Prípravná trhová konzultácia projektu

Konsolidovaná analytická vrstva

využitie veľkých dát pre zlepšenie fungovania inštitúcií verejnej správy

Obsah

[1 Základné informácie 1](#_Toc7768520)

[1.1 Prehľad 1](#_Toc7768521)

[1.2 Dôvod 3](#_Toc7768522)

[1.3 Rozsah 4](#_Toc7768523)

[1.4 Použité skratky a značky 5](#_Toc7768524)

[2 Manažérske zhrnutie 11](#_Toc7768525)

[3 Motivácia 21](#_Toc7768526)

[3.1.1 Súhrnný popis 21](#_Toc7768527)

[3.1.2 Riziká 23](#_Toc7768528)

[4 Popis aktuálneho stavu 24](#_Toc7768529)

[4.1 Legislatíva 24](#_Toc7768530)

[4.2 Architektúra 24](#_Toc7768531)

[4.2.1 Biznis architektúra 24](#_Toc7768532)

[4.2.2 Architektúra informačných systémov 28](#_Toc7768533)

[4.2.3 Technologická architektúra 28](#_Toc7768534)

[4.2.4 Bezpečnostná architektúra 29](#_Toc7768535)

[4.3 Prevádzka 29](#_Toc7768536)

[5 Popis budúceho stavu 30](#_Toc7768537)

[5.1 Legislatívne požiadavky 30](#_Toc7768538)

[5.2 Biznis architektúra navrhovaného riešenia 30](#_Toc7768539)

[5.3 Požiadavky na Technologická architektúra 47](#_Toc7768540)

[5.4 Požiadavky na Bezpečnostná architektúra 48](#_Toc7768541)

Zoznam obrázkov

[Obrázok 1: Koncept manažmentu údajov vo verejnej správe 12](#_Toc7768568)

[Obrázok 2: Základná štruktúra podporovaných dátových projektov 14](#_Toc7768569)

[Obrázok 5: Prehľad kľúčových používateľov analytickej vrstvy 32](#_Toc7768570)

[Obrázok 6: Životný postup dátovej vedy 33](#_Toc7768571)

Zoznam tabuliek

[Tabuľka 1 Základné informácie - zhrnutie 1](#_Toc7768577)

[Tabuľka 2: Dôležité termíny a pojmy 5](#_Toc7768578)

[Tabuľka 3: Skratky a značky 7](#_Toc7768579)

[Tabuľka 4: Logické komponenty Konsolidovanej analytickej vrstvy 13](#_Toc7768580)

[Tabuľka 5: Využitie Konsolidovanej analytickej vrstvy v praxi a realizované prípady použitia 16](#_Toc7768581)

[Tabuľka 6: Prehľad procesov, ktoré bude Konsolidovaná analytická vrstva podporovať 18](#_Toc7768582)

[Tabuľka 7: Architektúra Konsolidovanej analytickej vrstvy: Nasadené prvky dátovej infraštruktúry 18](#_Toc7768583)

[Tabuľka 8 Dátové zdroje dostupné po skončení projektu 20](#_Toc7768584)

[Tabuľka 9: Nastavenie biznis cieľov riešenia a merateľných ukazovateľov pre projekt 21](#_Toc7768585)

[Tabuľka 10: Inštitút pre stratégie a analýzy (ISA) 24](#_Toc7768586)

[Tabuľka 11: Odbor analýz vzdelávacej politiky (IVP) 25](#_Toc7768587)

[Tabuľka 12: Infostat 26](#_Toc7768588)

[Tabuľka 13: Monitorovanie strategických cieľov štátu a plnenia programového vyhlásenia 38](#_Toc7768589)

[Tabuľka 14: Strategické plánovanie rozvoja Slovenska 39](#_Toc7768590)

[Tabuľka 15: Hodnotenie vplyvu EŠIF 40](#_Toc7768591)

[Tabuľka 16: Hodnotenie výkonnosti vzdelávacích inštitúcií 41](#_Toc7768592)

[Tabuľka 17: Predpovedanie budúcich výsledkov žiakov 42](#_Toc7768593)

[Tabuľka 18: Predvídanie budúceho dopytu po službách (kapacitách vzdelávacích inštitúcií) 42](#_Toc7768594)

[Tabuľka 19: Prerozdeľovanie študentov do 1. ročníkov SŠ 43](#_Toc7768595)

[Tabuľka 20: Vzdelávanie ľudí s predčasne ukončenou školskou dochádzkou 44](#_Toc7768596)

[Tabuľka 21: Krátkodobé predikcie vývoja slovenskej ekonomiky 44](#_Toc7768597)

[Tabuľka 22: Rýchle odhady HDP a celkovej zamestnanosti v hospodárstve 45](#_Toc7768598)

[Tabuľka 23: Krátkodobé predikcie vývoja (indexov) priemyselnej produkcie 46](#_Toc7768599)

[Tabuľka 24: Populačné prognózy 46](#_Toc7768600)

# Základné informácie

## Prehľad

Tabuľka 1 Základné informácie - zhrnutie

|  |  |
| --- | --- |
| Zdôvodnenie využitia národného projektu a vylúčenia výberu projektu prostredníctvom výzvy | |
| Rada vlády pre digitalizáciu verejnej správy a digitálny jednotný trh schválila dňa 28.2.2017 dokument Manažment údajov, ktorý navrhuje riešiť využitia údajov  Projekt **Konsolidovaná analytická vrstva** je súčasťou riešenia princípov definovaných v strategickej priorite Manažment údajov. Problematika analytického spracovania údajov vo verejnej správe a ich ďalšieho využitia pre lepší návrh politík si vyžaduje odbornú koordináciu aktivít, ako je zber dát, čistenie a spracovanie dát, sprístupnenie údajov a zároveň technickú prevádzku riešení a odbornú asistenciu pri využívaní moderných analytických nástrojov.  Navrhované opatrenia pre tento projekt môže realizovať na centrálnej úrovni len jedna inštitúcia - Úrad podpredsedu vlády Slovenskej republiky pre investície a informatizáciu (ďalej len UPPVII).  Centrálne riešenie problému je najlepším variantom z nasledujúcich dôvodov:   * lepší prenos skúseností a praktík pri komunikácii s jednotlivými inštitúciami, * lepšia kontrola dosahovania výsledkov, * súčasný stav (neuspokojivý) je spôsobený najmä nedostatočnou centrálnou koordináciou aktivít. | |
| Prijímateľa/partnera národného projektu a dôvod jeho určenia | |
| Projekt bude implementovať prijímateľ ÚPPVII.  Podľa § 34a kompetenčného zákona je ÚPPVII ústredným orgánom štátnej správy pre oblasť informatizácie spoločnosti.  ÚPPVII v oblasti informatizácie spoločnosti zabezpečuje centrálne riadenie informatizácie spoločnosti a tvorbu politiky jednotného digitálneho trhu, rozhodovanie o využívaní finančných zdrojov vo verejnej správe pre informačné technológie, centrálnu architektúru integrovaného informačného systému verejnej správy a koordináciu plnenia úloh v oblasti informatizácie spoločnosti.  Kompetencie ÚPPVII tak umožňujú realizovať navrhnuté iniciatívy ako jedinej inštitúcii verejnej správy. | |
| Príslušnosť národného projektu k relevantnej časti PO7 OPII | Predkladaná štúdia je štúdiou uskutočniteľnosti pre programové obdobie 2014 až 2020 pre Operačný program Integrovaná infraštruktúra, Prioritná os číslo 07 Informatizácia spoločnosti.  Projekt je príslušný k špecifickým cieľom:  7.5 Zlepšovanie celkovej dostupnosti dát verejnej správy vo forme otvorených dát  7.7 Umožnenie modernizácie a racionalizácie verejnej správy IKT prostriedkami |

Z pohľadu úsekov verejnej správy rieši:

* U00029 - Informatizácia spoločnosti

##### Výsledkom projektu Konsolidovaná analytická vrstva bude:

* **Vytvorenie vybraných analytických výstupov** (produktov):
* pre podporu strategického riadenia štátu (“dashboard” o sledovaní strategických cieľov a merateľných ukazovateľov pre prioritné oblasti krajiny),
* pre podporu návrhu politík, testovanie politík a hodnotenie efektivity vybraných politík,
* pre podporu výpočtov hodnoty za peniaze,
* pre podporu operatívneho rozhodovania.
* **Vytvorenie spoločnej analytickej vrstvy pre analytické jednotky**:
* zabezpečenie vybraných datasetov a dátových zdrojov: ()
* zdieľanie analytických údajov,
* vrstva pre prístup k údajom,
* úložisko analytických údajov,
* prepojenie údajov na úrovni mikrodát,
* spracovanie údajov (využitie BI nástrojov, štatistických nástrojov, … ),
* podpora metód anonymizácie,
* publikovanie výsledkov analýz vo forme otvorených údajov.
* **Zavedenie SaaS nástrojov pre analytické spracovanie údajov** (pre inštitúcie, ktoré si chcú vytvárať vlastné riešenie):
* vizualizácia a publikácia výstupov,
* BI nástoje,
* matematické a štatistické nástroje,
* machine learning,
* nástroje na spoluprácu.

##### Dopadom projektu Konsolidovaná analytická vrstva bude:

* Výrazné zlepšenie možností a schopností analytických jednotiek vykonávať svoju činnosť: mať k dispozícii dáta v požadovanej kvalite a v požadovanom rozsahu, mať k dispozícii nástroje pre realizáciu analýz a publikovanie výsledkov,
* Podpora štátu fungujúceho na základe faktov a spracovania údajov: pri strategickom riadení a pri lepšom návrhu politík,
* Transformácia vybraných úsekov verejnej správy vďaka využitiu nástrojov umelej inteligencie.

## Dôvod

Štúdia uskutočniteľnosti vznikla, aby navrhla rámec pre projekt Konsolidovaná analytická vrstva (ďalej len KAV). Ide o zásadný krok k vytvoreniu možností pre lepšiu podporu využitia analytických údajov v rozhodovacích procesoch verejnej správy.

Dôvodom projektu KAV je poskytnutie:

* zabezpečenie kvalitných dát pre potreby analytických jednotiek,
* vytvorenia základnej dátovej infraštruktúry pre podporu analytického spracovania údajov vo verejnej správe,
* vytvorenie transparentného logovania všetkých prístupov k údajom,
* vytvorenie štatistických výstupov z operácií vykonávaných nad údajmi.

Implementáciu konceptu „Data-driven state“ (teda štátu, ktorý funguje na základe využívania dát) do praxe si vyžaduje výrazne zlepšenie využívania a spracovania údajov na analytické účely inštitúciami verejnej správy. Štát by mal prijímať rozhodnutia na základe najlepších znalostí, ktoré sú k dispozícii. Takáto transformácia si vyžaduje okrem riadenia životného cyklu dát aj zmenu spôsobu rozhodovania v samotných procesoch. Je potrebné zabezpečiť, aby inštitúcie vedeli, ako reálne používať údaje a tiež, aby rozhodovanie na základe údajov bolo možné (a kde sa dá i automatizované). Znamená to vytvorenie podmienok pre maximálne využitie potenciálu, ktorý je možné vyťažiť z údajov. Znamená to tiež transformáciu fungovania organizácií a procesov vo verejnej správe tak, aby boli tieto analýzy efektívne používané a zároveň zverejňované vo vhodnej vizuálnej podobe pre verejnosť.

V praxi to znamená, že inštitúcie verejnej správy musia mať dostatočnú kapacitu:

* Ľudské zdroje, ktoré vedia pracovať s údajmi,
* Nastavené procesy, ktoré podporujú vykonávanie analýz a rozhodovanie na základe analýzy dát,
* Technológie a algoritmy, vďaka ktorým je možné dáta spracovať,
* Prístup k zdrojom údajov (kvalitné údaje).

Pre jednotlivé oblasti, v ktorých je možné zlepšiť rozhodovanie (úseky výkonu verejnej správy) by preto mali byť zavádzané modely, dáta a nástroje, ktoré umožnia vytvoriť analýzy, na základe ktorých sa podporí rozhodovanie.

Transformácia inštitúcií verejnej správy by mala sledovať takzvané princípy Regulácie 2.0. Ide o nový prístup k výkonu verejnej správy, ktorého princípom je využívanie analytických údajov na podporu rozhodovania (najmä použitie analýzy rizík, odhaľovanie problémov a nesúladu s reguláciami v dátach a podobne). Prístup je tiež postavený na dynamickej interpretácii opatrení na základe sledovania dosahovania cieľov regulácie (vyhodnocovanie na základe analýzy údajov regulovaného prostredia). Základný prístup vo výkone verejnej moci sa tak posúva od udeľovania povolení k prevencii a predchádzaniu problémov. Systém postavený na princípoch Regulácie 2.0 je tak otvorenejší, flexibilnejší a má nižšiu administratívnu záťaž.

Účelom tohto projektu je preto pomôcť vybraným inštitúciám verejnej správy (ministerstvá, ústredné orgány štátnej správy a vybrané orgány verejnej správy) :

* Zabezpečiť základnú dátovú infraštruktúru pre analytické jednotky a ďalšie inštitúcie vo verejnej správe,
* Ponúknuť inštitúciám najmodernejšie analytické nástroje formou SaaS služieb,
* Zabezpečiť dostatočne kvalitné údaje, rýchlo a bez problémov (v rámci takzvanej konsolidovanej analytickej vrstvy),
* Implementovať pilotné prístupy k reforme fungovanie na základe lepšieho využitia údajov (v oblasti strategického riadenia štátu a vzdelávacích politík) a podporiť tak lepšie rozhodovanie na základe údajov v praxi.

Vďaka tejto aktivite budú mať všetky inštitúcie, ktoré majú záujem zlepšiť svoje fungovanie k dispozícii odbornú asistenciu, dáta a nástroje (takzvanú konsolidovanú analytickú vrstvu). Pre analytické jednotky (zákazníkov platformy) to znamená, že na požiadanie získajú potrebné dáta, na ktorých môžu postaviť svoje analýzy.

