



C2V9

Kontrola kvality otevřených dat obsahující analytické reporty jednotlivých kontrol kvality otevřených dat

Vytvořeno v rámci projektu

Rozvoj datových politik v oblasti zlepšování kvality
a interoperability dat veřejné správy
CZ.03.4.74/0.0/0.0/15_025/0013983

Klíčová aktivita: 03 Návrhy a realizace opatření pro zlepšování kvality
otevřených dat

Indikátor: 8 05 00 Počet napsaných a zveřejněných analytických a strategických
dokumentů (vč. evaluačních)

Verze výstupu: 01



Automatizovatelná pravidla kontroly kvality datových sad

Automatizovatelná pravidla kontroly kvality datových sad se řídí Otevřenou formální normou Rozhraní katalogů otevřených dat: DCAT-AP-CZ¹, kterou implementuje Národní katalog otevřených dat². Dělí se na dvě základní skupiny měření. První skupina počítá indikátory kvality nad metadaty registrovanými v NKOD. Druhá skupina pak kontroluje dostupnost zdrojů registrovaných v rámci metadatového záznamu na Webu, a také dostupnost techniky CORS na těchto zdrojích. Technika CORS³ zajišťuje, že s daty publikovanými na webu lze pracovat v aplikacích běžících ve webových prohlížečích na jiných doménách, než na kterých jsou publikována data. To je v případě otevřených dat žádoucí.

Měření probíhají zpravidla denně, a jejich výsledky jsou zobrazovány na webu <https://opendata.gov.cz/statistika:start>.

Měření nad metadatovými záznamy

Nad metadatovými záznamy z NKOD se měří následující indikátory:

Převzaté indikátory

Z minulého projektu Implementace strategií v oblasti otevřených dat II č.

CZ.03.4.74/0.0/0.0/15_025/0004172 byly převzaty následující indikátory kvality metadat v NKOD a způsoby jejich měření:

- Q1: Počet distribucí bez specifikovaných podmínek užití dle poskytovatele
- Q2: Počet datových sad, jejichž distribuce nemají specifikovány podmínky užití dle poskytovatele
- Q3: Počet záznamů datových sad nesplňujících povinné atributy dle poskytovatele
- Q4: Formáty dat distribucí dle poskytovatele
- Q5: Podmínky užití distribucí dle poskytovatele
- Q5b: Počty poskytovatelů dle podmínek užití distribucí
- Q6: Počty datových sad s danou periodicitou aktualizace dle poskytovatele
- Q7a: Počet datových sad a distribucí dle poskytovatele
- Q7b: Počty poskytovatelů datových sad s daným klíčovým slovem
- Q8: Nesprávné Media typy dle poskytovatele

Aktualizované indikátory

V souvislosti s výstupem C2V2, ve kterém navrhujeme úpravy NKOD pro kompatibilitu, a na základě sběru požadavků uživatelů NKOD došlo ke změně následujících indikátorů

- Q3 Počet záznamů datových sad nesplňujících povinné atributy dle poskytovatele
 - upravuje se tak, aby zahrnoval datové služby
- Q4 => Q4a Formáty dat distribucí dle poskytovatele

