Príručka pre realizáciu projektu v oblasti lepšieho využívania údajov inštitúcií verejnej správy

Obsah

1 Základné informácie 2

1.1 Manažérske zhrnutie 2

1.1.1 Postup definície projektu pre lepšie využívanie údajov 2

1.1.2 Oprávnené aktivity projektu 3

1.2 Zoznam obrázkov 4

2 Koncept projektu lepšieho využívania dát 5

2.1 Kedy je dátový projekt dobrý 10

2.2 Na čo určite nezabudnúť pri realizácií projektu 11

2.3 Výber témy a problémy 12

2.3.1 Základné témy a typy projektov 12

2.3.2 Vzorové príklady definovania problémov 13

2.3.3 Finančné alokácie a limity 14

2.4 Prípady použitia a analytické produkty 15

2.4.1 Prípady použitia 16

2.4.2 GAP analýza súčasného stavu 16

2.4.3 Výber štandardizovaných analytických metód 17

2.5 Požiadavky na dáta 19

2.5.1 Požadované dátové zdroje a ich vyhodnotenie 19

2.5.2 Vyhodnotenie technickej a dátovej pripravenosti (príklad) 20

2.6 Potrebná funkcionalita 22

2.7 Použitie v praxi 25

2.7.1 Používatelia riešenia 26

2.7.2 Dopad na rozhodovacie procesy a záväznosť rozhodnutí 26

2.8 Potrebné zmeny 27

2.8.1 Organizačné úpravy 27

2.8.2 Úpravy procesov 28

2.8.3 Úpravy legislatívy 28

2.9 Náklady riešenia 29

2.10 Prínosy navrhovaného riešenia 30

2.10.1 Otvorenosť a transparentnosť 31

2.10.2 Posilnenie dátovej ekonomika 31

2.10.3 Zvýšenie spotrebiteľskej a spoločenskej hodnoty 31

2.10.4 Využitie dátovej vedy na optimalizáciu 31

3 Služby konsolidovanej analytickej vrstvy 32

# Základné informácie

## Manažérske zhrnutie

Predstavený dokument sa venuje otázke, ako efektívne využiť možnosti, ktoré prinášajú analytické metódy a lepšie využívanie dát v praxi na jednotlivých organizačných jednotkách verejnej správy. Základnou odpoveďou je potreba sústrediť sa na výsledky a realizáciu, čo znamená navrhnúť vhodné projekty. Venujeme sa preto návrhu takzvaného vzorového projektu, ktorý posúdi dôležité aspekty podľa nasledujúcej schémy:

### Postup definície projektu pre lepšie využívanie údajov



Aby inštitúcia dokázala navrhnúť vhodný projekt, je potrebné:

* Stanoviť si základnú tému a identifikovať dôležité problémy, ktoré vďaka lepšiemu využívaniu dát dokážeme vyriešiť,
* Presne špecifikovať prípad použitia a stanoviť, kto bude používať výsledky analýz a akým spôsobom,
* Identifikovať potrebné dátové zdroje a ďalšie vstupy, ktoré bude potrebné zabezpečiť počas implementácie projektu, ako i prevádzky riešenia,
* Stanoviť, aké nástroje a technologické funkcie sú potrebné resp. vhodné v rámci danej metódy,
* Zamyslieť sa nad používaním výsledkov riešenia v praxi,
* Pripraviť plán zmien a nastaviť kroky pre ich implementáciu,
* Odhadnúť náklady projektu,
* Odhadnúť prínosy projektu.

Vyššie uvedený postup je pretransformovaný do Štúdie uskutočniteľnosti, ktorá predstavuje základný dokument na vypracovanie a posúdenie projektu. Definovaný postup je potrebné rozpracovať v nasledovných častiach štúdie:

|  |  |  |
| --- | --- | --- |
| **Krok definovania projektu** | 🡺 | **Časť štúdie v ktorej je potrebné krok popísať** |
| Téma a problémy | 🡺 | Výber rozsahu projektu |
| Prípady použitia a analytické produkty | 🡺 | Biznis architektúra TO BE |
| Potrebné údaje a vstupy | 🡺 | Aplikačná architektúra TO BE |
| Potrebné nástroje a funkcie | 🡺 | Aplikačná architektúra TO BE |
| Využitie riešenia a dopady | 🡺 | Motivácia |
| Potrebná zmena a jej implementácia | 🡺 | Implementácia a migrácia |
| 🡺 | Legislatíva TO BE |
| Náklady | 🡺 | Ekonomická analýza -> Rozpočet projektu |
| Benefity | 🡺 | Ekonomická analýza -> Analýza benefitov |

Aby sa dokázali spustiť projekty pre lepšie využitie dát, je potrebné zabezpečiť financovanie takého programu. V súčasnosti je najlepším spôsobom využitie zdrojov z Operačného programu Integrovaná infraštruktúra, Prioritná os 7 Informačná spoločnosť.

Pre definíciu projektu žiadatelia použijú vzorový dokument štúdie uskutočniteľnosti.

### Oprávnené aktivity projektu

Doporučené je využiť mechanizmus dopytových projektov na základe jasne a štruktúrovane stanovenej dopytovej výzvy a to s nasledujúcimi oprávnenými aktivitami:

* A1: Analýza prípadov použitia,
* A2: Zabezpečenie zdrojov dát,
* A3: Nastavenie funkcionalít,
* A4: Realizácia dátového modelu a spracovanie analýzy,
* A5: Publikovanie výstupov, analytických produktov a otvorených údajov
* A6: Zavedenie zmien do praxe.

Ako dôležitá podmienka realizácie je okrem zabezpečenie efektivity investícií a súladu s architektonickým konceptom, vnímaná najmä podpora otvorenosti dát a riešení a celkový priaznivý dopad na analytický ekosystém, ktorý v štáte postupne vzniká.

V zmysle oprávnených aktivít podľa príručky pre žiadateľa je zaradenie vecných aktivít nasledovné:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Aktivity | Analýza a návrh | Implementácia | Testovanie | Nasadenie | HW a SW |
| Analýza prípadov použitia | X |  |  |  |  |
| Zabezpečenie zdrojov dát | X | X | X | X |  |
| Nasadenie funkcionalít | X | X | X | X | X |
| Realizácia dátového modelu | X | X | X | X |  |
| Publikovanie výstupov | X | X | X | X |  |
| Zavedenie zmien do praxe |  |  |  | X |  |

Projekty pre lepšie využívanie dát by mali byť postavené nad Konsolidovanou analytickou vrstvou, ktorú začala budovať Dátová kancelária verejnej správy. Z tohto dôvodu budú všetky projekty koordinované Dátovou kanceláriou.

## Zoznam obrázkov

[Obrázok 1: Základná logická štruktúra projektu pre lepšie využitie dát 7](#_Toc17715917)

[Obrázok 2 - Koncept manažmentu údajov vo verejnej správe 8](#_Toc17715918)

[Obrázok 3: Logické pohľad na využitie analytickej vrstvy 9](#_Toc17715919)

[Obrázok 2: Checklist realizácie projektu 11](#_Toc17715920)

# Koncept projektu lepšieho využívania dát

Projekty v oblasti lepšieho využívania dát by mali mať svoju jasnú štruktúru a definované prínosy, dopady a prípady použitia v reálnom čase. Preto je potrebné pri koncipovaní projektu nájsť odpovede na základné otázky, ktoré ktoré obsahujú najdôležitejšie prvky projektu z rôznych uhlov pohľadu.

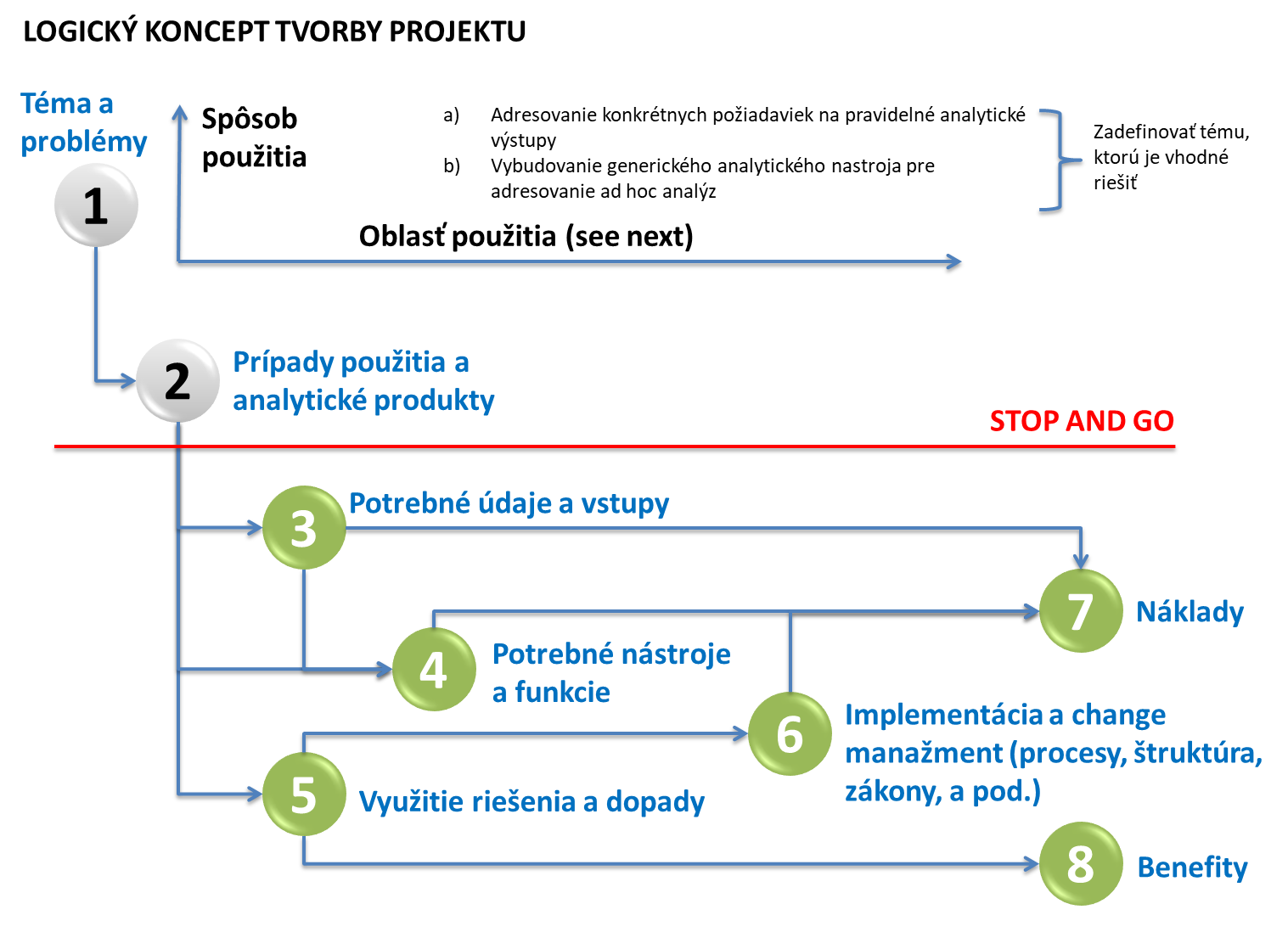
Jednoducho povedané, jedná sa o minimálny súbor otázok, na ktoré by mali byť schopné organizácie verejného sektora odpovedať pred samotných nadizajnovaním detailného projektu v oblasti využívania dát, Big Data a BI.