Riešenie priamo nadväzuje na nasledujúce strategické dokumenty:

* Strategická priorita Manažment údajov schválený dňa 28.2.2017, ktorý prijala Rada vlády pre digitalizáciu verejnej správy a digitálny jednotný trh,
* Strategická priorita Integrácia a orchestrácia schválená 28.2.2017, ktorý prijala Rada vlády pre digitalizáciu verejnej správy a digitálny jednotný trh,
* Strategická priorita Multikanálový prístup schválený 28.2.2017, ktorý prijala Rada vlády pre digitalizáciu verejnej správy a digitálny jednotný trh,
* Operačný program integrovaná infraštruktúra schválený Európskou komisiou dňa 28.10.2014,
* Národná koncepcia informatizácie verejnej správy schválená vládou 28.9.2016, kde sa definuje vízia, strategické ciele a smery e-Governmentu v SR,
* Národný program reforiem schválený uznesením vlády č. 198 z 24.4.2013, kde jedným z opatrení je modernizácia verejnej správy,
* Pozičný dokument Európskej komisie k vypracovaniu Partnerskej dohody a programov na roky 2014-2020, kde jednou z piatich priorít je moderná a odborná verejná správa,
* Strategický dokument pre oblasť rastu digitálnych služieb a oblasť infraštruktúry prístupovej siete novej generácie (2014 - 2020),
* Návrh centralizácie a rozvoja dátových centier v štátnej správe, ktorý bol schválený uznesením vlády SR č. 247/2014, pričom tento dokument ďalej rozpracováva časti popisujúce poskytovanie softvéru ako služby pre oblasť podporných a administratívnych činností vybraných subjektov verejnej správy.

## Rozsah

Štúdia uskutočniteľnosti vychádza z dokumentu Operačný program integrovaná infraštruktúra schválený Európskou komisiou dňa 28.10.2014 a nadväzuje na Strategické priority NKIVS schválené dňa 28.2.2017.

Predkladaný dokument popisuje:

* Legislatívu a jej potrebné zmeny, ktoré sú nevyhnutné pre implementáciu konceptu „Data-driven state“ a ktoré umožnia fungovanie Konsolidovanej analytickej vrstvy,
* Architektúru riešenia:
* Motivácia určuje základných používateľov, ich ciele, požiadavky a princípy,
* Biznis architektúra definuje biznis funkcie a biznis služby, ktoré budú ponúknuté ako v rámci Konsolidovanej analytickej vrstvy: aké dátové produkty pre verejnosť budú poskytované, aké prípady použitia budú podporované vďaka KAV, aké postupy dátovej vedy pre analytické jednotky budú umožnené, ktoré analytické jednotky sa zapoja do úvodného projektu, ako i procesy správy dátovej vrstvy a podporu pre manažment životného cyklu údajov.
* Architektúra informačných systémov znázorňuje kompozíciu Konsolidovanej analytickej vrstvy a integračné väzby systému s okolím: aké centrálne komponenty budú vytvorené a aké budú ich vlastnosti; aké nástroje budú nasadené vo vládnom cloude pre použite analytikmi a aké dátové zdroje budú do systému v rámci projektu zaradené.
* Keďže bude Konsolidovaná analytická vrstva umiestnená vo vládnom cloude, technologická architektúra definuje IaaS služby, ktoré sa plánujú využívať. V rámci možností sú stanovené i požiadavky na služby PaaS.
* V časti implementácia a migrácia sú vysvetlené základné etapy projektu KAV a ich súvislosť s postupom reformy, ktorú realizuje Dátová kancelária verejnej správy (v rámci projektu financovaného z OP EVS).
* Bezpečnostná architektúra definuje bezpečnostné požiadavky na Konsolidovanú analytickú vrstvu.
* Prevádzka riešenia popisuje, akým spôsobom bude zabezpečená podpora používateľov a inovácia systému KAV,
* V časti Ekonomická analýza sú kvantifikované prínosy a náklady, ktoré si realizácia cieľov reformy a implementácie projektu KAV vyžiadajú. Ich následná analýza dáva odpoveď o ekonomickej výhodnosti riešenia. V rámci časti sú špecifikované indikatívne náklady pre realizáciu systému KAC.
* Pre každú oblasť architektúry sú identifikované kritéria kvality, na základe ktorých je možné posudzovať návrhy a alternatívne riešenia. Obdobne sú identifikované riziká, ktoré bude potrebné v nasledujúcom období počas prípravy a realizácie projektu eliminovať.
* Štúdia uskutočniteľnosti analyzuje 4 základne alternatívy, ako je možné realizovať analytický systém a vyberá vhodné riešenie:
* Distribuované riešenie – pokračovanie v súčasnom stave,
* Outsourcing analytickej vrstvy súkromnej spoločnosti,
* Analytický market,
* Centrálne riešenie.

## Použité skratky a značky

Tabuľka 2: Dôležité termíny a pojmy

| Odborný termín | Definícia |
| --- | --- |
| Big Data | Pojem Big Data predstavuje obrovské množstvo dát rôzneho typu získavaných s vysokou rýchlosťou z vysokého počtu rôznych typov zdrojov. Big Data sa vyznačujú 4 charakteristikami: Objem, Rôznorodosť (napríklad videá, tweety, dáta zo senzorov), Rýchlosť (spracovávanie tokov dát v reálnom čase) a Hodnovernosť (s množstvom zbieraných dát sa spája neistota o ich presnosti). Spracovávanie takého množstva rôznorodých dát zbieraných v reálnom čase si vyžaduje nové nástroje a metódy ako výkonné procesory, nové softvérové riešenia a algoritmy. |
| Dataset | Ucelená a samostatne použiteľná skupina súvisiacich údajov vytvorených a udržiavaných na určitý účel a uložených spoločne podľa rovnakej schémy. |
| Dátový zdroj | Pôvodné miesto evidencie datasetu. Dátové zdroje verejnej správy sú vytvárané za účelom podpory služieb verejnej správy, služieb vo verejnom záujme alebo verejných služieb. Jednotlivé agendy súvisiace s výkonom verejnej správy môžu obsahovať viaceré dátové zdroje. Charakteristickým znakom dátového zdroja je jeho samostatná použiteľnosť (výpovedná hodnota) nezávisle od iných dátových zdrojov. Ďalšími znakmi popisujúcimi dátový zdroj sú názov, účel (zameranie), typy spracúvaných údajov a vzťahy medzi nimi, formát údajov a podobne – tzv. metadáta. Dátovým zdrojom môžu byť aj zariadenia produkujúce dáta v internete vecí, senzory alebo aplikačné rozhrania sociálnych sietí. |
| Metadáta | Štruktúrované údaje obsahujúce informácie o primárnych údajoch – dátovom zdroji, pričom primárne údaje spravidla reprezentujú určitý hmotný objekt alebo nehmotný objekt. sú určené najmä na vyhľadávanie, katalogizáciu a využívanie primárnych údajov. |
| Mikrodáta | V kontexte štatistiky predstavujú mikrodáta informáciu na úrovni individuálnych respondentov, ktorí sa zúčastňujú prieskumov alebo sčítaní ľudu. |
| Paradáta | Paradáta sú dáta dokumentujúce proces zberu údajov. V prípade vykonania prieskumu môže ísť o informácie zahŕňajúce časy, kedy sa vykonávali pohovory, dĺžku trvania pohovorov, spôsob komunikácie, ochotu respondenta a mnohé ďalšie. |
| Prístup cez aplikačné rozhranie - API | Štandardizovaný spôsob vzájomnej strojovej komunikácie, charakteristický určením koncového bodu poskytovania služby, množinou funkcií v tomto bode dostupných, ich parametrami a formátmi vymieňaných správ. |
| Jednotný referencovateľný identifikátor | Jednotný referencovateľný identifikátor (URI) je množina znakov tvoriacich jednotným spôsobom identifikátor ľubovoľných logických a fyzických objektov. |
| Surové dáta | Surové dáta, tiež nazývané primárne dáta, sú dáta získané priamo z dátového zdroja, ktoré nie sú spracované na použitie pre určitý konkrétny účel (ešte neprešli spracovaním ani žiadnou inou manipuláciou). |
| Dátová veda | Rozvíjajúci sa vedecký odbor, ktorý sa zameriava na získavanie použiteľných znalostí a informácií z veľkých datasetov. Dátová veda sa venuje štrukturovaným aj neštrukturovaným dátam a aplikuje rôzne prístupy, ako napríklad strojové učenie. |
| (Data mining) | Proces extrakcie a získania dôležitých vzorov z datasetov, aby bol vyriešený dobre definovaný problém. Tento proces je popísaný napríklad v metodike CRISP-DM[[1]](#footnote-2) (Cross-industry standard process for data mining), ktorý definuje rozsah pre „data-miningový projekt“. |
| Randomized control trials – RCT | Randomizovaná kontrolovaná štúdia je typ vedeckého experimentu, ktorého cieľom je znížiť skreslenie. Účastníci sú náhodne rozdelení do skupín, pričom pre každú skupinu sa vykonáva iná intervencia. |
| Strojové učenie (Machine learning) | Odbor počítačovej vedy, ktorý sa venuje návrhu a hodnoteniu algoritmov, ktoré dokážu v dátových množinách rozpoznať významné vzory. Algoritmus strojového učenia na vstupe spracuje dataset a vracia model, ktorý obsahuje vzor nachádzajúci sa v dátach. |

Tabuľka 3: Skratky a značky

| Skratka / Značka | Vysvetlenie |
| --- | --- |
| API | Application Platform Interface, Rozhranie aplikačnej platformy |
| AS IS | Aktuálny stav bez realizácie projektu |
| CBA | Nákladovo-výnosová analýza |
| DFŠ | Detailná funkčná špecifikácia |
| DPH | Daň z pridanej hodnoty |
| DWH | Data warehouse, úložisko údajov |
| eGov | eGovernement |
| eID | Elektronické identifikačné číslo |
| ENPV | Čistá súčasná ekonomická hodnota |
| EP | Európsky parlament |
| ETL | Extract, Transform, Load, Extrahovať, transformovať, načítať |
| EÚ | Európska únia |
| EUR, € | Mena EURO |
| G2B | Služby pre podnikateľov (Government to Business) |
| G2C | Služby pre občanov (Government to Citizens) |
| G2G | Služby pre verejnú správu, komunikácia systémov verejnej správy bez zásahu človeka (Government to Government) |
| GDPR | General Data Protection Regulation, 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 |
| GUI | Grafické používateľské rozhranie (Graphic User Inreface) |
| HW | Hardvér (Hardware) |
| IaaS | Infrastructure as a Service (Infraštruktúra ako služba) |
| ID | Identifikačné číslo |
| IKT | Informačné komunikačné technológie |
| IS | Informačný systém |
| IS CSRÚ | Informačný systém Centrálnej správy referenčných údajov |
| ISO | International Organization for Standardization |
| IT | Informačné technológie |
| ITIL | Information Technology Infrastructure Library |
| IS VS | IS verejnej správy |
| JSON | JavaScript Object Notation, Označenie objektu JavaScript |
| KPI | Key performance indicators, Kľúčové indikátory výkonnosti |
| LAN | Local area network |
| MDM | Master data management, Správa hlavných údajov |
| MDaV SR | Ministerstvo dopravy a výstavby SR |
| MF SR | Ministerstvo financií SR |
| MOU | Manažment osobných údajov |
| MS SR | Ministerstvo spravodlivosti Slovenskej republiky |
| MŠVVaŠ SR | Ministerstvo školstva, vedy, výskumu a športu SR |
| MV SR | Ministerstvo vnútra SR |
| MPaRV SR | Ministerstvo práce, sociálnych vecí a rodiny Slovenskej republiky |
| N/A | Not applicable, neaplikovateľné |
| NCZI SR | Národné centrum zdravotníckych informácií |
| NKIVS | Národná koncepcia informatizácie verejnej správy |
| NLZ | Návrh legislatívnych zmien |
| NPV | Čistá súčasná hodnota (Net Present Value) |
| OP EVS, OPEVS | Operačný program Efektívna verejná správa |
| OP II, OPII | Operačný program Integrovaná infraštruktúra |
| OPIS | Operačný program Informatizácia spoločnosti |
| PAT | Plán akceptačných testov |
| PaaS | Platform as a Service (Platforma ako služba) |
| PBP | Rok návratu investície |
| PIMS | Personal Information Management System, Manažment osobných údajov |
| REST | Representational State Transfer architectural style for distributed hypermedia systems, Reprezentatívny štatút pre štrukturálny štýl prenosu pre distribuované hypermedia systémy |
| RFO | Register fyzických osôb |
| ROI | Návratnosť investícií (Return of Investment) |
| RPO | Register právnických osôb a podnikateľov |
| RZ | Reformný zámer |
| RUP | Rational Unified Process |
| SAN | Storage area network |
| SaaS | Software as a Service (Softvér ako služba) |
| SLA | Service level agreement |
| SOA | Servisne orientovaná architektúra (Service Oriented Architecture) |
| SR | Slovenská republika |
| ŠU | Štúdia uskutočniteľnosti |
| SW | Softvér (Software) |
| TO BE | Cieľový stav po realizácii projektu |
| TOGAF | The Open Group Architecture Framework |
| TCO | Celkové náklady na vlastníctvo (Total Cost of Ownership) |
| URI | Uniform Resource Identifier, Identifikátor jednotného zdroja |
| ÚGKK | Úrad geodézie, kartografie a katastra SR |
| ÚPSVaR | Úrad práce, sociálnych vecí a rodiny |
| ÚPPVII, ÚPVII, ÚPPVIaI | Úrad podpredsedu vlády Slovenskej republiky pre investície a informatizáciu |
| ÚPVS | Ústredný portál verejnej správy |
| ÚV SR | Úrad vlády Slovenskej republiky |
| ÚVO | Úrad pre verejné obstarávanie |
| VO | Verejné obstarávanie |
| VS | Verejná správa |
| Z.z. | Zbierka zákonov |
| ŽoNFP, ŽNFP | Žiadosť o nenávratný finančný príspevok |

# Manažérske zhrnutie

Údaje sa stávajú “strategickou surovinou” a úspešné štáty musia fungovať na základe využívania znalostí a zaviesť metódy dátovej vedy do svojho fungovania

Výrazne lepšie využívanie dát vo verejnej správe predstavuje kľúčový cieľ programového obdobia 2014 až 2020. K dátam sa pristupuje ako k vzácnemu zdroju. Dátovej vrstve je preto v architektúre informačných systémov verejnej správy venovaná výrazná pozornosť. Hlavným zámerom bude dátová integrácia medzi jednotlivými systémami verejnej správy a zabezpečenie pravidelnej replikácie kvalitných a konsolidovaných transakčných dát do dátových úložísk. V oblasti informačných technológií dochádza k nástupu „Big data“ – kapacity zbierať, spracovávať a analyzovať obrovské množstvá údajov nielen dávkovo, ale aj v reálnom čase. Tento fenomén transformoval mnohé oblasti ekonomiky a z revolúcie informačných technológií sa stáva revolúcia v oblasti využívania dát. Nastáva teraz príležitosť, navrhnúť podobnú transformáciu aj vo verejnej správe: navrhnúť a realizovať iniciatívy, ktoré umožnia využiť potenciál lepších dát.

