



Evropská unie Evropský sociální fond Operační program Zaměstnanost

Kvalitní otevřená data

Jakub Klímek

Projekt:

Implementace strategií v oblasti otevřených dat II

Reg. číslo:

CZ.03.4.74/0.0/0.0/15_025/0004172



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

V současnosti si o českých
otevřených datech málokdo
myslí, že jsou kvalitní.

07:50

83%

← Tweet

Program největší české konference o otevřených datech je zveřejněn. Zahraniční hosté z Estonska a Slovenska, pohled digitálního šampiona v kontextu architektury veřejné správy, vybraní panelisté z řad uživatelů otevřených dat. opendata.gov.cz/edu:konference... #op...

15:09 · 06 Sep 19 · [Twitter Web App](#)

2 Retweets 14 Likes



Ondřej Kokeš @kondrej · 16h

Replying to [@kubanster](#)

"9:35 Kvalitní otevřená data"

Tohle je prosím co? To neznám!



1



1



2



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Otevřená data pohledem úředníka



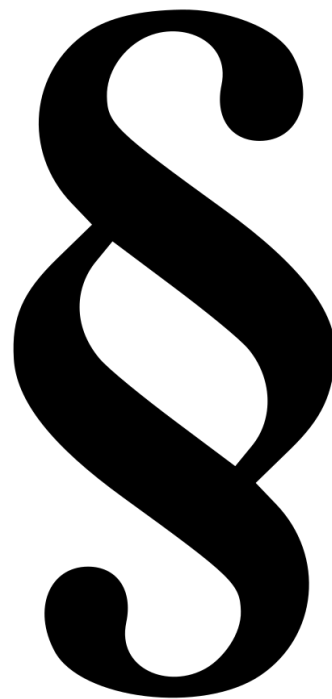


Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Otevřená data pohledem úředníka

Otevřená data jsou...

dle § 3 odst. 11 zákona č. 106/1999 Sb. o svobodném
přístupu k informacím





Otevřená data pohledem úředníka

Otevřená data jsou...

dle § 3 odst. 11 zákona č. 106/1999 Sb. o svobodném
přístupu k informacím

- informace zveřejňované způsobem
umožňujícím dálkový přístup
 - tedy na Webu





Otevřená data pohledem úředníka

Otevřená data jsou...

dle § 3 odst. 11 zákona č. 106/1999 Sb. o svobodném přístupu k informacím

- informace zveřejňované způsobem umožňujícím dálkový přístup
 - tedy na Webu
- v otevřeném formátu
 - tedy RDF, CSV, JSON, XML apod.





Otevřená data pohledem úředníka

Otevřená data jsou...

dle § 3 odst. 11 zákona č. 106/1999 Sb. o svobodném
přístupu k informacím

- informace zveřejňované způsobem umožňujícím dálkový přístup
 - tedy na Webu
- v otevřeném formátu
 - tedy RDF, CSV, JSON, XML apod.
- v strojově čitelném formátu
 - tedy žádné barevné Excely



Město	Změna	čtu
Praha		
Brno		arda
Ostrava		onů
Součet		a 33 milionů
Součet dávám jak		ek. ta čist strojic

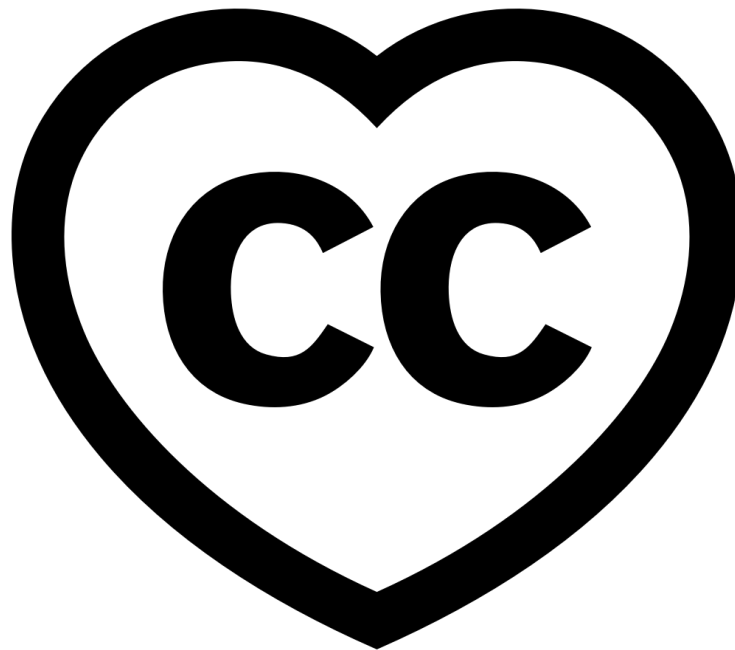


Otevřená data pohledem úředníka

Otevřená data jsou...