¹ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/>

² <https://data.gov.cz>

³ <https://developer.mozilla.org/en-US/docs/Web/HTTP/CORS>

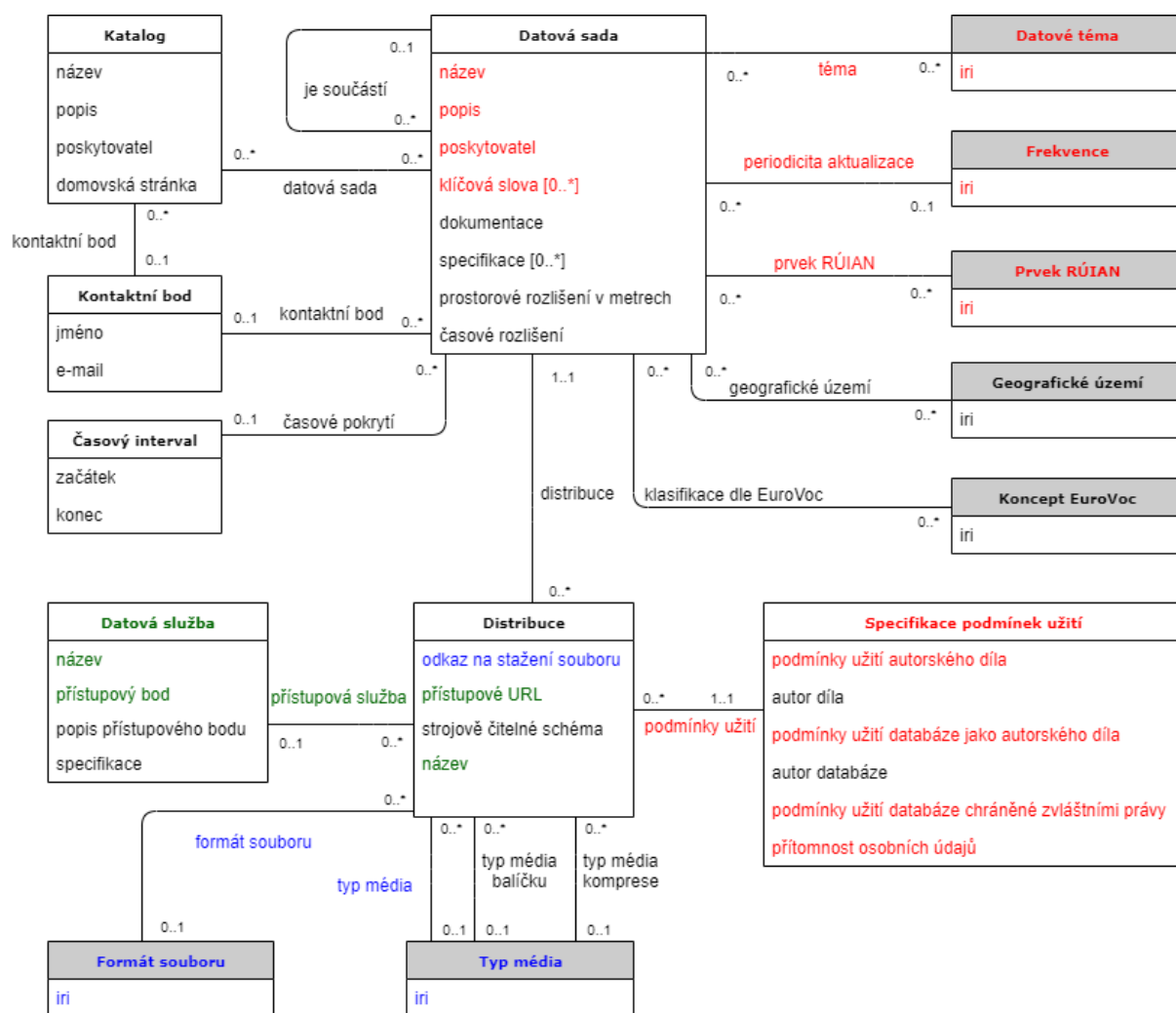


- omezuje se na distribuce v podobě souborů ke stažení
- Q8: Nesprávné Media typy dle poskytovatele
 - rozšířeno o podporu media typu komprese a balíčkování

Nové indikátory

V souvislosti s výstupem C2V2, ve kterém navrhujeme úpravy NKOD pro kompatibilitu, a na základě sběru požadavků uživatelů NKOD došlo k rozšíření měření kvality metadat a dostupnosti zdrojů odkazovaných z metadatového záznamu NKOD a přidání výsledného indikátoru:

1. Q4b Specifikace datových služeb dle poskytovatele

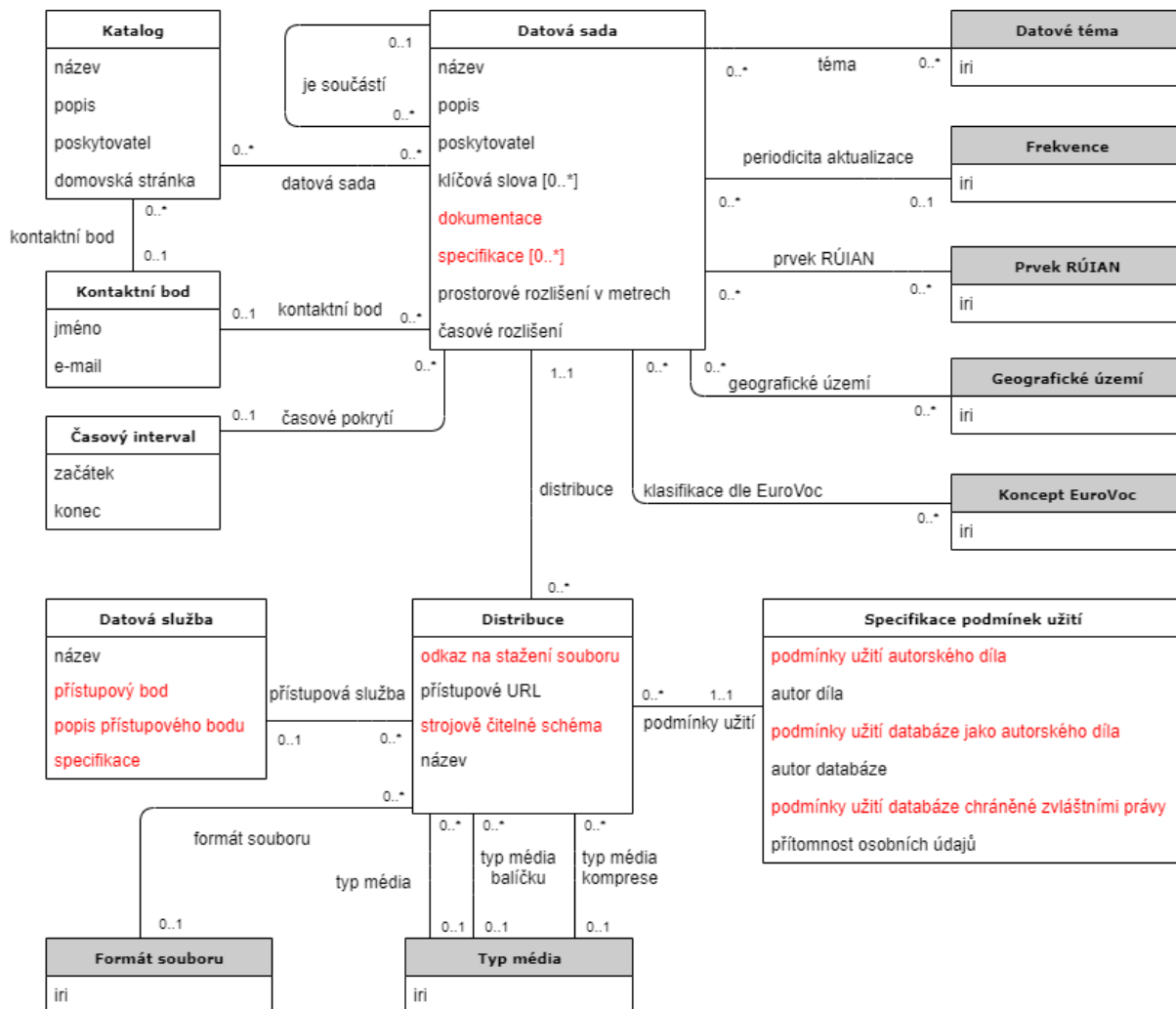


Povinné atributy metadatového záznamu dle OFN Rozhraní katalogů otevřených dat: DCAT-AP-CZ. Červeně jsou povinné atributy pro všechny datové sady. Zeleně jsou atributy povinné pro datové služby a modře povinné atributy pro soubory ke stažení.