Lepšie dáta znamenajú možnosť získavať kvalitné informácie, z nich vyplývajúce „insights“ (pohľady dovnútra problematiky), ktoré zas ďalej slúžia ako podklady pre tvorbu znalostí a lepšie rozhodovanie. Ak sa bude vo verejnej správe lepšie rozhodovať, pozitívne sa to prejaví na výsledkoch vládnutia a stave jednotlivých sektorov verejnej správy. Je preto potrebné podrobne, aby ľudia vo verejnej správe vedeli pracovať s dátami: rozumieť logike (ontológiám) spravovaných dát, vytvárať a udržiavať dátové modely, katalogizovať metadáta a paradáta, chápať prepojenia medzi jednotlivými systémami a podporovať zvyšovanie kvality dát ukladaných vo verejných databázach. Inštitúcie tiež musia zmeniť a reformovať svoje fungovanie, aby sa výsledky prejavili v praxi: či už vďaka lepším predikciám a modelom bude možné zefektívniť súčasne činnosti (napríklad efektívnejšie plánovanie siete škôlok, optimálnejšie zimná údržba), alebo vďaka dátam sa môže zmeniť celý model fungovania inštitúcie (strategické plánovanie na základe budúcich scenárov).

Počas posledných rokov sa postupne rozšíril rozsah techník, ktoré je možné vnímať ako dátovú vedu, k dispozícii sú nové nástroje pre analyzovanie údajov ako i nove typy dát a dátových zdrojov. V dnešnej dobe sa tieto nástroje využívajú najmä v akademickom, ale i súkromnom sektore. Hlavnou ambíciou tejto štúdie je navrhnúť vhodné riešenie a postup pre maximálne využitie dát vo verejnom sektore a overiť nové spôsoby práce založenej na tejto dátovej vede priamo vo verejnej správe. Inšpiráciou sú najproduktívnejšie prípady použitia dátovej vedy vo verejnom sektore v zahraničí ale i spôsob využívania dát vo verejnom sektore (napríklad dátovej vede sa venuje britský Behavioural Insights Team, ktorý vydal správu Použitie dátovej vedy pri návrhu politík[[2]](#footnote-3)).

Zlepšenie využívania dát v analytických úlohách je len jednou z tém, ako dokáže verejný sektor lepšie pracovať s dátami.

Koncepčný prístup k manažmentu údajov je hlavnou témou rozvoja e-Governmentu do roku 2020

Obrázok 1: Koncept manažmentu údajov vo verejnej správe



Údaje vo verejnej správe budú manažované systematicky. Riadenie údajov bude riešené centrálne, v kompetencii ÚPPVII, kde za týmto účelom bolo vytvorené nové organizačné oddelenie: **Dátová kancelária verejnej správy**. Za správu údajov a ich kvalitu naďalej zodpovedá príslušná inštitúcia verejnej správy. Navrhnuté opatrenia zabezpečia:

* **Zdieľanie referenčných údajov:** aby v konaniach verejnej správy boli k dispozícií všetky údaje, ktoré sú potrebné.
* **Manažment kvality údajov:** aby sme správne údaje a bolo možné sa spoľahnúť na ich správnosť.
* **Analytické využitie údajov:** aby verejná správa dokázala využívať svoje údaje pre potreby prípravy analýz (analytické spracovanie údajov), ktoré budú slúžiť ako podklad pre lepšie rozhodovanie. Počíta sa s využitím virtuálnych databáz a nástrojov pre anonymizáciu, pseudoanymizáciu (ktoré zabezpečia, že údaje osobného charakteru bude možné používať na analytické účely bez narušenia súkromia).
* **Manažment osobných údajov:** že podporujeme transparentnosť a „zmocniť“ občana/podnikateľa k rozhodovaniu o tom, ako budú využívané jeho dáta, možnosť opraviť údaje, preto navrhujeme službu Moje dáta: každý subjekt (občan a podnikateľský subjekt) získajú prístup k údajom, ktoré verejná správa o ňom eviduje, mohol s nimi v rámci možností manipulovať a zároveň vidieť, kto a prečo k takýmto údajom pristupoval.
* **Publikovanie otvorených údajov:** aby verejnosť mala prístup k údajom verejnej správy v otvorenom formáte vhodnom na opätovné použitie – okrem osobných údajov, citlivých údajov a utajovaných údajov.

##### Záber predkladanej štúdie pokrýva časť Analytické využitie údajov

Implementácia konceptu „Data-driven state“ (teda štátu, ktorý funguje na základe využívania dát) do praxe si vyžaduje výrazne zlepšenie využívania a spracovania údajov na analytické účely inštitúciami verejnej správy. Štát by mal prijímať rozhodnutia na základe najlepších znalostí, ktoré sú k dispozícii. Takáto transformácia si vyžaduje okrem riadenia životného cyklu dát aj zmenu spôsobu rozhodovania v samotných procesoch. Je potrebné zabezpečiť, aby inštitúcie vedeli, ako reálne používať údaje a tiež, aby rozhodovanie na základe údajov bolo možné (a kde sa dá i automatizované). Znamená to vytvorenie podmienok **pre maximálne využitie potenciálu, ktorý je možné vyťažiť z údajov**. Znamená to tiež transformáciu fungovania organizácií a procesov vo verejnej správe tak, aby boli tieto analýzy efektívne používané a zároveň zverejňované vo vhodnej vizuálnej podobe pre verejnosť.

Tabuľka 4: Logické komponenty Konsolidovanej analytickej vrstvy



Pre jednotlivé oblasti, v ktorých je možné zlepšiť rozhodovanie (úseky výkonu verejnej správy) by preto mali byť zavádzané modely, dáta a nástroje, ktoré umožnia vytvoriť analýzy, na základe ktorých sa podporí rozhodovanie.

Za základné oblasti, v ktorých sa môže využívať systematická analýza dát, je možné považovať:

* Strategické priority – príprava stratégií, analýza súčasného stavu a porovnávanie variantov a následné sledovanie napĺňania strategických priorít,
* Politiky – vyhodnocovanie dopadov politík, simulácie rozhodnutí,
* Regulácie – Hodnotenie dopadov regulácií a posudzovanie vplyvov návrhov (Regulatory impact assessment - RIA),
* Investície – vyhodnocovanie dopadov projektov a verejných investícií,
* Výkonnosť a kvalita – vyhodnocovanie výkonu verejnej správy pri poskytovaní služieb v jednotlivých sektoroch,
* Poskytovanie služieb – zlepšenie kvality služieb vďaka využitiu dát, napríklad pomocou zavádzania automatizovanej obsluhy, automatizovaného spracovania podaní, či vytvárania znalostnej bázy pre úradníkov.
* Podpora rozhodovania v operatívnych procesoch (analýzy rizík, spracovanie a posudzovanie podaní, fraud management, prediktívne analýzy a podobne).

Jednotlivé inštitúcie sa podľa svojho zamerania venujú rôznym z týchto oblastí. Spôsob požiadaviek a potrieb pre jednotlivé oblasti analýzy sa tiež líši, preto pri realizácii tejto aktivity pôjde o hľadanie vhodnej kombinácie nástrojov pre jednotlivé inštitúcie. Recept na úspech tiež nie je možné určiť hneď na začiatku, ale je potrebné experimentovať a postupne zlepšovať rozsah a šírku analýz a šíriť najlepšie skúsenosti.

##### „Analytický market“ a prípady použitia

Aby bolo možné účinne takýto zámer podporiť, je potrebné vybrať vhodný strategický prístup. V rámci štúdie boli navrhnuté rôzne koncepty, na základe odpovedí na otázky, ako nastaviť manažment analytickej vrstvy, akým spôsobom analytickú vrstvu vybudovať, ako zabezpečiť ochranu osobných údajoch pri ich analytickom spracovaní a akým spôsobom zabezpečiť potrebné analytické nástroje a dáta. Z kombinácií rôznych odpovedí sme definovali štyri alternatívy a víťazný model nazývame „Analytický market“:

* Konsolidovanú analytickú vrstvu bude spravovať verejná inštitúcia – Dátová kancelária verejnej správy.
* Dátová kancelária verejnej správy vytvorí otvorený ekosystém dát a aplikácií pre potreby analytických jednotiek. Pôjde súbor nástrojov a možností, z ktorých si budú môcť analytické jednotky nakombinovať systém podľa svojich potrieb a želaní. K dispozícii budú vo forme SaaS služieb nástroje pre štatistické analýzy, strojové učenie, ako R studio a Matlab, ako i klasické BI nástroje. Ekosystém nebude staticky, ale bude postupne narastať.
* Do analytickej vrstvy sa postupne integrujú významné zdroje dát, pričom budú prístupné všetkým používateľom. Dáta budú prepojené na „mikroúrovni“, pričom identita subjektov nebude prístupná, k dispozícii budú len pseudoanonymizované dáta.

Konsolidovaná analytická vrstva tak bude verejný otvorený technologický priestor. Ďalším dôležitým aspektom riešenia je orientácia na konkrétne prípady použitia, ktoré budú realizované dátovými projektami a ponúknu ako výstupy takzvané dátové produkty, s ktorými sa bude dať pracovať. Očakáva sa, že po skončení projektu budú nové prípady použitia neustále pribúdať.

Obrázok 2: Základná štruktúra podporovaných dátových projektov



Prechádzajúci obrázok znázorňuje jednotlivé vrstvy Konsolidovanej analytickej vrstvy vo vzťahu k realizácii dátového projektu. V prvom rade je potrebné, aby existoval tím, **analytická jednotka**, ktorá dokáže organizačne a odborne problém zastrešiť. Potom je potrebné pomenovať jasný **prípad použitia**:

* definovať oblasť verejnej politiky, v ktorej chceme dosiahnuť zlepšenie vďaka analytickému využitiu dát,
* definovať ciele intervencie a ukazovatele, ktoré budú v rámci prípadu použitia sledované,
* stanoviť potrebné procesné zmeny, pre zavedenie výsledkov analýzy do rozhodovacích procesov,

Ako náhle je jasný prípad použitia, je možne identifikovať **analytické modely**, ktoré sa pre tento účel v KAV vytvoria, vrátane definovaných výstupov, ktoré nazývame Dátové produkty. Pre prípravu analytických produktov budú k dispozícii **analytické nástroje** a informačné systémy a v neposlednom rade je nevyhnutné zabezpečiť potrebné **dáta**, či už nákupom, alebo integráciou na administratívne zdroje.

Takýto návrh je relatívne praktický, pretože umožňuje špecializáciu pri využití analytických metód a technológií. Analytické jednotky, organizované flexibilne so zameraním na riešenie problémov a experimentovanie sa môžu venovať podpore prípadov použitia a vytváraniu analytických modelov. Technickú stránku veci, akou je zabezpečenie kvalitných dát, dátovej infraštruktúry a vhodných nástrojov, bude riešiť Dátová kancelária.

##### Implementácia Konsolidovanej analytickej vrstvy a výsledky projektu

Implementácia Konsolidovanej dátovej vrstvy sa skladá z dvoch základných častí:

* **Vybudovanie dátovej infraštruktúry**: vytvorenie centrálneho dátového skladu, nasadenie analytických nástrojov vo forme dostupných SaaS služieb, konfigurácia riešenia, integrácia s Centrálnou integračnou platformou (CIP). Centrálna integračná platforma predstavuje implementáciu modulu procesnej a dátovej integrácie v rámci architektúry informačného prostredia verejnej správy.
* **Overenie vybraných prípadov použitia:** riešenie je postavené na zavádzaníanalytických metód do praxe. Boli vybrané 4 analytické jednotky a ich prípady použitia, ktoré budú implementované v rámci KAV, čo znamená: pre ktoré bude realizovaný dátový projekt s očakávaným výstupom vo forme dátového produktu. V rámci projektu tak budú vytvorené príslušné analytické modely, zabezpečenie potrebných dát a integrácia dátových zdrojov a zavedenie výsledkov do rozhodovacích procesov.

Ako pilotné analytické jednotky boli zvolené: Inštitút pre stratégie a analýzy (ISA), Inštitút vzdelávacej politiky, Infostat a Dátová kancelária verejnej správy. Ďalšie sa budú môcť postupne pripájať, aby získali prístup k analytickým nástrojom a zdieľaným dátam a mohli realizovať svoje dátové projekty (ktoré bude možné financovať z dopytovej výzvy OPII).

Nasledujúce tabuľky predstavujú podrobný prehľad výstupov tohto projektu:

* Zapojené analytické jednotky a realizované prípady použitia v praxi,
* Prehľad procesov, ktoré bude Konsolidovaná analytická vrstva podporovať,
* Nasadené prvky dátovej infraštruktúry,
* Dátové zdroje dostupné po skončení projektu.