dle § 3 odst. 11 zákona č. 106/1999 Sb. o svobodném přístupu k informacím

- informace zveřejňované způsobem umožňujícím dálkový přístup
 - tedy na Webu
- v otevřeném formátu
 - tedy RDF, CSV, JSON, XML apod.
- v strojově čitelném formátu
 - tedy žádné barevné Excely
- jejichž způsob ani účel následného využití není omezen
 - tedy volné užití, přinejhorším CC-BY





Otevřená data pohledem úředníka

Otevřená data jsou...

dle § 3 odst. 11 zákona č. 106/1999 Sb. o svobodném přístupu k informacím

- informace zveřejňované způsobem umožňujícím dálkový přístup
 - tedy na Webu
- v otevřeném formátu
 - tedy RDF, CSV, JSON, XML apod.
- v strojově čitelném formátu
 - tedy žádné barevné Excely
- jejichž způsob ani účel následného využití není omezen
 - tedy volné užití, přinejhorším CC-BY
- evidovány v NKOD

OTEVŘENÁ DATA

Poskytovatelé (36)

Český úřad zeměměřický a katastrální (127 330)

Ministerstvo vnitra (2 589)

Český statistický úřad (585)

HLAVNÍ MĚSTO PRAHA (242)

Statutární město Brno (165)

Statutární město Plzeň (163)

Ministerstvo zemědělství (102)

Zobrazit další

Témata (141)

parkovací plocha (133)

veřejné finance (79)

Novinky Datové sady Poskytovatelé Klíčová slova Další

OTEVŘENÁ DATA

131 844 datových sad nalezeno

1.Základní výkres

Statutární město Plzeň

Územní plán Plzeň - Výkres č. 1 - Základní členění území. Více na <http://www.ukr.plzen.eu/cz/uzemni-planovani/uzemni-plan-plzen/uzemni-plan-plzen.aspx> Více na <http://mapy.plzen.eu/aktuality/uzemni-plan-plzen-2016.aspx>

ZIP

100 letá voda

Statutární město Plzeň

Záplavové plochy n-leté vody na území Plzně.

KMZ ZIP

1od.Koordinální výkres

Statutární město Plzeň



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Datová kvalita
=
Připravenost dat k použití



Kvalitní otevřená data

=

otevřená data, která jsou **navíc**
dobře připravená k použití
uživateli



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

2 typy uživatelů otevřených dat



Uživatel typu “data šťoural” či “korporace”

- ví přesně, co za data máte
 - nepotřebuje katalog



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

2 typy uživatelů otevřených dat



Uživatel typu “data šťoural” či “korporace”

- ví přesně, co za data máte
- ví, co ta data znamenají
 - nepotřebuje dokumentaci



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

2 typy uživatelů otevřených dat



Uživatel typu “data šťoural” či “korporace”

- ví přesně, co za data máte
- ví, co ta data znamenají
- ví, jak by je použil, kdyby se k nim dostal
 - a pravděpodobně se k nim dostane i bez jejich otevření



2 typy uživatelů otevřených dat



Uživatel typu “data šťoural” či “korporace”

- ví přesně, co za data máte
- ví, co ta data znamenají
- ví, jak by je použil, kdyby se k nim dostal
- na kvalitě technického zpracování dat mu až tak nezáleží
 - má schopnosti, prostředky a motivaci pracovat i s nekvalitními daty



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

2 typy uživatelů otevřených dat



Uživatel typu “data šťoural” či “korporace”

- ví přesně, co za data máte
- ví, co ta data znamenají
- ví, jak by je použil, kdyby se k nim dostal
- na kvalitě technického zpracování dat mu až tak nezáleží
- bude mít ale konkrétní připomínky k věcnému obsahu a věcným problémům v datech
 - cenná zpětná vazba



2 typy uživatelů otevřených dat



Uživatel typu “data šťoural” či “korporace”

- ví přesně, co za data máte
- ví, co ta data znamenají
- ví, jak by je použil, kdyby se k nim dostal
- na kvalitě technického zpracování dat mu až tak nezáleží
- bude mít ale konkrétní připomínky k věcnému obsahu a věcným problémům v datech



Uživatel typu “nováček”

- je jich řádově více
- pravděpodobně o nich nevíte
 - to neznamená že neexistují



2 typy uživatelů otevřených dat



Uživatel typu “data šťoural” či “korporace”

- ví přesně, co za data máte
- ví, co ta data znamenají
- ví, jak by je použil, kdyby se k nim dostal
- na kvalitě technického zpracování dat mu až tak nezáleží
- bude mít ale konkrétní připomínky k věcnému obsahu a věcným problémům v datech



