

Nový portál otevřených dat Ministerstva financí

Martin Kopeček

Otevřená data ve veřejné správě 2020

23. 10. 2020

Ministerstvo financí
České republiky



https://opendata.mfcr.cz/

[Úvod](#)[Návody](#)[Aplikace a vizualizace](#)[Katalog datových sad](#)

OTEVŘENÁ DATA MINISTERSTVA FINANCÍ

6

organizací

687

datových souborů

9,00

GB dat

[ARES](#)[MONITOR](#)[CEDR](#)[Faktury a smlouvy](#)

Návody a vizualizace

Pohled na stavební a penzijní spoření

Ilustrace použití datových sad Vývoj stavebního spoření a Penzijní připojištění a spoření v programovacím jazyce Python.

Novinky

[Sledovat](#)

Otevřená data MFČR
@otevrenadatamf

Kromě nového portálu otevřených dat pracujeme i na zvýšení technické kvality publikace. Například datovou sadu Pokladní plnění státního rozpočtu jsme



Motivace



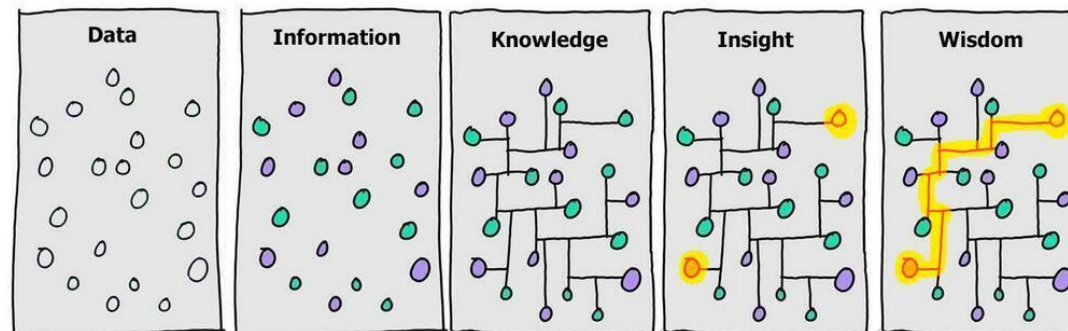
Motivace

1. Otevřená data děláme pro uživatele
2. Udržitelnost



Motivace

1. Otevřená data děláme pro uživatele

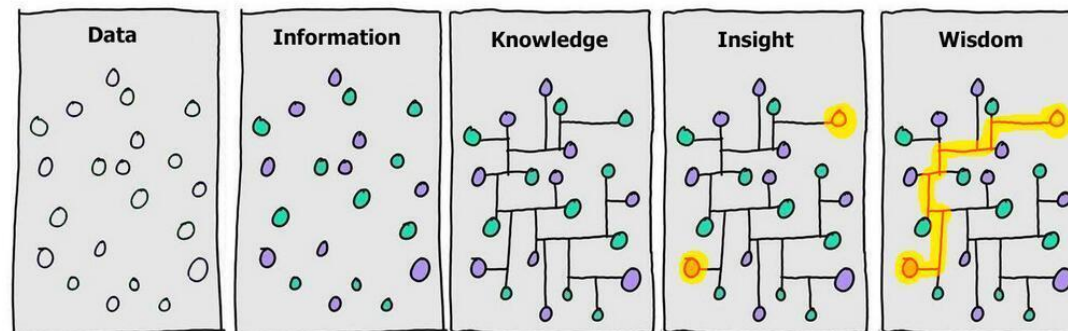


2. Udržitelnost



Motivace

1. Otevřená data děláme pro uživatele



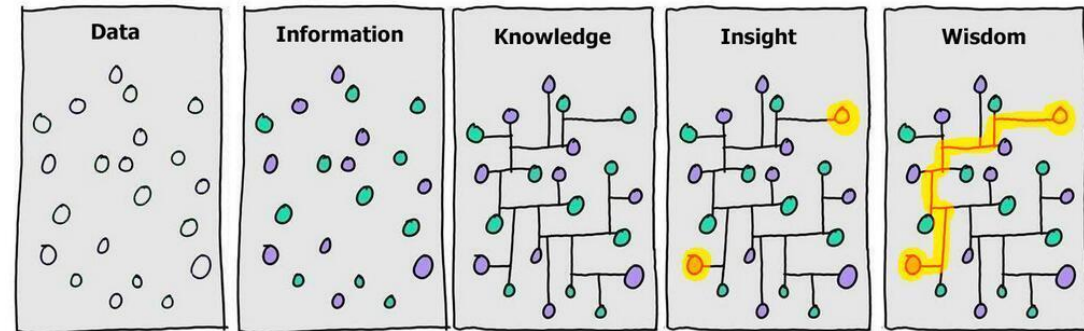
2. Udržitelnost

udržitelný = takový, který umožňuje změny



Motivace

1. Otevřená data děláme pro uživatele



2. Udržitelnost

udržitelný = takový, který umožňuje změny

3. (stav současného řešení)



Co náš portál umí?