Samotný kontext realizácie zamýšľaného projektu vytvára náročnosť zodpovedania predmetných otázok. Preto je potrebné venovať dostatočnú pozornosť práve pri plánovaní projektu a vyhodnotení jeho:

* Kontextu
* Komplexnosti
* Následkov
* Procesnej zložitosti
* Rizikovosti realizácie

|  |  |
| --- | --- |
| **OBLASŤ / OTÁZKA** | **ODPOVEĎ / Príklad** |
| Bullseye**ÚČEL**  Prečo je algoritmus / riešenie potrebný a aké výsledky má umožniť | Vlastný text |
| Workflow**VYUŽITIE**  V akých procesov a okolnostiach je vhodné projekt / riešenia využiť | Text |
| Family with boy**DOPAD**  Aké následky (dobré aj zlé) má použitie riešenia na ľudí | Text |
| Group brainstorm**PREDPOKLAD**  Na akých prepdokladoch je riešenie postavené a aké sú limity a bariéry požitia | Text |
| Checklist**DATA**  Na akých datsetov bude riešenie postavené a aké sú limity a bariéry | Text |
| Plug**VSTUPY**  Aké existujúce alebo nové údaje sú potrebné pre vytvorenie riešenia pre potreby rozhodovania | Text |
| Unlock**MITIGÁCIA**  Aké aktivity musia byť prijaté na zníženie negatívnych dopadov, ktoré vyplývajú z limitov a bariér využitia | Text |
| Head with gears**ETIKA RIEŠENIA**  Aké hodnotenie etiky využitia riešenia bolo zrealizované | Text |
| Magnifying glass**VÝHĽAD**  Do akej miery je potrebný ľudský úsudok pred algoritmom a kto je zodpovedný za jeho správne používanie | Text |
| Scales of justice**HODNOTENIE**  Ako a na základe akých kritérií bude riešenie hodnotené a kým | Text |

Tabuľka 1: 10 otázok na využitie riešení pre lepšie využívanie dát pri rozhodovaní vo verejnej správe



Obrázok 1: Základná logická štruktúra projektu pre lepšie využitie dát

Logický koncept tvorby projektu predstavuje základný rámec a nastavený postup, ako pristupovať k vytvoreniu dobrého projektu. V prvom rade je potrebné mať detailne pochopené a popísané problémové oblasti, ktoré sa realizáciou projektu majú vyriešiť. Je vhodné zamyslieť sa pri definovaní spôsobov a oblastí použitia aj nad nasledujúcimi princípmi resp. tézami:

**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. 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 vytváranej štúdie by mala byť snaha navrhnúť vhodné riešenie a postup pre maximálne využitie dát v definovanej problémovej oblasti a overiť definované spôsoby založené na dátovej vede a analytických prístupoch priamo vo rozhodovaní v predmetnej oblasti.

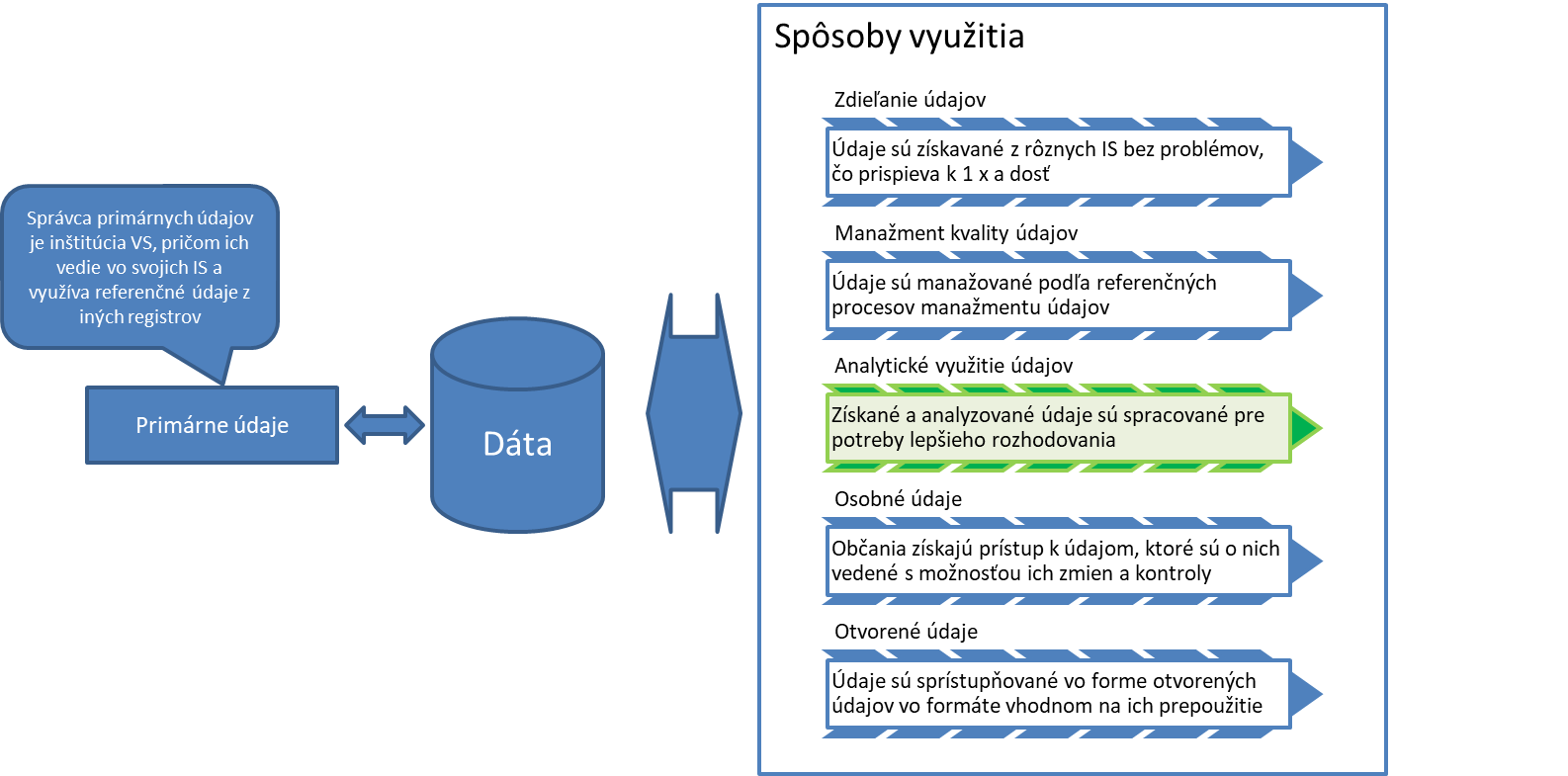
Inšpiráciou sú najproduktívnejšie prípady použitia dátovej vedy vo verejnom sektore v zahraničí ale i spôsob využívania dát vo verejnom sektore (napríklad dátovej vede sa venuje britský Behavioural Insights Team, ktorý vydal správu Použitie dátovej vedy pri návrhu politík[[1]](#footnote-2)).

PRÍKLAD Z REPORTU

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 2 - 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.

Štúdia by v tejto oblasti mala podporiť aj najmä nasledovné opatrenia súvisiace s manažmentom údajov, ktoré sú potrebné pre naplnenie analytických požiadaviek organizácie:

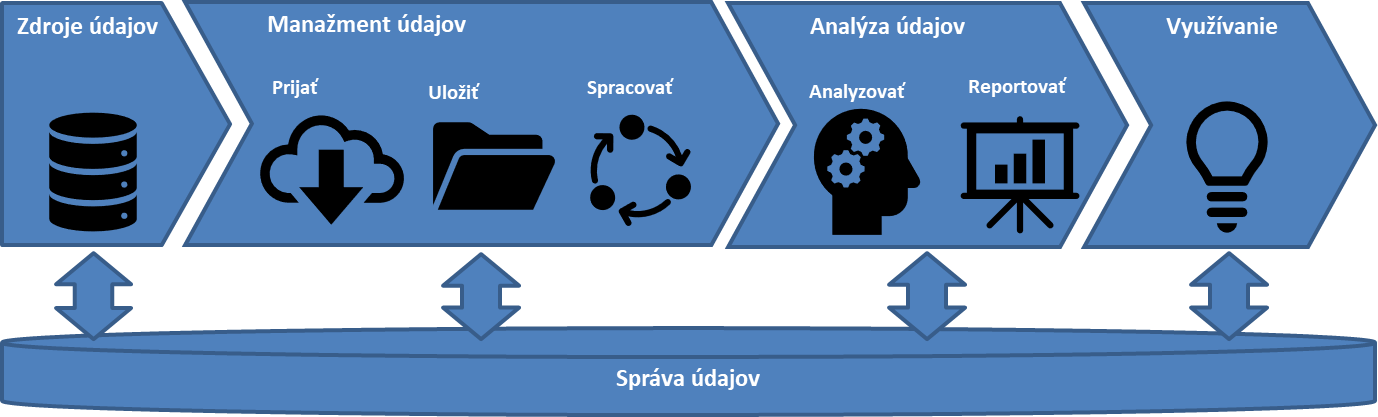
1. Analytické využitie údajov: aby organizácia dokázala využívať svoje údaje pre potreby prípravy analýz (analytické spracovanie údajov), ktoré budú slúžiť ako podklad pre lepšie rozhodovanie.
2. Manažment osobných údajov: navrhované riešenia (ak je to možné) sú adresné s využitím osobných údajov transparentným spôsobom. Výsledky sú prístupné občanovi/podnikateľovi, pričom na základe nich sa môže rozhodovať.
3. Publikovanie otvorených údajov: projekt vytvorí údaje, ktoré budú publikované ako otvorené údaje vo vhodnom na opätovné použitie – okrem osobných údajov, citlivých údajov a utajovaných údajov.
4. Manažment kvality údajov: procesy v rámci projektu sú nastavené tak, že budú aplikované správne postupy manažmentu údajov, pričom projekt bude využívať správne údaje a bude možné sa spoľahnúť na ich správnosť.

**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. Organizácia by mala 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.

Projekt by mal zabezpečiť, aby boli procesy a postupy v inštitúcií nastavené tak, aby boli využívané vhodné a správne údaje a 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 organizácie a jej procesov tak, aby boli tieto definované analýzy efektívne používané a zároveň zverejňované vo vhodnej vizuálnej podobe pre aj pre verejnosť.

Obrázok 3: Logické pohľad na využitie 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 smerovať projekt k jednej alebo viacerým oblastiam, v ktorých rozhodovanie na základe dát je opodstatnené. Jedná sa o nasledovné oblasti:

* 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 (analýzy rizík, spracovanie a posudzovanie podaní, fraud management, prediktívne analýzy a podobne).