Tabuľka 5: Využitie Konsolidovanej analytickej vrstvy v praxi a realizované prípady použitia

|  |  |  |
| --- | --- | --- |
| Organizačná jednotka | Prípad použitia | Dátový produkt/Výsledok |
| Inštitút pre stratégie a analýzy (ISA) | Monitorovanie strategických cieľov štátu | Dashboard strategického riadenia štátu:. |
| Plnenie programového vyhlásenia |
| Strategické plánovanie rozvoja Slovenska | Scenáre budúceho vývoja (simulácie): |
| Národná inovačná stratégia. |
| Hodnotenie vplyvu EŠIF a vládnych politík | Kontrafaktuálne hodnotenia vplyvu politík a intervencií  Kontrafaktuálne hodnotenie podpory z EŠIF |
| Inštitút vzdelávacej politiky | Hodnotenie výkonnosti vzdelávacích inštitúcií . identifikácia problémov a nedostatočne výkonných inštitúcií | Benchmark výkonnosti vzdelávania pre jednotlivé inštitúcie  Predikcia budúcej výkonnosti |
| Predpovedanie budúcich výsledkov žiakov | Globálny report o výsledkoch žiakov  Aplikácia pre individuálny odhad výsledkov |
| Predvídanie budúceho dopytu po službách (kapacitách vzdelávacích inštitúcií) | Mapa dopytu po vzdelávacích službách |
| Prerozdeľovanie študentov do 1. ročníkov SŠ | Návrh počtov žiakov prvého ročníka pre jednotlivé študijné odbory SŠ |
| Vzdelávanie ľudí s predčasne ukončenou školskou dochádzkou | Absolventská miera nezamestnanosti  Predikcia absolventskej miery nezamestnanosti |
| Infostat | Krátkodobé predikcie vývoja slovenskej ekonomiky | Krátkodobá predikcia vývoja slovenskej ekonomiky (analyticko-prognostický dokument) |
| Rýchle odhady HDP a celkovej zamestnanosti v hospodárstve | Odhad rastu a štruktúry HDP (výsledky s analytickým komentárom)  Odhad zamestnanosti, jej rastu a štruktúry (výsledky s analytickým komentárom) |
| Krátkodobé predikcie vývoja (indexov) priemyselnej produkcie | Predikcia vývoja indexov priemyselnej produkcie  (prognostický dokument) |
| Populačné prognózy | Regionálne populačné prognózy  Odvodené populačné prognózy |
| Dátová kancelária verejnej správy | Manažment kvality údajov vo verejnej správe | Benchmark dátovej kvality |

Tabuľka 6: Prehľad procesov, ktoré bude Konsolidovaná analytická vrstva podporovať

|  |  |  |
| --- | --- | --- |
| Organizačná jednotka | Funkcia | Procesy/Služby |
| Dátová kancelária verejnej správy | Manažment (Riadenie) analytickej vrstvy | Zabezpečenie zdieľania údajov medzi analytickými jednotkami  Zabezpečenie zberu údajov  Riadenie pripájania dátových zdrojov  Čistenie dát a dátová integrácia  Transformácia dát do potrebnej štruktúry a prepájanie údajov  Ochrana citlivých údajov  Anonymizácia a pseudonymizácia  Kontrola kvality a duplicít  Zabezpečenie verzionovania  Manažment úložísk  Archivovanie  Aplikovanie rôznych politík na ochranu citlivých údajov |
| Analytická jednotka | Použitie metód Dátovej vedy | Formulácia hypotéz  Správa analytického „sandboxu“  Výber vhodných dátových zdrojov  Zhromažďovanie a získavania dát:  Skladovanie dát  Hladanie informácií a vzťahov medzi dátami  Výber vhodných analytických metód:  Tvorba analytického dátového modelu:  Testovanie hypotéz (test relevantnosti a test istoty):  Hodnotenie výsledkov:  Komunikácia výsledkov a pripomienkovanie |

Tabuľka 7: Architektúra Konsolidovanej analytickej vrstvy: Nasadené prvky dátovej infraštruktúry

|  |  |
| --- | --- |
| Komponenty | * Centrálny dátový sklad |
| Nástroje | * tradičné BI nástroje * „machine learning“ * štatistické nástroje * nástroje pre simulácie“ * vizualizácia dát a publikácia: služby * publikácia otvorených údajov |

Tabuľka 8 Dátové zdroje dostupné po skončení projektu

| Zdroj | Gestor | Dátové objekty |
| --- | --- | --- |
| Daňový informačný systém | Finančná správa SR, | Register daňových subjektov |
| Zdravotnícka štatistika | NCZI | Poskytnutá zdravotná starostlivosť |
| Rezortný informačný systém | MŠVVaŠ SR | Centrálny register študentov |
| Testovanie NÚCEM | MŠVVaŠ SR | Údaje z testovaní NÚCEM (Testovanie 9, Testovanie 5, Maturita) |
|  | CVTI | Údaje z výkazov zbieraných CVTI. |
| IS Registra fyzických osôb | Ministerstvo vnútra SR -; | Fyzické osoby  Matrika (narodenia, úmrtia) |
| Evidencia nezamestnaných | Ministerstvo práce sociálnych vecí a rodiny SR | Nezamestnanosť/Evidencia nezamestnaných |
|  | Ministerstvo práce sociálnych vecí a rodiny SR/ Ústredie práce sociálnych vecí a rodiny SR | Zamestnanosť |
| Manažment údajov Sociálnej poisťovne | Sociálna poisťovňa | Sociálne dávky  Nemocenské dávky  Dávky v nezamestnanosti  Výber odvodov |
|  | Telekomunikačný operátori | Telekomunikačné údaje |
| ITMS+ | Úrad vlády SR | Údaje o štrukturálnych fondoch a monitorovanie ich výkonnosti  Údaje o výzvach  Údaje o projektoch  Údaje o financovaní |

# Motivácia

### Súhrnný popis

Lepšie rozhodovanie je možné dosiahnuť prostredníctvom aplikácie metód dátovej vedy v procesoch verejnej správy a odhaľovať tak nové poznatky a umožniť inštitúciám zefektívniť svoje fungovanie, lepšie navrhovať politiky, lepšie posudzovať vplyvy opatrení a investícií a riadiť sektor na základe reálnych faktov a skutočností.

Ambíciou projektu je podporiť dátovú kanceláriu a vybrané analytické jednotky potrebnými technológiami a nástrojmi, ktoré im umožnia plnohodnotne realizovať analytické úlohy. Zároveň za dôležité považujeme zabezpečiť potrebné kvalitné údaje z administratívnych a z externých zdrojov a zavedenie manažmentu prístupov k analytickým údajom.

V rámci projektu sa odskúšajú základné prípady použitia využitia dátovej vedy vo verejnej správe a tým sa nastaví priestor na ďalšiu digitálnu transformáciu.

Tabuľka 9: Nastavenie biznis cieľov riešenia a merateľných ukazovateľov pre projekt

| Ciel | Ukazovateľ cieľa | Spôsob dosiahnutia stanoveného cieľa | Výsledok | Obdobie |
| --- | --- | --- | --- | --- |
| **OPII 7.7 Umožnenie modernizácie a racionalizácie verejnej správy IKT prostriedkami** | Počet úsekov verejnej správy, v ktorých je rozhodovanie podporované analytickými systémami |  | 20 | 2021 |
| Počet dodatočných centrálne využitých podporných systémov vnútornej správy v rámci ISVS (ako služieb v cloude SaaS) |  | 4 | 2021 |
| **Zlepšiť rozhodovanie vo verejnej správe** | Analytické jednotky podporené riešením konsolidovanej analytickej vrstvy |  | 4 | 2021 |
| Počet prípadov použitia podporených analytickým spracovaním dát |  | 14 | 2021 |
| Počet realizovaných RCT |  | 2 | 2021 |
| Počet analytických výstupov generovaných v analytickej vrstve |  | 20 | 2021 |
| **Zvýšiť dostupnosť dát pre analytické spracovanie** | Počet pripojených dátových zdrojov (vo formáte umožňujúcom strojové spracovanie) |  | 30 | 2021 |

Koncept predstavený v strategickej priorite Manažment údajov predstavuje aplikáciu myšlienky „Data-driven state“ štátu – štátu fungujúcom na základe využívania dát a presnom riadení celého životného cyklu dát.

##### Prehľad hnacích faktorov („drivers“) pre motiváciu realizovať projekt

* Inštitúcie verejnej správy:
* Rozhodovať sa na základe dát a najlepších dôkazov,
* Byť schopný reagovať na výzvy a hrozby, chápať súvislosti vďaka analýze dát,
* Byť schopný využívať nové technológie a metódy práce,
* Optimalizovať svoje fungovanie (zvyšovať svoju efektivitu a produktivitu práce, schopnosť riešiť problémy).
* Analytické jednotky:
* Byť schopný predpovedať budúci vývoj,
* Mať k dispozícii kvalitné dáta bez prekážok,
* Mať k dispozícii potrebné nástroje pre dátové analýzy
* Spolupracovať a zdieľať výsledky analýz.
* Dátová kancelária verejnej správy:
* Spravovať analytickú platformu, ktorá spojí analytické jednotky, poskytovateľov dát a analytické nástroje,
* Prehľad o využívaní dát vo verejnej správe.
* Verejnosť
* Ochrana osobných údajov pri analytickom spracovaní dát pre verejný záujem.
* Mať prehľad o vykonávaných analýzach,
* Mať prehľad o predpokladaných dopadoch politík a regulácií,
* Mať prehľad o zdôvodnení investícii vo verejnom záujme,
* Mať prehľad o výkone verejnej správy v jednotlivých sektoroch (až na úroveň jednotlivých inštitúcií) – vrátane ratingov,
* Možnosť participatívnej účasti na návrhu politík, regulácií a inovácii služieb.
* Malí a strední podnikatelia:
* Dopyt po inovatívnych riešeniach s využítím dátovej vedy a umelej inteligencie,
* Mať prístup ku kvalitným analýzam a prognózam o vývoji ekonomiky.

##### Prečo realizovať projekt práve teraz:

* Zjednodušili sa možnosti získavať dáta vo veľkom objeme (z dát sa stáva komodita),
* Rozvoj algoritmov dátovej vedy, najmä strojového učenia a „deep learning“ za posledných 10 rokov,
* Dátová veda priniesla reálne výsledky v mnohých oblastiach hospodárstva a verejná správa sa môže učiť a aplikovať najlepšie skúsenosti,
* Výrazne klesla cena počítačového výkonu a úložiskovej kapacity potrebných na beh algoritmov dátovej vedy.

### Riziká

* Ochota a politická vôľa realizovať transformáciu fungovania inštitúcií so zámerom lepšieho využívania dát, bude chýbať motivácia realizovať reformy na lokálnej úrovni,
* Kapacita a ľudské zdroje potrebné pre efektívne využitie navrhovaných nástrojov a dát nebudú dostatočné, riešenie sa nebude využívať.
* Nedostatočná spolupráca, nevznikne ekosystém a dobrá prax sa nebude šíriť mezi jednotlivými inštitúciami.

# Popis aktuálneho stavu

## Legislatíva

##### Aké legislatívne prekážky sú kladené analytickému spracovaniu údajov?

* Neadekvátna úroveň regulácie čo sa týka hlavne oblasti zberu, produkcie a diseminácie dát,
* Zákon o ochrane osobných údajov a GDPR bránia využitiu dát (údaje nemôžu byť poskytnuté, jednotlivé databázy nie sú prepojené),
* V prípade Sociálnej poisťovne nie je zákonný nárok na prístup k osobným údajom (v školskom zákone a predovšetkým zákone o sociálnom poistení). V prípade Evidencie uchádzačov o zamestnanie sa aktuálne vedú rokovania. V prípade RFO sme rokovania z kapacitných dôvodov ani nezačali.

## Architektúra

### Biznis architektúra

##### Analytické jednotky

Analytické jednotky sú v súčasnosti zriadené len na niektorých ministerstvách a ich výkonnosť je až na výnimky (IFP, čiastočne IVP) obmedzená, čo je dané hlavne personálnymi kapacitami a kvalitou dostupných nástrojov a dát. Metodika posudzovania vplyvov sa dodržuje formálne a analýzy nie sú podrobované kontrole kvality. Vďaka iniciatívam reformy verejnej správy vyjadrených v Operačnom programe Efektívna verejná správa dochádza postupne k zmene tohto stavu a plánuje sa posilnenie agendy lepšej regulácie a vybudovanie analytických jednotiek.

V nasledujúcej tabuľke nájdete analýzu súčasného stavu vybraných analytických jednotiek, ktoré budú zapojené do navrhovaného projektu.

Tabuľka 10: Inštitút pre stratégie a analýzy (ISA)

| Analytická jednotka | Inštitút pre stratégie a analýzy (ISA) |
| --- | --- |
| Zaradenie | Úrad vlády SR |
| Počet zamestnancov aktuálny (1Q-2019) | 8 (plánujú sa obsadiť ďalšie 2 miesta) |
| Počet zamestnancov plánovaný (1Q-2020) | 12 |
| Politiky, ktoré analyzuje/ ovplyvňuje analytická jednotka | Vzhľadom na nadrezortný charakter ÚV SR sa analýzy týkajú verejnej politiky vo všeobecnosti. V minulosti ISA analyzoval hospodársku politiku SR a EÚ, regionálnu politiku, vzdelávaciu politiku, zdravotnú politiku a inovačnú politiku. |
| Finančný rozsah riešených politík | Celý štátny rozpočet |
| Využívané dátové zdroje | Niektoré rezortné databázy (ITMS, NÚCEM, Sociálna poisťovňa)  Voľne prístupné databázy (napríklad Datacube ŠÚ SR, Eurostat, Svetová banka, a podobne)  V príprave je prístup k mikroúdajom z Eurostat (napríklad LFS, EU-SILC a podobne.) |
| Dátové potreby (nie sú v súčasnosti k dispozícií) | Centrálna platforma údajov zhromažďujúca rezortné údaje, ako napríklad údaje zo Sociálnej poisťovne, ÚPSVaR, Ministerstva školstva, Ministerstva zdravotníctva a ďalšie vybrané štatistické údaje, ktoré zbierajú ústredné orgány štátnej správy pre svoje interné použitie. Podstatné je, aby boli (1) údaje očistené a spoľahlivé, a (2) jednotlivé databázy jednoducho prepojiteľné. |
| Relevantné prípady použitia | Vzhľadom na nadrezortný charakter ÚV SR nie sú analýzy ISA limitované na konkrétny rezort, ale pokrývajú väčšinu oblastí hospodárskej a sociálnej politiky. Opakujúce sa úlohy zahŕňajú:   * hodnotenie verejných politík, * hodnotenie vplyvu Európskych štrukturálnych a investičných fondov * hodnotenie konkurencieschopnosti slovenskej ekonomiky. |
| Najlepšie praxe v pracovaní s dátami |  |