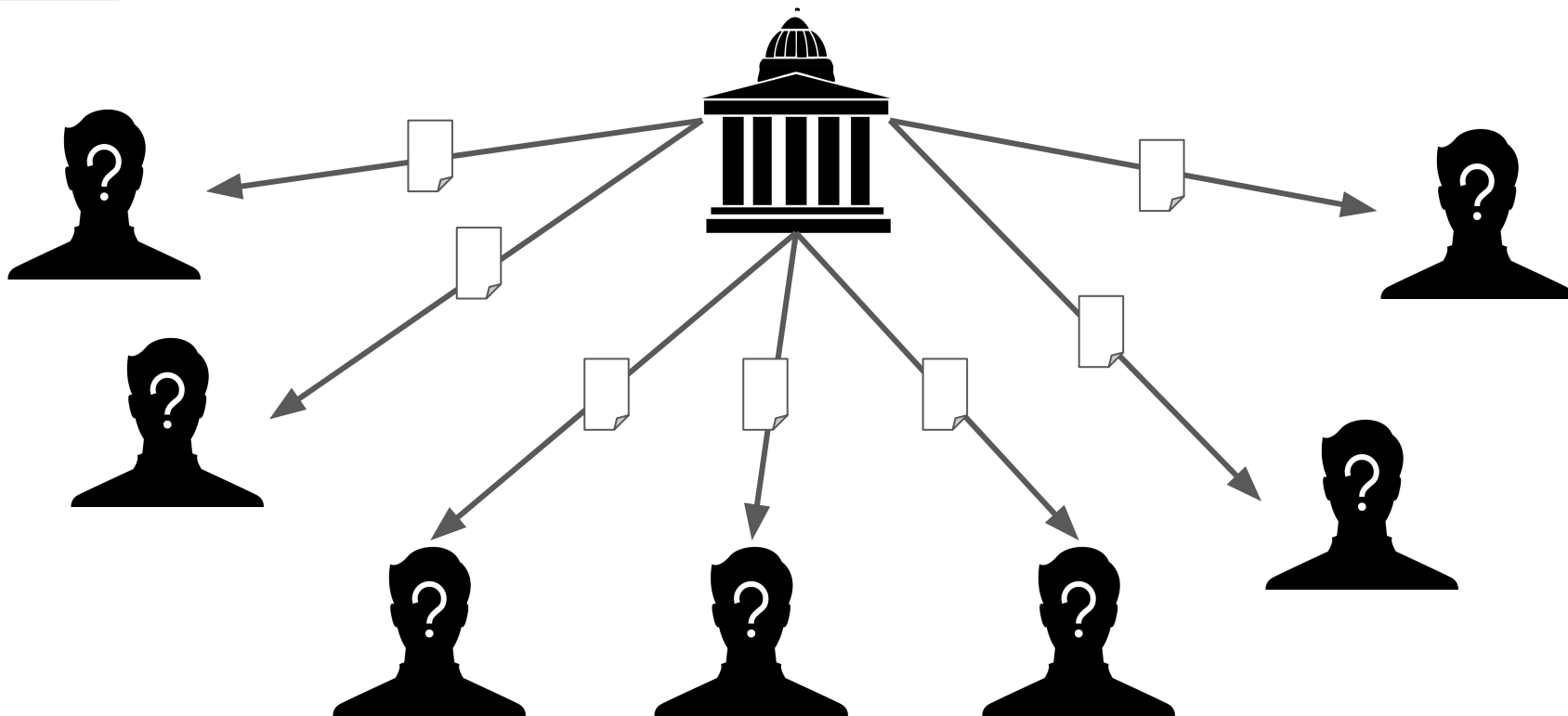
Uživatel typu “nováček”

- je jich řádově více
- pravděpodobně o nich nevíte
- data neznají, je třeba jim je představit a komfortně zprostředkovat
 - jinak půjdou dělat něco jiného a vaše data zůstanou bez užitku



