX, UI)e. Prípady použitia (use case model)(4) Dátová architektúra(5) Aplikačná architektúra<ul style="list-style-type: none">a. Existujúca a budúca aplikačná architektúrab. Aplikačné komponenty a ich vzťah k biznis komponentom a funkčným požiadavkámc. Integrácie – Komunikácia medzi komponentami (OpenAPI)(6) Technologická architektúra<ul style="list-style-type: none">a. Existujúca a budúca technologická architektúrab. Technologické komponenty riešenia a ich vzťah k aplikačným komponentom(7) Softvérové licencie a zdrojové kódy(8) Požiadavky na úrovne služieb (SLA) a výkonnosť(9) Zabezpečenie dostupnosti, zálohovanie a obnova riešenia(10) Bezpečnosť – riešenie požiadaviek na bezpečnosť(11) Migrácia dát(12) Harmonogram realizácie a nasadenia, závislosti
R2	R1-2 Plán a stratégia testovania
	<ul style="list-style-type: none">(1) Testovacie prípady (UC/TC)(2) Testovacie prostredia(3) Testovacie dáta(4) Defekt manažment, monitoring a reporting test
Nákup technických prostriedkov, programových prostriedkov a služieb	R2-1 Obstaranie technických prostriedkov a infraštruktúry
	R2-2 Obstaranie programových prostriedkov a služieb
R3	R3-1 Vývoj, migrácia údajov a integrácií
Implementácia a testovanie	R3-2 Testovanie
	<ul style="list-style-type: none">(1) Funkčné testovanie (FAT)(2) Systémové a integračné testovanie (SIT)(3) Zátťažové a výkonnostné testovanie(4) Bezpečnostné testovanie (SW/HW a kybernetická bezpečnosť)

		(5) Používateľské testy funkčného používateľského rozhrania (UX)
		(6) Používateľské akceptačné testovanie (UAT)
		(7) Školenia pre administrátorov
		(8) Testovacie dáta
		(9) Testovacie prostredia
		R3-3 Dokumentácia
		Aplikačná príručka
		Používateľská príručka
		Integračný manuál
		Inštalачný a konfiguračný manuál
		Testovacie scenáre
R4		R4-1 Nasadenie do produkčnej prevádzky
Nasadenie		<ul style="list-style-type: none"> • Príprava produkčného prostredia • Administratívna príprava produkčného prostredia (procesy, dokumentácia) • Inštalácia riešenia do produkčného prostredia • Sprístupnenie riešenia v produkčnom prostredí vybraným používateľom
		R4-2 Akceptácia spustenia do produkčnej prevádzky
M-02		Manažérsky produkt
Dokončovacia fáza projektu		M-02 Správa o dokončení projektu M-02 Správa o získaných poznatkoch M-02 Plán kontroly po odovzdaní projektu M-02 Odporúčanie nadväzných krokov M-02 Zoznam funkčných zdrojových kódov
Služby projektového riadenia		Manažérsky produkt M-01 Plán etapy M-02 Manažérske správy, plány, reporty, zoznamy a požiadavky M-03 Akceptačný protokol M-04 Audit kvality M-05 Analýza nákladov a prínosov Zoznam rizík a závislostí
		Tabuľka 31 Požadované výstupy etáp projektu

8. ODKAZY

Bez odkazov.

9. PRÍLOHY

Príloha 1: Zoznam rizík a závislostí projektu AI platforma pre Štatistický úrad SR (Excel)

Koniec dokumentu