**Katalóg požiadaviek**

**World Value Survey (DEKK)**

Získanie údajov, vizualizácia, výpočet korelácie,

aktualizácia dát, štatistické údaje

Zuzana Mačicová, Vladimír Ačjak, Marek Lichvár, Richard Mištík

**Obsah**

[1. Úvod 3](#_Toc87889546)

[1.1 Účel katalógu požiadaviek 3](#_Toc87889547)

[1.2 Rozsah využitia systému 3](#_Toc87889548)

[1.3 Slovník pojmov 3](#_Toc87889549)

[1.4 Odkazy a referencie 4](#_Toc87889550)

[1.5 Prehľad nasledujúcich kapitol 4](#_Toc87889551)

[2. Všeobecný popis 4](#_Toc87889552)

[2.1 Perspektíva projektu 4](#_Toc87889553)

[2.2 Funkcie systému 4](#_Toc87889554)

[2.3 Charakteristika používateľov 5](#_Toc87889555)

[2.4 Predpoklady, závislosti a obmedzenia 5](#_Toc87889556)

[3. Špecifikácia požiadaviek 5](#_Toc87889557)

[3.1 Funkčné požiadavky 5](#_Toc87889558)

[**3.1.1 Výber a prehľad datasetov** 5](#_Toc87889559)

[**3.1.3 Zobrazenie grafov** 6](#_Toc87889560)

[**3.1.4 Import údajov** 6](#_Toc87889561)

[3.2 Kvalitatívne požiadavky 7](#_Toc87889562)

[**3.2.1 Dlhodobosť** 7](#_Toc87889563)

[**3.2.2 Jednoduchosť** 7](#_Toc87889564)

[**3.2.3 Rýchlosť** 7](#_Toc87889565)

[**3.2.4 Modulárnosť** 7](#_Toc87889566)

[**3.2.5 Stabilita a spoľahlivosť** 7](#_Toc87889567)

[3.3 Požiadavky grafického rozhrania a mapy 8](#_Toc87889568)

[3.4 Požiadavky používateľského rozhrania 8](#_Toc87889569)

[3.6 Požiadavky databázového managementu 8](#_Toc87889570)

[4. Prílohy 8](#_Toc87889571)

[4.1 Zoznam používaných datasetov 9](#_Toc87889572)

# **1. Úvod**

## **1.1 Účel katalógu požiadaviek**

Dokument slúži na špecifikáciu všetkých požiadaviek k informačnému systému vytváraného pre World Value Survey (DEKK), ktorý vzniká v rámci predmetu Tvorba Informačných Systémov (1-AIN-131) na FMFI UK. Katalóg požiadaviek je určený pre všetkých stakeholderov a taktiež je záväzným pre zadávateľov projektu a jeho vývojárov. Katalóg slúži ako východiskový materiál pre následnú implementáciu projektu. Je spísaný zrozumiteľne, jednoduchým jazykom pre obe strany, tak aby došlo k minimalizácii nedorozumení medzi zadávateľmi a vývojármi.

## **1.2 Rozsah využitia systému**

Jedná sa o projekt v oblasti geo data science s reálnym využitím v rámci výskumu sociálnej kohézie. Projekt sa snaží prispieť k riešeniu problému nevhodnej sociálnej kohézie medzi ľuďmi, ktorý sa postupne zhoršuje a ktorý je pomerne ťažko analyzovateľný. Sociálna kohézia súvisí s kriminalitou, s platmi a s veľmi veľa inými faktormi. Hlavným cieľom je začať na Slovensku, no projekt by mal byť rozšíriteľný aj do sveta.

## **1.3 Slovník pojmov**

* **World Value Survey** - globálny výskumný projekt, ktorý skúma hodnoty a viery ľudí, ako sa v priebehu času menia a aký majú sociálny a politický dopad.
* **DEKK Inštitút** - reprezentant World Value Survey na Slovensku
* **Stakeholder** - zainteresovaná strana/“zúčastnený“
* **Data science**  - interdisciplinárny obor, ktorý využíva vedecké metódy, procesy, algoritmy a systémy pre získavanie znalostí a poznatkov z dát v rôznych podobách, ako štruktúrovaných, tak neštruktúrovaných
* **Dataset** - jeden číselný dátový atribút podľa okresov za jeden konkrétny rok, môže však byť parametrizovateľný niekoľkými atribútmi, napr. počet všetkých obyvateľov je základný atribút a je parametrizovateľný vekom a pohlavím (napr. vek=10, pohlavie=M)
* **Databanka** - úložisko informácií o jednom alebo viacerých subjektoch - databáze - ktorá je organizovaná spôsobom, ktorý uľahčuje miestne alebo vzdialené získavanie informácií a je schopný dlhodobo spracovávať mnoho neustálych dotazov
* **Geo data science** - disciplína data science zameraná špecificky na jej priestorový komponent
* **Sociálna kohézia** - charakteristika sociálnych pút a vzájomných väzieb členov spoločenskej skupiny alebo organizácie, daná prevahou spoločných záujmov/súdržnosť
* **Policy-maker** - skupina osôb/osoba zodpovedná za vytváranie a opravu nápadov/plánov/politiky
* **Slider** - posuvník/lišta/grafický ovládací prvok