(a co Váš portál ne)



Témata

- Tematicky související sady
- Struktury dat
- Popis obsahu dat

Hospodaření Ministerstva financí

Úvod

Datové sady

Přehled smluv MF

Přehled smluv MF - Vazby smluv a faktur

Faktury MF

Faktury ÚZSVM

Aplikace Supervisor

Náklady projektu Otevřená data MF

Úvod

Jednotlivé organizace resortu Ministerstva financí jsou samostatnými účetními jednotkami a mají tedy částečně odlišné vedení faktur a smluv. V současnosti se snažíme o sjednocení exportů otevření dat dle [otevřené formální normy pro zveřejňování faktur](#) Ministerstva vnitra. Tato norma je ale stále ve fázi draftu a tak se exporty budou bohužel ještě pravděpodobně měnit.

Datové sady

Přehled smluv MF

Přehled platných i neplatných smluv, které uzavřelo ministerstvo financí.

Název sloupce	Popis
cislo_smlouvy	evidenční číslo smlouvy přiřazené EKIS MF
typ_smlouvy	S: smlouva D: dodatek
povaha_smlouvy	O: odběratelská, D: dodavatelská
nazev_partnera	
predmet	
popis_smluvniho_typu	interní kategorizace smluv
castka	částka smlouvy
kod_stavu	V: vystavená (platná smlouva) U: ukončená (neplatná smlouva)
platnost_od	
platnost_do	

Datové sady

Smlouvy Ministerstva financí ČR

Tabulkový přehled platných i neplatných financí.

Datová sada CSV JSON

Neznámá datová sada

Nastala chyba při stahování informací o c
zobrazit informace kliknutím na tlačítko r

Datová sada

Faktury Ministerstva financí ČR

Přehled uhrazených faktur Ministerstva f

Datová sada CSV JSON

Přehled faktur ÚZSVM

Soubor představuje přehled uhrazených t
zastupování státu ve věcech majetkových
ÚZSVM je tzv. organizační složkou státu,
zcela ze zdrojů státního rozpočtu. Proto u



Návody a analýzy

- Návody na použití datových sad v různých jazycích
- Odkaz na GitHub, kde lze kód rovnou využít pro vlastní analýzu
- Plánujeme i vlastní analýzy dat

Návod: Pohled na stavební a penzijní spoření

V následujících řádcích se pokusíme ilustrovat použití otevřených dat publikovaných na našem portálu. Konkrétně půjde o datové sady [Vývoj stavebního spoření](#) a [Penzijní připojištění a spoření](#) a jejich zpracování bude prezentováno v programovacím jazyce Python. Mimo to se zde pokusíme poukázat na problémy, které se zpracováním dat souvisejí a které se neomezují pouze na tyto dvě datové sady.

Zabývat se budeme především tím, jak se do dat promítly konkrétní události a jaké trendy můžeme vypořizovat. Také budeme zkoumat, zdali spolu tato data nějak souvisí, tedy jestli se sklon k ukládání úspor formou stavebního spoření vyvíjí v čase podobně, jako je tomu u spoření penzijního.

Stavební spoření

Nahrajeme data přímo z našeho portálu a podíváme se na jejich strukturu:

```
In [1]: import pandas as pd
stavebni=pd.read_csv("https://opendata.mfcr.cz/exports/stavebni_sporeni/csv").sort_values(["rok","ctvrtleti"]).reset_index()
stavebni.dtypes

Out[1]: index                int64
rok                int64
ctvrtleti          int64
nove_smlouvy       int64
prumerna_cilova_castka  float64
smlouvy_sporeni    int64
statni_podpora     int64
prumerna_podpora   float64
nasporena_castka   int64
uvery_celkem       int64
uvery_ss           int64
uvery_preklenovaci int64
uvery_castka_celkem int64
uvery_castka_ss    int64
uvery_castka_preklenovaci int64
```



Katalog datových sad

- Pouze frontend řešení
 - Data přímo ze SPARQL endpointu
 - Možnost provozovat např. na Github Pages
 - Může být napojený i přímo na NKOD
- Nezávislé na organizaci
- Skrytí podřízených datových sad

Katalog datových sad

Téma

Vláda a veřejný sektor (520)
Hospodářství a finance (378)
Hospodaření Ministerstva financí (3)
Dotace (2)
Hospodaření státu (2)
více

Klíčová slova

rozpočet (370)
státní pokladna (369)
dotace (63)
predikce (19)
makroekonomika (14)

Nalezeno 52 datových sad. Podřízené datové sady jsou skryté. [Zobrazit](#)

[ARES - Administrativní registr ekonomických subjektů](#)

Data o ekonomických subjektech registrovaných v České republice. Zprostředková

[Centrální evidence dotací](#)

Centrální evidence dotací ze státního rozpočtu, státních fondů, státních finančních

[Centrální registr administrativních budov](#)

Export z informačního systému CRAB, který obsahuje informace o nemovitostech

[Čistá pozice ČR vůči EU](#)

Data ukazují příjmy, které Česká republika v jednotlivých letech od roku 2004 (o příjmových kapitol (Politika soudržnosti, Zemědělství, Politiky EU – prostřed infrastruktury, atd.). Údaje dále ukazují každoroční odvody České republiky do rozv odvody pohybují mezi 1 a 1,2 % českého hrubého národního důchodu. Data uk rozpočtu EU (objem příjmů z rozpočtu EU převyšuje odvody do evropského rozpoč v Evropské unii.