Dostupnost registrovaných zdrojů identifikovaných URL

Položky metadatového záznamu dle OFN Rozhraní katalogů otevřených dat: DCAT-AP-CZ, které obsahují URL odkazy na zdroje na Webu jsou kontrolovány na jejich dostupnost a na dostupnost techniky CORS.



Červeně jsou vyznačeny položky metadatového záznamu, které obsahují URL a jsou tak kontrolovány na dostupnost a dostupnost techniky CORS

Převzaté indikátory

Z minulého projektu Implementace strategií v oblasti otevřených dat II č.

CZ.03.4.74/0.0/0.0/15_025/0004172 byly převzaty následující indikátory kvality metadat v NKOD a způsoby jejich měření:

- A1.1: Nedostupnost distribucí datových sad
- A1.2: Nedostupné distribuce datových sad
- A2.1: Nedostupnost schémat distribucí datových sad
- A2.2: Nedostupná schémata distribucí datových sad
- A3.1: Nedostupnost podmínek užití distribucí datových sad
- A3.2: Nedostupné podmínky užití distribucí datových sad



- A4.1: Nedostupnost uživatelské dokumentace datových sad
- A4.2: Nedostupné uživatelské dokumentace datových sad
- A5.1: Neshoda mezi formátem distribuce v NKOD a formátem indikovaným serverem

Nové indikátory

V souvislosti s výstupem C2V2, ve kterém navrhujeme úpravy NKOD pro kompatibilitu, a na základě sběru požadavků uživatelů NKOD došlo k rozšíření měření kvality metadat a dostupnosti zdrojů odkazovaných z metadatového záznamu NKOD a přidání výsledných indikátorů:

1. A1.3 Nedostupnost techniky CORS u distribucí ve formě souboru ke stažení
2. A1.4 Distribuce ve formě souboru ke stažení s nedostupnou technikou CORS u souboru ke stažení
3. A2.3 Nedostupnost techniky CORS u schémat distribucí ve formě souboru ke stažení
4. A2.4 Distribuce ve formě souboru ke stažení s nedostupnou technikou CORS u schématu souboru ke stažení
5. A3.3.1 Nedostupnost techniky CORS u podmínek užití - autorské dílo
6. A3.3.2 Nedostupnost techniky CORS u podmínek užití - autorskoprávně chráněná databáze
7. A3.3.3 Nedostupnost techniky CORS u podmínek užití - databáze chráněná zvláštním právem pořizovatele databáze
8. A3.4.1 Podmínky užití s nedostupnou technikou CORS - autorské dílo
9. A3.4.2 Podmínky užití s nedostupnou technikou CORS - autorskoprávně chráněná databáze
10. A3.4.3 Podmínky užití s nedostupnou technikou CORS - databáze chráněná zvláštním právem pořizovatele databáze
11. A4.3 Nedostupnost techniky CORS u dokumentací datových sad
12. A4.4 Dokumentace datových sad s nedostupnou technikou CORS
13. A6.1 Nedostupnost přístupových bodů distribucí ve formě datové služby
14. A6.2 Nedostupné přístupové body distribucí ve formě datové služby
15. A6.3 Nedostupnost techniky CORS u přístupových bodů distribucí ve formě datové služby
16. A6.4 Přístupové body distribucí ve formě datové služby s nedostupnou technikou CORS
17. A7.1 Nedostupnost popisů přístupových bodů distribucí ve formě datové služby
18. A7.2 Nedostupné popisy přístupových bodů distribucí ve formě datové služby
19. A7.3 Nedostupnost techniky CORS u popisů přístupových bodů distribucí ve formě datové služby
20. A7.4 Popisy přístupových bodů distribucí ve formě datové služby s nedostupnou technikou CORS
21. A8.1 Nedostupnost specifikací datových služeb
22. A8.2 Nedostupné specifikace datových služeb
23. A8.3 Nedostupnost techniky CORS u specifikací datových služeb
24. A8.4 Specifikace datových služeb s nedostupnou technikou CORS
25. A9.1 Nedostupnost specifikací datových sad
26. A9.2 Nedostupné specifikace datových sad
27. A9.3 Nedostupnost techniky CORS u specifikací datových sad