## **1.4 Odkazy a referencie**

* Odkaz na github repozitár:

<https://github.com/TIS2021-FMFI/dekk>

* DEKK Inštitút: <https://dekk.sk/sk/>
* World Value Survey: <https://www.worldvaluessurvey.org/wvs.jsp>
* DATACUBE <http://datacube.statistics.sk>

## **1.5 Prehľad nasledujúcich kapitol**

V nasledujúcej kapitole je všeobecne popísaný projekt / informačný systém - jeho perspektíva, funkcia, funkcionalita, charakteristika, predpoklady, závislosti a obmedzenia používania a v tretej kapitole je podrobná špecifikácia kvalitatívnych, funkčných požiadaviek. Taktiež pojednáva o grafickom rozhraní. Štvrtá kapitola popisuje aké dokumenty budú poskytnuté k aplikácii a obsahuje taktiež zoznam odkazov na použité datasety.

# **2. Všeobecný popis**

## **2.1 Perspektíva projektu**

Projekt má ambíciu byť pomocným nástrojom na analýzu zhoršujúcej sa sociálnej kohézie na Slovensku. Má slúžiť ľuďom, ktorí majú moc rozhodovať, teda politikom, novinárom, policymaker-om, ktorým pomôže pomocou vizualizácie rôznych datasetov analyzovať príčiny zlej sociálnej kohézie pre možné následné hľadanie riešenia tohto problému. Všetky datasety, s ktorými aplikácia pracuje, obsahujú jeden údaj za každý okres a určitý rok. Aplikácia sa bude vedieť vysporiadať aj s chýbajúcimi údajmi (napr. za niektorý okres v niektorom roku).

## **2.2 Funkcie systému**

Výsledným produktom je webová aplikácia, prístupná cez web DEKK Inštitútu, ktorá umožňuje svojmu užívateľovi prehľadne graficky vizualizovať korelácie rôznych datasetov na mape Slovenska rozdelenej na okresy. Datasety bude získavať hlavne z verejných Slovenských databánk.

K funkciám aplikácie patrí aj filtrovanie v ponuke datasetov, vizualizácia zvolených datasetov na mapke Slovenska a ich prekrývania v oblastiach mapky, automatická aktualizácia dát každý rok a vykreslenie grafu na bočnom paneli pre dané filtre.

## **2.3 Charakteristika používateľov**

*Bežný používateľ* - môže si zobrazovať mapku, filtrovať, prehliadať všetky dáta na všetkých paneloch. Príkladom bežného používateľa sú zamestnanci, analytici štátnej správy, akademici, študenti a časom aj novinári a široká verejnosť. Tí sa na našu databázu a jej vizuálny výstup môžu obrátiť pri ich práci.

*Administrátor* - má navyše možnosť na strane servera manuálne nasadiť dáta do aplikácie.

## **2.4 Predpoklady, závislosti a obmedzenia**

## 

Medzi obmedzenia používania aplikácie patrí: prístup na internet, funkčnosť stránky DEKK Inštitútu, cez ktorú k nej bude umožnený prístup a funkčnosť a prístup k dátovému centru DATACUBE, taktiež predpokladá, že dáta budú do aplikácie poskytované v .csv formáte.

# **3. Špecifikácia požiadaviek**

Požiadavky označené \* sú požiadavky s nižšou prioritou a budú implementované, iba ak vývojový tím bude mať kapacitu.

## **3.1 Funkčné požiadavky**

### **3.1.1 Výber a prehľad datasetov**

1. V aplikácii sa bude zobrazovať zoznam prístupných datasetov, každý je určený svojím názvom
2. Používateľ môže zoznam datasetov filtrovať – zúžiť zoznam zobrazených na také, ktorých názov obsahuje nejaký (užívateľom zadaný) textový podreťazec
3. V zozname datasetov môže vybrať po jednom (aj pri rôznom zobrazení filtra) dva konkrétne datasety na zobrazenie korelácie, ktoré budú pre užívateľa zaujímavé
4. \* Ak dataset je parametrizovateľný, po jeho zvolení sa zobrazí pre každý parameter: názov parametra a vedľa neho drop-box v ktorom sa dá označiť jeden alebo viac riadkov. Takéto upresnenie datasetu sa použije pre zobrazenie a výpočet korelácie. (napríklad rozdelenie na mužov a ženy)