Vzhľadom na výber oblastí, v ktorých bude projekt aplikovaný je potrebné:

* nadefinovať správne požiadavky a potreby pre jednotlivé oblasti analýzy
* nájsť vhodnú kombináciu nástrojov pre definované problémy
* experimentovať a postupne zlepšovať rozsah a šírku analýz a šíriť najlepšie skúsenosti (Recept na úspech tiež nie je možné určiť hneď na začiatku)

**Spôsoby použitia a definovanie zdrojov a modelov**

V prvom rade je potrebné, aby existoval tím (analytická alebo iná jednotka), ktorý 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 v rámci projektu vytvoria, vrátane definovaných výstupov, ktoré nazývame Dátové produkty. Na prípravu analytických produktov bude potrebné zadefinovať analytické nástroje a informačné systémy a v neposlednom rade je nevyhnutné zabezpečiť potrebné dáta, či už nákupom, alebo integráciou na administratívne zdroje.

***Príklad definovania prípadov použitia a dátových produktov***

|  |  |  |
| --- | --- | --- |
| 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 |

## Kedy je dátový projekt dobrý

Pre kvalitný dátový projekt sú nevyhnutné nasledovné 4 zložky:

1. Predvídateľné problémy alebo potreba pochopenia neštruktúrovaných údajov – musí existovať merateľné správanie alebo hodnotenie, ktoré je možné predvídať a musí byť priamo relevantný pre riešenú oblasť.

Príklad: V prípade školských zariadení bolo predvídateľné na základe historických údajov, ktoré školy si vyžadujú zlepšenia alebo neprimeranú inšpekciu

2. Vysoko kvalitné, rozsiahle a vhodné údaje – v prípade údajov musí existovať dobrá zhoda medzi problémom a dátami, ktoré môžu byť použité na jeho vyriešenie. Aj napriek tomu, že jednotlivé projekty sú rozdielne majú dve spoločné témy:

* Údaje, ktoré sú na rovnakej úrovni ako správanie, ktoré nás zaujíma, sú užitočnejšie ako údaje, ktoré pokrývajú širšiu oblasť alebo skupinu
* Ak sa vyžadujú textové alebo obrazové údaje, je potrebné venovať osobitnú pozornosť tomu, aby sa zabezpečila vysoká kvalita údajov, primerane úplné a ľahko extrahovateľné údaje.

3. Schopnosť a buy-in v rezorte implementovať zistenia v praxi – tento fakt je rozhodujúci, aby sa rozhodovanie nevrátilo späť k rozhodovaniu založenému na tradičnom myslení

4. Právna a etická vôľa na implementáciu riešení – je veľmi dôležité na túto oblasť myslieť dostatočne včas:

* Veda o údajoch má obzvlášť silné dopady na súkromie
* Adoptovanie legislatívy pre podporu rozhodovania založenom na automatizovanom spôsobe vyhodnocovania faktov
* Údaje, ktoré môžu byť využívané na trénovanie algoritmov, môžu byť samy o sebe subjektom obmedzení / zaujatosti (príklad: ak boli iba Afroameričania historicky neprimerane obviňovaní určitými trestnými činmi kvôli zaujatosti polície, algoritmus zhorší model predpovedania na základe predsudkov, a to tým, že Afroameričania predstavujú vyššie riziko)
* Algoritmy, ktoré využívajú osobné údaje ako napr. pohlavie alebo rasa, musia byť vhodne navrhnuté tak, aby nedochádzalo k skresleniam vo vzťahu ku skupinám, ktorých sa údaje týkajú

## Na čo určite nezabudnúť pri realizácií projektu

Realizovaním projektu, ktorého výsledkom má byť predovšetkým zlepšenie rozhodovania na základe údajov a implementovaných algoritmov (pre rôzne sféry verejného života) musí byť zabezpečená rovnako transparentnosť a otvorenosť údajov. Preto bude nevyhnutné zodpovedať na nasledujúcich 10 otázok, ktoré vytvoria rámec transparentnosti projektu ako takého, ako aj výsledkov implementácie navrhovaných zmien a udržateľnosť výstupov projektu.

|  |  |
| --- | --- |
| # | Povinnosť |
| **1** | *Každý algoritmus / riešenie by mal byť doplnený opisom jeho funkcie, cieľov a zamýšľaného vplyvu, ktorý by bol k dispozícii tým, ktorí ho používajú.* |
| **2** | *Organizácie verejného sektora by mali zverejňovať podrobnosti o údajoch na základe, ktorých boli alebo sú algoritmy / riešenia založené, ako aj predpoklady použité na vytvorenie spoločne s opisom a hodnotením prípadných rizík alebo skreslení..* |
| **2** | *Algoritmy / riešenia by mali byť kategorizované podľa rizikovej stupnice a to od 1-5, pričom 5 sa týka tých, ktorých vplyv na jednotlivca by mohol byť veľmi vysoký a 1 je veľmi malý.* |
| **4** | *Bude zverejnený zoznam všetkých vstupov použitých pre vytvorenie algoritmu / riešenia pre realizáciu rozhodnutí* |
| **5** | *Občania musia vedieť, že riešenie ich problémov je úplne alebo čiastočne realizované algoritmom* |
| **6** | *Každý algoritmus / riešenie musí mať identické prostredie pre potreby auditu na otestovanie dopadov rôznych vstupných premenných.* |
| **7** | *Ak budú vytvorené algoritmy / riešenia tretími stranami v ich mene, mali by organizácie vybrať takých dodávateľov, ktorí splnia podmienky 1 – 6* |
| **8** | *Menovaný senior manažér by mal byť spolu zodpovedný za akékoľvek rozhodnutie podporené vytvoreným algoritmom resp. riešením* |
| **9** | *Organizácie verejného sektora, ktoré chcú prijať algoritmické rozhodovanie v oblastiach s vysokým rizikom, by sa mali poistiť, pre potreby kompenzácie chýb a negatívnych dopadov na jednotlivcov, ktorí budú negatívne ovplyvnení chybným rozhodnutím na základe algoritmu.* |
| **10** | *Organizácie by sa mali zaviazať k vyhodnocovaniu vplyvov algoritmov alebo riešení, ktoré využívajú na rozhodovanie a mali by publikovať ich výsledky* |

Obrázok 2: Checklist realizácie projektu

## Výber témy a problémy



**Táto časť sa nachádza vo vzorovej štúdii v časti Rozsah**

### Základné témy a typy projektov

Výber témy a definovanie problému predstavuje základný konceptuálny prvok, ktorý je potrebné popísať za účelom predstavenia celého projektového zámeru. Pri definovaná príkladov resp. oblasti použitia možno vychádzať z nasledovnej tabuľky, pričom v rámci vzorového projektu je potrebné vybrať jednu oblasť -> Typ projektu a popísať v kontexte vybraného typu projektu problém a jeho väzbu práce na danú oblasť.

|  |  |
| --- | --- |
| Typ projektu | Popis |
| Lepší návrh politík a regulácií | Vďaka online posudzovaniu vplyvov a využitie údajov na simulácie dopadov a testovanie účinnosti politík:  (Vznikne platforma na posudzovanie vplyvov a lepší návrh regulácii s využitím „big data“ a umelej inteligencie, okrem analýz vplyvov na podnikateľské prostredie bude potrebné vytvoriť modely sociálnych vplyvov, vplyvov na životné prostredie, vplyvov na zdravotný stav populácie a podobne) |
| Lepší dozor a dohľad nad regulovaním prostredím: | Využitie údajov pre online monitoring regulovaného prostredia a zavedenie princípov Regulácie 2.0, čo môže byť využité napríklad v procesoch verejného zdravotníctva, pri povoleniach životného prostredia, v podmienkach kontrol inšpektorátov práce, pri sledovaní telekomunikačného trhu, pri sledovaní finančných trhov a podobne |
| Spojenie úradníka a stroja: inovácie procesov | Vďaka zdieľaniu údajov a využitiu automatizovaných analýz prípadov, využitie podporných analytických nástrojov pre lepšie operatívne rozhodovanie (napríklad použitie metód „machine learning“ pre analýzu rizík a predikciu budúcich udalostí alebo analýzy sociálnych sietí pre pochopenie súvislostí). V princípe ide o rozšírenie znalostnej bázy úradníkov |
| Prediktívne kontroly | Napríklad využitie AI v kontrolnej činnosti NKÚ, v kontrolnej činnosti verejného obstarávania, daňové kontroly, colné kontroly |
| Automatizácia spracovania | Môže sa jednať napr. o podania, vďaka preskúmaniu podkladov a ich úvodného vyhodnotenia strojovo, napríklad pri podávaní žalôb, žiadostí o stavebné konanie a podobne |
| Určovanie opatrení na základe rizík: | Napríklad návrh vhodnej podoby trestu, sociálne opatrenia pre deti v núdzi, preventívne opatrenia pre minimalizáciu škôd krízových situácií |
| Lepšie riadenie zdrojov a plánovanie činností | Jedná sa o aplikáciu modelov na využitie predikcií budúcich udalostí |
| Plánovanie budúcich kapacít, | Na základe simulácie budúceho dopytu po verejných službách, napríklad počet miest v škôlkach, počet lôžok v nemocniciach, počet úradníkov |
| Prediktívna polícia a prediktívne hasičstvo | Využitie umelej inteligencie pre plánovanie policajných hliadok spôsobom, aby sa minimalizovala možná trestná činnosť |
| Územné plánovanie | Využitie modelov pre efektívne plánovanie územného rozloženia v závislostí od požiadaviek |
| Zvýšenie kvality služieb | Vďaka zavádzaniu automatizovanej obsluhy (cez „chatbotov“ alebo osobných asistentov pri využití hlasového rozhrania), vďaka automatizovanému spracovaniu podaní, vytvárania znalostnej bázy pre úradníkov a pracovníkov obsluhy |
| Zvýšenie výkonnosti vnútorných procesov | Využitie umelej inteligencie pre manažment ľudských zdrojov a/alebo celkové riadenie organizácie a jej výkonnosti (vrátane riadenia a zvyšovania výkonnosti podriadených organizácií), napríklad pre náber vhodných pracovníkov, pre plánovanie zmien, pre odmeňovanie, pre počítanie výkonnostných ukazovateľov, pre riešenie verejného obstarávania a podobne |
| Optimalizácia prevádzky organizácií verejnej správy | Identifikácia neefektívnosti v prevádzke jednotlivých inštitúcii, benchmarking nákladov na jednotlivé aktivity, optimalizácia využitia hmotného a nehmotného majetku, optimalizácia podporných a administratívnych činností |
| Využitie decentralizovaných technológií | Decentralizované riešenia, ktoré môžu vzniknúť vďaka technológii decentralizovanej hlavnej knihy alebo aj blockchain majú potenciál disruptovať inštitúcie, ako ich poznáme, vďaka vytvárania vrstvy dôvery medzi účastníkmi transakcií. Na zabezpečenie dôveryhodnosti transakcií už nie je potrebná dôveryhodná tretia strana. Okrem inštitucionálnej inovácie môže decentralizovaná architektúra prispieť k vytvoreniu spravodlivejšieho internetu, ktorý je viac v súlade s Európskymi hodnotami, ako súčasný model |