Tabuľka 11: Odbor analýz vzdelávacej politiky (IVP)

| Analytická jednotka | Odbor analýz vzdelávacej politiky (IVP) |
| --- | --- |
| Zaradenie | Ministerstvo školstva, vedy, výskumu a športu SR |
| Počet zamestnancov aktuálny (1Q-2019) | 7 |
| Počet zamestnancov plánovaný (1Q-2020) | 11 |
| Politiky, ktoré analyzuje/ ovplyvňuje analytická jednotka | Vzdelávacie politiky:   * inkluzívne vzdelávanie, * kurikulum, platy učiteľov, príprava učiteľov, * administratívna záťaž na školách, * nesúlad v zručnostiach na trhu práce, * racionalizícia siete základných škôl, * kapacity materských škôl, * národné či medzinárodné výsledky slovenských študentov, * revízia výdavkov na vzdelávanie. |
| Finančný rozsah riešených politík | Výdavky MŠVVŠ SR: 1 363 929 249 EUR (2018) |
| Využívané dátové zdroje | RIS, Register VŠ, Centrálny register študentov, NÚCEM databázy, CVTI zbery, medzinárodné databázy PISA, PIRLS, TIMSS, PIAAC |
| Dátové potreby (nie sú v súčasnosti k dispozícií) | Údaje Sociálnej poisťovne, Evidencie uchádzačov o zamestnanie, RFO. |
| Relevantné prípady použitia | * Prerozdeľovanie študentov do 1. ročníkov SŠ * Predikcia výsledkov vo vzdelávaní (na základe ekonomickej situácie v rodine) * Vzdelávacie politiky pre ľudí s predčasne ukončenou školskou dochádzkou |
| Najlepšie praxe v pracovaní s dátami | * EPRI-ESDC Tax Linkage Project (Kanada)[[3]](#footnote-4) * Tackling early leaving from education and training in Europe: Strategies, policies and measures[[4]](#footnote-5) |

Tabuľka 12: Infostat

| Analytická jednotka | Infostat |
| --- | --- |
| Zaradenie | Štatistický úrad Slovenskej republiky |
| Počet zamestnancov aktuálny (1Q-2019) | 35 |
| Počet zamestnancov plánovaný (1Q-2020) | 38 |
| Politiky, ktoré analyzuje/ ovplyvňuje analytická jednotka |  |
| Finančný rozsah riešených politík |  |
| Využívané dátové zdroje | * Sociálne štatistiky, demografické štatistiky, makro ekononomické dáta, priemyselné odvetvia, odvetvové údaje, národné účty, konjunkturálne prieskumy, kultúra, etc. |
| Dátové potreby (nie sú v súčasnosti k dispozícií) | * Finančná správa SR – daňové registre, * NCZI – zdravotná štatistika, * MŠVVaŠ SR, * Ministerstvo vnútra SR - Register fyzických osôb; * Ministerstvo práce sociálnych vecí a rodiny. |
| Relevantné prípady použitia | * Identifikácia rizikových daňových subjektov, * Makroekonomické a demografické prognózy, * Odvodené demografické prognózy - pracovná sila, * Ľudský kapitál, osoby v MRK, regionálne prognózy; * Trvaloudržateľná kvalita poskytovania zdravotnej starostlivosti vo všetkých regiónoch SR, * Analýza regionálnych disparít v rodinnom a reprodukčnom správaní a jej faktory, * Úmrtnosť, zdravotný stav a epidemiologická situácia v regionálnom aspekte. |
| Najlepšie praxe v pracovaní s dátami | * zvyšovanie úrovne dobrovoľného plnenia dňových povinností (Austrália), Rýchle odhady HDP (T+30dní) EUROSTAT, Demografické prognózy na národnej aj regionálnej úrovni; Aplikácia prospektívneho prístupu pri hodnotení procesu starnutia; Dekompozície faktorov rodinného a reprodukčného správania; Multivariačná faktorová analýza regionálnych diferenciácií demografického vývoja |

##### Použitie údajov

Údaje vo verejnej správe sa väčšinou nezbierajú pravidelne a väčšinou nie sú konzistentné a kvalitné. Zdieľanie dát je administratívne a technicky náročné (používa sa výmena dát na pevných nosičoch ako DVD). Neexistuje predstava o dátovom modeli verejnej správy: aké dáta majú jednotlivé inštitúcie verejnej správy k dispozícii a na aké účely sa používajú. Nie sú tiež jasne vyjasnené pravidlá na ochranu osobných údajov a zavedené spoľahlivé nástroje pre ich anonymizáciu a pseudonymizáciu.

##### Problémy, ktoré je potrebné vyriešiť:

* Publikácia údajov vo formáte otvorených dát.
* Formalizácia použitia analytických údajov v rozhodovacích procesoch.
* Zdieľanie analytických modelov.

### Architektúra informačných systémov

##### Aké nástroje využívajú analytické jednotky?

Každá jednotka používa vlastné nástroje, napríklad:

* ISA: R, Stata, v obmedzenej miere Python a No Sketch,
* Infostat: Microsoft office, EViews, modelové nástroje ekonometrického typu,
* IVP: : Microsoft office, Stata, R.

##### Aké údaje využívate?

Niektoré analytické jednotky majú prístup k zdrojovým databázam ISVS (napríklad IFP pre dáta s Finančnej správy SR).

Vo väčšine prípadov však pracujú s voľne dostupnými údajmi a štatistikami.

##### Akým spôsobom sa získavajú údaje?

* Anonymizované rezortné údaje. Ak údaje nie sú voľne dostupné, oslovujeme kompetentného zamestnanca na príslušnom rezorte. Údaje sa poskytujú na základe oficiálnej žiadosti alebo uzatvorenej zmluvy. Odozva na našu požiadavku trvá od niekoľkých dní po niekoľko mesiacov. Práve dlhá odozva býva obrovským problémom pri pružnom reagovaní na zadania. Ďalej, v prípade údajov, ktoré sa priamo nevykazujú v rezortných informačných systémoch, môže byť problémom, že údaje nie sú vo formáte, ktorý sa dá priamo využiť na analýzu (nejednotné alebo chýbajúce identifikátory sledovaných subjektov, papierová forma, pdf skeny, a podobne).
* Neanonymizované rezortné údaje. Postupuje sa v súlade s relevantnými právnymi predpismi.

##### Ako sa publikujú výsledky?

Údaje sa publikujú:

* na webových stránkach analytických jednotiek (napríklad <https://www.vlada.gov.sk/institut-pre-strategie-a-analyzy-isa/> ).
* na sociálnych sieťach, najmä facebook: (napríklad <https://www.facebook.com/ISA.UVSR/>),
* v zborníkoch z konferencií a v strategických dokumentoch (napríklad Akčné plány NRO).

Neexistuje štandard kvality publikovaných údajov a údaje sú publikované podľa potreby analytickej jednotky.

### Technologická architektúra

* V súčasnosti neexistuje spoločná technologická architektúra pre analytické spracovanie údajov, každá inštitúcia využíva vlastné možnosti a prostriedky.
* Vládny cloud neposkytuje PaaS a SaaS služby pre analytické nástroje, plánujú sa DB služby v rámci štúdie PaaS.

### Bezpečnostná architektúra

* Otázky bezpečnosti pri tvorbe analytických výstupoch sú riešené v rámci platných predpisov, samostatne každou inštitúciou.
* Centrálna koordinácia zabezpečenia analytických údajov nie je v praxi zavedená.

## Prevádzka

Prevádzkové náklady na tvorbu analytických výstupov si každá inštitúcia znáša samostatne.

# Popis budúceho stavu

## Legislatívne požiadavky

#### Súhrnný popis

Aby mohli analytické jednotky bez prekážok realizovať dátovú vedu a spracovanie údajov, je potrebné modernizovať regulačný rámec. Práve to je ambícia pripravovaného nového zákona o údajov nového zákona o údajoch, definovať pravidlá pre štát fungujúci na základe využívania znalostí. Je potrebné vytvoriť odborné podklady, ktoré umožnia nastaviť modernú legislatívu pre správne zadefinovanie údajov, ich vlastníctvo a použitie a ďalšie pravidlá tak, aby neboli narušené EÚ regulácie, ako GDPR, PSI, či voľný tok údajov v rámci krajín spoločenstva.

Nový zákon o údajoch vo verejnej správe sa bude venovať nasledujúcim otázkam:

* zadefinovať postavenie analytickej jednotky, jej práva a povinnosti pri analytickom spracovaní údajov verejnej správy, najmä vo vzťahu k ochrane osobných údajov,
* určiť povinnosti pri výkone verejnej moci na základe faktov a analytických podkladov,
* určiť povinnosti inštitúcii verejnej správy poskytovať údaje na analytické spracovanie,
* stanoviť rámec pre anonymizáciu údajov,
* právne definovať Konsolidovanú analytickú vrstvu.

Potrebné štandardy:

* spôsob anonymizácie údajov pre analytické spracovanie,
* prepojená databáza mikro-údajov Konsolidovanej analytickej vrstvy.

#### Riziká

* Výsledná kvalita legislatívy (takzvaného „zákona o údajoch“) bude nízka.
* Legislatívna príprava bude meškať, respektíve parlament neschváli reformné zákony v požadovanej forme.
* Neprijatie zákona o údajoch.
* Potrebné štandardy sa nepodarí pripraviť, resp. ich kvalita bude nízka

## Biznis architektúra navrhovaného riešenia

Biznis architektúra definuje biznis funkcie a biznis služby, ktoré budú ponúknuté ako v rámci Konsolidovanej analytickej vrstvy: aké dátové produkty pre verejnosť budú poskytované, aké prípady použitia budú podporované vďaka KAV, aké postupy dátovej vedy pre analytické jednotky budú umožnené, ktoré analytické jednotky sa zapoja do úvodného projektu, ako i procesy správy dátovej vrstvy a podporu pre manažment životného cyklu údajov.

V rámci biznis architektúry je potrebné vyriešiť:

* organizačné zabezpečenie: role v procesu spracovania a vyhodnocovania dát
* prípady použitia ktoré budú podporované konsolidovanou analytickou vrstvou
* postupy dátovej vedy, ktoré budú umožnené pre analytické jednotky
* biznis služby, ktoré budú poskytované v rámci konsolidovanej analytickej vrstvy
* procesy správy dátovej vrstvy

#### Organizačné zabezpečenie a ľudské zdroje

Využitie dátovej vedy v riadení štátu sa nedá automatizovať a kľúčovým faktorom úspechu je ľudský faktor. Až experti v kombinácií s nástrojmi a dátami dokážu priniesť reálne výsledky do praxe. Výrazným problémom je zabezpečiť kvalitné ľudské zdroje a ich organizáciu. Kľúčovou požiadavkou riešenia je preto podpora a rozvíjanie ľudského kapitálu v rámci verejnej správy, zabezpečením kvalitných pracovných podmienok a kultúry a prístup k najlepším školeniam a možnostiam ďalšieho vzdelávania.

Problematika si vyžaduje odbornú koordináciu procesov životného cyklu dát ako zberu, čistenia a spracovania dát, a zároveň technickú prevádzku dátových skladov a odbornú asistenciu pri využívaní moderných analytických nástrojov. Takéto úlohy by bolo vhodné riešiť centrálne a profesionálne. Vznikla preto Dátová kancelária verejnej správy, ktorá bude mať na starosti bázu dát verejnej správy a poskytovanie podpory a asistencie ostatným analytickým jednotkám v oblasti dátovej vedy.

##### Analytické jednotky:

V cieľovom stave budú analytické jednotky fungovať na každom ministerstve a významných ústredných orgánoch štátnej správy, ako i vo vybraných samosprávach. Od Analytickej jednotky sa očakáva doménová znalosť témy a dát. V rámci tímu by mali byť experti, ktorí dokážu využívať poskytované nástroje a poznajú postupy dátovej vedy.

V rámci riešenia budú odbremení od nutnosti získavať dáta a manuálne a vo vlastnej réžii riešiť kvalitu vstupných údajov. Budú klásť požiadavky na Dátovú kanceláriu aby im zabezpečila potrebné dáta a dáta, ktoré získajú budú môcť zdieľať s ostatnými analytickými jednotkami.

**Analytické jednotky budú používateľmi Konsolidovanej analytickej vrstvy.** Analytické jednotky vytvoria a implementujú vlastné analytické modely a nástroje, pričom budú využívať dostupné dáta najlepším spôsobom. Pravidlom bude, že každá analytická jednotka bude mať prístup k relevantným dátam v dátovom sklade (anonymizovaným alebo pseodoanonymizovaným, v prípade dostatočne odôvodenej potreby aj k neanonymizovaným), ako i k analytickým výstupom ostatných jednotiek, ktoré si budú môcť aj zdieľať na „peer review“.

##### Dátová kancelária:

Zriadenie Dátovej kancelárie verejnej správy, ktorá bude mať na starosti správu dátovej analytickej vrstvy a prípravu nástrojov pre analytické jednotky, ako aj správu Centrálneho dátového skladu a jeho informačnú a dátovú architektúru. Dátová kancelária verejnej správy je zriadená v rámci ÚPPVII. Dátová kancelária bude mať na starosti centrálnu koordináciu analytických jednotiek vo vzťahu k transformácií verejnej správy na štát fungujúci na základe využívania dát.

Dátová kancelária bude slúžiť aj ako poradný útvar pre ostatné inštitúcie verejnej správy v oblasti využitia dátových techník pre transformáciu fungovania organizácie a zlepšenie procesov formou odborného poradenstva.

**Dátová kancelária bude správcom Konsolidovanej analytickej vrstvy.**

##### Ďalšie organizačné jednotky

....

Obrázok 5: Prehľad kľúčových používateľov analytickej vrstvy



#### Systematické zlepšenie využívania dát vo verejnej správe

Systematického zlepšenia využívania dát vo verejnej správe bude založené na využívaní dôkazov a faktov v rozhodovacích procesoch. Organizačne je tvorba kvalitných analytických materiálov a výstupov úlohou **analytických jednotiek**, ktoré vznikli na jednotlivých ministerstvách. Analytické jednotky potrebujú pre svoju činnosť talentované ľudské zdroje, dáta a vhodné nástroje. V rámci projektu budú vybrané analytické jednotky, ktorých potreby po dátových zdrojoch budú zabezpečené a k dispozícii dostanú platformu na tvorbu a komunikáciu analytických výstupov. Ostatné analytické jednotky sa budú pripájať v ďalších fázach. Možnosti Konsolidovanej analytickej vrstvy budú využité aj v ďalších oblastiach fungovania verejnej správy (ako je podpora návrhu regulácií, posudzovanie investícií a podobne).