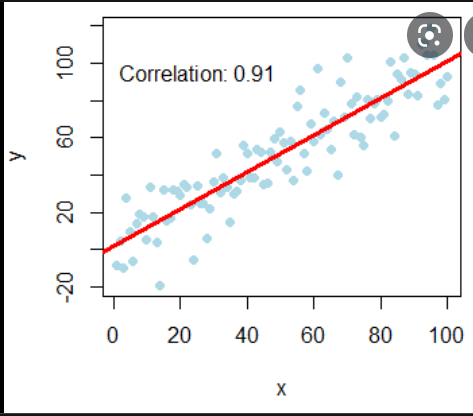
**3.1.2 Zobrazenie mapy**

1. Vedľa zoznamu sa zobrazuje mapa Slovenska s vyznačenými hranicami okresov, pokiaľ nie sú prístupné/označené žiadne datasety tak mapka bude biela.
2. Po výbere dvoch datasetov používateľom sa mapa Slovenska zafarbí podľa daných hodnôt (pre každý dataset jednou farbou a jej odtieňmi).
3. Pri mape bude zobrazená legenda, ktorá bude znázorňovať príslušnosť farieb k datasetom.
4. Pod mapou bude zobrazený slider, ktorý umožní zobraziť dáta z vybraného roku.
5. Pri mape je tlačidlo umožňujúce stiahnuť obrázok aktuálne zobrazenej mapy a korelácie.

### **3.1.3 Zobrazenie grafov**

1. Aplikácia bude zobrazovať graf, tvorený z dvoch datasetov (na x-ovej osi prvý, na y-ovej osi druhý dataset)
2. Predbežná podoba grafu:  
   Každý zobrazený bod grafu zodpovedá jednému okresu.

Súčasťou zobrazenia bude celkový korelačný koeficient pre celé dáta vybranej dvojice datasetov.



### **3.1.4 Import údajov**

1. Aplikácia bude poskytovať rozhranie pre administrátora (napr. samostatnú aplikáciu spúšťanú z príkazového riadka), ktorým je možné pridať nový dataset do databázy aplikácie. Používateľ zadá meno .csv súboru, názov datasetu a rok. Ak nezadá rok, predpokladá sa aktuálny rok.
2. Aplikácia bude poskytovať rozhranie pre administrátora (napr. samostatnú aplikáciu spúšťanú z príkazového riadka), ktorým je možné pridať nový dataset do databázy aplikácie automaticky downloadnutý zo servera Štatistického úradu SR. Používateľ zadá kód datasetu a rok. Ak nezadá rok, predpokladá sa predchádzajúci rok (napr. ak sa program pustí 1. januára 2022, tak sa downloadnu údaje z roku 2021). Typické použitie tohto programu je z automatických skriptov, ktoré ho pustia napríklad jeden raz ročne pre všetky požadované datasety.
3. \* Import popísaný 3.1.4 b. importuje kompletný dataset pre všetky hodnoty voliteľných parametrov.

## **Kvalitatívne požiadavky**

### **3.2.1 Dlhodobosť**

Chceme používať perspektívne platformy a nástroje tak, aby na nich bolo možné dlhodobo budovať. Nechceme meniť celú architektúru každých pár rokov.

### **3.2.2 Jednoduchosť**

Programátorské praktiky, ktoré budú udržateľné a jednoducho pochopiteľné pre všetkých (študentské projekty stavajúce na našom systéme, ďalší administrátor, nový developer).

### **3.2.3 Rýchlosť**

Je dôležité, aby sa mapky pre užívateľa nenačítali dlho a aby bola stránka rýchlo dostupná.

### **3.2.4 Modulárnosť**

Malo by byť možné jednoducho pridávať ďalšie nástroje. Poprípade staré vymeniť za nové bez zbytočne komplexných závislostí medzi modulmi. To platí zvlášť pre možnosť pridávať filtre.

### **3.2.5 Stabilita a spoľahlivosť**

Systém by nemal padať pri načítaní mapových filtrov a datasetov.

## **3.3 Požiadavky grafického rozhrania a mapy**

Aplikácia je prehľadná, grafy, štatistické údaje a mapa sú zrozumiteľné a rozdelené do jednotlivých panelov, korelácie na nej sú zobrazované a vysvetlené jednoznačne a jednoducho. Zvolené/aplikované datasety sú farebne rozlíšené od tých nezvolených. Grafy s vyobrazenými vrstvami dát korešpondujú.

## **3.4 Požiadavky používateľského rozhrania**

Používateľské rozhranie je intuitívne, prehľadné, dizajnovo ladiace so stránkou inštitútu a užívateľovi sú jednoducho a rýchlo prístupné informácie o daných vizuálnych častiach aplikácie a ich fungovaní, taktiež je užívateľovi prístupný stručný návod na používanie priamo v aplikácii.