### Vzorové príklady definovania problémov

V nasledujúcich tabuľkách sú uvedené definované problémy ako vzory resp. príklady. Vami definované problémy by mali byť popísané minimálne v nižšie uvedenej granularite tak, aby bolo jasné čo sa bude projektom riešiť a čo bude jeho výsledkom.

|  |  |
| --- | --- |
| Príklad 1 | Identifikácia nedostatočne výkonných praktických lekárov |
| Téma | Využitie rozhodovacích stromov ako nástroja benchmarkingu v rámci siete poskytovateľov primárnej zdravotnej starostlivosti |
| Problém | Nedostatočne výkonní praktickí lekári.  Praktickí lekári majú hlboký vplyv na naše zdravie - sú prvým kontaktným bodom pre mnoho neodkladných lekárskych prehliadok a dôležitý kanál pre propagáciu verejného zdravia.  Avšak kvôli veľkému počtu a širokej škále podmienok je možné vidieť rôznu úroveň profesionálnych zručností, ktoré nespĺňajú požadované štandardy. Samotné inšpekcie praktických lekárov sú zdĺhavé a neodhalia všetky problémové oblasti. |

|  |  |
| --- | --- |
| Príklad 2 | Predikcia budúcich problémov detí zo sociálne slabšieho prostredia |
| Téma | Vytvorenie predikčného modelu pre sociálnych pracovníkov, ktorý im dokáže predpovedať, že prípad je potrebné eskalovať |
| Problém | Náročné eskalovanie problémov sociálnymi pracovníkmi  Výkon činností sociálnych pracovníkov je jednou z najťažších prác v oblasti verejného a sociálneho priestoru. Sociálni pracovníci, ktorí vykonávajú posudzovanie zaobchádzania s deťmi, by mohli naraz vybaviť až 50 prípadov. Sociálny pracovníci sú zodpovedajú za rýchle posúdenie toho, či je dieťa vystavené riziku možného poškodenia alebo je potrebná jeho ochrana alebo je potrebné jeho vzatie do starostlivosti ako súčasť súdneho opatrenia. |

### Finančné alokácie a limity

Finančné limity na úrovni celého projektu sú stanovené v súlade s obdobnými dopytovými projektami a predpokladaným rozsahom projektov pre lepšie využívanie údajov inštitúciami verejnej správy:

Minimálna výška nenávratného finančného príspevku na projekt sa stanovuje v sume 80 000 EUR.

Vzhľadom na rozličný charakter jednotlivých typov projektov je maximálna výška nenávratného finančného príspevku na projekt stanovená špecificky pre každý typ projektu:

**Indikatívne limity pre jednotlivé typy projektov**

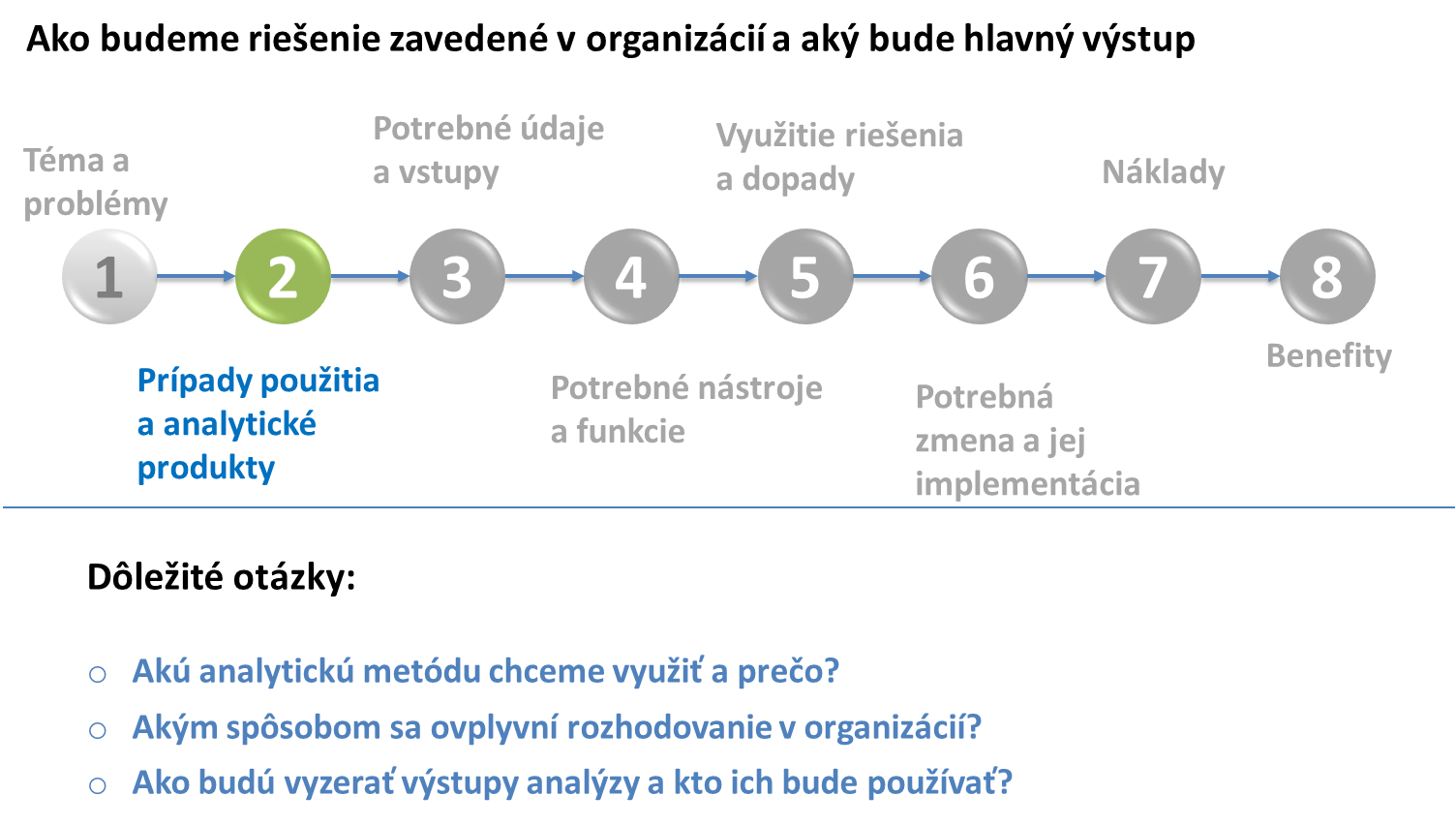
| **ID** | **Typ projektu** | **Finančný limit (EUR)** |
| --- | --- | --- |
| 1 | Lepší návrh politík a regulácií | <= 3 000 000.00 |
| 2 | Lepší dozor a dohľad nad regulovaním prostredím | <= 3 000 000.00 |
| 3 | Spojenie úradníka a stroja | <= 2 000 000.00 |
| 4 | Prediktívne kontroly | <= 2 000 000.00 |
| 5 | Automatizácia spracovania podaní | <= 1 000 000.00 |
| 6 | Určovanie opatrení na základe rizík | <= 1 000 000.00 |
| 7 | Lepšie riadenie zdrojov a plánovanie činností | <= 2 000 000.00 |
| 8 | Plánovanie budúcich kapacít | <= 2 000 000.00 |
| 9 | Prediktívna polícia a prediktívne hasičstvo | <= 1 000 000.00 |
| 10 | Územné plánovanie | <= 1 000 000.00 |
| 11 | Zvýšenie kvality služieb | <= 2 000 000.00 |
| 12 | Zvýšenie výkonnosti vnútorných procesov | <= 4 000 000.00 |
| 13 | Optimalizácia prevádzky | <= 5 000 000.00 |
| 14 | Využitie decentralizovaných technológií | <= 1 000 000.00 |

Tabuľka 6: Maximálna a minimálna výška príspevku

Pre projekty do 1 000 000 EUR za všetky zdroje financovania je potrebné predložiť Analýzu celkových nákladov na vlastníctvo (TCO).

Pre projekty nad 1 000 000 EUR za všetky zdroje financovania je potrebné predložiť Analýzu nákladov a prínosov (CBA) a Analýzu celkových nákladov na vlastníctvo (TCO).

## Prípady použitia a analytické produkty



Prípady použitia predstavujú konkrétne situácie, v ktorých alebo pre ktoré majú byť analytické činnosti realizované. V rámci projektu je pre každý prípad použitia definovaná oblasť v ktorej je prípad použitia aplikovaný. Ako ďalšou časťou je vypracovaná GAP analýza súčasného riešenia a to z pohľadu:

* Súčasnej situácie – čo existuje a ako prebieha daný prípad použitia s definovaním nasledovných oblastí:
  + Aké sú problémy technického riešenia
  + Aké sú problémy dátového riešenia
  + Aké sú problémy procesného a biznis riešenia
  + Aké sú problémy legislatívy
* Čo pre implementáciu navrhovanej situácie potrebujeme
* Následne je definovaný analytický produkt alebo výstup ako výsledok realizácie projektu.

### Prípady použitia

Každý prípad použitia má štruktúrovanú podobu, ktorá je znázornená v nasledujúcej tabuľke. Vo vzorovom projekte je potrebné popísať jednotlivé polia v zmysle nižšie uvedeného.

***Táto časť sa nachádza vo vzorovej štúdii v časti Popis budúceho stavu -> Biznis architektúra***

|  |  |  |  |
| --- | --- | --- | --- |
| **Prípad použitia** | **Text prípadu použitia** | | |
| Spôsob použitia | V tejto časti je potrebné detailnejšie popísať, ako sa daný use case bude „správať“ v realite | | |
| Dopad a využiteľnosť | V prípade dopadu je potrebné definovať ako budú prebiehať nasledovné kroky v súvislosti s vytvoreným modelom / algoritmom alebo riešením a čo to prinesie. Zároveň je potrebné popísať ako sa bude využívať | | |
| Využité analytické metódy[[2]](#footnote-3) | Jedná sa o popis analytických metód, ktoré budú v rámci riešenia použité na vytvorenie finálneho analytického produktu, resp. výstupu projektu. Malo by sa jednať o využitie štandardizovaných analytických metód a v tejto časti by malo byť popísané, ako sa daná metóda bude aplikovať. | | |
| Frekvencia analýz[[3]](#footnote-4) | Choose an item. | Výstup/ analytický produkt | V tejto časti bude zadefinovaný výstup prípadu použitia s krátkym popisom. Môže sa jednať o analytický produkt (model, scenáre a pod.) alebo o dokumente resp. aktualizovanú stratégiu a pod. |

V prípade, ak bude v rámci projektu riešených viac prípadov použitia, je možné nakopírovať tabuľku viac krát

### GAP analýza súčasného stavu

Jedná sa o posúdenie súčasného stavu riešenia definovaného prípadu použitia a jeho odchýlok od optimálneho riešenia, ktoré by sa malo práve realizáciou projektu dosiahnuť. ***GAP analýza sa nachádza v časti štúdie Popis súčasného stavu -> Architektúra***

| **Oblasť** | **Prípad použitia** | **GAP analýza** |
| --- | --- | --- |
|  | Potrebné identifikovať vždy prípad použitia, voči ktorému sa GAP analýza realizuje. | Ako prebieha v súčasnosti daný prípad? |
| Popis súčasného riešenia analyzovaného prípadu použitia. Jedná sa o definovanie procesu riešenia, zodpovedností, využívaných nástrojov a pod. |
| Problémy technické riešenia |
| V tejto časti je potrebné popísať technické obmedzenia a problémy resp. nedostatky |
| Problémy dátového riešenia |
| V prípade, ak existujúcemu riešeniu chýbajú práve údaje, alebo nie sú dostatočne kvalitné, je potrené túto skutočnosť opísať. |
| Problémy legislatívy |
| Ak existujú legislatívne obmedzenia na prípadné zavedenie inovácií alebo nových postupov, uveďte v tejto časti tieto skutočnosti. |
| Čo je pre vytvorenie riešenia potrebné |
| Pre každú oblasť v rámci GAP analýzy stručne popíšte aj návrh riešenia a prípadne doplňte ďalšie potreby, ktoré bude nevyhnutné pre realizáciu projektu zabezpečiť. |

### Výber štandardizovaných analytických metód

Analytická technika (analytická metóda) je postup alebo spôsob vykonania rozboru nejakého problému, stave alebo skutočnosti. Analytické techniky sú spravidla časovo i úlohovo obmedzené (jednorázovo používané pre riešenie danej témy) na rozdiel od metód riadenia, ktoré ovplyvňujú konanie v organizácii v dlhšom časovom horizonte.