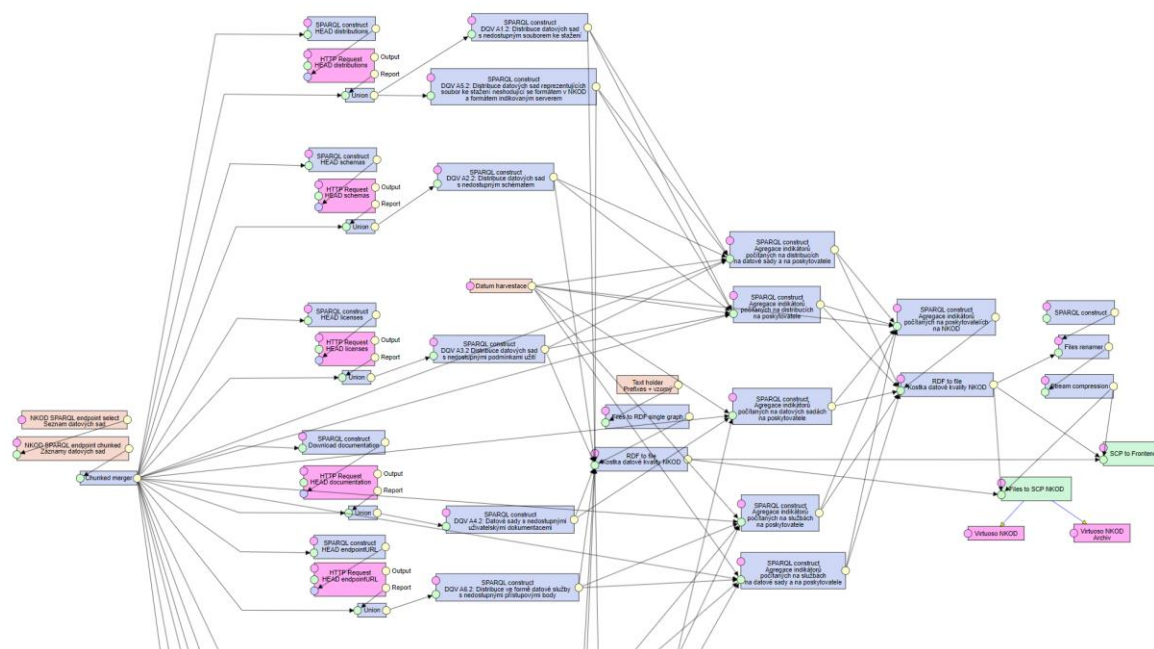
28. A9.4 Specifikace datových sad s nedostupnou technikou CORS

Měření jednotlivých indikátorů

Měření jednotlivých indikátorů jsou navrženy jako datové transformační procesy v nástroji LinkedPipes ETL⁴, který je použit i pro samotný Národní katalog otevřených dat. Jedná se o 3 procesy, které jsou popsány v této sekci a jsou k dispozici v příloze.

Měření dostupnosti zdrojů registrovaných v metadatových záznamech datových sad v NKOD

Tento proces přistupuje k metadatům NKOD a pak pro každý registrovaný zdroj kontroluje jeho dostupnost pomocí metody HTTP HEAD s timeoutem 3 vteřiny. Výsledky reprezentuje jako data v RDF pomocí Data Quality Vocabulary⁵.