## **3.6 Požiadavky databázového managementu**

Aplikácia má byť schopná uchovávať a pridávať veľké množstvo dát z viacerých rôznych databáz/datasetov. Navyše by mala periodicky-ročne preberať dáta zo zdrojov, ktoré sú pravidelne aktualizované. Tie budú pochádzať z vlastnej výskumnej produkcie (bude ich možné importovať vo formáte .csv), ale napríklad aj z dát od ministerstiev, štatistického úradu, DATACUBE.

# **4. Prílohy**

K aplikácii bude dodaná dokumentácia aj používateľský manuál a prezentácia zamýšľaného využitia/postupu užívania aplikácie.

## **4.1 Zoznam používaných datasetov**

Datasety označené \* nebudú zahrnuté do automatického skriptu.

Prehľad všetkých dátových tabuliek dostupných v API prostredníctvom URL odkazu https://data.statistics.sk/api/v2/collection?lang=sk vo formáte JSON-stat.

**Filtre z DataCube Štatistického úradu (na tieto treba pipeline)**

* základná geografická úroveň - okres
* základný časový rad - ročne
* filtre:

Počet obyvateľov podľa pohlavia - SR-oblasť-kraj-okres, m-v (ročne) [om7102rr]

<http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7102rr/v_om7102rr_00_00_00_sk>

\* Živonarodení podľa poradia - SR-oblasť-kraj-okres, m-v [om7029rr]

<http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7029rr/v_om7029rr_00_00_00_sk>

Priemerný vek osoby pri úmrtí podľa pohlavia - SR-oblasť-kraj-okres, m-v [om7038rr]

<http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7038rr/v_om7038rr_00_00_00_sk>

a

Zomrelí podľa veku a pohlavia - SR-oblasť-kraj-okres, m-v [om7034rr]

<http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7034rr/v_om7034rr_00_00_00_sk>

Vekové zloženie - SR-oblasť-kraj-okres, m-v [om7009rr] - .............

<http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7009rr/v_om7009rr_00_00_00_sk>

Prehľad stavu a pohybu obyvateľstva - SR-oblasť-kraj-okres, m-v [om7011rr]

<http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7011rr/v_om7011rr_00_00_00_sk>

Veľkostné skupiny obcí

<http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7023rr/v_om7023rr_00_00_00_sk>

Stredná dĺžka života

<http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7023rr/v_om7023rr_00_00_00_sk>

Hustota obyvateľstva - SR-oblasť-kraj-okres, m-v [om7015rr]

<http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7015rr/v_om7015rr_00_00_00_sk>

Vzdelanie - počty absolventov stredných a vysokých škôl podľa okresu

\* <http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/sv3012rr/v_sv3012rr_00_00_00_sk>

<http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/sv3024rr/v_sv3024rr_00_00_00_sk>

\* Nezamestnanosť podľa rokov podľa okresov

<https://www.upsvr.gov.sk/statistiky/nezamestnanost-mesacne-statistiky.html?page_id=1254>

\* Priemerné platy podľa rokov podľa okresov

<https://slovak.statistics.sk/wps/wcm/connect/910e7175-54da-4742-845e-19faee5b0314/Zamestnanci_a_priemerne_mesacne_mzdy_2018.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-910e7175-54da-4742-845e-19faee5b0314-mClEyAL>

Pracovná neschopnosť - prípady [zd3002rr]

<http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/zd3002rr/v_zd3002rr_00_00_00_sk>

Kriminalita podľa základných skupín trestných činov [sk3003rr]

<http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/sk3003rr/v_sk3003rr_00_00_00_sk>

\* Knižnice - knižničné jednotky a výpožičky [ku3006rr]

<http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/ku3006rr/v_ku3006rr_00_00_00_sk>

\* Repertoár a návštevníci divadiel [ku3002rr]

<http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/ku3002rr/v_ku3002rr_00_00_00_sk>

\* Emisie základných znečisťujúcich látok [zp3003rr]

<http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/zp3003rr/v_zp3003rr_00_00_00_sk>

\* Výsledky volieb do NR SR (ideálne všetky od počiatku SR)

<https://volby.statistics.sk/>

\* Atlas rómskych komunít 2019

<https://www.minv.sk/?atlas-romskych-komunit-2019>

stĺpce:

B - okres (bude treba sčítať/spájať jednotlivé mestá do jednotlivých okresov)

F - podiel rómskych obyvateľov (bude treba sčítať/spájať jednotlivé mestá do jednotlivých okresov)

AQ - výskyt komunitých centier

\* <https://www.europeansocialsurvey.org/data/country.html?c=slovakia>

country file za všetky dostupne roky