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

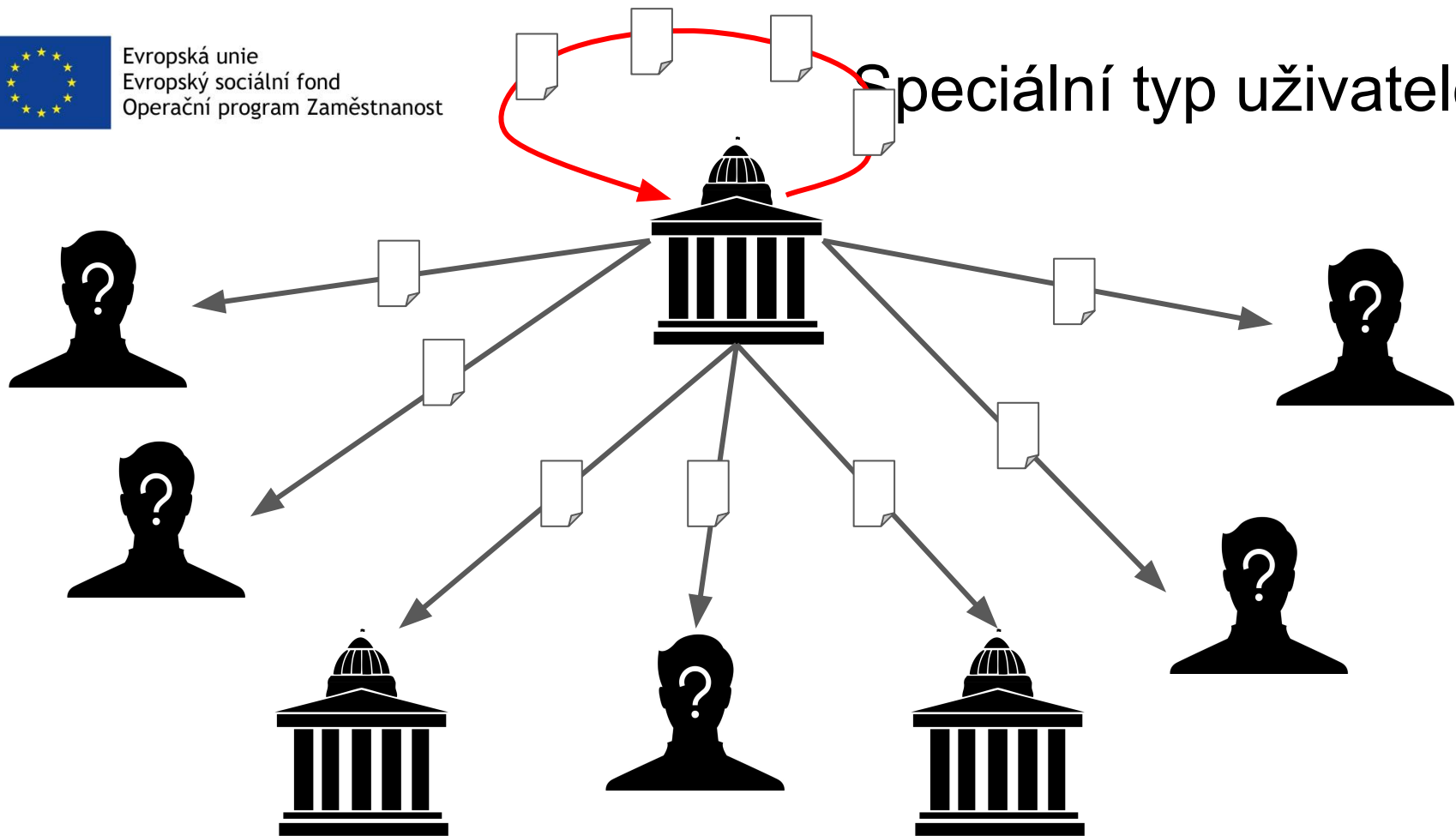
Speciální typ uživatele





Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Speciální typ uživatele





Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Otevřená data pohledem uživatelé “nováčka”



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Otevřená data...

1. Potřebuji data o X.

Otevřená data pohledem uživatele





Otevřená data...

1. Potřebuji data o X
2. Hledám tedy X v NKOD
 - Předpoklad: Poskytovatel data dobře popsal jako X
 - klíčové slovo
 - název
 - popis
 - téma

faktura (28)
smlouva (28)
drážní doprava, městská doprava, tramvaj, trolejbus, metro, lanovka, jízdní řády, CIS JŘ (1)
drážní doprava, železnice, železniční stanice, železniční zastávky, jízdní řády, CIS JŘ (1)

Otevřená data pohledem uživatele

Dálkové migrační koridory

Ministerstvo životního prostředí

Osy migračních koridorů vymezené v místech, která
T.; ANDREAS, M. (eds.) (2010). Migrační koridory pro
s.r.o. 1 s. ISBN 978-80-903787-6-6. Data byla dále ak

CSV

GeoJSON

HTML

KML

ZIP

Data50

Zeměměřický úřad

Data50 je digitální geografický
ČR 1 : 50 000. Data50 jsou posky
tvorena 59 typy geografických d

Esri Shape



Otevřená data...