[Daňová statistika](#)

„Daňová statistika“ zahrnuje roční přehledy ex-post o vývoji inkasa jednotlivých da ekonomických činností (CZ-NACE), počty poplatníků, počty daňových přiznání, p



Katalog datových sad

- Živé ukázky obsahu
- Poslední změny a velikost z HTTP hlavičky
- Plánujeme návody na stažení pro NodeJS, Python, Stata, R

Faktury Ministerstva financí ČR

Přehled uhrazených faktur Ministerstva financí ČR

Ukázka dat

CSV (text/csv)

```
typ_záznamu,iri,dodavatel_ičo,dodavatel_název,typ_dokladu,částka,částka_bez_dph,částka_v_měně,měna,datum_vystavení,dat  
souhrnný,https://opendata.mfcr.cz/lod/faktury/2088800217,28191226,"Economia, a.s.",Z,3630.00,3630.00,3630.00,CZK,2020-  
položkový,https://opendata.mfcr.cz/lod/faktury/2088800217/5167,28191226,"Economia, a.s.",Z,3630.00,3630.00,3630.00,CZK  
souhrnný,https://opendata.mfcr.cz/lod/faktury/2020000362,28402979,MA SI-CO s.r.o.,F,5489.00,4773.00,5489.00,CZK,2020-09  
položkový,https://opendata.mfcr.cz/lod/faktury/2020000362/,28402979,MA SI-CO s.r.o.,F,5489.00,4773.00,5489.00,CZK,2020-  
souhrnný,https://opendata.mfcr.cz/lod/faktury/2020000363,48109975,GOLD GROUP s.r.o.,F,3214.00,2794.00,3214.00,CZK,2020-  
položkový,https://opendata.mfcr.cz/lod/faktury/2020000363/,48109975,GOLD GROUP s.r.o.,F,3214.00,2794.00,3214.00,CZK,20  
souhrnný,https://opendata.mfcr.cz/lod/faktury/2088800216,00063410,
```

JSON (application/json)

```
[  
  {"typ_záznamu": "souhrnný", "iri": "https://opendata.mfcr.cz/lod/faktury/2088800217", "dodavatel_ičo": "28191226", "dodavate  
  {"typ_záznamu": "položkový", "iri": "https://opendata.mfcr.cz/lod/faktury/2088800217/5167", "dodavatel_ičo": "28191226", "do  
  {"typ_záznamu": "souhrnný", "iri": "https://opendata.mfcr.cz/lod/faktury/2020000362", "dodavatel_ičo": "28402979", "dodavate
```

Jak stáhnout data?

CSV

Ke stažení

CSV 37,6 MB

[Stáhnout](#) ⓘ

Poslední změna 4. 10. 2020 4:13:19

JSON 93,1 MB

[Stáhnout](#) ⓘ

Poslední změna 4. 10. 2020 4:13:33



Vyhledávání v portálu, v katalogu i v datech

- Use case: Hledám nějakou informaci, ale nevím v kterých datech je
- Jednoduchý výběr datových sad
- Automatická indexace
 - denně
 - cca 2 mil. záznamů ~ 2 minuty



Podřízené katalogy a SPARQL endpoint

- Větší informační systémy publikují vlastní katalogy
- Tyto katalogy jsou pravidelně centralizovány
 - konektory pro starší systémy
 - přiřazení odkazu na portál
- Publikace přes SPARQL endpoint
- V případě potřeby lze informační systémy napojit přímo na NKOD



Automatické exporty z datového skladu

- Jednotná struktura časových řad díky automatickým exportům z datového skladu.
- Aktualizace každý den.
- Datum změny se změní vždy jen se změnou dat.



Technické zpracování

→ DŮRAZ NA MODULARITU



Technické zpracování

- 13 oddělených služeb
- Postupně se budeme snažit poskytnout zdrojové kódy
- Možnost sdílení některých služeb mezi organizacemi

Služba		Technologie
Vyhledávání	- Automatizace	NodeJS (vlastní řešení)
	- Index	Elastic Search
SPARQL endpoint	- Endpoint	Apache Jena Fuseki2
	- Slučování zdrojů	NodeJS (vlastní řešení)
	- Konektory pro IS	NodeJS (vlastní řešení)
Portál	- Web	Jekyll
	- Administrace	NetlifyCMS
	- Statistiky	NodeJS (vlastní řešení)
Katalog otevřených dat		Angular (vlastní řešení)
Exporty dat z PostgreSQL databáze		NodeJS (vlastní řešení)
API pro vizualizace z datového skladu		PostgREST





Děkuji za pozornost

Twitter: [@otevrenadatamf](#)

E-mail: otevrenadata@mfcz.cz

Ministerstvo financí
České republiky