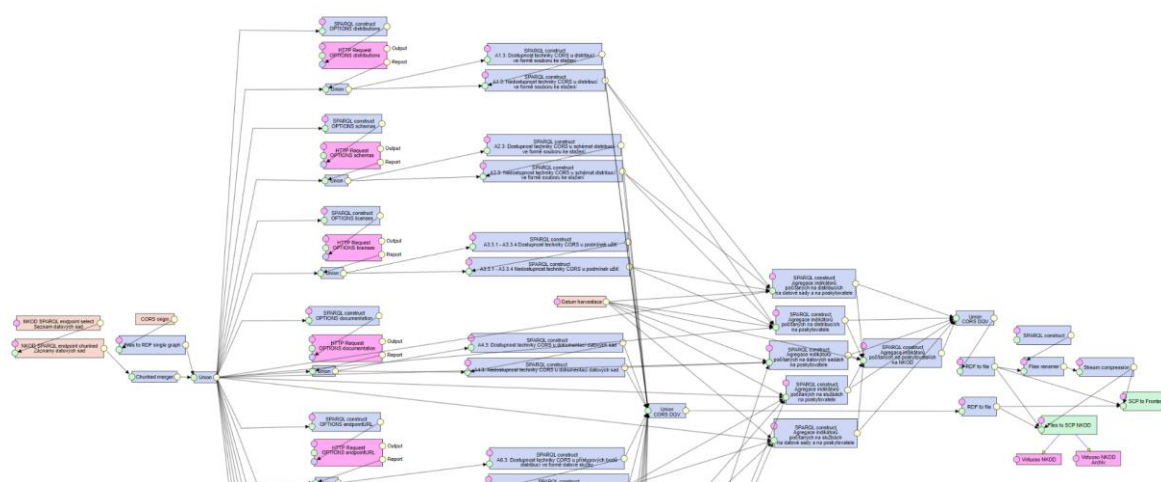
⁴ <https://etl.linkedpipes.com>

⁵ <https://www.w3.org/TR/vocab-dqv/>



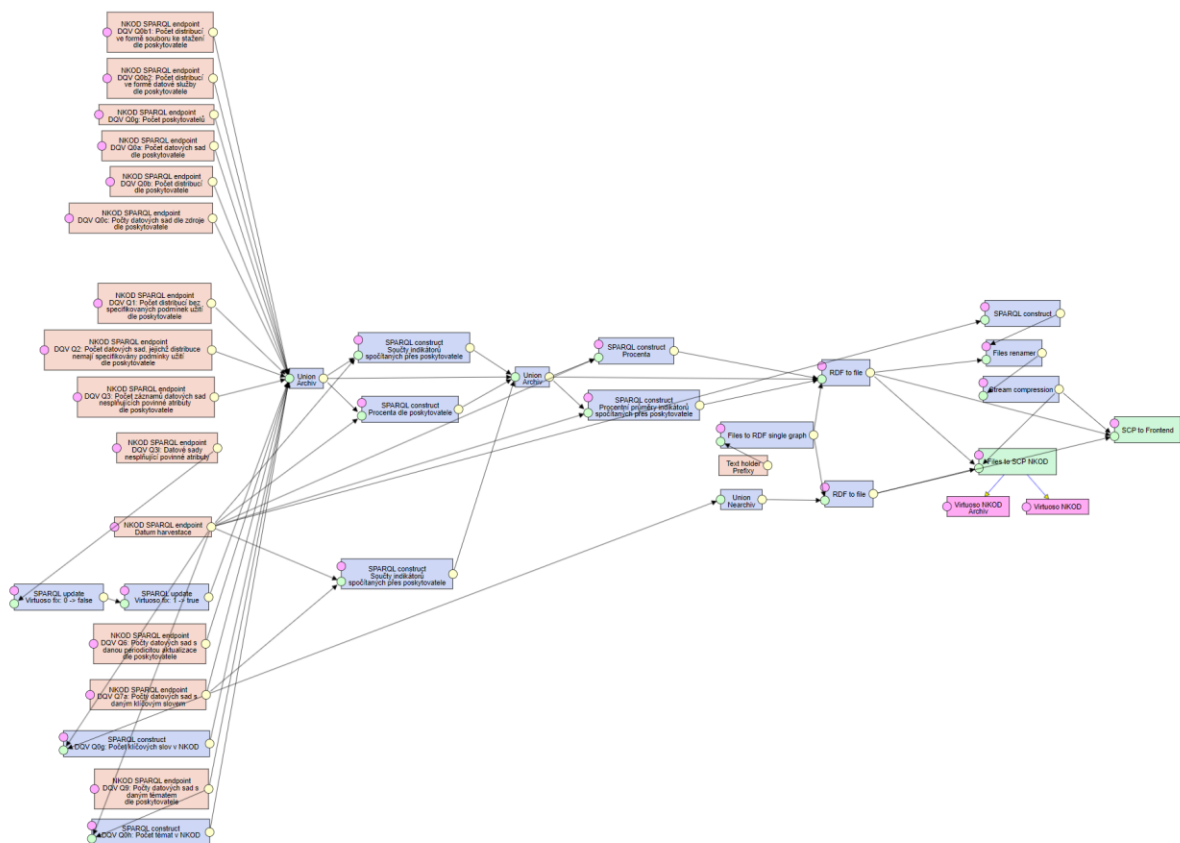
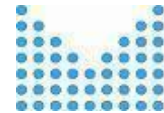
Měření dostupnosti techniky CORS u zdrojů registrovaných v metadatových záznamech datových sad v NKOD