1. Potřebuji data o X
2. Hledám tedy X v NKOD
3. Mohu data použít?
 - Předpoklady
 - Poskytovatel zvolil správné podmínky užití
 - Podmínkám užití rozumím

Podmínky užití distribuce
Vlastní licence Autorské dílo
Vlastní licence Originální databáze
Vlastní licence Zvláštní právo pořizovatele databáze
Může obsahovat osobní údaje (nespecifikováno) Osobní údaje

Podmínky užití distribuce
Neobsahuje Autorské dílo
Neobsahuje Originální databáze
Neobsahuje Zvláštní právo pořizovatele databáze
Neobsahuje Osobní údaje

Otevřená data pohledem uživatele

Podmínky užití distribuce
Nespecifikovány Autorské dílo
Nespecifikovány Originální databáze
Nespecifikovány Zvláštní právo pořizovatele databáze
Může obsahovat osobní údaje (nespecifikováno) Osobní údaje



Otevřená data...

1. Potřebuji data o X
2. Hledám tedy X v NKOD
3. Mohu data použít?
4. Jsem schopen data stáhnout?

404 : Not Found

404 : Zadaný zdroj nebyl nalezen.

no protocol

403 : Forbidden

Read timed out

400 : null

connect timed out

Otevřená data pohledem uživatele

Poskytovatel	Procento nedostupných distribucí
Ministerstvo zemědělství	95
Ministerstvo životního prostředí	30
Moravskoslezský kraj	25
Kraj Vysočina	25
Statutární město Brno	21
Ministerstvo pro místní rozvoj	19
Liberecký kraj	17
Statutární město Plzeň	9
Statutární město Ostrava	6
Český statistický úřad	6
Ministerstvo financí	5
Statutární město Děčín	3
Ministerstvo zdravotnictví	3
...	...

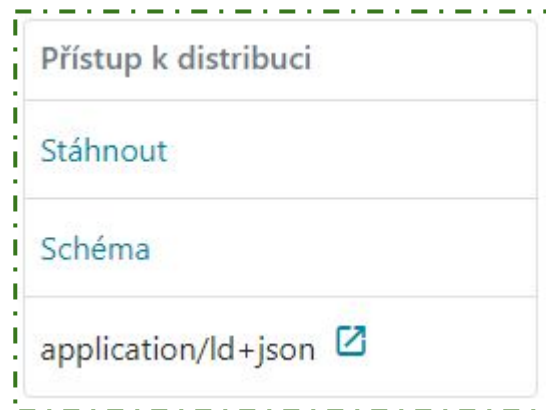
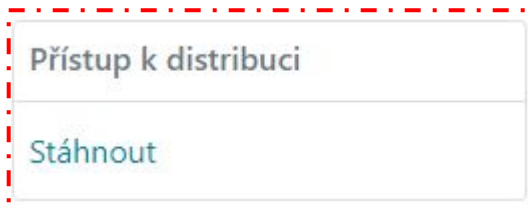
Stav k
2019-09-13



Otevřená data pohledem uživatele

Otevřená data...

1. Potřebuji data o X
2. Hledám tedy X v NKOD
3. Mohu data použít?
4. Jsem schopen data stáhnout?
5. Pochopí můj software, jak jsou data strukturována?





Otevřená data...

1. Potřebuji data o X
2. Hledám tedy X v NKOD
3. Mohu data použít?
4. Jsem schopen data stáhnout?
5. Pochopí můj software, jak jsou data strukturována?
6. Pochopím já, jak s daty pracovat a jak ne?

Otevřená data pohledem uživatele

Dokumentace

Zobrazit dokumentaci

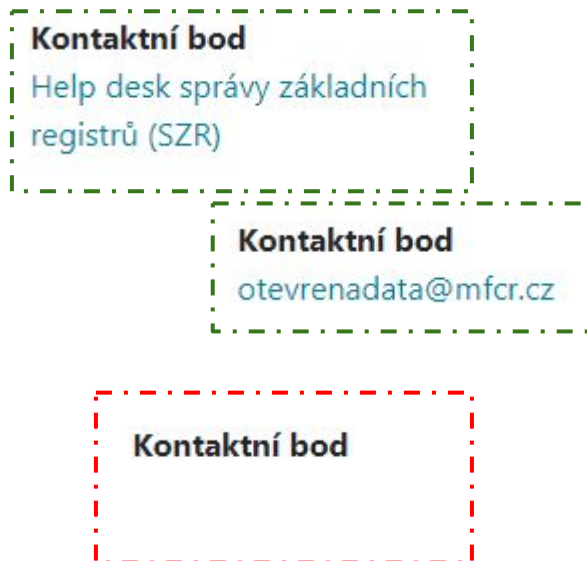
The screenshot shows a web browser window with the URL data.gov.cz/otevrene-formalni-normy/registr-prav-a-povinnosti/agency/2019-04-03/. The page is titled 'Registr práv a povinností - agendy' and is dated 'Doporučení 3. dubna 2019'. It features a 'TABLE OF CONTENTS' section on the left with a list of items including 'JSON struktura', 'Přehled JSON struktury', 'Detailní specifikace prvků JSON struktury', 'Agenda', and various 'Vlastnost' (Properties) entries. On the right, there is a section for 'Tato verze:' with a link to the document, 'Nejnovější verze:' with a link to the latest version, 'Editor:' with the name 'Martin Nečásky (Ministerstvo vnitra)', and 'Zúčastněte se:' with links to 'GitHub', 'opendata-mvcr/otevrene-formalni-normy', 'Nahlaste bug', 'Historie změn', and 'Pull requesty'. At the bottom, there is a section for 'Odkazy:' with a link to 'Portál otevřených dat'.