Nové riešenie bude navrhnuté tak, aby umožnilo zavádzanie analytického prístupu. Dátová kancelária pripravuje **systematický prehľad o analýzach vo verejnej správe:** pre jednotlivé sektory verejnej správy, úseky verejnej správy a oblasti analýz bude možné definovať množinu analytických produktov, ktoré sa budú vytvárať, pričom sa očakávajú typy analýz ako: ex-ante vyhodnocovanie opatrení a variantov, ex-post vyhodnocovanie prijatých opatrení, online vyhodnocovanie, prognózy, simulácie, štatistické zisťovania a podobne. Je vhodné, aby sa všetky typy analýz vykonávali nad spoločnou množinou dát, ktorá má vysokú kvalitu a jednotnú ontológiu a dátový model.

Dôležitým aspektom je zabezpečiť, aby analytické jednotky dostali správne dáta a bolo možné definovať správne problémy, ktoré bude potrebné riešiť. Preto sa podporia možnosti navrhovať a testovať predpoklady jednotlivých modelov a postupne ich zlepšovať. Pre každý úsek verejnej správy tak budeme poznať:

* aké sú analytické potreby na danom úseku,
* aké metódy budú použité na danom úseku,
* aké výstupy (analytické produkty) budú vytvárané (v akých intervaloch),
* aké sú požiadavky na zdroje údajov,
* aké sú požiadavky na analytické nástroje.

Základnou úlohou Konsolidovanej analytickej vrstvy je vytvoriť otvorenú platformu, ktorá umožní realizovať navrhnuté potreby, aby bolo možné postupne pripájať analytické jednotky, nasadzovať analytické produkty a zabezpečovať potrebné zdroje dát. Úlohou Dátovej Kancelárie verejnej správy je riadiť takýto systém a zabezpečovať vstupné dáta v požadovanej kvalite, ako i metodická podpora používateľov. V rámci riešenia bude kladený dôraz, aby riešenie podporovalo:

* relevantné metódy dátovej vedy a celý životný cyklus tvorby analytických produktov,
* lepšie možnosti pre realizáciu RCT, aby bolo možné hodnotiť intervencie,
* tvorbu prediktívnych služieb,
* profesionálny manažment konsolidovanej analytickej vrstvy.

#### Podporované postupy dátovej vedy

Analytické jednotky získajú prostredie, v ktorom budú môcť tvoriť kvalitné dátové modely a produkty na základe princípov a najlepších postupov dátovej vedy. Riešenie umožní tvorbu analytických prediktívnych modelov, ktoré môžu byť vždy aktualizované na základe aktuálnych dát (to znamená, že model je možné v každom okamihu v čase prepočítať na základe aktuálnych dát, čím je možné overiť jeho prediktívne schopnosti a prípadne ho vylepšiť, systém sa tak stará o neustálu kompatibilitu historických dát s aktuálnymi a o prepojenia modelov s aktualizovanými množinami dát).

Obrázok 6: Životný postup dátovej vedy



Podporovaný bude celý postup dátovej vedy:

* **Porozumenie doméne a formulácia hypotéz:** porozumenie situácií, definícia problémov, stanovenie cieľov, ktoré je potrebné dosiahnuť, definícia hypotéz.
* **Správa analytického „sandboxu“:**
* **Výber vhodných dátových zdrojov:** zabezpečenie prístupu k dátam na technickej úrovni, nákup dát,
* **Zhromažďovanie a získavania dát:** prehľadávanie dát a dátových zdrojov, zber dát, integrácia dát, popis dát, formátovanie dát, prepájanie dát, skladovanie dát, meranie dátovej kvality,
* **Príprava dát**: príprava a výber dát, čistenie dát, prieskum dát, hľadanie informácií a vzťahov medzi dátami,
* **Výber vhodných analytických metód**,
* **Tvorba analytického dátového modelu:** vypracovanie modelu, kontrola modelu, zdieľanie modelu,
* **Testovanie hypotéz** (test relevantnosti a test istoty),
* **Hodnotenie výsledkov:** produkcia finálneho reportu, posúdenie výsledkov a ich významu,
* **Komunikácia výsledkov a pripomienkovanie:** publikovanie reportu, zber spätnej väzby, vizualizácia výsledkov, príprava analytického produktu.

#### Realizácia „Randomized Control Trials“ - RCT

Randomized Control Trials predstavujú najlepší spôsob, ako sa dá stanoviť, či určitá politika funguje. RCT sa v dnešnej dobe vykonávajú extenzívne v medzinárodnej rozvojovej spolupráci, pri vývoji liečiv, ako i v komerčných prípadoch použitia a postupne sa stávajú odporúčaným spôsobom aj pri návrhu verejných politík a pri zvyšovaní výkonnosti verejného sektora. Lídrami v aplikácií RCT sú napríklad BIT a Demos Helsinki.

RCT funguje na princípe náhodného priradenia kontrolnej skupiny, čo umožní vyhodnocovať, aký efekt má dané opatrenie voči nečinnosti, alebo voči alternatívnej možnosti. Dôležitým faktorom správne nastaveného RCT je dôkladný zber dát počas doby realizácie testu, pričom táto doba môže byť i niekoľko mesiacov až rokov. Použitím algoritmov dátovej vedy pri spracovaní dát z RCT je možné identifikovať rozdielne vplyvy intervencií vo všetkých pozorovateľných charakteristikách, pričom sa zabezpečí, že nájdene riešenie bude predstavovať najlepšiu možnú intervenciu v danej oblasti. Riešenie podporí všetky relevantné kroky pre podporu RCT s využitím najnovších algoritmov dátovej vedy:

* Krok 1: Identifikácia dvoch alebo viacerých možností pre riešenie problému (návrh politiky),
* Krok 2: Definícia výstupov, ktoré bude politika ovplyvňovať a stanovenie požiadaviek na dátové zdroje,
* Krok 3: Určenie jednotky náhodného výberu a výpočet, koľko jednotiek je potrebných, aby výsledok bol robustný,
* Krok 4: Priradenie jednotiek k jednotlivým možnostiam politiky,
* Krok 5: Realizácia zásahov a intervencií,
* Krok 6: Meranie výsledkov a stanovenie dopadov, publikácia výsledkov,
* Krok 7: Prispôsobenie politík na základe výsledkov.

#### Tvorba prediktívnych služieb

Okrem návrhu politík dokáže dátová veda pomôcť aj pri poskytovaní službám občanom, najmä vďaka schopnosti predikcie. Zámerom je preto využiť Konsolidovanú analytickú vrstvu aj v manažmente verejných služieb a v operatívnom rozhodovaní na jednotlivých inštitúciách. Prediktívne modely je možné použiť:

* Na stanovenie budúceho vývoja trendov, napríklad poznanie štruktúry trhu práce umožní lepšie plánovať počty žiakov v jednotlivých odboroch,
* Na stanovenie budúceho dopytu po verejných službách, čo umožní optimalizovať poskytovanú ponuku (napríklad kapacity škôl, škôlok a jaslí v definovanej lokalite),
* Predpovedať výskyt nežiaducich javov: napríklad predpovedať výskyt epidémií a ochorení vo verejnom zdravotníctve alebo predpovedať kriminalitu vo vybranej lokalite,
* Predpovedať budúci vývoj prípadu alebo klienta a identifikovať riziko, na základe ktorého napríklad dieťa zo sociálne slabšieho prostredia bude mať problémy zvládať výučbu, čo umožní personalizovaný prístup k deťom.
* Predpovedať riziko podvodu, napríklad pri udeľovaní grantov, dotácií, alebo pri verejnom obstarávaní.

Konsolidovaná analytická vrstva bude poskytovať rámec pre tvorbu prediktívnych modelov, ktoré bude možné ľahko zakomponovať do biznis procesov organizácie.

#### Manažment analytickej vrstvy

Manažment analytickej vrstvy bude vykonávať Dátová kancelárie verejnej správy. Riešenie musí podporovať holistický súbor štandardizovaných procesov a usmernení, ktoré adresujú to, ako sa majú dáta získavať, spravovať, ukladať a ako sa k nim má pristupovať, aby používatelia mali vždy istotu, že majú správnu verziu dát zo spoľahlivého zdroja. Tieto procesy sa musia tiež odkazovať aj na patričné technológie, ktoré sú využívané na tieto účely a adresovať aspoň nasledujúce funkcie.

* **Zabezpečenie zdieľania údajov medzi analytickými jednotkami:** jednotlivé analytické jednotky potrebujú pracovať s výstupmi, alebo i zo zdrojovými údajmi, ktoré získali alebo pracujú s nimi ostatné analytické jednotky. Konsolidovaná analytická vrstva bude plniť funkciu zásobníka všetkých analytických dát a zabezpečí rýchly prístup pre každého, kto dáta potrebuje a má na ne oprávnenia.
* **Zabezpečenie zberu údajov:** v prípade, že potrebné údaje nie sú vo verejnej správe k dispozícií, je potrebné nastaviť jednoduchý a efektívny spôsob, ako takéto údaje získať a uložiť ich. Riešenie bude podporovať zber údajov z administratívnych zdrojov (z ostatných informačných systémov verejnej správy) cez centrálnu integračnú platformu alebo inými možnosťami ntegrácie, ale i pripojenie na externé zdroje dát, ako sú univerzity, sociálne siete, platformy a takzvaný „data-brokers“ a ďalšie externé inštitúcie (Eurostat, OECD, ... ). V neposlednej rade je dôležité zabezpečiť možnosti pre zber tokov dát cez IoT.
* **Riadenie pripájania dátových zdrojov**: jednotlivé integračné rozhrania sú manažované spôsobom, aby bolo možné sledovať a meniť frekvenciu zmien, spôsobom prístupu, oprávnenia a zmeny integračných rozhraní. Zámerom je aj znázorniť prehľad všetkých dátových zdrojov a metadát.
* **Čistenie dát a kontrola kvality**: ide o proces opravy alebo odstraňovania dát, ktoré sú nesprávne, nekompletné, zle naformátované ale duplicitné. V rámci procesu budú prebiehať opravy dát a vyhodnocovať kvalita jednotlivých dátových zdrojov.
* **Transformácia dát do potrebnej štruktúry a prepájanie údajov:** dátová kancelária bude vytvárať spoločnú databázu mikro-údajov, kde prepojí fakty a dátové objekty z rôznych zdrojov do spoločného dátového priestoru. Výrazne sa tým zvýšia možnosti pre analýzu a zjednoduší sa tiež následná práca analytických jednotiek s týmito dátami. Súčasťou úlohy je tiež transformácia veľkých dát z nízkou informačnou hodnotou a nízkou hustotou na veľké dáta s vysokou informačnou hodnotou a vysokou hustotou.
* **Ochrana citlivých údajov a aplikovanie rôznych politík na ochranu citlivých údajov:** konsolidovaná analytická vrstva bude obsahovať veľké množstvo osobných údajov. Osobné údaje a údaje, ktoré súvisia s daným subjektom sú predstavujú veľmi dôležitý vstup do analytických modelov. Aby mali predpovede zmysel a reálne použitie, je potrebné aby údaje mali čo najvyššiu informačnú hodnotu. Odporúčame preto pseudonymizovať osobné údaje, ide o vhodný kompromis medzi ochranou osobných údajov a ich využitím pre potreby lepšieho rozhodovania. Je preto potrebné, aby boli k dispozícii možnosti pre zabezpečenie ochrany osobných údajov a aplikovanie rôznych politík, ako sú: prístup k osobným údajom na len základe oprávnenia a logovanie týchto prístupov, využitie osobných údajov pre analýzy vo verejnom záujme, prístup k osobným údajom na základe udeleného súhlasu, prístup k anonymizovaným údajom.
* **Anonymizácia a pseudonymizácia:** pre analytickú prácu nie je potrebné pracovať s presnými osobnými údajmi a identifikátormi, stačí použiť virtuálny identifikátor, aby bolo možné prepájať záznamy a vytvárať tak obraz. Riešenie musí podporovať vhodné postupy pre anonymizáciu údajov, najmä pseudonymizáciu.
* **Zabezpečenie verzionovania:** sledovanie a správa historických verzií datasetov zo zdrojov, riešenie problémov s časovými radmi a nekonzistentnými verziami.
* **Manažment úložísk:** Správa jednotlivých typov úložísk tak, aby bol vždy k dispozícii optimálny priestor. Ako úložiská môže byť použitý vládny cloud, alebo privátne riešenia. Záleží na charaktere dát. V princípe budú podporované všetky významné typy úložísk, aby bolo možné zaznamenávať štruktúrované dáta s administratívnych zdrojov, štatistické dáta, dokumenty, obrázky, videá ako i dynamické dáta zo senzorov a IoT.
* **Publikovanie otvorených dát:** jedným z dôležitých princípov pre prácu s údajmi vo verejnom sektore bude, že všetky údaje, ktoré nemajú charakter utajených alebo osobných budú automaticky zverejňované v otvorenom formáte vhodnom na strojové spracovanie. Takýmto spôsobom bude fungovať aj Konsolidovaná analytická vrstva: ako otvorené údaje budú prístupné zdrojové dáta a aj výstupy analýz. Príslušné datasety budú registrované v centrálnom katalógu data.gov.sk, uložené v rámci Konsolidovanej analytickej vrstvy.
* **Archivovanie:** zdrojové aj spracované dáta budú vo vhodnom okamihu archivované. Archív dát bude použiteľný na historické bádanie.

#### Prípady použitia analytických nástrojov a analytickej vrstvy

##### Oblasti, ktoré je možné podporiť vďaka lepšiemu využitiu dát

Lepším využitím dát je možné zvýšiť výkonnosť Za základné oblasti verejnej správy, v ktorých sa bude využívať systematická analýza dát, je možné považovať:

* **Strategické priority** – príprava stratégií, analýza súčasného stavu a porovnávanie variantov a následné sledovanie napĺňania strategických priorít,
* **Politiky** – vyhodnocovanie dopadov politík, RCT, sledovanie výsledkov intervencií, scenáre budúceho vývoja, predvídanie budúceho dopytu po službách,
* **Regulácie** – Hodnotenie dopadov regulácií (Regulatory impact assessment - RIA), ktoré sa ďalej delí na:
* hodnotenie vplyvov na podnikateľské prostredie (Business impact assessment - BIA) – vrátane sledovania dopadov na malé a stredne podniky (MSP Test),
* hodnotenie vplyvov na životné prostredie (Environmental impact assessment - EIA),
* hodnotenie vplyvov v sociálnej oblasti (Social impact assessment - SIA),
* hodnotenie vplyvov na rozpočet verejnej správy (Financial impact assessment - FIA),
* hodnotenie vplyvov na zdravie (Health impact assessment HIA),
* hodnotenie vplyvov na informatizáciu spoločnosti,
* hodnotenie vplyvov na verejnú správu.
* **Investície** – vyhodnocovanie dopadov projektov a verejných investícií,
* **Výkonnosť a kvalita** – vyhodnocovanie výkonu verejnej správy pri poskytovaní služieb v jednotlivých sektoroch, identifikácia výkonnosti jednotlivých inštitúcii, predpovedanie budúceho vývoja,
* **Poskytovenie služieb** -
* **Podpora rozhodovania v operatívnych procesoch** (analýzy rizík, „fraud management“, ... ).