Tento proces přistupuje k metadatům NKOD a pak pro každý registrovaný zdroj kontroluje dostupnost techniky CORS pomocí metody HTTP OPTIONS s timeoutem 3 vteřiny. Výsledky reprezentuje jako data v RDF pomocí Data Quality Vocabulary.



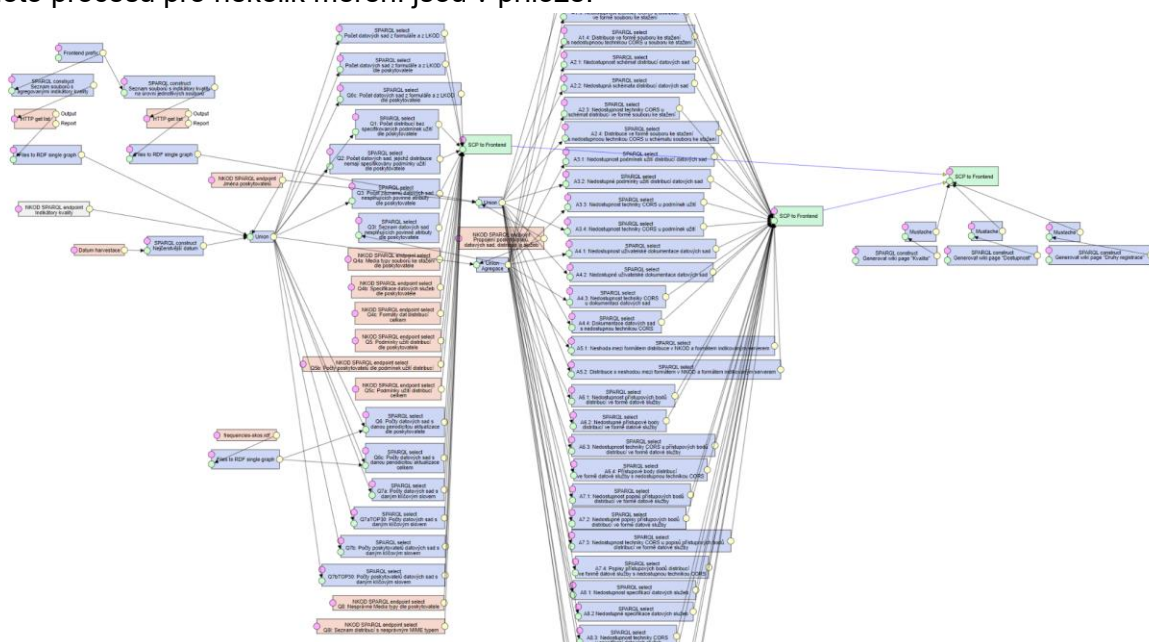
Měření kvality metadatových záznamů v NKOD