Existuje a v praxi sa používa mnoho úplne jednoduchých analytických techník, ktoré manažéri alebo analytici pri bežnej práci používajú. Často bez toho, aby ich nejako pomenovávali. Na také techniky stačí často systém “ceruzka-papier” čiže bežné kancelárske vybavenie - sú založené predovšetkým na skúsenostiach toho, kto ich používa. Existuje tiež celá rada špecializovaných analytických techník, ktoré sú vo veľkej väčšine založené na nejakom matematickom modeli alebo vyžadujú určité pomôcky či nástroje.[[4]](#footnote-5)

Táto časť predstavuje skôr pomôcku pre definovanie analytických metód, ktoré je možné použiť pri definovaní riešenia daného prípadu použitia.

|  |  |
| --- | --- |
| **Základné a v praxi najpoužívanejšie analytické metódy** | **Metódy analýzy systémov a metódy popisu architektúr organizácií:** |
| Bostonská matica (BCG matica)  Brainstorming  Benchmarking  Diferenčný analýza (Gap analýza)  Mentálna mapy (Mind Maps)  Paretovo pravidlo (pravidlo 80/20)  SWOT analýza  Šesť otázok | Analýza architektúry organizácie (EA)  Analýza sociálne siete  Analýza pracovných miest (Job Analysis)  Organizačná architektúra  Opis pracovného miesta  Reengineering  Sociogram  Sociometria  Špecifikácia pracovného miesta |
| **Analytické techniky a metódy analýzy a hodnotenia stavu organizácie** | **Analytické techniky pre hľadanie príčin negatívnych javov v organizáciách a systémoch** |
| EFQM  CAF  Procesný audit  Finančný audit  Personálny audit  Hodnotenie stupňov zrelosti CMM | ETA (Event tree analysis) - analýza stromu udalostí - analýza stromu udalostí  FTA (Fault Tree Analysis) - analýza stromu poruchových stavov  FMEA (Failure Modes and Effects Analysis) - analýza spôsobov poškodenia a účinkov  HAZOP (Hazard and Operability Study) - riziková a operačný analýza  Ishikawov diagram - diagram príčin a následkov  PHA (Preliminary Hazard Analysis) |
| **Technicky na riešenie problémov a metódy rozhodovania** | **Simulačné a optimalizačné metódy** |
| Analýza 5F (Five Forces)  Akčný výskum  Davidsonov zlom  Eisenhowerov princíp  Checklandova metodika  DBR (Drum Buffer Rope)  DOE (Design of Experiments)  G8D (Global Eight Disciplines)  Histogram  Box-Jenkinsova metodika  Ishikawov diagram - diagram príčin a následkov  Kontingenčný prístup (Contingency Approach)  Korelačný diagram (Scatter diagram)  Mentálna mapy (Mind Maps)  Metóda kritickej cesty CPM (Critical Path Method)  Metóda kritickej reťaze CCM (Critical Chain Method)  Marketingový mix 4P (Marketing Mix 4P)  Mintzbergov paradox (Paradox of Mintzberg)  Model NIMSAD  MSA (Measurement System Analysis)  Všeobecný prístup k riešeniu problémov  PESTLE analýza  Paralelné tímy  Paretovo pravidlo (pravidlo 80/20)  PPAP (Production Part Approval Process)  Princíp ekvifinality (Equifinality Principle)  QFD (Quality Function Deployment) - rozpracovanie požiadaviek zákazníka, dom kvality  Prekonanie skupinového myslenia  Priebehový diagram (Run chart)  Regresná analýza (Regression Analysis)  Riadiaci graf (Control Chart)  Sloanov filter (Sloanov princíp)  SPC (Statistical Process Control)  Swapping  TOC (Theory of Constraints) - teória obmedzenia  Value Stream Mapping (VSM)  VRIO analýza  Vývojový diagram  WIBI (Would I Buy It?)  RCT (Randomized controlled trial) | Dopadové analýzy (Impact Analysis)  Priestorová optimalizácia  Metóda Monte Carlo simulácie  Metódy sieťovej analýzy  Prognózovanie |

## Požiadavky na dáta



### Požadované dátové zdroje a ich vyhodnotenie

Všetky požadované údaje sú detailne rozpísané v nasledujúcich tabuľkách, pričom je uvedené všetky relevantné údaje pre realizáciu riešenia. V  prípade, ak sa údaje nachádzajú v jednom zdroji (napr. datasete), sú uvedené ako objekty evidencie resp. samostatné entity objektu evidencie v jednej tabuľke. Tabuľku je možné rozširovať podľa potreby dátových zdrojov pre daný prípad použitia.

***Táto časť sa nachádza v štúdií v časti Dátová architektúra projektu***

|  |  |  |  |
| --- | --- | --- | --- |
| **Požadovaný súbor / dataset** | **Názov požadovaného datasetu resp. údajov** | | |
| Opis a účel použitia | Potrebné popísať, aké údaje dataset obsahuje a prečo je potrebné získať práve tieto údaje | | |
| Požadované objekty evidencie | V tejto časti je potrebné identifikovať, ktoré konkrétne objekty evidencie alebo entity v rámci datasetu budú pre riešenie využívané | | |
| Zdroj | Stanoví a vlastník údajov | Potrebný nákup[[5]](#footnote-6) |  |
| Forma poskytovania údajov | Potrebné zadefinovať v akej štruktúre budú údaje požadované | Detail údajov | Jedná sa o definovanie, o aký detail údajov sa jedná, či ide o individuálne alebo je možné mať údaje agregované |
| Spôsob získavania (integrácie) | Jedná sa o stavenie spôsobu integrácie s poskytovateľom údajov | Frekvencia[[6]](#footnote-7) | Choose an item. |
| Dáta pre vytvorenie modelu | Určia sa údaje, ktoré budú slúžiť pre vytvorenie a otestovanie modelu / algoritmu | Dáta pre prevádzku modelu | Určia sa údaje, ktoré budú potrebné pre prevádzku a aktualizáciu modelu / algoritmu |
| Požadované historické údaje[[7]](#footnote-8) |  | Aké obdobie | V prípade, ak je potrebné historické údaje, je potrebné stanoviť obdobia v rokoch od |
| Zabezpečenie ochrany osobných údajov | Potrebné popísať, ako bude zabezpečená v prípade potreby ochrana osobných údajov | Dokumentácia údajov | Popísať aká je potrebná dokumentácia pre zbierané údaje |

### Vyhodnotenie technickej a dátovej pripravenosti (príklad)

Centre for Data science & Public policy (The University of Chicago) vypracovalo maticu hodnotenia pripravenosti projektov z pohľadu potreby údajov a technickej pripravenosti pre projekt Preventative Lead Inspections (preventívne kontroly olova). Zadefinovali 10 aspektov, ktoré sú hodnotené z pohľadu úrovne ich vyspelosti ako:

* Zaostávajúce
* Základné
* Pokročilé
* Vedúce

V nasledovnej matici je detailný popis, ako sa dá pozerať na dátovú a technickú pripravenosť projektu.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Kategória** | **Oblasť** | | **Hodnotenie** | | | |
| **Zaostávajúce** | **Základné** | **Pokročilé** | **Vedúce** |
| Ako sú údaje ukladané | Prístupnosť | | Dostupné iba v rámci softvéru, kde sa zhromažďujú | Sú prístupné aj mimo SW ale len v proprietárnom formáte, ktorý si vyžaduje špeciálny analytický SW | Štandardné otvorené formáty ako CSV, JSON, XML, database | Štandardný otvorený formát dostupný cez API |
| Úložisko | | Papier | PDF alebo obrázky | Textové súbory | Databázy |
| Integrácie | | Údaje neopustia zdrojové systémy | Údaje sú exportované príležitostne a integrované ad hoc spôsobom | Centrálny dátový sklad - automatická agregácia a prepojenie interných dát v reálnom čase | Externé údaje sú tiež integrovateľné |
| Čo je zbierané | Relevantnosť a dostatočnosť | | Nezbierate krvné testy alebo výsledky domácej prehliadky | Zbierate nejaké krvné testy alebo výsledky domácich prehliadok | Máte údaje, ktoré pomáhajú vyriešiť problém ale nie úplne.  Pre projekt sa jedná o nasledovné údaje:  Individuálne údaje o hladine olova v krvi  Dátum testu  Dátum narodenia  Adresa trvalého pobytu  Kontrolované budovy  Sanačná zhoda  Dátum postavenia budovy  145/5000  Rodné listy, WIC alebo imunizačné záznamy, EMR a ostatné zdroje údajov o tom, kde deti žijú do troch mesiacov od narodenia. | Máte všetky údaje o všetkých deťoch vrátane:  Meno  Demografické informácie (pohlavie, rasa, ...)  Typ testu (žily / kapiláry)  Laboratórium, ktoré vykonalo testy  Nebezpečenstvo identifikované pre interiér  Nebezpečenstvo identifikované pre exteriér  História zmien budovy  Stav budovy (priznaný / nepriznaný) |
| Kvalita | | Chýbajú riadky (napr. chýbajúce deti, krvné testy, alebo domové prehliadky) | Chýbajú stĺpce (chýbajúce premenné ako napríklad chýbajúce demografické údaje dieťaťa) | Žiadne chýbajúce údaje, ale chyby v zbere údajov, ako sú preklepy | Žiadne chýbajúce údaje ani chyby v zozbieraných údajoch |
| Frekvencia zberu | | Raz | Ročne | Štvrťročne | Real time |
| Granularita | | Agregácia na úrovni mesta | Agregácia na úrovni štvrtí / blokov | Údaje na úrovni detí a domov | Údaje na úrovni konkrétnych testov a prehliadok |
| História | | Žiadne historické údaje (vymazané) | Historické údaje sú uložené ale prepísané existujúcimi údajemi | Údaje sú uložené minimálne 3 roky a pridávajú sa nové údaje s časovou pečiatkou, pričom historické údaje sa zachovávajú | Údaje sú uložené minimálne 5 rokov, pričom nové údaje sa namapujú na staré schémy, čo umožní využiť aj historické údaje |
| Ostatné | | Ochrana osobných údajov | Žiadne zásady ochrany osobných údajov nie sú aplikované | Osobné údaje nemôžu byť použité | Ad-hoc schvaľovací proces, ktorý umožní využitie niektorých osobných údajov v projekte | Softvérom definovaná a kontrolovaná ochrana osobných údajov, pričom je možné vykonávať analýzy pri dodržaní preddefinovaných pravidiel |
| Dokumentácia | Žiadna digitálna dokumentácia alebo údaje o metadátach: údaje existujú ale neexistuje popis premenných alebo popis polí | Existuje slovník údajov (sú definované premenné a kategórie) | Existuje slovník údajov a úplné metadáta (vrátane podmienok, za ktorých boli údaje zaznamenané) | Existuje slovník údajov a úplné metadáta vrátane podmienok zberu a informácií o tom čo nebolo zbierané a aké sú potenciálne obmedzenia |