##### Prípady použitia, v ktoré budú podporené Konsolidovanou analytickou vrstvou

Konsolidovaná analytická vrstva bude používaná pre tvorbu „ad-hoc“ analýz, alebo štandardizovaných analytických produktov.

| Organizačná jednotka | Oblasť | Prípad použitia | Analytický produkt/Výstup |
| --- | --- | --- | --- |
| Inštitút pre stratégie a analýzy (ISA) | Strategické priority | Monitorovanie strategických cieľov štátu | **Dashboard strategického riadenia štátu**: online prehľad strategických cieľov (na základe schválených strategických dokumentov), ktoré ukazujú výkonnosť politík v jednotlivých sektoroch. |
| Plnenie programového vyhlásenia |
| Strategické priority | Strategické plánovanie rozvoja Slovenska | **Scenáre budúceho vývoja** (simulácie): |
| **Národná inovačná stratégia**. analýza „value-chains“ z cieľom identifikovať dopady rôznych inovačných politík. |
| Politiky | Hodnotenie vplyvu EŠIF a vládnych politík | **Kontrafaktuálne hodnotenia vplyvu politík a intervencií** v oblastiach zdravotníctvo, vzdelávanie, zamestnanosť a trh práce, rozvoj regiónov.  **Kontrafaktuálne hodnotenie podpory z EŠIF** v rámci operačného programu alebo konkrétnych projektov. |
| Inštitút vzdelávacej politiky | Výkonnosť a kvalita | Hodnotenie výkonnosti vzdelávacích inštitúcií . identifikácia problémov a nedostatočne výkonných inštitúcií | **Benchmark výkonnosti vzdelávania** pre jednotlivé inštitúcie  **Predikcia budúcej výkonnosti** |
| Výkonnosť a kvalita | Predpovedanie budúcich výsledkov žiakov | **Globálny report** o výsledkoch žiakov na základe relevantných faktorov  **Aplikácia pre individuálny odhad výsledkov** |
| Politiky | Predvídanie budúceho dopytu po službách (kapacitách vzdelávacích inštitúcií) | **Mapa dopytu po vzdelávacích službách** |
| Politiky | Prerozdeľovanie študentov do 1. ročníkov SŠ | **Návrh počtov žiakov prvého ročníka pre jednotlivé študijné odbory SŠ** |
| Politiky | Vzdelávanie ľudí s predčasne ukončenou školskou dochádzkou | **Absolventská miera nezamestnanosti**  **Predikcia absolventskej miery nezamestnanosti** |
| Infostat | Politiky | Krátkodobé predikcie vývoja slovenskej ekonomiky | **Krátkodobá predikcia vývoja slovenskej ekonomiky** (analyticko-prognostický dokument) |
| Politiky | Rýchle odhady HDP a celkovej zamestnanosti v hospodárstve | **Odhad rastu a štruktúry HDP** (výsledky s analytickým komentárom)  **Odhad zamestnanosti, jej rastu a štruktúry** (výsledky s analytickým komentárom) |
| Politiky | Krátkodobé predikcie vývoja (indexov) priemyselnej produkcie | **Predikcia vývoja indexov priemyselnej produkcie**  (prognostický dokument) |
| Strategické priority/  Politiky | Populačné prognózy | **Regionálne populačné prognózy**  **Odvodené populačné prognózy** |
| Dátová kancelária verejnej správy | Výkonnosť a kvalita | Manažment kvality dát | Prehľad kvality údajov v jednotlivých informačných systémoch |

##### Monitorovanie strategických cieľov štátu a plnenia programového vyhlásenia

Tabuľka 13: Monitorovanie strategických cieľov štátu a plnenia programového vyhlásenia

|  |  |
| --- | --- |
| Prípad použitia | Monitorovanie strategických cieľov štátu a plnenia programového vyhlásenia |
| Spôsob použitia | Denno-denné hodnotenie aktuálnej situácie na základe sledovaných KPI v jednotlivých rezortoch  Transparentné informovanie verejnosti o kvalite činnosti vlády, najmä vo vzťahu k plánovanému programovému vyhláseniu |
| Dopad | Možnosť rýchlej identifikácie oblastí, ktoré nespĺňajú nastavený výkonnostný rámec, čo môže znamenať schopnosť adekvátne reagovať  Prenesenie dôrazu na podstatné metriky v jednotlivých segmentoch a zmena celkovej diskusie o činnosti vlády |
| Frekvencia | Online |
| Potrebné dátové zdroje | * Finančná správa, * ÚPSVaR, * Sociálna poisťovňa * Zdravotnícka štatistika z NCZI * Ekonomické štatistiky |
| Analytické modely |  |
| Výstup/Analytický produkt | Dashboard strategického riadenia štátu: online prehľad strategických cieľov (na základe schválených strategických dokumentov), ktoré ukazujú výkonnosť politík v jednotlivých sektoroch. |
| Najlepšia prax | Dashboard v meste Boston[[5]](#footnote-6) |

Tabuľka 14: Strategické plánovanie rozvoja Slovenska

|  |  |
| --- | --- |
| Prípad použitia | Strategické plánovanie rozvoja Slovenska |
| Spôsob použitia | Príprava stratégií na základe realistických prognóz budúceho vývoja  Identifikácia kľúčových trendov a faktorov, na ktoré je potrebné sa sústrediť pri plánovaní  Simulácie geopolitických a ekonomických udalostí mimo kontroly Slovenska a ich dopadov |
| Dopad | Lepšia príprava na hrozby a výzvy, pred ktorými krajina stojí  Realistickejšie nastavovanie dopadových ukazovateľov  Lepšie porozumenie možnostiam ovplyvniť vývoj vďaka nastaveniu verejných politík |
| Frekvencia | Ad-hoc, online podľa potreby |
| Potrebné dátové zdroje | * Social Progress Index, * Ekonomické štatistiky, * Daňový informačný systém, * Údaje o pohybe tovarov, * Obchod, * Politické preferencie |
| Analytické modely | * Model hodnotenia národných a regionálnych politických, ekonomických, sociálnych, environmentálnych a technologických rizikových faktorov * Model „value-chains“ slovenského hospodárstva – agentovo-orientovaná simulácia |
| Výstup/Analytický produkt | Scenáre budúceho vývoja (simulácie),  Národná inovačná stratégia. analýza „value-chains“ z cieľom identifikovať dopady rôznych inovačných politík. |
| Najlepšia prax | <https://futuremap.io/foundation> |

Tabuľka 15: Hodnotenie vplyvu EŠIF

|  |  |
| --- | --- |
| Prípad použitia | Hodnotenie vplyvu EŠIF a vládnych politík |
| Spôsob použitia | Výstup slúži na identifikáciu politík a intervencií, ktoré prinášajú výsledky:   * napríklad pri vzdelávaní zlepšujú študijné výsledky študentov, * napríklad pri zdravotníctve zvyšujú efektívnosť nemocníc, znižujú úmrtnosť pacientov, a podobne, * napríklad pri trhu práce. zvyšujú zamestnanosť alebo zlepšujú postavenie marginalizovaných skupín.   V prípade hodnotenia EŠIF výstupy slúžia na hodnotenie operačných programov a realizovaných projektov. |
| Dopad | Efektívnejšia alokácia EŠIF.  Vyššia výkonnosť verejných politík vďaka porozumeniu efektu intervencií. |
| Frekvencia | Kontinuálne  Frekvencia hodnotení závisí od návrhov nových politík a opatrení alebo požiadaviek na hodnotenie existujúcich politík, opatrení a zdrojov podpory. |
| Potrebné dátové zdroje | * ITMS+ * ÚPSVaR, * Sociálna poisťovňa * Zdravotnícka štatistika z NCZI * Údaje z testovaní NÚCEM (Testovanie 9, Testovanie 5, Maturita) * Údaje z Rezortného informačného systému MŠVVŠ SR, * Údaje z výkazov zbieraných CVTI. |
| Analytické modely | * Prepojenie údajov z ITMS s rezortnými databázami. Predpokladá doplnenie a očistenie údajov z ITMS a zjednotenie identifikátorov sledovaných subjektov (organizácií, jednotlivcov a tak ďalej) * Mikromodely v zdravotníctve na úroveň pacientov, * Prepojenie vyššie uvedených databáz. * Zjednotenie identifikátorov tak, aby sa dali žiaci jednoducho prepojiť so školou, a školy so školskými charakteristikami. Tieto identifikátory by mali byť jednotné v čase. Nakoniec, ideálne by bolo prepojenie školských databáz s ďalšími databázami. Napríklad vieme, že vplyv rôznych politík a intervencií je moderovaný socio-ekonomickým zázemím študenta. Bolo by preto dobré mať o študentoch aj dodatočné informácie o ich rodine, napr. do ktorej príjmovej skupiny spadajú rodičia študenta alebo či majú študenti súrodencov. |
| Výstup/Analytický produkt | Kontrafaktuálne hodnotenia vplyvu politík a intervencií v oblastiach zdravotníctvo, vzdelávanie, zamestnanosť a trh práce.  Kontrafaktuálne hodnotenie podpory z EŠIF v rámci operačného programu alebo konkrétnych projektov. |
| Najlepšia prax |  |

Tabuľka 16: Hodnotenie výkonnosti vzdelávacích inštitúcií

|  |  |
| --- | --- |
| Prípad použitia | Hodnotenie výkonnosti vzdelávacích inštitúcií |
| Spôsob použitia | Identifikácia inštitúcii, u ktorých sa dá predpokladať nízka výkonnosť  Testovanie vzdelávacích programov  Podklady pre školskú inšpekciu |
| Dopad | Zvýšenie tlaku na kvalitu vzdelávacích inštitúcií a podpora voľby rodičov pre výber najlepších možností |
| Frekvencia | Raz mesačne |
| Potrebné dátové zdroje | Rezortný informačný systém |
| Analytické modely | Štatistické metódy a prediktívne modelovanie |
| Výstup/Analytický produkt | **Benchmark výkonnosti vzdelávania** pre jednotlivé inštitúcie (výkonnosť vo vzdelávaní, finančná efektivita, podmienky, participácia a zapojenie a podobne)  **Predikcia budúcej výkonnosti** vzdelávacích inštitúcií |
| Najlepšia prax | Global Partnership for Education (<https://www.globalpartnership.org/> ) |

Tabuľka 17: Predpovedanie budúcich výsledkov žiakov

|  |  |
| --- | --- |
| Prípad použitia | Predpovedanie budúcich výsledkov žiakov |
| Spôsob použitia | Ak budeme schopný na základe dostupných údajov predpovedať budúce výsledky, dokážeme nastaviť individuálne programy, navrhovať cieľené intervencie a minimalizovať riziká nevyhovujúcich výsledkov a zabezpečiť tak úspešný priebeh školskej dochádzky pre čo najširšiu skupinu žiakov.  Na základe údajov RFO a Sociálnej poisťovne sa dá zistiť ekonomická situácia v rodine. |
| Dopad | Ekonomická situácia rodiny je kľúčovým faktorom, ktorý ovplyvňuje výsledky vo vzdelávaní. Pri analytických výstupoch, ktoré sa zameriavajú na výsledky žiakov je preto nutné tento faktor zohľadňovať. To v súčanej situácii nie je možné, respektíve je možné iba na základe obmedzených medzinárodných údajov (kde tiež nie je evidovaný príjem, ale proxy premenné pre príjem). |
| Frekvencia | Ad-hoc, online podľa potreby |
| Potrebné dátové zdroje | * Centrálny register študentov (dostupný analytickej jednotke už teraz) * Údaje Sociálnej poisťovne, * RFO |
| Analytické modely | Predikčný algoritmus postavený na strojovom učení |
| Výstup/Analytický produkt | **Globálny report** o výsledkoch žiakov na základe relevenantných faktorov  **Aplikácia pre individuálny odhad výsledkov** (Decision aid tool) |
| Najlepšia prax | Using Data Science in Policy[[6]](#footnote-7) |

Tabuľka 18: Predvídanie budúceho dopytu po službách (kapacitách vzdelávacích inštitúcií)

|  |  |
| --- | --- |
| Prípad použitia | Predvídanie budúceho dopytu po službách (kapacitách vzdelávacích inštitúcií) |
| Spôsob použitia | Racionalizácia siete vzdelávacích inštitúcií  Manažment kvality vzdelávacích inštitúcií |
| Dopad | Efektívnejšia sieť vzdelávacích inštitúcií  Redukcia problémov s nedostatočnou kapacitou v jednotlivých lokalitách (napríklad pre materské školy)  Možnosť plánovať budúce počty vzdelávacích pracovníkov |
| Frekvencia | Ročne |
| Potrebné dátové zdroje | * Údaje zo sociálnej poisťovne * RFO * Demografické predikcie * Telekomunikačné údaje |
| Analytické modely | Sieťové analýzy |
| Výstup/Analytický produkt | Mapa dopytu po vzdelávacích službách (interaktívna aplikácia) |
| Najlepšia prax |  |