Otevřená data...

1. Potřebuji data o X
2. Hledám tedy X v NKOD
3. Mohu data použít?
4. Jsem schopen data stáhnout?
5. Pochopí můj software, jak jsou data strukturována?
6. Pochopím já, jak s daty pracovat a jak ne?
7. Víím, na koho se obrátit se zpětnou vazbou?

Otevřená data pohledem uživatele





Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Otevřená data...

8. Víím, jak často jsou data aktualizována?

Otevřená data pohledem uživatele

Periodicita aktualizace

nikdy 

Periodicita aktualizace

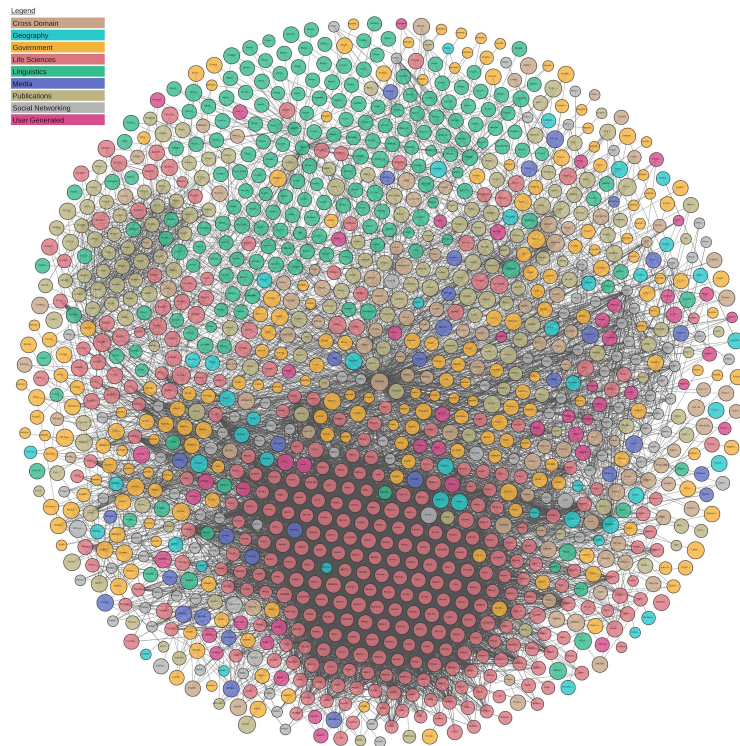
denní 



Otevřená data...

8. Víím, jak často jsou data aktualizována?
9. Víím, s jakými dalšími daty data v datové sadě souvisí?

Otevřená data pohledem uživatele

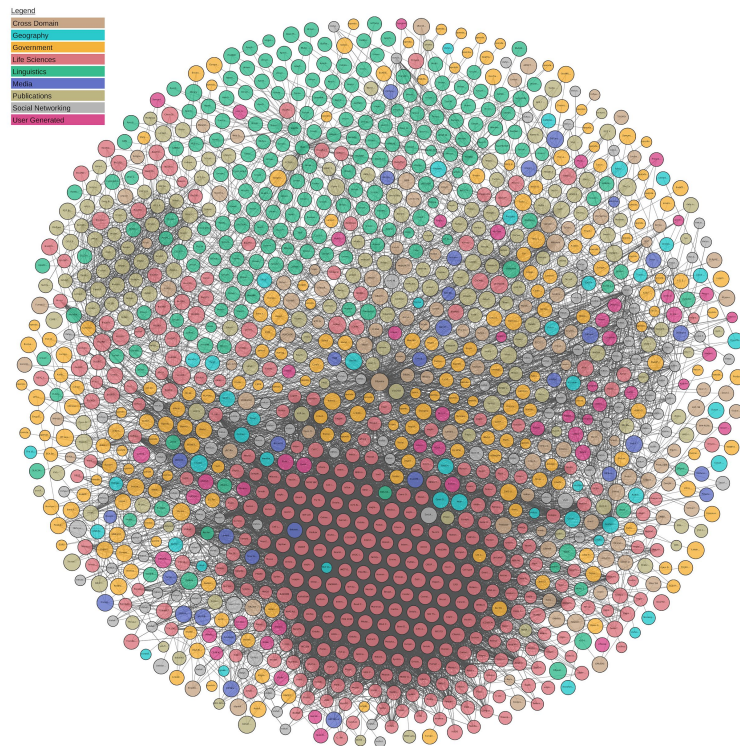




Otevřená data pohledem uživatele

Otevřená data...