## Potrebná funkcionalita



Potrebná funkcionalita predstavuje definovanie nástrojov, ktoré bude možné využiť na samotné spracovanie analýzy na základe vybraného analytického prístupu. Jedná sa o nástroje v nasledovných procesných krokoch / oblastiach:

* Výber zdrojov údajov
* Proces získavania údajov
* Ukladanie údajov
* Spracovanie údajov
* Analyzovanie údajov
* Príprava reportov a výstupov

Potrebné funkcionality je možné definovať prostredníctvom ***nasledujúcej tabuľky, ktorá sa nachádzať v časti štúdie – Architektúra informačných systémov.*** Jedná a o spôsob výberu jednotlivých komponentov, funkcionalít a služieb, ktoré je potom potrebné popísať v kontexte daného projektu / prípadu použitia.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Zdroje údajov | | | | Popis systému | | | Integrácia[[8]](#footnote-9) | | Spracovanie | |
| Centrálny IS VS | |  | | Stručný popis systému | | | Choose an item. | | Choose an item. | |
| IS VS | |  | | Stručný popis systému | | | Choose an item. | | Choose an item. | |
| Externé dátové zdroje | |  | | Stručný popis systému | | | Choose an item. | | Choose an item. | |
| Prioritizácia a kategorizácia | | | | | | | | | | |
| Služby prioritizácie |  | | Popis riešenia | | | Služby kategorizácie | |  | | Popis riešenia |
| Ukladanie údajov | | | | | | | | | | |
| Centrálny dátový sklad | | |  | | Popis riešenia | | | | | |
| Data lake | | |  | | Popis riešenia | | | | | |
| Data mart | | |  | | Popis riešenia | | | | | |
| Analytický sandbox | | |  | | Popis riešenia | | | | | |
| Iné | | |  | | Popis riešenia | | | | | |
| Požiadavky na spracovanie údajov | | | | | | | | | | |
| Čistenie údajov |  | | Popis riešenia | | | Procesy dátovej kvality | |  | | Popis riešenia |
| Transformácia údajov |  | | Popis riešenia | | | Optimalizácia údajov | |  | | Popis riešenia |
| Anonymizácia údajov |  | | Popis riešenia | | | Virtualizácia údajov | |  | | Popis riešenia |
| Obohacovanie údajov |  | | Popis riešenia | | | Orchestrácia údajov | |  | | Popis riešenia |
| Replikácia údajov |  | | Popis riešenia | | | Aktívny archív | |  | | Popis riešenia |
| Spracovanie v pamäti |  | | Popis riešenia | | | Iné | |  | | Popis riešenia |
| Požiadavky na analytické nástroje | | | | | | | | | | |
| Dátová agregácia |  | | Popis riešenia | | | Analýza sentimentu | |  | | Popis riešenia |
| Data query |  | | Popis riešenia | | | Regresná analýza | |  | | Popis riešenia |
| Štatistická analýza |  | | Popis riešenia | | | Štatistické modely | |  | | Popis riešenia |
| KPIs |  | | Popis riešenia | | | Simulácie | |  | | Popis riešenia |
| Data minig |  | | Popis riešenia | | | Machine learning | |  | | Popis riešenia |
| Vizualizácia |  | | Popis riešenia | | | Prediktívne modelovanie | |  | | Popis riešenia |
| Sémantická analýza |  | | Popis riešenia | | | Optimalizačné modely | |  | | Popis riešenia |
| Umelá inteligencia |  | | Popis riešenia | | | Neurónové siete | |  | | Popis riešenia |
| Heuristické metódy |  | | Popis riešenia | | | Iné | |  | | Popis riešenia |
| Požiadavky na reporty a produkty | | | | | | | | | | |
| Štandardné výkazy |  | | Popis riešenia | | | Dashboard | |  | | Popis riešenia |
| Kľúčové ukazovatele výkonnosti |  | | Popis riešenia | | | Nástroje pre mobilné zariadenia | |  | | Popis riešenia |
| Pravidelné výkazy |  | | Popis riešenia | | | Inteligentní agenti | |  | | Popis riešenia |
| OLAP reporty |  | | Popis riešenia | | | Spúšťače v reálnom čase | |  | | Popis riešenia |
| Užívateľské výkazy |  | | Popis riešenia | | | Iné | |  | | Popis riešenia |

## Použitie v praxi



Od realizácie dátového projektu sa očakáva, že zlepší fungovanie inštitúcie a zabezpečí, že rozhodovanie sa bude vykonávať kvalitnejšie. V rámci ďalšieho kroku je potrebné podrobne popísať dopad na jednotlivé úrovne rozhodovania a definovať, ako bude vyzerať budúci stav.

***Túto časť je potrebné popísať v časti Motivácia v štúdii***

### Používatelia riešenia

Úspech riešenia závisí od toho, kto sa k reálnemu analytickému produktu dostane a ako často. Preto je dôležité identifikovať kľúčových a potencionálnych používateľov riešenia a stanoviť frekvenciu rozhodovania, na základe účelu použitia predikcií.

V rámci štúdie sú preddefinovaní nasledovní používatelia, z ktorých je možné vybrať alebo v prípade potreby doplniť ďalších:

* Politické vedenie / Riadiaci pracovníci / Komisie
* Individuálni pracovníci organizácie
* Verejnosť
* Akademický sektor
* Inštitúcie verejnej správy / regulátori

Pre každú vybranú skupinu je potrebné stanoviť nasledovné:

* Počet - Jedná sa o stanovenie počtu používateľov, ktorý budú využívať výsledky projektu.
* Frekvencia - Predstavuje požadovanú frekvenciu výstupov tak, aby bolo možné rozhodnutia alebo procesy realizovať optimálne.
* Účel - V tejto časti sa definuje, aký je účel použitia výsledkov projektu resp. prípadu použitia.

### Dopad na rozhodovacie procesy a záväznosť rozhodnutí

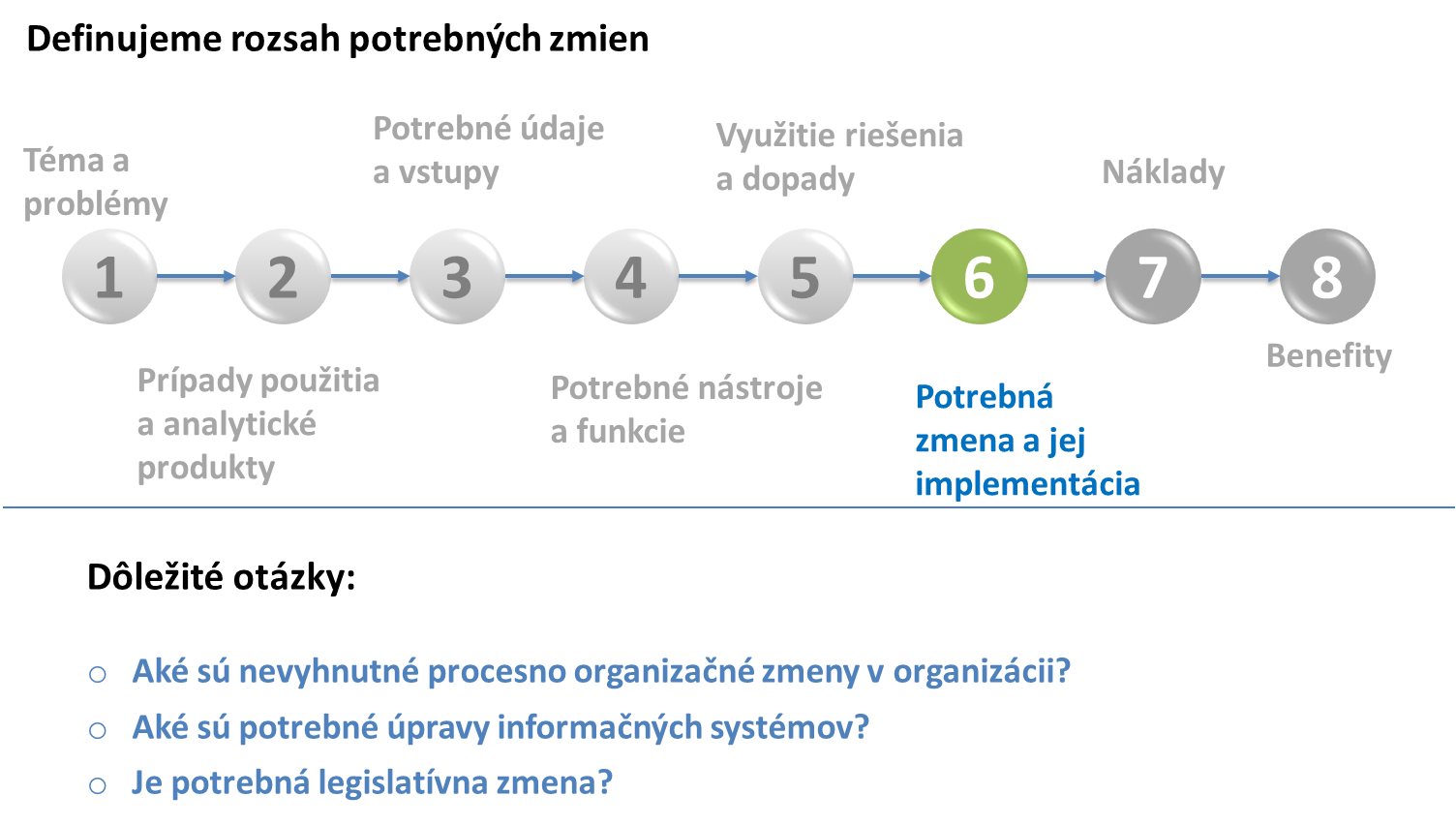
Prostredníctvom nasledujúcej tabuľky je potrebné určiť, aké procesy budú výsledkami projektu dotknuté alebo môžu byť dotknuté, pričom je potrebné definovať:

* Zmenu – čo nastane v prípade daného procesu
* Objem – aký je približný objem financií, ktorý sa daným procesom realizuje alebo je ovplyvnený (interné zdroje, zdroje o ktorých sa rozhoduje a pod.)
* Potenciál úspor - aký je potenciál úspor z pohľadu aplikácie riešenia na daný proces

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rozhodovací proces | | Zmena | Objem (EUR / ročne) | Potenciál optimalizácie |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

.