Tabuľka 19: Prerozdeľovanie študentov do 1. ročníkov SŠ

|  |  |
| --- | --- |
| Prípad použitia | Prerozdeľovanie študentov do 1. ročníkov SŠ |
| Spôsob použitia | Objektivizuje sa spôsob prerozdeľovania študentov do 1. ročníkov SŠ v jednotlivých SŠ.  V rámci vyhodnotenia kritérii vo Vyhláške č. 252/2018 Z. z. má ministerstvo povinnosť zohľadniť pomer miery zamestnanosti absolventov k miere zamestnanosti v príslušnom okrese a pomer priemernej mzdy absolventov k priemernej mzde v príslušnom okrese. |
| Dopad | Lepší informácie pre individuálny výber štúdijného odboru  Prispôsobenie kapacít stredného školstva objektívnym potrebám trhu práce  Lepšie plánovanie budúcich kapacít stredného školstva |
| Frekvencia | Ročne |
| Potrebné dátové zdroje | Centrálny register študentov (dostupný AJ)  Údaje Sociálnej poisťovne |
| Analytické modely |  |
| Výstup/Analytický produkt | Návrh počtov žiakov prvého ročníka pre jednotlivé štúdijné odbory SŠ  Vzorom môže byť súčasný výstup: Upravený počet žiakov prvého ročníka v dennej forme štúdia v členení na jednotlivé študijné odbory alebo na jednotlivé učebné odbory stredných škôl pre školský rok 2019/2020[[7]](#footnote-8) |
| Najlepšia prax |  |

Tabuľka 20: Vzdelávanie ľudí s predčasne ukončenou školskou dochádzkou

|  |  |
| --- | --- |
| Prípad použitia | Vzdelávanie ľudí s predčasne ukončenou školskou dochádzkou |
| Spôsob použitia | Podobne ako sa dajú hodnotiť vzdelávacie výstupy v podobe miery zamestnanosti a príjmu, dajú sa hodnotiť aj výstupy z hľadiska miery nezamestnanosti. |
| Dopad | Absolventská miera nezamestnanosti sa dá využiť ako proxy indikátor kvality SŠ prípadne VŠ. Miera nezamestnanosti by sa tiež mohla pridať do Vyhlášky č. 252/2018 Z. z. Miera nezamestnanosti sa tiež dá využiť ako indikátor pri nastavovaní vzdelávacích politík pre ľudí s predčasne ukončenou školskou dochádzkou. |
| Frekvencia | Ad-hoc, online podľa potreby |
| Potrebné dátové zdroje | Centrálny register študentov (dostupný AJ)  Údaje evidencie uchádzačov o zamestnanie |
| Analytické modely |  |
| Výstup/Analytický produkt | **Absolventská miera nezamestnanosti** (indikátor v reálnom čase podľa okruhov ako regióny, školy, štúdijné odbory)  **Predikcia absolventskej miery nezamestnanosti** |
| Najlepšia prax | <https://eacea.ec.europa.eu/national-policies/eurydice/content/tackling-early-leaving-education-and-training-europe-strategies-policies-and-measures_en>  EPRI-ESDC Tax Linkage Project (Kanada)[[8]](#footnote-9) |

Tabuľka 21: Krátkodobé predikcie vývoja slovenskej ekonomiky

|  |  |
| --- | --- |
| Prípad použitia | Krátkodobé predikcie vývoja slovenskej ekonomiky |
| Spôsob použitia | Použitie analytických metód znamená, že výsledky (predikcií) majú objektivizovaný charakter |
| Dopad |  |
| Frekvencia | Aktualizácia krátkodobej predikcie minimálne štyrikrát ročne (po uplynutí každého štvrťroka), predbežné údaje budú k dispozícii na mesačnej báze |
| Potrebné dátové zdroje | Časové rady vývoja základných makroekonomických a odvetvových ukazovateľov slovenskej ekonomiky kvantitatívneho a kvalitatívneho charakteru |
| Analytické modely | Štrukturálne ekonometrické modely na báze kointegrácie |
| Výstup/Analytický produkt | Krátkodobá predikcia vývoja slovenskej ekonomiky (analyticko-prognostický dokument) |
| Najlepšia prax | Krátkodobé predikcie sú v zahraničí štandardnou súčasťou rozhodovacích procesov |

Tabuľka 22: Rýchle odhady HDP a celkovej zamestnanosti v hospodárstve

|  |  |
| --- | --- |
| Prípad použitia | Rýchle odhady HDP a celkovej zamestnanosti v hospodárstve |
| Spôsob použitia | V rámci rozhodovacích procesov na makroúrovni pri návrhu finančných politík, daňových politík a hospodárskych politík |
| Dopad | Použitie analytických metód znamená, že výsledky (rýchlych odhadov) majú objektivizovaný charakter |
| Frekvencia | Online, na týždňovej báze (T+7) |
| Potrebné dátové zdroje | Časové rady vývoja základných makroekonomických a odvetvových ukazovateľov slovenskej ekonomiky kvantitatívneho a kvalitatívneho charakteru |
| Analytické modely | Štrukturálne ekonometrické modely na báze kointegrácie |
| Výstup/Analytický produkt | Odhad rastu a štruktúry HDP (výsledky s analytickým komentárom)  Odhad zamestnanosti, jej rastu a štruktúry (výsledky s analytickým komentárom) |
| Najlepšia prax | Zostavovanie rýchlych odhadov je v zahraničí štandardnou súčasťou štatistickej praxe. Najlepšiu prax napríklad predstadvuje štúdia „Forecasting GDP all over the World: Evidence from Comprehensive Survey Data (2017)[[9]](#footnote-10)“.  ... |

Tabuľka 23: Krátkodobé predikcie vývoja (indexov) priemyselnej produkcie

|  |  |
| --- | --- |
| Prípad použitia | Krátkodobé predikcie vývoja (indexov) priemyselnej produkcie |
| Spôsob použitia | V rámci rozhodovacích procesov na odvetvovej, resp. makroekonomickej úrovni |
| Dopad | Použitie analytických metód znamená, že výsledky (krátkodobej predikcie) majú objektivizovaný charakter. |
| Frekvencia | Klasicky: spravidla štyrikrát ročne  Inovačný prístup: Online, na týždňovej báze (T+7) |
| Potrebné dátové zdroje | Časové rady vývoja základných makroekonomických a odvetvových ukazovateľov slovenskej ekonomiky kvantitatívneho a kvalitatívneho charakteru |
| Analytické modely | Štrukturálne ekonometrické modely na báze kointegrácie |
| Výstup/Analytický produkt | Predikcia vývoja indexov priemyselnej produkcie (prognostický dokument) |
| Najlepšia prax | Krátkodobé predikcie sú v zahraničí štandardnou súčasťou rozhodovacích procesov |

Tabuľka 24: Populačné prognózy

|  |  |
| --- | --- |
| Prípad použitia | Populačné prognózy |
| Spôsob použitia | Konštruované scenáre a podrobné výstupy z modelov budú slúžiť ako vstupy pre ďalšie rozhodovacie procesy na regionálnej úrovni v rôznych oblastiach fungovania spoločnosti (napríklad školstvo, zdravotníctvo, trh práce a podobne) |
| Dopad | Výstupy umožnia formulovanie vývojových scenárov v ďalších oblastiach verejných politík na regionálnej úrovni - trh práce, školstvo, zdravotníctvo a pod. Podporné informácie pre ich nastavenie. |
| Frekvencia | Vzhľadom na dynamiku zmien: raz ročne. |
| Potrebné dátové zdroje | Primárne údaje demografických udalostí v rámci štatistických zisťovaní ŠÚ SR v rámci Hlásení Obyv 1-5/12; Štruktúry obyvateľstva podľa veku, pohlavia, rodinného stavu |
| Analytické modely | Kohortne-komponentný model pre konštrukciu populačných prognóz; Analytické modely za účelom analýzy vývojových trendov a hlavných charakteristických znakov jednotlivých procesov |
| Výstup/Analytický produkt | Regionálne populačné prognózy  Odvodené populačné prognózy  Analytické materiály v kombinácii s podrobnými online dostupnými výsledkami prognostických modelov |
| Najlepšia prax | Nastavenie rozhodovacích procesov na báze údajov/informácií, analýz a podrobných populačných prognóz je praxou vo viacerých zahraničných štátoch i nadnárodných organizáciách (OECD, EUROSTAT, národné banky a podobne) |

#### Riziká

* Nepodarí sa zabezpečiť dostatok kvalitných ľudských zdrojov,
* Analytické jednotky nebudú dostatočne spolupracovať,
* Dátová kancelária si nevybuduje potrebné kompetencie a nepriláka požadovaný talent a nenaplní vďaka tomu očakávania,
* Nepodarí sa získať dostatočný počet inštitúcii, ktoré sa zapoja do centrálneho riešenia (budú sa uprednostňovať lokálne riešenia).

## Požiadavky na Technologická architektúra

#### Súhrnný popis

##### Využitie vládneho cloudu

Riešenie IS KAV bude postavené a nasadené vo Vládnom cloude. Základom pre riešenie. IS KAV bude prevádzkovaný ÚPVIaI a bude poskytovať dostatočný výkon pre zabezpečenie spoľahlivej prevádzky. Systém bude v cieľovom stave pripojený na Centrálnu integračnú platformu.

Pri budovaní aplikačných komponentov v rámci navrhovaného riešenia sa predpokladá maximálne využitie služieb vládneho cloudu. Malo by ísť minimálne o model využívania IaaS:

* redundantné riešenie aplikačných a databázových serverov,
* zabezpečenie LAN infraštruktúry,
* zabezpečenie SAN infraštruktúry,
* zabezpečenie GovNet konektivity
* virtualizácie prostredia,
* zálohovania riešenia.

Pre úspešné nasadenie a prevádzku systému sa tiež odporúča využitie nasledujúcich eGovernment cloudových služieb PaaS:

* Služby konfiguračného manažmentu;
* Služby pre riadenie procesu nasadzovania nových verzií a ich aktualizácie;
* Služby vývojového aplikačného manažmentu a testovacieho prostredia;
* Správu testovacích scenárov a plánovanie testov;
* Služby správy a konfigurácie softvéru;
* Služby pre dohľad nad plynulou a bezpečnou prevádzkou systému.

#### Riziká

* Zložitosť a časová náročnosť riešenia s využitím len existujúcich IaaS služieb (PaaS sa môžu výrazne oneskoriť),
* Neuspokojivá škálovateľnosť riešenia,
* Vymáhanie dohodnutej SLA bude problematické.

## Požiadavky na Bezpečnostná architektúra

#### Súhrnný popis

Základnými východiskami pre rozvíjané riešenie bezpečnosti IS sú rovnako ako v súčasnom stave právne predpisy ako zákon č. 122/2013 o ochrane osobných údajov, zákon č. 275/2006 o informačných systémoch VS a s ním súvisiaci výnos Ministerstva financií Slovenskej republiky č. 55/2014 o štandardoch pre informačné systémy verejnej správy a ďalej ISO/IES 27000, Common Criteria a OWASP Guides a dodatočných požiadaviek prevádzkovateľa systému.

Bezpečnostná architektúra bude vychádzať z týchto pravidiel a v rámci pripraveného Bezpečnostného projektu a zámeru, ktorého vypracovanie a aplikovanie bude podmienkou sprevádzkovania navrhovaných nových, či rozvíjaných systémov. Výstupmi Bezpečnostného projektu budú najmä návrhy postupov pre riadenie prístupov, výkon prevádzky, riešenia incidentov, havarijné plánovanie, implementácie bezpečných zmien a monitorovanie SLA. Návrhy postupov budú zosúladené s už aplikovanými postupmi informačných systémov ÚPVII. Samozrejme bude potrebné zohľadniť aj špecifiká vládneho cloudu, keďže je predpoklad, že riešenie bude využívať služby IaaS a PaaS.

Prevádzkovateľovi KAV bude odporúčané formálne vykonať posúdenie potreby revízie existujúceho bezpečnostného zámeru (z hľadiska kritérií uvedených vyššie) aj v určitom pravidelnom časovom intervale, ktorý vzhľadom na charakter systému je stanovený nie dlhší ako 1 rok. V rovnakej periodicite odporúčame vykonať aj audit bezpečnosti KAV, ktorého súčasťou bude aj overenie spoľahlivosti poskytovania jeho služieb.

Prehodnotenie aktuálnosti bezpečnostného zámeru spolupracujúceho KAV je vzhľadom na povahu projektu budovania KAV a spôsobu spracúvania jeho bezpečnostného projektu navrhnuté vykonať vzhľadom k nasledujúcim situáciám:

* schválenie legislatívnych úprav, ktoré upravia zavedenie a používanie KAV – najmä vzhľadom na stanovené podrobnosti spracúvania osobných údajov a pravidlá prevádzkovania IS,
* po nasadení KAV do rutinnej prevádzky – v tomto prípade ide o overenie správnosti skutočností predpokladaných v tomto bezpečnostnom zámere a ich súlad s realitou.

Základnými požiadavkami na bezpečnosť riešenia KAV sú:

* zabezpečenie ochrany osobných údajov pri analytickom spracovaní,
* zabezpečenie autenticity dátových zdrojov,
* zabezpečenie autenticity výstupov a výsledkov analýz.

#### Riziká

* Umožnenie prístupu neoprávneným osobám a autorizačné nedostatky,
* Systém pseudonymizácie bude mať nedostatky a dôjde k úniku osobných údajov,
* Nedostatočné vybudovanie bezpečnostných technológií a komponentov vo vládnom cloude v čase spustenia projektu.

1. <https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining> [↑](#footnote-ref-2)
2. <https://www.bi.team/publications/using-data-science-in-policy/> [↑](#footnote-ref-3)
3. <https://www.epri.ca/tax-linkage/> [↑](#footnote-ref-4)
4. <https://eacea.ec.europa.eu/national-policies/eurydice/content/tackling-early-leaving-education-and-training-europe-strategies-policies-and-measures_en> [↑](#footnote-ref-5)
5. <https://www.cityofboston.gov/mayorsdashboard/> [↑](#footnote-ref-6)
6. <https://www.bi.team/wp-content/uploads/2017/12/BIT_DATA-SCIENCE_WEB-READY.pdf> [↑](#footnote-ref-7)
7. <https://www.minedu.sk/upraveny-pocet-ziakov-prveho-rocnika-v-dennej-forme-studia-v-cleneni-na-jednotlive-studijne-odbory-alebo-na-jednotlive-ucebne-odbory-strednych-skol-pre-skolsky-rok-20192020/> [↑](#footnote-ref-8)
8. <https://www.epri.ca/tax-linkage/> [↑](#footnote-ref-9)
9. <https://mpra.ub.uni-muenchen.de/81772/1/MPRA_paper_81772.pdf> [↑](#footnote-ref-10)