8. Víím, jak často jsou data aktualizována?
9. Víím, s jakými dalšími daty data v datové sadě souvisí?
10. A pochopí to i můj software?





Otevřená data jsou...

dle § 3 odst. 11 zákona č. 106/1999 Sb. o svobodném přístupu k informacím

- informace zveřejňované způsobem umožňujícím dálkový přístup
 - tedy na Webu
- v otevřeném formátu
 - tedy RDF, CSV, JSON, XML apod.
- v strojově čitelném formátu
 - tedy žádné barevné Excely
- jejichž způsob ani účel následného využití není omezen
 - tedy volné užití, přinejhorším CC-BY
- evidovány v NKOD

Kvalitní otevřená data

Kvalitní otevřená data jsou...

- dobře popsaná metadaty v NKOD
 - tedy tak, jak by to **uživatel** očekával
- mají jednoduché podmínky užití
 - tedy ty, které jsou v NKOD připraveny
- jsou dostupná
 - tedy žádné prošlé certifikáty a 404ky
- mají schéma a jsou validní
 - tedy je lze zpracovat
- mají dokumentaci
 - tedy je lze pochopit
- mají kontaktní osobu
 - tedy vím, komu hlásit problémy
- obsahují vazby na související data
 - tedy jsou propojená



Kvalita dat obecně - širší téma

- **Přístupnost**

- dostupnost, podmínky užití, propojenost
- bezpečnost, výkon

- Přesnost - syntaktická, sémantická

- **Validita**

- Úplnost
- Stručnost
- Konzistence
- ...

- Relevance

- Důvěryhodnost

- **Pochopitelnost**

- **Aktuálnost**

- **Interpretovatelnost**

- **Interoperabilita**

- Verzátilita

- ...



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Otevřená data teď nejsou příliš kvalitní.
Proč?



Kvalita OD v ČR není dobrá. Proč?

V prostředí otevřených dat jsou protichůdné zájmy.

Uživatelé dat

1. co nejjednodušeji data získat



Neosvícení poskytovatelé

1. data nepublikovat nejlépe vůbec



Kvalita OD v ČR není dobrá. Proč?

V prostředí otevřených dat jsou protichůdné zájmy.

Uživatelé dat

1. co nejjednodušeji data získat
2. co nejjednodušeji je zpracovat
 - dokumentace? validita?

Neosvícení poskytovatelé

1. data nepublikovat nejlépe vůbec
2. když už, tak s minimálním úsilím
 - dokumentace? validita?



Kvalita OD v ČR není dobrá. Proč?

V prostředí otevřených dat jsou protichůdné zájmy.

Uživatelé dat

1. co nejjednodušeji data získat
2. co nejjednodušeji je zpracovat
 - dokumentace? validita?
3. moci se na data spolehnout

Neosvícení poskytovatelé

1. data nepublikovat nejlépe vůbec
2. když už, tak s minimálním úsilím
 - dokumentace? validita?
3. mít za data minimální odpovědnost



Kvalita OD v ČR není dobrá. Proč?

V prostředí otevřených dat jsou protichůdné zájmy.

Uživatelé dat

1. co nejjednodušeji data získat
2. co nejjednodušeji je zpracovat
 - dokumentace? validita?
3. moci se na data spolehnout
4. moci data snadno kombinovat

Neosvícení poskytovatelé

1. data nepublikovat nejlépe vůbec
2. když už, tak s minimálním úsilím
 - dokumentace? validita?
3. mít za data minimální odpovědnost
4. neřešit, jak to dělají ostatní



Kvalita OD v ČR není dobrá. Proč?

V prostředí otevřených dat jsou protichůdné zájmy.

Uživatelé dat

1. co nejjednodušeji data získat
2. co nejjednodušeji je zpracovat
 - dokumentace? validita?
3. moci se na data spolehnout
4. moci data snadno kombinovat
- 5. výsledek patřičně prodat**
 - twitter, appka, hackathon, soutěž

Neosvícení poskytovatelé

1. data nepublikovat nejlépe vůbec
2. když už, tak s minimálním úsilím
 - dokumentace? validita?
3. mít za data minimální odpovědnost
4. neřešit, jak to dělají ostatní
- 5. výsledek patřičně prodat**
 - twitter, appka, hackathon, soutěž



Kvalita OD v ČR není dobrá. Proč?