## Potrebné zmeny



V rámci nasledujúcej časti sa definuje zoznam opatrení, ktoré je potrebné realizovať, aby bolo možné výsledky dátového projektu využívať v praxi. Zmeny sú rozdelené do troch kategórií:

* Organizačné úpravy
* Procesné úpravy
* Úpravy legislatívy

### Organizačné úpravy

V nasledujúcej tabuľke sú vybrané organizačné opatrenia, ktoré sú realizáciou projektu dotknuté, pričom sú definované konkrétne zmeny, ktoré budú pre potreby projektu alebo na základe výsledkov projektu zrealizované. V rámci štúdie v časti Implementácia a migrácia je možné vybrať z preddefinovaných zmien alebo doplniť vlastné:

* Vytvorenie novej organizácie – právnej entity
* Nová organizačná jednotka, zabezpečujúca podporu rozhodovania a spracovanie dát
* Nová organizačná jednotka zabezpečujúca zber a kvalitu dát
* Rozšírenie súčasných kapacít pre potreby zberu dát
* Rozšírenie súčasných kapacít pre potreby podpory rozhodovania
* Školenia a zvýšenie kvalifikácie personálu
* Zníženie počtu pracovníkov v súčasných organizačných útvaroch
* Zrušenie súčasnej organizačnej jednotky
* Zrušenie súčasnej organizácie – právnej entity

Ku každej vybranej zmene je potrebné definovať konkrétny popis danej zmeny, ktorá bude výsledkom projektu dosiahnutá alebo požadovaná.

### Úpravy procesov

Rovnako v prípade úpravy procesov sú preddefinované procesné úpravy, ktoré je možné samozrejme doplniť, pričom ku každej procesnej zmene je potrebné doplniť konkrétny popis zmeny. ***Úpravy procesov sa nachádzajú v časti štúdie Implementácia a migrácia:***

* Úprava riešenia vzťahov s klientami
* Zavedenie experimentovania do fungovania organizácie
* Nastavenie „dodávateľského reťazca“ pre spracovanie dát, manažment kvality údajov
* Úprava schvaľovacích procesov pri návrhu politík, regulácii alebo investícií
* Zavedenie publikovania analytických výstupov
* Automatizácia rozhodovacích činností
* Zavedenie systematickej podpory rozhodovania v expertnej činnosti

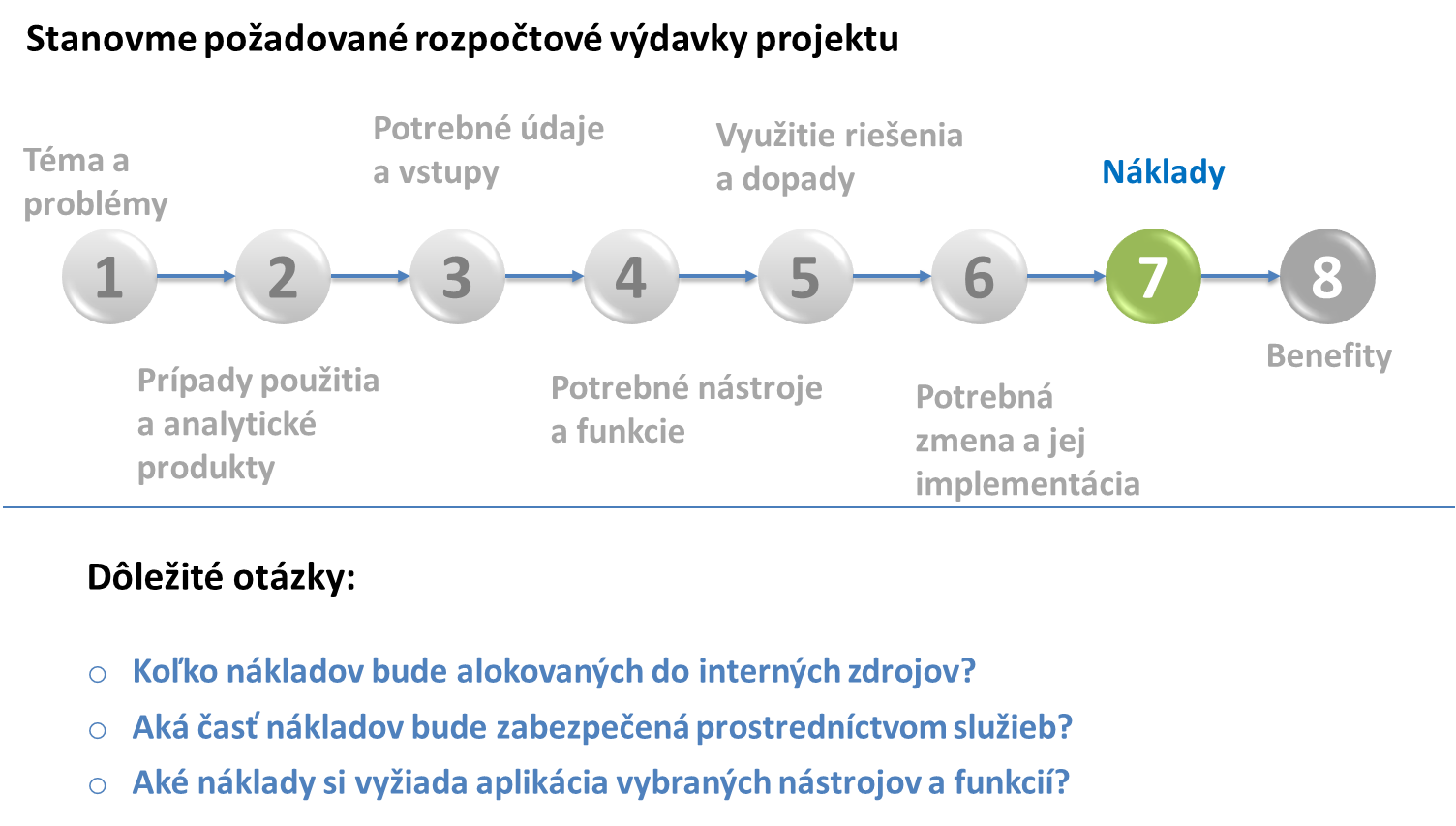
### Úpravy legislatívy

Táto časť sa nachádza v časti štúdie Popis budúceho stavu - > Legislatíva, pričom sú definované základné regulačné opatrenia, ktoré je možné realizovať. K týmto je v prípade výberu potrebné doplniť konkrétny popis zmeny. Jedná sa o nasledovné:

* Koncepčné zmeny regulačného rámca
* Nové povinnosti pre podnikateľské subjekty v súvislosti so zabezpečením zdrojov dát
* Nové povinnosti pre inštitúciu verejnej správy
* Nové povinnosti pre ostatné inštitúciu verejnej správy
* Zrušenie povinností pre podnikateľské subjekty
* Zrušenie povinností pre inštitúcie verejnej správy

V prípade potreby je možné doplniť ďalšie potrebné legislatívne oblasti resp. úpravy.

## Náklady riešenia



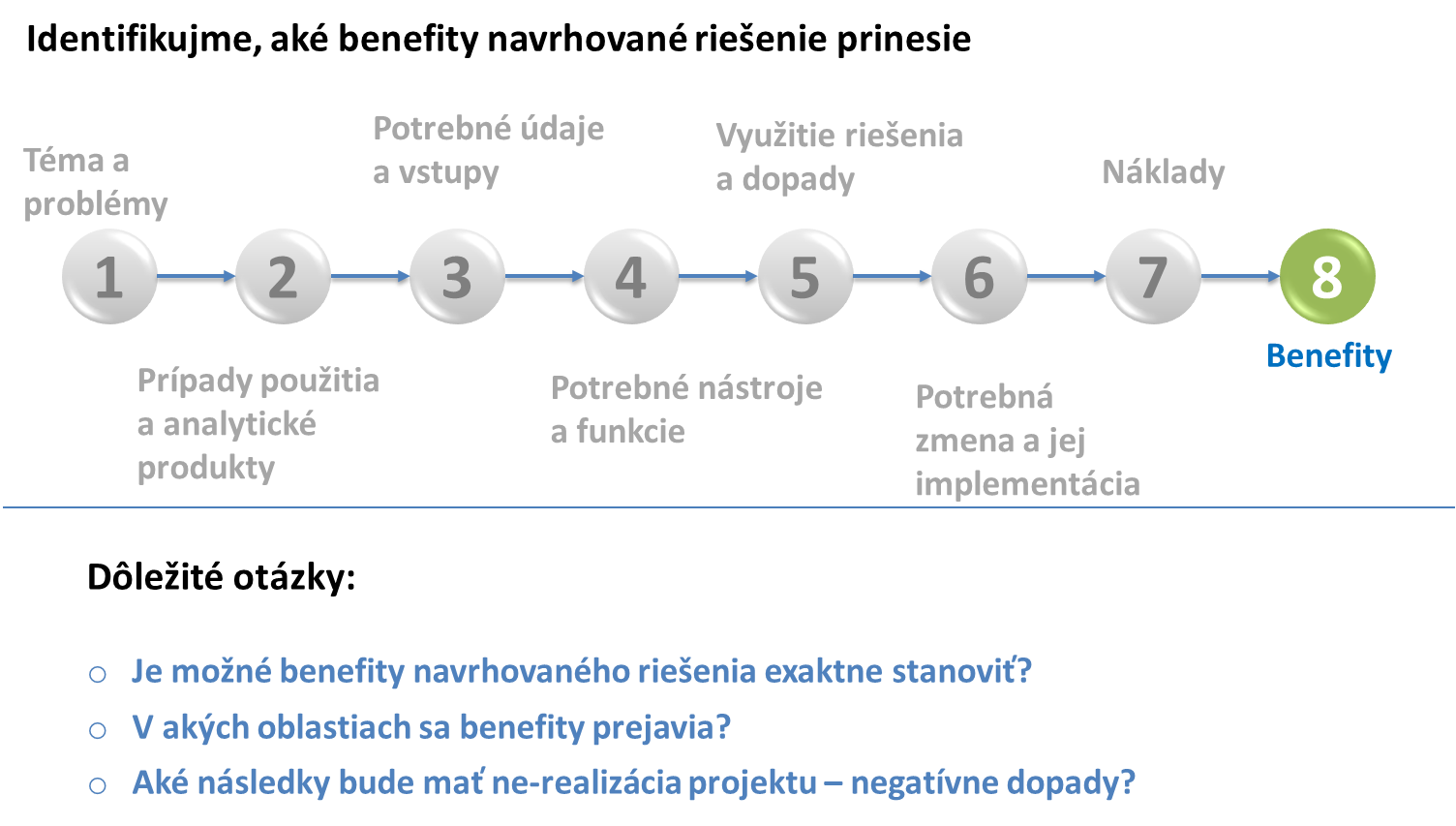
Náklady na realizáciu projektu vychádzajú z jeho rozsahu, požadovaných funkcionalít a aktivít, ktoré budú v rámci projektu realizované. Ako pomôcka bol vytvorený excel, v ktorom je možné definovať náklady tak, aby boli jednak v súlade s oprávnenými aktivitami dopytového projektu, ale aj oprávnenými aktivitami v zmysle príručky OP II.