Tento proces přistupuje k metadatům NKOD a pak pro každou registrovanou datovou sadu kontroluje, zda její metadata obsahují povinné položky a počítá statistické přehledy Q*. Výsledky reprezentuje jako data v RDF pomocí Data Quality Vocabulary.



Generování CSV výstupů jednotlivých indikátorů kvality datových sad

Tento proces načítá výsledky předchozích tří procesů a generuje jejich CSV reprezentaci, která je následně použita i pro prezentaci v podobě webových stránek v HTML. Výsledky tohoto procesu pro několik měření jsou v příloze.





Výsledky měření indikátorů

V přílohách jsou k dispozici výsledky měření stanovených indikátorů spočítaných procesy popsány výše. Ty probíhají pravidelně, z pravidla jednou denně, po aktualizaci Národního katalogu otevřených dat. Měření dostupnosti, dostupnosti techniky CORS a měření kvality metadatových záznamů jsou organizována do CSV souborů obsahujících statistiky o zjištěných problémech a seznamy zjištěných problémů. CSV soubory jsou pojmenovány po identifikátorech jednotlivých indikátorů, např. **a1-1-nedostupnost.csv**, **q4.csv** apod. Kromě naměřených dat daná příloha obsahuje soubory **a.pdf** a **q.pdf**, obsahující výtisky webových stránek prezentujících výsledky měření v HTML podobě.

Přílohy

V přílohách jsou zdrojové kódy použitého open-source software vyjma RDF úložiště a 2 prototypy LKOD podle 2 typů rozhraní DCAT-AP v Otevřené formální normě Rozhraní katalogů otevřených dat: DCAT-AP-CZ.

Použitý open-source software

1. LinkedPipes ETL (fork v <https://github.com/opendata-mvcr/etl>)
 - o soubor **etl-develop.zip**

Transformační procesy pro LinkedPipes ETL implementující měření

1. Statistika dostupnosti distribucí, schémat, podmínek užití a dokumentace
 - o počítá indikátory A* týkající se dostupnosti
 - o soubor **09 Statistika dostupnosti distribucí, schémat, podmínek užití a dokumentace - HEAD.jsonld**
2. Statistika dostupnosti CORS distribucí, schémat, podmínek užití a dokumentace
 - o počítá indikátory A* týkající se dostupnosti techniky CORS
 - o soubor **10 Statistika dostupnosti distribucí, schémat, podmínek užití a dokumentace - CORS.jsonld**
3. Kvalita metadatových záznamů v NKOD
 - o počítá indikátory Q*
 - o soubor **11 Kvalita metadatových záznamů v NKOD DQV.jsonld**
4. Generování reportů v CSV
 - o generuje výsledné CSV soubory na základě výsledků minulých procesů
 - o soubor **13 Generování reportů v CSV.jsonld**

Výsledky měření

1. 25.10.2021
 - o soubor **2021-10-25.zip**



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost



MINISTERSTVO VNITRA
ČESKÉ REPUBLIKY

2. 9.11.2021
 - soubor 2021-11-09.zip
3. 22.11.2021
 - soubor 2021-11-22.zip