V prostředí otevřených dat jsou protichůdné zájmy.

Uživatelé dat

1. co nejjednodušeji data získat
2. co nejjednodušeji je zpracovat
 - dokumentace? validita?
3. moci se na data spolehnout
4. moci data snadno kombinovat
- 5. výsledek patřičně prodat**
 - twitter, appka, hackathon, soutěž

Neosvícení poskytovatelé

1. data nepublikovat nejlépe vůbec
2. když už, tak s minimálním úsilím
 - dokumentace? validita?
3. mít za data minimální odpovědnost
4. neřešit, jak to dělají ostatní
- 5. výsledek patřičně prodat**
 - twitter, appka, hackathon, soutěž

Dogfooding otevřených dat - poskytovatelé se musí stát uživateli



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Kvalita OD v ČR není dobrá. Co s tím?



Kvalita OD v ČR není dobrá. Co s tím?



Nedělejte stejné chyby jako řada poskytovatelů před vámi. Před registrací datové sady se ujistěte, že neděláte nějaké z [nejčastějších chyb v otevřených datech](#)!



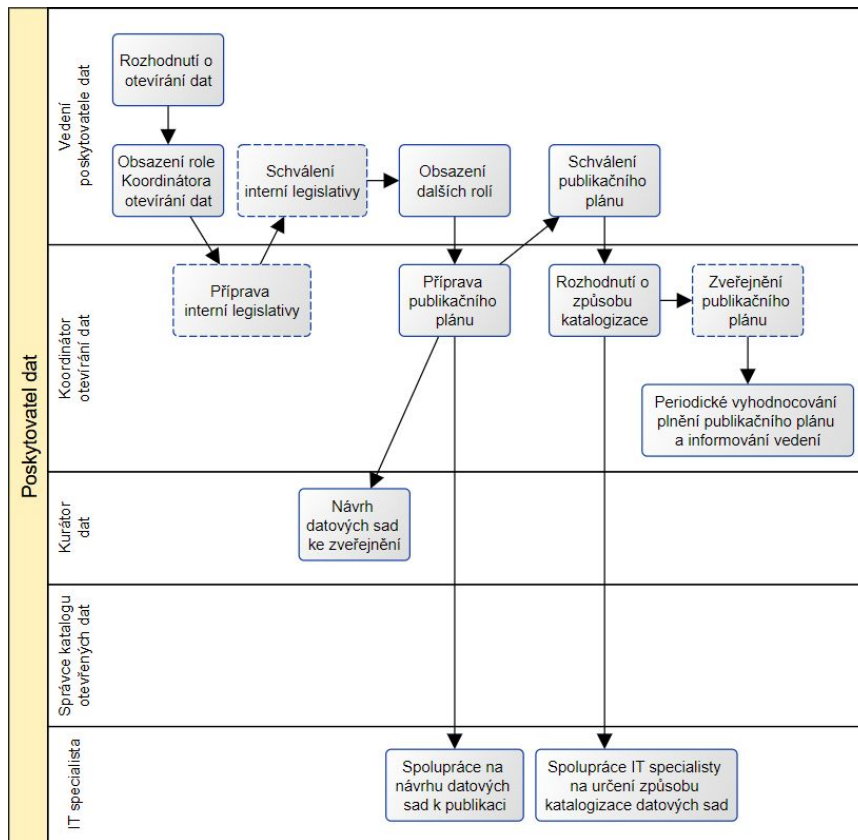
Kategorie špatné praxe	Počet opakujících se chyb
Právní překážky	6
Problémy s dokumentací a metadaty	4
Technické překážky	6
Datové formáty - CSV	16
Datové formáty - XML	1
Obsahové chyby	6



Kvalita OD v ČR není dobrá. Co s tím?

Předpoklady pro vznik kvalitních otevřených dat

- Kvalitní procesy práce s daty obecně
 - zakotvené v interní legislativě
- Data oddělená od SW
 - navržená a zdokumentovaná
- CDO - Chief Data Officer
 - a jeho tým - kurátoři dat
- Kvalitně zadávané a akceptované veřejné zakázky na SW
 - s datovou specifikací





Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Děkuji za pozornost

<https://opendata.gov.cz/špatná-praxe:start>

Projekt:

Implementace strategií v oblasti otevřených dat II

Reg. číslo:

CZ.03.4.74/0.0/0.0/15_025/0004172

Kvalitní otevřená data

Jakub Klímek