V danom excely je potrebné:

* vybrať v záložke Limity projektu typ projektu, aby sa aktivovali limity pre daný typ projektu
* vyplniť najmä záložku Vecný rozpočet, kde sú predpripravené aktivity dopytovej výzvy na aktivita podľa príručky a zároveň sú preddefinované aj typy výdavkov, ktoré sú pre danú aktivitu predpokladateľné. Samozrejme pre jednotlivé aktivity je možné doplniť aj iné typy výdavkov tak, aby boli v súlade s oprávnenými výdavkami OP II.
* Následne je potrebné do štúdie do časti Rozpočet projektu vložiť nasledujúcu tabuľku (tiež ako výstup excelu), v ktorej musia byť dodržané limity pre jednotlivé aktivity v zmysle podmienok uvedených vo výzve.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Typ projektu | | Finančný limit | Rozpočet projektu | Splnený linit? | |
|  | Prediktívne kontroly | 3 000 000 € | 236 500 € | ANO | |
| ID | Výsledok | Finančný limit (% z celkovej výšky <=) | Náklad v rozpočte | Reálny podiel v % | Podmienka splnená? |
| A1 | Analýza prípadov použitia | 5% | 236 500 € | 100% | NIE |
| A2 | Zabezpečenie zdrojov dát | 30% | 0 € | 0% | ANO |
| A3 | Nastavenie funkcionalít | 10% | 0 € | 0% | ANO |
| A4 | Realizácia dátového modelu | 50% | 0 € | 0% | ANO |
| A5 | Zavedenie zmien do praxe | 20% | 0 € | 0% | ANO |
| A6 | Publikovanie výstupov, analytických produktov a otvorených údajov | 20% | 0 € | 0% | ANO |

## Prínosy navrhovaného riešenia



V tejto časti sú popísané benefity ako aj riziká, ktoré vyplávajú z nerealizácie projektu. Prínosy sú definované do 4 základných kategórií a to:

* Otvorenosť a transparentnosť
* Posilnenie dátovej ekonomiky
* Zvýšenie spotrebiteľskej a spoločenskej hodnoty
* Využitie dátovej vedy na optimalizáciu

V prípade potreby je možné pre jednotlivé oblasti benefity dopĺňať.

Ak je definovaný daný typ prínosu, je tento posúdený z pohľadu nasledovných atribútov:

* Zdôvodnenie – jedná sa o popis zdôvodnenia, prečo dané riešenie prispeje práve k naplneniu daného typu prínosu
* Ukazovateľ – ide o stanovenie ukazovateľa, ktorým je možné daný atribút merať. Výber ukazovateľa je stanovaný na základe vedomostí, prípadne praxe alebo benchmarku tak, aby bol relevantný k danému typu prínosu.
* Súčasná hodnota – jedná sa o definovanie súčasnej hodnoty ukazovateľa. Hodnota a by mala byť overiteľná z existujúcich vlastných zdrojov alebo zdrojov inštitúcií, ktoré sú relevantné definované ukazovatele hodnotiť
* Cieľ a spôsob merania – v rámci tejto časti je potrebné definovať cieľovú hodnotu ukazovateľa a navrhnúť spôsob merania dopadov realizovaných riešení

***Táto časť sa nadchádza v štúdii v časti Analýza benefitov***

### Otvorenosť a transparentnosť

|  |  |
| --- | --- |
| Aspekt | Typ prínosu |
| Transparentnosť | Zníženie podvodov |
| Proaktívne služby klientom |
| Zníženie plytvania |
| Dôvera v štát | Dôveryhodnosť údajov |
| Legislatíva | Kvalitnejšie posúdenie vplyvov |
| Zdieľanie údajov | Nové znalosti a príležitosti |

### Posilnenie dátovej ekonomika

|  |  |
| --- | --- |
| Aspekt | Typ prínosu |
| Veľkosť trhu (Data market) | Zvyšovanie pridanej hodnoty |
| Tržby | Príjmy z daní |
| Pracovníci v oblasti dát (data workers) | Zvyšovanie zamestnanosti |
| Počet firiem | Zvyšovanie HDP |
| Hodnota dátovej ekonomiky | Pridaná hodnota |
| Globálna konkurencia | Posilnenie postavenia v rebríčkoch |

### Zvýšenie spotrebiteľskej a spoločenskej hodnoty

|  |  |
| --- | --- |
| Aspekt | Typ prínosu |
| Kvalitnejšie služby | Spotrebiteľský prebytok |
| Zvýšenie kvality rozhodovania | Zníženie počtu exekúcií |
| Spotrebiteľské riziko | Zníženie rizík |
| Používanie údajov | Eliminácia zlých rozhodnutí |

### Využitie dátovej vedy na optimalizáciu

|  |  |
| --- | --- |
| Aspekt | Typ prínosu |
| Zlepšenie rozhodovania | Zníženie počtu FTE |
| Efektivita procesov | Zníženie počtu FTE |
| Kvalita kontroly | Lepší dozor a dohľad nad regulovaním prostredím |
| Riziko rozhodovania | Znižovanie rizík na základe dát |
| Prediktívne modely | Zníženie časovej náročnosti |
| Eliminácia chybných rozhodnutí |
| Vyplácanie financií |
| Plánovanie zdrojov | Efektívne využitie nákladov |

# Služby konsolidovanej analytickej vrstvy

| **ID** | **Systém** | **Služba** | **Popis** | **Dostupnosť** |
| --- | --- | --- | --- | --- |
| 1 | Kolaboračná platforma | Úložisko dát | Kde je možné ukladať vstupné dátové zdroje, výsledky analýz a analytické modely (takzvaný „analytický dropbox“). | Január 2020 |
| 2 | Kolaboračná platforma | Zdieľanie | Služby zdieľania, na základe ktorých je možné zdieľať vstupné dátové zdroje, výsledky analýz alebo analytické modely nad definovanou množinou dát z Centrálneho dátového skladu. Základným pravidlom tvorby modelov je, že ich je vždy možné prepočítať na základe aktuálnych dát (to znamená, že predpovedané dáta sa zmenia na aktuálne, a ich porovnaním možno ďalej vylepšovať a kalibrovať modely). | Január 2020 |
| 3 | Kolaboračná platforma | Kolaborácia a spolupráca | Služby kolaborácie, pomocou ktorých analytici spolupracovať nielen interne na tvorbe analytického modelu, aj s externými špecialistami. | Január 2020 |
| 4 | Centrálny dátový sklad | Integrácia dát |  | September 2020 |
| 5 | Centrálny dátový sklad | RDBMS |  | September 2020 |
| 6 | Centrálny dátový sklad | Data Warehouse pre ukladanie mikroúdajov a mikrofaktov, s podporou OLAP |  | September 2020 |
| 7 | Centrálny dátový sklad | ODS | ODS – „Operational data store“, | September 2020 |
| 8 | Centrálny dátový sklad | MPP | Massively parallel processing database | September 2020 |
| 9 | Centrálny dátový sklad | Hadoop | Open source framework navrhnutý pre spracovanie big data na základe modelu MapReduce, distribuovaný systém | September 2020 |
| 10 | Centrálny dátový sklad | Storm | Open source framework pre spracovanie tokov dynamických dát | September 2020 |
| 11 | Centrálny dátový sklad | Spark |  | September 2020 |
| 12 | Centrálny dátový sklad | Analytické spracovanie dát priamo v pamäti | „in-memory analytics“ | September 2020 |
| 13 | Centrálny dátový sklad | Virtuálne DB |  | September 2020 |
| 14 | Centrálny dátový sklad | Ochrana dát | Riešenie zabezpečuje ochranu údajov (vďaka anonymizácii a pseudonymizácii, kryptografii a úrovniam prístupu s rôznou granularitou), správu dát (transakčné logy, audity na najnižšej úrovni granularity, data provenance) a ich integritu vďaka monitoringu v reálnom čase, validácii koncového bodu a filtrovaniu | Január 2021 |
| 15 | Analytické nástroje | Tradičné BI | podporujú analytické funkcie nad klasickými štruktúrovanými dátovými zdrojmi (faktov), OLAP, podpora dátových kociek – „data cubes“, prieskum dát a reporty, | Jún 2020 |
| 16 | Analytické nástroje | „machine learning“ | množina nástrojov a knižnice ready to run modelov, ktoré podporujú spracovanie údajov metódami umelej inteligencie – strojového učenia | Jún 2020 |
| 17 | Analytické nástroje | štatistické nástroje | umožňujú tvorbu štatistických dátových modelov, testovanie hypotéz, faktorové analýzy, korelácie, regresie a podobne | Jún 2020 |
| 18 | Analytické nástroje | nástroje pre simulácie | umožňujú simulovať historický aj budúci priebeh modelovaných udalostí nad vybranou množinou historických a aktuálnych údajov. Simulačné scenáre môžu byť založené na metodikách, ako je prognóza trendov, modelovanie založené na agentoch, analýza systémov, „croud-sourcing“ a „red teaming“. | Jún 2020 |
| 19 | Analytické nástroje | vizualizácia dát a publikácia | umožňujú pre zamestnancov alebo pre verejnosť vidieť výsledky analýz v interaktívnych grafoch vo webovom priestore | Jún 2020 |
| 20 | Analytické nástroje | publikácia otvorených údajov | zdrojové údaje a výsledky vo formáte otvorených dát, dostupné cez OpenAPI, registrácia v centrálnom katalógu otvorených údajov | Jún 2020 |

1. <https://www.bi.team/publications/using-data-science-in-policy/> [↑](#footnote-ref-2)
2. Zoznam najpoužívanejších metód sa nachádza v časti 5.3.4Výber štandardizovaných analytických metód [↑](#footnote-ref-3)
3. Výberové pole [↑](#footnote-ref-4)
4. https://managementmania.com/sk/analyzy-analyticke-techniky [↑](#footnote-ref-5)
5. Zaškrtne sa v prípade, ak je potrebné údaje nakupovať [↑](#footnote-ref-6)
6. Výberové pole [↑](#footnote-ref-7)
7. Zaškrtne sa v prípade, ak sú požadované historické údaje [↑](#footnote-ref-8)
8. Typom integrácie je myslené nasledovné:

   * **Dávkové spracovanie** – asynchrónna výmena veľkého množstva údajov, platných k určitému času. Len v nevyhnutných prípadoch (pre veľmi veľké objemy údajov z dôvodu časovej efektívnosti)
   * **Asynchrónna komunikácia** – preferované z pohľadu rýchlej optimalizácie procesov a ďalšieho nastavenia riešenia
   * **Real-time (Synchrónna komunikácia)** – ideálne z pohľadu konzumenta – dá sa očakávať, že bude využitá po nasadení interaktívneho riešenia životných situácií
   * **Virtualizácia dát** – pre komplexnú integráciu údajov z viacerých zdrojov a typov

   [↑](#footnote-ref-9)